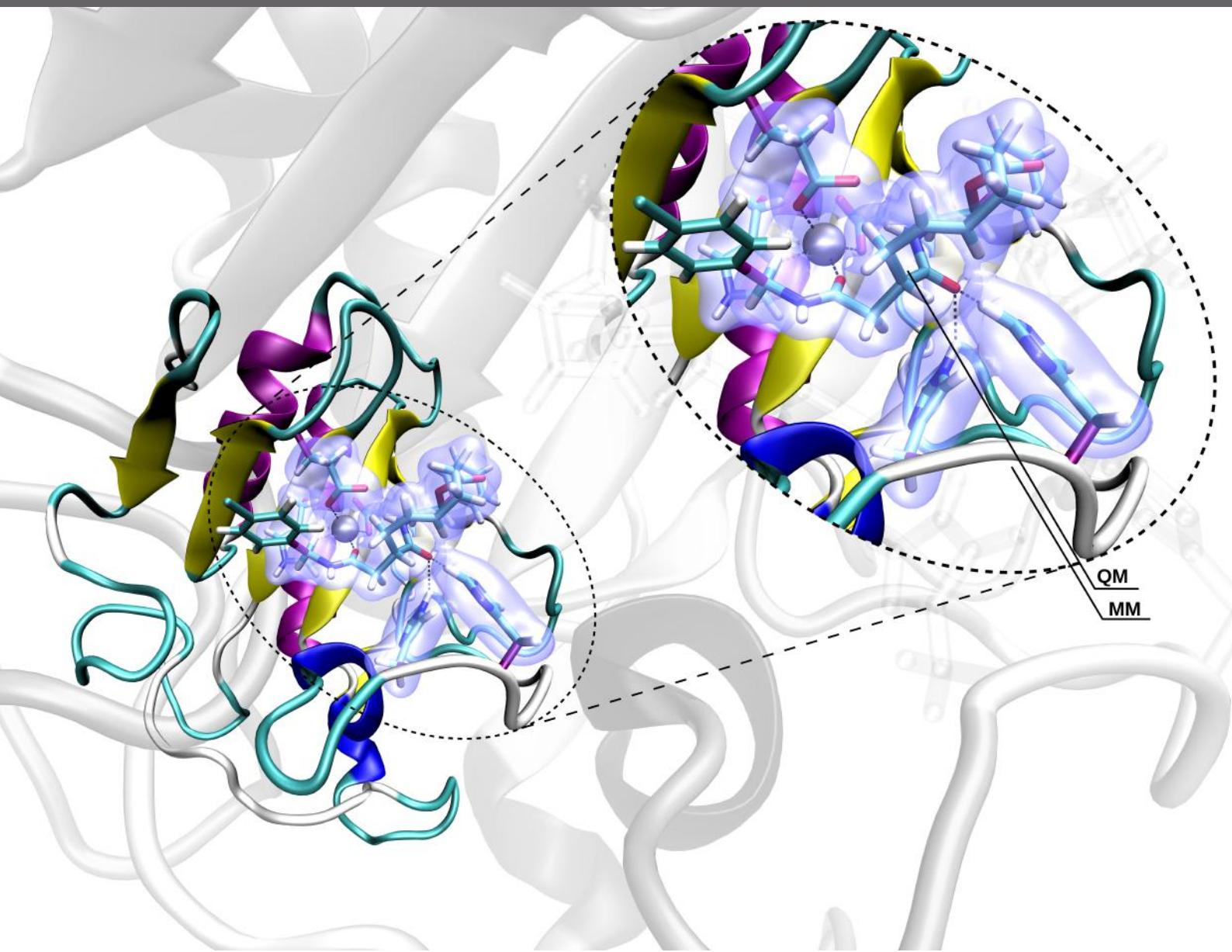
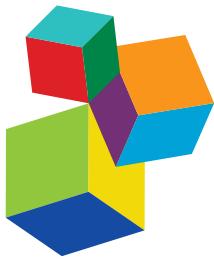


# QUANTUM MECHANICAL/MOLECULAR MECHANICAL APPROACHES FOR THE INVESTIGATION OF CHEMICAL SYSTEMS – RECENT DEVELOPMENTS AND ADVANCED APPLICATIONS

EDITED BY: Thomas S. Hofer and Sam P. De Visser  
PUBLISHED IN: *Frontiers in Chemistry*





#### Frontiers Copyright Statement

© Copyright 2007-2018 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714  
ISBN 978-2-88945-626-0  
DOI 10.3389/978-2-88945-626-0

#### About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

#### Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

#### Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

#### What are Frontiers Research Topics?

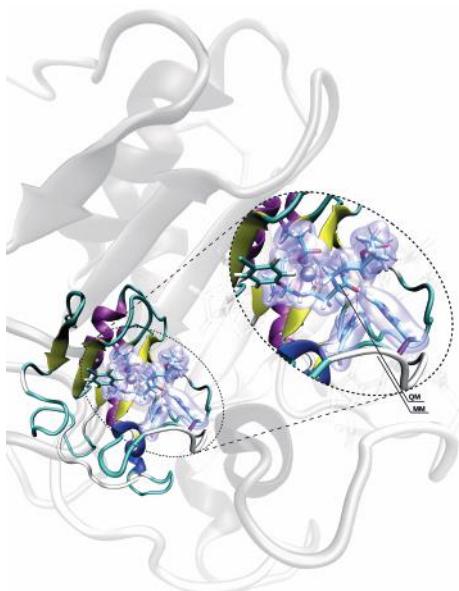
Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)

# QUANTUM MECHANICAL/MOLECULAR MECHANICAL APPROACHES FOR THE INVESTIGATION OF CHEMICAL SYSTEMS – RECENT DEVELOPMENTS AND ADVANCED APPLICATIONS

Topic Editors:

**Thomas S. Hofer**, University of Innsbruck, Austria

**Sam P. De Visser**, University of Manchester, United Kingdom



The binding mode of the only known ligand robotnikinin to the Zn-site of the extracellular signalling protein Sonic Hedgehog as obtained from a hybrid QM/MM simulations (Hitzenberger et al., 2017). The comparably large QM zone shown as blue volume contains the central peptide-like ring of robotnikinin, the  $Zn^{2+}$ -ion and the side chains of five amino acids critical for protein-ligand (H134, H135) and protein-ion interaction (D148, E177 and H183). A total of 7 QM/MM link bonds indicated in purple are present in the QM sub-system. Image: "Shh-Robotnikinin Complex" by Thomas S. Hofer.

The QM/MM method, short for quantum mechanical/molecular mechanical, is a highly versatile approach for the study of chemical phenomena, combining the accuracy of quantum chemistry to describe the region of interest with the efficiency of molecular mechanical potentials to represent the remaining part of the system. Originally conceived in the 1970s by the influential work of the the Nobel laureates Martin Karplus, Michael Levitt and Arieh Warshel, QM/MM techniques have evolved into one of the most accurate and general approaches to investigate the properties of chemical systems via computational methods. Whereas the first applications have been focused on studies of organic and biomolecular systems, a large variety of QM/MM implementations have been developed over the last decades, extending the range of applicability to address research questions relevant for both solution and solid-state chemistry as well.

Despite approaching their 50th anniversary in 2022, the formulation of improved QM/MM methods is still an active field of research, with the aim to (i) extend the applicability to address an even broader range of research questions in chemistry and related disciplines, and (ii) further push the accuracy achieved in the QM/MM description beyond that of established formulations. While being a highly successful approach on its own, the combination of the QM/MM strategy with other established theoretical techniques greatly extends the capabilities of the computational approaches. For instance the integration of a suitable QM/MM technique into the highly successful Monte-Carlo and molecular dynamics simulation protocols enables the description of the chemical systems on the basis of an ensemble that is in part constructed on a quantum-mechanical basis.

This eBook presents the contributions of a recent Research Topic published in *Frontiers in Chemistry*, that highlight novel approaches as well as advanced applications of QM/MM method to a broad variety of targets. In total 2 review articles and 10 original research contributions from 48 authors are presented, covering 12 different countries on four continents. The range of research questions addressed by the individual contributions provide a lucid overview on the versatility of the QM/MM method, and demonstrate the general applicability and accuracy that can be achieved for different problems in chemical sciences. Together with the development of improved algorithms to enhance the capabilities of quantum chemical methods and the continuous advancement in the capacities of computational resources, it can be expected that the impact of QM/MM methods in chemical sciences will be further increased already in the near future.

**Citation:** Hofer, T. S., De Visser, S. P., eds. (2018). Quantum Mechanical/Molecular Mechanical Approaches for the Investigation of Chemical Systems – Recent Developments and Advanced Applications. Lausanne: Frontiers Media.  
doi: 10.3389/978-2-88945-626-0

# Table of Contents

## CHAPTER I

### EDITORIAL

- 06 Editorial: Quantum Mechanical/Molecular Mechanical Approaches for the Investigation of Chemical Systems – Recent Developments and Advanced Applications**

Thomas S. Hofer and Sam P. de Visser

## CHAPTER II

### REVIEW ARTICLES

- 11 Chemical Reactivity and Spectroscopy Explored From QM/MM Molecular Dynamics Simulations Using the LIO Code**

Juan P. Marcolongo, Ari Zeida, Jonathan A. Semela, Nicolás O. Foglia, Uriel N. Morzan, Dario A. Estrin, Mariano C. González Lebrero and Damián A. Scherlis

- 23 Steady-State Linear and Non-linear Optical Spectroscopy of Organic Chromophores and Bio-macromolecules**

Marco Marazzi, Hugo Gattuso, Antonio Monari and Xavier Assfeld

## CHAPTER III

### NOVEL METHODOLOGIES AND BENCHMARKING OF QM/MM TECHNIQUES

- 42 On the Difference Between Additive and Subtractive QM/MM Calculations**

Lili Cao and Ulf Ryde

- 57 Automated Fragmentation QM/MM Calculation of NMR Chemical Shifts for Protein-Ligand Complexes**

Xinsheng Jin, Tong Zhu, John Z. H. Zhang and Xiao He

- 68 A Force Balanced Fragmentation Method for ab Initio Molecular Dynamic Simulation of Protein**

Mingyuan Xu, Tong Zhu and John Z. H. Zhang

- 76 QM Cluster or QM/MM in Computational Enzymology: The Test Case of LigW-Decarboxylase**

Mario Prejanò, Tiziana Marino and Nino Russo

- 85 Interfacing the Core-Shell or the Drude Polarizable Force Field With Car-Parrinello Molecular Dynamics for QM/MM Simulationse**

Sudhir K. Sahoo and Nisanth N. Nair

## CHAPTER IV

### ADVANCED APPLICATIONS OF QM/MM TECHNIQUES

- 95 The Binding Mode of the Sonic Hedgehog Inhibitor Robotnikinin, a Combined Docking and QM/MM MD Study**

Manuel Hitzenberger, Daniela Schuster and Thomas S. Hofer

- 109 How are Substrate Binding and Catalysis Affected by Mutating Glu<sub>127</sub> and Arg<sub>161</sub> in Prolyl-4-hydroxylase? A QM/MM and MD Study**

Amy Timmins and Sam P. de Visser

- 122 *Chloride Ion Transport by the E. coli CLC Cl<sup>-</sup>/H<sup>+</sup> Antiporter: A Combined Quantum-Mechanical and Molecular-Mechanical Study***  
Chun-Hung Wang, Adam W. Duster, Baris O. Aydintug, MacKenzie G. Zarecki and Hai Lin
- 139 *Modeling Chemical Reactions by QM/MM Calculations: The Case of the Tautomerization in Fireflies Bioluminescent Systems***  
Romain Beraud-Pache, Cristina Garcia-Iriepa and Isabelle Navizet
- 152 *Exploring the Interaction Mechanism Between Cyclopeptide DC3 and Androgen Receptor Using Molecular Dynamics Simulations and Free Energy Calculations***  
Huimin Zhang, Tianqing Song, Yizhao Yang, Chenggong Fu and Jiazhong Li
- 167 *Conceptual DFT Study of the Local Chemical Reactivity of the Colored BISARG Melanoidin and its Protonated Derivative***  
Juan Frau and Daniel Grossman-Mitnik
- 176 *QM/MM Investigation of the Role of a Second Coordination Shell Arginine in [NiFe]-Hydrogenases***  
Andrés M. Escorcia and Matthias Stein



# Editorial: Quantum Mechanical/Molecular Mechanical Approaches for the Investigation of Chemical Systems – Recent Developments and Advanced Applications

Thomas S. Hofer<sup>1\*</sup> and Sam P. de Visser<sup>2</sup>

<sup>1</sup> Theoretical Chemistry Division, Institute of General, Inorganic and Theoretical Chemistry, University of Innsbruck, Innsbruck, Austria, <sup>2</sup> School of Chemical Engineering and Analytical Science, Manchester Institute of Biotechnology, The University of Manchester, Manchester, United Kingdom

**Keywords:** hybrid QM/MM, ab initio methods, quantum chemistry, force fields, molecular mechanics, quantum chemistry, density functional theory

## Editorial on the Research Topic

### OPEN ACCESS

#### Edited and reviewed by:

Hans Martin Senn,  
University of Glasgow,  
United Kingdom

**\*Correspondence:**  
Thomas S. Hofer  
t.hofer@uibk.ac.at

#### Specialty section:

This article was submitted to  
Theoretical and Computational  
Chemistry,  
a section of the journal  
*Frontiers in Chemistry*

**Received:** 29 June 2018

**Accepted:** 30 July 2018

**Published:** 13 September 2018

#### Citation:

Hofer TS and de Visser SP (2018)  
Editorial: Quantum Mechanical/Molecular Mechanical Approaches for the Investigation of Chemical Systems – Recent Developments and Advanced Applications. *Front. Chem.* 6:357.  
doi: 10.3389/fchem.2018.00357

## Quantum Mechanical/Molecular Mechanical Approaches for the Investigation of Chemical Systems – Recent Developments and Advanced Applications

With the advent of microprocessor technology in the late 1960s (Moore, 1965; Whitworth, 1979; Brinkman et al., 1997) the foundation to a novel interdisciplinary field of research known today as scientific computing was laid. Located at an intersection between mathematics and computational sciences on the one hand and scientific disciplines such as physics and chemistry on the other, computational methods assumed a dominant role in modern science and engineering, enabling investigations of a broad variety of phenomena in effectively every sub-discipline of these vast fields of research.

One of the main challenges for the successful application of computational models to study chemical systems rests with the accurate description of the interaction between atoms and molecules, the two main approaches being quantum mechanics (QM) (Parr and Yang, 1994; Szabo and Ostlund, 1996; Helgaker et al., 2000; Koch and Holthausen, 2002; Cook, 2005; Sholl and Steckel, 2009) and molecular mechanics (MM) (Leach, 2001; Jensen, 2006; Ramachandran et al., 2008) methods. The latter employs empirical (*i.e.*, parametrised), classical representations of the interactions, based on comparably simple potential formulations such as harmonic springs to describe bonds and valence angles as well as Coulomb and Lennard-Jones interactions to account for charge-charge and non-bonded contributions, respectively. These approaches, often referred to as molecular force fields (FFs) provide a versatile and efficient description of chemical systems, provided that the large number of involved parameters are perfectly adjusted and balanced among each other. Typical applications for FF-based studies are located in the realm of biomolecular simulations such as proteins and nucleic acids. Nevertheless, some systems relevant for material sciences can be treated equally well with these approaches and particularly polymer chemistry studies (comprised of the same elements with similar functional groups as found for instance in peptides and proteins) often include FF-based methods.

MM models have also been extended to include more complex phenomena such as polarization (Yu and van Gunsteren, 2005; Baker, 2015; Lemkul et al., 2016) and many-body contributions (Stone, 1995), extending their applicability to areas in which simplistic force field approaches are unreliable, such as studies of solid-state interfaces and semi-conducting systems as well as metals and alloys. However, a key shortcoming inherent to the majority of FF approaches is the inability to describe the formation and cleavage of covalent chemical bonds. While so-called reactive force fields (van Duin et al., 2001; Mahadevan and Garofalini, 2007; Hartke and Grimme, 2010; Liang et al., 2013) have been developed to make such processes accessible in the regime of molecular mechanics, a quantum description of the system is often the natural choice.

QM-based descriptions of chemical systems (Parr and Yang, 1994; Szabo and Ostlund, 1996; Helgaker et al., 2000; Koch and Holthausen, 2002; Cook, 2005; Sholl and Steckel, 2009) partition the atoms into the nuclei and the surrounding electrons, inherently taking all shifts in the electron density resulting from polarization, many-body contributions and even charge-transfer into account. Although a key challenge of QM methods is the accurate description of the correlated motion of electrons (Raghavachari and Anderson, 2010; Popelier, 2011; McDonagh et al., 2017), the hierarchy of quantum chemical approaches established over the last decades (Szabo and Ostlund, 1996; Helgaker et al., 2000; Cook, 2005) provides a versatile framework for the study of challenging chemical phenomena. Since no empirical parameters are required in a QM-based description, quantum chemical methods are not restricted to a particular class of molecules and, thus, generally applicable to achieve first principle descriptions of chemical systems. Unfortunately, these benefits come with a cost, which in this case is a substantially increased computational effort over MM-based approaches and thereby, dramatically limiting their treatable system size.

In order to combine the advantages of MM and QM methods, hybrid QM/MM approaches (Gao, 1993; Bakowies and Thiel, 1996; Lin and Truhlar, 2007; Senn and Thiel, 2007, 2009; Metz et al., 2014; Pezeshki and Lin, 2015; Zheng and Waller, 2016) have been devised: In this framework the most relevant part of the chemical system is treated on the basis of a suitable quantum chemical method, while classical MM potentials are considered sufficiently accurate to model the remaining part of the system. This innovative idea was pioneered by Martin Karplus, Michael Levitt and Arieh Warshel (Warshel and Levitt, 1976; Field et al., 1990; Lyne et al., 1990; Aqvist and Warshel, 1993; Warshel, 2002) in the 1970s, who were awarded the Nobel prize in chemistry for the development of multiscale models for complex chemical systems in 2013. Today, four decades after these influential developments, QM/MM methods are regarded as one of the most influential approaches for the description of challenging chemical phenomena. Initially conceived in the framework of biomolecular simulations (Friesner and Guallar, 2005; Hu and Yang, 2009; van der Kamp and Mulholland, 2013; de Visser et al., 2014; Cui, 2016; Lu et al., 2016; Quesne et al., 2016), the range of QM/MM methods has been substantially extended to include other areas accessing *inter alia*

solid-state chemistry and material science (Gonis and Garland, 1977; Krüger and Rösch, 1994; Stefanovich and Truong, 1996; Jacob et al., 2001; Herschend et al., 2004; Keal et al., 2011; Bjornsson and Bühl, 2012; Golze et al., 2013, 2015; Hofer and Tirler, 2015) and solution chemistry (Staib and Borgis, 1995; Tuñón et al., 1995, 1996; Gao, 1996; Hofer et al., 2010, 2011, 2012; Weiss and Hofer, 2012; Hofer, 2014) as well. These QM/MM studies have given insight into how Nature works, and, for instance, explain regio- and stereochemical selectivities during substrate activation (Faponle et al., 2016, 2017; Timmins et al., 2017). Furthermore, using computational modeling, predictions can be made to engineer proteins and enzymes and in a recent example the computationally proposed change led to a full enantioselectivity reversal (Pratter et al., 2013a,b).

Despite their success and widespread recognition, the development of advanced QM/MM methodologies is still an active field of research, aiming to push the accuracy and applicability of this versatile approach even further. This article collection aims to present an overview of present research activities focused on the development and application of modern QM/MM formulations, demonstrating the versatile capabilities of this celebrated methodology.

A total of 12 exciting contributions by 48 authors from 12 different countries in four continents have been included in this article collection that contains ten original research contributions and two review articles.

Scherlis and his team compiled a review article that covers recent applications of their graphical-processor accelerated LIO code for density functional theory calculations (Marcolongo et al.). The presented examples include the decomposition of nitroxyl in aqueous solution, a comparison of the reactivity of thiols against peroxides in aqueous solution and the active site of a peroxiredoxin enzyme. The latter studies are linked to molecular spectroscopy such as the vibrational spectrum of aqueous peroxynitrite anion and  $\text{LiAlH}_4$  and  $\text{AlH}_4^-$ . Furthermore, using their novel code they predict the UV/Vis spectra of  $(\text{HO})\text{NS}_2$  and the so-called  $\text{NO}/\text{H}_2\text{S}$  “cross-talk” system.

The second review article by Marazza et al. presents a comprehensive overview of recent research activities in studying electronic spectroscopy via QM/MM approaches. A broad variety of examples in the fields of linear absorption, non-linear optical properties and circular dichroism applied to organic molecules, proteins and nucleic acid systems is presented.

Prejanò et al. compared the application of a QM cluster model to the decarboxylation of 5-carboxyvanillate by LigW with results obtained from a more elaborate QM/MM setup. Their study indicates that the reaction is mainly influenced by the constituents of the active site and the cluster model already delivers a reliable description for the presented system.

The article of Escorcio and Stein explored the influence of a conserved arginine residue of the *E. coli* Hyd-1 [NiFe]-Hydrogenase on the  $\text{H}_2$  oxidation reaction via DFT-based QM/MM calculations. This study highlights the key influence of this Arg-residue in promoting both the access of molecular

hydrogen to the catalytically active Ni-atom as well as the associated proton transfer to nearby terminal cysteine residues.

Xu et al. present a novel force balanced simulation approach, separating a protein system into hydrogen-bonded fragments that are computed quantum-mechanically, while the AMOEBA force field is employed to describe long-range non-bonded interactions. To conserve the total energy of the system, a force balancing of the hydrogen link-atoms is carried out. The applicability of this approach is demonstrated for linear ACE-(ALA)<sub>9</sub>-NME as well as the 56 residue GB3 peptides.

The contribution of Sahoo and Nair presents a combination of polarizable Drude oscillators with the well-established Car-Parrinello Molecular Dynamics framework via an extended Lagrangian QM/p-MM method. The approach is demonstrated for a H<sub>2</sub>O(QM)+4H<sub>2</sub>O(p-MM) test system and applied to study an O-vacancy in  $\alpha$ -cristobalite, the hydrogenation of ethene via Y-Zeolite-supported Rh-Clustes and the H<sup>+</sup>-exchange between methane and a H-ZSM-5 zeolite.

The difference between additive and subtractive QM/MM protocols has been highlighted in the contribution by Cao and Ryde, focusing *inter alia* on the different correction schemes to account for errors introduced by the application of link-atoms. Three different systems of increasing complexity have been studied, namely an isolated ethanol molecule, sulfite oxidase and the conversion of oxophlorin to verdohaem by haem oxygenase.

Berraud-Pache et al. studied the keto-enol tautomerisation reaction of oxyluciferin representing the emissive species in the bioluminescent system of fireflies. By combining classical molecular dynamics studies of the active species in a polarisable continuum and explicit QM/MM calculations, the keto-OxyLH<sup>-</sup> species was identified as the most likely candidate to act as emitter in bioluminescence.

He and colleagues applied an automated fragmentation QM/MM protocol to study <sup>1</sup>H chemical shifts of the apo- and holo-neocarzinostatin-chromophore binding complex (Jin et al.). The calculated NMR data obtained by the fragmented QM/MM approach proved to be in good agreement with results of large-scale calculations as well as experimental data.

The research team of Lin studied the migration of Cl<sup>-</sup> through the transmembrane domain of a prototypical *E. coli* chloride-channel (Wang et al.), thereby including the entire pore section

into the quantum-mechanically treated zone. The obtained results demonstrate that the influence of electron delocalization, inherently taken into account at the QM level of theory, appear to be more critical than previously considered.

Timmins and De Visser investigated the impact of different mutations in prolyl-4-hydroxylase via a combined QM/MM and MD study. Based on the results of this extensive study two mutants with the potential of displaying notably changes in the regio- and stereoselectivity could be identified.

Hitzenberger et al. provided a contribution combining docking and pharmacophore modeling with QM-based molecular dynamics simulations to investigate the binding of the only known inhibitor robotnikinin to the Zn-site of the extracellular signaling protein Sonic Hedgehog. Comparison to a purely classical molecular dynamics highlights the substantially improved description of the binding observed in the QM/MM MD simulation.

In the contribution of Frau and Grossman-Mitnik the influence of different range-separated hybrid DFT methods in the prediction of chemical reactivity descriptors was evaluated.

Finally, Li and coworkers investigated the interaction mechanism between cyclopeptide DC3 and an androgen receptor via free energy calculations and extensive molecular dynamics simulations Zhang et al..

Clearly, the scope of topics covered by the contributions in this article collection demonstrates the widespread capabilities of the QM/MM technique as a general and versatile approach to address a broad spectrum of research questions. Moreover, the research field is as active as ever and moving into many different research directions. In conjunction with the formulation of advanced theoretical approaches, efficient simulation programs and the development of improved computational infrastructure, QM/MM methods proved to be an indispensable tool in modern chemical research, providing a highly successful alternative route for the study of complex chemical phenomena, which can be expected to play an even more dominant role in the coming years.

## AUTHOR CONTRIBUTIONS

The article collection was edited SdV and TH. The respective editorial was compiled by TH.

## REFERENCES

- Aqvist, J., and Warshel, A. (1993). Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. *Chem. Rev.* 93:2523. doi: 10.1021/cr00023a010
- Baker, C. M. (2015). Polarizable force fields for molecular dynamics simulations of biomolecules. *WIREs Comp. Mol. Sci.* 5, 241–254. doi: 10.1002/wcms.1215
- Bakowies, D., and Thiel, W. (1996). Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.* 100:10580. doi: 10.1021/jp9536514
- Bjornsson, R., and Bühl, M. (2012). Modeling molecular crystals by QM/MM: self-consistent electrostatic embedding for geometry optimizations and molecular property calculations in the solid. *J. Chem. Theory Comput.* 8, 498–508. doi: 10.1021/ct200824r
- Brinkman, W. F., Haggan, D. E., and Troutman, W. W. (1997). A history of the invention of the transistor and where it will lead us. *IEEE J. Solid St. Circ.* 32, 1858–1865. doi: 10.1109/4.643644
- Cook, D. B. (2005). *Handbook of Computational Chemistry*. New York, NY: Dover Publications.
- Cui, Q. (2016). Perspective: Quantum mechanical methods in biochemistry and biophysics. *J. Chem. Phys.* 145, 140901/1–140901/12. doi: 10.1063/1.4964410
- de Visser, S. P., Quesne, M. G., Martin, B., Comba, P., and Ryde, U. (2014). Computational modelling of oxygenation processes in enzymes and biomimetic model complexes. *Chem. Commun.* 50, 262–282. doi: 10.1039/C3CC47148A
- Faponle, A. S., Quesne, M. G., and de Visser, S. P. (2016). Origin of the regioselective fatty acid hydroxylation versus decarboxylation by a cytochrome

- P450 peroxygenase: what drives the reaction to biofuel production? *Chem. Eur. J.* 22, 5478–5483. doi: 10.1002/chem.201600739
- Faponle, A. S., Seebeck, F. P., and de Visser, S. P. (2017). Sulfoxide synthase versus cysteine dioxygenase reactivity in a nonheme iron enzyme. *J. Am. Chem. Soc.* 139, 9259–9270. doi: 10.1021/jacs.7b04251
- Field, M. J., Bash, P. A., and Karplus, M. (1990). A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* 11:700. doi: 10.1002/jcc.540110605
- Friesner, R. A., and Guallar, V. (2005). Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. *Ann. Rev. Phys. Chem.* 56, 389–427. doi: 10.1146/annurev.physchem.55.091602.094410
- Gao, J. (1993). Potential of mean force for the isomerization of DMF in aqueous solution: a monte carlo QM/MM simulation study. *J. Am. Chem. Soc.* 115:2930. doi: 10.1021/ja00060a047
- Gao, J. (1996). Hybrid quantum and molecular mechanical simulations An alternative avenue to solvent effects in organic chemistry. *Acc. Chem. Res.* 29, 298. doi: 10.1021/ar950140r
- Golze, D., Hutter, J., and Iannuzzi, M. (2015). Wetting of water on hexagonal boron nitride@Rh(111): a QM/MM model based on atomic charges derived for nano-structured substrates. *Phys. Chem. Chem. Phys.* 17:14307. doi: 10.1039/C4CP04638B
- Golze, D., Iannuzzi, M., Nguyen, M., Passerone, D., and Hutter, J. (2013). Simulation of adsorption processes at metallic interfaces: An image charge-augmented QM/MM approach. *J. Chem. Theory Comput.* 9:5086. doi: 10.1021/ct400698y
- Gonis, A., and Garland, J. W. (1977). Multishell method: exact treatment of a cluster in an effective medium. *Phys. Rev. B* 16:2424. doi: 10.1103/PhysRevB.16.2424
- Hartke, B. and Grimme, S. (2010). Reactive force fields made simple. *Phys. Chem. Chem. Phys.* 17, 16715–16718. doi: 10.1039/C5CP02580J
- Helgaker, T., Jørgensen, P., and Olsen, J. (2000). *Molecular Electronic Structure Theory*. Chichester: Wiley.
- Herschend, B., Baudin, M., and Hermansson, K. (2004). A combined molecular dynamics+quantum mechanics method for investigation of dynamic effects on local surface structures. *J. Chem. Phys.* 120:4939. doi: 10.1063/1.1635802
- Hofer, T. S. (2014). Perspectives for hybrid ab initio/molecular mechanical simulations of solutions: from complex chemistry to proton-transfer reactions and interfaces. *Pure Appl. Chem.* 86:105. doi: 10.1515/pac-2014-5019
- Hofer, T. S., Hitzenberger, M., and Randolph, B. R. (2012). Combining a Dissociative Water Model with a Hybrid QM/MM Approach - A Simulation Strategy for the Study of Proton Transfer Reactions in Solutions. *J. Chem. Theory Comput.* 8:3586. doi: 10.1021/ct300062k
- Hofer, T. S., Rode, B. M., Pribil, A. B., and Randolph, B. R. (2010). Simulations of Liquids and Solutions Based On Quantum Mechanical Forces. *Adv. Inorg. Chem.* 62:143. doi: 10.1016/S0898-8838(10)62004-1
- Hofer, T. S. and Tirler, A. O. (2015). Combining 2d-periodic quantum chemistry with molecular force fields: a novel QM/MM procedure for the treatment of solid-state surfaces and interfaces. *J. Chem. Comput. Theory* 11, 5873–5887. doi: 10.1021/acs.jctc.5b00548
- Hofer, T. S., Weiss, A. K. H., Randolph, B. R., and Rode, B. M. (2011). Hydration of highly charged ions. *Chem. Phys. Lett.* 512:139. doi: 10.1016/j.cplett.2011.05.060
- Hu, H. and Yang, W. (2009). Development and application of ab initio QM/MM methods for mechanistic simulation of reactions in solution and in enzymes. *J. Mol. Struct. Theochem* 898, 17–30. doi: 10.1016/j.theochem.2008.12.025
- Jacob, T., Fricke, B., Anton, J., Varga, S., Bastug, T., Fritzsche, S., et al. (2001). Cluster-embedding method to simulate large cluster and surface problems. *Eur. Phys. J. D* 16, 257–260. doi: 10.1007/s100530170105
- Jensen, F. (2006). *Introduction to Computational Chemistry*, 2nd Edition. Chichester: John Wiley & Sons Ltd.
- Keal, T. W., Sherwood, P., Dutta, G., Sokol, A. A., and Catlow, C. R. A. (2011). Characterization of hydrogen dissociation over aluminium-doped zinc oxide using an efficient massively parallel framework for QM/MM calculations. *Proc. R. Soc. A* 467:1900. doi: 10.1098/rspa.2010.0613
- Koch, W. and Holthausen, M. C. (2002). *A Chemist's Guide to Density Functional Theory*, 2nd Edition. Weinheim: Wiley-VCH.
- Krüger, S., and Rösch, N. (1994). The moderately-large-embedded-cluster method for metal surfaces: a density-functional study of atomic adsorption. *J. Phys. Condens. Matter* 6:8149. doi: 10.1088/0953-8984/6/40/007
- Leach, A. R. (2001). *Molecular Modelling*, 2nd Edn. Harlow: Prentice-Hall.
- Lemkul, J. A., Huang, J., Roux, B., and MacKerell, A. D. Jr. (2016). An empirical polarizable force field based on the classical drude oscillator model: development history and recent applications. *Chem. Rev.* 116, 4983–5013. doi: 10.1021/acs.chemrev.5b00505
- Liang, T., Shin, Y. K., Cheng, Y., Yilmaz, D. E., Vishnu, K. G., Verners, O., et al. (2013). Reactive potentials for advanced atomistic simulations. *Ann. Rev. Mater. Res.* 43:109. doi: 10.1146/annurev-matsci-071312-121610
- Lin, H. and Truhlar, D. G. (2007). Qm/mm: what have we learned, where are we, and where do we go from here? *Theor. Chim. Acta* 117:185. doi: 10.1007/s00214-006-0143-z
- Lu, X., Fang, D., Ito, S., Okamoto, Y., Ovchinnikov, V., and Cui, Q. (2016). QM/MM free energy simulations: recent progress and challenges. *Mol. Sim.* 42, 1056–1078. doi: 10.1080/08927022.2015.1132317
- Lyne, P. D., Hodosek, M., and Karplus, M. (1990). A hybrid QM-MM potential employing Hartree-Fock or density functional methods in the quantum region. *J. Phys. Chem. A* 103:3462. doi: 10.1021/jp982115j
- Mahadevan, T. S. and Garofalini, S. H. (2007). Dissociative Water Potential for Molecular Dynamics Simulations. *J. Phys. Chem. B* 111:8919. doi: 10.1021/jp072530
- McDonagh, J. L., Silva, A. F., Vincent, M. A., Popelier, P., and Paul, L. A. (2017). Quantifying Electron Correlation of the Chemical Bond. *J. Phys. Chem. Lett.* 8, 1937–1942. doi: 10.1021/acs.jpclett.7b00535
- Metz, S., Kästner, J., Sokol, A. A., Keal, T. W., and Sherwood, P. (2014). ChemShell—a modular software package for QM/MM simulations. *WIREs Comput. Mol. Sci.* 4:101. doi: 10.1002/wcms.1163
- Moore, G. (1965). Cramming more components onto integrated circuits. *Electronics* 38, 114–117.
- Parr, R. G. and Yang, W. (1994). *Density-Functional Theory of Atoms and Molecules*. Oxford: Oxford University Press.
- Pezeshki, S., and Lin, H. (2015). “Recent progress in adaptive-partitioning QM/MM methods for Born-Oppenheimer molecular dynamics,” in *Quantum Modeling of Complex Molecular Systems, Challenges and Advances in Computational Chemistry and Physics*, Vol. 21, eds J-L. Rivail, M. Ruiz-Lopez, and X. Assfeld (Cham: Springer International Publishing), 93–113.
- Popelier, P. (ed.) (2011). *Solving the Schrödinger Equation: Has Everything Been Tried?* Singapore: World Scientific.
- Pratter, S. M., Konstantinovics, C., DiGiuro, C. L. M., Leitner, E., Kumar, D., de Visser, S. P., et al. (2013a). Inversion of enantio-selectivity of a mononuclear non-heme iron(II)-dependent hydroxylase by tuning the interplay of metal-center geometry and protein structure. *Angew. Chem. Int. Ed.* 52, 9677–9681. doi: 10.1002/anie.201304633
- Pratter, S. M., Konstantinovics, C., DiGiuro, C. L. M., Leitner, E., Kumar, D., de Visser, S. P., et al. (2013b). Inversion of enantio-selectivity of a mononuclear non-heme iron(II)-dependent hydroxylase by tuning the interplay of metal-center geometry and protein structure. *Angew. Chem.* 125, 9859–9863. doi: 10.1002/ange.201304633
- Quesne, M. G., Borowski, T., and de Visser, S. P. (2016). Quantum Mechanics/Molecular Mechanics Modelling of Enzymatic Processes: Caveats and Breakthroughs. *Chem. Eur. J.* 22, 2562–2581. doi: 10.1002/chem.201503802
- Raghavachari, K. and Anderson, J. B. (2010). Electron Correlation Effects in Molecules. *J. Phys. Chem.* 100, 12960–12973. doi: 10.1021/jp953749i
- Ramachandran, K. I., Deepa, G., and Namboori, K. (2008). *Computational Chemistry and Molecular Modeling: Principles and Applications* (Berlin: Springer.)
- Senn, H. M., and Thiel, W. (2007). QM/MM studies of enzymes. *Current Opinion in Chemical Biology* 11:182. doi: 10.1016/j.cbpa.2007.01.684
- Senn, H. M., and Thiel, W. (2009). QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed.* 48, 1198–1229. doi: 10.1002/anie.200802019
- Sholl, D. S., and Steckel, J. A. (2009). *Density Functional Theory-A Practical Introduction*. Hoboken, NY: Wiley.
- Staib, A., and Borgis, D. (1995). Molecular dynamics simulation of an excess charge in water using mobile Gaussian orbitals. *J. Chem. Phys.* 103:2642. doi: 10.1063/1.470524

- Stefanovich, E. V., and Truong, T. N. (1996). Embedded density functional approach for calculations of adsorption on ionic crystals. *J. Chem. Phys.* 104, 2946–2955. doi: 10.1063/1.471115
- Stone, A. J. (1995). *The Theory of Intermolecular Forces*. Oxford, UK: Oxford University Press.
- Szabo, A., and Ostlund, N. S. (1996). *Modern Quantum Chemistry*. New York, NY: Dover Publications.
- Timmins, A., Saint-André, M., and de Visser, S. P. (2017). Understanding how prolyl-4-hydroxylase structure steers a ferryl oxidant toward scission of a strong C-H bond. *J. Am. Chem. Soc.* 139, 9855–9866. doi: 10.1021/jacs.7b02839
- Tuñón, I., Martins-Costa, M. T. C., Millot, C., and Ruiz-López, M. F. (1995). Coupled density functional/molecular mechanics Monte Carlo simulations of ions in water. The bromide ion. *Chem. Phys. Lett.* 241:450. doi: 10.1016/0009-2614(95)00615-B
- Tuñón, I., Martins-Costa, M. T. C., Millot, C., Ruiz-López, M. F., and Rivail, J. L. (1996). A coupled density functional-molecular mechanics Monte Carlo simulation method: the water molecule in liquid water. *J. Comput. Chem.* 17:19. doi: 10.1002/(SICI)1096-987X(19960115)17:1<19::AID-JCC2>3.0.CO;2-3
- van der Kamp, M. W. and Mulholland, A. J. (2013). Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry* 52, 2708–2728. doi: 10.1021/bi400215w
- van Duin, A. C. T., Dasgupta, S., Lorant, F., and Goddard, W. A. (2001). ReaxFF A reactive force field for hydrocarbons. *J. Phys. Chem. A* 105, 9396. doi: 10.1021/jp004368u
- Warshel, A. (2002). Molecular dynamics simulations of biological reactions. *Acc. Chem. Res.* 35:385. doi: 10.1021/ar010033z
- Warshel, A., and Levitt, M. (1976). Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbenium Ion in the reaction of lysozyme. *J. Mol. Biol.* 103:227. doi: 10.1016/0022-2836(76)90311-9
- Weiss, A. K., and Hofer, T. S. (2012). Exploiting the capabilities of quantum chemical simulations to characterise the hydration of molecular compounds. *RSC Adv.* 3:1606. doi: 10.1039/C2RA21873A
- Whitworth, I. (1979). Review of microprocessor architecture. *Microprocess Microsys.* 3, 21–28. doi: 10.1016/0141-9331(79)90088-7
- Yu, H. and van Gunsteren, W. F. (2005). Accounting for polarization in molecular simulation.. *Comp. Phys. Commun.* 172, 69–85. doi: 10.1016/j.cpc.2005.01.022
- Zheng, M., and Waller, M. P. (2016). Adaptive quantum mechanics/molecular mechanics methods. *WIREs Comput. Mol. Sci.* 6, 369–385. doi: 10.1002/wcms.1255

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Hofer and de Visser. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Chemical Reactivity and Spectroscopy Explored From QM/MM Molecular Dynamics Simulations Using the LIO Code

Juan P. Marcolongo<sup>1†</sup>, Ari Zeida<sup>1,2†</sup>, Jonathan A. Semelak<sup>1</sup>, Nicolás O. Foglia<sup>1</sup>, Uriel N. Morzan<sup>1</sup>, Dario A. Estrin<sup>1</sup>, Mariano C. González Lebrero<sup>1\*</sup> and Damián A. Scherlis<sup>1\*</sup>

## OPEN ACCESS

### Edited by:

Sam P. De Visser,  
University of Manchester,  
United Kingdom

### Reviewed by:

Yong Wang,  
Lanzhou Institute of Chemical Physics  
(CAS), China  
Nino Russo,  
Dipartimento di Chimica e Tecnologie  
Chimiche, Università della Calabria,  
Italy

### \*Correspondence:

Mariano C. González Lebrero  
nanolebrero@qi.fcen.uba.ar  
Damián A. Scherlis  
damian@qi.fcen.uba.ar

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Theoretical and Computational  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

Received: 02 February 2018

Accepted: 05 March 2018

Published: 21 March 2018

### Citation:

Marcolongo JP, Zeida A, Semelak JA,  
Foglia NO, Morzan UN, Estrin DA,  
González Lebrero MC and Scherlis DA  
(2018) Chemical Reactivity and  
Spectroscopy Explored From QM/MM  
Molecular Dynamics Simulations  
Using the LIO Code.  
Front. Chem. 6:70.  
doi: 10.3389/fchem.2018.00070

<sup>1</sup> DQIAyQF, INQUIMAE-CONICET, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina, <sup>2</sup> Departamento de Bioquímica and Center for Free Radical and Biomedical Research, Facultad de Medicina, Universidad de la República, Montevideo, Uruguay

In this work we present the current advances in the development and the applications of LIO, a lab-made code designed for density functional theory calculations in graphical processing units (GPU), that can be coupled with different classical molecular dynamics engines. This code has been thoroughly optimized to perform efficient molecular dynamics simulations at the QM/MM DFT level, allowing for an exhaustive sampling of the configurational space. Selected examples are presented for the description of chemical reactivity in terms of free energy profiles, and also for the computation of optical properties, such as vibrational and electronic spectra in solvent and protein environments.

**Keywords:** QM/MM, DFT, GPU, free energy, TDDFT

## INTRODUCTION

The accurate prediction of the physicochemical properties of molecules within a realistic description of the surrounding environment is essentially the driving force in the development of QM/MM methods. The evolution of this field is not only associated with theoretical progresses, but is also very much related, and often constrained, by the advances in computer power. The last few decades have witnessed the birth and growth of two large categories of QM/MM strategies: those based on expensive QM levels of theory [like Density Functional Theory (DFT) or perturbation methods], but only applicable to dozens of quantum-mechanical atoms and a few nuclear configurations; and those that may account for thermal effects, i.e., thousands or even millions of configurations, affordable at the expense of a reduction in the accuracy of the electronic structure calculations, mostly by means of semiempirical approaches.

In this contribution, we present and review the current advances in the development of LIO, a lab-made code designed for electronic structure calculations exploiting the computational advantages of graphic processing units. This code can be coupled to molecular dynamics (MD) engines to perform QM/MM simulations, allowing for exhaustive MD sampling at the DFT level. In the next section, we provide an overview on the program structure and its capabilities, accompanied by some benchmarks illustrating efficiency. Then, we review a series of representative applications of our methodology, namely the description of chemical reactivity in terms of free energy profiles, and the prediction of optical properties, such as vibrational and electronic spectra.

Closing remarks are given in the final section, along with some observations concerning current and future directions.

## THE LIO CODE

*LIO* (<https://github.com/MALBECC/LIO>) is a highly efficient tool to solve the electronic structure problem in molecules using DFT and Time Dependent DFT (TDDFT) frameworks. *LIO* is designed to be used as a standalone application or as a library to be combined with other molecular dynamics packages, to run QM/MM simulations. This versatility constitutes a design advantage to maximize its applicability and power. While its roots go back to the early 90's, to the program originally authored by Estrin et al. (1993), with the first QM/MM applications on chemical reactivity dating from only a few years later (Elola et al., 1999), today *LIO* is a project in continuous evolution with a focus in performance, and the outcome of an interdisciplinary effort bringing together developers from chemistry and computer science backgrounds. *LIO* has been interfaced with the *Amber* package (Pearlman et al., 1995), to perform Born-Oppenheimer molecular dynamics and electron dynamics simulations in a hybrid QM/MM framework. In this context, the total energy is obtained according to the electrostatic embedding, additive QM/MM formulation:

$$E = E_{QM} + E_{MM} + E_{QM-MM} \quad (1)$$

where the first term on the right hand side corresponds to the QM Kohn-Sham energy, the second one to the MM force field potential, and the third one is the coupling energy between the classical and quantum regions.

$$E_{QM} [\rho] = T_s [\rho] + \sum_I \int \frac{\rho Z_I}{|\mathbf{r} - \mathbf{R}_I|} d\mathbf{r} + \frac{1}{2} \int \int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 + E_{xc} + \sum_I \sum_A \frac{Z_I Z_A}{|\mathbf{R}_I - \mathbf{R}_A|} \quad (2)$$

$$E_{QM-MM} = E_{LJ}^{QM-MM} (|\mathbf{R}_A - \mathbf{R}_I|) + \sum_{A \in MM} q_A \int \frac{\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_I|} d\mathbf{r} + \sum_{I \in QM} \sum_{A \in MM} \frac{Z_I q_A}{|\mathbf{R}_A - \mathbf{R}_I|} \quad (3)$$

In the equations above,  $\rho$  represents the electron density,  $T_s$  the electronic kinetic energy,  $E_{xc}$  the exchange correlation term,  $Z_I$  the atomic number of quantum atom  $I$ , and  $q_A$  the partial charge of the classical atom  $A$ .  $E_{LJ}^{QM-MM}$  is a non-electrostatic term, which describes dispersion and short range repulsion effects between QM and MM atoms, using Lennard-Jones potentials, consistently with the MM force field.

The kinetic energy and nuclear attraction contributions are calculated in terms of analytical one electron integrals, which are derived using Obara-Saika recursive equations (Obara and Saika, 1986). The electron repulsion term is computed in terms of two-electron repulsion integrals (ERI), which are also derived recursively. The exchange correlation energy is calculated using

numerical spherically centered grids (Becke, 1988). The flow diagram of the computation scheme is depicted in **Figure 1**.

In the development of efficient algorithms for electronic structure calculations it is very important to consider the size of the systems to be treated. Hence, in *LIO* we put our focus on medium-sized systems (a few tens of atoms), which are the typical dimensions of the QM region in hybrid calculations. As can be seen in **Figure 1**, in this QM/MM scheme the major computational cost goes into the calculation of exchange-correlation, electron repulsion (ERIs) and QM/MM energy terms (in that order) and the corresponding forces contribution. In medium-size systems the cost of the computation is dominated by the exchange-correlation integral which scales linearly, whereas for larger systems the calculation of ERIs (which scale quadratically) and the diagonalization of the Fock matrix (which scales cubically) have an increasing importance.

Several optimization schemes were implemented to improve performance, like a linear scaling algorithm for exchange-correlation (Stratmann et al., 1996) or the use of auxiliary basis functions for the ERIs (Stratmann et al., 1996). In addition, these terms are totally or partially computed in the GPU (Nitsche et al., 2014). The overall result is a code that performs QM/MM molecular dynamics with high efficiency allowing for the computation of systems and/or properties at reduced computational costs. **Table 1** shows the computation timings for some selected QM/MM systems discussed along the coming sections, and also for the case of the activation of copper-translocating P-type ATPase from *Archaeoglobus fulgidus* (*AfCopA*), an enzyme that couples the energy of ATP hydrolysis to catalyze  $\text{Cu}^{+}$  translocation across cellular membranes (see **Figure 2**) (Tsuda and Toyoshima, 2009).

Aside from standard molecular dynamics simulations, *LIO* can propagate the electron density as a function of time for a fixed molecular configuration, becoming, to the best of our knowledge, the first real-time TDDFT (RT-TDDFT) implementation in a QM/MM setting (Morzan et al., 2014). The evolution of the density matrix  $\rho$  is performed according to the Liouville-von Neumann equation:

$$\frac{\partial \rho}{\partial t} = \frac{1}{i\hbar} [H, \rho] \quad (4)$$

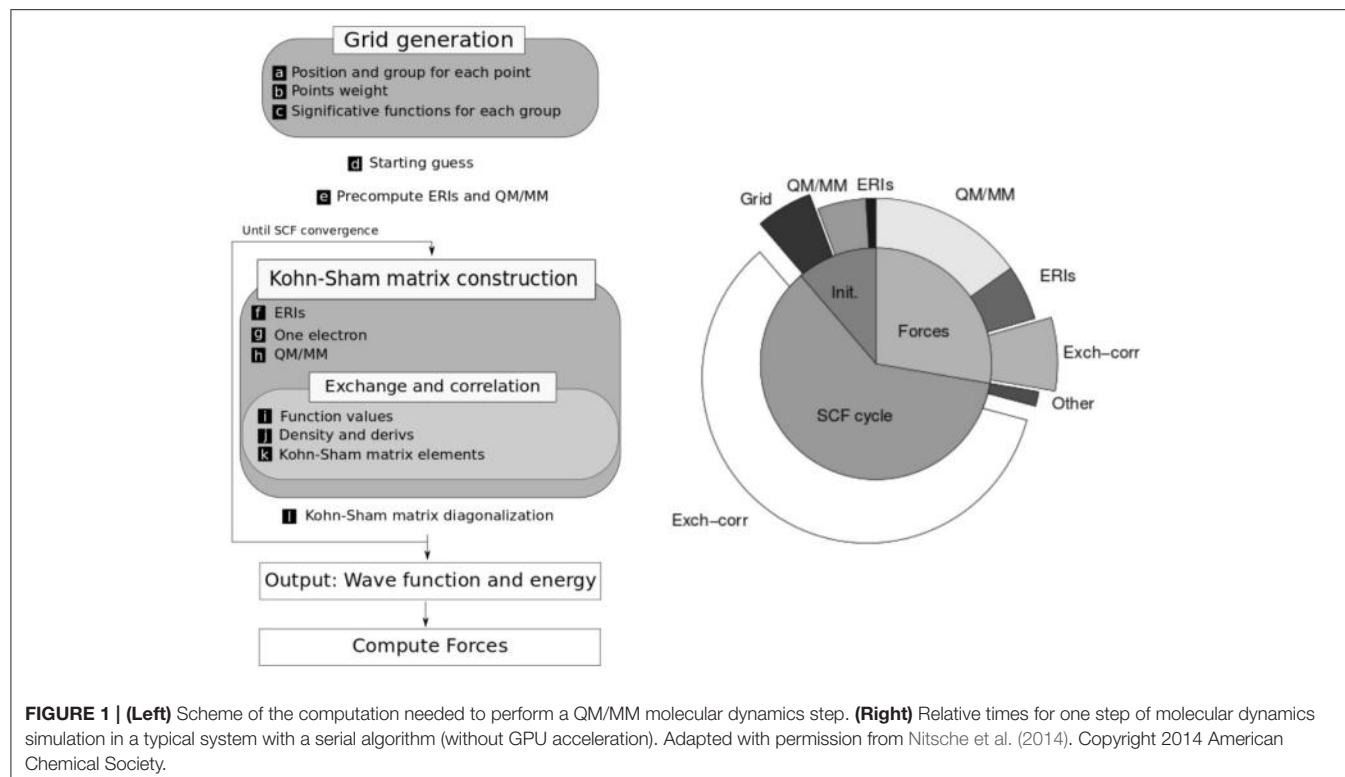
where  $H$  is the Kohn-Sham matrix. The integration of this equation can be realized with either a Verlet integration scheme,

$$\rho(t + \Delta t) = \frac{2}{i\hbar} [H(t), \rho(t)] \Delta t + \rho(t - \Delta t) \quad (5)$$

or the Magnus expansion to first order,

$$\begin{aligned} \rho(t + \Delta t) = & \rho(t) - i\Delta t [H(t + \Delta t/2), \rho(t)] \\ & - \frac{\Delta t^2}{2!} [H(t + \Delta t/2), [H(t + \Delta t/2), \rho(t)]] \\ & + i \frac{\Delta t^3}{3!} [H(t + \Delta t/2), [H(t + \Delta t/2), [H(t + \Delta t/2), \rho(t)]]] + \dots \end{aligned} \quad (6)$$

The advantage associated with the Magnus expansion is that it allows a greater  $\Delta t$  than the Verlet algorithm (usually



**TABLE 1 |** L/O performance in ground state molecular dynamics simulations, in terms of timings per QM/MM steps.

System	QM atoms	MM atoms	Number of basis functions	Time per step (L/O) (s)	Time per step total (s)	Ps/day
CH <sub>3</sub> S <sup>-</sup> + H <sub>2</sub> O <sub>2</sub>	9	11445	79	0.622	0.74	113.9
MtAhpE + H <sub>2</sub> O <sub>2</sub>	22	240000	178	1.55	3.86	22.4
AfCopA	34	61000	364	4.65	5.30	16.3

Calculations were conducted in an Intel(R) Core(TM) i5-7400 CPU @ 3.00 GHz and a Gforce 980 TI GPU.

in the order of 10 to 20 times larger), reducing the total number of steps needed for a fixed simulation time. On the other hand, the computational burden associated with the Magnus algorithm is much larger because it entails a greater number of matrix multiplications. However, these operations can be efficiently handled by GPUs, with a particularly high impact in the evaluation of the Magnus expansion which is usually truncated at order 10 to 50. Therefore, the use of the Magnus integrator turns out to be much more efficient than the use of the Verlet scheme when running on GPU.

The TDDFT scheme involves the evaluation of the Fock matrix, so all the optimizations made for SCF calculations are exploited here as well. Moreover, when working with fixed nuclear positions, other optimizations can be applied, especially in the QM/MM coupling integral and ERIs, significantly reducing their cost and leaving a quasi-linear scaling method. The efficiency of TDDFT depends on both the computation of one simulation step, and the maximum time step ( $\Delta t$ ) that can be used. The later varies with the type of atoms in the QM zone

and can be extended by using better propagators and/or freezing the inner electron density through the use of pseudopotentials. For a more detailed discussion please refer to Foglia et al. (2017). Table 2 illustrates the performance of TDDFT calculations comparing the full electron scheme with a pseudopotential approach.

## EXAMPLES

### Chemical Reactivity

The elucidation of the thermodynamics and kinetics of chemical reactions is one of the main goals of theoretical chemistry. In this context, the computation of free energy profiles assisted by QM/MM schemes has proved extremely useful to obtain mechanistic information (Kollman, 1993; Chipot and Pearlman, 2002; Hu and Yang, 2008; Carvalho et al., 2014). The two key ingredients for obtaining a meaningful free energy profile are the selection of the QM region and the QM level of theory on one hand, and the quality of the sampling on the other. Many sampling methodologies were developed and tested in order to

optimize the tradeoff between cost and accuracy (Chipot and Pohorille, 2007). Each scheme presents advantages and caveats, but all of them rely in an extensive sampling of configurations of the system along the reactive process. The speedups described in the previous section, and the fact that *LIO* can be coupled to different MD engines, like *Gromacs* (Van Der Spoel et al., 2005) or *Amber* (Pearlman et al., 1995), allow us to attain this kind of sampling at the DFT level. It is worth noting that obtaining each of these profiles often requires  $\sim 0.5\text{--}1$  ns of QM/MM MD sampling. The following sections show selected examples of chemical reactions studied with our code, both in aqueous solution and in enzyme catalyzed processes. In these particular cases, the QM region computations were performed at the generalized gradient approximation (GGA) using the *PBE* combination of exchange and correlation functionals, with a *dzvp* basis set (Godbout et al., 1992). The MM region was treated with the *Amber99* forcefield (Lindorff-Larsen et al., 2010).

### Reactivity in Aqueous Solution: Nitrous Oxide Formation Upon Nitroxyl Decomposition

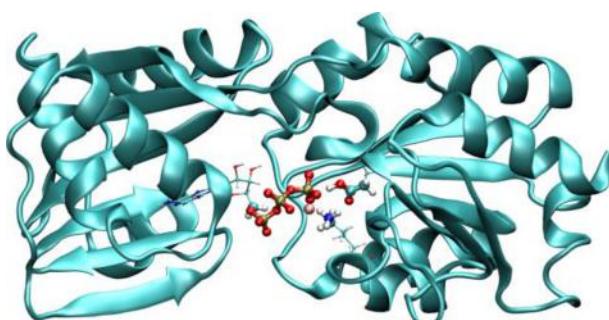
Modeling solvent effects on chemical reactivity, and the influence of aqueous solvation in particular, has been among the main goals of computational chemistry, since much of the most relevant processes in chemistry, biochemistry and materials sciences, take place in solution or at a solid-liquid interface. Here we illustrate

the application of our code to aqueous chemistry through an example involving nitroxyl, a key species in redox biochemistry. Nitroxyl (*HNO*) is a species playing different roles in nitrosative stress processes with great interest in the pharmacology field due to its potential use in heart failure treatment, as well as its vasodilator properties and its role in cellular metabolism (Ma et al., 1999; Miranda, 2005; Miranda et al., 2005). Nitroxyl rapidly decomposes in aqueous solution yielding nitrous oxide (*N<sub>2</sub>O*) (Shafirovich and Lymar, 2002). This is a fast reaction that competes with many other chemical processes in which *HNO* may be involved in a cellular context. Then, a detailed molecular description of this phenomenon is of general interest in biochemistry.

Great efforts have previously been done from an experimental and theoretical point of view to study the different steps leading to *HNO* decomposition (Shafirovich and Lymar, 2002; Fehling and Friedrichs, 2011). It was proposed that the mechanism involved a dimerization of *HNO* followed by a cleavage of one of the N-O bonds, to yield *N<sub>2</sub>O* and water. Our aim was to fully characterize the energetics and the molecular determinants of the reaction mechanism, taking into account the influence of the environment, which could be extremely important especially because an acid-base equilibrium was proposed to be involved in the mechanism. Therefore, we studied both steps of the reaction mechanism, evaluating different protonation states and possible isomers, by means of multiple QM/MM MDs, using the umbrella sampling scheme to determine the corresponding free energy profiles (Bringas et al., 2016).

Our results showed that the dimerization is an exergonic process that occurs without a significant activation barrier (**Figure 3A**), and that the *cis* isomer of the dimer (*HONNOH*) is more stable than the *trans* one by  $\sim 2$  kcal/mol (data not shown). The dimer intermediate might be involved in different acid-base equilibria, so we investigated the second step of the reaction, starting from different protonation states (**Figure 3B**). The anionic path showed a much lower activation barrier ( $\sim 7$  kcal/mol) than the one corresponding to the neutral path ( $\sim 14$  kcal/mol). This effect is mainly explained by a more stable transition state for the anionic pathway, due to specific interactions with water molecules (Bringas et al., 2016).

The available experimental data and this analysis allowed us to propose an overall reaction mechanism for *HNO* decomposition that can be written as a consecutive reactions scheme with

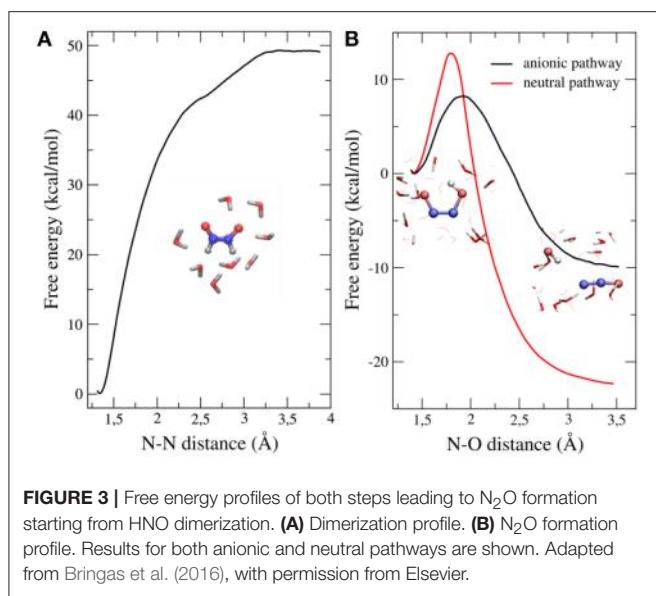


**FIGURE 2 |** Representation of AfCopA QM/MM system. The QM region is defined by the atoms in the reactive region, which includes the phosphates of ATP, Mg<sup>2+</sup>, the aspartic 424 (which is phosphorylated) and part of the atoms of lysine 600. Adapted with permission from Nitsche et al. (2014). Copyright 2014 American Chemical Society.

**TABLE 2 |** *LIO* RT-TDDFT performance for selected systems.

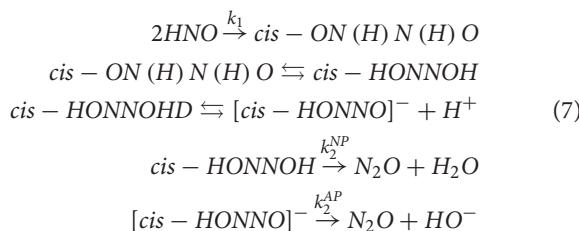
System	Atoms	Full electron			ECP			Speedup
		Time Step (a.u.)	Average time per step (s)	fs/day	Time Step (a.u.)	Average time per step (s)	fs/day	
CH <sub>4</sub>	5	0.076	0.022	7624	0.52	0.015	70500	9.25
CCl <sub>4</sub>	5	0.0095	0.030	680	0.16	0.022	15000	21.66
Cl <sub>4</sub>	5	0.00075	0.059	24	0.16	0.022	15144	586.46
[Fe(CN) <sub>6</sub> ] <sup>4-</sup>	13	0.0038	0.141	48	0.0072	0.099	152	3.13
[Ru(CN) <sub>6</sub> ] <sup>4-</sup>	13	0.0011	0.178	17	0.018	0.108	348	20.58

Effective core potentials and full-electron calculations were carried out with CEP and DZVP basis sets, respectively. Data from Foglia et al. (2017).



**FIGURE 3 |** Free energy profiles of both steps leading to  $\text{N}_2\text{O}$  formation starting from  $\text{HNO}$  dimerization. **(A)** Dimerization profile. **(B)**  $\text{N}_2\text{O}$  formation profile. Results for both anionic and neutral pathways are shown. Adapted from Bringas et al. (2016), with permission from Elsevier.

quick acid-base equilibria connecting products and reactants of different steps:



where  $k_1$ ,  $k_2^{\text{NP}}$ , and  $k_2^{\text{AP}}$  are the constants for the dimerization and for the dissociation via the neutral and anionic pathway elemental steps, respectively. Using transition state theory and the predicted  $pK_a$ s for the dimer intermediate (Fehling and Friedrichs, 2011), we estimated that the anionic pathway will be the only one operative near physiological pH.

In summary, the detailed molecular description of the reaction mechanism at the QM/MM level, showed that specific interactions between the reactive species and the water molecules turn out to be determinant in the stabilization of transition states, thereby modifying the free energy barriers. We predicted a strong pH-dependence of the overall kinetics of  $\text{N}_2\text{O}$  formation, related with the fraction of reactive species available in solution, in agreement with previous experimental data.

### Protein Catalysis

The investigation of the molecular basis of enzyme catalysis has attracted the attention of many researchers both in the experimental as well as the computational sides. To illustrate the application of our code to this objective, we have chosen hydroperoxides reduction as case study. The reduction of cellular endogenous or exogenous hydroperoxides is a key biochemical reaction associated not only with the cellular redox homeostasis, but also with signaling and regulation processes (Hopkins, 2017).

One of the most important antioxidant mechanisms in biological systems is the reduction of peroxides through the oxidation of low molecular weight thiols (LMW) and/or reactive cysteine (Cys) residues in proteins:

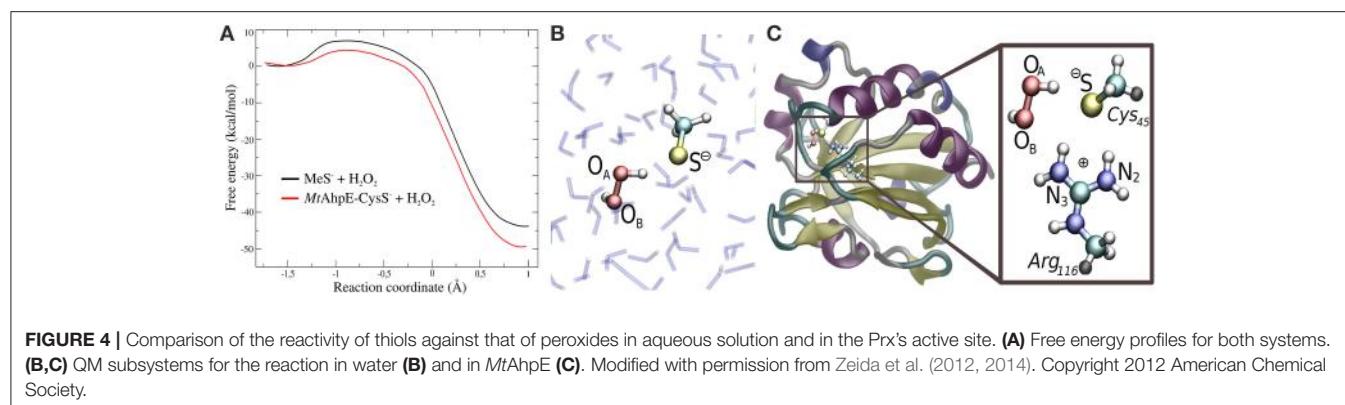


where the reactive species are the thiolate and the hydroperoxide. Although the reaction of peroxides with LMW thiols is usually slow ( $\sim 10 \text{ M}^{-1} \text{ s}^{-1}$  for  $\text{H}_2\text{O}_2$ ) (Winterbourn and Metodiewa, 1999), some enzyme thiols react several orders of magnitude faster in terms of second order rate constants. Among these Cys-based peroxidases, peroxiredoxins (Prx) are the most significant ones, given their reactivity, distribution, and concentration (Hofmann et al., 2002; Winterbourn, 2008).

In recent years, we have applied QM/MM techniques to shed light on the molecular determinants that govern this extremely important biochemical reaction, comparing the reactivity of LMW thiols with Prx, and assessing also different hydroperoxides (Zeida et al., 2012, 2013, 2014, 2015). In particular, aiming to gain microscopic insight onto the Prxs active site's properties that could explain the catalytic effect of these systems, we performed QM/MM MDs to determine the free energy profiles of reaction 8 for  $\text{H}_2\text{O}_2$ , with the thiolate being  $\text{CH}_3\text{S}^-$  or the reactive Cys of the alkyl hydroperoxide reductase E from *Mycobacterium tuberculosis* (*MtAhpE*), the 1-Cys Prx of the mycobacteria genome as a Prx model (Zeida et al., 2012, 2014). The umbrella sampling approach was applied, choosing the reaction coordinate as the difference between the  $\text{O}_\text{A}-\text{O}_\text{B}$  and  $\text{S}-\text{O}_\text{A}$  distances (Figure 4A). We have also measured the activation thermodynamics parameters of the catalyzed reaction by means of temperature dependence stopped-flow kinetics experiments and the Eyring's formalism (Table 3).

**Figure 4A** shows the energy profile for the uncatalyzed and catalyzed processes, while **Figures 4B,C** provide a schematic representation of the QM subsystem for each case. The corresponding  $\Delta G^\ddagger$ s of  $\sim 8$  and  $\sim 4$  kcal/mol turn out to be underestimated in comparison with the one determined experimentally (see Table 3), which can be possibly attributed to the flaws of DFT at the GGA level for determining activation energies (Zhao and Truhlar, 2008). However, the catalytic effect, meaning the difference between the activation free energies ( $\Delta\Delta G^\ddagger$ ), is about  $\sim 4$  kcal/mol, consistently with the experimentally determined  $\Delta\Delta G^\ddagger = 5.4$  kcal/mol and with the  $\sim 5000$ -fold increase in reactivity observed (Luo et al., 2005; Hugo et al., 2009).

It is worth noticing that these profiles are not exactly the same as those published earlier (Zeida et al., 2012, 2014), due to improvements in our computing capabilities and significant advances in sampling protocols. Both new profiles display lower free energy barriers, which is expected for better explorations of the free energy landscape. Nevertheless, the reaction mechanism and properties along the reaction coordinate do not show significant changes, and, most importantly, the  $\Delta\Delta G^\ddagger$  remain unaffected, and so the catalytic effect is still being reproduced in fair agreement with the experimental data.



**FIGURE 4** | Comparison of the reactivity of thiols against that of peroxides in aqueous solution and in the Prx's active site. **(A)** Free energy profiles for both systems. **(B,C)** QM subsystems for the reaction in water **(B)** and in *MtAhpE* **(C)**. Modified with permission from Zeida et al. (2012, 2014). Copyright 2012 American Chemical Society.

**TABLE 3** | Experimental kinetic and thermodynamics activation parameters for the reduction of H<sub>2</sub>O<sub>2</sub> by free Cys and *MtAhpE*.

<i>k</i> <sub>2</sub> (M <sup>-1</sup> s <sup>-1</sup> ) <sup>a</sup>	<i>E</i> <sub>act</sub> (kcal/mol)	$\Delta H^\ddagger$ (kcal/mol)	$\Delta S^\ddagger$ (cal/K.mol)	$\Delta G^\ddagger$ (kcal/mol)
Free Cys <sup>b</sup>	14.9	17.0	16.4 (0.3)	1.7 (1.1)
<i>MtAhpE</i> -Cys <sub>45</sub> <sup>c</sup>	$8.2 \times 10^4$ (1.5)	5.7 (0.7)	4.8 (0.5)	-19.1 (1.9)

Standard deviations in parenthesis when available.

<sup>a</sup>pH-independent rate constants.

<sup>b</sup>data taken from Luo et al. (2005).

<sup>c</sup>data taken from Zeida et al. (2014).

The exploration of the reaction coordinate allow us to identify key events during the reaction to explain the differences in the  $\Delta G^\ddagger$ s. Specifically, the strong interactions of the thiolate and the peroxide with Arg<sub>116</sub> and Thr<sub>42</sub> residues, which are extremely conserved among the Prx family (Soito et al., 2011; Perkins et al., 2015), are the main factors responsible for the transition state stabilization and the concomitantly significant reduction in  $\Delta H^\ddagger$ , which in turn results in a decrease in  $\Delta G^\ddagger$  in spite of the unfavorable entropic contribution (Table 3). Our calculations support the idea of a bimolecular nucleophilic S<sub>N</sub>2 type substitution mechanism, with an internal proton transfer and no acid-base catalysis. The catalytic ability of Prxs lies on the stabilization of the transition state due to an active site design that configures a complex H-bond network activating both reactive species, the thiolate and the peroxide.

## Spectroscopic Studies

UV-vis or FTIR spectrophotometers are ubiquitous tools in research laboratories, routinely employed to characterize new species, to establish the identity of a chemical systems, to perform both thermal and photochemical reactivity studies, or even to analyse the presence of impurities in a given sample. Ambiguities in the characterization of unknown species are often present, and so experimental chemists are increasingly relying on theoretical calculations to complement their measurements.

The vast majority of the computational studies addressing electronic or vibrational spectra involve the geometry optimization of the molecule at a given level of theory, and the study of its spectroscopic properties at those frozen nuclear coordinates. Within this framework, the emulation of the environment is carried out by different approaches,

as the Polarizable Continuum Model (PCM) (Tomasi et al., 2005), often placing a few explicit solvent molecules in some fixed position in the first solvation shell interacting with the chromophore (Zuehlsdorff et al., 2016) or, more recently, using a classical and explicit description of the solvent by employing a QM/MM hybrid Hamiltonian (Barone et al., 2010). The most common approach to calculate infrared spectra is the harmonic oscillator approximation, through the diagonalization of the Hessian matrix at different levels of theory. This process yields the vibrational modes, their characteristic frequencies, and an estimation of the intensity of those bands in the infrared spectrum (Bloino et al., 2016). However, this scheme may present flaws in cases in which there are strong anharmonicities or specific solute-solvent interactions. The IR spectrum can also be computed directly from a molecular dynamics simulation including explicit solvent molecules. The absorption spectrum can be determined as the Fourier Transform of the temporal autocorrelation function of the dipole moment (Futrelle and McGinty, 1971; McQuarrie, 1976):

$$I(\omega) = \frac{1}{2} \int dt e^{i\omega t} \langle \vec{\mu}(0) | \vec{\mu}(t) \rangle \quad (9)$$

In the previous equation, the brackets denote an equilibrium ensemble average, which can be obtained from molecular dynamics simulations:

$$\langle \vec{\mu}(0) | \vec{\mu}(t) \rangle \approx \frac{t - t_i}{\Delta t} \sum_i \vec{\mu}(t_i) \vec{\mu}(t_i + t) \quad (10)$$

where  $\Delta t$  is the integration time-step. Similarly, the Fourier Transform of the temporal autocorrelation function of the polarizability yields the Raman spectrum of a given system.

In the case of the electronic spectra, Time Dependent Density Functional Theory in the linear response formulation (LR-TDDFT) has become in the last two decades the most popular approach to predict the electronic excitation frequencies of middle-size systems, due to its modest computational cost and relatively good performance in comparison with highly correlated schemes (Runge and Gross, 1984; Marques et al., 2006). Alternatively, it is also possible to compute an electronic spectrum from electron dynamics simulations, through real-time TDDFT. This methodology, which is incorporated in *LIO*, is not so commonly used to calculate electronic frequencies. This is related to the fact that the electron dynamics simulations needed to recover the absorption frequencies with RT-TDDFT require to excite and propagate the density matrix of the electronic system for tens or hundreds of femtoseconds, and then perform an *ad-hoc* post-processing analysis. This scheme tends to be more demanding, and somehow prevents RT-TDDFT from being used in a black-box format. However, the real-time approach exhibits some appealing features in comparison with LR-TDDFT, that include the possibility to study intense perturbations beyond the linear response regime (Marques et al., 2006; Lopata and Govind, 2011), the scaling of the computation burden with respect to the number of electrons, that can be made quasi-linear, or the avoidance of the computation of the exchange-correlation kernel. Therefore, RT-TDDFT could be a competitive alternative to the linear response method, especially in the case of large systems, since the computation time in typical LR-TDDFT implementations grows as  $N^3$  or  $N^4$ .

Beyond the applied methodology and the level of sophistication of the electronic structure calculations, one of the main causes of the failures in the prediction of spectroscopic properties are due to the consideration of a frozen nuclear geometry, obtained by an initial optimization, thus neglecting thermal fluctuations. A proper exploration of configurational space produces spectra averaged on the degrees of freedom of both the chromophore and its environment, providing smooth lineshapes instead of the discrete, multiple-lines spectra corresponding to the particular transitions energies obtained from a single geometry. The use of molecular dynamics or Monte Carlo simulations where the spectra are obtained from statistical averages within an ensemble has become an increasingly chosen strategy for different types of systems (Valsson et al., 2013).

### Vibrational Spectra of Simple Aqueous Species: Early Results and Benchmarks

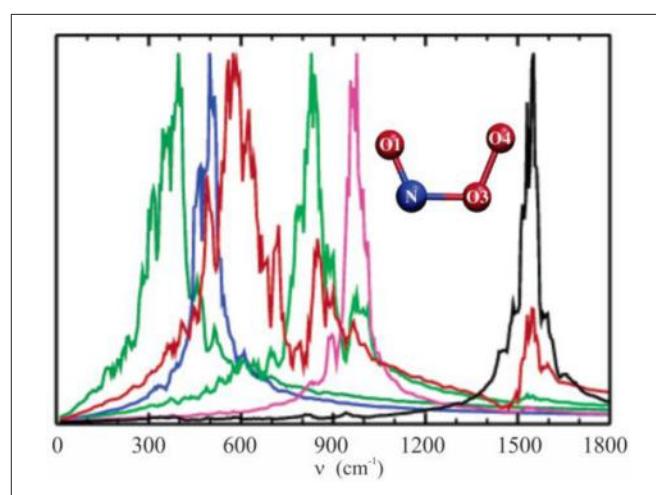
In 2005 we investigated the structure and the vibrational spectrum of peroxy nitrite anion in solution via QM/MM molecular dynamics simulations, calculating the temporal autocorrelation functions of the atomic velocities (Gonzalez Lebrero et al., 2005). At that time, the computation capabilities allowed us to perform production dynamics of 20 ps in a reasonable time window. This study provided a picture of the complex interactions between the  $\text{ONOO}^-$  anion and the solvent, and how these were reflected in the vibrational spectrum in solution, as shown in **Figure 5**. Our results yielded frequency values much closer to the experimental ones than those obtained using standard methodologies, and also helped to assign a

controversial band centered at  $642 \text{ cm}^{-1}$  as corresponding to  $\text{NO}_3$  stretching.

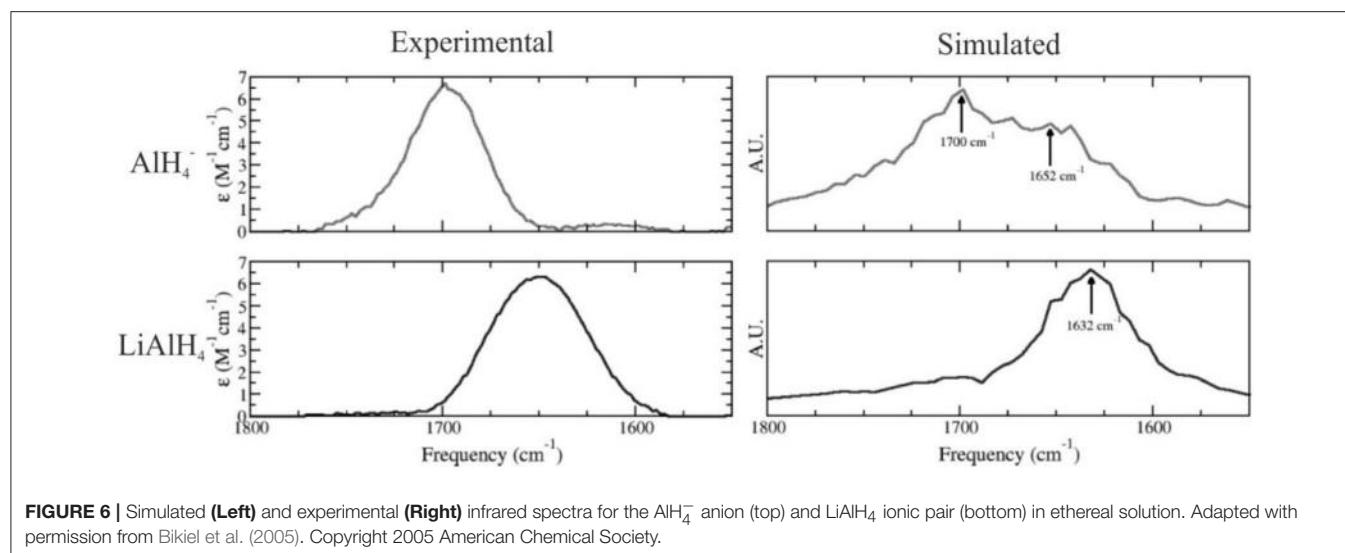
The nature of the solute species present in ethereal solutions of  $\text{LiAlH}_4$  is of crucial importance to understand the mechanism for the reduction of ketones and other functional groups by  $\text{LiAlH}_4$ . In 2005, we have employed a combination of experimental and theoretical techniques to investigate the structure of this system in ethereal solutions, using QM/MM simulations in which  $\text{LiAlH}_4$  was modeled at the DFT PBE level using dzvp basis sets, and the solvent was described using a three site potential (Bikiel et al., 2005). Our results were consistent with a dissociation equilibrium displaced to the associated species. However, a significant amount of the dissociated species is also expected to exist. We calculated the infrared spectra for both  $\text{LiAlH}_4$  and  $\text{AlH}_4^-$  species performing molecular dynamics simulations, reproducing the main features of the experimental spectra, as shown in **Figure 6**.

### The Electronic Spectrum and the Sampling Issue

In 2014 we have developed a powerful scheme to perform real-time TDDFT electron dynamics in a QM/MM framework (Morzan et al., 2014). This implementation can easily handle quantum subsystems in the order of 100 atoms surrounded by thousands of classical nuclei, enabling the investigation of the effect of a complex environment, such as a solvent or a protein matrix, on the UV-vis spectra of molecular systems. Our starting point was the validation on isolated species by comparison of the absorption maxima obtained using the real time and the linear response methodologies (some of these results are presented in **Table 4**). These molecules were chosen as a benchmark to test our code and verify that the



**FIGURE 5** | Vibrational density of states computed as the Fourier transform of the velocity autocorrelation function, evaluated in the isolated species normal modes coordinates. Molecular dynamics simulations were done with the PBE functional and the TIP4P model for water.  $\text{NO}_1$  stretching, black line;  $\text{O}_1\text{NO}_3$  bending, pink line;  $\text{O}_3\text{O}_4$  stretching, green line;  $\text{NO}_3$  stretching, red line;  $\text{OONO}$  torsion, blue line;  $\text{NO}_3\text{O}_4$  bending, dark green line. Adapted with permission from Gonzalez Lebrero et al. (2005). Copyright 2005 American Chemical Society.



**FIGURE 6 |** Simulated (Left) and experimental (Right) infrared spectra for the  $\text{AlH}_4^-$  anion (top) and  $\text{LiAlH}_4$  ionic pair (bottom) in ethereal solution. Adapted with permission from Bikiel et al. (2005). Copyright 2005 American Chemical Society.

obtained spectra were in very good agreement with experimental results.

The effect of the environment becomes crucial within proteins and here is where the QM/MM methodologies make the difference. In the work mentioned above, the spectrum of the CO hexa-coordinated heme group in Flavohemoglobin of *Escherichia Coli* (*Ec*FlavoHb) was analyzed. The results showed that the Soret band of the chromophore experiences a notorious blue shift ( $\Delta\lambda \sim 35 \text{ nm}$ ) when going from vacuum to the protein environment. Despite the lack of a direct experimental comparison, the observed shift in the *Ec*FlavoHb heme with respect to the gas phase is in accordance with the expected trend.

The examples given so far correspond to electronic spectra extracted from electron dynamics simulations on a single geometry. To generate a realistic electronic spectrum that considers thermal effects, we calculate the spectra of a set of configurations extracted from a QM/MM molecular dynamics trajectory. These configurations must be separated from one another by a time frame  $\eta$  long enough as to decorrelate the molecular vibrations (for simple molecular systems in solution it is usually enough with  $\eta = 5\text{--}10 \text{ ps}$ ). The number of configurations needed to get a converged spectrum in solution is typically of a few dozens. A healthy practice is to combine, i.e., to intercalate, the QM/MM molecular dynamics simulations with some purely classical sampling, especially in systems with many degrees of freedom and multiple solvation structures, to amplify the exploration and avoid the system to get trapped around a local minimum. **Figure 7** shows a typical convergence sequence in the computation of an electronic spectrum using QM/MM dynamics.

As seen in **Figure 7**, it is of a huge importance to perform an adequate sampling of the nuclear configurations visited along the dynamics to obtain a spectrum with the absorption maximum located in the correct position and with the correct band shapes. Unfortunately, there is no magic number associated with the required amount of nuclear configurations to reach convergence.

**TABLE 4 |** Calculated and experimental absorption maximum ( $\lambda_{\max}$  (nm)) of different molecules.

Compound	RT-TDDFT	LR-TDDFT	Experimental
$\text{H}_2\text{O}$	161	159	170
CO	147	143	145
$\text{C}_6\text{H}_6$	175	172	180
Tryptophan	298	311	286

Real-time and linear-response TDDFT calculations were performed with the PBE exchange-correlation functional and *dzvp* basis set.

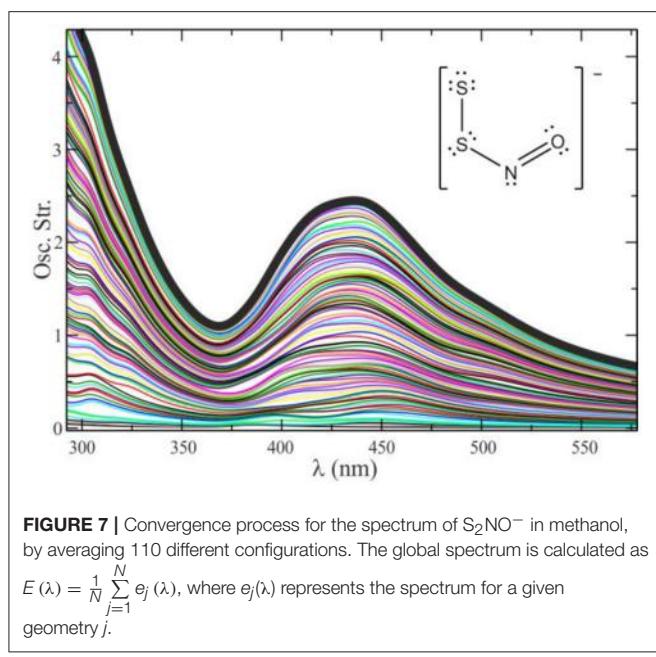
This number is usually tied to several factors, mainly the rigidity of the molecular system under study (and therefore the number of accessible local minima) and the magnitude and lability of the interactions between the solute and the solvent.

**Figure 8** shows that the convergence of the electronic spectrum of the  $(\text{HO})\text{NS}_2$  molecule in acetonitrile is achieved by averaging less than 10 spectra taken at snapshots separated by 5 ps of QM/MM dynamics, while for more flexible species like  $\text{S}_4^{2-}$  (with high rotational freedom) said process is not satisfactorily converged after averaging 70 spectra.

Ensuring that a spectrum is statistically converged is not only crucial to find the correct values of  $\lambda_{\max}$ ; it is also important because, if not converged, the spectra may present shoulders similar to those found in experiments, thus leading to incorrect interpretations.

#### Solving Specific Questions: The Case of the “Crosstalk” Between NO and $\text{H}_2\text{S}$

We will illustrate the capabilities of our code, by showing results related to the crosstalk chemistry between NO and  $\text{H}_2\text{S}$ . In addition to the diverse biological roles of nitric oxide (NO) and hydrogen sulfide ( $\text{H}_2\text{S}$ ), there is a growing appreciation that both molecules have interdependent biological actions resulting in either mutual attenuation or potentiation responses, the so-called NO/ $\text{H}_2\text{S}$  “cross-talk” (Marcolongo et al., 2017).



The study of the interaction pathways between these two molecules has led to a large number of publications throughout this decade, and there have been many controversies around the appearance and assignation of spectroscopic signals in different experiments. In particular, there has been much debate around the appearance of a transient signal at  $\sim 409$  nm produced in the reaction of S-nitrosoglutathione (GSNO) with  $\text{HS}^-$  at  $\text{pH} = 7.4$  (Filipovic et al., 2012). Filipovic and colleagues argue that this signal corresponds to the presence of a mixture of polysulfides while other groups claim that the yellow signal corresponds to perthionitrite ( $\text{S}_2\text{NO}^-$ ), the sulfur analog of peroxy nitrite (Seel and Wagner, 1988; Munro and Williams, 2000; Filipovic et al., 2012; Cortese-Krott et al., 2015; Bailey et al., 2016; Bolden et al., 2016). This anion has been well-characterized taking part of solid salts and in organic solvents (Seel et al., 1985; Filipovic et al., 2012; Wedmann et al., 2015), though their clear identification and chemistry in aqueous solutions remains elusive.

In 2016, we tested the use of QM/MM dynamics in conjunction with the application of RT-TDDFT implemented in *LIO* to try to help in the elucidation of this controversial signal, tilting the balance toward the proposal of  $\text{SSNO}^-$  as the responsible for the yellow signal (Marcolongo et al., 2016). Our calculations for this species (together with a large number of related benchmark molecules) showed that  $\text{SSNO}^-$  is a good candidate to absorb in that region and also reproduces quantitatively the solvatochromic shift of the absorption band while going from water to a set of organic solvents where the species is well-characterized. The QM/MM dynamics performed by our group showed that the specific interactions between  $\text{SSNO}^-$  and the solvent are responsible for the modulation of the  $\theta(\text{N}_1\text{-S}_1\text{-S}_2)$  angle which strongly affects the spectroscopic properties of this molecule, as shown in Table 5. For these simulations, a TIP3P potential was utilized to describe the water molecule and the force fields for acetonitrile,

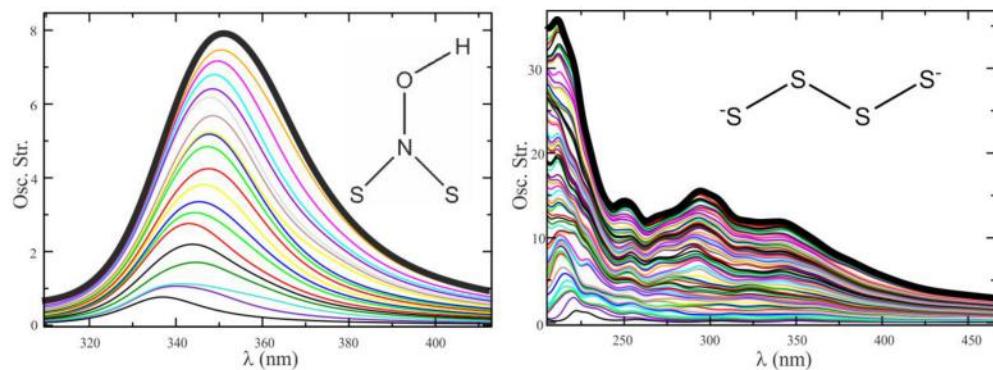
acetone and methanol were generated following the restricted electrostatic potential (RESP) technique and DFT calculations at the PBE/dzvp level. Equilibrium distances and angles, as well as force constants were computed using the same electronic structure scheme.

The results obtained for this system showed how the statistical study of the interactions resulting from the QM/MM sampling is the key to obtain results that can be compared to the experimental values almost quantitatively.

## CONCLUDING REMARKS AND PERSPECTIVES

In this article, we have provided an overview on the basic features of the *LIO* program, focusing on some representative applications. This review has emphasized the fundamental importance of the code performance in the study of chemical reactivity and free energy profiles, and in the simulation of vibrational and electronic spectra. Currently, most QM/MM DFT programs are capable of handling molecular systems having, in the quantum and classical domains, in the order of a few hundreds and a few thousands of atoms, respectively. However, the computational cost associated with systems of this size restrains their applicability to single-point calculations or partial geometry optimizations, which may be helpful to analyze some structural or energetic aspects at zero temperature, whereas the kind of applications considered in this article require extensive sampling. A data point along a reaction coordinate profile, or the construction of an auto-correlation function to extract a vibrational spectrum, typically require molecular dynamics simulations in the order of at least several picoseconds. To reach these time-scales, the cost per iteration cannot exceed a few seconds. The goal in developing the *LIO* program is not to save computation time in the calculation of a given property, but to make it feasible when it was not. Thanks to recent advances in the code, it is currently possible to study models containing about 30 QM atoms in the quantum region for time windows in the order of the nanosecond, using a single commercial GPU.

In *LIO*, the most expensive parts of the QM/MM DFT scheme, including the calculation of the exchange-correlation energy, Coulomb interactions, and forces, have been thoroughly optimized, reaching an almost linear-scaling performance. As a consequence of these improvements, those operations typically inexpensive start to become the new limiting steps in the overall speed. One example of this is the diagonalization of the Hamiltonian, which, except in the case of very large systems, takes up a negligible fraction of the SCF iteration cost in Gaussian basis codes. At the present stage, diagonalization, which scales as  $N^3$ , represents the next potential bottleneck for big systems. For this reason, *LIO* is not a “linear-scaling” implementation, although it does behave approximately linearly within a certain size range, when the diagonalization still does not contribute appreciably to the computational load. However, the diagonalization dominates the overall performance in the case of very large systems exceeding 200 or 300 atoms in the quantum



**FIGURE 8 |** Convergence sequence for the spectrum of (HO)NS<sub>2</sub> in acetonitrile (**Left**) and S<sub>4</sub><sup>2-</sup> in water (**Right**). Ten configurations provide a converged profile in one case, whereas more than 70 do not seem enough in the other.

**TABLE 5 |** Calculated and experimental absorption maximums for S<sub>2</sub>NO<sup>-</sup> in different solvents, as well as the mean value of the  $\theta(N_1-S_1-S_2)$  angle.

Solvent	$\theta(N_1-S_1-S_2)$ (°)	$\lambda_{\text{max}}$ (nm) (experimental)	$\lambda_{\text{max}}$ (nm) (RT-TDDFT)	Number of averaged spectra
Water	111	409–412	411	54
Acetone	113	448	458	43
Acetonitrile	117	450	458	63
Methanol	117	431	425	110

region. Thus, to go beyond these sizes, the implementation of an order  $N$  diagonalization algorithm should be made available.

As discussed in the Examples section, a linear-scaling behavior is particularly appealing for the computation of electronic spectra of large molecules. Presently, most UV-visible spectroscopy calculations involve TDDFT in the linear response formulation. This implementation, typically scaling as  $N^3$  or  $N^4$ , is often easier to use in comparison with the real time approach, and more efficient than this one for systems of small and moderate size. Nevertheless, for large systems in complex environments, real-time TDDFT simulations with an approximately linear scaling might be a smart alternative to surpass the size limitations of the linear response method. Besides, the feasibility to sample the structural degrees of freedom in solution or in a biomolecule with the same Hamiltonian employed for the electron dynamics, constitutes an asset of the *LIO* code.

Aside from spectroscopic applications, real-time TDDFT offers the possibility to perform quantum transport simulations. Recently, we have implemented in *LIO* an approach to compute molecular conductance in open quantum conditions (Morzan et al., 2017). The combination of RT-TDDFT with a QM/MM framework provides a unique platform to investigate challenging phenomena related to electron dynamics in realistic environments, which are very difficult to address from the experimental side or even with other modeling strategies. In our group, we are studying the electron transfer dynamics to and from the CuA active site in cytochrome C oxidase. Charge transport between redox sites in an enzyme is a fascinating phenomenon that might be studied with our approach, providing

that the number of atoms involved in the likely paths foreseen for the electrons is not excessive to be treated within the QM region. The conductance of conjugated polymers in a disordered matrix, or the electron exchange at an electrochemical interface, are also problems that could be explored using the present approach, where the solvent, the counterions, or, more generally, the surrounding media, can be described at the MM level.

The use of pseudopotentials, recently made available in the code (Foglia et al., 2017), is fundamental to perform real time TDDFT simulations with transition metals and heavy atoms. The time-propagation of the core electrons, with high characteristic frequencies, demands integration time-steps that may be orders of magnitude below those required for the integration of the valence density. Thus, all-electron TDDFT dynamics with transition metals or atoms below the second period are extremely costly because of the small time-steps necessary for energy conservation. The incorporation of pseudopotentials has relaxed this constraint, making it feasible the electron dynamics simulations of the active sites of metalloenzymes or metallic wires. The pseudopotential implementation of the forces, necessary to perform MD simulations, is currently underway.

To summarize, the *LIO* code is a very active project and a valuable resource, open to the community via github (<https://github.com/MALBECC/LIO>), to perform DFT molecular dynamics and real time TDDFT simulations in a QM/MM framework. Beyond those objectives concerning performance, which have concentrated much of the efforts of recent years, there is presently a strong drive toward the development of new capabilities. Some of the features or implementations currently in progress are:

- the linear response TDDFT method;
- pseudopotential forces;
- intensities in resonant Raman spectra;
- exact exchange terms for hybrid DFT in GPU;
- schemes to compute quantum transport;
- Ehrenfest dynamics;
- coupling with libxc (Marques et al., 2012) in order to extend the variety of available DFT functionals.

Hopefully, this review has made it clear how much realistic chemical and spectroscopic predictions rely on sampling efficiency, necessary to introduce both thermal and environment effects. This is the spirit that has guided the evolution of LIO, which has been optimized for QM/MM simulations where the typical size of the QM domain reaches a few tens of atoms. Perhaps the most important challenge for the development of molecular simulation methods is to keep up with the advances in HPC platforms, algorithmic optimization, and theory. To take the proper advantage from all these worlds is only possible through a collaborative endeavor involving theoretical chemists and computer scientists, with the continuous feedback from the experimental side.

## REFERENCES

- Bailey, T. S., Henthorn, H. A., and Pluth, M. D. (2016). The intersection of NO and H<sub>2</sub>S: persulfides generate NO from nitrite through polysulfide formation. *Inorg. Chem.* 55, 12618–12625. doi: 10.1021/acs.inorgchem.6b01660
- Barone, V., Małgorzata, B., and Giuseppe, B. (2010). Extending the range of computational spectroscopy by QM/MM approaches: time-dependent and time-independent routes. *Adv. Quantum Chem.* 59, 17–57. doi: 10.1016/S0651-3276(10)59002-6
- Becke, A. D. (1988). A multicenter numerical integration scheme for polyatomic molecules. *J. Chem. Phys.* 88, 2547–2553. doi: 10.1063/1.454033
- Bikiel, D. E., Di Salvo, F., González Lebrero, M. C., Doctorovich, F., and Estrin, D. A. (2005). Solvation and structure of LiAlH<sub>4</sub> in ethereal solvents. *Inorg. Chem.* 44, 5286–5292. doi: 10.1021/ic050330q
- Bloino, J., Baiardi, A., and Biczysko, M. (2016). Aiming at an accurate prediction of vibrational and electronic spectra for medium-to-large molecules: an overview. *Int. J. Quantum Chem.* 116, 1543–1574. doi: 10.1002/qua.25188
- Bolden, C., King, S. B., and Kim-Shapiro, D. B. (2016). Reactions between nitrosopersulfide and heme proteins. *Free Radic. Biol. Med.* 99, 418–425. doi: 10.1016/j.freeradbiomed.2016.09.005
- Bringas, M., Semelak, J., Zeida, A., and Estrin, D. A. (2016). Theoretical investigation of the mechanism of nitroxyl decomposition in aqueous solution. *J. Inorg. Biochem.* 162, 102–108. doi: 10.1016/j.jinorgbio.2016.06.016
- Carvalho, A. T., Barrozo, A., Doron, D., Kilshtain, A. V., Major, D. T., and Kamerlin, S. C. L. (2014). Challenges in computational studies of enzyme structure, function and dynamics. *J. Mol. Graph. Model.* 54(Suppl. C), 62–79. doi: 10.1016/j.jmgm.2014.09.003
- Chipot, C., and Pearlman, D. A. (2002). Free energy calculations. The long and winding gilded road. *Mol. Simul.* 28, 1–12. doi: 10.1080/08927020211974
- Chipot, C., and Pohorille, A. (2007). *Free Energy Calculations: Theory and Applications in Chemistry and Biology*. Berlin; Heidelberg: Springer-Verlag.
- Cortese-Krott, M. M., Kuhnle, G. G. C., Dyson, A., Fernandez, B. O., Grman, M., DuMond, J. F., et al. (2015). Key bioactive reaction products of the NO/H<sub>2</sub>S interaction are S/N-hybrid species, polysulfides, and nitroxyl. *Proc. Natl. Acad. Sci. U.S.A.* 112, E4651–E4660. doi: 10.1073/pnas.1509277112
- Elola, M. D., Estrin, D. A., and Laria, D. (1999). Hybrid quantum classical molecular dynamics simulation of the proton-transfer reaction of HO- with HBr in aqueous clusters. *J. Phys. Chem. A* 103, 5105–5112. doi: 10.1021/jp983678y
- Estrin, D. A., Corongiu, G., and Clementi, E. (1993). *Methods and Techniques in Computational Chemistry*. Cagliari: Stef.
- Fehling, C., and Friedrichs, G. (2011). Dimerization of HNO in aqueous solution: an interplay of solvation effects, fast acid-base equilibria, and intramolecular hydrogen bonding? *J. Am. Chem. Soc.* 133, 17912–17922. doi: 10.1021/ja2075949
- Filipovic, M. R., Miljkovic, J. L., Nauser, T., Royzen, M., Klos, K., Shubina, T., et al. (2012). Chemical characterization of the smallest S-Nitrosothiol, HSNO; cellular cross-talk of H<sub>2</sub>S and S-Nitrosothiols. *J. Am. Chem. Soc.* 134, 12016–12027. doi: 10.1021/ja3009693
- Foglia, N. O., Morzan, U. N., Estrin, D. A., Scherlis, D. A., and Gonzalez Lebrero, M. C. (2017). Role of core electrons in quantum dynamics using TDDFT. *J. Chem. Theor. Comput.* 13, 77–85. doi: 10.1021/acs.jctc.6b00771
- Futrelle, R. P., and McGinty, D. J. (1971). Calculation of spectra and correlation functions from molecular dynamics data using the fast Fourier transform. *Chem. Phys. Lett.* 12, 285–287. doi: 10.1016/0009-2614(71)85065-0
- Godbout, N., Salahub, D. R., Andzelm, J., and Wimmer, E. (1992). Optimization of Gaussian-type basis sets for local spin density functional calculations. Part I. Boron through neon, optimization technique and validation. *Can. J. Chem.* 70, 560–571. doi: 10.1139/v92-079
- Gonzalez Lebrero, M. C., Perissinotti, L. L., and Estrin, D. A. (2005). Solvent effects on peroxy nitrite structure and properties from QM/MM simulations. *J. Phys. Chem. A* 109, 9598–9604. doi: 10.1021/jp0542241
- Hofmann, B., Hecht, H. J., and Flohé, L. (2002). Peroxiredoxins. *Biol. Chem.* 383, 347–364. doi: 10.1515/BC.2002.040
- Hopkins, R. Z. (2017). Hydrogen peroxide in biology and medicine: an overview. *Reactive Oxygen Species* 3, 26–37. doi: 10.20455/ros.2017.809
- Hu, H., and Yang, W. (2008). Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. *Annu. Rev. Phys. Chem.* 59, 573–601. doi: 10.1146/annurev.physchem.59.032607.093618
- Hugo, M., Turell, L., Manta, B., Botti, H., Monteiro, G., Netto, L. E., et al. (2009). Thiol and sulfenic acid oxidation of AhpE, the one-cysteine peroxiredoxin from *Mycobacterium tuberculosis*: kinetics, acidity constants, and conformational dynamics. *Biochemistry* 48, 9416–9426. doi: 10.1021/bi901221s
- Kollman, P. (1993). Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* 93, 2395–2417. doi: 10.1021/cr00023a004
- Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., et al. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78, 1950–1958. doi: 10.1002/prot.22711
- Lopata, K., and Govind, N. (2011). Modeling fast electron dynamics with real-time time-dependent density functional theory: application to small molecules and chromophores. *J. Chem. Theor. Comput.* 7, 1344–1355. doi: 10.1021/ct200137z
- Luo, D., Smith, S. W., and Anderson, B. D. (2005). Kinetics and mechanism of the reaction of cysteine and hydrogen peroxide in aqueous solution. *J. Pharm. Sci.* 94, 304–316. doi: 10.1002/jps.20253
- Ma, X. L., Gao, F., Liu, G.-L., Lopez, B. L., Christopher, T. A., Fukuto, J. M., et al. (1999). Opposite effects of nitric oxide and nitroxyl on postischemic myocardial injury. *Proc. Natl. Acad. Sci. U.S.A.* 96, 14617–14622. doi: 10.1073/pnas.96.25.14617
- Marcolongo, J. P., Morzan, U. N., Zeida, A., Scherlis, D. A., and Olabe, J. A. (2016). Nitrosodisulfide [S<sub>2</sub>NO]<sup>-</sup> (perthionitrite) is a true intermediate during the “cross-talk” of nitroxyl and sulfide. *Phys. Chem. Chem. Phys.* 18, 30047–30052. doi: 10.1039/C6CP06314D

## AUTHOR CONTRIBUTIONS

AZ, JM, JS, NF, and UM selected and analyzed the data. AZ, JM, MG, DE, and DS wrote the paper. MG, DE, and DS supervised research.

## FUNDING

This research was supported by grants of the Universidad de Buenos Aires, UBACYT 20020130100097BA and Agencia Nacional de Promoción Científica y Tecnológica, PICT 2015-0672, PICT 2014-1022, PICT 2015-2761, and CONICET grant 11220150100303CO. AZ, JS, NF, and JM gratefully acknowledge CONICET for fellowships.

- Marcolongo, J. P., Zeida, A., Slep, L. D., and Olabe, J. A. (2017). Thionitrous acid/thionitrite and perthionitrite intermediates in the “Crosstalk” of NO and H<sub>2</sub>S, in *Adv. Inorg. Chem.* 70, 277–309. doi: 10.1016/bs.adioch.2017.02.002
- Marques, M. A. L., Oliveira, M. J. T., and Burnus, T. (2012). Libxc: a library of exchange and correlation functionals for density functional theory. *Comput. Phys. Commun.* 183, 2272–2281. doi: 10.1016/j.cpc.2012.05.007
- Marques, M. A. L., Ullrich, C. A., Nogueira, F., Rubio, A., Burke, K., and Gross, E. K. U. (eds.). (2006). *Time-Dependent Density Functional Theory*. Berlin; Heidelberg: Springer.
- McQuarrie, D. A. (1976). *Statistical Mechanics*. London, UK: HarperCollins Publishing, Inc.
- Miranda, K. M. (2005). The chemistry of nitroxyl (HNO) and implications in biology. *Coord. Chem. Rev.* 249, 433–455. doi: 10.1016/j.ccr.2004.08.010
- Miranda, K. M., Katori, T., Torres De Holding, C. L., Thomas, L., Ridnour, L. A., McLendon, W. J., et al. (2005). Comparison of the NO and HNO donating properties of diazeniumdiolates: primary amine adducts release HNO *in vivo*. *J. Med. Chem.* 48, 8220–8228. doi: 10.1021/jm050151i
- Morzan, U. N., Ramírez, F. F., González Lebrero, M. C., and Scherlis, D. A. (2017). Electron transport in real time from first-principles. *J. Chem. Phys.* 146:044110. doi: 10.1063/1.4974095
- Morzan, U. N., Ramírez, F. F., Oviedo, M. B., Sánchez, C. G., Scherlis, D. A., and Lebrero, M. C. G. (2014). Electron dynamics in complex environments with real-time time dependent density functional theory in a QM-MM framework. *J. Chem. Phys.* 140:164105. doi: 10.1063/1.4871688
- Munro, A. P., and Williams, D. L. H. (2000). Reactivity of sulfur nucleophiles towards S-nitrosothiols. *J. Chem. Soc. Perkin Trans. 2*, 1794–1797. doi: 10.1039/b004415f
- Nitsche, M. A., Ferreria, M., Mocskos, E. E., and González Lebrero, M. C. (2014). GPU accelerated implementation of density functional theory for hybrid QM/MM simulations. *J. Chem. Theor. Comput.* 10, 959–967. doi: 10.1021/ct400308n
- Obara, S., and Saika, A. (1986). Efficient recursive computation of molecular integrals over cartesian gaussian functions. *J. Chem. Phys.* 84, 3963–3974. doi: 10.1063/1.450106
- Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., DeBolt, S., et al. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* 91, 1–41. doi: 10.1016/0010-4655(95)00041-D
- Perkins, A., Nelson, K. J., Parsonage, D., Poole, L. B., and Karplus, P. A. (2015). Peroxiredoxins: guardians against oxidative stress and modulators of peroxide signaling. *Trends Biochem. Sci.* 40, 435–445. doi: 10.1016/j.tibs.2015.05.001
- Runge, E., and Gross, E. K. (1984). Density-functional theory for time-dependent systems. *Phys. Rev. Lett.* 52, 997. doi: 10.1103/PhysRevLett.52.997
- Seel, F., Kuhn, R., Simon, G., and Wagner, M. (1985). PNP-Perthionitrit und PNP-Monothionitrit / PNP-Perthionitrite and PNP-Monothionitrite. *Z. Naturforschung B* 40, 1607–1617. doi: 10.1515/znb-1985-1203
- Seel, F., and Wagner, M. (1988). Reaction of sulfides with nitrogen monoxide in aqueous solution. *ChemInform* 19, 189–192. doi: 10.1002/chin.198824008
- Shafirovich, V., and Lymar, S. V. (2002). Nitroxyl and its anion in aqueous solutions: spin states, protic equilibria, and reactivities toward oxygen and nitric oxide. *Proc. Natl. Acad. Sci. U.S.A.* 99, 7340–7345. doi: 10.1073/pnas.112202099
- Soito, L., Williamson, C., Knutson, S. T., Fetrow, J. S., Poole, L. B., and Nelson, K. J. (2011). PREX: peroxiredoxin classification indEX, a database of subfamily assignments across the diverse peroxiredoxin family. *Nucleic Acids Res.* 39(Suppl. 1), D332–D337. doi: 10.1093/nar/gkq1060
- Stratmann, R. E., Scuseria, G. E., and Frisch, M. J. (1996). Achieving linear scaling in exchange-correlation density functional quadratures. *Chem. Phys. Lett.* 257, 213–223. doi: 10.1016/0009-2614(96)00600-8
- Tomasi, J., Mennucci, B., and Cammi, R. (2005). Quantum mechanical continuum solvation models. *Chem. Rev.* 105, 2999–3093. doi: 10.1021/cr9904009
- Tsuda, T., and Toyoshima, C. (2009). Nucleotide recognition by CopA, a Cu<sup>+</sup>-transporting P-type ATPase. *EMBO J.* 28, 1782–1791. doi: 10.1038/embj.2009.143
- Valsson, O., Campomanes, P., Tavernelli, I., Rothlisberger, U., and Filippi, C. (2013). Rhodopsin absorption from first principles: bypassing common pitfalls. *J. Chem. Theor. Comput.* 9, 2441–2454. doi: 10.1021/ct3010408
- Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005). GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26, 1701–1718. doi: 10.1002/jcc.20291
- Wedmann, R., Zahn, A., Shubina, T. E., Dürr, M., Heinemann, F. W., Bugenhagen, B. E., et al. (2015). Does perthionitrite (SSNO<sup>−</sup>) account for sustained bioactivity of NO? A (bio)chemical characterization. *Inorg. Chem.* 54, 9367–9380. doi: 10.1021/acs.inorgchem.5b00831
- Winterbourn, C. C. (2008). Reconciling the chemistry and biology of reactive oxygen species. *Nat. Chem. Biol.* 4, 278–286. doi: 10.1038/nchembio.85
- Winterbourn, C. C., and Metodiewa, D. (1999). Reactivity of biologically relevant thiol compounds with superoxide and hydrogen peroxide. *Free Radic. Biol. Med.* 27, 322–328. doi: 10.1016/S0891-5849(99)00051-9
- Zeida, A., Babbush, R., González Lebrero, M. C., Trujillo, M., Radi, R., and Estrin, D. A. (2012). Molecular basis of the mechanism of thiol oxidation by hydrogen peroxide in aqueous solution: challenging the S<sub>N</sub>2 paradigm. *Chem. Res. Toxicol.* 25, 741–746. doi: 10.1021/tx200540z
- Zeida, A., González Lebrero, M. C., Radi, R., Trujillo, M., and Estrin, D. A. (2013). Mechanism of cysteine oxidation by peroxynitrite: an integrated experimental and theoretical study. *Arch. Biochem. Biophys.* 539, 81–86. doi: 10.1016/j.abb.2013.08.016
- Zeida, A., Reyes, A. M., González Lebrero, M. C., Radi, R., Trujillo, M., and Estrin, D. A. (2014). The extraordinary catalytic ability of peroxiredoxins: a combined experimental and QM/MM study on the fast thiol oxidation step. *Chem. Commun.* 50, 10070–10073. doi: 10.1039/C4CC02899F
- Zeida, A., Reyes, A. M., Lichtig, P., Hugo, M., Vazquez, D. S., Santos, J., et al. (2015). Molecular basis of hydroperoxide specificity in peroxiredoxins: the case of AhpE from *Mycobacterium tuberculosis*. *Biochemistry* 54, 7237–7247. doi: 10.1021/acs.biochem.5b00758
- Zhao, Y., and Truhlar, D. G. (2008). Density functionals with broad applicability in chemistry. *Acc. Chem. Res.* 41, 157–167. doi: 10.1021/ar700111a
- Zuehlsdorff, T. J., Haynes, P. D., Hanke, F., Payne, M. C., and Hine, N. D. M. (2016). Solvent effects on electronic excitations of an organic chromophore. *J. Chem. Theor. Comput.* 12, 1853–1861. doi: 10.1021/acs.jctc.5b01014

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Marcolongo, Zeida, Semelak, Foglia, Morzan, Estrin, González Lebrero and Scherlis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Steady-State Linear and Non-linear Optical Spectroscopy of Organic Chromophores and Bio-macromolecules

Marco Marazzi <sup>1,2,3</sup>, Hugo Gattuso <sup>1,2</sup>, Antonio Monari <sup>1,2\*</sup> and Xavier Assfeld <sup>1,2\*</sup>

<sup>1</sup> Laboratoire de Physique et Chimie Théoriques, Université de Lorraine–Nancy, UMR 7019, Vandoeuvre-lès-Nancy, France

<sup>2</sup> Laboratoire de Physique et Chimie Théoriques, Centre National de la Recherche Scientifique, UMR 7019, Vandoeuvre-lès-Nancy, France, <sup>3</sup> Departamento de Química, Centro de Investigación en Síntesis Química (CISQ), Universidad de La Rioja, Logroño, Spain

## OPEN ACCESS

### Edited by:

Sam P. De Visser,  
University of Manchester,  
United Kingdom

### Reviewed by:

Etienne Derat,  
Université Pierre et Marie Curie,  
France

Artur Nenov,  
Università degli Studi di Bologna, Italy

### \*Correspondence:

Antonio Monari  
antonio.monari@univ-lorraine.fr  
Xavier Assfeld  
xavier.assfeld@univ-lorraine.fr

### Specialty section:

This article was submitted to  
Theoretical and Computational  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

Received: 31 January 2018

Accepted: 12 March 2018

Published: 03 April 2018

### Citation:

Marazzi M, Gattuso H, Monari A and Assfeld X (2018) Steady-State Linear and Non-linear Optical Spectroscopy of Organic Chromophores and Bio-macromolecules. *Front. Chem.* 6:86.  
doi: 10.3389/fchem.2018.00086

Bio-macromolecules as DNA, lipid membranes and (poly)peptides are essential compounds at the core of biological systems. The development of techniques and methodologies for their characterization is therefore necessary and of utmost interest, even though difficulties can be experienced due to their intrinsic complex nature. Among these methods, spectroscopies, relying on optical properties are especially important to determine their macromolecular structures and behaviors, as well as the possible interactions and reactivity with external dyes—often drugs or pollutants—that can (photo)sensitize the bio-macromolecule leading to eventual chemical modifications, thus damages. In this review, we will focus on the theoretical simulation of electronic spectroscopies of bio-macromolecules, considering their secondary structure and including their interaction with different kind of (photo)sensitizers. Namely, absorption, emission and electronic circular dichroism (CD) spectra are calculated and compared with the available experimental data. Non-linear properties will be also taken into account by two-photon absorption, a highly promising technique (i) to enhance absorption in the red and infra-red windows and (ii) to enhance spatial resolution. Methodologically, the implications of using implicit and explicit solvent, coupled to quantum and thermal samplings of the phase space, will be addressed. Especially, hybrid quantum mechanics/molecular mechanics (QM/MM) methods are explored for a comparison with solely QM methods, in order to address the necessity to consider an accurate description of environmental effects on spectroscopic properties of biological systems.

**Keywords:** nucleic acids, lipid membrane, (poly)peptide, circular dichroism, (two-photon) absorption, fluorescence, hybrid quantum mechanics/molecular mechanics, solvent and dynamics effects

## INTRODUCTION

The frontier between computational biochemistry and computational chemistry is now becoming blurred thanks to the development of novel more efficient modeling methods able to tackle very large systems also thanks to new more powerful hardware architectures. For a long time, biological systems, due to the enormous number of particles (atoms) to deal with, were studied only by means of statistical simulations based on classical force fields. Hence electronic phenomena

were explicitly excluded from such studies. Contrarily, chemical systems were treated by means of quantum chemistry tools since electronic motion was involved. However, most calculations were done placing the system in isolated (infinitely diluted gas phase) conditions. Nowadays, taking into account surroundings effects, like solvent effects, on chemical reactivity is routinely achieved either by implicit methods (Ruiz-Lopez et al., 1993; Tomasi et al., 2005) or explicit one. Conversely, studying electronic phenomena in biological macromolecules is also widely spread. These enlargements of both field, make them quite similar from the methodological point of view, since one needs to combine high accuracy quantum chemical calculations with statistical thermodynamics simulations in order to get meaningful information on electronic phenomena (chemical reactions, photochemical reactions, or electronic spectroscopy) in large systems. The appearance of hybrid quantum mechanics/molecular mechanics methods (QM/MM) in the 90's is at the origin of this merging between chemistry and biology, as recognized by the 2013 Nobel Prize (The Nobel Prize in Chemistry, 2013)<sup>1</sup>. It opens the way toward a brand new unexplored field: tackling electronic excited states in complex systems.

In this review we gather several applications carried out mainly by our group during the past 5 years, or so, that deal with different electronic spectroscopies, namely linear absorption, non-linear spectroscopy [especially two-photon absorption (TPA)], and circular dichroism on chemical or biochemical systems. All the chosen applications have a very tight link to critical biological problems.

This review is organized as follows. In section Linear Absorption, the importance of the surroundings and of the dynamical effects on electronic absorption spectra will be presented for various DNA photosensitizers. This is particularly important since DNA, although quite stable (almost no absorption) with respect to visible light irradiation, can be damaged by energy/charge transfer from small neighboring molecules (photosensitizers) that absorb the visible light. Section non-linear Spectroscopy collected mainly our studies on TPA. This method, arising from material science, starts now to be widely employed in photodynamic therapy since it allows (i) a deeper penetration of the beam, thanks to the use of red or infrared wavelengths, and (ii) a more precise localization of the action, since the density of photons needed for a twin absorption is high enough only in the focal region of the laser beam. The various tenets of the simulations will be discussed and their validity highlighted by numerical results. Finally, in section Circular Dichroism Modeling, a recently developed tool to compute electronic circular dichroism spectra of large macromolecular systems will be presented. Its application to DNA and peptides conformations will be shown and compared to other existing methods.

<sup>1</sup>Available online at: [https://www.nobelprize.org/nobel\\_prizes/chemistry/laureates/2013/](https://www.nobelprize.org/nobel_prizes/chemistry/laureates/2013/) (Accessed Jan 27, 2018)

## LINEAR ABSORPTION

Modeling linear absorption spectra of complex system has been a crucial and very important task and many efforts have been devoted to its realization in the past. In particular the role of the environment should be precisely taken into account when devoting to the study of complex realistic systems.

One of the most straightforward methods to take the environment into accounts has been by means of continuum polarizable methods (Rinaldi and Rivail, 1973; Tomasi, 2004), such as COSMO (Klamt and Schüürmann, 1993) or PCM (Tomasi et al., 2005). Even though the former have been, and still are, extremely popular and proven efficient in modeling homogeneous media, more recently the use of QM/MM methods has gain appeal in particular for the possibility to treat inhomogeneous environment that may include chromophores interacting with complex biological systems, such as nucleic acids and proteins, or materials.

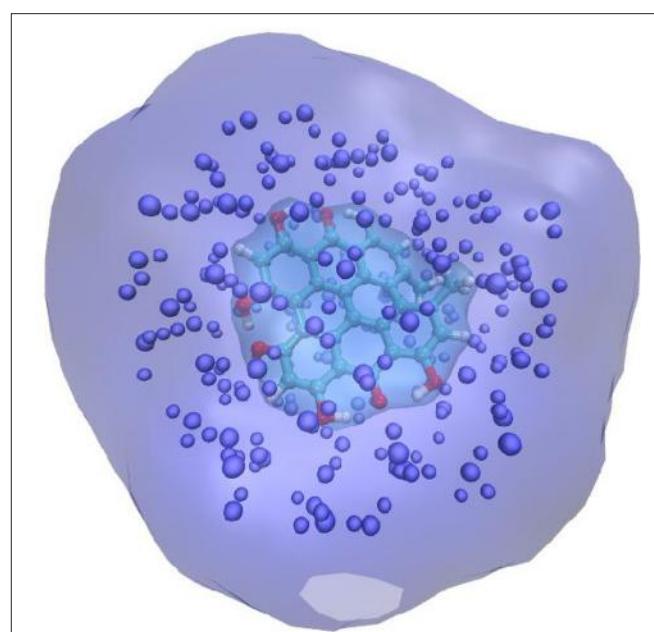
However, going from an implicit to an explicit description of the molecular surroundings, brings important conceptual problems that should be properly addressed, and that are partially different from the one observed for ground state problems. On that respect, indeed, when considering QM/MM methods one can distinguish a hierarchy in the treatment of the environment (Monari et al., 2013; Rivail et al., 2015). In particular one may distinguish mechanical embedding (ME) in which only the geometrical constraints imposed by the MM partition on the QM geometry are taken into account; the electrostatic embedding (EE) in which the polarization of the QM wave function by the MM point charges is allowed; and finally the polarizable embedding (PE) in which the back polarization of the MM potential by the QM partition is included. While, force field parameterization allows to recover the polarization of the bulk when dealing with ground state problems, and hence EE is usually sufficient to achieve a correct description of the complex environment, the situation is much more complex in the case of electronic excited state problems. Indeed, by definition, an electronic excited state will involve a sudden, and in some instance important, change in the electronic density distribution, hence, and as a consequence, the response, i.e., the polarization, of the nearby molecular surroundings will become important and cannot be neglected anymore. One straightforward strategy to include polarizable effects is to switch from a fixed-charge force field to a polarizable one (Caprasecca et al., 2014; Shi et al., 2015). Different strategies exist, based on the inclusion of atom multipole moments in addition to the charges, and have been interfaced with a number of very used codes such as Dalton and Gaussian QM/MMPol (Orlando and Jorgensen, 2010; Loco et al., 2016b). However, the parameterization of polarizable force field can in some instance be rather cumbersome, while the calculation overhead can become important.

An alternative strategy developed about 10 years ago (Jacquemin et al., 2009; Monari et al., 2013) and called the electrostatic response of the surrounding (ERS) proposes to tackle polarization issues by surrounding the QM partition in a polarizable cavity described by the fast component of the dielectric constants, i.e., the dielectric constant extrapolated

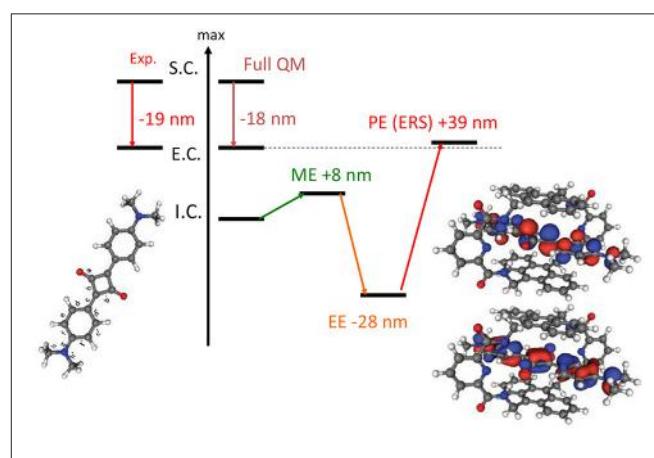
to infinite frequencies. To be precise, let us remind that the former continuum will be embedded in the MM partition but will not interact with the MM point charges (Figure 1). The idea underlining this strategy is due to the fact that the fast component of the dielectric constant represents the instantaneous rearrangement of the surrounding to the change in the QM electronic density, hence the use of a self-consistent reaction field (SCRF) approach, like in PCM, will allow to optimize the surroundings charge distribution to the change induced by the electronic transition. The advantages of this approach are two-fold: first no particular parameterization is required since the fast component of the dielectric constant is almost a constant value comprised between 1.50 and 2.00 a.u. for every non-conductive material. Secondly the computational overhead compared to EE is absolutely negligible and of exactly the same magnitude as a PCM calculation. Those aspects may hence qualify ERS as an efficient and universal strategy to include PE in complex systems. Indeed, its performance has been extensively proven in a number of different systems, including chromophore embedded in protein (Monari et al., 2012a,b), dyes interacting with nucleic acids (Chantzis et al., 2013, 2014; Etienne et al., 2013; Véry et al., 2014; Dumont and Monari, 2015) as well as native proteins (Etienne et al., 2014b), and its capacity to yield extremely accurate results has been confirmed. Even though conceptually generalizable to every quantum chemistry method able to provide electronic excitation energies, most of the applications have been performed in the framework of time dependent density functional theory (TD-DFT) approach. This fact is definitively understandable considering from the one hand the good ratio between accuracy and computational cost, and on the other hand the fact that TD-DFT provides a well balanced description of a relatively large manifold of excited states independently on the choice of an active space. However, as in all DFT methods, the drawback is the necessity to preliminary choose an exchange-correlation functionals whose effects on the accuracy of the results can be rather important. The general good practices concerning the choice of the functionals for excited state calculations that have been established for isolated systems also hold when the environment is taken into account, as an example while hybrid represent an obvious improvement over the performance of pure LDA and GGA functionals, charge-transfer states will necessitate the inclusion of long-range corrected functionals. Furthermore, the use of diagnostic indexes (Peach et al., 2008; Le Bahers et al., 2011; Etienne et al., 2014c) giving a numerical representation of the amount of charge transfer is strongly encouraged.

Indeed in a most stunning application (Jacquemin et al., 2009), involving the TD-DFT study of a caged dye, the necessity of including PE in the QM/MM calculation of absorption spectra has been absolutely evidenced. It has been shown that the effects of EE and PE were of the same order of magnitude but while the former was inducing a red-shift the latter is blue-shifting the absorption back as compared to the ME results (Figure 2). Hence, the inclusion of EE alone would have strongly deteriorated the quality of the computed results.

Once the necessity to include ME, EE, and PE in the QM/MM spectra calculations has been firmly assessed one has



**FIGURE 1** | Schematic representation of the QM/MM + ERS partition for a solvated chromophore. The QM chromophore is represented in balls and sticks and atom color, the explicit MM charges are indicated as van der Waals spheres. The QM chromophore is placed in a cavity created in the polarizable continuum that is schematized by the cyan transparent surface.



**FIGURE 2** | Illustration of the importance of the inclusion of PE in the calculation of absorption spectrum. The results are obtained at TD-DFT level using PBE0 for an encapsulated squaraine as reported by Jacquemin et al. (2009). S.C. stands for solvated chromophore, I.C. for isolated chromophore i.e., in gas phase, E.C. for encapsulated chromophore. Full QM refers to the inclusion of both the chromophore and the cage in the QM partition.

to take into account that, especially in complex systems and complex environments, a good representation of the electronic excited states and hence of the absorption spectra requires to go beyond the usual vertical representation of the excitation energy, in which one consider that absorption or emission spectra can be obtained as vertical transition from the starting state equilibrium geometry. Usually, and as a first approximation, the

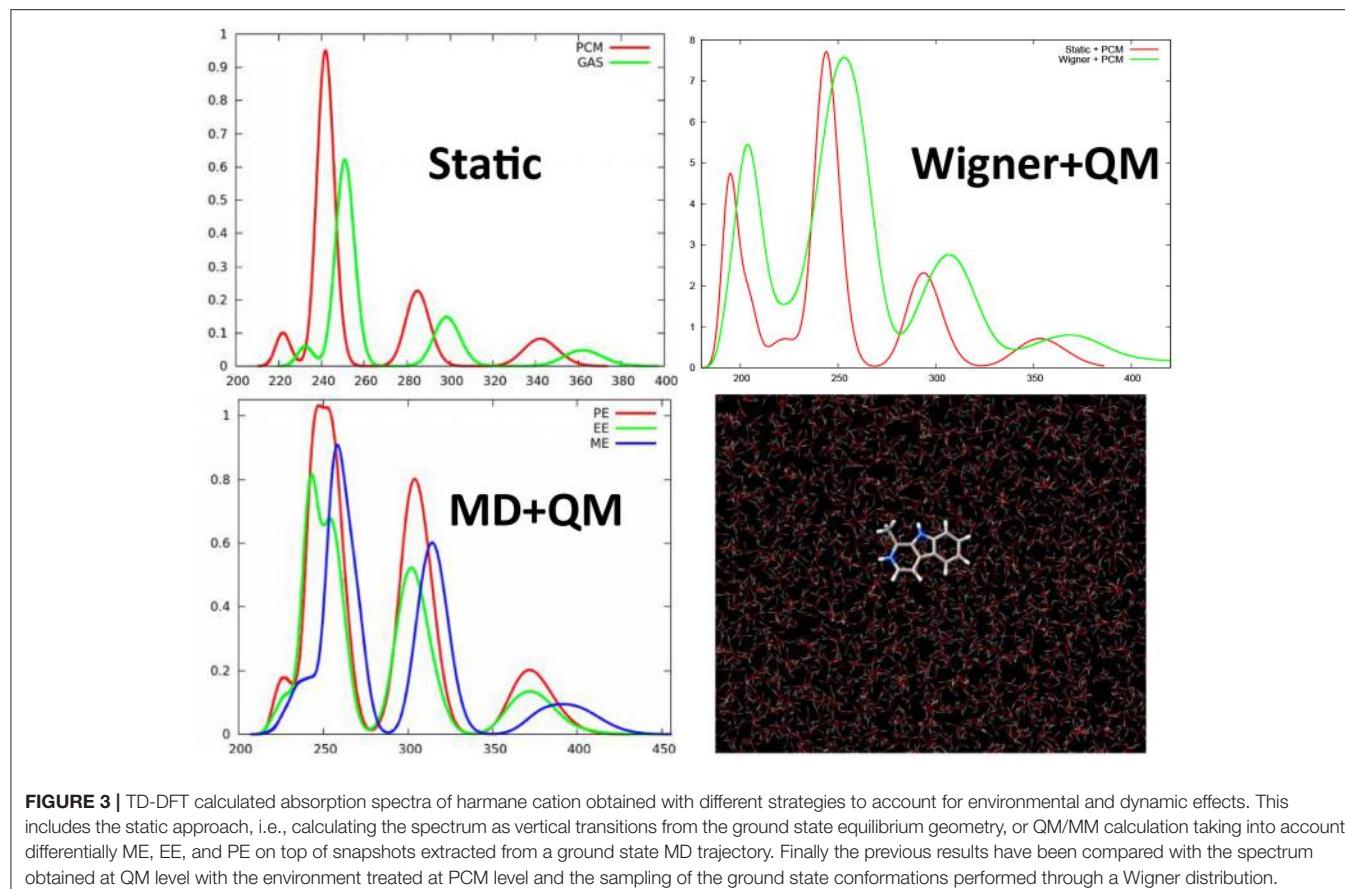
explicit calculation of the difference in energy between the first vibrational states ( $E_{0-0}$ ) can be used, however this strategy does not allow to take into account the full vibrational structure. An alternative strategy is based on the explicit calculation of the quantum based vibronic coupling, i.e., the Franck-Condon and Herzberg-Teller factors, based on the overlap between the vibrational wavefunctions and whose convolution will give the vibronic electronic spectra, i.e., the exact coupling between the vibrational and electronic states (Impronta et al., 2007; Santoro et al., 2007). Some codes allowing to calculate such factors exist, and have shown considerable success in particular in the description of the coupling with the high frequency modes and the corresponding change in the spectral band shape (Cerezo et al., 2015, 2016). However, also because of the numerical calculation of the vibrational overlap integrals, they show limitations in the case of low-frequency large-amplitude vibrational modes, such as out-of-plane bending of  $\pi$ -conjugated systems. Furthermore, the explicit inclusion of the quantum effects, provided by the Franck-Condon principle, is less important for the latter mode, which due to the large amplitude behaves much closer to the classical limit. Hence an alternative strategy to take into account such vibrational mode is to perform an accurate sampling of the potential energy surface landscape, allowing for the extraction of snapshots from which vertical excited states can be obtained (Etienne et al., 2013). The final spectrum will then be constituted by the convolution of all the vertical transitions, weighted by the corresponding intensities, for each of the snapshots. Even though the quantum chemistry calculation for the excited states will have to be repeated in order to obtain a reasonable statistical sampling of the conformational space, the explicit calculation of the Franck-Condon factors is avoided. Furthermore, different sampling techniques can be used ranging from classical or QM/MM molecular dynamics, up to semi-classical distributions such as the Wigner one (Dahl and Springborg, 1988). Most notably, in Wigner distribution the snapshots are obtained from the chromophore equilibrium geometry and its vibrational harmonic frequencies, taking into account the energies of a set of quantum harmonic oscillators.

We have shown that even for simple organic molecules such as harmane cations whose spectrum was simulated at TD-DFT level (Etienne et al., 2013, 2014d), the inclusion of vibrational effects via the calculation of the excited states on snapshots extracted from a classical molecular dynamics trajectory in a water box, induces a red-shift of about 20 nm and hence allows to perfectly recover the experimental results (Figure 3). Remarkably enough the same red-shift, and hence the same agreement with experimental results is obtained when the excited states of harmane are computed: (i) at QM/MM level from snapshots extracted from a molecular dynamic trajectory; (ii) at QM + PCM level from snapshots extracted from a molecular dynamic trajectory; and (iii) at QM + PCM level from snapshots extracted from a Wigner distribution. Hence, it appears clearly that the crucial element to recover experimental results will be the correct treatment of vibrational effects rather than the treatment of the environment. However, it has to be underlined that, even though to a less extent than for encapsulated squaraine, EE and PE

produce different shifts on the absorption wavelength (Figure 3). The same importance of the dynamic effects has also been found in the calculation of emission, i.e., fluorescence, spectra of harmane (Etienne et al., 2014d). On the same point, and also due to the local nature of the bright states involved in the spectrum the choice of the functional appears as not crucial, and generally hybrid ones may be considered as accurate and provide vertical excitation energies results comparable with the one obtained at equation of motion coupled cluster (EOM-CC) level (Etienne et al., 2014d).

Vibrational effects also play a crucial role in the case of other organic compounds such as the natural occurring drug palmatine (Dumont and Monari, 2015), known to interact with DNA and triggering singlet oxygen production (Hirakawa and Hirano, 2008). It has been shown (Figure 4) that the calculation of vertical transitions from the equilibrium ground state geometry provides an excellent agreement with experimental values when hybrid functionals, such as PBE0, are used. On the other hand, including dynamic effects via a molecular dynamic sampling leads to an excessive red-shift of the spectrum. However, this agreement is only due to the unphysical stabilization of charge transfer states induced by the use of hybrid functional and hence is only related to errors cancelation. Indeed, when using long-range corrected or meta-GGA functionals, the effects are reversed and in the case of M06-2X, the spectrum modeled including the dynamic effects gives results in perfect agreement with the experimental one while the static one appears way too blue-shifted (Figure 4). This tendency can be perfectly rationalized considering that long-range corrected functionals solve the problem related to the overestabilization of the charge transfer states, and hence globally blue-shift the absorption spectrum avoiding error cancelation with the dynamic effects. Hence, palmatine theoretical spectroscopy represent a very nice example of the subtle interplay of different interlocking effects in the treatment of excited states in complex systems, that necessitate to finely tune the computational strategy to avoid spurious errors and a bad description of the dominant states. Furthermore, the sampling of the palmatine conformational space through molecular dynamics has also allowed to recover its interaction modes with DNA and to model the change in absorption spectrum due to this interaction (Dumont and Monari, 2015).

In the same context it has been shown that the inclusion of vibrational and dynamics effects may also be crucial in determining not only the absorption spectrum but also the photophysically allowed pathways of many chromophores. As an example the sampling of the conformational space of the hypericin drug (Gattuso et al., 2017a) via the extraction of snapshots through Wigner and molecular dynamics distribution has allowed to calculate also other crucial parameters such as the spin-orbit coupling (SOC) between the low lying excited singlet and triplet states. In particular it has been shown that the SOC is generally increased due to geometry distribution and may reach values up to some tenths of  $\text{cm}^{-1}$ . This sort of vibrationally allowed SOC can rationalize the relatively efficient intersystem-crossing experienced by hypericin and hence its efficient production of singlet oxygen upon irradiation that constitutes the basis of some of its pharmacological activity.



**FIGURE 3** | TD-DFT calculated absorption spectra of harmane cation obtained with different strategies to account for environmental and dynamic effects. This includes the static approach, i.e., calculating the spectrum as vertical transitions from the ground state equilibrium geometry, or QM/MM calculation taking into account differentially ME, EE, and PE on top of snapshots extracted from a ground state MD trajectory. Finally the previous results have been compared with the spectrum obtained at QM level with the environment treated at PCM level and the sampling of the ground state conformations performed through a Wigner distribution.

Remarkably enough those properties have been observed both in the case of aqueous solution and when interacting with lipid membrane bilayers.

Hence the combination of a proper treatment of the environment, allowing for the inclusion of environmental and dynamics effects, allows the precise determination of spectroscopic properties as well as photophysical key parameters, of a number of complex systems including optical materials such as solar cells dyes (Sengul et al., 2017) and poly-thiophene units (Turan et al., 2016), as well as biological systems and drugs. Furthermore, the effects of solvent relaxation and their influence on the spectroscopical properties can also be assessed through similar protocols (Zvereva et al., 2018).

## NON-LINEAR SPECTROSCOPY

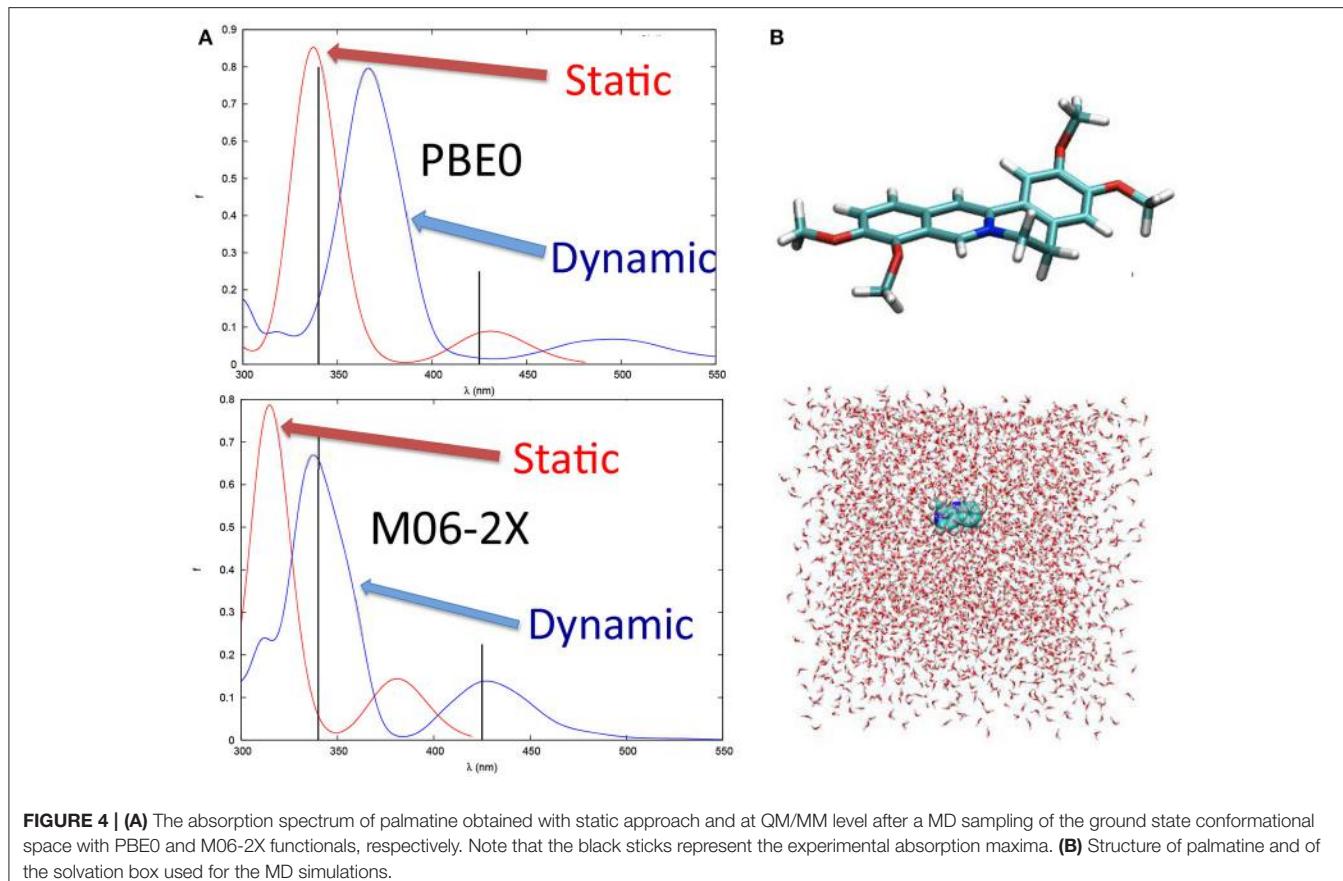
### Two-Photon Absorption and Second-Order Harmonic Generation: General Principles, Experimental Techniques, and Computational Methods

If linear absorption properties are essential to characterize bio-macromolecules, non-linear absorption processes lead to important features, especially when looking at possible applications. Highly investigated in the past with the goal to enhance materials properties, the scientific interest has recently

turned toward bio-macromolecules, in some cases reaching the bio-medicine field. Especially, two phenomena are of utmost interest: multi-photon absorption and high-order harmonic generation. Among them, because of the lower symmetry usually found in biological media compared to condensed materials, two-photon absorption (TPA) and second-order harmonic generation (SHG) are mainly concerned (Antoine and Bonačić-Koutecký, 2018). In both cases, two photons interact simultaneously with the same molecule resulting, in the case of TPA, in the absorption to an electronic excited state corresponding to the sum of the energies of the two incoming photons or, in the case of SHG, in the emission of a photon arising from the sum of the incoming photon energies. Moreover, we can discern between degenerate two-photon phenomena, in which both photons have the same energy (hence same frequency and same wavelength) and non-degenerate phenomena, in which the two photons differ in their energies (Figure 5).

Concerning applications in the biological field, TPA is an especially desired feature since it solves two problems at once:

- i) The simultaneous absorption of two photons allows, in the case of degenerate TPA, to divide the required energy by a factor of two, i.e., to double the vertical transition wavelength. This permit to notably shift the required incoming photon energy toward the red part of the visible spectrum (i.e.,



**FIGURE 4 | (A)** The absorption spectrum of palmatine obtained with static approach and at QM/MM level after a MD sampling of the ground state conformational space with PBE0 and M06-2X functionals, respectively. Note that the black sticks represent the experimental absorption maxima. **(B)** Structure of palmatine and of the solvation box used for the MD simulations.

bathochromic shift), eventually allowing to enter in the so-called near-infrared therapeutic window (from 650 to 1350 nm), where the penetration of biological tissues is maximal (Tsai et al., 2001; Mojzisova and Vermot, 2011).

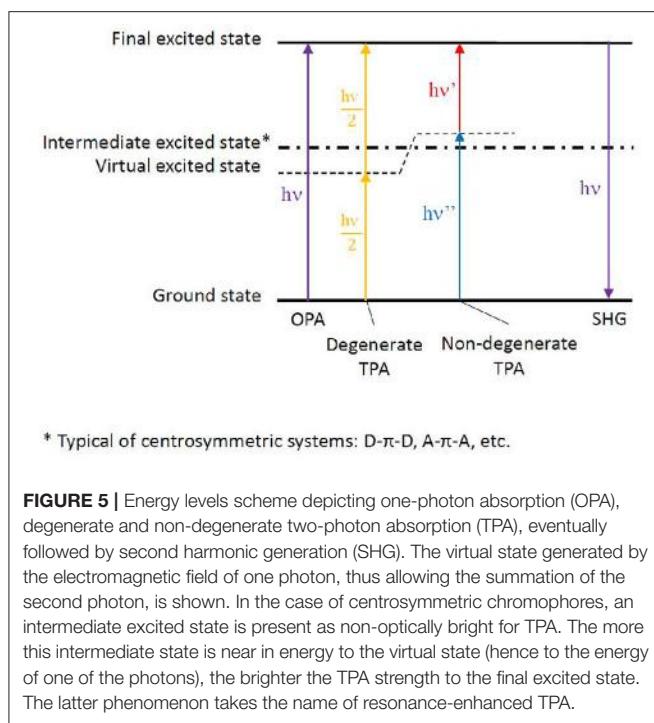
- ii) The probability for the two photons to be absorbed simultaneously is proportional to the square of the light source intensity. This will definitely increase the spatial precision, since TPA will most likely play a role only at the laser focal point, decreasing in intensity much faster than one-photon absorption (OPA), i.e., much faster than linear absorption processes discussed in the previous section. This results in a crucial advantage for bio-medical applications, since TPA can focus precisely on the lesion area, reducing the side effects (Sun and Dalton, 2008; Benninger and Piston, 2013).

On the other hand, SHG is mainly used as a principle to build powerful microscopy techniques of interest in bio-imaging. Hence, the key factor stands in the possibility to generate visible light at high-resolution. This allowed high resolution imaging from deep inside biological tissues, as lipid bilayers of cell membranes, by designing active chromophores with a required balance between hydrophobic and hydrophilic moieties (Barsu et al., 2009; Reeve et al., 2010).

Finally, the use of TPA and/or SHG should also be considered as highly promising in theranostics, a fusion of therapeutics and

diagnostics (Jeelani et al., 2014). Indeed, solely TPA coupled to red/infrared emission can constitute the basis of emissive drugs combining treatment and imaging properties. Likewise, specific chromophores, as retinal analogs among others (Theer et al., 2011), can combine TPA and SHG to match absorption, emission, and reactivity purposes (Barsu et al., 2006). In the following, we will focus on non-linear absorption, hence on TPA.

Even though a detailed understanding of the underlying physical principles is out of the scope of the present review, some fundamental concepts should be recalled in order to interpret experimental and theoretical results, thus possibly predict TPA properties. In linear absorption spectra, the intensity values are obtained by the Beer–Lambert law, by which the logarithm of the intensity is proportional to the molar absorption coefficient and to two experimental parameters: the solution concentration and the optical length, i.e., the size of the spectroscopic cell. On the other hand, when theoretically determining linear absorption spectra, the oscillator strength is calculated, a dimensionless quantity proportional to the square of the transition dipole moment between the ground and the excited state of interest. Hence, a straightforward comparison between experimental and theoretical linear spectra is not possible, and usually relies on normalization of the computed spectra. This is not the case for TPA, since absolute Göppert–Mayers units ( $1 \text{ GM} = 10^{-50} \text{ cm}^4 \text{ s photons}^{-1} \text{ molecule}^{-1}$ ) can be both measured experimentally



**FIGURE 5 |** Energy levels scheme depicting one-photon absorption (OPA), degenerate and non-degenerate two-photon absorption (TPA), eventually followed by second harmonic generation (SHG). The virtual state generated by the electromagnetic field of one photon, thus allowing the summation of the second photon, is shown. In the case of centrosymmetric chromophores, an intermediate excited state is present as non-optically bright for TPA. The more this intermediate state is near in energy to the virtual state (hence to the energy of one of the photons), the brighter the TPA strength to the final excited state. The latter phenomenon takes the name of resonance-enhanced TPA.

and computed by theory, being proportional to the so-called molecular TPA cross-section (Göppert-Mayer, 1931; Kaiser and Garrett, 1961). This makes in principle the comparison between theory and experiment easier, even though much less data are available in the literature compared to OPA, due to the more sophisticated experimental setups and the computational implementations required.

Concerning experiments, after the first demonstration of TPA by organic dyes (Peticolas et al., 1963), different techniques were developed. Indeed, an added requirement makes TPA direct measurements more complicated than OPA ones: the necessity to determine the source irradiance for each probed wavelength. This means that not only the energy of the laser beam, but also its spatial distribution (including possible changes of propagation through the sample) and its pulse-width (including eventual temporal modulation) need to be accurately monitored all over the spectral window to be probed (Negres et al., 2002). The main technique presently in use for TPA direct measurements is the z-scan technique. It derives its name from the required movement of the sample along the z axis defined as the distance between the focused laser and the detector (Sheik-Bahae et al., 1990). Nevertheless, as a way to overcome the difficulties caused by TPA direct measurements, an alternative technique has emerged, based on two-photon excited fluorescence (Xu and Webb, 1996): if OPA and TPA spectra of a reference compound are known (de Reguardati et al., 2016), then a simple comparison of one- and two-photon excited fluorescence spectra between the sample and the reference will allow to cancel out most of the variables required by the direct measurement, finally leading to the TPA spectrum of the sample. Thanks to this indirect approach, it is not necessary to manage all the issues

raised by a direct laser focus, even though fluorescence is required in order to detect a signal, hence limiting the amount and type of molecules that can be measured. Moreover, TPA measurements cannot be obtained in the spectral region where OPA is dominant.

As for theory, general rules can be followed to attempt structure-property relationships, even though the inclusion of dynamics effects by molecular dynamics, coupled to explicit solvent and complex bio-chemical environments, can lead to additional and unexplored informations. More in detail, chromophores can be divided in centrosymmetric and non-centrosymmetric ones. Especially, while a  $\pi$ -conjugated backbone is necessary for an organic molecule to absorb light at long wavelengths, donor (D), and acceptor (A) groups can be added to the ends or in the middle of the  $\pi$ -conjugated chromophore, generating centrosymmetric (e.g., D- $\pi$ -D, A- $\pi$ -A, D- $\pi$ -A- $\pi$ -D, A- $\pi$ -D- $\pi$ -A) and non-centrosymmetric (e.g., D- $\pi$ -A) structures. Indeed, symmetry plays a role in the photophysical transition selection rules (Heflin et al., 1988; Dixit et al., 1991): in centrosymmetric chromophores, a virtual state can be generated while the molecule experiences the field of the first photon, allowing the following—ca. 5 fs delay (Birge and Pierce, 1986)—photon to reach the final state, hence favoring TPA over OPA. This is possible thanks to the presence of an intermediate state next to the virtual state, a resonance condition which is generally not fulfilled in non-centrosymmetric chromophores, where in principle TPA and OPA are both possible with the same probability. (Figure 5; Pawlicki et al., 2009). While the presence of charge transfer states can indeed improve TPA cross section, their presence is however not strictly necessary to induce non-linear absorption (Beerepoot et al., 2014).

This can partially explain why, in the past, most of the attention focused on centrosymmetric modified molecules—based initially on *trans*-stilbene (Parthenopoulos and Rentzepis, 1989; Ehrlich et al., 1997; Makarov et al., 2008) and azo-aromatic compounds (Antonov et al., 2003; De Boni et al., 2005) among others. This lead to the development of compounds being more appealing for material scientists compared to biochemists and biologists: natural and bio-inspired systems rarely accomplish the requirement to be totally centrosymmetric; moreover, strong D and/or A groups cannot be always added, depending on the frequently delicate balance between chemical structure and biological function.

Computationally, two-photon transition moments are available for EOM-EE-CCSD (Equation-Of-Motion for Excitation Energies Coupled-Cluster C with Single and Double substitutions; Krylov and Gill, 2013) and ADC (Algebraic Diagrammatic Construction; Knippenberg et al., 2012) methods within the Q-Chem package (Shao et al., 2015), but also at CC2 (second-order approximate Coupled-Cluster singles and doubles model; Christiansen et al., 1995) level of theory through the TURBOMOLE program package (Furche et al., 2014), and at TD-DFT (Time Dependent-Density Functional Theory) through the DALTON2016 program suite (Aidas et al., 2014). In the latter case, since TPA active molecules usually include D and A groups as explained above, it is important to choose functionals which

can describe the displacement of charge during the transition from D-centered orbitals to A-centered ones, as hybrid and long-range corrected exchange-correlation functionals.

## Theoretical Predictions: Level of Theory and Environmental Effects

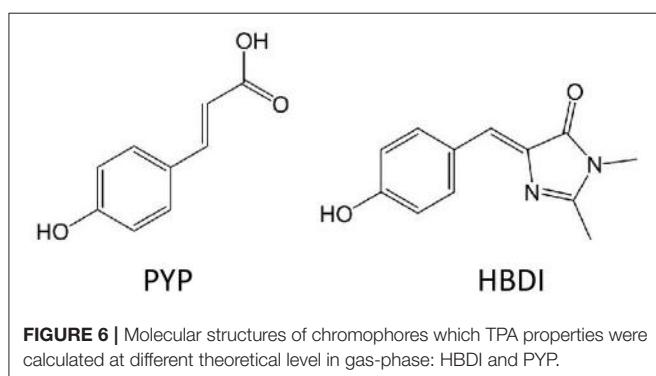
Two benchmark studies compare EOM-EE-CCSD, CC2 and TD-DFT/CAM-B3LYP (Yanai et al., 2004) two-photon cross sections for the chromophores of the Photoactive Yellow Protein (PYP) and of the Green Fluorescent Protein, GFP, in its neutral form (HBDI) (Figure 6; Beerepoot et al., 2015; Nanda and Krylov, 2015). Being all calculations performed in gas-phase, the results point toward a general qualitative agreement between the three levels of theory, with quantitative discrepancies: TD-DFT/CAM-B3LYP TPA cross-sections were found to be 1.5–3 times smaller than EOM-EE-CCSD and CC2 ones (Beerepoot et al., 2015). Of course, the divergence could lie in the different description of the excited state dipole moments (Bednarska et al., 2013). Nevertheless, also because the effect of the environment is not taken into account by this study, conclusions cannot be raised on the basis of a comparison with experimental results.

On the other hand, the computationally more affordable TD-DFT (compared to Equation-Of-Motion and Coupled Cluster approximations) allows to treat the environment by PCM and QM/MM methods at a much easier cost, including dynamics effects, finally allowing to establish how the TPA maximum intensity is affected by a realistic environment. Moreover, the spectral window and spectral shape can be recovered, determining the applicability of the chromophore for therapeutic purposes or as constituent of bio-materials.

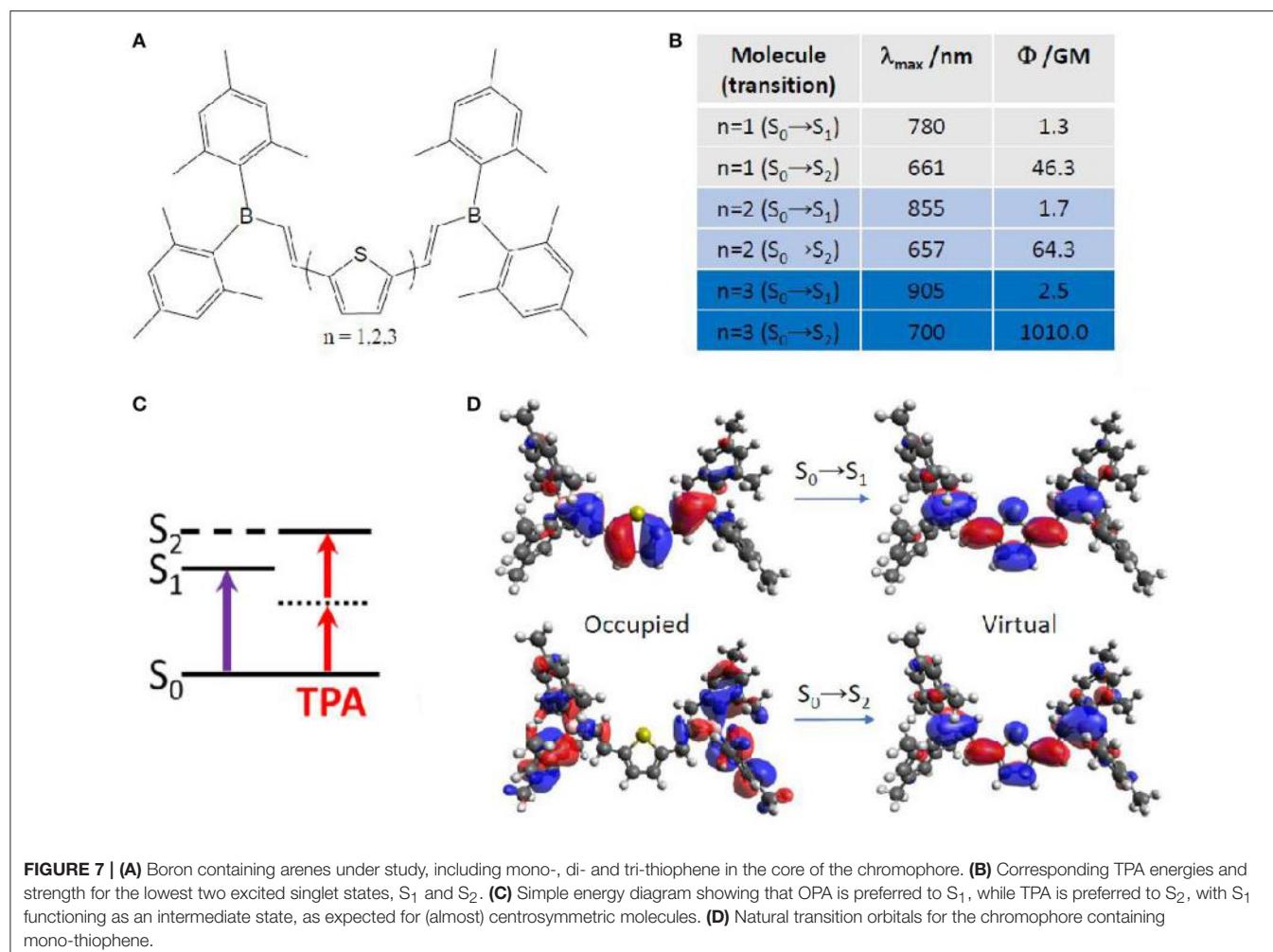
More in detail, TPA properties of boron containing arenes were evaluated at TD-DFT level, including the solvent by linear response PCM (Turan et al., 2016). In this case, a vertical approach was used, meaning that all molecules were optimized on the ground state followed by calculation of the vertical transition to the excited state. The results point toward a  $S_2$  optically bright state for TPA, while  $S_1$  is the bright state for OPA (Figure 7). This can be explained in terms of molecular orbitals. Especially, electronic density reorganization in the excited state can be efficiently monitored through Natural Transition Orbitals (NTOs) (Martin, 2003), in this case obtained with the Nancy\_EX code (Etienne et al., 2014a,c). The  $S_0 \rightarrow S_1$  transition is of

local  $\pi, \pi^*$  character, while the  $S_0 \rightarrow S_2$  transition clearly shows the charge transfer from the *E*-dimesitylborylethetyl lateral groups toward the central part of the conjugated chromophore, resembling to a  $A-\pi-A$  centrosymmetric system mentioned in the previous section. Indeed,  $S_0 \rightarrow S_1$  TPA values are almost negligible (not more than 2.5 GM), while  $S_0 \rightarrow S_2$  TPA values reach up to 1010 GM. Interestingly, we should mention that TPA values for this kind of compounds is predicted to change by two orders of magnitude, when looking at the type and number of arenes contained. Especially, thiophene based compounds correspond to the highest cross section values, compared to phenyl and fluorinated phenyl. Moreover, increasing the number of linearly linked thiophene rings induced a strong non-linear effect, passing from 46 GM (one thiophene) to 64 GM (two thiophenes) to 1010 GM (three thiophenes). In this last case, the experimental value is available [1500 GM (Entwistle et al., 2009)], denoting an underestimation by the TD-DFT/PCM prediction, but keeping a good qualitative agreement.

An alternative method to obtain TPA properties is by performing classical MD, followed by QM/MM TPA calculations of single snapshots, randomly selected along the MD trajectory. This approach, explained also in section Linear Absorption for OPA, has the advantage to include the solvent explicitly, even though usually force field parameters have to be extracted for the chromophore under study. Especially, the proper parameterization of chromophores with an extended cyclic  $\pi$ -conjugated backbone is far from being trivial (Li et al., 1989; Autenrieth et al., 2004; Zhang et al., 2012), as is the case of porphyrin-like systems. Indeed, the correct description of low-frequency modes is crucial, since they are expected to notably impact the prediction of the TPA spectrum. An example is chlorin-e6, already reported as photodynamic and antibacterial drug (Zenkevich et al., 1996; Fernandez et al., 1997; Nyman and Hynninen, 2004; Paul et al., 2013; Winkler et al., 2016). In this case, a careful validation of the force field was performed by comparison with the Wigner distribution approach, when calculating OPA properties (Gattuso et al., 2017b). Specifically, a maximization of the overlap between force field and Wigner computed absorption spectra, coupled to a minimization of the Root Mean Square Deviation (RMSD) was accomplished. This force field optimization further allowed TPA properties calculations in explicit water at QM(TD-DFT)/MM level. The spectrum (Figure 8) shows two peaks corresponding—as for all OPA spectra of porphyrin-like systems—to the Soret band (at *ca.* 730 nm and 60 GM) and to the Q band (at *ca.* 1100 nm and 20 GM). TPA intensity values are therefore much lower than boron containing thiophenes aforementioned (Turan et al., 2016), even though still acceptable for bio-applications. Lower TPA intensities can be expected since chlorin-e6 partially loses the centrosymmetry of porphyrins, and can be further rationalized based on its NTOs: both OPA and TPA are possible, with a Soret ( $S_0 \rightarrow S_2$ ) band much more intense than the Q band ( $S_0 \rightarrow S_1$ ) in both cases. Indeed, the NTOs describe for both electronic transitions a charge reorganization centered in the core ring, also explaining the lower change in TPA intensity compared to the three orders of magnitude observed for thiophene based chromophores (Figure 7). Anyway, only a relatively small



**FIGURE 6 |** Molecular structures of chromophores which TPA properties were calculated at different theoretical level in gas-phase: HBDI and PYP.

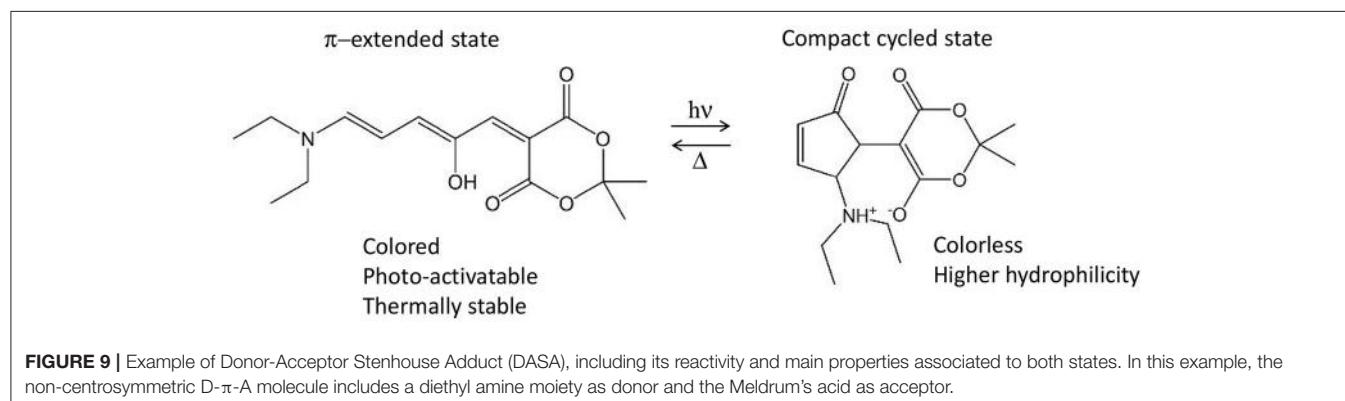
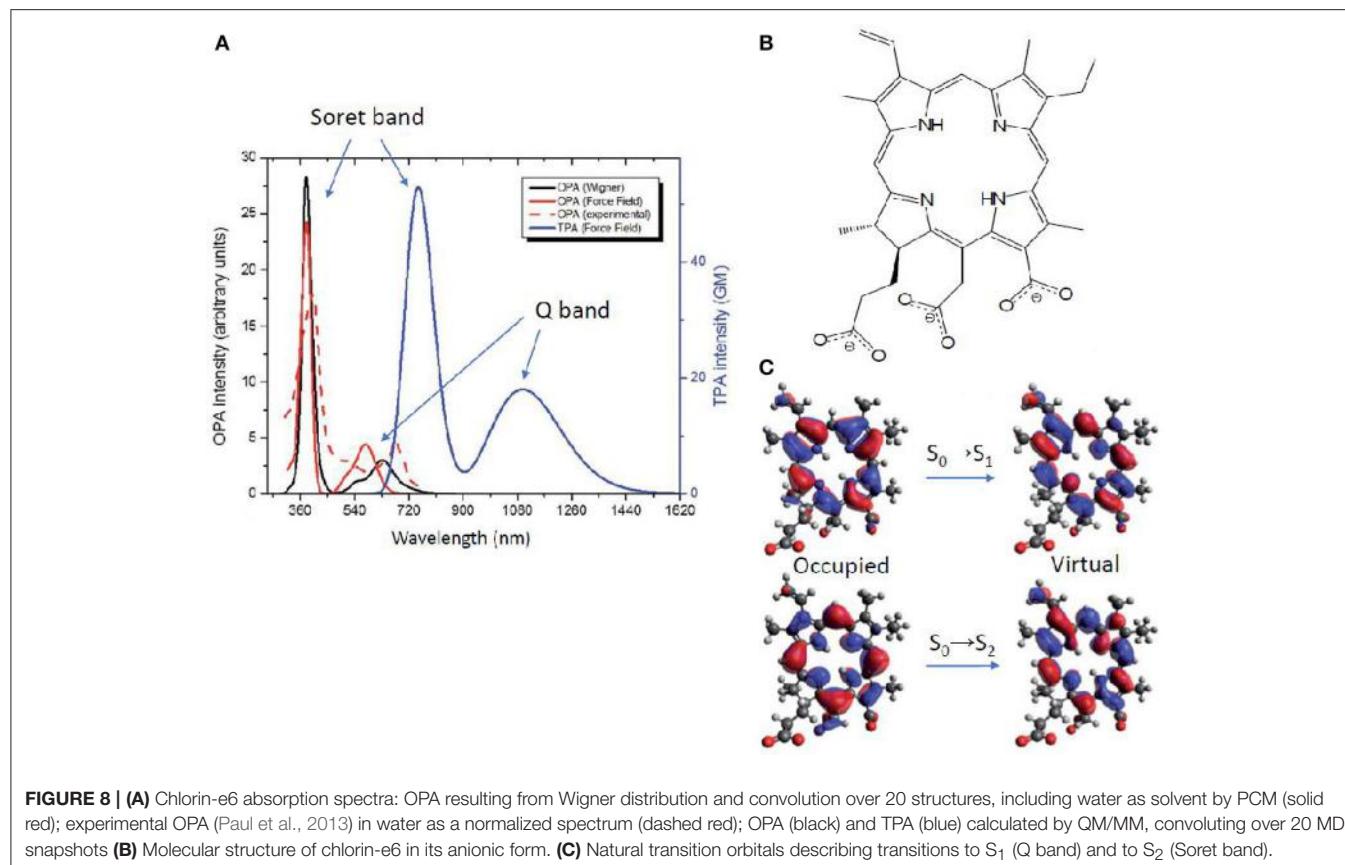


**FIGURE 7 | (A)** Boron containing arenes under study, including mono-, di- and tri-thiophene in the core of the chromophore. **(B)** Corresponding TPA energies and strength for the lowest two excited singlet states,  $S_1$  and  $S_2$ . **(C)** Simple energy diagram showing that OPA is preferred to  $S_1$ , while TPA is preferred to  $S_2$ , with  $S_1$  functioning as an intermediate state, as expected for (almost) centrosymmetric molecules. **(D)** Natural transition orbitals for the chromophore containing mono-thiophene.

overlap between OPA and TPA spectra is expected, thus justifying the use of the indirect—and easier to afford—experimental setup, also thanks to the fluorescent properties typical of these molecules (Gattuso et al., 2017b; Liu et al., 2018). This could hence encourage more experimental groups to measure this type of molecules, to better assess theoretical methods and proposed advancements in various applications (Ryu et al., 2018).

Nonetheless, several photo-active chromophores are totally non-centrosymmetric. Examples are the retinal and the Donor-Acceptor Stenhouse Adduct (DASA). The former is known as the *cis-trans* photo-isomerizable switch of rhodopsins, G-protein coupled receptors responsible for the process of vision (Wald et al., 1968; Okada et al., 2004) and recently employed as optogenetic tools (Deisseroth, 2011; Tischer and Weiner, 2014; Guo et al., 2016; Hontani et al., 2017). The latter is a recently discovered type of photo-switch (Helmy et al., 2014a,b) which converts from an initial  $\pi$ -extended colored state into a final compact colorless state, moreover accompanied by a notable polarity change (Figure 9). Both retinals and DASAs are characterized by a lowest-lying singlet excited state ( $S_1$ ) of partial charge transfer character, even though of different

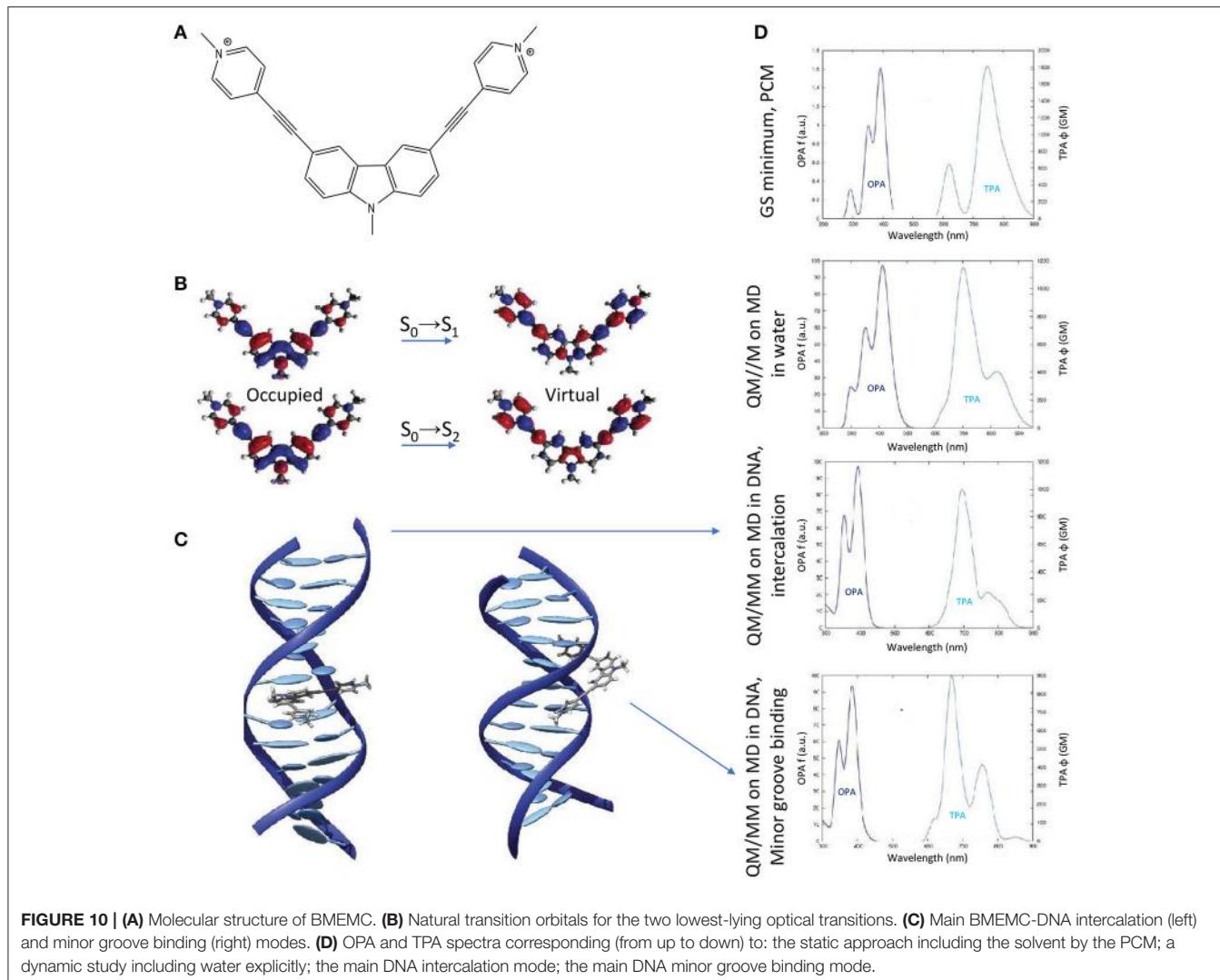
origin: retinal is a protonated Schiff-base which, in biological media, is surrounded by the opsin pocket of rhodopsin, i.e., an active site where the charges and polarity of the amino acid side chains enhances the  $S_1$  charge transfer character (a sort of induced D- $\pi$ -A system), compared to a retinal molecule in solution. In the case of DASAs, no macromolecular entity is responsible for the efficiency of its photo-activity, since it is a D- $\pi$ -A system itself. This symmetry reasonings explain why for both systems a low  $S_0 \rightarrow S_1$  TPA intensity is theoretically predicted [ca. 2 and 3.32 GM for 11-*cis* retinal and DASA, respectively (Palczewska et al., 2014; García-Iriepa and Marazzi, 2017)]. Even so, retinal isomerization by TPA is expected to play a role to trigger human infrared vision (Artal et al., 2017). Hence, some additional QM/MM studies including polarizable embedding could be worth. Also, it should be mentioned that bio-mimetic photo-switches based on retinal are available, even though still limited to the blue/UV part of spectrum (Sampedro et al., 2004; García-Iriepa et al., 2013, 2016), and could hence greatly benefit from computational design to improve TPA properties. In this context it is also important to precise that cross-sections of  $< 10$  GM can indeed be detected experimentally, however their increase is fundamental



in order to permit the use of TPA switches in practical applications.

Finally, the impact of the biological media—in the specific case of B-DNA—on the TPA properties of an organic photosensitizer, was established by QM(TD-DFT)/MM calculations (including electrostatic and mechanical embedding), after having determined the photosensitizer-DNA interaction modes by classical MD. Especially, a centrosymmetric A- $\pi$ -D- $\pi$ -A dication was selected as photosensitizer (the 3,6-bis[2-(1-methylpyridinium)-]9-methylcarbazole, abbreviated as BMEMC) as it was experimentally shown to cause DNA strand breaks upon infrared irradiation, also in hypoxic conditions

(Zheng et al., 2015). The BMEMC-DNA interaction modes and the sensitization mode of action (Gattuso et al., 2016a; Figure 10) were then rationalized on theoretical basis: BMEMC can undergo a photoinduced spontaneous ionization, leading to the production of a solvated electron and a radical cation. Since it is a centrosymmetric structure with relatively strong A groups, TPA intensity values are high, overcoming 1000 GM for the  $S_0 \rightarrow S_2$  transition, when computed in water. We should note that in this specific case a proper spectroscopic description of TPA properties is possible only by averaging QM/MM results over a reasonable number of MD trajectory snapshots (20 in this case). Indeed, when compared with the static approach,



i.e., by considering only the Franck–Condon structure obtained with the PCM model, we note an apparent inversion of  $S_0 \rightarrow S_1$  and  $S_0 \rightarrow S_2$  bands in the TPA calculated spectrum (Figure 10). This can be explained in terms of low-frequency modes that are strongly affecting the TPA active  $S_2$  state, for which only a dynamical approach results in a proper description. Indeed, as shown by the NTOs, the  $S_2$  virtual orbital is the main responsible for the electron displacement toward the pyrimidinium A edges, which are also the most affected by low-frequency modes. When looking at the interactions with DNA, four stable binding modes are found by MD: two intercalation modes and two minor groove binding modes, depending if the carbazole D core or a pyrimidinium A edge attempts first the contact with the bio-macromolecule. The QM/MM TPA spectra calculated in the four cases, by convoluting the same number of trajectory snapshots as in water, shows that the non-linear response of BMEMC does not significantly changes in presence of DNA, regarding both spectral regions and peak intensities. Such description matches qualitatively the experimental measurements (Zheng

et al., 2015). Comparing intercalation modes, the only notable difference regards the  $S_0 \rightarrow S_1$  band shape, broader and less structured in intercalation than in minor groove binding. Again, this is the result of electrostatic and mechanical embedding: in intercalation, the core of BMEMC is in the hydrophobic pocket in the middle of DNA, and it is more constrained by the environment. Even if the present example is showing that the impact of the biological media is affecting TPA cross-section only marginal and is by the way more oriented toward photodynamic therapy than diagnosis, it is important to remind that possible applications of TPA fluorescence, and its modulation by the environment, are of great importance for bio-imaging and are paving the way to a new dimension in theranostic, i.e., the combined treatment and diagnosis (Hu et al., 2018).

## Effect of the Polarizable Embedding Scheme

The polarizable embedding scheme has become recently available to include not only polarization of the QM region due to MM

charges, but also vice-versa, thus reaching an even more realistic description of the environment. Especially, TD-DFT can be applied to calculate multi-photon absorption properties in a modified version of the DALTON program (Steindal et al., 2016). This can definitely improve the quality in the treatment of the environmental effects, even though increasing the computational cost. Indeed, the same neutral GFP chromophore previously studied in gas-phase (Figure 6; Beerepoot et al., 2015) was found to absorb with a bathochromic shift (of 0.09–0.16 eV) when including the polarizable embedding scheme for two-, three- and four-photon absorption. Nevertheless, we should note that TPA strengths are more sensitive to the size of the QM region than OPA strengths (Steindal et al., 2016), hence envisaging two alternatives to maintain a reasonable balance between computational quality and cost: (i) increase notably the size of the QM region, keeping an electrostatic QM/MM embedding scheme; (ii) keep an acceptable size of the QM region, applying a polarizable QM/MM embedding scheme. Moreover, for future tests and benchmarks, alternative QM/MM border treatments rather than the hydrogen link-atom approach could be proposed, as a way to limit the size of the QM region and improve the quality of TPA properties prediction.

## CIRCULAR DICHROISM MODELING

Even though linear and non-linear absorption spectroscopies proved to be efficient in probing the binding propensity of organic compounds with complex bio-macromolecular systems (Jiang et al., 2014; Li et al., 2015), they are unable to provide information about the global structure of bio-macromolecular systems. Indeed, the sensitivity of these techniques is not sufficient to differentiate between different arrangements. In contrary, electronic Circular Dichroism (CD), which is a fast and sensitive spectroscopic method, provides specific optical signature for each polypeptide secondary structure arrangement (alpha-helix, beta-sheet, random coils,) (Greenfield, 2007) or nucleic acids macromolecular structure (A/B/Z-DNA, G-quadruplexes, i-motifs, and more; Kypr et al., 2009; Vorlíčková et al., 2012). Hence, it allows an unequivocal differentiation of the systems configurations. Moreover, since the resulting optical signal is directly correlated to the global structure of the system, specific phenomena can be efficiently probed such as the formation of aggregates between molecular compounds and proteins or DNA. It can be used to investigate structural reorganizations from a configuration to another induced by external effects. Moreover, in the case of DNA/ligand and DNA/protein binding, CD is able to provide structural information at two levels: (i) on one side, considering the spectral fingerprint of well-characterized macromolecular arrangements, the structural modifications (local or global) induced by the binding can be probed by simply following the spectral features while increasing the concentration of the substrate (Basu and Suresh Kumar, 2014); (ii) on the other side, the induced CD (for non-chiral compounds) or band shape modifications (for chiral compounds) while bound to a biomolecular system can also provide insights in the binding configurations or be used

to perform spectroscopic titration and hence recover binding free-energies (Holmgaard List et al., 2017).

In the case of nucleic acids aggregates, (Kypr et al., 2009; Vorlíčková et al., 2012) extensive experimental studies of the ability of CD to compare and differentiate most of the well-known conformations such as A/B/Z-DNA, guanine, and cytosine quadruplexes or even i-motifs have been reported. Moreover, the results demonstrated the efficiency of CD to follow induced transitions between two configurations. Indeed, they described the trifluoroethanol-induced transitions from B-DNA to A-DNA (sequence poly[GCGGCGACTGGTGAGTACGC]) and B-DNA to Z-DNA (sequence poly[d(GC)]). Also, they described the evident spectral modifications during either transitions between quadruplexes or the formations of i-motifs while varying the pH or increasing the ionic strength. This work demonstrated how robust the information given by CD can be and paved the way to more advanced studies on DNA/ligand and DNA/protein bindings.

Considering natural nucleic and amino acid arrangements such as B-DNA or alpha helices, their CD spectral features arises from the interaction of optically active chromophores organized in a specific layout. The chromophoric units, giving rise to the most prominent CD signals, are the nucleobases in the case of DNA, and the -(N-H)-(C=O)- peptide bonds in the case of proteins. Assembled in larger macromolecular ensembles owning intrinsic chirality (such as the B-DNA double helix or alpha helical peptides) the interaction between these achiral molecular chromophores will induce circularly polarized optical signals. The interactions influencing the electronic properties of these units can either be (i) hydrogen bonds, (ii)  $\pi$ -stacking interactions, (iii) electrostatic interactions. Furthermore, upon the optical excitation of the aggregate one can evidence, (iv) excitons formation or (v) population of Charge-Transfer (CT) states. In the case of B-DNA, the delocalization of excitons is not extended to more than three stacked nucleobases and doublets of stacked nucleobase pairs can be considered as a model accurate enough to provide a reasonable description of the B-DNA CD spectra (Nogueira et al., 2017) as will be presented in the following sections.

Modeling the CD spectra of large macromolecular systems such as proteins and DNA is challenging since it requires to overcome numerous technical drawbacks, some of them being common also to the calculations of linear and non-linear absorption but being emphasized by the inherent extreme sensitivity of CD spectroscopy to subtle structural and electronic effects:

- Sufficient and accurate sampling of the conformational space. The increase of computing resources allows nowadays extensively long Molecular Dynamics (MD) simulations, reaching from microsecond to millisecond time scales, directly comparable to experimental results (Galindo-Murillo et al., 2015). Moreover, the accuracy of current force fields describing these biological systems have been demonstrated and the sampled potential energy surfaces can be considered close to *in-vitro* experiments (Dans et al., 2017). The recently available parmbsc1 correction (Ivani et al., 2015) available for

- AMBER is currently one of the most robust force field for nucleic acids.
- ii) Computation of the excited states. TD-DFT, even considering its limitations, is a widely employed computational tool for obtaining of electronic properties in biological systems. It has proved to be accurate enough to be comparable with experiments (Laurent and Jacquemin, 2013). Also, more advanced methods such as CASSCF, CASPT2, or CC have been employed but still suffer from their much higher computational cost which is a drawback when considering large biological systems made up of hundreds to thousands of atoms.
  - iii) Taking into account the environment. Even though the use of common continuum models such as PCM or COSMO, provide relatively accurate results and are computationally cheap, many specific cases have required to consider explicit solvations to account for more specific interactions between solvents and the molecular region of interest. This is overcome using hybrid QM/MM methods.

The protocols that have been employed by our group to model CD spectra of biomolecular systems will be described more in details in the following. Nevertheless, they can be summarized in four steps: (i) the starting structure can be obtained from X-ray or NMR studies or from *in-silico* builders of biomolecules (NAB, tleap Case et al., 2017); (ii) MD simulation to sample the potential energy surface of the system; (iii) computation of excited states at *ab-initio* level and eventually of their coupling over the full macromolecular system via semiempirical Hamiltonians; (iv) convolution of excited states energies and transition strengths using Gaussian or Lorentzian shaped functions. In particular concerning the crucial point of the coupling between the individual chromophores we chose to use the simplest dipole moment approach. Hence, the coupling is estimated simply by the scalar product of the transition dipole moments calculated for each of the individual chromophores weighted by the distance of their center of charges. This strategy differs from the one employed by other authors (Jurinovich et al., 2014) who chose to explicitly determine the coupling via the excited states' density matrices, and although much simpler both conceptually and from a computational point of view is able to yield quite accurate representation of the macromolecular CD signal. However, care should be taken in using the approximate dipole model, especially in case of strongly coupled systems, for instance closely packed chromophore, for which the explicit calculation of the coupling via the density matrix approach may be necessary to avoid the model breakdown. Indeed, it is known that for distances of the interacting units smaller than 5 Å an overestimation of the excitonic coupling by the dipole model can be experienced.

## Nucleic Acids

### Circular Dichroism of B-DNA

The specific CD signals of B-DNA are directly associated to its famous double helical structure. Modeling the CD spectral features that can be directly matched with experiments is still far from being straightforward and suffers many drawbacks: (i) most

experimental studies of phenomena in DNA are performed on non-specific sequences such as calf thymus DNA (Kankia et al., 2001). Since the CD signal is directly related to the sequence, it is thus impossible to provide a meaningful model without considering all the possible nucleobases sequences. Moreover, in the case of the interaction of DNA with an organic molecule, the binding will often be non-specific. Since the length of B-DNA strands that are commonly modeled using MD ranges from 10 to 32 base-pairs, the number of possible binding sites increases drastically, thus reproducing every possible binding configuration becomes an unfeasible task. Based on this, model systems must be considered made up of a selected sequence with a specific binding configuration (Basu and Suresh Kumar, 2014). That is why, most of the CD modeling that have been performed so far were of either well-characterized sequences (such as the X-ray crystal structure—pdb code: 1BNA—or NMR structure—pdb code: 1K9H) or of the simplest nucleic acid sequences, being hetero- and homo-polymers of adenine (A)-thymine (T) and guanine (G)-cytosine (C) (Drew et al., 1981).

The first reported modeling of B-DNA CD employing *ab-initio* methods has been performed by (Miayahara et al., 2013) on the structure of a poly(dGCCGGGC) double strand obtained by X-ray crystallography (Heinemann et al., 1987). In this study, the excited states energies were computed using the Symmetry-Adapted Cluster (SAC)-Configuration Interaction (CI) theory after pre-optimization of the system at the DFT/B3LYP/6-31G(d,p) level. Solvation has been taken into account using Polarizable Continuum Model (PCM) implemented in Gaussian. In this work, either eight excited states have been computed in dimer models (two stacked nucleobases to study the influence of  $\pi$ -stacking interactions or a nucleobase pair including hydrogen bonding) or 14 excited states have been considered in a tetramer model made up of two stacked nucleobase pairs (containing both  $\pi$ -stacking and hydrogen bonding interactions). For the smaller dimer models, it has been demonstrated, on one side, that hydrogen bonding within base pairs is accountable for the change in excitation energies compared to single nucleobases and, on the other side, that  $\pi$ -stacking is responsible for the sign of the CD signal. Finally, the computation performed on the tetramer model resulted in a CD spectrum in very good agreement with experiment.

Employing a Complex Polarization Propagator (CPP), Di Meo et al. modeled the CD spectra of a poly(dA.dT)<sub>20</sub> B-DNA double strand (Di Meo et al., 2015). Excited states of model base pairs dimers and trimers have been computed at the TD-DFT/CAM-B3LYP/aug-cc-pVDZ level of theory considering 30 excited states. Also, the sampling of the B-DNA conformational space has been performed through 100 ns long MD simulations using the amberff12-bsc0 force field. Their results show a very good agreement with the experimental CD fingerprint of the adenine-thymine base pairs homo-polymers in the near-ultraviolet region, corresponding to two positive bands at 260 and 283 nm and a more intense negative band at 249 nm. Furthermore, Norman et al. (2015) tackled the case of the 147 base pairs long nucleosomal DNA (pdb code: 1KX5) employing a similar procedure. The results demonstrated the influence of the super-helical configuration on the CD spectra, including a

hypsochromic shift of the band at 269 nm and a strong decrease of its intensity.

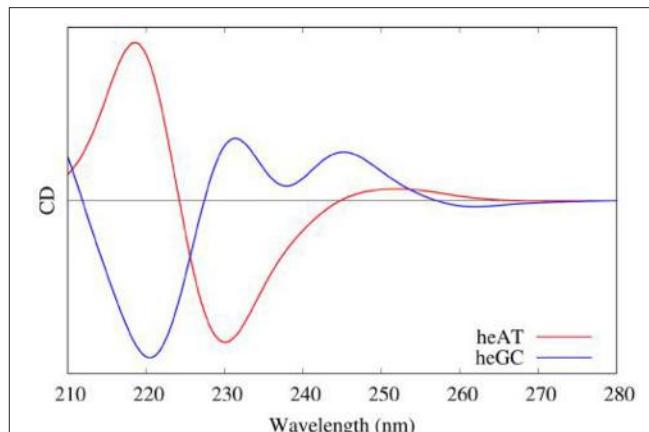
Another approach by Padula et al. to model the CD spectrum of two 10 base-pairs long B-DNA guanine-cytosine hetero- and homo-polymers, relied on the explicit calculation of an exciton coupling in the DNA double strand (Padula et al., 2016). Experimental structures obtained by X-ray crystallography or NMR have been used as starting configurations. The excited states were computed at the TD-DFT/M06-2X/6-31G(d) level of theory taking into account the environment effects and interactions between chromophores using the MMPol method, which allow to treat the environment through a polarizable embedding. Moreover, only bright  $\pi-\pi^*$  excited states have been selected in this protocol. Afterwards, the excitonic Hamiltonian was generated coupling the full transition densities of each partition to obtain, after its diagonalization, the final CD spectra. The final spectrum results (averaged over 270 snapshots extracted from a 90 ns long MD simulation employing the parmbsc0 force field for DNA) recover the overall band shapes and is comparable with experiment.

Our protocol for the simulation of nucleic acids' CD signals, also relied on the 4 steps described hereinbefore with first 10 ns of MD simulation on an *in-silico* built 15 base pairs long DNA double strand (bsc0 force field for nucleic acids). Considering that excited states in B-DNA are relatively local, they can be treated in the framework of the Frenkel excitons theory. In such case, excited states can be computed on a series of decoupled partitions (for example single nucleobases) which will be recoupled by building an exciton coupling matrix. Each excited state for each partition will be coupled with all the other excited states using the methodology presented in (Gattuso et al., 2015, 2016b). The excited states have been obtained using TD-DFT/M06-2X/6-311+G(d) on 40 snapshots, divided in up to 8 different partitions made up of nucleobase pairs.

Moreover, QM/MM embedding was used to take into account the environment, in order to tune the electronic properties of each partition. Polarization was accounted using ERS as presented before. Then the resulting rotatory strength have been convoluted using Gaussian shaped functions of 0.2 eV Full Width at Half Maximum (FWHM). The results obtained on heteropolymers of AT and GC base pairs are presented in **Figure 11** and nicely underline the sequence effects on the DNA CD spectra.

### Circular Dichroism of G-quadruplexes

G-quadruplexes (G4) can be sorted in three families,—parallel, hybrid, and antiparallel—corresponding to their specific nucleic acids arrangements with each family possessing sub-members (Randazzo et al., 2012). Their CD spectra have been deeply described (del Villar-Guerra et al., 2017) and can be easily decomposed in a series of nucleic acids interactions components with a major contribution from the central guanine core. Each family can be differentiated by a specific sequence of spectral bands: (i) parallel, characterized by a positive band at 264 nm and a negative one at 245 nm; (ii) hybrid, characterized by positive bands at 295 and 260 nm and a negative one at 245 nm; (iii) antiparallel, characterized by a positive band at 295 nm and a negative one at 260 nm.



**FIGURE 11** | Modeled CD spectra of a heteropolymer of adenine-thymine (red) and a heteropolymer of guanine-cytosine (blue) base pairs.

In (Loco et al., 2016a) a similar procedure as in (Padula et al., 2016), has been used to model the CD spectrum of an antiparallel G4 (pdb code: 143D). However, this study did not consider a sampling of the conformations using MD simulations, but was instead performed directly on the experimental (crystallographic) structure. In the excitonic Hamiltonian, guanines have been considered as individual partitions brought in interaction using the MMPol polarizable embedding (TD-DFT/M06-2X/6-311+G(d) level, 10 excited states calculated for each guanine). The resulting spectrum almost matches the experiment one.

Using the same protocol as for B-DNA, including sampling the conformational space through MD, and considering as partitions the four triplets of stacked guanines of the G4 core, we also proved (Gattuso et al., 2016b) that the excitons coupling gives a satisfactory simulated CD spectrum of the three G4 families with a clear evidence of their spectral features (see **Figure 12**). Our starting structures were X-ray crystal structures (pdb code: 1KF1 for parallel, 2HY9 for hybrid and 143D for antiparallel configurations) which possess the main spectral features detailed before. Moreover, our results demonstrated that to model the CD spectra of G4, the explicit environment has to be considered with great care, since these aggregates are stabilized by two central cations ( $K^+$  for hybrid and parallel configurations and  $Na^+$  for the antiparallel one) electrostatically bound to the guanines of the core.

### Peptides

#### Circular Dichroism of Alpha-Helices

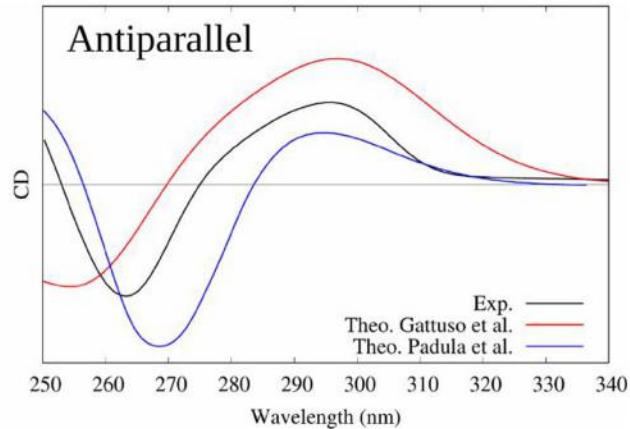
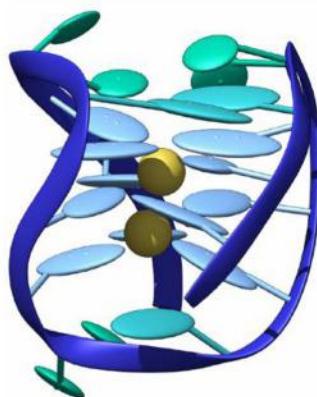
Circular dichroism of protein secondary structures also draws strong interest since it can be, as DNA, directly related to the global structure of the bio-system (Greenfield, 2007). Indeed, it provides information about open and closed conformations in the case of binding with a ligand, or even protein-protein and protein-DNA bindings (Carpenter et al., 2009). Moreover, CD can probe reorganizations in the case of environmental changes, such as increase of

ionic strength or temperature (Jirgensons, 1980). So far, the main attempts to model the CD spectra of macromolecular amino acids arrangements have targeted alpha helices and their specific spectral band shapes, consisting of two negative bands at 208 and 222 nm and an intense negative band at 193 nm.

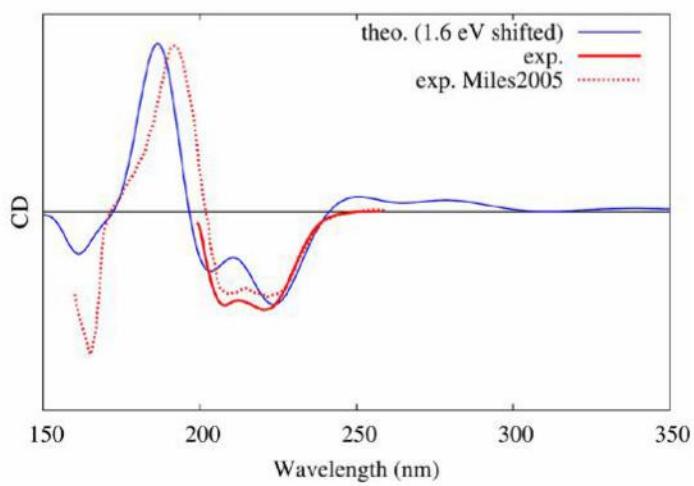
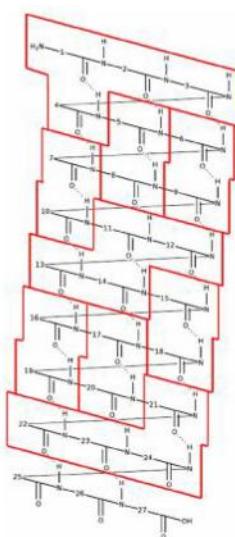
Kaminský et al. (2011) performed the first CD modeling of relatively long alpha helices. Their protocol previewed to first sampling the conformational space of *in-silico* built alpha helices using MD simulations, followed by a computation of the excited states and rotatory strengths. Especially, Ac-(Ala)<sub>n</sub>-NH-Me structures (with *n* going from 2 to 19) were studied by Transition Dipole Coupling (TDC) and TD-DFT, using the B3LYP functional and various basis sets, focusing on the difference between vacuum and the COSMO

method to include solvent effects. Even though the resulting band shape is reasonably comparable with experimental CD, the MD simulations has been performed on *in-silico* built starting configurations for a maximum of 50 ps. Considering the high flexibility and necessity to allow the system (and solvent) to reorganize, this time length cannot definitely be considered as sufficient to sample the ensemble of configurations of such system. They mainly concluded that considering several consecutive amino acids is important to allow exciton delocalization.

Our protocol (Gattuso et al., 2017b), described hereinbefore, combined, as it was the case for nucleic acids, conformational sampling through MD, followed by the diagonalization of an excitonic coupling Hamiltonian whose elements are obtained by QM/MM calculations on individual subunits. In detail, in



**FIGURE 12** | Experimental and modeled CD spectra of the antiparallel G4 structure (pdb code: 143D).



**FIGURE 13** | Experimental (Miles and Wallace, 2006) and modeled (Gattuso et al., 2017b) CD spectra of an alpha helical peptide. On the left, the selected partitioning scheme to apply the Frenkel excitons scheme is shown.

the case of a polypeptide comprising 27 amino acids and assuming a dominant alpha-helix conformation, we performed: (i) 80 ns of MD sampling using amber99 force field; (ii) computation of excited states properties on partitions at the TD-DFT/M06-2X/6-311++G(d,p) level of theory (each computation was performed using hybrid QM/MM method including ERS); (iii) coupling of excitons transition dipole moments; (iv) convolution of rotatory strengths by Gaussian shaped functions of 0.4 eV at FWHM. In this study, we demonstrated that the final CD spectra are highly sensitive to the partitioning scheme with a convergence obtained for loop subunits (four consecutive amino acids, see **Figure 13**). Also, the simplicity of the Hamiltonian requires to consider with great care the arrangement of partitions. Indeed, the most accurate results were obtained considering a hybrid partitioning including loops of amino acids (quadruplets) and hydrogen bounded dimers (couples). In fact, this allowed to retain the higher accuracy obtained for bigger molecular partitions, while keeping an alpha-helical arrangement of subunits' centers of charges. Most notably, the coupling of MD sampling and QM/MM calculations, also allowed to recover the modification of CD signals induced by the partial breaking of the alpha-helix, due to the coupling of the polypeptide with a photoactive switch undergoing *cis/trans* isomerization.

## GENERAL CONCLUSION

In this review we gather recent work done in our group dealing with the description of electronic excited states in complex molecular systems. We present three main applications, namely linear absorption, non-linear optical properties (mainly

two-photon absorption) and circular dichroism together with the specific methodological development that has been made. We show that when an adequate description of both surrounding and dynamical effects is chosen, our calculations agree nicely with the available experimental data and can thus be used for prediction when experimental data are missing, or at least, as in the case of the structurally sensitive electronic CD, they allow for a simple and straightforward one-to-one mapping with the system structural features. As a general take home message, one should be aware that a good sampling of the conformational space and a polarizable embedding scheme are required to obtain reliable results.

## AUTHOR CONTRIBUTIONS

AM: was responsible for the part on linear absorption; MM: took care of the non-linear optical properties; HG: handled the section devoted to circular dichroism; XA: was responsible for the remaining parts of the manuscript.

## FUNDING

MM is thankful to the French National Research Agency (ANR) for a grant under the DeNeTheor project, and to the Universidad de La Rioja (UR) for a postdoctoral contract.

## ACKNOWLEDGMENTS

The authors would like to gratefully thank the Phi-Science student association for providing an adequate working space to the authors while they wrote this manuscript.

## REFERENCES

- Aidas, K., Angeli, C., Bak, K. L., Bakken, V., Bast, R., Boman, L., et al. (2014). The Dalton quantum chemistry program system. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 4, 269–284. doi: 10.1002/wcms.1172
- Antoine, R., and Bonačić-Koutecký, V. (2018). “Measurement techniques of optical nonlinearities—two-photon absorption/fluorescence and hyper-rayleigh scattering,” in *Liganded Silver and Gold Quantum Clusters, Towards a New Class of Nonlinear Optical Nanomaterials* (Cham: Springer), 49–62.
- Antonov, L., Kamada, K., Ohta, K., and Kamounah, F. S. (2003). A systematic femtosecond study on the two-photon absorbing D-π-A molecules-π-bridge nitrogen insertion and strength of the donor and acceptor groups. *Phys. Chem. Chem. Phys.* 5, 1193–1197. doi: 10.1039/b211260d
- Artal, P., Manzanera, S., Komar, K., Gambín-Regadera, A., and Wojtkowski, M. (2017). Visual acuity in two-photon infrared vision. *Optica* 4, 1488–1491. doi: 10.1364/OPTICA.4.001488
- Autenrieth, F., Tajkhorshid, E., Baudry, J., and Luthey-Schulten, Z. (2004). Classical force field parameters for the heme prosthetic group of cytochrome c. *J. Comput. Chem.* 25, 1613–1622. doi: 10.1002/jcc.20079
- Barsu, C., Cheaib, R., Chambert, S., Queneau, Y., Maury, O., Cottet, D., et al. (2009). Neutral push-pull chromophores for nonlinear optical imaging of cell membranes. *Org. Biomol. Chem.* 8, 142–150. doi: 10.1039/B915654B
- Barsu, C., Fortrie, R., Nowika, K., Baldeck, P. L., Vial, J.-C., Barsella, A., et al. (2006). Synthesis of chromophores combining second harmonic generation and two photon induced fluorescence properties. *Chem. Commun.* 89, 4744–4746. doi: 10.1039/B610557B
- Basu, P., and Suresh Kumar, G. (2014). Elucidation of the DNA binding specificity of the natural plant alkaloid chelerythrine: a biophysical approach. *J. Photochem. Photobiol. B Biol.* 138, 282–294. doi: 10.1016/j.jphotobiol.2014.06.005
- Bednarska, J., Roztoczynska, A., Bartkowiak, W., and Zaleśny, R. (2013). Comparative assessment of density functionals for excited-state dipole moments. *Chem. Phys. Lett.* 584, 58–62. doi: 10.1016/j.cplett.2013.08.079
- Beerepoot, M. T., Friese, D. H., List, N. H., Kongsted, J., and Ruud, K. (2015). Benchmarking two-photon absorption cross sections: performance of CC2 and CAM-B3LYP. *Phys. Chem. Chem. Phys.* 17, 19306–19314. doi: 10.1039/C5CP03241E
- Beerepoot, M. T. P., Friese, D. H., and Ruud, K. (2014). Intermolecular charge transfer enhances two-photon absorption in yellow fluorescent protein. *Phys. Chem. Chem. Phys.* 16, 5958–5964. doi: 10.1039/c3cp55205e
- Benninger, R. K. P., and Pistor, D. W. (2013). Two-photon excitation microscopy for the study of living cells and tissues. *Curr. Protoc. Cell Biol.* Chapter 4, Unit 4.11.1–24. doi: 10.1002/0471143030.cb0411s59
- Birge, R. R., and Pierce, B. M. (1986). Semiclassical time-dependent theory of two-photon spectroscopy. The effect of dephasing in the virtual level on the two-photon excitation spectrum of isotachysterol. *Int. J. Quantum Chem.* 29, 639–656. doi: 10.1002/qua.560290407
- Caprasecca, S., Jurinovich, S., Viani, L., Curutchet, C., and Mennucci, B. (2014). Geometry optimization in polarizable QM/MM models: the induced dipole formulation. *J. Chem. Theory Comput.* 10, 1588–1598. doi: 10.1021/ct50021d
- Case, D. A. C., Darden, T. E., Gohlke, T., Luo, H., Merz, R., Onufriev, K. M., et al. (2017). *AMBER 15*. San Francisco, CA: University of California.

- Carpenter, M. L., Oliver, A. W., and Kneale, G. G. (2009). Circular dichroism for the analysis of protein – DNA interactions. *Methods Mol. Biol.* 148, 503–510. doi: 10.1385/0-89603-256-6:339
- Cerezo, J., Avila Ferrer, F. J., Prampolini, G., and Santoro, F. (2015). Modeling solvent broadening on the vibronic spectra of a series of coumarin dyes. from implicit to explicit solvent models. *J. Chem. Theory Comput.* 11, 5810–5825. doi: 10.1021/acs.jctc.5b00870
- Cerezo, J., Mazzeo, G., Longhi, G., Abbate, S., and Santoro, F. (2016). Quantum-classical calculation of vibronic spectra along a reaction path: the case of the ecd of easily interconvertible conformers with opposite chiral responses. *J. Phys. Chem. Lett.* 7, 4891–4897. doi: 10.1021/acs.jpclett.6b02484
- Chantzis, A., Very, T., Daniel, C., Monari, A., and Assfeld, X. (2013). Theoretical evidence of photo-induced charge transfer from DNA to intercalated ruthenium (II) organometallic complexes. *Chem. Phys. Lett.* 578, 133–137. doi: 10.1016/j.cplett.2013.05.068
- Chantzis, A., Very, T., Despax, S., Issenhuth, J.-T., Boeglin, A., Hébraud, P., et al. (2014). UV-vis absorption spectrum of a novel Ru(II) complex intercalated in DNA:  $[\text{Ru}(2,2'\text{-bipy})(\text{dppz})(2,2'\text{-ArPy})]^+$ . *J. Mol. Model.* 20:2082. doi: 10.1007/s00894-014-2082-2
- Christiansen, O., Koch, H., and Jørgensen, P. (1995). The second-order approximate coupled cluster singles and doubles model CC2. *Chem. Phys. Lett.* 243, 409–418. doi: 10.1016/0009-2614(95)00841-Q
- Dahl, J. P., and Springborg, M. (1988). The Morse oscillator in position space, momentum space, and phase space. *J. Chem. Phys.* 88, 4535–4547. doi: 10.1063/1.453761
- Dans, P. D., Ivani, I., Hospital, A., Portella, G., González, C., and Orozco, M. (2017). How accurate are accurate force-fields for B-DNA? *Nucleic Acids Res.* 45, 4217–4230. doi: 10.1093/nar/gkw1355
- De Boni, L., Misoguti, L., Zilio, S. C., and Mendonça, C. R. (2005). Degenerate two-photon absorption spectra in azoaromatic compounds. *Chem. Phys. Chem.* 6, 1121–1125. doi: 10.1002/cphc.200400391
- Deisseroth, K. (2011). Optogenetics. *Nat. Methods* 8, 26–29. doi: 10.1038/nmeth.f.324
- del Villar-Guerra, R., Trent, J. O., and Chaires, J. B. (2017). G-quadruplex secondary structure from circular dichroism spectroscopy. *Angew. Chemie Int. Ed.* doi: 10.1002/anie.201709184. [Epub ahead of print].
- de Reguadati, S., Pahapill, J., Mikhailov, A., Stepanenko, Y., and Rebane, A. (2016). High-accuracy reference standards for two-photon absorption in the 680–1050 nm wavelength range. *Opt. Express* 24, 9053–9066. doi: 10.1364/OE.24.009053
- Di Meo, F., D. M., Pedersen, M. N., Rubio-Magnieto, J., Surin, M., Linares, M., and Norman, P. (2015). DNA electronic circular dichroism on the inter-base pair scale: an experimental-theoretical case study of the at homo-oligonucleotide. *J. Phys. Chem. Lett.* 6, 355–359. doi: 10.1021/jz502696t
- Dixit, S. N., Guo, D., and Mazumdar, S. (1991). Essential-states mechanism of optical nonlinearity in  $\pi$ -conjugated polymers. *Phys. Rev. B* 43, 6781–6784. doi: 10.1103/PhysRevB.43.6781
- Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K., et al. (1981). Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 78, 2179–2183. doi: 10.1073/pnas.78.4.2179
- Dumont, É., and Monari, A. (2015). Interaction of palmitate with DNA: an environmentally controlled phototherapy drug. *J. Phys. Chem. B* 119, 410–419. doi: 10.1021/jp5088515
- Ehrlich, J. E., Wu, X. L., Lee, I.-Y. S., Hu, Z.-Y., Röckel, H., Marder, S. R., et al. (1997). Two-photon absorption and broadband optical limiting with bis-donor stilbenes. *Opt. Lett.* 22:1843. doi: 10.1364/OL.22.001843
- Entwistle, C. D., Collings, J. C., Steffen, A., Pálsson, L.-O., Beeby, A., Albesa-Jové, D., et al. (2009). Syntheses, structures, two-photon absorption cross-sections and computed second hyperpolarisabilities of quadrupolar A- $\pi$ -A systems containing E-dimesitylborylethenyl acceptors. *J. Mater. Chem.* 19:7532. doi: 10.1039/b905719f
- Etienne, T., Assfeld, X., and Monari, A. (2014a). New insight into the topology of excited states through detachment/attachment density matrices-based centroids of charge. *J. Chem. Theory Comput.* 10, 3906–3914. doi: 10.1021/ct500400s
- Etienne, T., Assfeld, X., and Monari, A. (2014b). QM/MM calculation of absorption spectra of complex systems: the case of human serum albumin. *Comput. Theor. Chem.* 1040–1041, 360–366. doi: 10.1016/j.comptc.2014.01.009
- Etienne, T., Assfeld, X., and Monari, A. (2014c). Toward a quantitative assessment of electronic transitions" charge-transfer character. *J. Chem. Theory Comput.* 10, 3896–3905. doi: 10.1021/ct5003994
- Etienne, T., Gattuso, H., Monari, A., and Assfeld, X. (2014d). QM/MM modeling of Harmane cation fluorescence spectrum in water solution and interacting with DNA. *Comput. Theor. Chem.* 1040–1041, 367–372. doi: 10.1016/j.comptc.2014.03.026
- Etienne, T., Very, T., Perpète, E. A., Monari, A., and Assfeld, X. (2013). A QM/MM study of the absorption spectrum of harmane in water solution and interacting with DNA: the crucial role of dynamic effects. *J. Phys. Chem. B* 117, 4973–4980. doi: 10.1021/jp4017882
- Fernandez, J. M., Bilgin, M. D., and Grossweiner, L. I. (1997). Singlet oxygen generation by photodynamic agents. *J. Photochem. Photobiol. B Biol.* 37, 131–140. doi: 10.1016/S1011-1344(96)07349-6
- Forche, F., Ahlrichs, R., Hättig, C., Klopper, W., Sierka, M., and Weigend, F. (2014). Turbomole. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 4, 91–100. doi: 10.1002/wcms.1162
- Galindo-Murillo, R., García-Ramos, J. C., Ruiz-Azuara, L., Cheatham, T. E., and Cortés-Guzmán, F. (2015). Intercalation processes of copper complexes in DNA. *Nucleic Acids Res.* 43, 5364–5376. doi: 10.1093/nar/gkv467
- García-Iriepa, C., Gueye, M., Léonard, J., Martínez-López, D., Campos, P. J., Frutos, L. M., et al. (2016). A biomimetic molecular switch at work: coupling photoisomerization dynamics to peptide structural rearrangement. *Phys. Chem. Chem. Phys.* 18, 6742–6753. doi: 10.1039/C5CP07599H
- García-Iriepa, C., and Marazzi, M. (2017). Level of theory and solvent effects on DASA absorption properties prediction: comparing TD-DFT, CASPT2 and NEVPT2. *Materials* 10:1025. doi: 10.3390/ma10091025
- García-Iriepa, C., Marazzi, M., Frutos, L. M., and Sampedro, D. (2013). E/Z Photochemical switches: syntheses, properties and applications. *RSC Adv.* 3:6241. doi: 10.1039/c2ra22363e
- Gattuso, H., Assfeld, X., and Monari, A. (2015). Modeling DNA electronic circular dichroism by QM/MM methods and Frenkel Hamiltonian. *Theor. Chem. Acc.* 134:36. doi: 10.1007/s00214-015-1640-8
- Gattuso, H., Dumont, E., Marazzi, M., and Monari, A. (2016a). Two-photon-absorption DNA sensitization: via solvated electron production: unraveling photochemical pathways by molecular modeling and simulation. *Phys. Chem. Chem. Phys.* 18, 18598–18606. doi: 10.1039/C6CP02592G
- Gattuso, H., Marazzi, M., Dehez, F., and Monari, A. (2017a). Deciphering the photosensitization mechanisms of hypericin towards biological membranes. *Phys. Chem. Chem. Phys.* 19, 23187–23193. doi: 10.1039/C7CP03723F
- Gattuso, H., Monari, A., and Marazzi, M. (2017b). Photophysics of chlorin e6: from one- and two-photon absorption to fluorescence and phosphorescence. *RSC Adv.* 7, 10992–10999. doi: 10.1039/C6RA28616J
- Gattuso, H., Spinello, A., Terenzi, A., Assfeld, X., Barone, G., and Monari, A. (2016b). Circular dichroism of DNA G-quadruplexes: combining modeling and spectroscopy to unravel complex structures. *J. Phys. Chem. B* 120, 3113–3121. doi: 10.1021/acs.jpch.6b00634
- Göppert-Mayer, M. (1931). Über elementarakte mit zwei quantensprüngen. *Ann. Phys.* 401, 273–294. doi: 10.1002/andp.193140101303
- Greenfield, N. J. (2007). Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* 1, 2876–2890. doi: 10.1038/nprot.2006.202
- Guo, Y., Beyle, F. E., Bold, B. M., Watanabe, H. C., Koslowski, A., Thiel, W., et al. (2016). Active site structure and absorption spectrum of channelrhodopsin-2 wild-type and C128T mutant. *Chem. Sci.* 7, 3879–3891. doi: 10.1039/C6SC00468G
- Heflin, J. R., Wong, K. Y., Zamani-Khamiri, O., and Garito, A. F. (1988). Nonlinear optical properties of linear chains and electron-correlation effects. *Phys. Rev. B* 38, 1573–1576. doi: 10.1103/PhysRevB.38.1573
- Heinemann, U., Lauble, H., Frank, R., and Blöcker, H. (1987). Crystal structure analysis of an A-DNA fragment at 1.8 Å resolution: d(GCCCGGGC). *Nucleic Acids Res.* 15, 9531–9550. doi: 10.1093/nar/15.22.9531
- Helmy, S., Leibfarth, F. A., Oh, S., and Poelma, J. E., (2014a). Photoswitching using visible light: a new class of organic photochromic molecules. *J. Am. Chem. Soc.* 136, 8169–8172. doi: 10.1021/ja503016b
- Helmy, S., Oh, S., Leibfarth, F. A., Hawker, C. J., and Read de Alaniz, J. (2014b). Design and synthesis of donor–acceptor stenhouse adducts: a visible

- light photoswitch derived from furfural. *J. Org. Chem.* 79, 11316–11329. doi: 10.1021/jo502206g
- Hirakawa, K., and Hirano, T. (2008). The microenvironment of DNA switches the activity of singlet oxygen generation photosensitized by berberine and palmatine. *Photochem. Photobiol.* 84, 202–208. doi: 10.1111/j.1751-1097.2007.00220.x
- Holmgård List, N., Knoops, J., Rubio-Magnieto, J., Idé, J., Beljonne, D., Norman, P., et al. (2017). Origin of DNA-induced circular dichroism in a minor-groove binder. *J. Am. Chem. Soc.* 139, 14947–14953. doi: 10.1021/jacs.7b05994
- Hontani, Y., Marazzi, M., Stehfest, K., Mathes, T., van Stokkum, I. H. M., Elstner, M., et al. (2017). Reaction dynamics of the chimeric channelrhodopsin C1C2. *Sci. Rep.* 7:7217. doi: 10.1038/s41598-017-07363-w
- Hu, W., Xie, M., Zhao, H., Tang, Y., Yao, S., He, T., et al. (2018). Nitric oxide activatable photosensitizer accompanying extremely elevated two-photon absorption for efficient fluorescence imaging and photodynamic therapy. *Chem. Sci.* 9, 999–1005. doi: 10.1039/C7SC04044J
- Impróta, R., Barone, V., and Santoro, F. (2007). Ab initio calculations of absorption spectra of large molecules in solution: coumarin C153. *Angew. Chemie Int. Ed.* 46, 405–408. doi: 10.1002/anie.200602907
- Ivani, I., Dans, P. D., Noy, A., Pérez, A., Faustino, I., Hospital, A., et al. (2015). Parmbsc1: a refined force field for DNA simulations. *Nat. Methods* 13, 55–58. doi: 10.1038/nmeth.3658
- Jacquemin, D., Perpète, E. A., Laurent, A. D., Assfeld, X., and Adamo, C. (2009). Spectral properties of self-assembled squaraine-tetralactam: a theoretical assessment. *Phys. Chem. Chem. Phys.* 11, 1258–1262. doi: 10.1039/b817720a
- Jeelan, S., Reddy, R. C. J., Maheswaran, T., Asokan, G. S., Dany, A., and Anand, B. (2014). Theranostics: a treasured tailor for tomorrow. *J. Pharm. Biomed. Sci.* 6, S6–S8. doi: 10.4103/0975-7406.137249
- Jiang, J., Lai, Z., Wang, J., and Mukamel, S. (2014). Signatures of the protein folding pathway in two-dimensional ultraviolet spectroscopy. *J. Phys. Chem. Lett.* 5, 1341–1346. doi: 10.1021/jz5002264
- Jirgensons, B. (1980). Circular dichroism study on structural reorganization of lectins by sodium dodecyl sulfate. *BBA Protein Struct.* 623, 69–76. doi: 10.1016/0005-2795(80)90008-2
- Jurinovich, S., Pescitelli, G., Di Bari, L., and Mennucci, B. (2014). A TDDFT/MMPol/PCM model for the simulation of exciton-coupled circular dichroism spectra. *Phys. Chem. Chem. Phys.* 16, 16407–16418. doi: 10.1039/C3CP55428G
- Kaiser, W., and Garrett, C. G. B. (1961). Two-photon excitation in CaF<sub>2</sub>: Eu 2+. *Phys. Rev. Lett.* 7, 229–231. doi: 10.1103/PhysRevLett.7.229
- Kaminský, J., Kubelka, J., and Bour, P. (2011). Theoretical modeling of peptide  $\alpha$ -helical circular dichroism in aqueous solution. *J. Phys. Chem. A* 115, 1734–1742. doi: 10.1021/jp110418w
- Kankia, B. I., Buckin, V., and Bloomfield, V. A. (2001). Hexamminecobalt(III)-induced condensation of calf thymus DNA: circular dichroism and hydration measurements. *Nucleic Acids Res.* 29, 2795–2801. doi: 10.1093/nar/29.13.2795
- Klamt, A., and Schüürmann, G. (1993). COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Trans. 2*, 799–805. doi: 10.1039/P29930000799
- Knippenberg, S., Rehn, D. R., Wormit, M., Starcke, J. H., Rusakova, I. L., Trofimov, A. B., et al. (2012). Calculations of nonlinear response properties using the intermediate state representation and the algebraic-diagrammatic construction polarization propagator approach: two-photon absorption spectra. *J. Chem. Phys.* 136:64107. doi: 10.1063/1.3682324
- Krylov, A. I., and Gill, P. M. W. (2013). Q-Chem: an engine for innovation. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 3, 317–326. doi: 10.1002/wcms.1122
- Kypr, J., Kejnovská, I., Renčík, D., and Vorlíčková, M. (2009). Survey and Summary circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* 37, 1713–1725. doi: 10.1093/nar/gkp026
- Laurent, A. D., and Jacquemin, D. (2013). TD-DFT benchmarks: a review. *Int. J. Quantum Chem.* 113, 2019–2039. doi: 10.1002/qua.24438
- Le Bahers, T., Adamo, C., and Ciofini, I. (2011). A qualitative index of spatial extent in charge-transfer excitations. *J. Chem. Theory Comput.* 7, 2498–2506. doi: 10.1021/ct200308m
- Li, J., Deng, M., Voronine, D. V., Mukamel, S., and Jiang, J. (2015). Two-dimensional near ultraviolet (2DNUV) spectroscopic probe of structural-dependent exciton dynamics in a protein. *J. Phys. Chem. B* 119, 1314–1322. doi: 10.1021/jp509314y
- Li, X. Y., Czernuszewicz, R. S., Spiro, T. G., and Kincaid, J. R. (1989). Consistent porphyrin force field. 3. out-of-plane modes in the resonance raman spectra of planar and ruffled nickel octaethylporphyrin. *J. Am. Chem. Soc.* 111, 7012–7023. doi: 10.1021/ja00200a018
- Liu, R., Liu, M., Hood, D., Chen, C.-Y., MacNevin, C. J., Holten, D., et al. (2018). Chlorophyll-Inspired red-region fluorophores: building block synthesis and studies in aqueous media. *Molecules* 23:130. doi: 10.3390/molecules23010130
- Loco, D., Jurinovich, S., Di Bari, L., and Mennucci, B. (2016a). A fast but accurate excitonic simulation of the electronic circular dichroism of nucleic acids: how can it be achieved? *Phys. Chem. Chem. Phys.* 18, 866–877. doi: 10.1039/C5CP06341H
- Loco, D., Polack, É., Caprasecca, S., Lagardère, L., Lipparrini, F., Piquemal, J. P., et al. (2016b). A QM/MM approach using the AMOEBA polarizable embedding: from ground state energies to electronic excitations. *J. Chem. Theory Comput.* 12, 3654–3661. doi: 10.1021/acs.jctc.6b00385
- Makarov, N. S., Drobizhev, M., and Rebane, A. (2008). Two-photon absorption standards in the 550–1600 nm excitation wavelength range. *Opt. Express* 16:4029. doi: 10.1364/OE.16.004029
- Martin, R. L. (2003). Natural transition orbitals. *J. Chem. Phys.* 118, 4775–4777. doi: 10.1063/1.1558471
- Miles, A. J., and Wallace, B. A. (2006). Synchrotron radiation circular dichroism spectroscopy of proteins and applications in structural and functional genomics. *Chem. Soc. Rev.* 35, 39–51. doi: 10.1039/B316168B
- Miyahara, T., Nakatsuji, H., and Sugiyama, H. (2013). Helical structure and circular dichroism spectra of DNA: a theoretical study. *J. Phys. Chem. A* 117, 42–55. doi: 10.1021/jp3085556
- Mojzisova, H., and Vermot, J. (2011). When multiphoton microscopy sees near infrared. *Curr. Opin. Genet. Dev.* 21, 549–557. doi: 10.1016/j.gde.2011.08.004
- Monari, A., Rivail, J. L., and Assfeld, X. (2013). Theoretical modeling of large molecular systems. Advances in the local self consistent field method for mixed quantum mechanics/molecular mechanics calculations. *Acc. Chem. Res.* 46, 596–603. doi: 10.1021/ar300278j
- Monari, A., Very, T., Rivail, J. L., and Assfeld, X. (2012a). A QM/MM study on the spinach plastocyanin: redox properties and absorption spectra. *Comput. Theor. Chem.* 990, 119–125. doi: 10.1016/j.comptc.2011.11.026
- Monari, A., Very, T., Rivail, J. L., and Assfeld, X. (2012b). Effects of mutations on the absorption spectra of copper proteins: a QM/MM study. *Theor. Chem. Acc.* 131, 1–9. doi: 10.1007/s00214-012-1221-z
- Nanda, K. D., and Krylov, A. I. (2015). Two-photon absorption cross sections within equation-of-motion coupled-cluster formalism using resolution-of-the-identity and Cholesky decomposition representations: theory, implementation, and benchmarks. *J. Chem. Phys.* 142:64118. doi: 10.1063/1.4907715
- Negres, R. A., Hales, J. M., Kobayakov, A., Hagan, D. J., and Van Stryland, E. W. (2002). Experiment and analysis of two-photon absorption spectroscopy using a white-light continuum probe. *IEEE J. Quantum Electron.* 38, 1205–1216. doi: 10.1109/JQE.2002.802448
- Nogueira, J. J., Plasser, F., and González, L. (2017). Electronic delocalization, charge transfer and hypochromism in the UV absorption spectrum of polyadenine unravelled by multiscale computations and quantitative wavefunction analysis. *Chem. Sci.* 8, 5682–5691. doi: 10.1039/C7SC01600J
- Norman, P., Parello, J., Polavarapu, P. L., and Linares, M. (2015). Predicting near-UV electronic circular dichroism in nucleosomal DNA by means of DFT response theory. *Phys. Chem. Chem. Phys.* 17, 21866–21879. doi: 10.1039/C5CP02481A
- Nyman, E. S., and Hynnninen, P. H. (2004). Research advances in the use of tetrapyrrolic photosensitizers for photodynamic therapy. *J. Photochem. Photobiol. B Biol.* 73, 1–28. doi: 10.1016/j.jphotobiol.2003.10.002
- Okada, T., Sugihara, M., Bondar, A., Elstner, M., Entel, P., and Buss, V. (2004). The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *J. Mol. Biol.* 342, 571–583. doi: 10.1016/j.jmb.2004.07.044
- Orlando, A., and Jorgensen, W. L. (2010). Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions. *Acc. Chem. Res.* 43, 142–151. doi: 10.1021/ar900171c
- Padula, D., Jurinovich, S., Di Bari, L., and Mennucci, B. (2016). Simulation of electronic circular dichroism of nucleic acids: from the structure to the spectrum. *Chem. A Eur. J.* 22, 17011–17019. doi: 10.1002/chem.201602777
- Palczewska, G., Vinberg, F., Stremplewski, P., Bircher, M. P., Salom, D., Komar, K., et al. (2014). Human infrared vision is triggered by two-photon

- chromophore isomerization. *Proc. Natl. Acad. Sci. U.S.A.* 111, E5445–E5454. doi: 10.1073/pnas.1410162111
- Parthenopoulos, D. A., and Rentzepis, P., M. (1989). Three-dimensional optical storage memory. *Science* 245, 843–845. doi: 10.1126/science.245.4920.843
- Paul, S., Heng, P. W., and Chan, L. W. (2013). Optimization in solvent selection for chlorin e6 in photodynamic therapy. *J. Fluoresc.* 23, 283–291. doi: 10.1007/s10895-012-1146-x
- Pawlak, M., Collins, H. A., and Denning, R. G., (2009). Two-photon absorption and the design of two-photon dyes. *Angew. Chemie Int. Ed.* 48, 3244–3266. doi: 10.1002/anie.200805257
- Peach, M. J. G., Benfield, P., Helgaker, T., and Tozer, D. J. (2008). Excitation energies in density functional theory: an evaluation and a diagnostic test. *J. Chem. Phys.* 128:44118. doi: 10.1063/1.2831900
- Peticolas, W. L., Goldsborough, J. P., and Rieckhoff, K. E. (1963). Double photon excitation in organic crystals. *Phys. Rev. Lett.* 10, 43–45. doi: 10.1103/PhysRevLett.10.43
- Randazzo, A., Spada, G. P., and da Silva, M. W. (2012). “Circular dichroism of quadruplex structures,” in *Quadruplex Nucleic Acids*, eds B. Jonathan and G. ChairesDavid (Berlin; Heidelberg: Springer), 67–86.
- Reeve, J. E., Anderson, H. L., and Clays, K. (2010). Dyes for biological second harmonic generation imaging. *Phys. Chem. Chem. Phys.* 12:13484. doi: 10.1039/c003720f
- Rinaldi, D., and Rivail, J. L. (1973). Molecular polarizability and dielectric effect of medium in the liquid phase. Theoretical study of the water molecule and its dimers. *Theor. Chim. Acta* 32, 57–70. doi: 10.1007/BF01209416
- Rivail, J.-L., Monari, A., and Assfeld, X. (eds.). (2015). “The non empirical local self consistent field method: application to Quantum Mechanics/Molecular Mechanics (QM/MM) modeling of large biomolecular systems,” in *Quantum Modeling of Complex Molecular Systems* (Cham: Springer), 343–365.
- Ruiz-Lopez, M. F., Assfeld, X., Garcia, J. I., Mayoral, J. A., and Salvatella, L. (1993). Solvent effects on the mechanism and selectivities of asymmetric Diels-Alder reactions. *J. Am. Chem. Soc.* 115, 8780–8787. doi: 10.1021/ja00072a035
- Ryu, J. H., Jeong, Y.-I., Kim, H. Y., Son, G. M., Lee, H. L., Chung, C.-W., et al. (2018). Enhanced photosensing and photodynamic treatment of colon cancer cells using methoxy poly(ethylene glycol)-conjugated chlorin e6. *J. Nanosci. Nanotechnol.* 18, 1131–1136. doi: 10.1166/jnn.2018.14872
- Sampedro, D., Migani, A., Pepi, A., Busi, E., Basosi, R., Latterini, L., et al. (2004). Design and photochemical characterization of a biomimetic light-driven Z/E switcher. *J. Am. Chem. Soc.* 126, 9349–9359. doi: 10.1021/ja038859e
- Santoro, F., Improta, R., Lami, A., Bloino, J., and Barone, V. (2007). Effective method to compute Franck-Condon integrals for optical spectra of large molecules in solution. *J. Chem. Phys.* 126:84509. doi: 10.1063/1.2437197
- Sengul, O., Boydas, E. B., Pastore, M., Sharmouk, W., Gros, P. C., Catak, S., et al. (2017). Probing optical properties of thiophene derivatives for two-photon absorption. *Theor. Chem. Acc.* 136:67. doi: 10.1007/s00214-017-2094-y
- Shao, Y., Gan, Z., Epifanovsky, E., Gilbert, A. T. B., Wormit, M., Kussmann, J., et al. (2015). Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* 113, 184–215. doi: 10.1080/00268976.2014.952696
- Sheik-Bahae, M., Said, A. A., Wei, T.-H., Hagan, D. J., and Van Stryland, E. W. (1990). Sensitive measurement of optical nonlinearities using a single beam. *IEEE J. Quantum Electron.* 26, 760–769. doi: 10.1109/3.53394
- Shi, Y., Ren, P., Schnieders, M., and Piquemal, J. P. (2015). “Polarizable force fields for biomolecular modeling,” in *Reviews in Computational Chemistry*, eds A. L. Parrill and K. B. Lipkowitz (Hoboken, NJ: John Wiley & Sons, Inc), 51–86.
- Steindal, A. H., Beerepoot, M. T., Ringholm, M., List, N. H., Ruud, K., Kongsted, J., et al. (2016). Open-ended response theory with polarizable embedding: multiphoton absorption in biomolecular systems. *Phys. Chem. Chem. Phys.* 18, 28339–28352. doi: 10.1039/C6CP05297E
- Sun, S., and Dalton, L. (eds.). (2008). *Introduction to organic electronic and optoelectronic materials and devices*. Boca Raton, FL: CRC Press.
- Theer, P., Denk, W., Sheves, M., Lewis, A., and Detwiler, P. B. (2011). Second-harmonic generation imaging of membrane potential with retinal analogues. *Biophys. J.* 100, 232–242. doi: 10.1016/j.bpj.2010.11.021
- Tischer, D., and Weiner, O. D. (2014). Illuminating cell signalling with optogenetic tools. *Nat. Rev. Mol. Cell Biol.* 15, 551–558. doi: 10.1038/nrm3837
- Tomasi, J. (2004). ‘Thirty years of continuum solvation chemistry: a review, and prospects for the near future. *Theor. Chem. Acc.* 112:184. doi: 10.1007/s00214-004-0582-3
- Tomasi, J., Mennucci, B., and Cammi, R. (2005). Quantum mechanical continuum solvation models. *Chem. Rev.* 105, 2999–3093. doi: 10.1021/cr9904009
- Tsai, C.-L., Chen, J.-C., and Wang, W.-J. (2001). Near-infrared absorption property of biological soft tissue constituents. *J. Med. Biol. Eng.* 21, 7–14.
- Turan, H. T., Eken, Y. Y., Marazzi, M., Pastore, M., Aviyente, V., and Monari, A. (2016). Assessing one- and two-photon optical properties of boron containing arenes. *J. Phys. Chem. C* 120, 17916–17926. doi: 10.1021/acs.jpcc.6b05493
- Véry, T., Ambrosek, D., Otsuka, M., Gourlaouen, C., Assfeld, X., Monari, A., et al. (2014). Photophysical properties of ruthenium(II) polypyridyl DNA intercalators: effects of the molecular surroundings investigated by theory. *Chem. Eur. J.* 20, 12901–12909. doi: 10.1002/chem.201402963
- Vorlíčková, M., Kejnovská, I., Bednárová, K., Rencík, D., and Kypr, J. (2012). Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes. *Chirality* 24, 691–698. doi: 10.1002/chir.22064
- Wald, G., Peteanu, L., Mathies, R., and Shank, C. (1968). Molecular basis of visual excitation. *Science* 162, 230–239. doi: 10.1126/science.162.3850.230
- Winkler, K., Simon, C., Finke, M., Blees, K., Birke, M., Szentmáry, N., et al. (2016). Photodynamic inactivation of multidrug-resistant *Staphylococcus aureus* by chlorin e6 and red light ( $\lambda = 670$  nm). *J. Photochem. Photobiol. B Biol.* 162, 340–347. doi: 10.1016/j.jphotobiol.2016.07.007
- Xu, C., and Webb, W. W. (1996). Measurement of two-photon excitation cross sections of molecular fluorophores with data from 690 to 1050 nm. *J. Opt. Soc. Am. B* 13:481. doi: 10.1364/JOSAB.13.000481
- Yanai, T., Tew, D. P., and Handy, N. C. (2004). A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* 393, 51–57. doi: 10.1016/j.cplett.2004.06.011
- Zenkevich, E., Sagun, E., Knyukshto, V., Shulga, A., Mironov, A., Efremova, O., et al. (1996). Photophysical and photochemical properties of potential porphyrin and chlorin photosensitizers for PDT. *J. Photochem. Photobiol. B Biol.* 33, 171–180. doi: 10.1016/1011-1344(95)07241-1
- Zhang, L., Silva, D.-A., Yan, Y., and Huang, X. (2012). Force field development for cofactors in the photosystem, II. *J. Comput. Chem.* 33, 1969–1980. doi: 10.1002/jcc.23016
- Zheng, Y.-C., Zheng, M.-L., Li, K., Chen, S., Zhao, Z.-S., Wang, X.-S., et al. (2015). Novel carbazole-based two-photon photosensitizer for efficient DNA photocleavage in anaerobic condition using near-infrared light. *RSC Adv.* 5, 770–774. doi: 10.1039/C4RA11133H
- Zvereva, E., Segarra-Martí, J., Marazzi, M., Brazard, J., Nenov, A., Weingart, O., et al. (2018). The effect of solvent relaxation in the ultrafast time-resolved spectroscopy of solvated benzophenone. *Photochem. Photobiol. Sci.* 17, 323–331. doi: 10.1039/C7PP00439G

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer, AN, declared a past co-authorship with several of the authors, XA, AM, MM, to the handling Editor.

Copyright © 2018 Marazzi, Gattuso, Monari and Assfeld. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# On the Difference Between Additive and Subtractive QM/MM Calculations

Lili Cao and Ulf Ryde\*

Department of Theoretical Chemistry, Chemical Centre, Lund University, Lund, Sweden

The combined quantum mechanical (QM) and molecular mechanical (MM) approach (QM/MM) is a popular method to study reactions in biochemical macromolecules. Even if the general procedure of using QM for a small, but interesting part of the system and MM for the rest is common to all approaches, the details of the implementations vary extensively, especially the treatment of the interface between the two systems. For example, QM/MM can use either additive or subtractive schemes, of which the former is often said to be preferable, although the two schemes are often mixed up with mechanical and electrostatic embedding. In this article, we clarify the similarities and differences of the two approaches. We show that inherently, the two approaches should be identical and in practice require the same sets of parameters. However, the subtractive scheme provides an opportunity to correct errors introduced by the truncation of the QM system, i.e., the link atoms, but such corrections require additional MM parameters for the QM system. We describe and test three types of link-atom correction, viz. for van der Waals, electrostatic, and bonded interactions. The calculations show that electrostatic and bonded link-atom corrections often give rise to problems in the geometries and energies. The van der Waals link-atom corrections are quite small and give results similar to a pure additive QM/MM scheme. Therefore, both approaches can be recommended.

## OPEN ACCESS

**Edited by:**

Sam P. De Visser,  
University of Manchester,  
United Kingdom

**Reviewed by:**

Albert Poater,  
University of Girona, Spain  
Jiayun Pang,  
University of Greenwich,  
United Kingdom

**\*Correspondence:**

Ulf Ryde  
Ulf.Ryde@teokem.lu.se

**Specialty section:**

This article was submitted to  
Theoretical and Computational  
Chemistry,  
a section of the journal  
*Frontiers in Chemistry*

**Received:** 31 January 2018

**Accepted:** 14 March 2018

**Published:** 03 April 2018

**Citation:**

Cao L and Ryde U (2018) On the  
Difference Between Additive and  
Subtractive QM/MM Calculations.  
*Front. Chem.* 6:89.  
doi: 10.3389/fchem.2018.00089

## INTRODUCTION

Combined quantum mechanics and molecular mechanics (QM/MM) is a popular method to study biological macromolecules, as well as homogeneous catalysis and nanostructures, with computational methods (Balcells and Maseras, 2007; Lin and Truhlar, 2007; Ramos and Fernandes, 2008; Stoyanov et al., 2008; Senn and Thiel, 2009; Keal et al., 2011; Chung et al., 2015; Jover and Maseras, 2016; Ryde, 2016). In this approach, a small region of central interest (typically 20–300 atoms) is treated with quantum mechanical (QM) methods, whereas the remainder of the macromolecule, as well as a considerable amount of explicit solvent are treated by molecular mechanics (MM). This is supposed to combine the accuracy of QM methods with the speed of MM methods. Moreover, the entire macromolecule is included in the calculations [in contrast to the alternative QM-cluster approach (Blomberg et al., 2014), in which most parts of the macromolecule are omitted], reducing the risk of making a biased choice in the selection of the considered system and allowing for a detailed study of how the surroundings affect the properties of interest.

A problem with QM/MM approaches is that there exists so many variants and that the details of these are seldom discussed. For example, QM/MM approaches can use either subtractive or additive schemes (Senn and Thiel, 2009). In a subtractive scheme, three separate calculations are performed:

One QM calculation with the QM region (system 1;  $E_1^{\text{QM}}$ ) and two MM calculations, one for the entire system (systems 1 and 2;  $E_{12}^{\text{MM}}$ ) and one for the QM region ( $E_1^{\text{MM}}$ ) (Maseras and Morokuma, 1995; Ryde, 1996b; Svensson et al., 1996):

$$E_{\text{QM/MM}}^{\text{sub}} = E_1^{\text{QM}} + E_{12}^{\text{MM}} - E_1^{\text{MM}} \quad (1)$$

The advantage with this approach is the simplicity: It automatically ensures that no interactions are double-counted and it can be set up for any QM and MM software (provided that they can write out energies and forces), without the need of any modification of the code. Thereby, the QM/MM software is updated every time the underlying QM or MM software is updated. Moreover, it can be easily extended to more than two computational methods and regions (Svensson et al., 1996). The typical example of a subtractive scheme is ONIOM (Svensson et al., 1996), but other software use similar methods, e.g., ComQum (Ryde, 1996b).

In the additive scheme, only two calculations are performed: the same QM calculation for the QM region, but only a single MM calculation ( $E_{2-1}^{\text{MM}}$ ) (Sherwood et al., 2003; Senn and Thiel, 2009):

$$E_{\text{QM/MM}}^{\text{add}} = E_1^{\text{QM}} + E_{2-1}^{\text{MM}} \quad (2)$$

although the latter is often formally divided into two terms, a MM energy of system 2 and a QM/MM interface energy ( $E_{2-1}^{\text{MM}} = E_2^{\text{MM}} + E_{12}^{\text{QM/MM}}$ ). In this case, it is up to the developer to ensure that no interactions are omitted or double-counted. Therefore, an additive scheme requires a special MM software, in which the user or developer can select which MM terms to include. The advantage of the additive QM/MM scheme is that no MM parameters for the QM atoms are needed, because those energy terms are calculated by QM.

Further differences may arise if the QM region is covalently connected to the MM region. Then, the QM system needs to be properly truncated. This can be done by special localized orbitals (Levitt, 1976; Théry et al., 1994; Gao et al., 1998; Murphy et al., 2000; Senn and Thiel, 2009), but it is more common that the QM system is simply truncated by hydrogen atoms, the hydrogen link-atom approach (Singh and Kollman, 1986; Field et al., 1990; Reuter et al., 2000; Senn and Thiel, 2009; Ryde, 2016). In the subtractive scheme, MM parameters for the link atoms are needed (Senn and Thiel, 2009).

The interaction between the QM and MM regions is typically dominated by electrostatics. This interaction can also be treated at different levels of approximation (Senn and Thiel, 2009; Ryde, 2016). In mechanical embedding, it is calculated at the MM level (Maseras and Morokuma, 1995; Svensson et al., 1996). In electrostatic embedding, the electrostatic QM-MM interaction is instead treated at the QM level by including a point-charge model (i.e., atomic partial MM charges) of system 2 in the QM calculations (Singh and Kollman, 1986; Field et al., 1990; Ryde, 1996b; Dapprich et al., 1999). Thereby, system 1 is polarized by system 2, but not vice versa. In

polarized embedding, both systems are mutually and self-consistently polarized in the QM calculations (Poulsen et al., 2001; Söderhjelm et al., 2009; Olsen et al., 2010). This requires a polarizable MM force field for system 2 (Lopes et al., 2009) and a QM software that can treat polarizabilities, which are still rather unusual. Therefore, such calculations are less common and typically restricted to single-point calculations of accurate properties. Mechanical embedding is normally considered to be less accurate than electrostatic embedding (Senn and Thiel, 2009), and the latter has therefore been the most widely used approximation, although it involves polarization of only parts of the system and is more sensitive to the treatment of the link atoms (Hu et al., 2011b). Strictly, Equations (1, 2) apply only to mechanical embedding, but they can easily be adapted to electrostatic embedding by including a point-charge model of system 2 in the QM term ( $E_{\text{QM1+ptch2}}$ ) and setting the charges of system 1 to zero in the MM calculations (Ryde, 1996b).

Unfortunately, the distinction between the subtractive and additive schemes in literature is often unclear and confused. In many cases, the subtractive scheme is equated with mechanical embedding and the additive scheme with electrostatic embedding (Senn and Thiel, 2009; Götz et al., 2014). In other cases, the subtractive scheme is equated with the ONIOM method (Roßbach and Ochsenfeld, 2017). We prefer the definition in Equations (1, 2), emphasizing that the subtractive scheme employs two MM calculations with an external MM program, whereas the additive scheme employs a single MM calculation with an internal MM program, allowing the developer to cherry-pick the MM terms actually needed. In particular, both additive and subtractive schemes may use either mechanical or electrostatic (or even polarized) embedding.

It is normally assumed that the subtractive scheme is harder to set up and requires accurate MM parameters for the QM region and link atoms. For example, Roßbach and Ochsenfeld state in a recent article comparing subtractive and additive QM/MM (Roßbach and Ochsenfeld, 2017): “The (additive) QM/MM approach has the advantage that parameters for QM and link atoms, saturating covalent bonds between QM and MM, are unnecessary, as these are never described by the force field. The subtractive ONIOM approach requires accurate parameters for all atoms, including link atoms, because an MM calculation of the QM region is also necessary to avoid double counting.” On the other hand, Sousa et al. present the opposite view that the subtractive scheme is more advantageous, because of the “lack of a requirement for a parameterized expression describing the interaction of the various regions, and the fact that all systematic errors in the treatment of the inner regions by the lower levels of theory are canceled out” (Sousa et al., 2017).

In this article, we aim at clarifying the difference between the two schemes and compare their performance. We will show that with a proper setup, additive, and subtractive schemes should give identical results with a similar effort and that they require the same set of MM parameters. However, the subtractive scheme may be tuned to correct errors introduced by the link atoms and then additional parameters are needed.

## METHODS

### The ComQum QM/MM Software and Its Subtractive Scheme

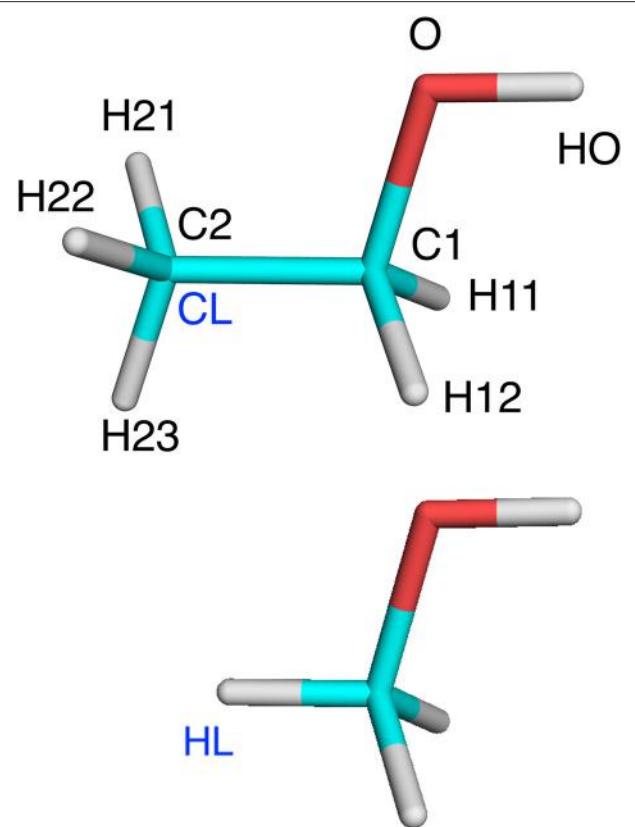
A problem when comparing QM/MM methods is that there exist so many variants and that the details of the calculations are seldom discussed (Ryde, 1996b; Senn and Thiel, 2009). Therefore, we here give a thorough discussion of our QM/MM software and details of all QM/MM variants implemented. All QM/MM calculations in this article were performed with the ComQum software (Ryde, 1996b; Ryde and Olsson, 2001). ComQum is a modular program, combining the QM software Turbomole (Ahrlrichs et al., 1989; Furche et al., 2014; TURBOMOLE version 7.1, 2016) and the MM software AMBER (Case et al., 2014). It consists of five small Fortran programs that read, write, manipulate, and transfer coordinates, energies, forces, and charges between the MM and QM software. It employs a subtractive scheme and it was developed in 1992–1995, concurrently and independently from the ONIOM software (Ryde, 1995, 1996a,b). It has always used electrostatic embedding, in contrast to ONIOM. Junctions are treated by the hydrogen link-atom approach (Ryde, 1996b; Reuter et al., 2000; Senn and Thiel, 2009).

To make the discussion clear, we use the following conventions (Senn and Thiel, 2009): The QM region is called system 1, whereas atoms in the MM region are called system 2. The QM region is terminated by hydrogen link atoms, called HL. They replace the corresponding carbon link (CL) atoms in the real system. HL and CL are different representations of the same atom and never appear in the same (MM or QM) calculation. This is illustrated in **Figure 1**. A superscript HL or CL show which representation is used. We will use XL to denote either HL or CL. The HL atom is covalently bound to a single  $Q_1$  atom in the QM region, whereas the CL atom is connected to (typically several)  $M_2$  atoms in the MM region (we use this somewhat illogical notation, because several other QM/MM descriptions use  $M_1$  to denote CL; we prefer our notation because in ComQum, HL and CL are two different representations of the same atom, which belongs to the QM region). The  $Q_1$  atoms are covalently bound to  $Q_2$  atoms, which are covalently bound to  $Q_3$  atoms, and so on. Likewise, the  $M_2$  atoms are covalently bound to  $M_3$  atoms, and so on.

The HL atom is placed along the  $Q_1$ –CL bond, with the  $Q_1$ –HL bond length ( $r_{Q_1-HL}$ ) calculated from (Ryde, 1996b):

$$r_{Q_1-HL} = r_{Q_1-CL} \frac{r_{Q_1-HL}^0}{r_{Q_1-CL}^0} \quad (3)$$

where  $r_{Q_1-CL}$  is the current  $Q_1$ –CL bond length,  $r_{Q_1-CL}^0$  is the optimum  $Q_1$ –CL bond length in the MM force field and  $r_{Q_1-HL}^0$  is the  $Q_1$ –HL bond length in a model of the isolated truncated residue, optimized with the current QM method and basis set.  $r_{Q_1-CL}^0$  can be found in the MM force field libraries and  $r_{Q_1-HL}^0$  is easily obtained by a simple QM geometry optimisation (which typically takes less than a minute). Equation (3) can also be used in reverse during the QM/MM geometry optimization to calculate the CL coordinates from the HL coordinates and it is



**FIGURE 1** | Ethanol and methanol with atom names indicated, as well as the CL and HL atoms.

also used together with the chain rule to obtain the MM forces on the HL atom (Ryde and Olsson, 2001). Thus, the HL atoms do not introduce any additional degrees of freedom.

The total QM/MM energy in standard ComQum (which uses a subtractive scheme with electrostatic embedding) is calculated from Equation (4) (Ryde, 1996b; Ryde and Olsson, 2001):

$$E_{QM/MM}^{sub,EE} = E_{QM1+ptch2}^{HL} + E_{MM12,q_1=0}^{CL} - E_{MM1,q_1=0}^{HL} \quad (4)$$

where  $E_{QM1+ptch2}^{HL}$  is the QM energy of the QM system truncated by HL atoms and embedded in the set of point charges modeling system 2 (but excluding the self-energy of the point charges). All atoms in system 2 are included in the point-charge model (but not the CL atoms, which do not belong to system 2 in our view). In our original development, charges of some additional atoms were excluded (Ryde, 1996b), but a comparison of several different charge distribution schemes did not show any advantage of more complicated schemes (Hu et al., 2011b).  $E_{MM1,q_1=0}^{HL}$  is the MM energy of the QM system 1, still truncated by HL atoms, but without any electrostatic interactions. Finally,  $E_{MM12,q_1=0}^{CL}$  is the MM energy of all atoms in the system with CL atoms and with the charges of the QM system set to zero (to avoid double counting of the electrostatic interactions).

In the original implementation of ComQum (Ryde, 1996b), it was necessary to set up two AMBER parameter and topology (prmtop) files for the MM calculations, one for the full system ( $E_{MM12,q_1=0}^{CL}$ ) and one for the truncated QM region ( $E_{MM1,q_1=0}^{HL}$ ). This involved development of MM parameters for all junctions, which is tedious and error prone, even if the same truncated residues can be used for several different proteins. Initially, some attempts were made to allow the calculations on the small system compensate for the truncations and the introduction of link atoms (i.e., for the conversion of CL atoms to HL atoms). However, we did not see any consistent improvement in the results; instead such a treatment often introduced instability in the calculations.

In particular, we soon realized that the parameters of the HL–Q<sub>1</sub> bond cannot be freely selected. The deterministic relation between HL and CL in ComQum (Equation 3) implies that we should use  $r_{Q_1-HL}^0$  as the equilibrium bond length and the force constant must be

$$k_{Q_1-HL} = k_{Q_1-CL} \left( \frac{r_{Q_1-HL}^0}{r_{Q_1-CL}^0} \right)^2 \quad (5)$$

where  $k_{Q_1-CL}$  is the force constant of the Q<sub>1</sub>–CL bond in the MM force field. Otherwise, a spurious force will be introduced. Moreover, for the Q<sub>2</sub>–Q<sub>1</sub>–HL angle and the Q<sub>3</sub>–Q<sub>2</sub>–Q<sub>1</sub>–HL dihedral parameters we simply used the corresponding MM Q<sub>2</sub>–Q<sub>1</sub>–CL angle and the Q<sub>3</sub>–Q<sub>2</sub>–Q<sub>1</sub>–CL parameters, obtained by copying these entries from the MM force field.

Even if this was a completely mechanical procedure, it was still somewhat tedious and error prone, because it had to be redone every time the MM force field, QM method, or basis set was changed. Therefore, we implemented in 2006 a program (changeparm) that performed this task automatically: Reading the AMBER prmtop file of the full system and a file with the ideal QM bond length ( $r_{XL-HL}^0$ ), it automatically generates the prmtop and coordinate files for the MM calculation of the QM system ( $E_{MM1,q_1=0}^{HL}$ ) according to these rules. Thereby, only a single MM calculation needs to be set up (that of the full system, which typically is already done, because QM/MM studies of macromolecules normally start with an equilibration of the structure with molecular dynamics) and no special parameters needs to be developed for the truncated system or the link atoms. This removes one of the disadvantages with the subtractive scheme (but it also discards a potential advantage with the method, as will be discussed below). However, the changeparm program required procedures to read and write the prmtop file, as well as a complete understanding of the meaning of all entries in it, a significant programming effort. Still, it has been so valuable that we recently have implemented the corresponding program (Cao et al., 2018) also for the crystallography and NMR system (CNS; Brünger et al., 1998; Brünger, 2007) software for quantum refinement (Ryde et al., 2002; Ryde and Nilsson, 2003).

Thus, with our implementation, also the subtractive QM/MM scheme requires only a single prmtop for the entire system. However, this file must contain parameters for all atoms, including those in the QM region. This is so because (the *leap*

module in) AMBER refuses print the file if any parameter is missing. It may be that other MM software is less restrictive. However, this is a rather minor restriction, because the parameters do not need to be accurate. On the contrary, for all interactions involving only QM atoms (except van der Waals interactions involving the link atoms), the MM energies cancel exactly (owing to the  $E_{MM12,q_1=0}^{CL} - E_{MM1,q_1=0}^{HL}$  terms in Equation 4). Therefore, dummy (e.g., zeroed) parameters may be used. Moreover, with the general MM force fields available in most MM software (Vanommeslaeghe et al., 2010), including AMBER (Wang et al., 2004), most parameters already exist. The only problem may be metal sites, but if no explicit bonds are defined between the metals and any ligand atoms, only van der Waals parameters for the metal are needed and these are necessary also in the additive scheme. Thus, the need of parameters for the QM region is a very minor restriction of the subtractive scheme and the setup of dummy parameters can easily be automatized (although we have never felt such a need).

## Additive ComQum

Recently, we implemented a simple software (calcforce), which calculates MM energies and forces, based on AMBER coordinate and parameter–topology files. This was done to allow calculations without any cut-off for non-bonded interactions, even for very large systems (AMBER employs a non-bonded pair list that can become too large for the memory with more than  $\sim 10^5$  atoms) and to gain control over exactly what forces are written out by the program (for example, turning on the dumfrc option in the AMBER software changes the electrostatic energy). As a by-product, it also gives us full control over the energy function and allowed us to implement an additive QM/MM scheme.

In our additive approach, we use the energy function:

$$E_{QM/MM}^{\text{add,EE}} = E_{QM1+ptch2}^{HL} + E_{MM\ 2-1}^{CL} \quad (6)$$

where the  $E_{QM1+ptch2}^{HL}$  QM term is identical to that used in the subtractive scheme. As shown by the superscript, all terms in  $E_{MM\ 2-1}^{CL}$  employ CL atoms, coordinates and parameters, never any HL atoms. We employed the following rules to determine what MM terms to include in  $E_{MM\ 2-1}^{CL}$  (note again that in our notation, the CL atoms belong to the QM region):

- All interactions involving only QM atoms are excluded (already included in the QM term).
- All interactions involving only MM atoms are included.
- Bonded interactions involving at least one MM atom are included.
- Electrostatic interactions involving one MM and one QM atom are excluded (already included in the QM calculation).
- Van der Waals interactions involving one MM and one QM atom are included.

These rules are based on the simple philosophy that we should calculate by MM all terms that are not already considered in the QM calculations. This seems very natural and should represent a typical implementation of additive QM/MM (Sherwood et al., 2003), although the rules are seldom discussed explicitly. This selection is illustrated in **Table 1** for the simple ethanol model

**TABLE 1** | Illustration of which terms are included in the various energies for the ethanol molecules in **Figure 1** (atom names are shown in that figure, except that H1 indicates either H11 or H12 and H2 indicates H21, H22, or H23).

	$E_{\text{QM1+ptch2}}^{\text{HL}}$	$E_{\text{MM1},q_1=0}^{\text{HL}}$	$E_{\text{MM1,ptch2}}^{\text{HL}}$	$E_{\text{MM12}}^{\text{CL}}$	$E_{\text{MM 2-1}}^{\text{CL}}$
<b>BONDS</b>					
HO-O	QM	MM	MM	MM	
O-C1	QM	MM	MM	MM	
C1-H1	QM	MM	MM	MM	
C1-C2	QM,HL	MM,HL	MM,HL	MM,CL	
C2-H2				MM,CL	MM,CL
<b>ANGLES</b>					
HO-O-C1	QM	MM	MM	MM	
O-C1-H1	QM	MM	MM	MM	
O-C1-C2	QM,HL	MM,HL	MM,HL	MM,CL	
H1-C1-C2	QM,HL	MM,HL	MM,HL	MM,CL	
C1-C2-H2				MM,CL	MM,CL
<b>DIHEDRALS</b>					
HO-O-C1-H1	QM	MM	MM	MM	
HO-O-C1-C2	QM,HL	MM,HL	MM,HL	MM,CL	
O-C1-C2-H2				MM,CL	MM,CL
H1-C1-C2-H2				MM,CL	MM,CL
<b>NON-BONDED</b>					
HO-O	QM				
HO-C1	QM				
HO-H1	QM	Sc	Sc	Sc	
HO-C2	QM,HL	Sc,HL	Sc,CL	Sc,CL	
HO-H2	Ptch	Ptch	MM	MM	
O-C1	QM				
O-H1	QM				
O-C2	QM,HL				
O-H2	Ptch	Ptch	Sc	Sc	
C1-H1	QM				
C1-C2	QM,HL				
C1-H2	Ptch	Ptch			
H1-C2	QM,HL				
H1-H2	Ptch	Ptch	Sc	Sc	
C2-H2	Ptch,HL	Ptch,HL			

QM means that the term is included by the QM calculation, MM that it is included as a MM term, HL or CL that it is treated as a HL or CL atom, Sc that the non-bonded interaction is scaled down, and Ptch that it is treated by point charges.

In **Figure 1** (with CB as the link atom). Note that the first four rules are identical to the rules we use for setting up the truncated prmtop file in the subtractive scheme, so that the two schemes should give identical bonded and electrostatic energies. In particular, the two schemes give identical energies for the XL-Q<sub>1</sub> bond term (assuming the harmonic term  $E_{\text{Q}_1-\text{XL}}^{\text{bond}} = k_{\text{Q}_1-\text{XL}}(r_{\text{Q}_1-\text{XL}} - r_{\text{Q}_1-\text{XL}}^0)^2$  employed in AMBER and most other macromolecular force fields) even if the subtractive scheme uses HL coordinates and the additive scheme CL coordinates, because of the relations between the coordinates in Equation (3) and the force constants in Equation (5) (the force constants in Equation 5 were constructed with this aim).

The only difference between the two schemes is the van der Waals interactions involving the link atoms and another atom

in the QM system (possibly other link atoms): In the additive scheme, no such interactions are calculated by MM (because both atoms belong to the QM system). However, in the subtractive scheme, all van der Waals interactions involving link atoms are calculated twice: In  $E_{\text{MM12},q_1=0}^{\text{CL}}$ , they are obtained with MM parameters and coordinates for CL atoms, whereas in  $E_{\text{MM1},q_1=0}^{\text{HL}}$ , they are obtained with MM parameters and coordinates for HL atoms. In variance to all the other QM atoms, these two terms are not identical and therefore will not cancel. Instead, they provide a MM correction to the link atom, i.e., to the fact that the HL atoms in the QM region are H atoms and not the correct C atoms (i.e., they have smaller van der Waals radii) and that they are at incorrect positions. Thus, these van der Waals terms in  $E_{\text{MM1},q_1=0}^{\text{HL}}$  can be seen as a correction to the corresponding energy in the QM calculation ( $E_{\text{QM1+ptch2}}^{\text{HL}}$ ), which also involves the incorrect HL coordinates and atoms. We will call it the *van der Waals link-atom correction* (VLAC). MM van der Waals parameters are normally quite accurate, so this approach is used in most subtractive schemes, but it cannot be included in a strict additive scheme, which may be a disadvantage. It should be noted that these interactions are only within the QM region, so with a small QM region, it involves only a few interactions and the correction is small.

In fact, we can exactly reproduce the additive QM/MM calculations within a subtractive scheme by replacing the  $E_{\text{MM1},q_1=0}^{\text{HL}}$  term in Equation (4) with a  $E_{\text{MM1},q_1=0}^{\text{CL}}$  term, in which parameters and coordinates corresponding to the CL atoms, rather than the HL atoms, are used. This was done manually to confirm that the implementations are correct, but it has never been implemented for production calculations (because the additive scheme gives the same results).

## Mechanical Embedding

For comparison, we have also implemented mechanical embedding (ME) in ComQm [we have calculated single-point ME energies before (Hu et al., 2011a,b), but not done full geometry optimizations]. ME calculations can be run by simply deleting the point-charge model from the QM calculations (removing the \$point\_charges keyword from the Turbomole control file) and (re-)inserting charges of the QM system in the prmtop files for the two MM calculations in Equation (4). This gives the energy function:

$$E_{\text{QM/MM}}^{\text{sub,ME}} = E_{\text{QM1}}^{\text{HL}} + E_{\text{MM12}}^{\text{CL}} - E_{\text{MM1}}^{\text{HL}} \quad (7)$$

Two issues need to be settled in this implementation. The first is how to treat the link atoms. If the charge of the HL atom is identical to that of the CL atom, the electrostatics within the QM system cancel exactly in the  $E_{\text{MM12}}^{\text{CL}} - E_{\text{MM1}}^{\text{HL}}$  terms in Equation (7). However, then the total energy reflects electrostatics involving HL atoms, from the  $E_{\text{QM1}}^{\text{HL}}$  term. In analogy with the van der Waals energy correction described in the previous section, we prefer to have different charges for the HL and CL atoms, with those of the HL atoms being representative for a H atom and those of the CL atoms being representative for the true (typically C) atoms.

This leads us to the second issue, viz. how the charges are calculated for the QM system, including the HL and CL atoms. Since QM calculations are done for the QM system, it is natural to use some sort of QM-derived charges. Originally, ComQum employed Mulliken charges (these charges are used also with electrostatic embedding when parts of system 2 is optimized by MM; Ryde, 1996b). However, it is well-known that charges fitted to the electrostatic potential (ESP) give (by construction) more accurate electrostatic interaction energies (Sigfridsson and Ryde, 1998). Therefore, we have used such ESP charges [obtained with the Merz–Kollman scheme (Besler et al., 1990)] ever since ESP charges were implemented in Turbomole (note that in Turbomole, the point-charge model needs to be removed before the charges are calculated, without reoptimizing the wavefunction). These charges can be directly used in the  $E_{MM1}^{HL}$  term and therefore also for the HL atoms, since they were obtained for a QM system with HL atoms.

However, for the  $E_{MM12}^{CL}$  term, these charges need to be adapted so that the charge of the total system remains integer, meaning that the HL charges need to be adapted to apply for CL atoms instead. It is not evident how this should be done and it is seldom discussed, although this needs to be done for essentially all QM/MM and MD simulations employing QM charges calculated for QM systems with link atoms.

We have selected to follow this procedure:

1. Start from the MM charges of the entire system with formal (integer) charges on metals and other inorganic groups.
2. Use the QM charges for all QM atoms, except the HL atoms.
3. Use the original MM charges for all MM atoms.
4. For each amino acid (or other biochemical building unit) with link atoms: Add a constant offset to the original MM charges of each CL atom in the unit so that the total charge of this unit is the same as the sum of the QM charge of these atoms, possibly with the addition of an integer charge from atoms outside the QM system.

The procedure is illustrated in **Table 2**. It is fully automatic, except that a possible integer charge outside the QM system needs to be specified. This way, charge transfer within the QM system is allowed, meaning that none of the QM residues has an integer charge. The modification of the charges on the CL atoms is also kept to a minimum. However, alternative approaches are conceivable, e.g., dividing the remaining charge equally over all link atom, after the charges of the non-HL QM atoms have been set. Thus, the CL charges are somewhat ambiguous. Importantly, the procedure keeps all charges of the QM system equal between  $E_{MM12}^{CL}$  and  $E_{MM1}^{HL}$ , except for the HL/CL atoms, allowing for a proper cancelation of those electrostatic terms in the ME approach, in analogy with the VLAC correction for the subtractive scheme.

## Electrostatic Link-Atom Corrections

In the electrostatic-embedding variant of ONIOM (Vreven et al., 2006), as well as in our QTCP approach (QM/MM thermodynamic cycle perturbation; Rod and Ryde, 2005), further attempts are made to correct errors introduced by the link atoms. In the QM calculations, the HL atoms are of the wrong element,

**TABLE 2** | Illustration of the method to determine charges for methanol and ethanol (shown in **Figure 1** with atom names; H1 is H11 and H12; H2 is H21, H22, and H23).

Atom	Methanol		Ethanol	
	Set1	Set2	Set1	Set2
C2	-0.0108		-0.2324	-0.1935
H2		<b>0.0609</b>	<b>0.0609</b>	
C1	<b>0.2120</b>		0.3841	<b>0.2120</b>
H1	<b>-0.0043</b>		-0.0534	<b>-0.0043</b>
O	<b>-0.5903</b>		-0.5847	<b>-0.5903</b>
HO	<b>0.3977</b>		0.3570	<b>0.3977</b>

The methanol charges were obtained from a QM RESP calculation. The ethanol Set1 charges were obtained in the same way. For Set2, charges of the C1, H1, O, and H atoms were taken from methanol (marked in bold face) and the charges of H2 were taken from Set1 (marked in bold face and italics). Finally, the charge on C2 was determined to give a vanishing net charge.

located at the incorrect position (compared to the CL atom) and they may make Coulombic interactions with point charges of nearby atoms that are not included or are scaled down in normal MM calculations (viz. interactions with the M<sub>2</sub>, M<sub>3</sub>, and M<sub>4</sub> atoms). These errors can be compensated by calculating exactly the same interactions in the  $E_{MM1}^{HL}$  term and replace them with the corresponding interactions in the  $E_{MM12}^{CL}$  term.

In practice, this is accomplished by using QM charges for the QM system in both MM calculations and including the same point-charge model of the surroundings in the  $E_{MM1}^{HL}$  term. We call this approach electrostatic link-atom correction (ELAC) and it gives the following energy function:

$$E_{QM/MM}^{sub,ELAC} = E_{QM1+ptch2}^{HL} + E_{MM12}^{CL} - E_{MM1+ptch2}^{HL} \quad (8)$$

For these calculations, we used the same QM charges for the QM system and the same MM charges for MM system as described for the mechanical-embedding calculations.

## Bonded Link-Atom Corrections

Finally, the subtractive scheme allows for a third type of link-atom corrections, viz. for the bonded terms. It is likely that a HL atom will give rise to slightly different bonded terms than the corresponding CL atom, e.g., smaller XL–Q<sub>1</sub>–Q<sub>2</sub> ideal angles. Again, we may try to use the  $E_{MM1}^{HL}$  calculation to correct for these errors (i.e., so that the HL bonded terms would cancel between the  $E_{QM1+ptch2}^{HL}$  and  $E_{MM1}^{HL}$  terms and the corresponding CL result in  $E_{MM12}^{CL}$  would remain, instead of the exact cancelation of these terms between  $E_{MM1}^{HL}$  and  $E_{MM12}^{CL}$  as in both the standard additive and subtractive approaches). This is done by using different parameters for the Q<sub>2</sub>–Q<sub>1</sub>–HL and Q<sub>2</sub>–Q<sub>1</sub>–CL angles and the Q<sub>3</sub>–Q<sub>2</sub>–Q<sub>1</sub>–HL and Q<sub>3</sub>–Q<sub>2</sub>–Q<sub>1</sub>–CL dihedral parameters (and possibly also for the Q<sub>1</sub>–HL and Q<sub>1</sub>–CL bond parameters). We will call this bonded link-atom corrections (BLAC).

Of course, the parameters need to be accurate for there to be a hope of any improved results. In this paper, we tried two approaches. In the first (BLAC1), we used standard AMBER parameters for both the CL and HL terms, the latter typically

coming from the GAFF (Wang et al., 2004) force field. These parameters involved also the bonded Q<sub>1</sub>-XL terms.

In the second approach (BLAC2), we instead performed a parametrisation of both the full and truncated systems, based on a QM frequency calculation on each system. The bonded parameters were then extracted with the Seminario approach (Seminario, 1996), using the Hess2FF program (Nilsson et al., 2003; Hu and Ryde, 2011). We used the same atom types as in the AMBER files in BLAC1. This means that in principle all bonded parameters will differ in the two MM calculations, not only those involving the XL atom. Finally, BLAC1J and BLAC2J was obtained from BLAC1 and BLAC2 by changing the Q<sub>1</sub>-HL force constants according to Equation (5), keeping everything else the same.

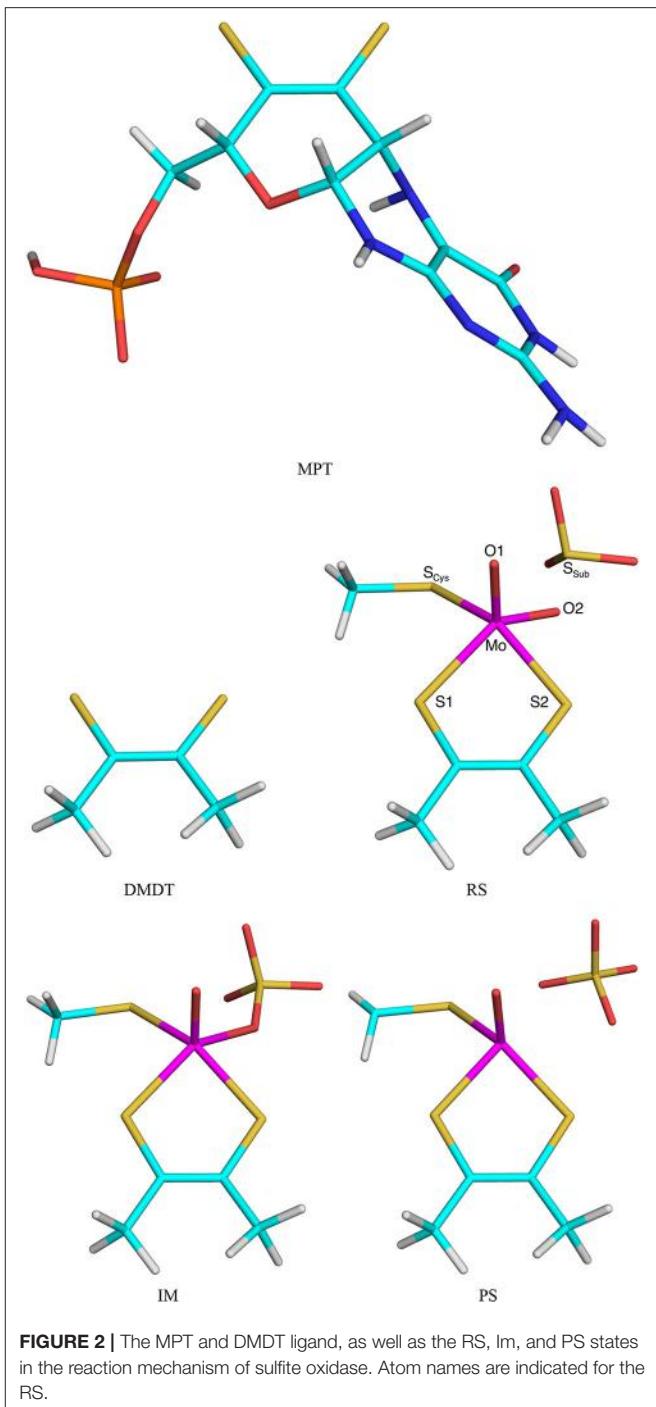
## Test Systems

All QM calculations were carried out using the Turbomole software (versions 7.1 and 7.2; Ahlrichs et al., 1989; Furche et al., 2014). They were performed using the TPSS (Tao et al., 2003) functional in combination with def2-SV(P) (Schäfer et al., 1992) basis set, including empirical dispersion corrections with the DFT-D3 approach (Grimme et al., 2010) with Becke-Johnson damping (Grimme et al., 2011), as implemented in Turbomole. The MM calculations were performed with the AMBER ff14SB (Maier et al., 2015) force field for protein residues, GAFF (Wang et al., 2004) for non-protein molecules and TIP3P (Jorgensen et al., 1983) for water. In all QM/MM calculations, the MM system was kept fixed to simplify the interpretation of the results.

The various approaches were tested on three systems. The first was an isolated ethanol molecule. The reference system was ethanol, optimized with QM. In the QM/MM calculations, the QM system was methanol with C2 converted to a HL atom (**Figure 1**). The terminal methyl group was in the MM system.

The second test system was sulfite oxidase. The calculations were taken from our recent study of this enzyme (Caldararu et al., 2018). The active site contains a Mo ion coordinated to a molybdopterin (MPT) molecule, as well as a Cys residue and two oxo groups. All these groups were included in the QM system (Cys modeled as CH<sub>3</sub>S<sup>−</sup>), as well as the sulfite substrate [note that this QM system is smaller than in our previous study (Caldararu et al., 2018), in which nine additional residues and five water molecules were also included]. In this paper we compared the effects of either including the full MPT residue (in the reduced and protonated form, called MPH in our previous paper) or truncating it to a dimethyldithiolene molecule [DMDT, (CH<sub>3</sub>CS)<sub>2</sub><sup>2−</sup>; both shown in **Figure 2**]. Thus, the QM/MM calculations with the full MPT molecule were the reference structures and QM/MM calculations with DMDT in the QM system were run to compare the performance of the various QM/MM variants.

In DMDT two of the CL atoms are covalently bonded. This gives a complication when automatically setting up the prmtop file for the truncated system: This bond needs to be explicitly removed from the file (together with the corresponding angles and dihedrals); otherwise, spurious MM forces will cause the calculations to crash with distorted structures. Normally, we



**FIGURE 2** | The MPT and DMDT ligand, as well as the RS, IM, and PS states in the reaction mechanism of sulfite oxidase. Atom names are indicated for the RS.

discourage from having covalently connected CL atoms, but in this study, it provides a hard test for the various methods to provide junction corrections.

Five states were studied in the reaction: The Mo<sup>VI</sup> =O+SO<sub>3</sub><sup>2−</sup> reactant state (RS) with sulfite in the second coordination sphere of Mo, the Mo<sup>IV</sup>-SO<sub>4</sub><sup>2−</sup> intermediate (IM) with sulfate coordinated to Mo, the Mo<sup>IV</sup>+SO<sub>4</sub><sup>2−</sup> product state (PS) with sulfate in the second sphere of Mo (shown in **Figure 2**), as well as the two transition states (TS1 and TS2) connecting these three

states. The transition states were obtained from potential-energy scans along the S–O1 (TS1, 2.0 Å) and Mo–O1 (TS2, 3.7 Å) reaction coordinates. PS was also obtained with a restraint in the Mo–O2 distance of 3.85 Å, taken from calculations with an appreciably larger QM system (Caldararu et al., 2018). Besides the QM system, the setup was identical to that in our previous study (Caldararu et al., 2018).

The third test system was the conversion of oxophlorin to verdohaem by haem oxygenase. Again, the calculations were taken from a recent study of this enzyme (Alavi et al., 2017). The QM system consisted of the oxophlorin group (an oxidized haem molecule), as well as O<sub>2</sub> and His (truncated to imidazole) as axial ligands of the Fe ion [again, this QM system is smaller than in our previous study (Alavi et al., 2017), in which two additional residues and six water molecules were also included]. We compared the effects of including either the full oxophlorin ring (OXF) with its eight peripheral propionate, vinyl and methyl substituents or truncating all substituents to HL atoms (OXT; both shown in **Figure 3**). Thus, the calculations with the full OXF were the reference structures and QM/MM calculations with OXT were run to compare the performance of the various QM/MM variants.

Nine states were studied in the reaction as is shown in **Figure 3**: The Fe–O<sub>2</sub> reactant state (**1**), the first intermediate, in which O<sub>2</sub> is bridging between Fe and one of the OXF carbon atoms (C4B; **2**), the second intermediate (**3**) in which the O1–O2 bond is cleaved and the C4B–O2 bond is formed, the third intermediate (**4**) in which the O<sub>2</sub> atom has formed a bond with the C1C OXF atom, giving a four-membered C4B–O2–C1B–CMC ring, the verdohaem product (**5**), in which CO has dissociated, but the C4B–O2–C1B bonds are kept, as well as the four connecting transition states (**T1–T4**). The transition states were obtained from potential-energy scans along the O2–C4B (**T1**, 1.8 Å), O1–O2 (**T2**, 1.6 Å), O2–C1B (**T3**, 1.9 Å), and the C4B–C (**T4** 1.8 Å). Besides the QM system, the setup was identical to that in our previous study (Alavi et al., 2017).

## RESULT AND DISCUSSION

In this paper, we clarify the difference between additive and subtractive variants of the QM/MM approach. In principle, the two approaches can be tuned to give exactly the same results, as the additive approach can freely pick almost any energy in the  $E_{2-1}^{\text{MM}}$  term in Equation (2). However, in a typical implementation of the two approaches, the primary difference between the two approaches is that the additive scheme employs only a single MM term for each interaction, whereas in the subtractive scheme, there are two MM terms for the QM system, one in  $E_{12}^{\text{MM}}$  and one in  $E_1^{\text{MM}}$ . Depending on the implementation, these duplicate terms can either be selected to be identical (and therefore canceling, which would give exactly the same results as in the additive scheme) or they can be different, in particular with the aim of correcting the errors introduced by the HL link atoms in the QM system. We have investigated three different levels of link-atom corrections, involving van der Waals terms (VLAC), electrostatic terms (ELAC), and bonded terms (BLAC). In the following, we

test the performance of the various correction schemes for three different systems: ethanol, sulfite oxidase, and haem oxidase. The results are described in three separate sections. For each system, we study six different approaches: the additive scheme (Add, i.e., without any link-atom corrections), the subtractive scheme with van der Waals (VLAC), electrostatic (ELAC), and bonded link-atom corrections (BLAC), the latter in two variants (BLAC1 or BLAC2 and BLAC2J) and mechanical embedding (ME, using a subtractive scheme and VLAC). ELAC and the three variants of BLAC always also include VLAC. Therefore, we use the abbreviation Sub for the subtractive scheme involving only VLAC (emphasizing that this is the standard approach for the subtractive scheme in ComQum). For ethanol, we tested a few additional combinations.

### Ethanol

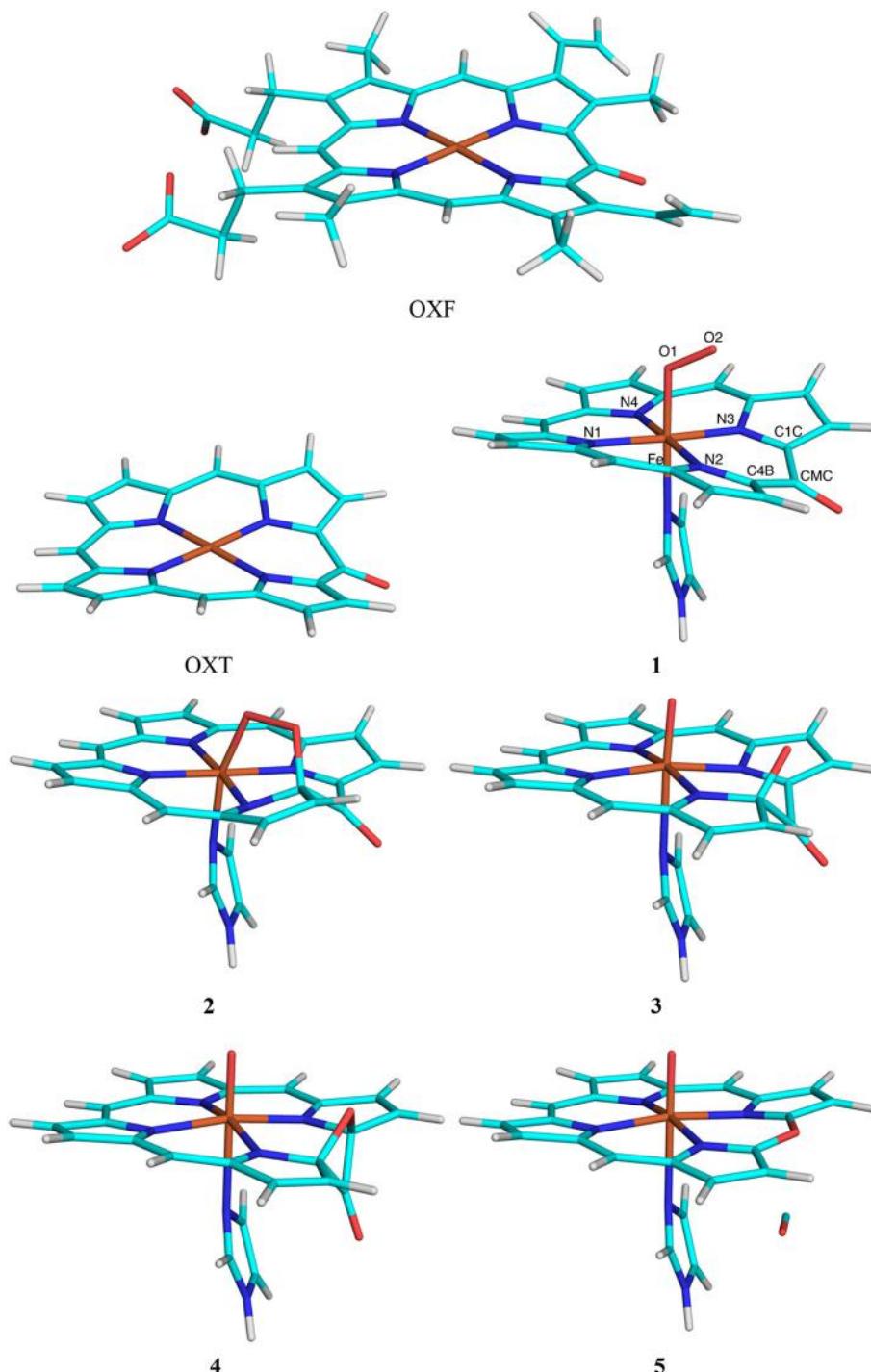
We first tested all methods on a very simple model system, viz. ethanol, in which methanol was used as the QM system in QM/MM calculations and the results were compared to a QM calculation on the full ethanol molecule. Nine different calculations were run for this system: Add, Sub, two variants of ELAC, BLAC1, BLAC1J, BLAC2, BLAC2J, and ME, as well as ELAC combined with BLAC2J. The two variants of ELAC used the same ESP charges for the QM system. However, for the MM system, they used either the ESP charges of ethanol (ELAC2) or the ESP charges for methanol for all methanol atoms except HL, the ethanol ESP charges for the H21–H23 atoms on the terminal methyl group, whereas the charge on the CL atom was adapted to give a vanishing net charge (ELAC1). The latter approach is similar to what is used for the two enzyme systems, for which ESP charges for the full MM system are not available. The two sets of charges are shown in **Table 2**.

The results of the QM/MM calculations on ethanol are collected in **Table 3**. The first column gives the total QM/MM energies, relative to the Add calculations. This is only to illustrate that the QM/MM energy depends on the MM force field and therefore give different results for the various calculations.

The second column gives the QM energy of the QM/MM optimized structure of ethanol. It can be seen that BLAC1J and ME give the lowest energy, 1.7 kJ/mol above the QM minimum. On the other hand, ELAC+BLAC1 gives the highest energy, 3.3 kJ/mol. Thus, the variation in energies is small, showing that all structures give excellent structures. A MM minimisation with the GAFF force field gives a slightly higher energy, 3.8 kJ/mol (row MM in **Table 3**).

The third column shows the root-mean-squared deviation (RMSD) of the coordinates from the optimized QM structure of ethanol. ELAC2 gives the lowest RMSD (0.012 Å), followed by the two BLAC1 variants (0.013 Å), as well as Add, Sub and ME (0.014–0.016 Å). The three variants involving BLAC2 give slightly higher values, 0.036–0.039 Å.

Finally, the three last columns give the mean absolute deviation (MAD) for the 8 bonds, the 13 angles and the 12 dihedrals in the molecule. For the bonds, the three BLAC2 variants give minimal errors (0.001–0.002 Å), whereas the other methods give slightly higher errors, 0.005–0.006 Å. The same



**FIGURE 3 |** The OXF and OXT ligands, as well as the five states in the reaction mechanism of haem oxygenase. Atom names are indicated for the state **1**.

applies for the angles and the dihedrals, the three BLAC2 variants are still the best ( $0.8$  and  $0.4^\circ$ ), whereas the other methods give slightly larger errors,  $0.8\text{--}1.2^\circ$  and  $0.5\text{--}0.7^\circ$ .

In conclusion, the test calculations show that the BLAC approaches give the best result, but it depends on the force field

used. The best structures are obtained with the Hess2FF force field, which is tailored for the molecule and the QM method. ELAC sometimes improves the results, sometimes not. Sub and Add give similar results and in the differences among the various method are minimal for this small test molecule.

## Sulfite Oxidase

Next, we studied a more realistic enzyme system, viz. sulfite oxidase. Based on our recent QM-cluster (Van Severen et al., 2014) and QM/MM (Caldararu et al., 2018) studies, we considered the S → OMo mechanism, in which the S atom of the sulfite substrate attacks the equatorial oxo group of Mo<sup>VI</sup>, directly forming a Mo<sup>IV</sup>-sulfate intermediate (Im), via a first transition state TS1 (Figure 2). The sulfate product then dissociates into the second coordination sphere of the Mo ion via a second transition state (TS2).

For all five states, we tested six different methods: Add, Sub, ELAC, BLAC2, BLAC2J, and ME. For all methods, we compare the QM/MM results obtained with the small DMDT model of the molybdopterin ligand with those obtained with a standard (subtractive with VLAC) QM/MM calculation with the full MPT ligand (Figure 2). Initially, we tested also BLAC1, but it failed for all systems. The RS state with ME could be obtained only if the S<sub>Sub</sub>–O<sub>2</sub> distance was fixed at 2.43 Å (taken from the reference structure).

**Table 4** shows the RMSD deviations of the small QM system between the MPT and DMDT calculations. It can be seen that all calculations give similar results, with a RMSD of 0.02–0.10 Å. The RMSD is typically lowest for the Im and TS1 states.

**TABLE 3 |** Results for the QM/MM calculations on ethanol.

	$E_{QM/MM}$	$E_{EtOH}$	RMSD	$MAD_{bond}$	$MAD_{angle}$	$MAD_{dihed}$
Add	0.0	2.2	0.014	0.006	1.15	0.67
Sub	0.0	2.2	0.014	0.006	1.14	0.66
ELAC	−52.1	2.1	0.015	0.006	1.19	0.66
ELAC2	−5.1	2.0	<b>0.012</b>	0.006	1.07	0.74
BLAC1	−0.1	2.0	0.013	0.006	1.01	0.58
BLAC1J	0.3	<b>1.7</b>	0.013	0.006	1.02	0.58
BLAC2	−8.3	2.9	0.036	<b>0.001</b>	0.80	0.38
BLAC2J	−8.1	2.9	0.036	0.002	0.80	0.38
ELAC+BLAC2J	−60.2	3.3	0.039	0.002	<b>0.77</b>	<b>0.37</b>
ME	−56.5	<b>1.7</b>	0.016	0.005	1.12	0.49
MM		3.8	0.015	0.014	0.78	0.83

$E_{QM/MM}$  is the total QM/MM energy (kJ/mol),  $E_{EtOH}$  is the QM energy of the whole ethanol molecule (kJ/mol), RMSD is the RMS difference of the coordinates,  $MAD_{bond}$  (Å),  $MAD_{angle}$  (°), and  $MAD_{dihed}$  (°) are the mean absolute deviation compared to the ethanol molecule optimized by QM. The best value in each column is marked in bold face.

**TABLE 4 |** RMS deviations (Å) of the various QM systems for sulfite oxidase, compared to the QM/MM structures optimized with the full MPT ligand.

	Add	Sub	ELAC	BLAC1	BLAC2J	ME
RS	0.079	0.085	0.085	0.094	0.094	0.101
Ts1	0.025	0.027	0.028	0.040	0.041	0.053
Im	0.019	0.016	0.020	0.036	0.036	0.071
Ts2	0.063	0.070	0.074	0.083	0.083	0.087
PS	0.064	0.070	0.075	0.082	0.083	0.087
Average	0.047	0.050	0.052	0.063	0.064	0.078

The smallest RMSD is always found for Add and Sub methods, with an average RMSD of 0.05 Å for the five states. However, ELAC also gives a similar average, whereas that of the other three methods is somewhat larger, 0.06 and 0.08 Å for BLAC and ME, respectively. This increase in the RMSD is not caused by a single structure, but is seen for all structures.

In Table S1, the Mo-ligand and S<sub>Sub</sub>–O<sub>2</sub> distances for all structures are listed. It can be seen that these key distances are well preserved in the truncated calculations. The best results are again obtained for Add and Sub, for which the average difference for the nine distances and five sets of structures is only 0.008 Å. The maximum deviation is 0.14 Å for Add and 0.16 Å for Sub, in both cases obtained for the non-bonded S<sub>Sub</sub>–O<sub>2</sub> distance in the RS structure. Besides this distance, the largest deviation is only 0.02–0.03 Å. The results for the other four sets of calculations were slightly worse, with an average error of 0.017 Å for ME and 0.009 Å for the other three approaches. The maximum error is 0.14–0.17 Å, again for S<sub>Sub</sub>–O<sub>2</sub> distance in the RS structure, except for ME (Mo–O<sub>2</sub> distance of the Im state).

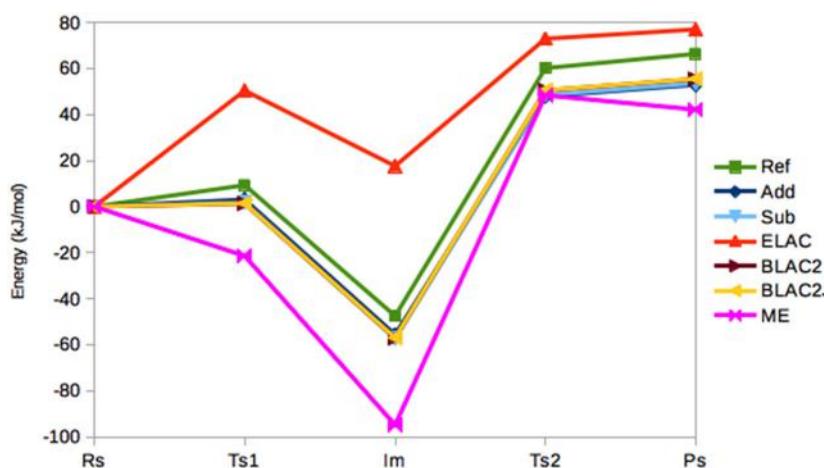
**Figure 4** shows the energies (relative to the RS state) in the seven sets of calculations. It can be seen that the Add, Sub, BLAC2, and BLAC2J methods give very similar results. In fact, the two sets of BLAC2 energies differ by only 0.1 kJ/mol for all states and these methods differ by 0–2 kJ/mol from Sub, and slightly more from Add (2 kJ/mol on average). The Add and Sub results agree within 2 kJ/mol. All these curves follow quite closely the reference with a systematic underestimation that increases from 6–8 kJ/mol for TS1 to 11–13 kJ/mol for PS. On average, all methods give a MAD of 10 kJ/mol, lowest for BLAC2J.

On the other hand, ME and ELAC give appreciably larger errors, with MADs of 29 and 32 kJ/mol, respectively and maximum errors of 47 and 65 kJ/mol. For ME, the problem is related to the failure to find the RS state—different restraints for this state may translate the curve upwards and therefore reduce the error, but it always remains worse than the other four methods.

Thus, we can conclude that all methods give reasonable structures for sulfite oxidase, although ME has problem with one of the states. However, ME and ELAC give quite large errors for the energies. In general, Add and Sub seem to give the best (and similar) results.

## Haem Oxygenase

The third test case is haem oxygenase, for which we studied the conversion of oxophlorin (OXF) to verdohaem (Alavi et al., 2017). As is shown in **Figure 3**, this involves five states and four transition states. It starts from the Fe<sup>III</sup>-OXF-O<sub>2</sub><sup>−</sup> complex in the doublet state (**1**), which has one unpaired electron on each of the three moieties, in analogy with previous studies (Alavi et al., 2017; Gheidi et al., 2017). In the first step of the reaction, the terminal oxygen atom in O<sub>2</sub> (O2) reacts with the C4B atom of OXF, forming a bridging intermediate (**2**). In the next step, the O–O bond is cleaved, giving intermediate **3**. Next, the O<sub>2</sub> atom reacts with another atom in the OXF ring (C1C), forming a four-membered ring in intermediate **4**. Finally, the C4B–CMC and C1C–CMC bonds are cleaved, and CO dissociates, giving rise to verdohaem (**5**). These states are separated by four transition



**FIGURE 4 |** Relative energies (kJ/mol) of the five states in the sulfite oxidase reaction, obtained with the various QM/MM methods.

states (**T1–T4**). We study the effect of moving the side chains of the OXF ring from the QM to the MM system.

This test case is challenging for at least two reasons. First, the electronic structure is more complicated than for the other two test cases, with several antiferromagnetically coupled open-shell moieties. Second, the OXF ring, for which we test the effect of truncation, is involved in the reaction. In fact, the C1C and C4B atoms are both only two bonds away from the HL atoms in the truncated OXT model.

As for sulfite oxidase, we tested six different methods with the truncated OXT model, Add, Sub, ELAC, BLAC1, BLAC2J, and ME. In contrast to sulfite oxidase, the simplest BLAC1 approach, with standard GAFF parameters for both OXF and OXT worked well. Considering the similar results for BLAC2 and BLAC2J for sulfite oxidase, we did not test BLAC2.

As for sulfite oxidase, we used the QM/MM calculations (subtractive with VLAC) with the full OXF ligand as the reference and study how the various QM/MM calculations with OXT reproduce these calculations in terms of the RMSD deviation for the entire (truncated) QM system, key distances and energies. The RMSD deviations of the nine different systems are shown in **Table 5**. It can be seen that most methods give similar results with an average RMSD of 0.06 Å. Add gives the lowest RMSD for most systems, but that of Sub is very similar and sometimes lower. The largest RMSD (0.09–0.10 Å) is typically found for state **3**, for which ME actually gives the best results and the latter method also gives the lowest maximum RMSD.

However, for the BLAC1 method, the RMSD is slightly larger, with an average of 0.07 Å and a maximum of 0.11 Å (still for **3**). For BLAC2J, the results are even worse (0.08 Å on average and a maximum of 0.13 Å for **T3**). In particular, BLAC2J failed to converge to any reasonable structure for the product (**5**). This most likely reflects that the force fields, especially that of BLAC2J, were determined for the starting structure **1** (note that BLAC2J gives the second-best structure for that state) and was then used unchanged for the other states. This is clearly suboptimal for structures later in the reaction mechanism. Of course, we

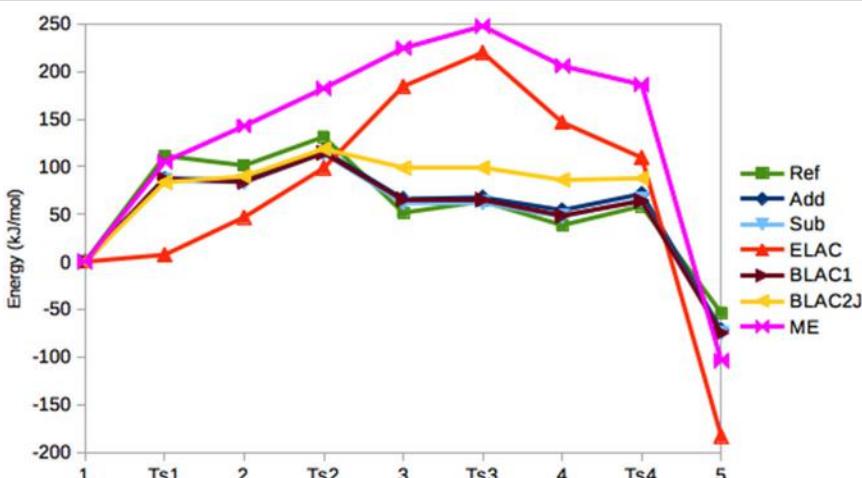
**TABLE 5 |** RMS deviations (Å) of the various QM systems of haem oxygenase, compared to the QM/MM structures optimized with the full OXF ligand.

	Add	Sub	ELAC	BLAC1	BLAC2J	ME
1	0.053	0.066	0.072	0.071	0.058	0.073
T1	0.046	0.053	0.057	0.057	0.069	0.057
2	0.042	0.050	0.053	0.055	0.065	0.053
T2	0.047	0.052	0.054	0.058	0.064	0.056
3	0.096	0.098	0.099	0.106	0.110	0.091
T3	0.082	0.083	0.087	0.089	0.126	0.083
4	0.086	0.086	0.091	0.092	0.104	0.094
T4	0.030	0.031	0.035	0.043	0.069	0.037
5	0.032	0.031	0.033	0.043		0.034
Average	0.057	0.061	0.065	0.068	0.083	0.064

could have determined new force fields for each intermediate and transition state in the mechanism, but this would have required much large computational and manual effort. Moreover, it would have given problems in the calculated energies, because the force field would be different for every state, making the energies not comparable. Thus, we do not recommend the BLAC approaches if the reaction is within three bonds of the XL atoms.

These trends in the RMSD are reflected also in the individual distances in the complexes. In Table S2, we examine the six Fe–N/O distances, as well as distances involving the reacting O<sub>2</sub>, C1C, C4B, and CMC atoms. It can be seen that all methods give similar MAD and maximum deviations from the reference distances (0.02–0.03 and 0.29–0.32 Å, respectively). However, Add and Sub still give the best results on average and BLAC2J the worst. All methods give large errors for the Fe–N<sub>His</sub> distances in the **3**, **T4** and **4** states (0.26–0.32 Å too short). All methods also give a large error for the O<sub>2</sub>–C1C distance in state **3** (0.18–0.27 Å too short).

In **Figure 5**, the relative energies of the various states are shown. It can be seen that the Add and Sub methods still give



**FIGURE 5 |** Relative energies (kJ/mol) of the five states in the haem oxygenase reaction, obtained with the various QM/MM methods.

similar results, with a MAD of only 4 kJ/mol. However, the Sub and BLAC1 methods give even more similar results with a MAD of only 1.4 kJ/mol. This reflects that GAFF parameters for OXT and OXF differ only for a few bonds, angles, and dihedrals around the periphery of the ring, which apparently do not affect the results much. These three methods also reproduce the reference calculations fairly well, with MADs of 14–15 kJ/mol, with Sub giving the smallest error. The maximum error, 24–25 kJ/mol, is obtained for T2.

BLAC2J gives somewhat worse energies (MAD = 30 kJ/mol, with a maximum error of 47 kJ/mol for both 3 and 4), especially for the later states in the reaction, reflecting that the force field is worse for these states. As for sulfite oxidase, the results for ELAC and ME are much worse, with MADs of 96–100 kJ/mol and maximum errors of 156–184 kJ/mol.

Finally, we checked also the electronic structures in the various calculations. However, they showed only a small variation between the various methods. For example, for the spin densities on Fe (shown in Table S3), Add, Sub, BLAC1 and BLAC2J give the same MAD from the reference calculations, 0.06  $e$ , and the MADs of ELAC and ME are only slightly larger, 0.07 and 0.08  $e$ , respectively. The spin density is lower for all calculations with OXT, except for the 1 state and the difference is largest for 3 and T3 (0.09–0.12  $e$ ).

In conclusion, Sub and Add again give the best results, but those of BLAC1 are also good. Again, ELAC and ME give large errors in the energies and BLAC2J has problems with geometries of the later states in the reaction.

## CONCLUSIONS

In this paper, we have tried to clarify similarities and differences between the subtractive and additive QM/MM schemes. In our view, the primary difference is that the subtractive scheme allows for an attempt to correct for errors introduced by the link

atoms. This correction can be introduced for three different types of interactions: van der Waals, electrostatic, and bonded interactions. Different software implements different corrections and it is also partly up to the user to define the level of correction (when setting up the force field for the QM system). For example, our ComQum software automatically implements the van der Waals correction, whereas ONIOM with electrostatic embedding implements also the electrostatic correction. Both software can implement the bonded correction, but typically do not do so.

Of course, the corrections come at some extra cost. For the van der Waals correction (VLAC), the cost is minimal: It requires van der Waals parameters of the HL atoms, which can almost always be taken from standard parameters for hydrogen atoms in the used MM force field (using the rules for atom types of the force field). Beside these parameters, the calculations can be automatically set up from a prmtop file for the full system with no extra effort, as in the ComQum implementation.

For the electrostatic link-atom corrections (ELAC), QM charges of the QM region are required, both with HL and CL atoms. This can be obtained from most QM software at a small extra cost and be automatized. Moreover, such charges are normally already available, because most QM/MM studies start with a MD equilibration of the full system (including solvent), for which a proper charge model of the QM system is needed. However, the charges on the CL atoms are ambiguous and ESP charges of buried atoms are poorly defined, which becomes a serious problem for large QM regions. Therefore, we have not seen any advantage for the electrostatic correction, at least not when the link atoms are rather close to the reactive atoms (Hu et al., 2011b). Moreover, the test calculations in this article indicate that ELAC can give rise to problems with energies in a reaction sequence. ELAC is implemented in the ONIOM software, but in practice it gives often severe convergence problems and is therefore seldom used. Instead, alternative approaches have been implemented, based on iterative calculations with

mechanical embedding and updated charges (Kawatsu et al., 2011; Dutta and Mishra, 2014; Wójcik et al., 2014). It could therefore be recommended that ONIOM implements electrostatic embedding with only VLAC, which is a very stable approach in ComQum.

For the bonded link-atom correction (BLAC), an accurate MM force field for the QM region is required. Nowadays, the general MM force field for organic and drug-like molecules often provide the required parameters. However, it is unclear whether these (together with the MM parameters of the full system) are accurate enough to give any advantage of this approach. The alternative is to make a tailored force field for the QM region, both when truncated and in the full protein. In this article, we have tested both approaches. For the simple ethanol test case, the best results were actually obtained with BLAC2, i.e., with the optimized Hess2FF force field. However, for the two enzyme systems, the results were worse, especially if the reactive site is close to the link atoms. Therefore, we cannot recommend BLAC for general use.

Beside the parameters needed for these link-atom corrections, there is no difference in the requirement of MM parameters for the subtractive and additive QM/MM schemes; without any link-atom corrections, the two schemes should give identical results if correctly implemented and require exactly the same MM parameters. However, in practice there may be differences because the subtractive scheme is typically based on a standard (general-purpose) MM software, whereas the additive scheme is based on a software tailored for QM/MM. In particular, most MM software refuse to run if any MM parameters are missing. Therefore, subtractive calculations normally require a full set of MM parameters, also for the QM region. However, these can be dummy (zeroed) parameters, because they cancel in the QM/MM calculations. Moreover, also the additive scheme requires these parameters for an initial MD equilibration of the full solvated system. The same parameters can normally be used throughout a reaction mechanism (again because these MM terms cancel in Equation 4).

Thus, our conclusion is that intrinsically, the subtractive and additive QM/MM schemes are equivalent if properly implemented. The subtractive scheme allows the introduction of various link-atom corrections, at the expense of requiring more MM parameters. For van der Waals and electrostatic corrections, the extra cost is minimal and fully automatic, whereas for the bonded corrections, significant extra effort may be needed. Of course, the same corrections may be implemented also in an additive scheme, by picking proper terms, but this goes outside a standard implementation (and also a strict definition) of the additive scheme and therefore should be thoroughly specified.

In practice, the subtractive scheme is easier to implement and maintain (standard QM and MM software are used). On the other hand, the additive scheme may be somewhat easier to set

up and can be tailored for QM/MM calculations. Moreover, if (a major part of) the MM system is fixed, calculations can be somewhat sped up by not calculating the MM energy and forces for the fixed atoms. In our test calculations, the additive and subtractive calculations with VLAC, but no other corrections (i.e., Add and Sub) give closely similar results, showing that the VLAC has only a minor influence on the results. Thus, both approaches can be recommended for QM/MM calculations.

We have also included mechanical embedding (within a subtractive scheme with VLAC) in the comparison. However, this approach gave rather poor structures, especially for sulfite oxidase. Moreover, the energies were quite poor, although this may be partly attributed to the fact that the reference energies employed electrostatic embedding and not mechanical embedding. For a more fair comparison, reference energies obtained with very large QM systems should be used, as in our previous study (Hu et al., 2011b).

Finally, we want to emphasize the importance of specifying exactly what is done in the QM/MM calculations, owing to the many different implementations. Obviously, it is not enough to say that a subtractive or additive scheme is used. Instead, for a subtractive scheme, it must be specified what type of link-atom corrections is applied (van der Waals, electrostatic, or bonded). In addition, the treatment of QM-MM electrostatics (mechanical or electrostatic embedding) must be specified, together with a detailed account of what charges are included in the point-charge model, if any charge redistribution scheme is employed and how charges on CL atoms are obtained. Finally, the treatment of link atoms need to be specified, as well as the relation between the coordinates of the HL and CL atoms.

## AUTHOR CONTRIBUTIONS

LC: Performed most of the calculations; UR: Did some calculations, designed the project, and wrote the article.

## ACKNOWLEDGMENTS

This investigation has been supported by grants from the Swedish research council (project 2014-5540), from Knut and Alice Wallenberg Foundation (KAW 2013.0022) and from COST through Action CM1305 (ECOSTBio). The computations were performed on computer resources provided by the Swedish National Infrastructure for Computing (SNIC) at Lunarc at Lund University.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2018.00089/full#supplementary-material>

## REFERENCES

- Ahrlrichs, R., Bär, M., Häser, M., Horn, H., and Kölmel, C. (1989). Electronic structure calculations on workstation computers: the program system Turbomole. *Chem. Phys. Lett.* 162, 165–169. doi: 10.1016/0009-2614(89)85118-8
- Alavi, F. S., Zahedi, M., Safari, N., and Ryde, U. (2017). QM/MM study of the conversion of oxophlorin into verdoheme by heme oxygenase. *J. Phys. Chem. B* 121, 11427–11436. doi: 10.1021/acs.jpcb.7b08332
- Balcells, D., and Maseras, F. (2007). Computational approaches to asymmetric synthesis. *New J. Chem.* 31, 333–343. doi: 10.1039/b615528f
- Besler, B. H., Merz, K. M., and Kollman, P. A. (1990). Atomic charges derived from semiempirical methods. *J. Comput. Chem.* 11, 431–439. doi: 10.1002/jcc.540110404
- Blomberg, M. R. A., Borowski, T., Himo, F., Liao, R.-Z., and Siegbahn, P. E. M. (2014). Quantum chemical studies of mechanisms for metalloenzymes. *Chem. Rev.* 114, 3601–3658. doi: 10.1021/cr400388t
- Brunger, A. T. (2007). Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.* 2, 2728–2733. doi: 10.1038/nprot.2007.406
- Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 54, 905–921. doi: 10.1107/S0907449898003254
- Caldararu, O., Feldt, M., Cioloboc, D., van Severen, M.-C., Starke, K., Mata, R. A., et al. (2018). QM/MM study of the reaction mechanism of sulfite oxidase. *Sci. Rep.* 8:4684. doi: 10.1038/s41598-018-22751-6
- Cao, L., Caldararu, O., Rosenzweig, A. C., and Ryde, U. (2018). Quantum refinement does not support dinuclear copper sites in crystal structures of particulate methane monooxygenase. *Angew. Chem. Int. Ed.* 57, 162–166. doi: 10.1002/anie.201708977
- Case, D. A., Berryman, J. T., Betz, R. M., Cerutti, D. S., Cheatham, T. E., Darden, T. A., et al. (2014). *AMBER 14*. San Francisco, CA: University of California. Available online at: <http://ambermd.org>
- Chung, L. W., Sameera, W. M. C., Ramozzi, R., Page, A. J., Hatanaka, M., Petrova, G. P., et al. (2015). The ONIOM Method and its applications. *Chem. Rev.* 115, 5678–5796. doi: 10.1021/cr5004419
- Dapprich, I. K., Komáromia, I., Morokuma, and, M. J., Frisch, K. S. B. (1999). A New ONIOM Implementation in Gaussian 98. 1. The calculation of energies, gradients and vibrational frequencies and electric field derivatives. *J. Mol. Struct.* 462, 1–21. doi: 10.1016/S0166-1280(98)00475-8
- Dutta, D., and Mishra, S. (2014). The structural and energetic aspects of substrate binding and the mechanism of action of the DapE-encoded N-succinyl-[l,]-diaminopimelic acid desuccinylase (DapE) investigated using a hybrid QM/MM method. *Phys. Chem. Chem. Phys.* 16, 26348–26358. doi: 10.1039/C4CP03986F
- Field, M. J., Bash, P. A., and Karplus, M. (1990). A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* 11, 700–733. doi: 10.1002/jcc.540110605
- Furche, F., Ahrlrichs, R., Hättig, C., Klopper, W., Sierka, M., and Weigend, F. (2014). Turbomole. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 4, 91–100. doi: 10.1002/wcms.1162
- Gao, J., Amara, P., Alhambra, C., and Field, M. J. (1998). A Generalized Hybrid Orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem. A* 102, 4714–4721. doi: 10.1021/jp9809890
- Gheidi, M., Safari, N., and Zahedi, M. (2017). Density functional theory studies on the conversion of hydroxyheme to iron-verdoheme in the presence of dioxygen. *Dalt. Trans.* 46, 2146–2158. doi: 10.1039/C6DT04250C
- Götz, A. W., Clark, M. A., and Walker, R. C. (2014). An extensible interface for QM/MM molecular dynamics simulations with AMBER. *J. Comput. Chem.* 35, 95–108. doi: 10.1002/jcc.23444
- Grimme, S., Antony, J., Ehrlich, S., and Krieg, H. (2010). A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* 132:154104. doi: 10.1063/1.3382344
- Grimme, S., Ehrlich, S., and Goerigk, L. (2011). Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* 32, 1456–1465. doi: 10.1002/jcc.21759
- Hu, L., and Ryde, U. (2011). Comparison of methods to obtain force-field parameters for metal sites. *J. Chem. Theory Comput.* 7, 2452–2463. doi: 10.1021/ct100725a
- Hu, L., Farrokhnia, M., Heimdal, J., Shleev, S., Rulíšek, L., and Ryde, U. (2011a). Reorganization energy for internal electron transfer in multicopper oxidases. *J. Phys. Chem. B* 115, 13111–13126. doi: 10.1021/jp205897z
- Hu, L., Söderhjelm, P., and Ryde, U. (2011b). On the convergence of QM/MM energies. *J. Chem. Theory Comput.* 7, 761–777. doi: 10.1021/ct100530r
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869
- Jover, J., and Maseras, F. (2016). QM/MM calculations on selectivity in homogeneous catalysis. *Struct. Bond.* 167, 59–79. doi: 10.1007/430\_2015\_188
- Kawatsu, T., Lundberg, M., and Morokuma, K. (2011). Protein free energy corrections in ONIOM QM/MM modeling: a case study for isopenicillin N synthase (IPNS). *J. Chem. Theory Comput.* 7, 390–401. doi: 10.1021/ct1005592
- Keal, T. W., Sherwood, P., Dutta, G., Sokol, A. A., and Catlow, C. R. A. (2011). Characterization of hydrogen dissociation over aluminium-doped zinc oxide using an efficient massively parallel framework for QM/MM calculations. *Proc. R. Soc. Lond. A Math. Phys. Eng. Sci.* 467, 1900–1924. doi: 10.1098/rspa.2010.0613
- Levitt, M. (1976). A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* 104, 59–107. doi: 10.1016/0022-2836(76)90004-8
- Lin, H., and Truhlar, D. G. (2007). QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* 117, 185–199. doi: 10.1007/s00214-006-0143-z
- Lopes, P. E. M., Roux, B., and MacKerell, A. D. (2009). Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: theory and applications. *Theor. Chem. Acc.* 124, 11–28. doi: 10.1007/s00214-009-0617-x
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713. doi: 10.1021/acs.jctc.5b00255
- Maseras, F., and Morokuma, K. (1995). IMOMM: a new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem.* 16:1170. doi: 10.1002/jcc.540160911
- Murphy, R. B., Philipp, D. M., and Friesner, R. A. (2000). A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. *J. Comput. Chem.* 21, 1442–1457. doi: 10.1002/1096-987X(200012)21:16<1442::AID-JCC3>3.0.CO;2-O
- Nilsson, K., Lecerof, D., Sigfridsson, E., and Ryde, U. (2003). An automatic method to generate force-field parameters for hetero-compounds. *Acta Crystallogr. D Biol. Crystallogr.* 59, 274–289. doi: 10.1107/S0907444902021431
- Olsen, J. M., Aidas, K., and Kongsted, J. (2010). Excited states in solution through polarizable embedding. *J. Chem. Theory Comput.* 6, 3721–3734. doi: 10.1021/ct1003803
- Poulsen, T. D., Kongsted, J., Osted, A., Ogilby, P. R., and Mikkelsen, K. V. (2001). The combined multiconfigurational self-consistent-field/molecular mechanics wave function approach. *J. Chem. Phys.* 115, 2393–2400. doi: 10.1063/1.1374559
- Ramos, M. J., and Fernandes, P. A. (2008). Computational enzymatic catalysis. *Acc. Chem. Res.* 41, 689–698. doi: 10.1021/ar7001045
- Reuter, N., Dejaegere, A., Maigret, B., and Karplus, M. (2000). Frontier Bonds in QM/MM Methods: a comparison of different approaches. *J. Phys. Chem. A* 104, 1720–1735. doi: 10.1021/jp9924124
- Rod, T. H., and Ryde, U. (2005). Accurate QM/MM free energy calculations of enzyme reactions: methylation by catechol O-methyltransferase. *J. Chem. Theory Comput.* 1, 1240–1251. doi: 10.1021/ct0501102
- Roßbach, S., and Ochsenfeld, C. (2017). Influence of coupling and embedding schemes on QM size convergence in QM/MM approaches for the example of a proton transfer in DNA. *J. Chem. Theory Comput.* 13, 1102–1107. doi: 10.1021/acs.jctc.6b00727
- Ryde, U. (1995). On the role of Glu-68 in alcohol dehydrogenase. *Protein Sci.* 4, 1124–1132. doi: 10.1002/pro.5560040611
- Ryde, U. (1996a). The coordination chemistry of the structural zinc ion in alcohol dehydrogenase studied by ab initio quantum chemical calculations. *Eur. Biophys. J.* 24, 213–221. doi: 10.1007/BF00205102

- Ryde, U. (1996b). The coordination of the catalytic zinc in alcohol dehydrogenase studied by combined quantum-chemical and molecular mechanics calculations. *J. Comput. Aided. Mol. Des.* 10, 153–164. doi: 10.1007/BF00402823
- Ryde, U. (2016). QM/MM calculations on proteins. *Methods Enzymol.* 577, 119–158. doi: 10.1016/bs.mie.2016.05.014
- Ryde, U., and Nilsson, K. (2003). Quantum chemistry can locally improve protein crystal structures. *J. Am. Chem. Soc.* 125, 14232–14233. doi: 10.1021/ja0365328
- Ryde, U., and Olsson, M. H. M. (2001). Structure, strain, and reorganization energy of blue copper models in the protein. *Int. J. Quantum Chem.* 81, 335–347. doi: 10.1002/1097-461X(2001)81:5<335::AID-QUA1003>3.0.CO;2-Q.
- Ryde, U., Olsen, L., and Nilsson, K. (2002). Quantum chemical geometry optimizations in proteins using crystallographic raw data. *J. Comput. Chem.* 23, 1058–1070. doi: 10.1002/jcc.10093
- Schäfer, A., Horn, H., and Ahlrichs, R. (1992). Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.* 97, 2571–2577. doi: 10.1063/1.463096
- Seminario, J. M. (1996). Calculation of intramolecular force fields from second-derivative tensors. *Int. J. Quantum Chem.* 30, 1271–1277. doi: 10.1002/(SICI)1097-461X(1996)60:7<1271::AID-QUA8>3.0.CO;2-W
- Senn, H. M., and Thiel, W. (2009). QM/MM methods for biomolecular systems. *Angew. Chemie* 48, 1198–1229. doi: 10.1002/anie.200802019
- Sherwood, P., De Vries, A. H., Guest, M. F., Schreckenbach, G., Catlow, C. R. A., French, S. A., et al. (2003). QUASI: a general purpose implementation of the QM/MM approach and its application to problems in catalysis. *J. Mol. Struct. THEOCHEM* 632, 1–28. doi: 10.1016/S0166-1280(03)00285-9
- Sigfridsson, E., and Ryde, U. (1998). Comparison of methods for deriving atomic charges from the electrostatic potential and moments. *J. Comput. Chem.* 19, 377–395. doi: 10.1002/(SICI)1096-987X(199803)19:4<377::AID-JCC1>3.0.CO;2-P
- Singh, U. C., and Kollman, P. A. (1986). A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: applications to the CH<sub>3</sub>Cl + Cl<sup>-</sup> exchange reaction and gas phase protonation of polyethers. *J. Comput. Chem.* 7, 718–730. doi: 10.1002/jcc.540070604
- Söderhjelm, P., Husberg, C., Strambi, A., Olivucci, M., and Ryde, U. (2009). Protein influence on electronic spectra modeled by multipoles and polarizabilities. *J. Chem. Theory Comput.* 5, 649–658. doi: 10.1021/ct800459t
- Sousa, S. F., Ribeiro, A. J. M., Neves, R. P. P., Brás, N. F., Cerqueira, N. M. F. S. A., Fernandes, P. A., et al. (2017). Application of quantum mechanics/molecular mechanics methods in the study of enzymatic reaction mechanisms. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 7:e1281. doi: 10.1002/wcms.1281
- Stoyanov, S. R., Gusalov, S., Kuznicki, S. M., and Kovalenko, A. (2008). Theoretical modeling of zeolite nanoparticle surface acidity for heavy oil upgrading. *J. Phys. Chem. C* 112, 6794–6810. doi: 10.1021/jp075688h
- Svensson, M., Humbel, S., Froese, R. D. J., Matsubara, T., Sieber, S., and Morokuma, K. (1996). ONIOM: a multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. A test for diels-alder reactions and Pt(P(t-Bu)<sub>3</sub>)<sub>2</sub> + H<sub>2</sub> oxidative addition. *J. Phys. Chem.* 100, 19357–19363. doi: 10.1021/jp962071j
- Tao, J., Perdew, J. P., Staroverov, V. N., and Scuseria, G. E. (2003). Climbing the density functional ladder: non-empirical meta-generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.* 91:146401. doi: 10.1103/PhysRevLett.91.146401
- Théry, V., Rinaldi, D., Rivail, J.-L., Maigret, B., and Ferenczy, G. G. (1994). Quantum mechanical computations on very large molecular systems: the local self-consistent field method. *J. Comput. Chem.* 15, 269–282. doi: 10.1002/jcc.540150303
- TURBOMOLE version 7.1 (2016). *University of Karlsruhe and Forschungszentrum Karlsruhe GmbH*. Available online at: <http://www.turbomole.com/>
- Van Severen, M.-C., Andrejic, M., Li, J., Starke, K., Mata, R. A., Nordlander, E., et al. (2014). A quantum-mechanical study of the reaction mechanism of sulfite oxidase. *J. Biol. Inorg. Chem.* 19, 1165–1179. doi: 10.1007/s00775-014-1172-z
- Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., et al. (2010). CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* 31, 671–690. doi: 10.1002/jcc.21367
- Vreven, T., Byun, K. S., Komáromi, I., Dapprich, S., Montgomery, J. A., Jr., Morokuma, K., et al. (2006). Combining quantum mechanics methods with molecular mechanics methods in ONIOM. *J. Chem. Theory Comput.* 2, 815–826. doi: 10.1021/ct050289g
- Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general Amber force field. *J. Comput. Chem.* 25, 1157–1174. doi: 10.1002/jcc.20035
- Wójcik, A., Broclawik, E., Siegbahn, P. E. M., Lundberg, M., Moran, G., and Borowski, T. (2014). Role of substrate positioning in the catalytic reaction of 4-Hydroxyphenylpyruvate Dioxygenase—A QM/MM Study. *J. Am. Chem. Soc.* 136, 14472–14485. doi: 10.1021/ja506378u

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Cao and Ryde. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Automated Fragmentation QM/MM Calculation of NMR Chemical Shifts for Protein-Ligand Complexes

Xinsheng Jin<sup>1</sup>, Tong Zhu<sup>1,2</sup>, John Z. H. Zhang<sup>1,2,3</sup> and Xiao He<sup>1,2,4\*</sup>

<sup>1</sup> State Key Laboratory of Precision Spectroscopy, School of Chemistry and Molecular Engineering, Shanghai Engineering Research Center of Molecular Therapeutics and New Drug Development, East China Normal University, Shanghai, China,

<sup>2</sup> NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai, China, <sup>3</sup> Department of Chemistry, New York University, New York, NY, United States, <sup>4</sup> National Engineering Research Centre for Nanotechnology, Shanghai, China

## OPEN ACCESS

### Edited by:

Sam P. De Visser,  
University of Manchester,  
United Kingdom

### Reviewed by:

Ahmet Altun,  
Max-Planck-Institut für  
Kohlenforschung, Germany  
Giovanni La Penna,  
Consiglio Nazionale Delle Ricerche  
(CNR), Italy

### \*Correspondence:

Xiao He  
xiaohe@phy.ecnu.edu.cn

### Specialty section:

This article was submitted to  
Theoretical and Computational  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

Received: 30 January 2018

Accepted: 16 April 2018

Published: 08 May 2018

### Citation:

Jin X, Zhu T, Zhang JZH and He X  
(2018) Automated Fragmentation  
QM/MM Calculation of NMR Chemical  
Shifts for Protein-Ligand Complexes.  
Front. Chem. 6:150.  
doi: 10.3389/fchem.2018.00150

In this study, the automated fragmentation quantum mechanics/molecular mechanics (AF-QM/MM) method was applied for NMR chemical shift calculations of protein-ligand complexes. In the AF-QM/MM approach, the protein binding pocket is automatically divided into capped fragments (within ~200 atoms) for density functional theory (DFT) calculations of NMR chemical shifts. Meanwhile, the solvent effect was also included using the Poission-Boltzmann (PB) model, which properly accounts for the electrostatic polarization effect from the solvent for protein-ligand complexes. The NMR chemical shifts of neocarzinostatin (NCS)-chromophore binding complex calculated by AF-QM/MM accurately reproduce the large-sized system results. The <sup>1</sup>H chemical shift perturbations (CSP) between apo-NCS and holo-NCS predicted by AF-QM/MM are also in excellent agreement with experimental results. Furthermore, the DFT calculated chemical shifts of the chromophore and residues in the NCS binding pocket can be utilized as molecular probes to identify the correct ligand binding conformation. By combining the CSP of the atoms in the binding pocket with the Glide scoring function, the new scoring function can accurately distinguish the native ligand pose from decoy structures. Therefore, the AF-QM/MM approach provides an accurate and efficient platform for protein-ligand binding structure prediction based on NMR derived information.

**Keywords:** AF-QM/MM, NMR chemical shift, protein-ligand binding, scoring function, structure prediction

## INTRODUCTION

Structure-based computational methods are useful tools for analyzing the binding modes and affinities for protein-ligand complexes (Grinter and Zou, 2014). With the development of X-ray crystallography and nuclear magnetic resonance (NMR) technology, more than 100,000 high resolution three-dimensional protein structures have been determined, which is helpful for finding lead compounds and therapeutic targets (Ferreira et al., 2015). As compared to experimental methods, computational approaches, such as molecular docking, are the fast and efficient ways for drug discovery. In molecular docking programs, the scoring functions are used to approximate the binding free energy and hence to rank the simulated decoy poses (Sliwoski et al., 2014). Most scoring functions are roughly derived from force-field-based (Morris et al., 1998; Englebienne and Moitessier, 2009), empirical (Eldridge et al., 1997; Murray et al., 1998) or knowledge-based

potentials (Gohlke et al., 2000; Huang and Zou, 2006). However, based on parameterized functions, these scoring methods are usually not accurate enough to differentiate the experimental structure from the docked decoy structures, and sometimes the rankings from different software suites may be inconsistent (Śledź and Caflisch, 2018). For solving this problem, much effort has been devoted to the development of docking methods by introducing the experimental structural information or quantum mechanical (QM) calculations for scoring the native and predicted poses (Mohan et al., 2005; Grinter and Zou, 2014; Adeniyi and Soliman, 2017).

The chemical shift is one of the most effective and precise NMR parameters in reflecting the local chemical environment around the atom, which plays an important role in structure determination and refinement (Zhu et al., 2014; Bratholm and Jensen, 2017). For NMR chemical shift calculations, the empirical chemical shift prediction softwares include ShiftS (Xu and Case, 2001; Moon and Case, 2007), ShiftX (Neal et al., 2003), ShiftX2 (Han et al., 2011), CamShift (Robustelli et al., 2010), PROSHIFT (Meiler, 2003; Meiler and Baker, 2003), SHIFTCALC (Williamson and Craven, 2009), ProCS (Christensen et al., 2014; Bratholm and Jensen, 2017), CheShift (Vila et al., 2009; Garay et al., 2014) and Sparta+ (Shen and Bax, 2010). These empirical methods are fast in computational speed, and have been successful in predicting backbone chemical shifts. As the empirical formulas for these models are derived from fitting the experimental or QM calculated chemical shift database and the high-quality structures, these models are not well suited for accurate prediction of NMR chemical shifts for some complex systems such as protein-ligand complexes, non-standard protein residues or non-canonical base pairs in nucleic acid systems (Swails et al., 2015). The quantum mechanical chemical shift calculations are in principle able to predict the NMR chemical shifts for any complex systems (Lodewyk et al., 2012; Hartman and Beran, 2014; Merz, 2014). For protein NMR chemical shift calculations, Cui and Karplus had proposed a very effective QM/MM approach (Cui and Karplus, 2000), Gao et al. developed fragment molecular orbital (FMO) method (Gao et al., 2007, 2010), Exner and coworkers utilized the adjustable density matrix assembler (ADMA) approach (Frank et al., 2012; Victora et al., 2014), Tan and Bettens developed the combined fragmentation method (CFM) (Tan and Bettens, 2013), and He and coworkers developed the automated fragmentation quantum mechanics/molecular mechanics (AF-QM/MM) method (He et al., 2009, 2014; Zhu et al., 2012, 2013, 2014, 2015; Swails et al., 2015; Jin et al., 2016). These fragment-based QM methods have been successfully applied for NMR chemical shift calculation of proteins and nucleic acids.

On the basis that chemical shifts or chemical shift perturbations (CSP) are sensitive to the variations of chemical environment, these parameters are quite suitable for structure determination (Case, 1998; Cavalli et al., 2007; Shen and Bax, 2015). Many NMR-based methods have been developed for prediction of protein-ligand binding modes (Medek et al., 2000; Cioffi et al., 2008; Riedinger et al., 2008; Aguirre et al., 2014). McCoy and Wyss utilized proton CSP data, induced by aromatic ring current effect in the ligands, to locate the ring position

of docking structures (McCoy and Wyss, 2002). Recently, Ten Brink et al. compared the experimental and simulated CSPs to verify protein conformational changes and developed the CSP-based docking method (Ten Brink et al., 2015). Merz et al. developed CSP-based scoring functions to determine the binding poses for protein-ligand complexes (Wang et al., 2007; Yu et al., 2017). However, most of these scoring functions only calculate the proton chemical shift on proteins. The scoring functions could be more efficient by taking ligand <sup>1</sup>H chemical shifts into consideration.

In this work, we applied the AF-QM/MM approach for NMR chemical shift calculations on protein-ligand binding complexes. Based on DFT calculations, the <sup>1</sup>H chemical shifts on both protein and ligand are available for structure determination and improving the scoring functions. In the framework of the AF-QM/MM approach, the ligand is also divided into smaller fragments, and hence it significantly reduced the computational cost for <sup>1</sup>H NMR chemical shift calculation on the large ligand. In this study, the neocarzinostatin (NCS) protein is selected as the test case for the AF-QM/MM method because of its importance in cancer therapy. NCS has experimental chemical shifts for both apo and holo forms (Myers et al., 1988; Mohanty et al., 1994; Schaus et al., 2001; Takashima et al., 2005; Wang and Merz, 2010). Furthermore, by comparing the AF-QM/MM calculated chemical shifts with experiment data, a chemical shift based scoring function was developed to rank the native and predicted protein-ligand binding poses.

This paper is organized as follows: first, a benchmark test was performed using the AF-QM/MM method for NMR chemical shift calculations of protein-ligand complex. The computed results are compared to the large-sized system NMR chemical shift calculations. Subsequently, AF-QM/MM calculated chemical shifts are compared with the experimental results for both apo and holo NCS structures. Next, the performance of chemical shift based and conventional energy based scoring functions on the rankings of predicted protein-ligand binding poses is discussed. Finally, the hybrid scoring function, that combines the calculated NMR chemical shifts and binding energy, is applied to rank the experimental structure and other docked binding poses.

## COMPUTATIONAL APPROACHES

### Structure Preparation

The X-ray structures of apo and holo NCS were download from the Protein Data Bank (PDB ids: 1NOA and 1NCO, respectively). The experimental chemical shift data of apo protein and chromophore were obtained from previous studies (Myers et al., 1988; Mohanty et al., 1994). The holo NCS experiment chemical shifts were downloaded from Biological Magnetic Resonance Data Bank (BMRB entry: 5969). The structure minimization of the protein X-ray structures was performed using the AMBER12 program (Case et al., 2012) with the ff99SB force field. The apo and holo NCS structures were solvated in a truncated octahedral periodic box of TIP3P water molecules with each side at least 10 Å from the nearest solute atom (Jorgensen and Jenson, 1998). After the entire system was neutralized with counter ions, 1,000

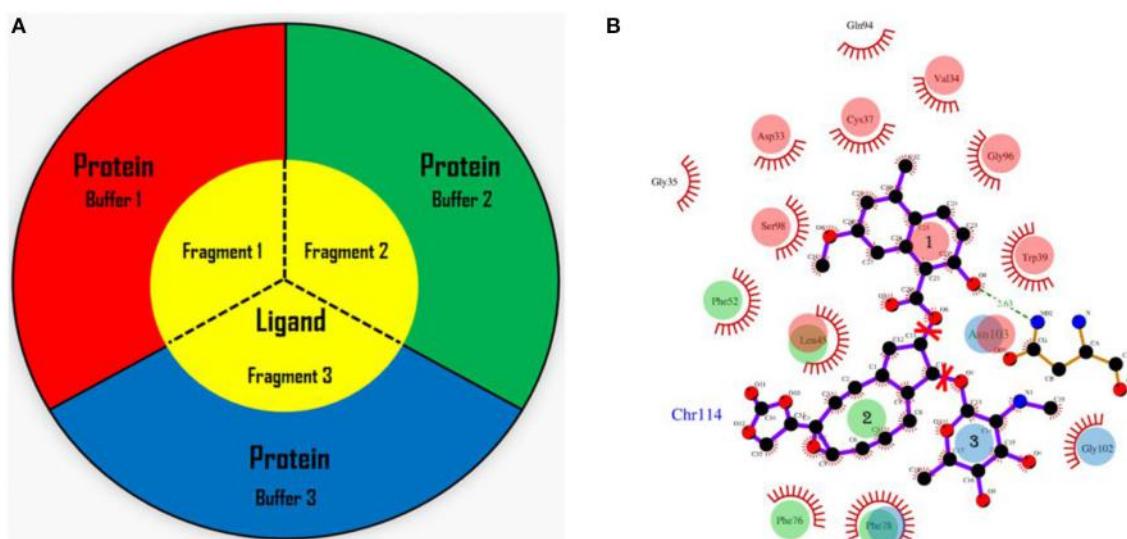
steps of steepest descent algorithm following with 4,000 steps of conjugate gradient method were used to remove the improper contacts of the system. For obtaining the force field parameters of the ligand, the general AMBER force field (GAFF) (Wang et al., 2004) and AM1-bond charge correlations (AM1-BCC) charge model were utilized for the ligand (Jakalian et al., 2002). The molecular docking was performed using the Glide module in the Schrödinger program (Friesner et al., 2004; Halgren et al., 2004). The scoring function used in this study was Glide XP. In this study, the protein structure was fixed when the ligand was docked into the binding site. Therefore, we did not include the flexibility of the protein during molecular docking. Based on the optimized experimental structure using the molecular force field, 38 docking poses of the ligand predicted by Glide, whose RMSDs range from 1.5 to 10.5 Å with reference to the native position, were selected for subsequent chemical shift calculations at the QM level.

## The AF-QM/MM Method for NMR Chemical Shift Calculation of the Protein-Ligand Complex

In the AF-QM/MM approach (He et al., 2009, 2014; Zhu et al., 2012, 2013; Swails et al., 2015; Jin et al., 2016), the apo protein is divided into individual residue by cutting through the peptide bonds. The number of fragments is the same as the number of residues in the protein. Each fragment contains a core region (each amino acid) in the protein, and the buffer region which contains the nearby residues surrounding the core region. Both the core and buffer regions are treated with quantum mechanics (QM) while the residues outside the buffer region are described

by molecular mechanics (MM). The details for the definition of the buffer regions are described in our previous work (He et al., 2009; Zhu et al., 2012, 2013). For the holo protein studied in this work, we developed the fragmentation scheme for the ligand and its surrounding protein residues. As shown in **Figure 1**, we also divided the chromophore into three parts by cutting the C-O single bond. Fragment 1 contains the naphthoate group, fragment 2 includes the enediyne ring and fragment 3 has the aminosugar group, respectively. For each fragment of the ligand (taken as each core region), the rest part of the ligand and the protein residues surrounding the core region (each fragment of the ligand) are taken as the buffer region (see **Figure 1**). The distance criteria for selecting the buffer region for each fragment of the ligand is the same as that for each residue in the protein.

In this study, we adopt the following distance-dependent criteria to include residues within the buffer region of each core region for ligand: (1) if a heavy atom of the residue is less than 3.5 Å away from any atom in the core region, (2) if the distance of a hydrogen atom of the residue is less than 3.0 Å away from any atom in the core region. The cutoff enables a sufficient size of the buffer region for the convergence of chemical shift calculations on the core region (Flaig et al., 2012). The remaining residues beyond the buffer region are described by embedding charges to account for the electrostatic field outside the QM region. The protein charges were obtained directly from the ff99SB force field. For assigning the buffer region for each residue (where each residue is defined as the core region), the ligand is treated as a whole molecule (non-fragmented), and the definition of the buffer region follows the same criteria as the apo protein. The dangling bonds are capped with hydrogen atoms for constructing the closed-shell fragment.



**FIGURE 1 |** The fragmentation scheme for the ligand in AF-QM/MM. **(A)** The ligand is divided into three fragments. Each fragment of the ligand is taken as the core region. The buffer region contains remaining part of the ligand and protein residues within the certain distance threshold from the core region (see the text for more details). The core and buffer regions are calculated at the QM level while rest of the system are described by embedding charges. **(B)** The definition of each core region of the chromophore. Fragment 1: the naphthoate group; fragment 2: the enediyne ring part; fragment 3: the aminosugar group. The buffer region has the same color as the core region.

The fragment QM calculations were carried out in parallel at the B3LYP/6-31G\*\* level. All QM calculations were performed using the Gaussian 09 package (Frisch et al., 2009). Only the NMR isotropic shieldings of the core region atoms were collected from each fragment QM calculation. The  $^1\text{H}$  chemical shifts are obtained by referencing to that of the tetramethylsilane (TMS) at the same computational level, which is 31.66 ppm. The implicit solvation model was applied to approximate the solvent effect. The protein charge distribution polarizes the dielectric solution and creates a reaction field to act back on the solute until equilibrium is reached. The reaction field acting on the solute can be effectively represented by the induced charges on the cavity surface. In this work, the surface charges are calculated by the Poisson-Boltzmann (PB) model using the Delphi program (Rocchia et al., 2001). The set of point charges of the MM environment and on the molecular surface, which represents the reaction field, are used as the background charges in the QM calculation. Because the computational cost of QM chemical shift calculations will be dramatically increased on multiple configurations when the conformational sampling effect is taken into account. In this study, the optimized X-ray structure using molecular force field was taken as a representative configuration for the ensemble averaging structure.

## Scoring Functions

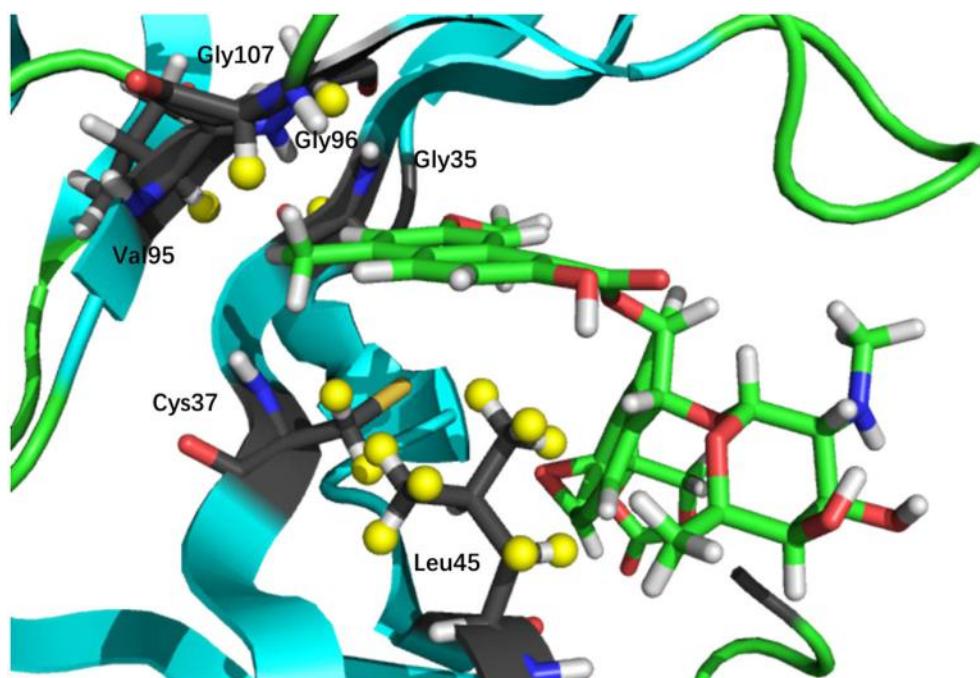
To differentiate the native protein-ligand binding structure from decoy poses, here we propose a scoring function based on NMR chemical shifts ( $\text{CS}_{\text{score}}$ ), which is simply the root-mean-square deviation (RMSD) of computed

chemical shifts with reference to the experimental values,

$$\text{CS}_{\text{score}} = \sqrt{\frac{\sum_{i=1}^N (\delta_{\text{H}}^i - \delta_{\text{exp}}^i)^2}{N}} \quad (1)$$

where  $\delta_{\text{H}}^i$  is the chemical shift of  $i$ th hydrogen atom on the ligand and nearby residues, and  $\delta_{\text{exp}}^i$  is the experimental chemical shift of the corresponding atom in the native complex (holo NCS).  $N$  is the number of atoms whose chemical shifts were selected as molecular probe to characterize the NCS-chromophore binding structure. In this study,  $N$  was set to 31 for holo NCS, 21 of which are non-amide protons on the chromophore, and the other 10 hydrogen atoms are those with experimental chemical shift perturbations (CSP, between the bound and unbound complexes) greater than 0.5 ppm from residues in the binding site of the protein (see **Figure 2** and Table S1 of the Supplementary Materials).

It is worth noting that, in Equation (1), we could also add the chemical shifts of the hydrogen atoms that experimentally do not change upon ligand binding. A false docking pose may cause significant deviations in CSPs for those hydrogen atoms. However, there are many of such protons on the residues in the binding pocket, which will average out the final score to make the scoring function incapable of distinguishing the native structure from the decoy sets. Protons with experimental perturbations greater than 0.5 ppm are more sensitive to the binding pose, therefore we took those atoms into account in the scoring function. Furthermore, although NMR chemical shifts of amide



**FIGURE 2** | 3D structure of the binding pocket. The nonpolar hydrogen atoms on the ligand, and protein protons whose CSP values (between bound and unbound complexes) are greater than 0.5 ppm (denoted in yellow) are chosen in the chemical shift based scoring functions (namely,  $\text{CS}_{\text{score}}$  and  $\text{CS}_{\text{Gs}}$ ).

protons are also very sensitive to the local chemical environment of the binding pocket, these atoms were excluded owing to the lack of experimental data.

The second scoring function ( $CS_{Gscore}$ ) we propose here, is a linear combination of  $CS_{score}$  and Glide score,

$$CS_{Gscore} = CS_{score} + \alpha \text{Glide Score} \quad (2)$$

where  $\alpha$  is a weighting factor. In this study, the ranges of  $CS_{score}$  and Glide Score are  $0.42 \sim 2.90$  and  $-10.96 \sim 10.98$  (see Table S2 of the Supplementary Materials), respectively, and thus we choose  $\alpha = (2.90 - 0.42) / (10.98 - (-10.96)) \approx 0.1$ . By adding the Glide score to  $CS_{score}$ , the unphysical structure with an unfavorable binding energy will be avoided. Since both the  $CS_{score}$  and Glide score will be smaller as the docking pose gets closer to the experimental structure, the native docking pose will give the lowest  $CS_{Gscore}$  value.

## RESULTS AND DISCUSSION

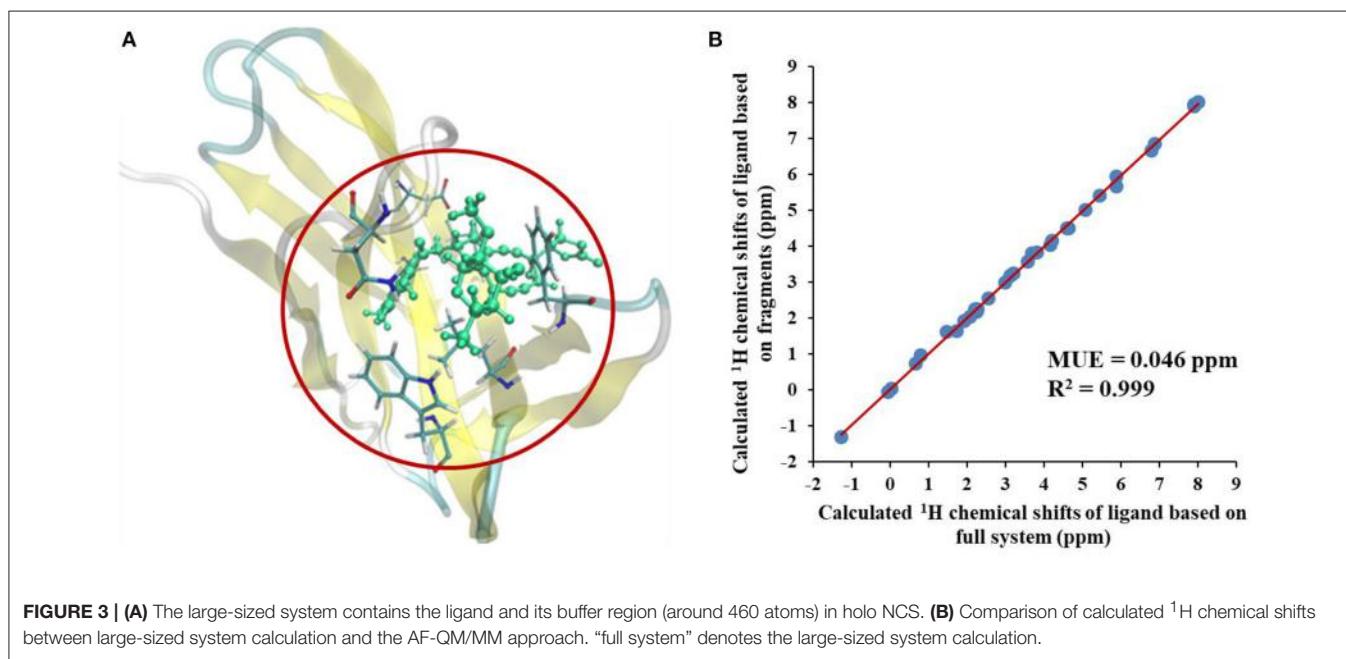
### Benchmark Test of AF-QM/MM on the Native NCS-Chromophore Binding Complex

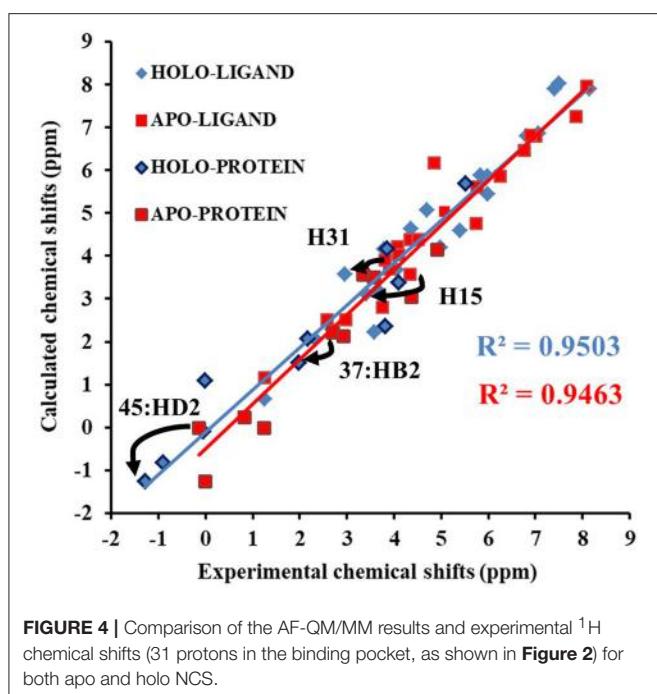
We first compared the calculated chemical shifts on the chromophore between AF-QM/MM and large-sized system calculations. Because the holo NCS contains more than 1500 atoms and was too large to perform full system QM calculations, we alternatively used the entire ligand and its buffer region for large-sized QM calculation. The other atoms beyond the buffer region are taken as background charges, and the PB surface charges for the entire complex are also placed to approximate the implicit solvent. The computed chemical shifts from such a model system (around 460 atoms in the QM region) are taken as the reference values. Here, we only compare the

chemical shifts on the ligand between AF-QM/MM calculation and large-sized system calculation. As shown in **Figure 3**, the  $^1\text{H}$  chemical shifts on the ligand calculated by the AF-QM/MM method (where the ligand is divided into three parts) are in good agreement with large-sized system calculation. The mean unsigned error (MUE) between AF-QM/MM results and chemical shifts from the large-sized system calculation is 0.046 ppm, and the RMSD between them is 0.051 ppm. The results demonstrate that the AF-QM/MM approach can accurately reproduce the large-sized system calculation. Furthermore, at the DFT level, the total computational cost was reduced by 36%, from 5,601 min (CPU time) by the large-sized system calculation to 3,585 min by dividing the ligand into 3 fragments. In addition, the 3 fragment-based QM/MM calculations were carried out in parallel. Therefore, the computational wall time could be further reduced by approximately 2/3.

Next, we compare the calculated chemical shifts for apo and holo NCS (31 protons in the binding pocket, as shown in **Figure 2**) with the experimental values. In this benchmark test, the AF-QM/MM results correlate well with the experiment (see **Figure 4**). For the bound complex, the MUE between the calculated and experimental chemical shifts is 0.44 ppm, and the RMSD is 0.57 ppm. For the unbound protein and ligand, the MUE and RMSD between calculated and experimental chemical shifts are 0.45 and 0.62 ppm, respectively.

The prediction of chemical shift perturbations (CSP) between apo and holo NCS could further validate the accuracy of the AF-QM/MM approach. Among the hydrogen atoms in the binding pocket, H15, H31 of the ligand, HD2 of Leu45 and HB2 of Cys37 are significantly influenced by the ring current effect, where they are close to the aromatic rings in the native holo structure (see **Figure 5**). As a result, large upfield shifts upon ligand binding were observed for those atoms. The experimental CSPs of those four atoms (namely, H15, H31, Leu45:HD2, and Cys37:HB2) are





**FIGURE 4** | Comparison of the AF-QM/MM results and experimental  $^1\text{H}$  chemical shifts (31 protons in the binding pocket, as shown in **Figure 2**) for both apo and holo NCS.

–0.93, –0.86, –1.15, and –0.72 ppm, respectively, while the AF-QM/MM results of them are –0.45, –0.34, –1.25, and –0.71 ppm. The results show that large chemical shift perturbations between apo and holo protein-ligand systems could be accurately predicted by the AF-QM/MM approach.

## Performance of the $\text{CS}_{\text{score}}$ Scoring Function

The Glide scores of the 38 docked decoys and the native holo NCS are shown in **Figure 6**. The energy based scoring function is capable of distinguishing the experimental structure from the docked poses whose structural RMSDs are larger than 4 Å with reference to the native pose. However, for the docked poses whose RMSDs are between 2 and 4 Å, the Glide scores of them are sometimes very close to the experimental structure (see Pose 7 in **Figure 6A**). In contrast, the  $\text{CS}_{\text{score}}$  is easier to discriminate the experimental structure from the decoy sets whose structural RMSDs are around 2 Å. This is mainly due to that chemical shifts are quite sensitive to the local chemical environment at the binding site. In  $\text{CS}_{\text{score}}$ , protons in ligand and the selected hydrogen atoms in protein residues serve as molecular probes to detect the binding environment. When the protons have different close contacts between the native and decoy structures, such as the interactions with aromatic rings or hydrogen bonding, the calculated chemical shift of certain protons may have substantial deviations between different binding modes. Therefore, the change of NMR chemical shifts of protons could clearly reflect the corresponding binding interactions between the protein and ligand.

The comparison of calculated chemical shifts between the native and docking poses could probe the structural changes of ligand binding poses among them. **Figure 7** shows that, for

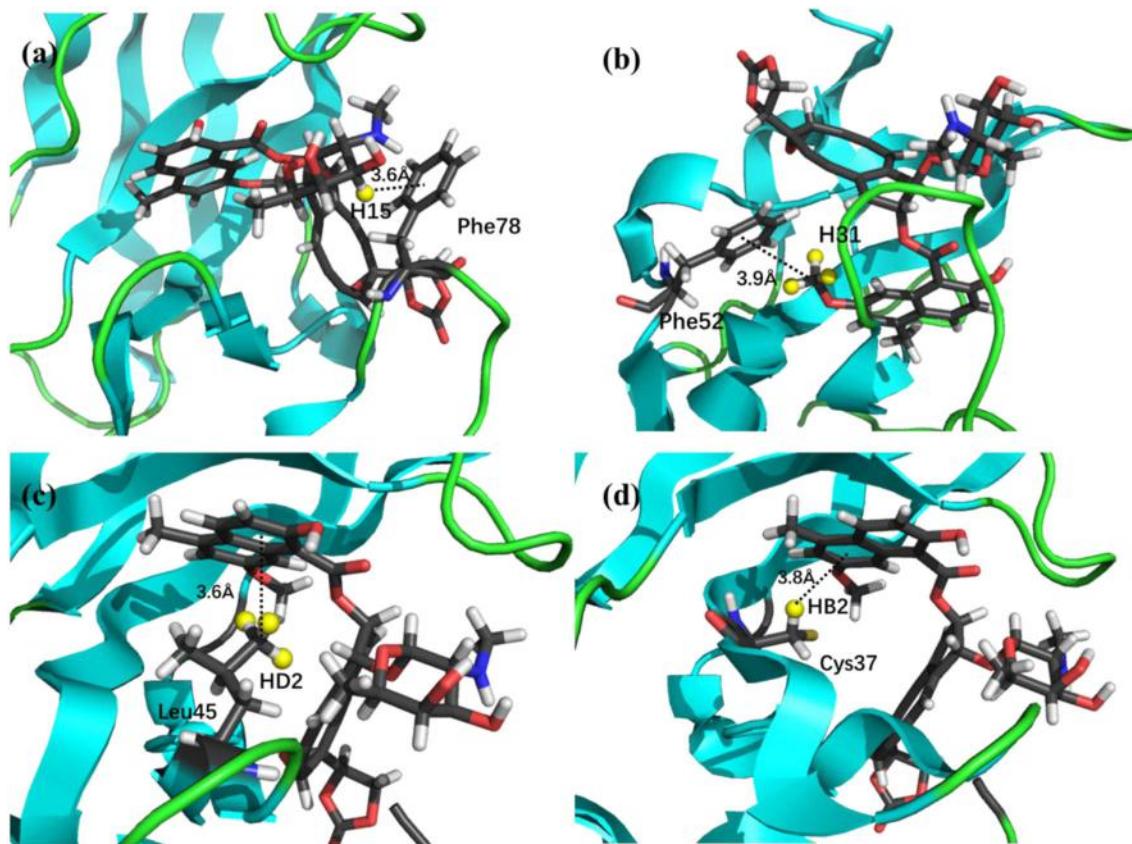
the positions of the naphthoate group and enediyne ring in Pose 7, the chromophore is very close to the native binding structure. However, the aminosugar group is pointing to a different direction, which results in the large chemical shifts deviation for H13 and H15 on the ligand (see **Table 1**).

The weakness of  $\text{CS}_{\text{score}}$  is that for decoys with larger structural RMSDs, the chemical shift based scoring function might be not as efficient as those with low structural RMSDs. In high structural RMSD range, some ligand poses might be close to the apo state (fewer interactions with the protein), and the  $\text{CS}_{\text{score}}$  score for the apo state of chromophore is 0.60. Therefore, the rankings of those poses are not very sensitive to the structural RMSDs using  $\text{CS}_{\text{score}}$ . The example cases for poses 10 and 18 will be discussed in Section Improvement of the hybrid  $\text{CS}_{\text{Gscore}}$  scoring function.

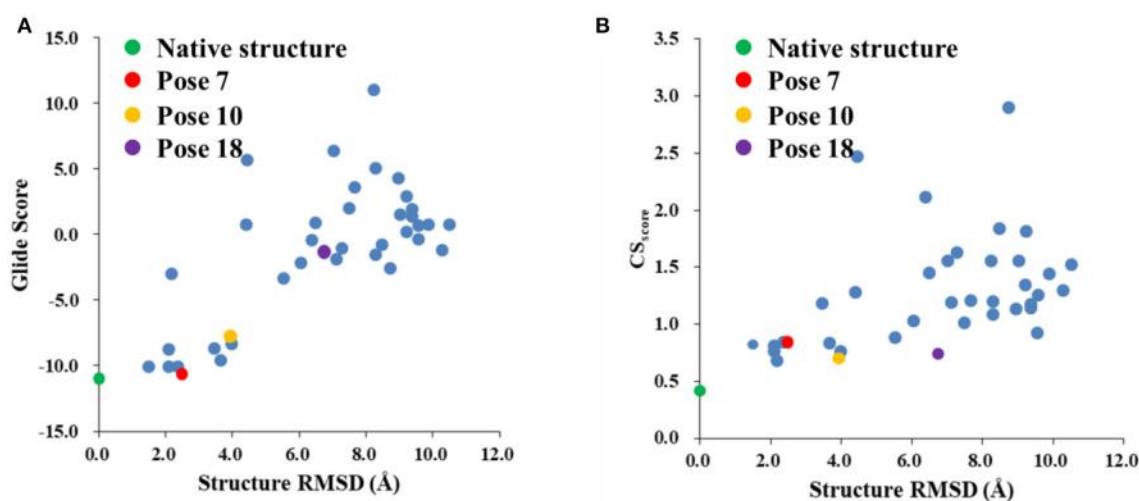
## Improvement of the Hybrid $\text{CS}_{\text{Gscore}}$ Scoring Function

**Figure 8** shows that the  $\text{CS}_{\text{Gscore}}$  score is capable of differentiating the experimental structure from the decoys for the protein-ligand complex. In this work, the weighting factor  $\alpha$  in Equation (2) was set to 0.1 to make the  $\text{CS}_{\text{score}}$  and Glide scores on the same scale. As shown in **Figure 8**, for poses whose structural RMSDs are around 2–4 Å, the  $\text{CS}_{\text{score}}$  score from NMR chemical shifts dominates the scoring function. Even though the Glide score for the decoy structures are close to the experimental structure (see **Figure 6B**), the  $\text{CS}_{\text{score}}$  could discriminate the experimental structure from the decoys, resulting that the  $\text{CS}_{\text{Gscore}}$  (combination of  $\text{CS}_{\text{score}}$  and Glide scores) ranked the native binding pose clearly as the most favorable structure. On the other hand, for decoy poses whose structural RMSDs are larger than 4 Å, the energy based scoring function (Glide score) has the major impact on the  $\text{CS}_{\text{Gscore}}$ , which makes the decoys with large structural RMSDs deviates more from the experimental structure as compared to the  $\text{CS}_{\text{score}}$  score (**Figure 6B**).

The pose 10 was previously scored low in  $\text{CS}_{\text{score}}$ , whose chemical shift RMSD is 0.70 ppm between the calculated and experimental data. The naphthoate group of the ligand was flipped in the docking structure, while locations of the other regions of the chromophore are similar to the native binding pose (see **Figure 9**). On the naphthoate group, H23 has the largest chemical shift deviation between the native binding structure and pose 10. In the native binding pose, H23 is beside the aromatic ring, which causes downfield chemical shift on H23. While in Pose 10, the H23 atom moves away from the indole ring of Trp39, resulting that its calculated chemical shift is much lower than that of the experimental structure (see **Table 2**). Furthermore, **Figure 9a** shows that in Pose 10, H31 moves away from the phenyl ring of Phe52, and the corresponding chemical shielding decreased, resulting in higher NMR chemical shift. Furthermore, because the aromatic ring position of the naphthoate group moved in the docking structure,  $^1\text{H}$  chemical shift on surrounding protein residues, such as Cys37:HB3, also changed significantly. Meanwhile, since the naphthoate ring strongly influenced the  $^1\text{H}$  chemical shifts from the unbound



**FIGURE 5** | X-ray structure of the holo NCS. The CSPs for protons of H15 (a) and H31 (b) in the chromophore, Leu45:HD2 (c) and Cys37:HB2 (d) in the NCS, are significantly influenced by the ring current effect.



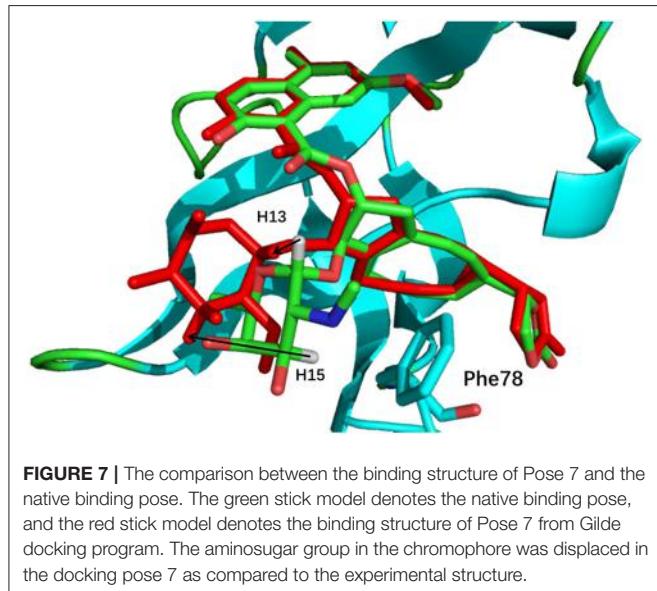
**FIGURE 6** | The Glide (A) and CSscore (B) scores with different structural RMSDs of the chromophore in the holo NCS. The green dot represents the score of the binding pose in the X-ray structure of the holo NCS. The red, yellow and violet dots denote the docking poses 7, 10 and 18, respectively.

to bound states, straying away of the naphthoate group in Pose 10 caused that the chemical shifts are less affected upon ligand binding, which results in slightly higher CS<sub>score</sub> score

(see **Figure 6B**, but other groups are close to the native state). Considering the physical non-bonded interactions between NCS and chromophore, as the naphthoate group stays deep inside the

binding pocket in the native state, which contributes most to the NCS-chromophore binding energy. Therefore, the structural deformation in pose 10 caused that the interaction energy between them became weaker, and the corresponding Glide score gave the lower rank. In the hybrid scoring function  $CS_{Gscore}$ , the ranking of pose 10 has a clear separation from the experimental structure, because it incorporates both the NMR chemical shift deviations from the experimental data and the physical interaction energy between the protein and ligand.

The docking pose 18 is also the case that the energy function is more important than the  $CS_{score}$  in ranking the docking poses. **Figure 10** shows that the ligand position of pose 18 almost translated to the direction away from the binding pocket. As the naphthoate ring moved away, the chemical shieldings of

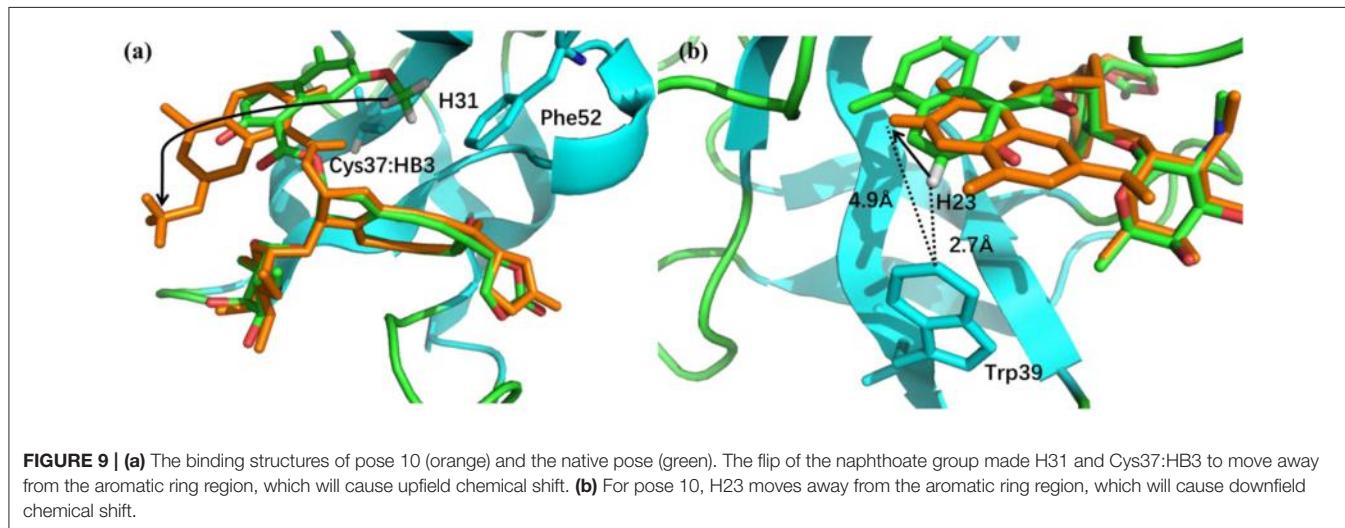
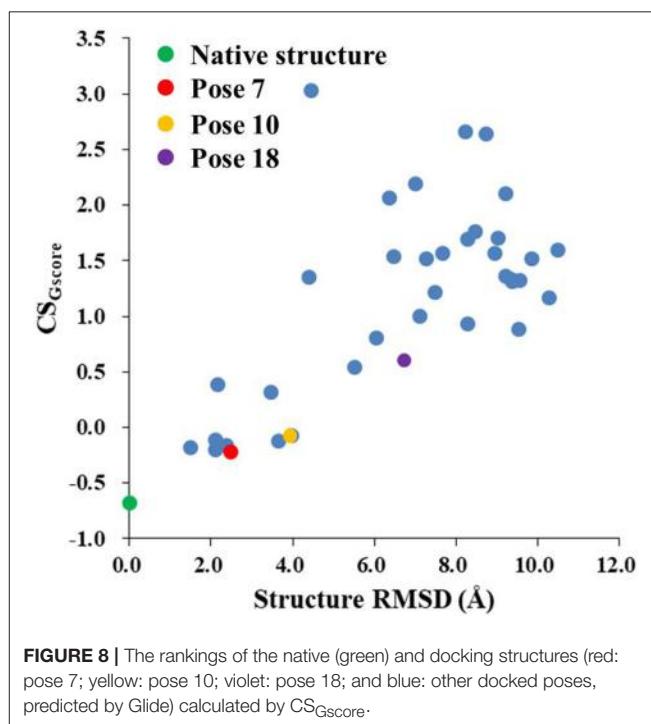


Cys37:HB2, Cys37:HB3 and Leu45:HD2 decreased as compared to the native state (see **Table 3**). The ligand pose 18 is closer

**TABLE 1** | The comparison between the experimental and calculated chemical shifts (in ppm).

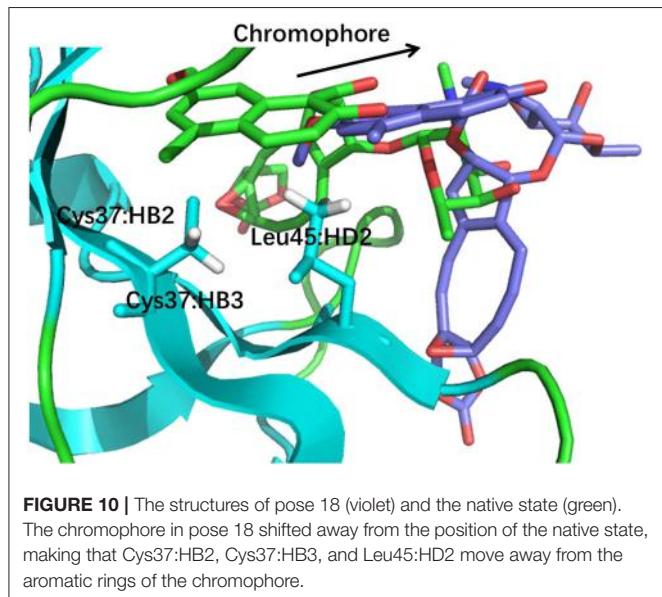
	Experiment	Native	Pose 7
LIG:H13	5.98	5.87	4.45
LIG:H15	3.42	3.12	4.31

"Native" and "Pose 7" denote the calculated chemical shifts for the native binding pose and the docking pose of Pose 7, respectively.



**TABLE 2** | Comparison between the experimental and calculated chemical shifts on the native binding model and pose 10 (in ppm).

Experiment	Native	Pose 10
LIG:H23	7.40	7.91
LIG:H31	2.96	3.58
Cys37:HB3	2.17	2.08
		3.19

**TABLE 3** | Comparison between the experimental and calculated chemical shifts on the native state and the docking pose 18 (in ppm).

Experiment	Native	Pose 18
Cys37:HB2	1.98	2.60
Cys37:HB3	2.17	2.08
Leu45:HD2	-0.90	-1.25
		0.69

to the apo state, which resulted in high  $CS_{score}$  score, but the interaction energy between the chromophore and NCS would be obviously weaker than that of the native state, and the Glide score for the pose 18 is substantially higher than the native binding pose (see **Figure 6A**). Therefore, in the hybrid scoring function  $CS_{Gscore}$ , the ranking of pose 18 is clearly lower than the native state, which correctly reflects that the rankings decrease as the structural RMSDs become larger.

## CONCLUSION

In this work, we applied the automated fragmentation method for QM/MM calculation of NMR chemical shifts for protein-ligand binding complexes. In the AF-QM/MM approach, the atomic NMR chemical shifts were obtained by dividing the protein automatically into residue-centric fragments. In order to

reduce the computational cost for the ligand, the chromophore that contains 81 atoms was also divided into three smaller fragments to make the QM size for ligand calculation comparable to the protein fragments. The AF-QM/MM approach with the implicit solvation treatment is computationally efficient and linear-scaling with a low pre-factor. Moreover, the approach is massively parallel and can be applied to routinely calculate the *ab initio* NMR chemical shifts for protein-ligand complexes of any size.

The  $^1H$  chemical shifts calculated by the AF-QM/MM approach at the DFT level are in good agreement with large-sized system calculation, where the entire ligand and its buffer region are treated by QM, and the remaining atoms of the protein are described by background charges. The MUE between AF-QM/MM and large-sized system calculation is 0.046 ppm. Furthermore, the MUEs between calculated and experimental  $^1H$  chemical shifts in the binding pocket of apo and holo NCS are 0.45 and 0.44 ppm, respectively. Our results demonstrate that the AF-QM/MM approach is capable of reproducing the large-sized system *ab initio* calculations of NMR chemical shifts for protein-ligand complexes, and the calculated chemical shifts are in good agreement with the experimental results.

The results of  $CS_{score}$  scores show that chemical shifts could be utilized as molecular probes to detect the binding conformation of the protein-ligand complex. The experimental structure has the clear leading score as compared to the decoy binding poses. By investigating the CSP patterns of decoy structures, the position changes of the ligand could be detected by variations of chemical shifts in different local chemical environment.

In this study, we further proposed the hybrid scoring function  $CS_{Gscore}$  which combines  $CS_{score}$  and the energy-based scoring function of Glide score. The hybrid  $CS_{Gscore}$  scoring function can help to distinguish the native ligand structure from the decoy docking poses.  $CS_{Gscore}$  can also clearly separate the scores of decoy structures, which have significantly large structural RMSD values and give relatively low  $CS_{score}$  scores, from the native docking pose. The  $CS_{Gscore}$  incorporates both the experimental NMR chemical shift information and the energy-based scoring method, which could better determine the binding site structure of the protein-ligand complex. Therefore, the AF-QM/MM approach provides an accurate and efficient platform for protein-ligand binding structure prediction based on NMR derived information.

## AUTHOR CONTRIBUTIONS

XH designed research; XJ, TZ, and XH performed research; XJ, TZ, JZ, and XH analyzed data; and XJ and XH wrote the paper.

## ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (Grant No. 2016YFA0501700), National Natural Science Foundation of China (Nos. 21673074, 21761132022 and 21433004), Youth Top-Notch Talent Support Program of

Shanghai, NYU-ECNU Center for Computational Chemistry at NYU Shanghai, and Shanghai Putuo District (Grant 2014-A-02). We thank the Supercomputer Center of East China Normal University for providing us with computational time.

## REFERENCES

- Adeniyi, A. A., and Soliman, M. E. S. (2017). Implementing QM in docking calculations: is it a waste of computational time? *Drug Discov. Today* 22, 1216–1223. doi: 10.1016/j.drudis.2017.06.012
- Aguirre, C., ten Brink, T., Cala, O., Guichou, J.-F., and Krimm, I. (2014). Protein-ligand structure guided by backbone and side-chain proton chemical shift perturbations. *J. Biomol. NMR* 60, 147–156. doi: 10.1007/s10858-014-9864-9
- Bratholm, L. A., and Jensen, J. H. (2017). Protein structure refinement using a quantum mechanics-based chemical shielding predictor. *Chem. Sci.* 8, 2061–2072. doi: 10.1039/C6SC04344E
- Case, D. A. (1998). The use of chemical shifts and their anisotropies in biomolecular structure determination. *Curr. Opin. Struct. Biol.* 8, 624–630. doi: 10.1016/S0959-440X(98)80155-3
- Case, D., Darden, T., Cheatham, T. E. III., Simmerling, C., Wang, J., Duke, R., et al. (2012). *AMBER 12*. San Francisco, CA: University of California.
- Cavalli, A., Salvatella, X., Dobson, C. M., and Vendruscolo, M. (2007). Protein structure determination from NMR chemical shifts. *Proc. Nat. Acad. Sci. U.S.A.* 104, 9615–9620. doi: 10.1073/pnas.0610313104
- Christensen, A. S., Linnet, T. E., Borg, M., Boomsma, W., Lindorff-Larsen, K., Hamelryck, T., et al. (2014). Protein structure validation and refinement using amide proton chemical shifts derived from quantum mechanics. *PLoS ONE* 8:e84123. doi: 10.1371/journal.pone.0084123
- Cioffi, M., Hunter, C. A., Packer, M. J., Pandya, M. J., and Williamson, M. P. (2008). Use of quantitative <sup>1</sup>H NMR chemical shift changes for ligand docking into barnase. *J. Biomol. NMR* 43, 11–19. doi: 10.1007/s10858-008-9286-7
- Cui, Q., and Karplus, M. (2000). Molecular properties from combined QM/MM methods. 2. chemical shifts in large molecules. *J. Phys. Chem. B* 104, 3721–3743. doi: 10.1021/jp994154g
- Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., and Mee, R. P. (1997). Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided Mol. Des.* 11, 425–445.
- Englebienne, P., and Moitessier, N. (2009). Docking ligands into flexible and solvated macromolecules. 5. Force-field-based prediction of binding affinities of ligands to proteins. *J. Chem. Inf. Model.* 49, 2564–2571. doi: 10.1021/ci900251k
- Ferreira, L., dos Santos, R., Oliva, G., and Andricopulo, A. (2015). Molecular docking and structure-based drug design strategies. *Molecules* 20, 13384–13421. doi: 10.3390/molecules200713384
- Flaig, D., Beer, M., and Ochsenfeld, C. (2012). Convergence of electronic structure with the size of the QM region: example of QM/MM NMR shieldings. *J. Chem. Theor. Comput.* 8, 2260–2271. doi: 10.1021/ct300036s
- Frank, A., Möller, H. M., and Exner, T. E. (2012). Toward the quantum chemical calculation of NMR chemical shifts of proteins. 2. Level of theory, basis set, and solvents model dependence. *J. Chem. Theor. Comput.* 8, 1480–1492. doi: 10.1021/ct200913r
- Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., et al. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *J. Med. Chem.* 47, 1739–1749. doi: 10.1021/jm0306430
- Frisch, M., Trucks, G., Schlegel, H. B., Scuseria, G., Robb, M., Cheeseman, J., et al. (2009). *Gaussian 09*. Wallingford, CT: Gaussian Inc.
- Gao, Q., Yokojima, S., Fedorov, D. G., Kitaura, K., Sakurai, M., and Nakamura, S. (2010). Fragment-molecular-orbital-method-based ab initio NMR chemical-shift calculations for large molecular systems. *J. Chem. Theor. Comput.* 6, 1428–1444. doi: 10.1021/ct100006n
- Gao, Q., Yokojima, S., Kohno, T., Ishida, T., Fedorov, D. G., Kitaura, K., et al. (2007). Ab initio NMR chemical shift calculations on proteins using fragment molecular orbitals with electrostatic environment. *Chem. Phys. Lett.* 445, 331–339. doi: 10.1016/j.cplett.2007.07.103
- Garay, P. G., Martin, O. A., Scheraga, H. A., and Vila, J. A. (2014). Factors affecting the computation of the <sup>13</sup>C shielding in disaccharides. *J. Comput. Chem.* 35, 1854–1864. doi: 10.1002/jcc.23697
- Gohlke, H., Hendlich, M., and Klebe, G. (2000). Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* 295, 337–356. doi: 10.1006/jmbi.1999.3371
- Grinter, S., and Zou, X. (2014). Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design. *Molecules* 19, 10150–10176. doi: 10.3390/molecules190710150
- Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., et al. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. enrichment factors in database screening. *J. Med. Chem.* 47, 1750–1759. doi: 10.1021/jm030644s
- Han, B., Liu, Y., Ginzinger, S. W., and Wishart, D. S. (2011). SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* 50, 43–57. doi: 10.1007/s10858-011-9478-4
- Hartman, J. D., and Beran, G. J. O. (2014). Fragment-based electronic structure approach for computing nuclear magnetic resonance chemical shifts in molecular crystals. *J. Chem. Theor. Comput.* 10, 4862–4872. doi: 10.1021/ct500749h
- He, X., Wang, B., and Merz, K. M. Jr. (2009). Protein NMR chemical shift calculations based on the automated fragmentation QM/MM approach. *J. Phys. Chem. B* 113, 10380–10388. doi: 10.1021/jp901992p
- He, X., Zhu, T., Wang, X., Liu, J., and Zhang, J. Z. (2014). Fragment quantum mechanical calculation of proteins and its applications. *Acc. Chem. Res.* 47, 2748–2757. doi: 10.1021/ar500077t
- Huang, S.-Y., and Zou, X. (2006). An iterative knowledge-based scoring function to predict protein-ligand interactions: I. derivation of interaction potentials. *J. Comput. Chem.* 27, 1866–1875. doi: 10.1002/jcc.20504
- Jakalian, A., Jack, D. B., and Bayly, C. I. (2002). Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. parameterization and validation. *J. Comput. Chem.* 23, 1623–1641. doi: 10.1002/jcc.10128
- Jin, X., Zhu, T., Zhang, J. Z., and He, X. (2016). A systematic study on RNA NMR chemical shift calculation based on the automated fragmentation QM/MM approach. *RSC Adv.* 6, 108590–108602. doi: 10.1039/C6RA22518G
- Jorgensen, W. L., and Jenson, C. (1998). Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: seeking temperatures of maximum density. *J. Comput. Chem.* 19, 1179–1186. doi: 10.1002/(SICI)1096-987X(19980730)19:103.0.CO;2-J
- Lodewyk, M. W., Siebert, M. R., and Tantillo, D. J. (2012). Computational prediction of <sup>1</sup>H and <sup>13</sup>C chemical shifts: a useful tool for natural product, mechanistic, and synthetic organic chemistry. *Chem. Rev.* 112, 1839–1862. doi: 10.1021/cr200106v
- McCoy, M. A., and Wyss, D. F. (2002). Spatial localization of ligand binding sites from electron current density surfaces calculated from NMR chemical shift perturbations. *J. Am. Chem. Soc.* 124, 11758–11763. doi: 10.1021/ja026166c
- Medek, A., Hajduk, P. J., Mack, J., and Fesik, S. W. (2000). The use of differential chemical shifts for determining the binding site location and orientation of protein-bound ligands. *J. Am. Chem. Soc.* 122, 1241–1242. doi: 10.1021/ja993921m
- Meiler, J. (2003). PROSHIFT: protein chemical shift prediction using artificial neural networks. *J. Biomol. NMR* 26, 25–37. doi: 10.1023/A:1023060720156
- Meiler, J., and Baker, D. (2003). Rapid protein fold determination using unassigned NMR data. *Proc. Nat. Acad. Sci. U.S.A.* 100, 15404–15409. doi: 10.1073/pnas.2434121100
- Merz, K. M. (2014). Using quantum mechanical approaches to study biological systems. *Acc. Chem. Res.* 47, 2804–2811. doi: 10.1021/ar5001023

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2018.00150/full#supplementary-material>

- Mohan, V., Gibbs, A. C., Cummings, M. D., Jaeger, E. P., and DesJarlais, R. L. (2005). Docking: successes and challenges. *Curr. Drug Metab.* 11, 323–333. doi: 10.2174/1381612053382106
- Mohanty, S., Sieker, L. C., and Drobny, G. P. (1994). Sequential  $^1\text{H}$  NMR assignment of the complex of aponeocarzinostatin with ethidium bromide and investigation of protein-drug interactions in the chromophore binding site. *Biochemistry* 33, 10579–10590. doi: 10.1021/bi00201a003
- Moon, S., and Case, D. A. (2007). A new model for chemical shifts of amide hydrogens in proteins. *J. Biomol. NMR* 38, 139–150. doi: 10.1007/s10858-007-9156-8
- Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., et al. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 19, 1639–1662. doi: 10.1002/(SICI)1096-987X(19981115)19:143.0.CO;2-B
- Murray, C. W., Auton, T. R., and Eldridge, M. D. (1998). Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand-receptor binding affinities and the use of bayesian regression to improve the quality of the model. *J. Comput. Aided Mol. Des.* 12, 503–519.
- Myers, A. G., Proteau, P. J., and Handel, T. M. (1988). Stereochemical assignment of neocarzinostatin chromophore. Structures of neocarzinostatin chromophore-methyl thioglycolate adducts. *J. Am. Chem. Soc.* 110, 7212–7214.
- Neal, S., Nip, A. M., Zhang, H., and Wishart, D. S. (2003). Rapid and accurate calculation of protein  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts. *J. Biomol. NMR* 26, 215–240. doi: 10.1023/A:1023812930288
- Riedinger, C., Endicott, J. A., Kemp, S. J., Smyth, L. A., Watson, A., Valeur, E., et al. (2008). Analysis of chemical shift changes reveals the binding modes of isoindolinone inhibitors of the MDM2-p53 interaction. *J. Am. Chem. Soc.* 130, 16038–16044. doi: 10.1021/ja8062088
- Robustelli, P., Kohlhoff, K., Cavalli, A., and Vendruscolo, M. (2010). Using NMR chemical shifts as structural restraints in molecular dynamics simulations of proteins. *Structure* 18, 923–933. doi: 10.1016/j.str.2010.04.016
- Rocchia, W., Alexov, E., and Honig, B. (2001). Extending the applicability of the nonlinear Poisson-Boltzmann equation: multiple dielectric constants and multivalent ions. *J. Phys. Chem. B* 105, 6507–6514. doi: 10.1021/jp010454y
- Schaus, S. E., Cavalieri, D., and Myers, A. G. (2001). Gene transcription analysis of *Saccharomyces cerevisiae* exposed to neocarzinostatin protein–chromophore complex reveals evidence of DNA damage, a potential mechanism of resistance, and consequences of prolonged exposure. *Proc. Nat. Acad. Sci. U.S.A.* 98, 11075–11080. doi: 10.1073/pnas.191340698
- Shen, Y., and Bax, A. (2010). SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR* 48, 13–22. doi: 10.1007/s10858-010-9433-9
- Shen, Y., and Bax, A. (2015). Homology modeling of larger proteins guided by chemical shifts. *Nat. Methods* 12, 747–750. doi: 10.1038/nmeth.3437
- Śledź, P., and Caflisch, A. (2018). Protein structure-based drug design: from docking to molecular dynamics. *Curr. Opin. Struct. Biol.* 48, 93–102. doi: 10.1016/j.sbi.2017.10.010
- Sliwoski, G., Kothiwale, S., Meiler, J., and Lowe, E. W. (2014). Computational methods in drug discovery. *Pharmacol. Rev.* 66, 334–395. doi: 10.1124/pr.112.007336
- Swails, J., Zhu, T., He, X., and Case, D. A. (2015). AFNMR: automated fragmentation quantum mechanical calculation of NMR chemical shifts for biomolecules. *J. Biomol. NMR* 63, 125–139. doi: 10.1007/s10858-015-9970-3
- Takashima, H., Yoshida, T., Ishino, T., Hasuda, K., Ohkubo, T., and Kobayashi, Y. (2005). Solution NMR structure investigation for releasing mechanism of neocarzinostatin chromophore from the holoprotein. *J. Biol. Chem.* 280, 11340–11346. doi: 10.1074/jbc.M411579200
- Tan, H.-J., and Bettens, R. P. (2013). Ab initio NMR chemical-shift calculations based on the combined fragmentation method. *Phys. Chem. Chem. Phys.* 15, 7541–7547. doi: 10.1039/c3cp50406a
- Ten Brink, T., Aguirre, C., Exner, T. E., and Krimm, I. (2015). Performance of protein-ligand docking with simulated chemical shift perturbations. *J. Chem. Inf. Model.* 55, 275–283. doi: 10.1021/ci500446s
- Victoria, A., Möller, H. M., and Exner, T. E. (2014). Accurate ab initio prediction of NMR chemical shifts of nucleic acids and nucleic acids/protein complexes. *Nucleic Acids Res.* 42, e173–e173. doi: 10.1093/nar/gku1006
- Vila, J. A., Arnaudova, Y. A., Martin, O. A., and Scheraga, H. A. (2009). Quantum-mechanics-derived  $^{13}\text{C}^\alpha$  chemical shift server (CheShift) for protein structure validation. *Proc. Nat. Acad. Sci. U.S.A.* 106, 16972–16977. doi: 10.1073/pnas.0908833106
- Wang, B., and Merz, K. M. Jr. (2010). Importance of loop dynamics in the neocarzinostatin chromophore binding and release mechanisms. *Phys. Chem. Chem. Phys.* 12, 3443–3449. doi: 10.1039/b924951f
- Wang, B., Westerhoff, L. M., and Merz, K. M. (2007). A critical assessment of the performance of protein-ligand scoring functions based on NMR chemical shift perturbations. *J. Med. Chem.* 50, 5128–5134. doi: 10.1021/jm070484a
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157–1174. doi: 10.1002/jcc.20035
- Williamson, M. P., and Craven, C. J. (2009). Automated protein structure calculation from NMR data. *J. Biomol. NMR* 43, 131–143. doi: 10.1007/s10858-008-9295-6
- Xu, X.-P., and Case, D. A. (2001). Automated prediction of  $^{15}\text{N}$ ,  $^{13}\text{C}^\alpha$ ,  $^{13}\text{C}^\beta$  and  $^{13}\text{C}^\gamma$  chemical shifts in proteins using a density functional database. *J. Biomol. NMR* 21, 321–333. doi: 10.1023/A:1013324104681
- Yu, Z., Li, P., and Merz, K. M. (2017). Using ligand-induced protein chemical shift perturbations to determine protein-ligand structures. *Biochemistry* 56, 2349–2362. doi: 10.1021/acs.biochem.7b00170
- Zhu, T., He, X., and Zhang, J. Z. (2012). Fragment density functional theory calculation of NMR chemical shifts for proteins with implicit solvation. *Phys. Chem. Chem. Phys.* 14, 7837–7845. doi: 10.1039/C2CP23746F
- Zhu, T., Zhang, J. Z. H., and He, X. (2013). Automated fragmentation QM/MM calculation of amide proton chemical shifts in proteins with explicit solvent model. *J. Chem. Theor. Comput.* 9, 2104–2114. doi: 10.1021/ct300999w
- Zhu, T., Zhang, J. Z. H., and He, X. (2014). Correction of erroneously packed protein's side chains in the NMR structure based on ab initio chemical shift calculations. *Phys. Chem. Chem. Phys.* 16, 18163–18169. doi: 10.1039/C4CP02553A
- Zhu, T., Zhang, J. Z. H., and He, X. (2015). “Quantum calculation of protein NMR chemical shifts based on the automated fragmentation method,” in *Advance in Structural Bioinformatics*, eds D. Wei, Q. Xu, T. Zhao and H. Dai (Dordrecht: Springer), 827, 49–70.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Jin, Zhu, Zhang and He. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# A Force Balanced Fragmentation Method for *ab Initio* Molecular Dynamic Simulation of Protein

Mingyuan Xu<sup>1</sup>, Tong Zhu<sup>1,2\*</sup> and John Z. H. Zhang<sup>1,2,3,4\*</sup>

<sup>1</sup> State Key Lab of Precision Spectroscopy, Shanghai Engineering Research Center of Molecular Therapeutics & New Drug Development, Shanghai Key Laboratory of Green Chemistry & Chemical Process, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai, China, <sup>2</sup> NYU-ECNU Center for Computational Chemistry at New York University Shanghai, Shanghai, China, <sup>3</sup> Department of Chemistry, New York University, New York, NY, United States, <sup>4</sup> Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan, China

## OPEN ACCESS

### Edited by:

Thomas S. Hofer,  
University of Innsbruck, Austria

### Reviewed by:

Antonio Monari,  
Université de Lorraine, France  
Hans Martin Senn,  
University of Glasgow,  
United Kingdom

### \*Correspondence:

Tong Zhu  
tzhu@lps.ecnu.edu.cn  
John Z. H. Zhang  
john.zhang@nyu.edu

### Specialty section:

This article was submitted to  
Theoretical and Computational  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

Received: 06 March 2018

Accepted: 09 May 2018

Published: 30 May 2018

### Citation:

Xu M, Zhu T and Zhang JZH (2018) A Force Balanced Fragmentation Method for *ab Initio* Molecular Dynamic Simulation of Protein. *Front. Chem.* 6:189.  
doi: 10.3389/fchem.2018.00189

A force balanced generalized molecular fractionation with conjugate caps (FB-GMFCC) method is proposed for *ab initio* molecular dynamic simulation of proteins. In this approach, the energy of the protein is computed by a linear combination of the QM energies of individual residues and molecular fragments that account for the two-body interaction of hydrogen bond between backbone peptides. The atomic forces on the capped H atoms were corrected to conserve the total force of the protein. Using this approach, *ab initio* molecular dynamic simulation of an Ace-(ALA)<sub>9</sub>-NME linear peptide showed the conservation of the total energy of the system throughout the simulation. Further a more robust 110 ps *ab initio* molecular dynamic simulation was performed for a protein with 56 residues and 862 atoms in explicit water. Compared with the classical force field, the *ab initio* molecular dynamic simulations gave better description of the geometry of peptide bonds. Although further development is still needed, the current approach is highly efficient, trivially parallel, and can be applied to *ab initio* molecular dynamic simulation study of large proteins.

**Keywords:** quantum fragment method, *ab initio* molecular dynamics, force balanced, GB3, protein dynamics, MFCC

## INTRODUCTION

Molecular dynamic (MD) simulation plays an increasingly important role in the study of structural and dynamical properties of biomolecules at the atomic level (Karplus and Petsko, 1990; Cheatham and Kollman, 2000; Karplus and McCammon, 2002). With the ever-increasing power of computer hardware and the development of enhanced sampling methods, a significant advance in MD simulations with larger systems and longer simulation time have been achieved over the past decades (Shaw et al., 2010; Prinz et al., 2011). However, the accuracy and reliability of MD results are highly dependent on the accuracy of the force field employed in the simulation (Weiner et al., 1984, 1986; Ponder and Case, 2003). Despite widely successful applications of the current force fields in bio-molecular simulations, these simplified, pre-defined pairwise force fields have serious drawbacks. The most widely known deficiency is that the atomic charges in most of these force fields are pre-fixed, and there is no explicit treatment of electrostatic polarization and charge transfer (Duan et al., 2010; Tong et al., 2010; Ji and Mei, 2014). In the past few decades, significant efforts have been devoted to the development of polarizable force fields. However, although great achievements have been made,

the accuracy of these polarizable force fields still have a lot of room for improvement. In addition, many force fields have a bias toward the secondary structure of the protein. For example, the  $\alpha$ -helical propensity of the AMBER03 force field is too high relative to experimental measurements, while that of AMBER99SB is arguably too low (Best et al., 2008).

Compared with classical force fields, QM calculation can provide much more accurate potential energy function for the studied system, and include all important quantum effects. The advantage or need of the so-called *ab initio* molecular dynamic (AIMD) simulation over classical force fields in the study of proteins have been reported by various researchers (Wei et al., 2001; Dal Peraro et al., 2005; Ufimtsev and Martinez, 2009; Ufimtsev et al., 2011; Isborn et al., 2013). In these AIMD calculations, the atomic forces of the studied protein were calculated by QM methods, normally on the DFT level, whereas the motion of the nuclei was handled by classical mechanics. However, QM calculation needs a large amount of computational cost, which means that it can only be used for proteins with relatively small size.

So far, considerable efforts have been made to extend the applicability of QM calculation to large systems. Among existing approaches, the fragment-based QM methods has attracted much attention (Gordon et al., 2012; Collins et al., 2014; Li et al., 2014; Pruitt et al., 2014; Ramabhadran and Raghavachari, 2014; Chung et al., 2015; Collins and Bettens, 2015; Raghavachari and Saha, 2015). These approaches based on the chemical locality of molecular system, which assumes that the local regions of a molecular system can only be influenced weakly by atoms that are far away from it (Xu et al., 1998; Fedorov et al., 2014; Gao et al., 2014; He et al., 2014; Pruitt et al., 2014). In this kind of methods, the studied system is divided into small subsystems (fragments); the properties of these fragments such as energy are calculated separately by QM method. Then the property of the whole system can be obtained by taking a proper combination of the properties of these individual fragments. The fragment-based QM method is attractive in several aspects, such as easy implementation of parallelization without extensively modifying the existing QM programs and can be combined with all levels of *ab initio* electronic structure theories. In our previous study (Liu et al., 2015), a fragment based approach is presented for AIMD simulation of protein. In this approach, the potential energy and atomic forces of the studied protein are calculated by a recently developed electrostatically embedded generalized molecular fractionation with conjugate caps (EE-GMFCC) method (Wang et al., 2013). This AIMD approach had been applied to MD simulation of a small benchmark protein Trpcage in both gas phase and in solution. Compared with AMBER force field, this method can give more stable protein structure in simulation, and capture quantum effects that are missing in standard classical MD simulations.

To further improve the accuracy and efficiency of the AIMD simulations for protein, in this work, we presented a force balanced generalized molecular fractionation with conjugate caps (FB-GMFCC) method and checked its performance in the AIMD simulations for several systems. The paper is organized as follows. The next section provides a description of the FB-GMFCC

approach. In section Result and Discussion, we performed AIMD simulations on two selected proteins to validate the new method, and finally, a brief summary will be given in section Conclusion.

## THEORY AND METHOD

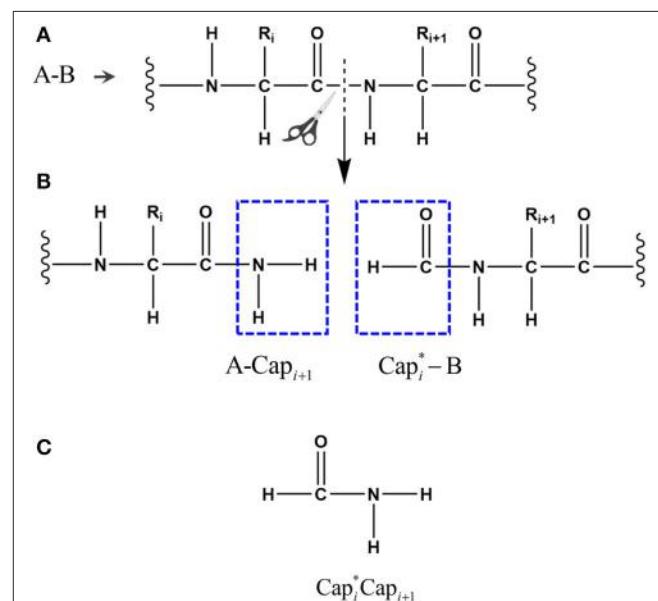
The FB-GMFCC method was developed based on the framework of molecular fractionation with conjugate caps (MFCC) approach (Zhang and Zhang, 2003). The computation procedure of FB-GMFCC can be roughly divided into two steps. Firstly, the given protein is cut into capped molecular fragments, including individual residues and residues that form backbone hydrogen bonds. Then the energy and atomic forces of each fragment are calculated by QM methods separately. Secondly, the AMOEBA polarizable force field (Ren and Ponder, 2002; Ponder et al., 2010; Ren et al., 2011; Wu et al., 2012) was employed to describe the long-range non-bonded interactions. Thus, the total energy of the protein system is obtained by a summation of quantum and classical components,

$$E_{FB-GMFCC} = E_{QM} + E_{MM} \quad (1)$$

Computational details of these energy components are described below.

### Calculation of $E_{QM}$

To calculate the energy  $E_{QM}$ , a given protein with  $N$  amino acids (defined as  $A_1A_2A_3 \dots A_N$ ) is decomposed into  $N$  individual fragments by cutting through the peptide bonds (Figure 1). At every cut point, a pair of molecular caps were designed to



**FIGURE 1** | The peptide bond is cut in the upper panel (A) and the fragments are capped with  $\text{Cap}_{i+1}$  and its conjugate  $\text{Cap}_i^*$  in the middle panel (B), where  $i$  represents the index of  $i$ th amino acid in the given protein. The atomic structure of the conjugate caps is shown in the bottom panel (C), defined as the fused molecular species.

saturate each fragment in order to preserve the local chemical environment. To minimize the computational cost, we simply use the amine and formyl group from the peptide bond as molecular caps, which are conjugate to each other (by forming a peptide bond) and are much smaller than that used in the EE-GMFCC approach. To avoid dangling bond, hydrogen atoms were added to terminate the molecular cap, the position of these extra H atoms are determined from the coordinates of the corresponding  $C_{\alpha}$  atoms.

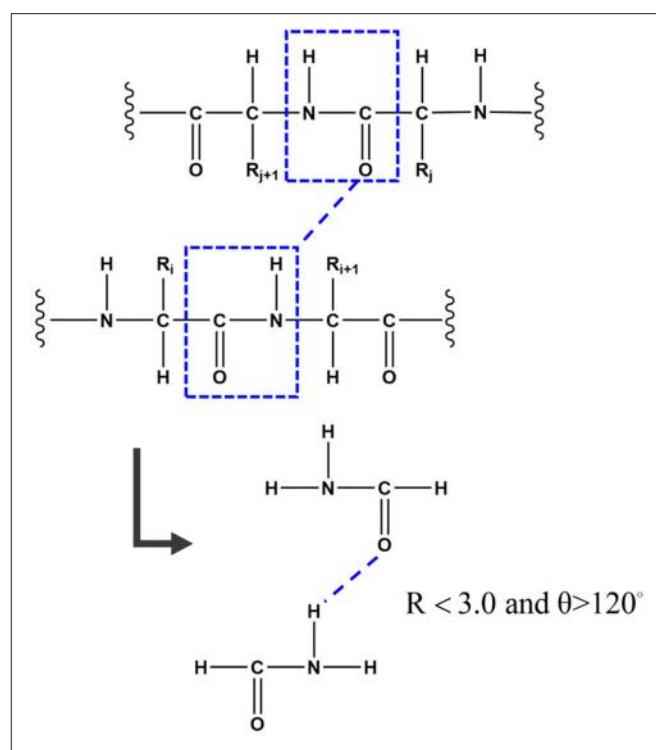
Hydrogen bond is one of the most important structural elements of protein and the dominant factor that stabilizes the protein secondary structures. Many previous works demonstrated that the strength of hydrogen bond from simulations under non-polarizable force fields is underestimated due to the lack of polarization effect (Ji et al., 2008; Gao et al., 2012). In the FB-GMFCC method, the backbone hydrogen bond was considered by two-body QM calculation. To reduce computational cost, only the H-saturated peptide bond which contains the donor or the accepter (which actually is a formamide as shown in Figure 2) was kept in the two-body QM calculation. To reduce computational cost, the position of the extra H atoms are also determined from the coordinates of the corresponding  $C_{\alpha}$  atoms. If the distance between donor H atom and acceptor O atom is  $<3.0 \text{ \AA}$  and the angle  $\theta$  of  $N-H_N-O$  is larger than  $120^\circ$ , the 2-body correction will be considered. Thus,  $E_{\text{QM}}$  can be expressed by the following formula:

$$\begin{aligned} E_{\text{QM}} &= E_{\text{fragment}} - E_{\text{concap}} + E_{\text{two-body}} \\ &= \sum_{i=2}^{N-1} E(\text{Cap}_{i-1}^* A_i \text{Cap}_{i+1}) - \sum_{i=2}^{N-2} E(\text{Cap}_i^* \text{Cap}_{i+1}) \\ &\quad + \sum_{\substack{i,j > i+2 \\ |R_{H_N-O}| \leq \lambda \\ \angle_{N-H_N-O} \geq \theta}} (E(A_i^P A_j^P) - E(A_i^P) - E(A_j^P)) \end{aligned} \quad (2)$$

Where the  $i$  and  $j$  represent the index of  $i$ th and  $j$ th residues, respectively. If the formyl or amide group of residue A is included in a backbone hydrogen bond,  $A^P$  represents the H-saturated peptide bond which contains this group. The first term  $E(\text{Cap}_{i-1}^* A_i \text{Cap}_{i+1})$  in Equation (2) represents the self-energy of fragment  $i$  (the  $i$ th residue  $A_i$  capped with a left cap  $\text{Cap}_{i-1}^*$  and a right cap  $\text{Cap}_{i+1}$ ). And it is clear that the self-energy of conjugate caps  $E(\text{Cap}_i^* \text{Cap}_{i+1})$  are double counted in first term of Equation (2) and it should be deducted.

## Calculation of $E_{\text{MM}}$

The  $E_{\text{QM}}$  term includes the self-energy of individual residue and the two-body correction of the interaction energy between residues that form backbone hydrogen bonds. To obtain the total energy expression for proteins, the classical force field was introduced to represent the long-range non-bonded interactions. In our previous study, we found that the electrostatic polarization arising from the environment also plays a critical role for including the many body effect in fragmentation methods (Wang et al., 2013). To describe the electrostatic polarization effect



**FIGURE 2 |** To better describe the backbone hydrogen bond, if the distance between the donor H atom and the acceptor O atom is  $<3.0 \text{ \AA}$  and the angle of  $N-H_N-O$  is larger than  $120^\circ$ , the 2-Body correction will be considered. To reduce computational cost, only the H-saturated peptide bond which contains the donor or the accepter was kept in the two-body QM calculation.

efficiently, we employed the polarizable atomic multipole-based AMOEBA force field (Ren and Ponder, 2002; Ponder et al., 2010; Ren et al., 2011; Wu et al., 2012). The expression of  $E_{\text{MM}}$  is as the following:

$$E_{\text{MM}} = \sum_{\substack{i,j \notin \text{same} \\ \text{QM zone}}} E_{\text{ele}}^{\text{perm}}(i,j) + E_{\text{ele}}^{\text{ind}}(i,j) + E_{\text{vdW}}(i,j) \quad (3)$$

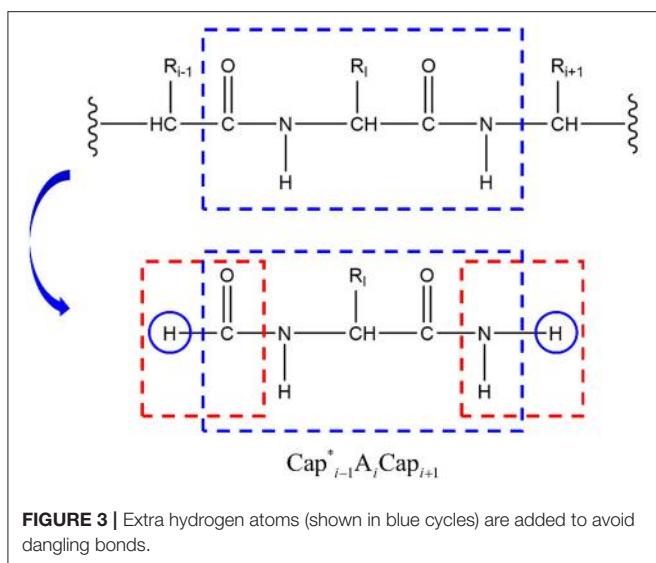
Details about calculations of Van der Waals interactions, permanent and induced electrostatic energies of the AMOEBA force field could be found in Refs (Ren and Ponder, 2002; Ponder et al., 2010; Ren et al., 2011; Wu et al., 2012). For any two atoms that have not been calculated in the same QM zone, these non-neighboring interactions between them should be added to the total energy expression.

## Balance the Force

To obtain atomic forces, we need to compute the derivative of FB-MFCC with respect to nuclear coordinates. For a given atom  $m$ , the atomic force can be expressed as following:

$$\mathbf{F}_m = -\nabla_m E_{\text{FB-GMFCC}} = -\nabla_m E_{\text{QM}} - \nabla_m E_{\text{MM}} \quad (4)$$

It should be noted, however, that we employed extra hydrogen atoms to avoid dangling bonds (Figure 3) and their coordinates



**FIGURE 3** | Extra hydrogen atoms (shown in blue cycles) are added to avoid dangling bonds.

were determined from those of the corresponding  $C_{\alpha}$  atoms. Because the forces on these extra hydrogen atoms in capped fragments  $E(Cap_{i-1}^* A_i Cap_{i+1})$  cannot be canceled exactly by subtracting those in the caps  $E(Cap_i^* Cap_{i+1})$ , it will not exactly conserve the energy. In order to fix this problem, we balance the forces on the corresponding  $C_{\alpha}$  atoms by adding the differences of forces on these extra hydrogen atoms. For instance, the difference of the forces of the H atoms added to the carbonyl group of the residue  $i-1$  (H atom in the left blue cycle of Figure 3) is

$$\Delta \mathbf{F} = \mathbf{F}_{Cap_i}^{ex-H}(Cap_{i-1}^* A_i Cap_{i+1}) - \mathbf{F}_{Cap_i}^{ex-H}(Cap_{i-1}^* Cap_{i+1}) \neq 0, \quad (5)$$

Which is added to the force of the  $C_{\alpha}$  atom of residue  $i-1$ ,

$$\mathbf{F}_{C_{\alpha}}^{\text{final}} = \mathbf{F}_{C_{\alpha}} + \Delta \mathbf{F} \quad (6)$$

This approach will balance the forces of the fragments and conserve the total energy of the system.

## RESULT AND DISCUSSION

### Performance of FB-GMFCC on Pure Proteins

To validate the FB-GMFCC method, we checked its performance for four protein systems and compared the calculated results with that calculated by conventional full-system QM calculations. An Ace-(ALA)<sub>9</sub>-NME linear peptide was constructed by the TLEAP software in the AMBER16 package, and three small proteins with different secondary structures are selected from the protein data bank (ID: 2I9M, 1LE1, and 2OED). Energy minimization (by using gradient descent and conjugate gradient algorithms with Amber ff14SB force field) was performed to remove bad contacts in these structures before QM calculations. Comparisons of job CPU times of FB-GMFCC and full system QM calculations were shown in Table 1. We can see that FB-GMFCC method is 4 or

**TABLE 1** | Comparison of the computational cost of FB-GMFCC and full-system QM calculation on the Linux server with two Intel E5-2680v3 CPUs (14 cores, 2.50 GHz).

System	Protein type	Atom No.	CPU time (hour)	
			Full-system QM	FB-GMFCC
Ace-(ALA) <sub>9</sub> -NME	Linear peptide	102	0.63	0.27
2I9M	$\alpha$ -helix	237	4.97	0.90
1LE1	$\beta$ -sheet	194	4.06	0.76
2OED	Mixed structure	862	N/A*	3.16

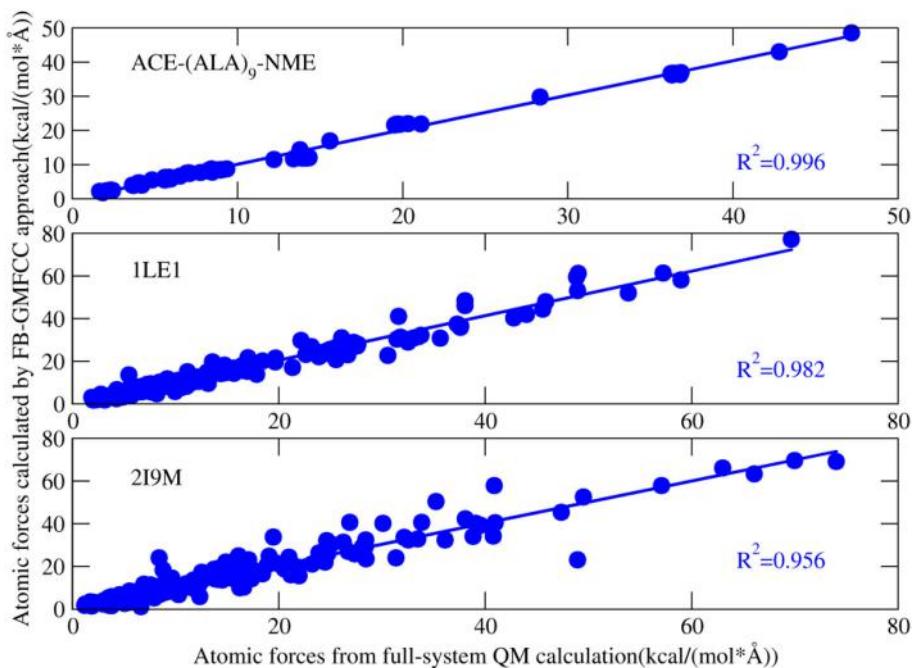
All calculations were performed with GAUSSIAN09 at M062X/6-31G\* level. \*Full-system QM calculation is not possible on our current machine due to large size of the system.

5 times faster than full QM calculation for a real protein with about 200 atoms. For the larger 2OED protein, the full-system QM calculation is not possible on our computer system due to its large size. It should be noted that the computational time in the present approach is essentially linear with the system size as shown in Table 1.

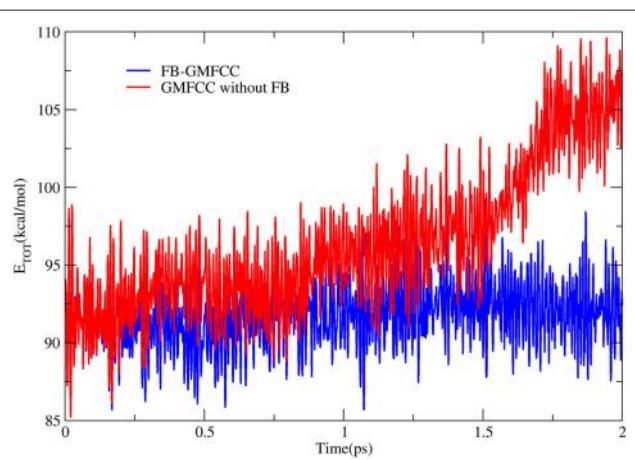
Figure 4 shows the comparison of computed atomic forces with those from full-system QM calculation. Overall, the atomic forces are in good agreement with the full-system calculations except a few points. For example, there is a bad point in the calculated atomic force of 2I9M (Figure 4), which correspond to one H atom on the  $\epsilon$ -amino group of LYS8. After carefully checking the structure, we found that a salt bridge is formed between this group and the side chain of GLU4. As a result, this salt bridge cannot be accurately described by the force field which is used in the present method to describe interaction between non-neighboring residues.

### Ab Initio Molecular Dynamic in Gas Phase

To further check the performance of FB-GMFCC method, we performed an AIMD simulation for the linear peptide ACE-(ALA)<sub>9</sub>-NME in the NVE ensemble and gas phase. The simulation was performed by combining the FB-GMFCC and the TINKER program. Before the AIMD simulation, an energy minimization (by using gradient descent and conjugate gradient algorithms with Amber ff14 force field), a 400 ps heating simulation which heated the system from 0 to 300 K and a 5 ns equilibrium MD simulation (by using velocity verlet algorithm and the same force field) were performed. The AIMD simulation lasted 2 ps with 1 fs time step and without any constraints. Another AIMD simulation without force balance was also performed as a reference. The total energy fluctuations in these two AIMD simulations were shown in Figure 5. As can be seen, the total energy in the AIMD simulation based on GMFCC without force balance is gradually increased in gas phase NVE ensemble, which means that the energy is not conserved if atomic forces of extra cap hydrogen atoms are not compensated to corresponding  $C_{\alpha}$  atoms. However, the total energy of FB-GMFCC can be maintained well and conserved at 91 kcal/mol. On average, the extra H atoms can import extra forces as large as 11 kcal/(mol $\cdot$ Å) to the system, which lead to additional works



**FIGURE 4** | Correlation between atomic forces of three proteins [ACE-(ALA)<sub>9</sub>-NME linear peptide, 1LE1, 2I9M] calculated by full-system QM calculation and that calculated by FB-GMFCC. All the QM calculations were performed at the M062X/6-31G\* level.



**FIGURE 5** | Time evolution of total energies in AIMD simulations of a linear peptide ACE-(ALA)<sub>9</sub>-NME with FB-GMFCC method (blue) and GMFCC without force balance (red) in gas phase and NVE ensemble.

on the system. Thus it is necessary to balance the atomic forces on these extra H atoms in the AIMD simulation.

## AIMD in Explicit Water

Since water plays an important role in protein structure and dynamics, the study of protein should be carried out in the solvent environment. The FB-GMFCC approach can also be used to perform AIMD simulations for proteins in explicit solvent environment. The total energy of protein-solvent system with

FB-GMFCC can be expressed by the following formula.

$$E_{\text{total}} = E_{\text{Protein}}^{\text{FB-GMFCC}} + E_{\text{water}}^{\text{MM}} + E_{\text{Protein-water}}^{\text{MM}} \quad (7)$$

To save the computation cost, the inter- and intra-interactions of water molecules and their interactions with proteins are described by classical force field (Amber ff14SB), as mechanical embedding in the QM/MM framework.

We performed 110 ps AIMD simulation for the relatively larger protein (2OED, 56 residues, 862 atoms) in explicit water. The protein was solvated in a water ball consisting of 3084 TIP3P water molecules. Before AIMD simulation, energy minimizations were performed to remove bad contacts in the system, and a 25 ps heating simulation was performed to heat the system slowly to 300 K. A restraint of 50 kcal/mol was used on the backbone to avoid large unphysical structural change in heating process. Then the system underwent AIMD simulation with 1 fs time step and without any constraints. The Langevin thermostat with the collision frequency 2.0 ps<sup>-1</sup> was applied to control the temperature. In addition, there was a 20 kcal/mol half-harmonic restrain used on the boundary of water ball to avoid the escaping of water molecules.

This AIMD simulation was performed on a linux server cluster with 30 nodes and each node has dual Intel Xeon E5-2680v3 CPUs. To balance the computational cost and accuracy, the combination of BLYP functional and 6-31G\* basis set was used in the calculation. It took 55 days to complete the simulation. The time evolution of temperature in the simulation was shown in **Figure 6**. We can see that the temperature is very stable in the trajectory. The backbone RMSD with respect to the

X-ray structure is no larger than 1.5 Å, which means that the structure of 2OED protein are relatively stable during the 110 ps AIMD simulation.

Recently, many researchers discovered that considerable deviations from planarity of peptide bond ( $\omega = 180^\circ$ ) can be identified in atomic resolution X-ray structures, sometimes even exceeding  $10\text{--}15^\circ$  (Wlodawer et al., 1984; MacArthur and Thornton, 1996; Ulmer et al., 2003). As the resolution of the structure (2OED) used in this work is very high (1.1 Å) and the coordinates of hydrogen atoms in this structure were further refined with NMR experiment (Ulmer et al., 2003), it is worth to compare the planarity of peptide bonds in the experimental structure and the AIMD trajectory. Six peptide bonds were selected from the experimental structure as their  $O_{i-1}\text{--}C_{i-1}\text{--}N_i\text{--}H_i^N$  dihedral angles deviate the most from the peptide plane. The result can be found in **Table 2** and **Figure 7**. We can see that four of the six peptide bonds still maintained their large deviations. The results calculated by FB-GMFCC generally agree well with the experiments, especially for VAL21, TYR3, and PHE52. For comparison, we also test the performance of MD with classical force field (Amber ff14SB) at the same conditions. Not surprisingly, the Amber ff14SB force field prefers planer peptide bonds, which was predefined. As based on QM calculation, the AIMD simulation generally describes the intra-protein interactions more accurately.

## CONCLUSION

In this study, a force balanced generalized molecular fractionation with conjugate caps (FB-GMFCC) method

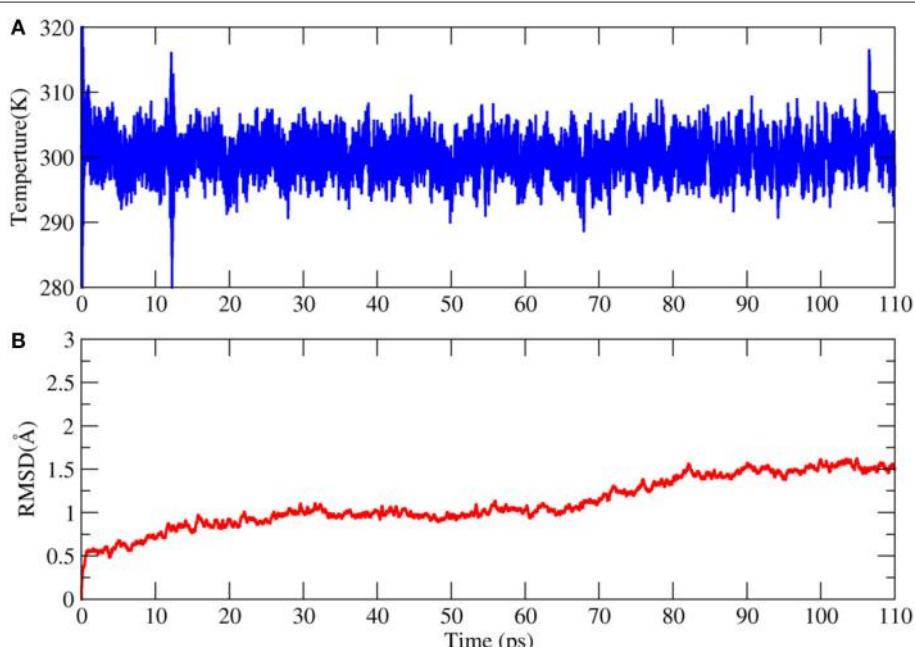
was presented. In this approach, fragment-based energies of individual residues and interaction energies of residues that form backbone hydrogen bonds are calculated by quantum mechanics. Other non-bonded interactions are considered by the polarizable AMOEBA force field. The calculated atomic forces of this method showed good agreements with that calculated by the conventional full-system QM calculations. A key element of the FB-GMFCC method is that the atomic forces of capped H atoms are corrected to achieve the conservation of the total force of the studied system.

We also demonstrated the applicability of FB-GMFCC method for performing *ab initio* molecular dynamic (AIMD) simulations for proteins. The results of an Ace-(ALA)<sub>9</sub>-NME linear protein showed that only the balanced force can keep

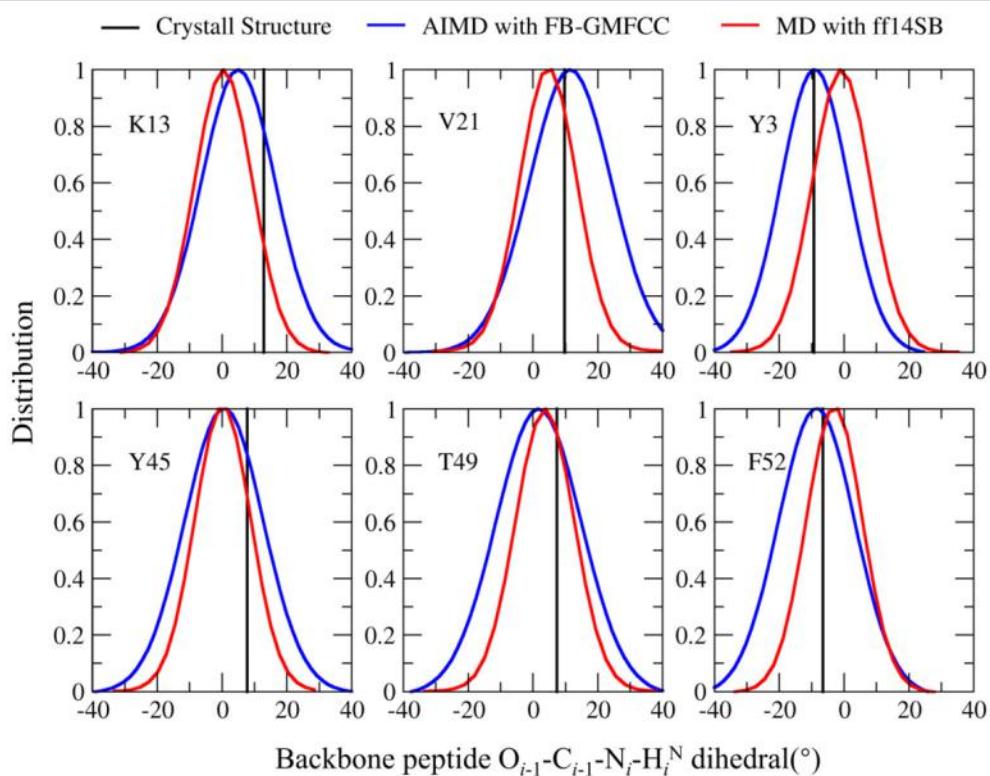
**TABLE 2 |** Comparison of six selected  $O_{i-1}\text{--}C_{i-1}\text{--}N_i\text{--}H_i^N$  dihedral angles in both the AIMD and AMBER MD calculations with experimental measurements.

Residue	$O_{i-1}\text{--}C_{i-1}\text{--}N_i\text{--}H_i^N$ dihedral ( $^\circ$ )		
	Crystall structure	AIMD with FB-GMFCC	AMBER MD with ff14SB
LYS13	12.88	5.91	1.84
VAL21	9.69	11.91	6.27
TYR3	-9.34	-8.23	0.74
TYR45	7.70	1.21	1.64
THR49	7.36	2.36	5.32
PHE52	-6.54	-7.72	-1.50

The calculated values were averaged from the trajectories.



**FIGURE 6 |** Time evolution of the temperature (A) and root mean square deviation (RMSD) of backbone atoms with respect to the X-ray structure in the AIMD simulation of 2OED protein (B).



**FIGURE 7 |** Comparison of the  $O_{i-1}-C_{i-1}-N_i-H_i^N$  dihedral angle distribution in the FB-GMFCC AIMD (blue lines) and classical MD simulations (with Amber ff14SB force field, red lines) for 6 selected peptide bonds. Black lines are experimental values.

the conservation of the total energy during the simulation. An 110 ps AIMD simulation was also performed for a relatively large protein with 56 residues and 862 atoms in explicit water. Compared with the classical force field, the AIMD simulations gave better description about the geometry of peptide bonds. It should be noted that the accuracy of the FB-MFCC method still have room to be improved. Further development of this method will focus on the consideration of strong short-range interactions such as salt bridges and hydrogen bond including side chains, relevant work is underway in our laboratory.

These results have shown that the FB-GMFCC approach is potentially powerful and attractive for studying protein dynamics. As a fragment based approach, the FB-GMFCC method is linear-scaling and trivially parallelizable. With further development and improvement, this method will become more and more practical for AIMD simulation of larger proteins.

## REFERENCES

- Best, R. B., Buchete, N.-V., and Hummer, G. (2008). Are current molecular dynamics force fields too helical? *Biophys. J.* 95, L07–L09. doi: 10.1529/biophysj.108.132696
- Cheatham, T. E. III, and Kollman, P. A. (2000). Molecular dynamics simulation of nucleic acids. *Annu. Rev. Phys.*

## AUTHOR CONTRIBUTIONS

MX performed theoretical calculation and analysis of result. TZ developed the theory and contributed to the writing and discussion of the paper. JZ organized the project and contributed to the discussion and writing of the paper.

## ACKNOWLEDGMENTS

This work was supported by National Key R&D Program of China (Grant no. 2016YFA0501700), National Natural Science Foundation of China (Grant nos. 21433004, 91641116, 91753103), Shanghai Putuo District (Grant 2014-A-02), Innovation Program of Shanghai Municipal Education Commission (201701070005E00020), and NYU Global Seed Grant. We thank the Supercomputer Center of East China Normal University for providing us computer time.

*Chem.* 51, 435–471. doi: 10.1146/annurev.physchem.51.1.435

Chung, L. W., Sameera, W. M. C., Ramozzi, R., Page, A. J., Hatanaka, M., Petrova, G. P., et al. (2015). The ONIOM method and its applications. *Chem. Rev.* 115, 5678–5796. doi: 10.1021/cr5004419

Collins, M. A., and Bettens, R. P. A. (2015). Energy-based molecular fragmentation methods. *Chem. Rev.* 115, 5607–5642. doi: 10.1021/cr500455b

- Collins, M. A., Cvitkovic, M. W., and Bettens, R. P. A. (2014). The combined fragmentation and systematic molecular fragmentation methods. *Acc. Chem. Res.* 47, 2776–2785. doi: 10.1021/ar500088d
- Dal Peraro, M., Raugei, S., Carloni, P., and Klein, M. L. (2005). Solute-solvent charge transfer in aqueous solution. *ChemPhysChem* 6, 1715–1718. doi: 10.1002/cphc.200500039
- Duan, L. L., Mei, Y., Zhang, D., Zhang, Q. G., and Zhang, J. Z. H. (2010). Folding of a helix at room temperature is critically aided by electrostatic polarization of intraprotein hydrogen bonds. *J. Am. Chem. Soc.* 132, 11159–11164. doi: 10.1021/ja102735g
- Fedorov, D. G., Asada, N., Nakanishi, I., and Kitaura, K. (2014). The use of many-body expansions and geometry optimizations in fragment-based methods. *Acc. Chem. Res.* 47, 2846–2856. doi: 10.1021/ar500224r
- Gao, J., Truhlar, D. G., Wang, Y., Mazack, M. J. M., Loeffler, P., Provorov, M. R., et al. (2014). Explicit polarization: a quantum mechanical framework for developing next generation force fields. *Acc. Chem. Res.* 47, 2837–2845. doi: 10.1021/ar5002186
- Gao, Y., Lu, X., Duan, L. L., Zhang, J. Z. H., and Mei, Y. (2012). Polarization of intraprotein hydrogen bond is critical to thermal stability of short helix. *J. Phys. Chem. B* 116, 549–554. doi: 10.1021/jp208953x
- Gordon, M. S., Fedorov, D. G., Pruitt, S. R., and Slipchenko, L. V. (2012). Fragmentation methods: a route to accurate calculations on large systems. *Chem. Rev.* 112, 632–672. doi: 10.1021/cr200093j
- He, X., Zhu, T., Wang, X., Liu, J., and Zhang, J. Z. H. (2014). Fragment quantum mechanical calculation of proteins and its applications. *Acc. Chem. Res.* 47, 2748–2757. doi: 10.1021/ar500077t
- Isborn, C. M., Mar, B. D., Curchod, B. F. E., Tavernelli, I., and Martinez, T. J. (2013). The charge transfer problem in density functional theory calculations of aqueously solvated molecules. *J. Phys. Chem. B.* 117, 12189–12201. doi: 10.1021/jp4058274
- Ji, C. G., Mei, Y., and Zhang, J. Z. H. (2008). Developing polarized protein-specific charges for protein dynamics: MD free energy calculation of pK(a) shifts for Asp(26)/Asp(20) in thioredoxin. *Biophys. J.* 95, 1080–1088. doi: 10.1529/biophys.108.131110
- Ji, C., and Mei, Y. (2014). Some practical approaches to treating electrostatic polarization of proteins. *Acc. Chem. Res.* 47, 2795–2803. doi: 10.1021/ar500094n
- Karplus, M., and McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9, 646–652. doi: 10.1038/nsb0902-646
- Karplus, M., and Petsko, G. A. (1990). Molecular dynamics simulations in biology. *Nature* 347, 631–639. doi: 10.1038/347631a0
- Li, S., Li, W., and Ma, J. (2014). Generalized energy-based fragmentation approach and its applications to macromolecules and molecular aggregates. *Acc. Chem. Res.* 47, 2712–2720. doi: 10.1021/ar500038z
- Liu, J. F., Zhu, T., Wang, X. W., He, X., and Zhang, J. Z. H. (2015). Quantum fragment based *ab initio* molecular dynamics for proteins. *J. Chem. Theor. Comput.* 11, 5897–5905. doi: 10.1021/acs.jctc.5b00558
- MacArthur, M. W., and Thornton, J. M. (1996). Deviations from planarity of the peptide bond in peptides and proteins. *J. Mol. Biol.* 264, 1180–1195. doi: 10.1006/jmbi.1996.0705
- Ponder, J. W., and Case, D. A. (2003). Force fields for protein simulations. *Protein Simulat.* 66, 27–85. doi: 10.1016/S0065-3233(03)66002-X
- Ponder, J. W., Wu, C., Ren, P., Pande, V. S., Chodera, J. D., Schnieders, M. J., et al. (2010). Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* 114, 2549–2564. doi: 10.1021/jp910674d
- Prinz, J. H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., et al. (2011). Markov models of molecular kinetics: generation and validation. *J. Chem. Phys.* 134:174105. doi: 10.1063/1.3565032
- Pruitt, S. R., Bertoni, C., Brorsen, K. R., and Gordon, M. S. (2014). Efficient and accurate fragmentation methods. *Acc. Chem. Res.* 47, 2786–2794. doi: 10.1021/ar500097m
- Raghavachari, K., and Saha, A. (2015). Accurate composite and fragment-based quantum chemical models for large molecules. *Chem. Rev.* 115, 5643–5677. doi: 10.1021/cr500606e
- Ramabhadran, R. O., and Raghavachari, K. (2014). The successful merger of theoretical thermochemistry with fragment-based methods in quantum chemistry. *Acc. Chem. Res.* 47, 3596–3604. doi: 10.1021/ar500294s
- Ren, P., and Ponder, J. W. (2002). Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *J. Comput. Chem.* 23, 1497–1506. doi: 10.1002/jcc.10127
- Ren, P., Wu, C., and Ponder, J. W. (2011). Polarizable atomic multipole-based molecular mechanics for organic molecules. *J. Chem. Theor. Comput.* 7, 3143–3161. doi: 10.1021/ct200304d
- Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., et al. (2010). Atomic-level characterization of the structural dynamics of proteins. *Science* 330, 341–346. doi: 10.1126/science.1187409
- Tong, Y., Mei, Y., Li, Y. L., Ji, C. G., and Zhang, J. Z. H. (2010). Electrostatic polarization makes a substantial contribution to the free energy of avidin-biotin binding. *J. Am. Chem. Soc.* 132, 5137–5142. doi: 10.1021/ja909575j
- Ufimtsev, I. S., Luehr, N., and Martinez, T. J. (2011). Charge transfer and polarization in solvated proteins from *ab initio* molecular dynamics. *J. Phys. Chem. Lett.* 2, 1789–1793. doi: 10.1021/jz200697c
- Ufimtsev, I. S., and Martinez, T. J. (2009). Quantum chemistry on graphical processing units. 3. analytical energy gradients, geometry optimization, and first principles molecular dynamics. *J. Chem. Theor. Comput.* 5, 2619–2628. doi: 10.1021/cr9003004
- Ulmer, T. S., Ramirez, B. E., Delaglio, F., and Bax, A. (2003). Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy. *J. Am. Chem. Soc.* 125, 9179–9191. doi: 10.1021/ja0350684
- Wang, X., Liu, J., Zhang, J. Z. H., and He, X. (2013). Electrostatically embedded generalized molecular fractionation with conjugate caps method for full quantum mechanical calculation of protein energy. *J. Phys. Chem. A.* 117, 7149–7161. doi: 10.1021/jp400779t
- Wei, D. Q., Guo, H., and Salahub, D. R. (2001). Conformational dynamics of an alanine dipeptide analog: an *ab initio* molecular dynamics study. *Phys. Rev. E.* 64:0111907. doi: 10.1103/PhysRevE.64.0111907
- Weiner, S. J., Kollman, P. A., Case, D. A., Chandra Singh, U., Ghio, C., Alagona, G., et al. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 106, 765–784. doi: 10.1021/ja00315a051
- Weiner, S. J., Kollman, P. A., Nguyen, D. T., and Case, D. A. (1986). An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* 7, 230–252. doi: 10.1002/jcc.540070216
- Wlodawer, A., Walter, J., Huber, R., and Sjolin, L. (1984). Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and X-ray refinement of crystal form II. *J. Mol. Biol.* 180, 301–329. doi: 10.1016/S0022-2836(84)80006-6
- Wu, J. C., Chattree, G., and Ren, P. (2012). Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor. Chem. Acc.* 131:1138. doi: 10.1007/s00214-012-1138-6
- Xu, X., Nakatsuji, H., Ehara, M., Lü, X., Wang, N. Q., and Zhang, Q. E. (1998). Cluster modeling of metal oxides: the influence of the surrounding point charges on the embedded cluster. *Chem. Phys. Lett.* 292, 282–288. doi: 10.1016/S0009-2614(98)00687-3
- Zhang, D. W., and Zhang, J. Z. H. (2003). Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein-molecule interaction energy. *J. Chem. Phys.* 119, 3599–3605. doi: 10.1063/1.1591727

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Xu, Zhu and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# QM Cluster or QM/MM in Computational Enzymology: The Test Case of LigW-Decarboxylase

Mario Prejanò, Tiziana Marino\* and Nino Russo

Dipartimento Di Chimica e Tecnologie Chimiche, Università della Calabria, Rende, Italy

The catalytic mechanism of the decarboxylation of 5-carboxyvanillate by LigW producing vanillic acid has been studied by using QM cluster and hybrid QM/MM methodologies. In the QM cluster model, the environment of a small QM model is treated with a bulky potential while two QM/MM models studies include partial and full protein with and without explicitly treated water solvent. The studied reaction involves two sequential steps: the protonation of the carbon of the 5-carboxy-vanillate substrate and the decarboxylation of the intermediate from which results deprotonated vanillic acid as product. The structures and energetics obtained by using three structural models and two density functionals are quite consistent to each other. This indicates that the small QM cluster model of the presently considered enzymatic reaction is appropriate enough and the reaction is mainly influenced by the active site.

**Keywords:** QM, QM/MM, decarboxylation, enzymatic catalysis, reaction mechanism, LigW

## OPEN ACCESS

### Edited by:

Sam P. De Visser,  
University of Manchester,  
United Kingdom

### Reviewed by:

Artur Nenov,  
Università degli Studi di Bologna, Italy

Ahmet Altun,  
Max-Planck-Institut für  
Kohlenforschung, Germany

### \*Correspondence:

Tiziana Marino  
tiziana.marino65@unical.it

### Specialty section:

This article was submitted to  
Theoretical and Computational  
Chemistry, a section of the journal  
Frontiers in Chemistry

**Received:** 11 April 2018

**Accepted:** 08 June 2018

**Published:** 28 June 2018

### Citation:

Prejanò M, Marino T and Russo N (2018) QM Cluster or QM/MM in Computational Enzymology: The Test Case of LigW-Decarboxylase. *Front. Chem.* 6:249.  
doi: 10.3389/fchem.2018.00249

## INTRODUCTION

Enzymes are biological machines that efficiently catalyze a huge number of chemical reactions in the very short time steps required by the physiological processes. In the last decades, computational enzymology has become a very useful tool for studying enzyme activity since it allows to determine the energies and structures of short-lived intermediates and transition states. Through computational enzymology, different reaction pathways can be analyzed, and their feasibility can be established by a careful analysis of calculated energy barriers. A crucial issue in computational enzymology is the choice of the model to be used in the simulations. The choice is not so obvious because it depends on the nature of enzyme (without or with metal cofactor) on the catalytic pocket and on the amino acids implicated in the chemical reaction. In fully quantum mechanical (QM) treatment (Himo, 2006; Ramos and Fernandes, 2008; Siegbahn and Himo, 2011; Merz, 2014), a cluster that contains all the residues around the active site is considered.

In metalloenzymes, the construction of the cluster model is facilitated by the presence of metal ions and all the residues of their inner coordination sphere. One may also need to include some other surrounding residues involved in the chemical process. Atoms at the periphery of the model, where truncation is made, are normally frozen in their original positions present in the crystallographic structure for avoiding artificial expansion or other rearrangements (Blomberg et al., 2014). The surrounding protein environment not directly implicated in the chemical transformation, is modeled with implicit dielectric constant-based solvation models (Warshel, 1991). This method is highly versatile and widely applied to a large variety of enzyme families and to different classes of enzymes (Ramos and Fernandes, 2008; Liao et al., 2010; Amata et al., 2011a; Himo, 2017; Piazzetta et al., 2017; Prejanò et al., 2017a). A different approach developed in 1976

(Warshel and Levitt, 1976) is the hybrid quantum mechanics/molecular mechanics (QM/MM) (Senn and Thiel, 2009; Quesne et al., 2016; Ryde, 2016). In this procedure, other than the QM portion a large number of residues (or the whole enzyme sequence) is treated at molecular mechanics level (MM) (Senn and Thiel, 2009). Convergence studies performed by different research groups indicated that QM-cluster models (Siegbahn and Himo, 2011; Ryde, 2017) gives reliable energetics when the size of the model is large enough. Herein we perform a theoretical study using both QM cluster and QM/MM approaches on the gene product of LigW of 5-carboxyvanillate decarboxylase (5CVA) (Vladimirova et al., 2016). The QM part in all the models has been treated in the framework of density functional theory (DFT) and by using two different exchange-correlation functionals.

The LigW belongs to the amidohydrolase (AHS) superfamily including a high number of enzymes catalyzing the hydrolysis of a wide range of substrates. In all AHS members, a mononuclear or binuclear metal binding site is found (Gerlt and Babbitt, 2001; Seibert and Raushel, 2005). All AHS members have a  $(\beta/\alpha)_8$ -barrel structural fold and catalyze the metal-dependent hydrolysis of phosphate and carboxylate esters (Jackson et al., 2005; Shapir et al., 2006; Elias et al., 2008; Khurana et al., 2009; Duarte et al., 2011; Tobimatsu et al., 2013). LigW catalyzes the C-C bond cleavage of 5-CV to vanillate (VAN) in an oxidant-independent fashion. The 5-carboxyvanillate (5-CV) represents one of the final product of the multienzymatic degradation of the biphenyl lignin derivatives. The lignin degradation of microbial origin represents an interesting process from both commercial and biotechnological point of view owing to the plant biomass conversion in renewable aromatic chemicals and biofuels (Liu and Zhang, 2006). Furthermore, decarboxylation represents a process of widespread occurrence in nature and therefore it is of relevant biological interest (Faponle et al., 2016).

## COMPUTATIONAL METHODS

All the calculations were carried out by using the Gaussian 09 program (Gaussian 09, Revision D.01, 2011)<sup>1</sup>. The QM portions were treated with the B3LYP (Lee et al., 1988; Becke, 1993) hybrid density functional. 6-31+G(d,p) basis set was used for the C, N, O, and H atoms, whereas the SDD pseudo-potential and corresponding orbital basis set (Andrae et al., 1990) were employed for Mn atom. Our own N-layered integrated molecular orbital and molecular mechanics (ONIOM) method was applied as the QM/MM method in the framework of electronic embedding scheme, in which the effects of the fixed MM charges are incorporated in the QM hamiltonian (Svensson et al., 1996; Vreven et al., 2006). As shown in Figure S1, the enzyme-substrate complex (ES) is a high-spin sextet species while its low-spin doublet and the intermediate-spin quartet states are energetically not accessible. The sextet state does not suffer from any spin contamination ( $\langle S^2 \rangle$  equal to 8.75). The optimized minima and transition states on the potential energy surfaces were confirmed by the analysis of the corresponding

Hessian matrices. Zero-point-energy corrections were calculated and added to the final energies. In order to obtain more accurate energies, single point calculations on the optimized structures were performed with the larger basis set 6-311+G(2d,2p) taking into account the effects of the protein environment by using the solvation model density (SMD) (Marenich et al., 2009), with a dielectric constant ( $\epsilon = 4$ ) of the enzyme environment, for the cluster simulations (Alberto et al., 2010; Liao et al., 2010; Amata et al., 2011a,b; Himo, 2017; Piazzetta et al., 2017; Prejanò et al., 2017a,b). Energetics presented includes D3 dispersion correction (Grimme et al., 2011). To evaluate the effect of the exchange-correlation functionals single point calculations on the B3LYP optimized geometries have been performed by using the M06-L functional that was previously demonstrated to be accurate for describing metal containing systems properties (Zhao and Truhlar, 2006, 2008) (see Table S1). NBO analysis (NBO, version 3.1, 2001)<sup>2</sup> was performed on all intercepted stationary points at QM and QM/MM levels with B3LYP functional. Furthermore, the noncovalent interactions on the minima of the PES have been assessed by using the NCIPILOT tool (NCIPILOT, version 3.0, 2011)<sup>3</sup>.

## COMPUTATIONAL SETUP AND QM MODEL DEFINITIONS

The model of the LigW active site, used for both QM and QM/MM calculations, was obtained from the three-dimensional structure of wild-type LigW in the presence of the substrate-like inhibitor 5-nitrovanillate (5-NV) isolated by *N. aromaticivorans* (PDB id: 4QRN, resolution: 1.07 Å). Vladimirova et al. (2016) due to the very small difference (one atom) between the inhibitor (5-NV) and substrate (5-CV). This choice has been already shown sufficient when structurally compared with larger QM clusters (Sheng et al., 2017). In the active site, (see Figure 1) the manganese ion is octahedrally coordinated to Glu-19, His-188, Asp-314, one water molecule w1 and the substrate. Two water molecules, (w2 and w3), located at about 5 Å from the substrate and other residues of the active site pocket not directly bound to the metal ion are retained in QM region (Arg58, Phe212, His241, Arg252, and Tyr317).

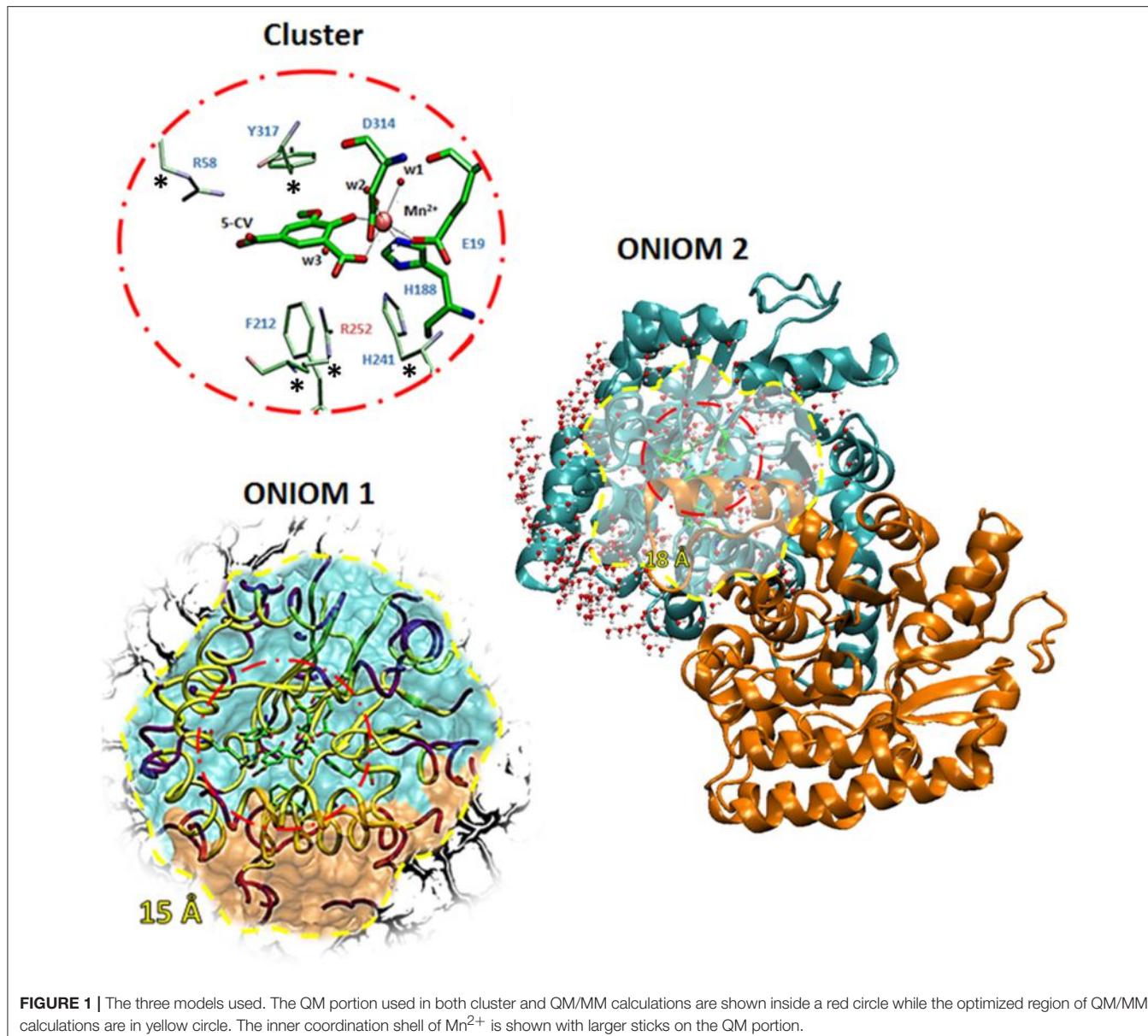
In the QM/MM models, the Amber ff14SB force field (Maier et al., 2015) as implemented in AMBER16 software was used. The missing MM parameters for the substrate 5-CV were created from single molecule optimization at HF/6-31G(d) level of theory with the Antechamber tool, as implemented in AMBER16 (AMBER version 16, 2016)<sup>4</sup>. At this purpose the General Amber Force Field (GAFF) (Wang et al., 2004) and the Restrained Electrostatic Potential (RESP) (Bayly et al., 1993) methods were used to derive intramolecular and Lennard-Jones parameters and atomic charges, respectively (see Table S2).

<sup>2</sup>NBO, version 3.1 (2001).

<sup>3</sup>NCIPILOT, version3.0 (2011). Download: <http://www.lct.jussieu.fr/pagesperso/contrera/ncipilot.html>

<sup>4</sup>AMBER 16 (2016), University of California, San Francisco.

<sup>1</sup>Gaussian 09, Revision D.01 (2011), Gaussian, Inc., Wallingford CT.



## QM Cluster

All the amino acids of the QM region were truncated at the  $\alpha$ -carbons, and hydrogen atoms were added manually. In order to avoid unrealistic movements of the groups during the geometry optimizations, the truncated  $\alpha$ -carbons of the outer coordination shell labeled by stars in **Figure 1** were kept fixed to their crystallographic positions. The residues were modeled according to standard procedure (Liao et al., 2010; Amata et al., 2011a,b; Siegbahn and Himo, 2011; Blomberg et al., 2014; Himo, 2017; Prejanò et al., 2017a) considering the protonation states coming from the experimental evidences (Vladimirova et al., 2016). The obtained model consists of 126 atoms with a total charge equal to zero. The size of the cluster is adequate enough to represent the chemistry involved in the considered reaction mechanisms for formation or breaking bonds.

## ONIOM-1

In this model, the QM region is surrounded by the residues present in radius of 15 Å from the metal ion center. In this way, the interactions between  $\alpha$  and  $\beta$  subunits of the homodimer were included. Inside the considered sphere, an outer shell of residues with a thickness of 2 Å was fixed, and only the inner 13 Å shell was allowed to move during the QM/MM geometry optimizations. This strategy is commonly used to avoid drifting through multiple minima unrelated to the reaction coordinate. This model includes in the MM region also a number of water molecules (20) present in the crystallographic structure. The obtained model consists of 2,154 atoms with 118 atoms in QM region (**Figure 1**).

## ONIOM-2

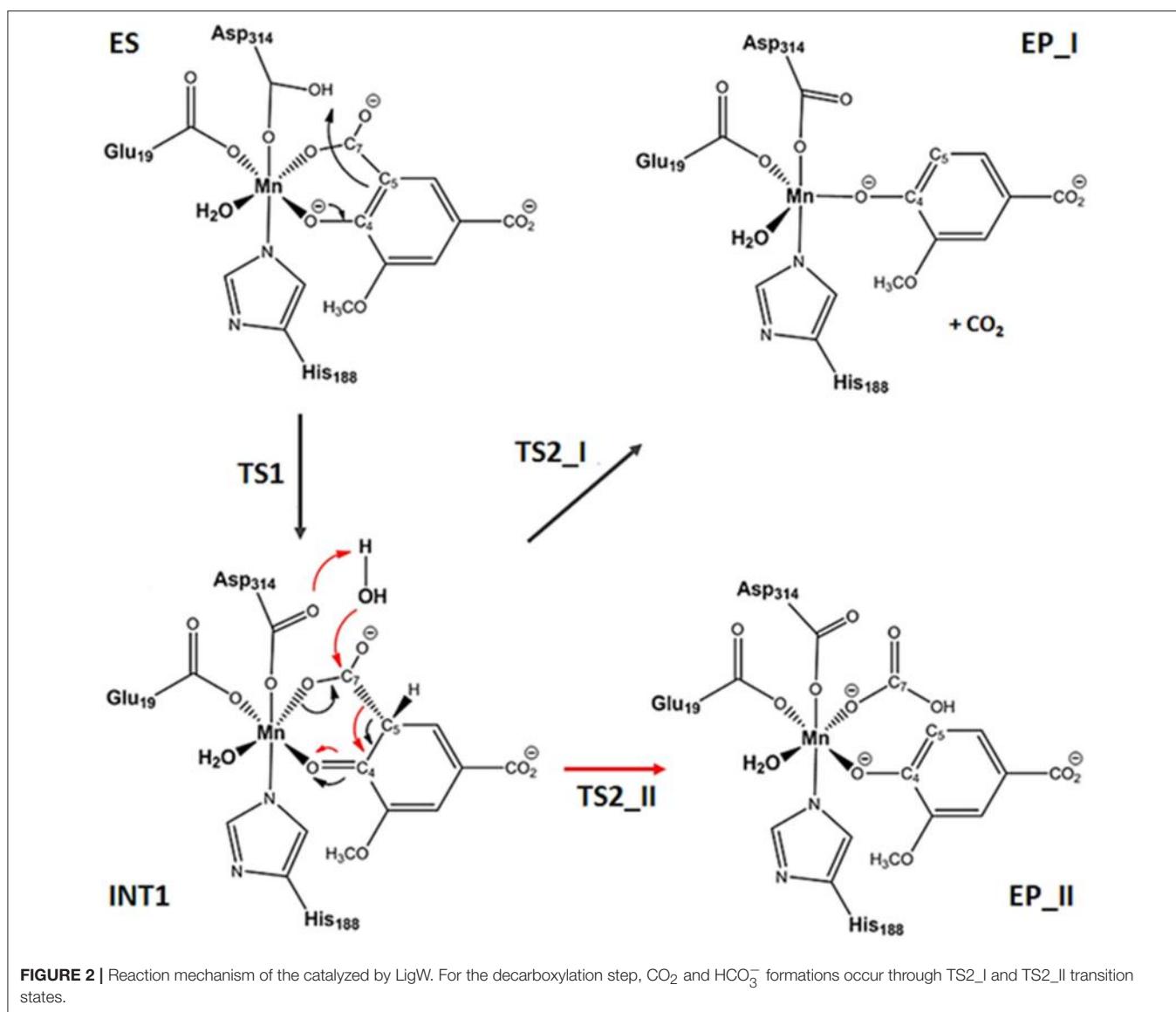
A rectangular box was used to solvate the system up to 12.0 Å of the metal center. During the optimizations, all the water molecules and protein atoms in the 18 Å from the active site were kept frozen, as proposed by a recent work (Medina et al., 2017). The final model contains 11,895 atoms with 118 atoms of QM region. In this case, the MM region includes the whole protein and a number of water molecules within 5 Å around of catalytic domain as depicted in **Figure 1**.

## RESULTS AND DISCUSSION

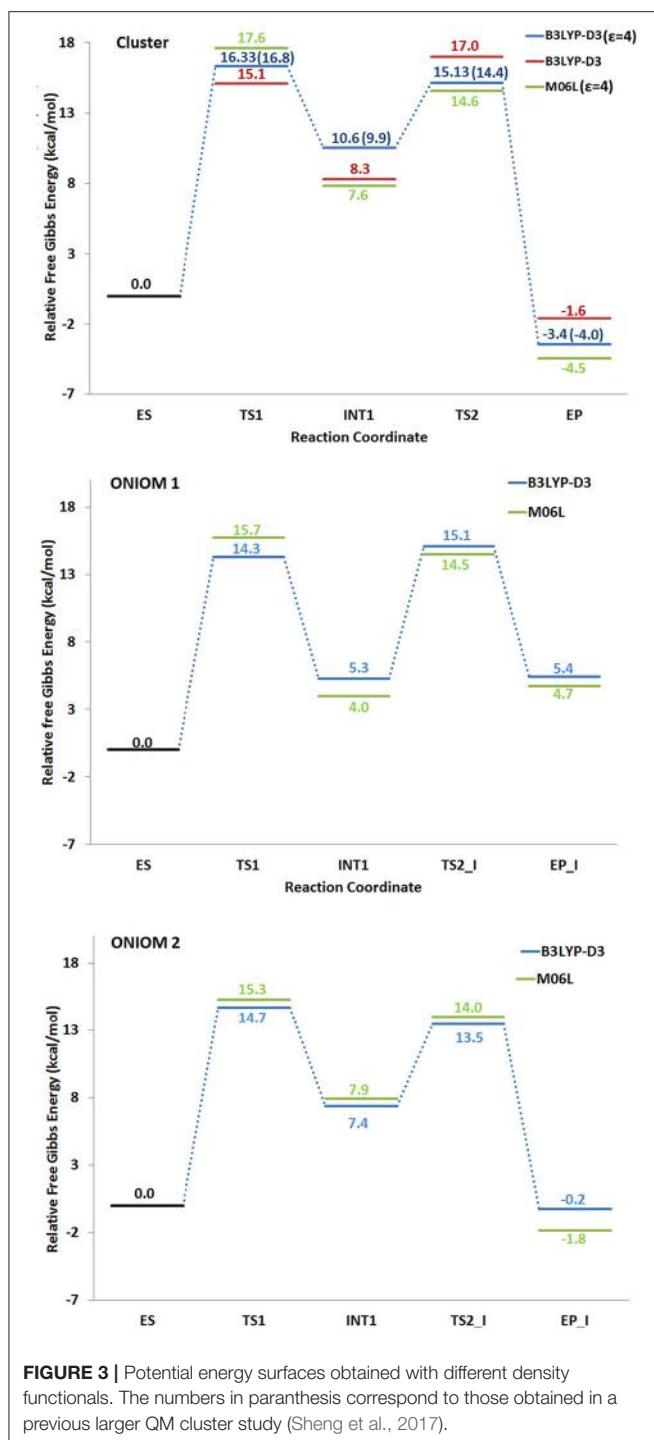
The reaction can follow two paths with the formation of  $\text{CO}_2$  or  $\text{HCO}_3^-$  products (see **Figure 2**). After the formation of the ES, the reaction proceeds with the proton transfer from Asp314 to C5 of the substrate generating the INT1 species, that acts as common intermediate for the formation of EP\_I

or EP\_II complexes in which  $\text{CO}_2$  or  $\text{HCO}_3^-$  product should be released. In both decarboxylation pathways, it is clear that the enzyme must generate an adjacent electron sink (such as the ketone carbonyl C4 since the formation of the new carbon-hydrogen bond) to stabilize the incipient carbanion at C5 prior to decarboxylation. This mechanism corresponds to that explored in the recent combined experimental and theoretical work (Sheng et al., 2017) where the membrane inlet mass spectrometry (MIMS) based assay is applied to study the LigW mechanism. The above-mentioned MIMS-based strategy (Sheng et al., 2017) was able to establish  $\text{CO}_2$  and not  $\text{HCO}_3^-$  as reaction product. We have considered also the path for the bicarbonate release but our calculated PESs with the three models used give very high energetic barriers (see Table S1) that are not compatible with the enzymatic kinetics.

All the obtained PESs with the used models are depicted in **Figure 3**. Those concerning the QM one will be compared with



**FIGURE 2 |** Reaction mechanism of the catalyzed by LigW. For the decarboxylation step,  $\text{CO}_2$  and  $\text{HCO}_3^-$  formations occur through TS2\_I and TS2\_II transition states.



**FIGURE 3 |** Potential energy surfaces obtained with different density functionals. The numbers in parenthesis correspond to those obtained in a previous larger QM cluster study (Sheng et al., 2017).

the values arising from the previous larger QM-cluster model study (Sheng et al., 2017). B3LYP optimized structures obtained employing the ONIOM-2 model of all the species are given in Figure 4 while that for the QM and ONIOM-1 models are given in Figures S2, S3.

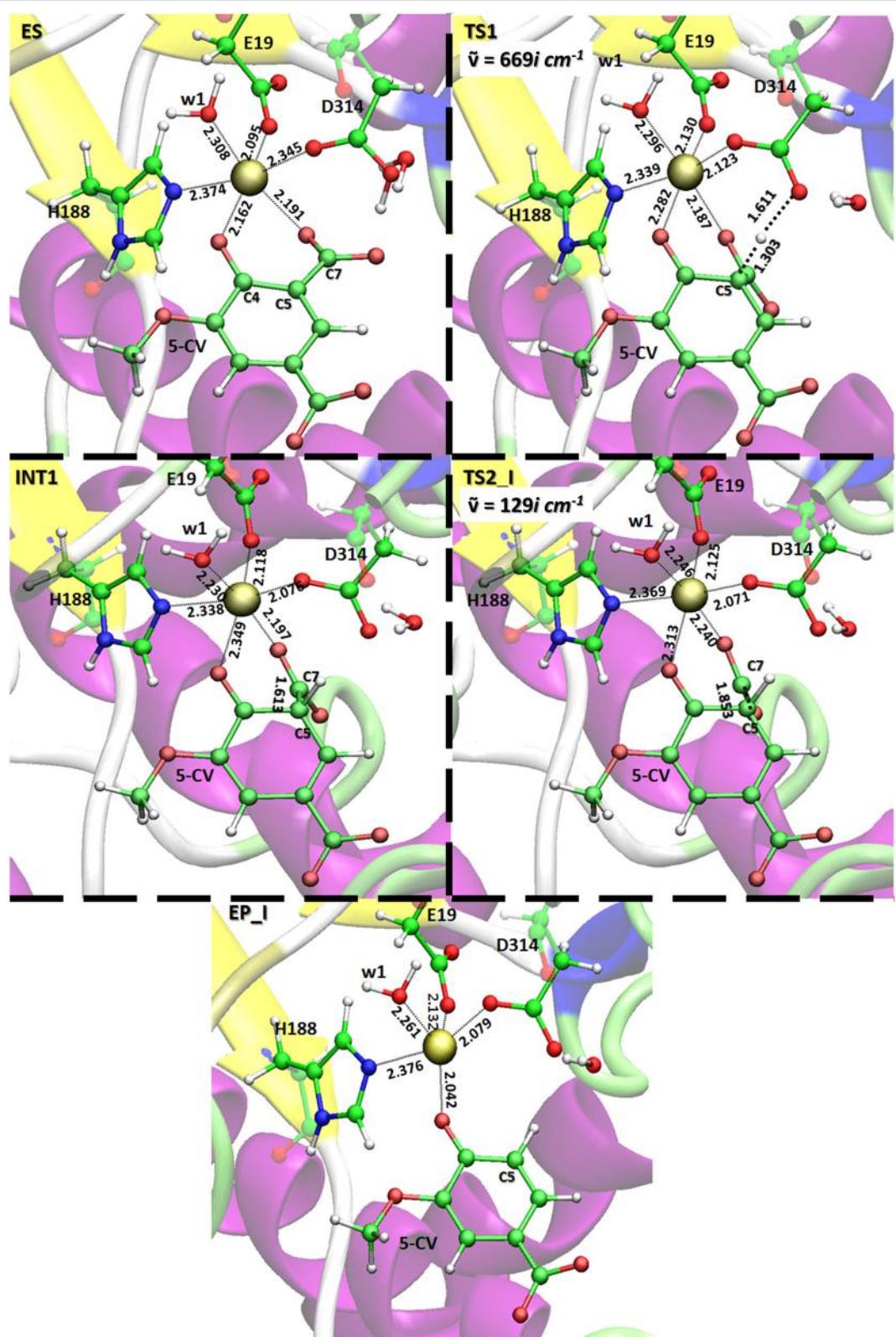
In ES complex, the Asp314, as in the original X-ray structure, is oriented in a suitable way to deliver the proton to C5 of the substrate ( $\text{H}_{\text{Asp314}}\text{-C5}$  3.103 Å). **w2** and **w3** water

molecules originally bonded to the metal ion and displaced upon the substrate entrance, lie in proximity to the reaction site establishing H- bonds network with the surrounding amino acid residues (see Figure S4). The bond lengths in the active site of the present (126 atoms) and previous (Sheng et al., 2017) larger (308 atoms) QM cluster study agree very well.

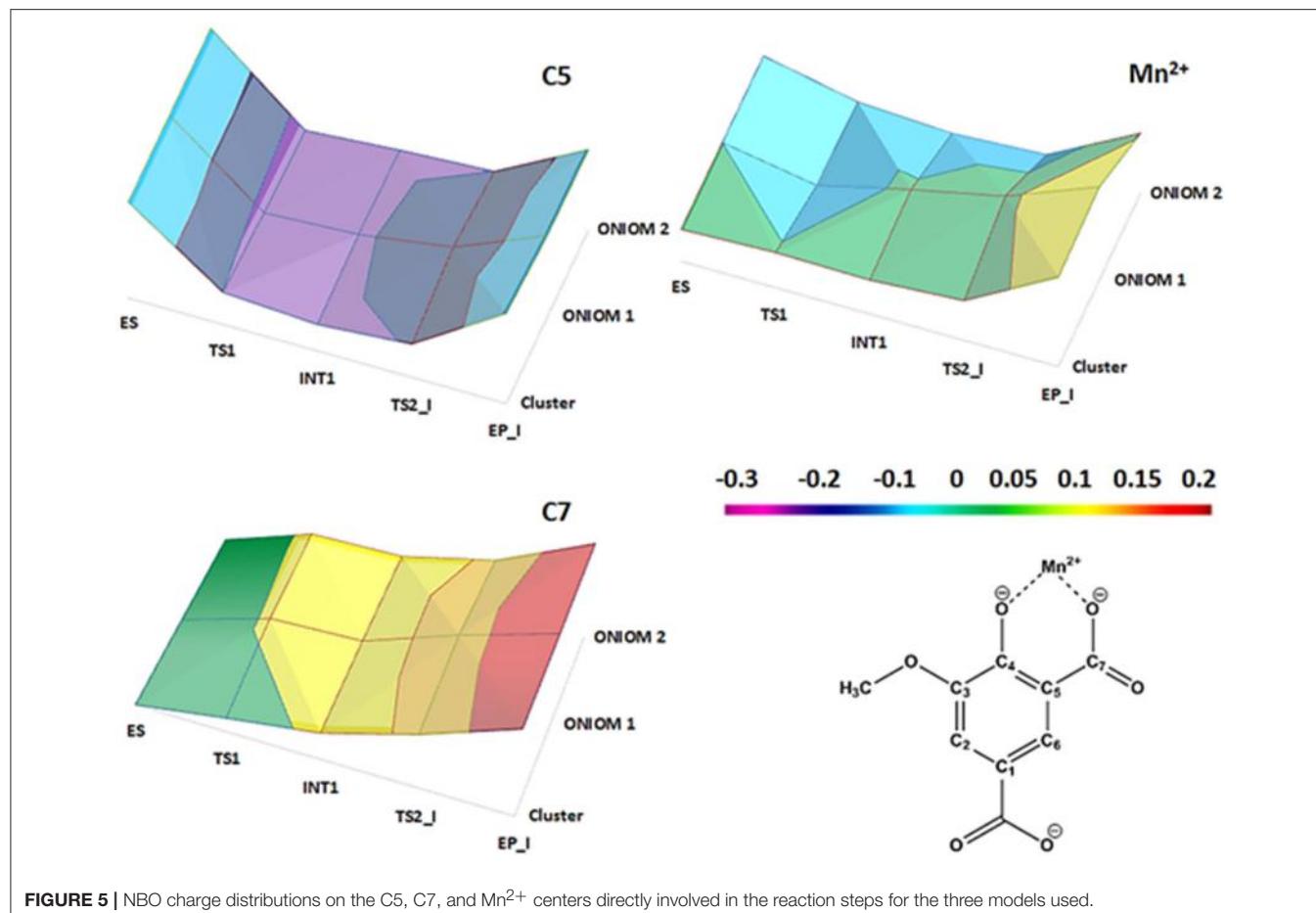
The formation of INT1 takes place through the transition state TS1 that describes the proton transfer from the Asp314 to the carbon atom of 5-CV. The related imaginary frequency ( $669i \text{ cm}^{-1}$ ) well accounts for this process since it is associated to the stretching vibrational motions of the proton transfer (O-H and H-C5). The analysis of the TS1 optimized structure (Figure 4 for ONIOM-2 and Figure S2 for QM cluster) reveals that the formation of the C5-H bond (1.303 Å) is more advanced in the case of ONIOM-2 calculation. In fact, the breaking bond between hydrogen and oxygen of Asp314 (1.611 Å) is more elongated than the usual  $\text{sp}^3$  O-H bond. Furthermore, a major distortion of the  $-\text{COO}^-$  moiety out of plane of the phenyl ring of the substrate can be observed (76 degrees in ONIOM-2 vs. 19 degrees in QM cluster). These geometrical differences may be responsible from the slight variations in the TS1 barrier (14.7 kcal/mol and 16.3 kcal/mol for ONIOM-2 and QM cluster, respectively). INT1 (Figure 4) is characterized by a C5-C7 single bond with a distance slightly elongated (1.613 Å) with respect to the single canonical bond (C-C) and a  $\text{sp}^3$  C5 hybridized prone for the subsequent decarboxylation step. The barrier for the  $\text{CO}_2$  formation (TS2\_I) is calculated to be 13.4 kcal/mol above ES complex, (only 6 kcal/mol relative to the INT1). The present QM cluster model obtains this barrier as 15.1 kcal/mol, analogous to the result (14.4 kcal/mol) of the previous (Sheng et al., 2017) cluster study with larger QM size.

The TS2\_I is characterized by the C5-C7 distance of 1.853 Å associated with a relative imaginary C-C stretching frequency of  $129i \text{ cm}^{-1}$  (Figure 4). The already formed carbon dioxide is still coordinated to the metal ion (2.240 Å) and the manganese ion is still hexa-coordinate in octahedral geometry fashion (Figure 4). This topology is present in all our used models and in the previous larger QM cluster. (Sheng et al., 2017) At the end of the decarboxylation process, one molecule of carbon dioxide is released and the EP\_I complex is generated (see Figure 4). The manganese ion assumes a trigonal bipyramidal geometry due to the loss of the sixth ligand ( $\text{CO}_2$ ). The created vacancy will be filled by one of the two water molecules present in active site (**w2** and **w3**) and essential to restore the catalytic cycle. ONIOM-2 offers a better value of the reaction energy (0.2 kcal/mol below the ES complex, see Figure 3) while at QM level it is exergonic ( $-3.5$  kcal/mol, see Figure 3). In order to verify the role of the bulk potential on the cluster model, single point computations were performed on the previous optimized structures removing all the environmental effects. Results, reported in Figure 3, show that the PES behavior is almost retained. The largest effect ( $-2.3$  kcal/mol) concerns the INT1 species.

NBO charges trend illustrated in Figure 5 confirms the nonoxidative nature of the decarboxylation process as evinced



**FIGURE 4 |** Optimized QM layer in ball and sticks (ONIOM 2 model) of all species present on the potential energy surface obtained with the B3LYP/6-31G(d,p)|SDD:FF99SB level. For clarity, only the amino acid residues of the inner coordination shell of the metal center are retained. The distances are indicated in Å. Imaginary frequencies of the transition states are also reported.



**FIGURE 5 |** NBO charge distributions on the C5, C7, and  $\text{Mn}^{2+}$  centers directly involved in the reaction steps for the three models used.

from the average value of the charges of the  $\text{Mn}^{2+}$  ( $1.117\text{ e}$ ), the C5 ( $-0.417\text{ e}$ ) and the C7 ( $0.918\text{ e}$ ) atoms in all the species intercepted on the PES. From the Figure S5, it can be also evidenced that the nonbonded interactions (characterizing the amino acid residues of the inner coordination shell with the metal ion) as well as the stacking interactions between the substrate (product) and Tyr317 are retained during the reaction.

All the models propose the **TS1** which describes the formation of the C5-protonated intermediate, as the rate limiting step ( $14.7\text{ kcal/mol}$ ). Based on the experimental  $k_{\text{cat}}$  value of  $27\text{ s}^{-1}$  for *Sphingomonas paucimobilis* LigW (Sheng et al., 2017), the reaction barrier is expected to be  $\sim 16\text{ kcal/mol}$ . The closeness of the experimental estimate of the reaction barrier and computational **TS1** barrier suggest the appropriateness of the present and previous computational protocols.

The optimized species intercepted along the PES for the bicarbonate release (step II) are shown in Figure S6. The **w3** molecule comes into play in the reaction since it performs a nucleophilic attack on the carbon (C7) ( $\text{O}_{\text{w3}} - \text{C7}$  distance of  $1.944\text{ \AA}$ ) for generating the  $\text{HCO}_3^-$  species and simultaneously donating a proton to Asp314 ( $\text{H}_\text{w} - \text{O}_{\text{Asp314}}$  distance of  $1.532\text{ \AA}$ ). The obtained energy barrier is  $30.3\text{ kcal/mol}$  (see Table S1).

## CONCLUSION

In this work, we have investigated the reaction mechanism of LigW by using three different models and two exchange correlation functionals. This allowed us to assess the influence of the employed model on the computed structures and energetics compared with available experimental data. The models include full structure, its partial solvation, and reactive center with the rest represented by a bulk potential including geometrical restraints at the border. Since the results of these three models, previous larger QM cluster and experimental studies are consistent to each other, the amino acids and waters outside the reactive center act on the reaction energetics in an average way for the present enzyme system. A similar behavior was also observed in many other enzymes (Himo, 2006, 2017; Blomberg et al., 2014). However, one should keep in mind that every enzyme system acts differently and thus one should avoid the generalization of the result despite its validity on a large variety of enzyme systems.

## AUTHOR CONTRIBUTIONS

MP, TM, and NR have analyzed the results, edit and reviewed equally the manuscript. MP, TM, and NR approved it for publications.

## ACKNOWLEDGMENTS

Financial support from the Università degli Studi della Calabria -Dipartimento di Chimica e Tecnologie Chimiche (CTC) is acknowledged.

## REFERENCES

- Alberto, M. E., Marino, T., Ramos, M. J., and Russo, N. (2010). Atomistic details of the Catalytic Mechanism of Fe(III)-Zn(II) Purple Acid Phosphatase. *J. Chem. Theory Comput.* 6, 2424–2433. doi: 10.1021/ct00187c
- Amata, O., Marino, T., Russo, N., and Toscano, M. (2011a). Catalytic activity of a  $\zeta$ -class zinc and cadmium containing carbonic anhydrase. Compared work mechanisms. *Phys. Chem. Chem. Phys.* 13, 3468–3477. doi: 10.1039/c0cp01053g
- Amata, O., Marino, T., Russo, N., and Toscano, M. (2011b). A proposal for mitochondrial processing peptidase catalytic mechanism. *J. Am. Chem. Soc.* 133, 17824–17831. doi: 10.1021/ja207065v
- Andrae, D., Haußermann, H., Dolg, M., Stoll, H., and Preuß, H. (1990). Energy-adjusted *ab initio* pseudopotentials for the second and third row transition elements. *Theor. Chim. Acta* 77, 123–141. doi: 10.1007/BF0114537
- Bayly, C. I., Cieplak, P., Cornell, W. D., and Kollman, P. A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* 97, 10269–10280. doi: 10.1021/j100142a004
- Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98, 5648–5652. doi: 10.1063/1.464913
- Blomberg, M. R. A., Borowski, T., Himo, F., Liao, R. Z., and Siegbahn, P. E. M. (2014). Quantum chemical studies of mechanisms for metalloenzymes. *Chem. Rev.* 114, 3601–3658. doi: 10.1021/cr400388t
- Duarte, M., Ferreira-da-Silva, F., Lünsdorf, H., Junca, H., Gales, L., Pieper, D. H., et al. (2011). *Gulosisibacter molinativorax* ON4T molinate hydrolyase, a novel cobalt-dependent amidohydrolase. *J. Bacteriol.* 193, 5810–5816. doi: 10.1128/JB.05054-11
- Elias, M., Dupuy, J., Merone, L., Mandrich, L., Porzio, E., Moniot, S., et al. (2008). Structural basis for natural lactonase and promiscuous phosphotriesterase activities. *J. Mol. Biol.* 379, 1017–1028. doi: 10.1016/j.jmb.2008.04.022
- Faponle, A. S., Quesne, M. G., and de Visser, S. P. (2016). Origin of the regioselective fatty-acid hydroxylation versus decarboxylation by a cytochrome P450 peroxygenase: what drives the reaction to biofuel production? *Chem. Eur. J.* 22, 5478–5483. doi: 10.1002/chem.201600739
- Gerlt, J. A., and Babbitt, P. C. (2001). Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu. Rev. Biochem.* 70, 209–246. doi: 10.1146/annurev.biochem.70.1.209
- Grimme, S., Ehrlich, S., and Goerigk, L. (2011). Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* 32, 1456–1465. doi: 10.1002/jcc.21759
- Himo, F. (2006). Quantum chemical modeling of enzyme active sites and reaction mechanisms. *Theor. Chem. Acc.* 116, 232–240. doi: 10.1007/s00214-005-0012-1
- Himo, F. (2017). Recent trends in quantum chemical modeling of enzymatic reactions. *J. Am. Chem. Soc.* 139, 6780–6786. doi: 10.1021/jacs.7b02671
- Jackson, C. J., Liu, J. W., Coote, M. L., and Ollis, D. L. (2005). The effects of substrate orientation on the mechanism of a phosphotriesterase. *Org. Biomol. Chem.* 3, 4343–4350. doi: 10.1039/b512399b
- Khurana, J. L., Jackson, C. J., Scott, C., Pandey, G., Horne, I., Russell, R. J., et al. (2009). Characterization of the phenylurea hydrolases A and B: founding members of a novel amidohydrolase subgroup. *Biochem. J.* 418, 431–441. doi: 10.1042/BJ20081488
- Lee, C., Yang, W., and Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* 37, 785–789. doi: 10.1103/PhysRevB.37.785
- Liao, R. Z., Yu, J. G., and Himo, F. (2010). Mechanism of tungsten-dependent acetylene hydratase from quantum chemical calculations. *Proc. Natl. Acad. Sci. U.S.A.* 107, 22523–22527. doi: 10.1073/pnas.1014060108
- Liu, A., and Zhang, H. (2006). Transition metal-catalyzed nonoxidative decarboxylation reactions. *Biochemistry* 45, 10407–10411. doi: 10.1021/bi061031v
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713. doi: 10.1021/acs.jctc.5b00255
- Marenich, A. V., Cramer, C. J., and Truhlar, D. G. (2009). Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* 113, 6378–6396. doi: 10.1021/jp810292n
- Medina, F. E., Neves, R. P. P., Ramos, M. J., and Fernandes, P. A. (2017). A QM/MM study of the reaction mechanism of human  $\beta$ -ketoadyl reductase. *Phys. Chem. Chem. Phys.* 19, 347–355. doi: 10.1039/C6CP07014K
- Merz, K. M. (2014). Using quantum mechanical approaches to study biological systems. *Acc. Chem. Res.* 47, 2804–2811. doi: 10.1021/ar5001023
- Piazzetta, P., Marino, T., Russo, N., and Salahub, D. R. (2017). The role of metal substitution in the promiscuity of natural and artificial carbonic anhydrases. *Coord. Chem. Rev.* 345, 73–85. doi: 10.1016/j.ccr.2016.12.014
- Prejanò, M., Marino, T., Rizzuto, C., Madrid, J. C. M., Russo, N., and Toscano, M. (2017a). Reaction mechanism of low-spin iron(III)- and cobalt(III)-containing nitrile hydratases: a quantum mechanics investigation. *Inorg. Chem.* 56, 13390–13400. doi: 10.1021/acs.inorgchem.7b02121
- Prejanò, M., Marino, T., and Russo, N. (2017b). How can methanol dehydrogenase from *Methylacidiphilum fumariolicum* work with the alien Ce<sup>III</sup> Ion in the active center? A theoretical study. *Chem. Eur. J.* 23, 8652. doi: 10.1002/chem.201700381
- Quesne, M. G., Borowski, T., and de Visser, S. P. (2016). Quantum mechanics/molecular mechanics modeling of enzymatic processes: caveats and breakthroughs. *Chem. Eur. J.* 22, 2562–2581. doi: 10.1002/chem.201503802
- Ramos, M. J., and Fernandes, P. A. (2008). Computational enzymatic catalysis. *Acc. Chem. Res.* 41, 689–698. doi: 10.1021/ar7001045
- Ryde, U. (2016). QM/MM calculations on proteins. *Methods Enzymol.* 577, 119–158. doi: 10.1016/bs.mie.2016.05.014
- Ryde, U. (2017). How many conformations need to be sampled to obtain converged QM/MM energies? The Curse of Exponential Averaging. *J. Chem. Theory Comput.* 13, 5745–5752. doi: 10.1021/acs.jctc.7b00826
- Seibert, C. M., and Raushel, F. M. (2005). Structural and catalytic diversity within the amidohydrolase superfamily. *Biochemistry* 44, 6383–6391. doi: 10.1021/bi047326v
- Senn, H. M., and Thiel, W. (2009). QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.* 48, 1198–1229. doi: 10.1002/anie.200802019
- Shapir, N., Pedersen, C., Gil, O., Strong, L., Seffernick, J., Sadowsky, M. J., et al. (2006). TrzN from *Arthrobacter aurescens* TC1 Is a zinc amidohydrolase. *J. Bacteriol.* 188, 5859–5864. doi: 10.1128/JB.00517-06
- Sheng, X., Zhu, W., Huddleston, J., Xiang, D. F., Raushel, F. M., Richards, N. G. J., et al. (2017). A combined experimental-theoretical study of the LigW-catalyzed decarboxylation of 5-carboxyvanillate in the metabolic pathway for lignin degradation. *ACS Catal.* 7, 4968–4974. doi: 10.1021/acscatal.7b01166
- Siegbahn, P. E. M., and Himo, F. (2011). The quantum chemical cluster approach for modeling enzyme reactions. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 1, 323–336. doi: 10.1002/wcms.13
- Svensson, M., Humbel, S., Froese, R. D. J., Matsubara, T., Sieber, S., and Morokuma, K. (1996). ONIOM: a multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. a test for diels-alder reactions and Pt(P*t*-Bu)<sub>3</sub>)<sub>2</sub> + H<sub>2</sub> oxidative addition. *J. Phys. Chem.* 100, 19357–19363. doi: 10.1021/jp962071j
- Tobimatsu, Y., Chen, F., Nakashima, J., Escamilla-Treviño, L. L., Jackson, L., Dixon, R. A., and et al. (2013). Coexistence but independent biosynthesis of catechyl and guaiacyl/syringyl lignin polymers in seed coats. *Plant. Cell.* 25, 2587–2600. doi: 10.1105/tpc.113.113142
- Vladimirova, A., Patskovsky, Y., Fedorov, A. A., Bonanno, J. B., Fedorov, E. V., Toro, R., et al. (2016). Substrate distortion and the catalytic reaction

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2018.00249/full#supplementary-material>

- mechanism of 5-carboxyvanillate decarboxylase. *J. Am. Chem. Soc.* 138, 826–836. doi: 10.1021/jacs.5b08251
- Vreven, T., Byun, K. S., Komáromi, I., Dapprich, S., Montgomery, J. A. Jr., Morokuma, K., et al. (2006). Combining quantum mechanics methods with molecular mechanics methods in ONIOM. *J. Chem. Theory Comput.* 2, 815–826. doi: 10.1021/ct050289g
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comp. Chem.* 25, 1157–1174. doi: 10.1002/jcc.20035
- Warshel, A. (1991). *In Computer Modeling of Chemical Reactions in Enzymes and Solutions*. New York, NY: Wiley.
- Warshel, A., and Levitt, M. (1976). Theoretical studies of enzymatic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* 103, 227–249. doi: 10.1016/0022-2836(76)90311-9
- Zhao, Y., and Truhlar, D. G. (2006). A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.* 125, 1–18. doi: 10.1063/1.2370993
- Zhao, Y., and Truhlar, D. G. (2008). The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* 120, 215–241. doi: 10.1007/s00214-007-0310-x

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Prejanò, Marino and Russo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Interfacing the Core–Shell or the Drude Polarizable Force Field With Car–Parrinello Molecular Dynamics for QM/MM Simulations

Sudhir K. Sahoo<sup>†</sup> and Nisanth N. Nair<sup>\*</sup>

Department of Chemistry, Indian Institute of Technology Kanpur, Kanpur, India

## OPEN ACCESS

### Edited by:

Sam P. De Visser,  
University of Manchester,  
United Kingdom

### Reviewed by:

Thomas S. Hofer,  
University of Innsbruck, Austria  
Hugo Gattuso,  
University of Liège, Belgium

### \*Correspondence:

Nisanth N. Nair  
nnair@iitk.ac.in

### †Present Address:

Sudhir K. Sahoo,  
Department Chemie, Universität  
Paderborn, Paderborn, Germany

### Specialty section:

This article was submitted to  
Theoretical and Computational  
Chemistry,  
a section of the journal  
*Frontiers in Chemistry*

Received: 22 February 2018

Accepted: 18 June 2018

Published: 10 July 2018

### Citation:

Sahoo SK and Nair NN (2018)  
Interfacing the Core–Shell or the Drude  
Polarizable Force Field With  
Car–Parrinello Molecular Dynamics for  
QM/MM Simulations.  
*Front. Chem.* 6:275.  
doi: 10.3389/fchem.2018.00275

We report a quantum mechanics/polarizable–molecular mechanics (QM/p–MM) potential based molecular dynamics (MD) technique where the core–shell (or the Drude) type polarizable MM force field is interfaced with the plane-wave density functional theory based QM force field which allows Car–Parrinello MD for the QM subsystem. In the QM/p–MM Lagrangian proposed here, the shell (or the Drude) MM variables are treated as extended degrees of freedom along with the Kohn–Sham (KS) orbitals describing the QM wavefunction. The shell and the KS orbital degrees of freedom are then adiabatically decoupled from the nuclear degrees of freedom. In this respect, we also present here the Nosé–Hoover Chain thermostat implementation for the dynamical subsystems. Our approach is then used to investigate the effect of MM polarization on the QM/MM results. Especially, the consequence of MM polarization on reaction free energy barriers, defect formation energy, and structural and dynamical properties are investigated. A low point charge polarizable potential (p–MZHB) for pure siliceous systems is also reported here.

**Keywords:** QMMM simulations, MD, POLARIZED MM, catalysis, CPMD-GULP

## 1. INTRODUCTION

Hybrid quantum mechanical/molecular mechanical (QM/MM) calculations offer a powerful way to bridge the length scales in a chemically complex system where a small region of the system of interest is treated by QM techniques, while the rest is described by computationally cheap MM force-fields. Widely used MM force fields employ a fixed point charge model for accounting the electrostatic interactions between MM atoms. The QM/MM implementations with fixed charge MM models enable polarization of QM charge density due to MM electrostatic potential. However, such approaches cannot take into account the polarization of MM atoms due to the QM electrostatic potential. Inclusion of polarization of MM atoms in QM/MM simulations demands usage of polarizable MM force fields, i.e., QM/polarized-MM (QM/p–MM) methods. It was reported that inclusion of polarization of MM atoms has significant effects on various properties (Bakowies and Thiel, 1996; Illingworth et al., 2006; Geerke et al., 2007; Lu and Zhang, 2008; Boulanger and Thiel, 2014),

for instance, free energy barriers of chemical reactions are affected by about 10% (Lu and Zhang, 2008; Boulanger and Thiel, 2014).

The shell model (Dick and Overhauser, 1958) (or the Drude oscillator model) is widely used to describe polarization of MM atoms. Molecular dynamics (MD) simulations with the shell model based MM force fields can be carried out in two ways. In the conventional scheme, the position of the shells are minimized (Sangster and Dixon, 1976; Lindan and Gillan, 1993) at every MD step while keeping the core coordinates fixed. In an alternative scheme, the shells are treated as extended degrees of freedom and these are propagated classically to avoid minimization of their positions (Sprik, 1991; Mitchell and Fincham, 1993; Wilson and Madden, 1993) in the spirit of the Car-Parrinello MD method (Car and Parrinello, 1985). Often, the shell variables are assigned a mass smaller compared to that of the nuclear masses. Due to this reason, a smaller time step than used in a conventional MD is required for this approach. In practice, the shell temperature is kept close to 0 K, and most importantly, dynamics of shell degrees of freedom is made adiabatically decoupled from the rest of the system.

Different QM/MM MD schemes have been proposed to interfere the shell model with *ab initio* methods (Sulimov et al., 2002; Nasluzov et al., 2003; Woodcock et al., 2007; Geerke et al., 2007; Lu and Zhang, 2008; Lev et al., 2010; Boulanger and Thiel, 2012; Rowley and Roux, 2012; Boulanger and Thiel, 2014; Riahi and Rowley, 2014). Ideally, wavefunction of the QM subsystem and positions of shells are minimized at every MD step. In the approach by Lu and Zhang (2008) the positions of the shells were either minimized or updated only once in every MD step. The shell variables are treated as extended degrees of freedom in the QM/MM scheme proposed by Boulanger and Thiel (2012). Similar method was also reported by Rowley and Roux (2012). Recently Loco et al. (2016, 2017) have presented a QM/MM coupling method using the AMOEBA polarizable force fields. Here, SCF (self consistent field) calculations were carried out for the induced dipoles, while either SCF or the extended Lagrangian variant of Born-Oppenheimer MD (Niklasson et al., 2009) technique was employed for the wavefunction update. Incorporating the shell model in the extended Lagrangian scheme for Car-Parrinello MD within a QM/MM implementation is, however, not straightforward, and has not been attempted before, to the best of our knowledge.

In this paper we present an extended Lagrangian scheme to carry out Car-Parrinello MD for the QM subsystem which is coupled to a polarizable shell model based MM force field. First, we discuss the theory and the technical details of our method. Next, the implementation is validated by taking a system of water cluster composed of five water molecules. A new polarizable MM potential for silica with low point charges is then developed. Using our implementation and this new force-field for silica, we study three problems: (a) Oxygen vacancy in  $\alpha$ -cristobalite silica; (b) Hydrogenation of ethene catalyzed by Rh cluster supported in Y-zeolite; (c) Proton exchange between methane and H-ZSM-5 zeolite.

## 2. THEORY

### 2.1. Formulation of the Extended Lagrangian QM/p-MM Method

The Lagrangian for the conventional QM/MM Car-Parrinello MD simulation is,

$$\mathcal{L}_{\text{CP/QMMM}}(\mathbf{R}, \dot{\mathbf{R}}, \phi, \dot{\phi}) = \sum_I \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 + \sum_i \frac{1}{2} \mu \langle \dot{\phi}_i | \dot{\phi}_i \rangle - E_{\text{KS}}(\mathbf{R}, \phi) - E_{\text{MM}}(\mathbf{R}) - E_{\text{QM/MM}}(\mathbf{R}, \phi) + \sum_{i,j} \Delta_{ij} (\langle \phi_i | \phi_j \rangle - \delta_{ij}), \quad (1)$$

where  $\{M_I\}$  and  $\{\mu_i\}$  are the masses of the ionic and the orbital degrees of freedom, respectively, and  $\{\mathbf{R}_I\}$  and  $\{\phi_i\}$  are the nuclear coordinates (in Cartesian) and the Kohn-Sham orbitals, respectively. Here,  $E_{\text{KS}}$ ,  $E_{\text{MM}}$ , and  $E_{\text{QM/MM}}$  are the energy of the QM subsystem, the energy of the MM subsystem and the energy due to QM-MM non-bonding interactions, respectively. The last term in the Lagrangian invokes orthonormality constraints during the classical evolution of the Kohn-Sham orbitals (Marx and Hutter, 2009). For details of this implementation, see Lai et al. (2002) and Sahoo and Nair (2016).

In the case of QM/p-MM implementation, MM polarization is included by augmenting the degrees of freedom by the shells  $\{\mathbf{r}_k\}$ , which are connected to a selected set of polarizable atoms  $\mathcal{P}$ . We propose the QM/p-MM Lagrangian,

$$\mathcal{L}_{\text{CP/Shell}}(\mathbf{R}, \dot{\mathbf{R}}, \mathbf{r}, \dot{\mathbf{r}}, \phi, \dot{\phi}) = \sum_I \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 + \sum_i \frac{1}{2} \mu \langle \dot{\phi}_i | \dot{\phi}_i \rangle + \sum_{k \in \mathcal{P}} \frac{1}{2} m_k \dot{\mathbf{r}}_k^2 - E_{\text{KS}}(\mathbf{R}, \phi) - E_{\text{MM}}(\mathbf{R}, \mathbf{r}) - E_{\text{QM/MM}}(\mathbf{R}, \mathbf{r}, \phi) + \sum_{i,j} \Delta_{ij} (\langle \phi_i | \phi_j \rangle - \delta_{ij}), \quad (2)$$

where,  $n_s$  is the number of shells, and  $m_k$  is the fictitious mass of a shell  $k$ . Also,

$$E_{\text{MM}}(\mathbf{R}, \mathbf{r}) = E_b(\mathbf{R}, \mathbf{r}) + E_{\text{nb}}(\mathbf{R}, \mathbf{r}) + \sum_k^{n_s} \frac{1}{2} \kappa_k s_k^2, \quad (3)$$

where

$$s_k = |\mathbf{s}_k| = |\mathbf{R}_I - \mathbf{r}_k|,$$

having the shell  $k$  harmonically bound to the core atom  $I \in \mathcal{P}$ . Here  $E_b$  refers to the sum of all the bonding terms in the MM potential, which is conventionally defined over the shell variables. The total non-bonding interaction energy,  $E_{\text{nb}}$ , is the sum of the dispersive and the electrostatic interactions within the MM subsystem. The dispersive interactions are defined over the shell degrees of freedom, while the electrostatic interactions span over the cores and the shells degrees of freedom. Further,

$$E_{\text{QM/MM}}(\mathbf{R}, \mathbf{r}) = E'_b(\mathbf{R}, \mathbf{r}) + E'_{\text{vdw}}(\mathbf{R}, \mathbf{r}) - \sum_I q_I^c \int d\bar{\mathbf{r}} \rho(\bar{\mathbf{r}}) \frac{1}{|\mathbf{R}_I - \bar{\mathbf{r}}|} - \sum_k q_k^s \int d\bar{\mathbf{r}} \rho(\bar{\mathbf{r}}) \frac{1}{|\mathbf{r}_k - \bar{\mathbf{r}}|}, \quad (4)$$

where  $E'_b$  and  $E'_{\text{vdw}}$  are the energy contributions due to the bonding and the dispersive interactions between the QM and the MM atoms, respectively. The last two terms in the above equation account for the electrostatic interaction between the point charges of the core ( $\{q_I^c\}$ ) and the shell ( $\{q_k^s\}$ ) degrees of freedom with the electronic density  $\rho(\bar{\mathbf{r}})$ , respectively. Electrostatic interactions are computed in the real space with the modified Coulomb kernel as in Laio et al. (2002).

For the success of this approach it is crucial that the Lagrangian in Equation (2) leads to dynamics close to that on the Born–Oppenheimer surface. This is taken care by starting the MD simulations with the optimized  $\{\phi\}$  and  $\{\mathbf{r}\}$ , while maintaining the temperatures of the orbitals ( $T_\phi$ ) and the shells ( $T_s$ ) degrees of freedom close to zero, considering

$$\min_{\{\phi, \mathbf{r}\}} \lim_{T_\phi, T_s \rightarrow 0} \mathcal{L}_{\text{CP/Shell}}(\mathbf{R}, \dot{\mathbf{R}}, \mathbf{r}, \dot{\mathbf{r}}, \phi, \dot{\phi}) \rightarrow \mathcal{L}(\mathbf{R}, \dot{\mathbf{R}}). \quad (5)$$

The physical temperature ( $T_{\text{phys}}$ ) is defined as,

$$T_{\text{phys}} = \frac{1}{N_f k_B} \sum_I \mathcal{M}_I \dot{\mathbf{S}}_I^2, \quad (6)$$

while the shell temperature is defined as,

$$T_s = \frac{1}{3n_s k_B} \sum_k \bar{m}_k \dot{\mathbf{s}}_k^2. \quad (7)$$

In the above equations,  $\dot{\mathbf{S}}_I$  is the velocity of the center of mass of a core–shell pair ( $I, k$ ), and  $\dot{\mathbf{s}}_k$  is the relative velocity of the shell  $k$  connected to a core atom  $I$ . Here,

$$\dot{\mathbf{S}}_I = \frac{1}{\mathcal{M}_I} (M_I \mathbf{R}_I + m_k \mathbf{r}_k),$$

$\mathcal{M}_I$  and  $\bar{m}_k$  are the total mass of a core–shell pair ( $I, k$ ), and the reduced mass of a shell  $k$ , respectively:

$$\mathcal{M}_I = M_I + m_k$$

$$\bar{m}_k = \frac{M_I m_k}{M_I + m_k}$$

In the above,  $k_B$  is the Boltzmann constant,  $N_f$  is the total nuclear degrees of freedom and  $n_s$  is the total number of shell variables, respectively. We have implemented this method in the CPMD/GULP QM/MM interface program, as developed in Sahoo and Nair (2016), where the plane wave density functional theory (DFT) based CPMD (CPMD, 132) code is interfaced with the MM based GULP (Gale, 1997) program.

At this stage, the following points are noted:

- Both the orbital and the shell degrees of freedom have to be adiabatically separated from the nuclear degrees of freedom. While the nuclei could be kept hot, both the orbitals and the shells variables should remain cold throughout the dynamics.
- As per Equation (5), it is not critical to have an adiabatic separation between shell and orbital degrees of freedom, provided that both sets of variables remain cold.

Accordingly, we have strategized the application of our implementation. The time step of integration and the masses of both shell and orbital degrees of freedom have to be chosen such that adiabatic separation between the nuclear subsystem and the subsystem containing shells and orbitals is maintained. We choose  $m_k = \mu$  (i.e., the masses of the shell and the orbital degrees of freedom are taken to be the same), which allows us to choose the same time step for integrating the equations of motion for all the subsystems.

## 2.2. Implementation of Nosé–Hoover Chain Thermostat for Shell Dynamics

For obtaining stable dynamics and to achieve a canonical ensemble, it is crucial to couple the dynamical subsystems with thermostats. We have implemented three separate sets of Nosé–Hoover Chain (NHC) thermostats (Martyna et al., 1992). The system temperature is maintained at  $T_{\text{phys}}$  using one set of NHC thermostats whereas the shell and the orbital variables are maintained close to 0 K using two separate thermostats. We coupled the nuclear and the shell NHC thermostats to the center of mass motion and the relative motion of the core–shell pairs, and the corresponding equations of motion are given by,

$$\begin{aligned} \mathcal{M}_I \ddot{\mathbf{S}}_I &= \mathbf{F}_I^{(S)} - \mathcal{M}_I \dot{\mathbf{S}}_I \frac{p_{\eta_1}}{q_1} \\ \dot{p}_{\eta_1} &= \sum_I \mathcal{M}_I \dot{\mathbf{S}}_I^2 - N_f k_B T_{\text{phys}} - \frac{p_{\eta_2}}{q_2} p_{\eta_1} \\ \dot{p}_{\eta_j} &= \frac{p_{\eta_{j-1}}^2}{q_{j-1}} - k_B T_{\text{phys}} - \frac{p_{\eta_{j+1}}}{q_{j+1}} p_{\eta_j}, \quad j = 2, \dots, n_c - 1, \\ \dot{p}_{\eta_j} &= \frac{p_{\eta_{j-1}}^2}{q_{j-1}} - k_B T_{\text{phys}}, \quad j = n_c, \\ \dot{\eta}_j &= \frac{p_{\eta_j}}{q_j}, \quad j = 1, \dots, n_c. \end{aligned}$$

Here  $\mathbf{F}_I^{(S)}$  is the force acting on the center of mass coordinates  $\mathbf{S}_I$ ,  $\{p_{\eta_i}\}$  and  $\{q_i\}$  are the momentum and the masses of the thermostat variables  $\eta$ . Number of thermostat variables,  $n_c$ , is chosen to be more than one.

In order to thermostat the shell dynamics, we write the shell equations of motion in relative coordinates as,

$$\begin{aligned} \bar{m}_k \ddot{\mathbf{s}}_k &= \mathbf{f}_k^{(s)} - \bar{m}_k \dot{\mathbf{s}}_k \frac{p_{\eta_1^*}}{q_1^*} \\ \dot{p}_{\eta_1^*} &= \sum_k \bar{m}_k \dot{\mathbf{s}}_k^2 - 3n_s k_B T_s - \frac{p_{\eta_2^*}}{q_2^*} p_{\eta_1^*} \\ \dot{p}_{\eta_j^*} &= \frac{p_{\eta_{j-1}^*}^2}{q_{j-1}^*} - k_B T_s - \frac{p_{\eta_{j+1}^*}}{q_{j+1}^*} p_{\eta_j^*}, \quad j = 2, \dots, n_c^* - 1, \end{aligned}$$

$$\dot{p}_{\eta_j^*} = \frac{p_{\eta_{j-1}^*}^2}{q_{j-1}^*} - k_B T_s, \quad j = n_c^*,$$

$$\dot{\eta}_j^* = \frac{p_{\eta_j}}{q_j^*}, \quad j = 1, 2, \dots, n_c^*.$$

Here  $\mathbf{f}_k^{(s)}$  is the force acting on the relative coordinate  $\mathbf{s}_k$  of a core-shell pair  $(I, k)$ . The thermostat variables for the relative motion are  $\{\eta_j^*\}$ , having masses  $\{q_j^*\}$ , and their conjugate momenta are given by  $\{p_{\eta_j^*}\}$ . In our practical implementation, we transform the Cartesian coordinates of a core-shell pair to the corresponding relative and the center of mass coordinates at every MD steps. Transformation of forces for a core-shell pair  $(I, k)$  is achieved by,

$$\mathbf{F}_I^{(s)} = \mathbf{F}_I + \mathbf{f}_k^{(r)},$$

$$\mathbf{f}_I^{(s)} = \frac{1}{M_I} \left( m_k \mathbf{F}_I - M_I \mathbf{f}_k^{(r)} \right).$$

Here,  $\mathbf{F}_I$  and  $\mathbf{f}_k^{(r)}$  are the Cartesian forces on an atom (or core)  $I$  and on a shell  $k$  connected to core  $I$ , respectively. A similar coordinate transformation from cartesian to normal mode and vice-versa was reported by Marx et al. (1999) and can be also found in the work of Lamoureux and Roux (2003).

### 3. LOW POINT CHARGE POLARIZABLE FORCE FIELD FOR SILICA

Herein, the previously reported low point charge potential (MZHB) for silica (Sahoo and Nair, 2015) is further extended to include polarization of O atoms using the core shell model. This new potential is termed as p-MZHB hereafter. We have re-parameterized  $k_\theta$  of O-Si-O and Si-O-Si angles and optimized the core-shell coupling parameters  $\{\kappa_k\}$ . During the parameterization, point charges of the core and the shell of O atoms were allowed to vary, while their sum was fixed to  $-0.35\text{ e}$ . The re-parameterization was done using the GULP (Gale, 1997) program to reproduce the experimental structure of  $\alpha$ -Quartz (Jorgensen, 1978). The final set of parameters are given in Table 1. The detailed validation of the p-MZHB potential is discussed in the Supporting Information.

## 4. RESULTS AND DISCUSSION

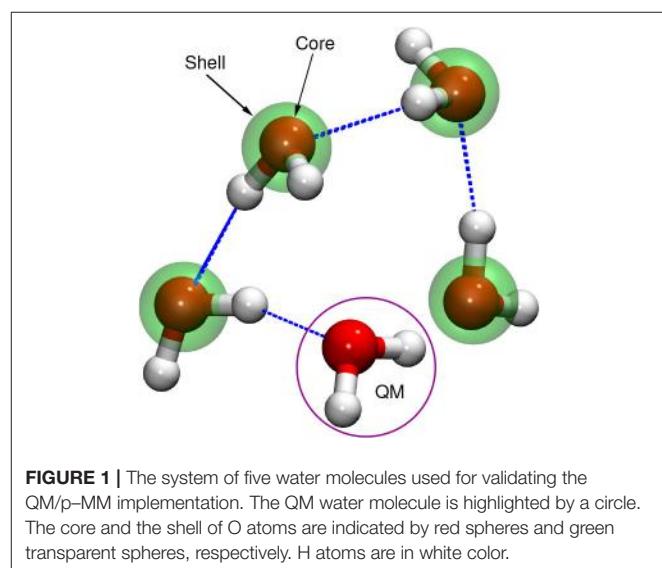
### 4.1. Validation of the Implementation in CPMD/GULP Interface Program

Total energy conservation during MD runs is tested to validate the method and the implementation. For the benchmarking purpose, we used  $1\text{H}_2\text{O}(\text{QM})+4\text{H}_2\text{O}(\text{p-MM})$  system (Figure 1), and carried out NVE MD runs for 36 ps using DFT/PBE (Perdew et al., 1996) to describe the QM subsystem and polarizable de Leeuw-Parker (de Leeuw and Parker, 1998) potential to treat the MM subsystem; see Supporting Information for other technical details. In this MM potential, shells are added on the oxygen atoms of the water molecules.

**TABLE 1 |** The p-MZHB force field parameters for siliceous zeolites.

	$k_r(\text{eV}/\text{\AA}^2)$	$r_0(\text{\AA})$
Si–O	23.3	1.62
	$k_\theta(\text{eV}/\text{rad}^2)$	$\theta_0(\text{^\circ})$
O–Si–O	6.061057	109.4
Si–O–Si	1.766554	149.8
	Species	Charge (e)
Si	core	0.70
O	core	1.387258
O	shell	-1.737258
	Lennard–Jones	
	$\epsilon(\text{eV})$	$\sigma^0(\text{\AA})$
Si	0.00864	2.200
O	0.00324	1.770
	Species	$\kappa(\text{eV}/\text{\AA}^2)$
O		99.4732

For the cross Lennard–Jones parameters, the Lorentz–Berthelot (Leach, 2010; Schlick, 2010) combination rule is applied.



**FIGURE 1 |** The system of five water molecules used for validating the QM/p-MM implementation. The QM water molecule is highlighted by a circle. The core and the shell of O atoms are indicated by red spheres and green transparent spheres, respectively. H atoms are in white color.

The total energy, the orbital kinetic energy and the shell temperature of the system were monitored; see Figure 2. The drift in total energy is only of the order of  $10^{-7}\text{ a.u. atom}^{-1}\text{ ps}^{-1}$ , indicating that the total energy conservation is fairly good. The plot of kinetic energy of orbitals shows only a small long-time drift, while the temperature of the shell variables are maintained at low temperature. Such small drifts in kinetic energy could be controlled by connecting the orbital degrees of freedom with a thermostat, as demonstrated below.

As next, we repeated the same simulation, but in the NVT ensemble. Here, three thermostats were added on the nuclear, the orbitals and the shells degrees of freedom, and these

were thermostated to 300 K ( $T_{\text{phys}}$ , 0.0007 Hartree and 1 K, respectively). Since the NHC thermostat possesses a conserved quantity, drift in this conserved quantity allows us to verify our implementation further. The drift in the conserved quantity is only of the order of  $10^{-7}$  a.u. atom $^{-1}$  ps $^{-1}$ , confirming the correctness of our implementation (see **Figure 3A**). The orbital kinetic energy and the shell temperature (see **Figure 3B–D**) plots clearly show that the dynamics of the extended variables is stable and is well thermostated.

## 4.2. Benchmark Studies Using $\alpha$ -Cristobalite Silica

To further benchmark the performance of the developed method, we carried out MD simulation of  $\alpha$ -cristobalite silica in the

NVT ensemble at 300 K. Here we use the DFT/PBE (Perdew et al., 1996) level of theory for the QM part and p-MZHB potential for the MM part. We used a supercell of  $8 \times 8 \times 8$  (Si<sub>2048</sub>O<sub>4096</sub>) for QM/p-MM calculations. Optimized structure and lattice parameters using the p-MZHB MM potential were used here. Multiple QM/p-MM calculations were carried out with different QM sizes: 2T (Si<sub>2</sub>O<sub>7</sub>), 8T (Si<sub>8</sub>O<sub>25</sub>), 14T (Si<sub>14</sub>O<sub>40</sub>), and 26T (Si<sub>26</sub>O<sub>67</sub>), where T stands for SiO<sub>4</sub> tetrahedral unit (see **Figure 4**).

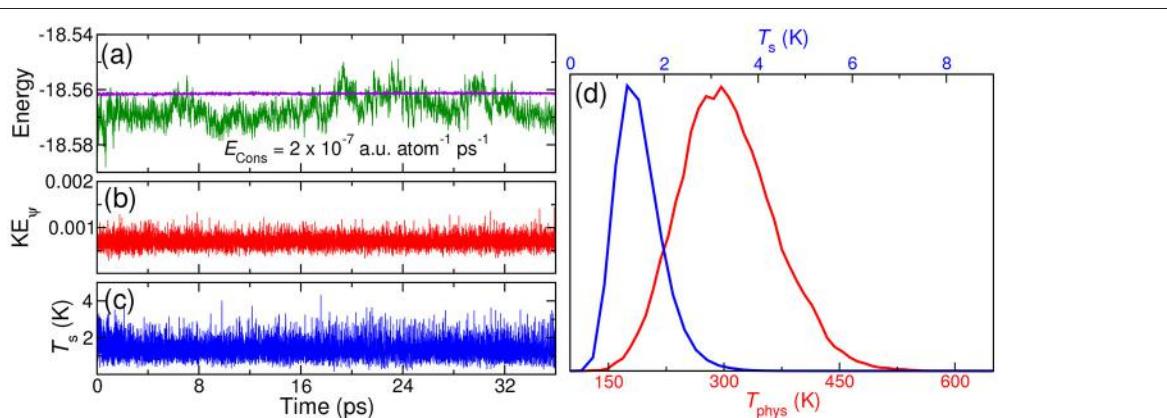
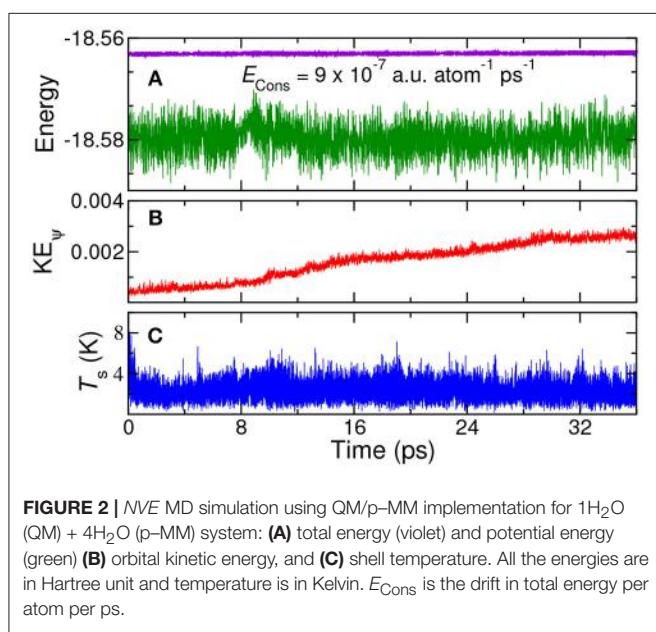
The structure (inner QM atoms only) obtained from QM/p-MM simulation was compared with QM data (“all-QM”) and MM data (“all-MM”) data; see **Table 2**. Difference between “all-QM” and QM/p-MM data for Si–O bond length is only 0.01 Å, O–Si–O angle is only 0.2° and Si–O–Si angle is only 0.8°. However, structures near the boundary are deviating more from the “all-QM” data; see Supporting Information. This is expected due to the boundary effects in QM/MM calculations (as also seen in Sahoo and Nair, 2016).

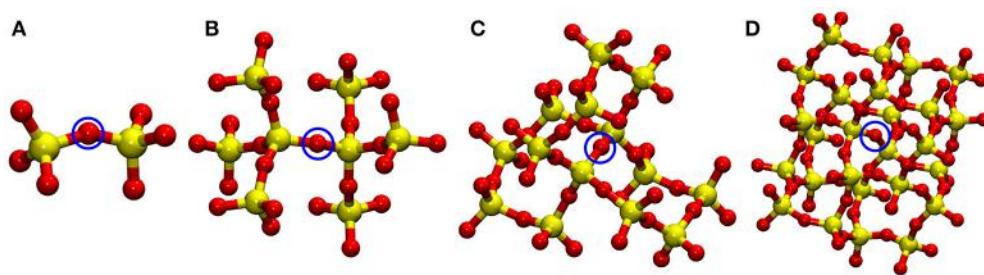
Next, the vacancy formation energy, ( $\Delta E_f$ ) in  $\alpha$ -cristobalite silica was then computed, as

$$\Delta E_f = \frac{x}{2} [E(O_2) + E_{\text{diss}}(O_2)] + E(SiO_{2-x}) - E(SiO_2) \quad (8)$$

where  $E(O_2)$ ,  $E_{\text{diss}}(O_2)$ ,  $E(SiO_{2-x})$ , and,  $E(SiO_2)$  are the energies of O<sub>2</sub> molecule (in the triplet electronic ground state), the dissociation energy of O<sub>2</sub> molecule, the energy of bulk silica with oxygen vacancy, and the energy of pure bulk silica, respectively.  $E(O_2)$  was computed from “all-QM” calculations whereas  $E(SiO_{2-x})$  and  $E(SiO_2)$  were computed either by “all-QM” or QM/p-MM calculations.  $E_{\text{diss}}(O_2) = 5.16$  eV was taken from the available experimental data (Lide, 2005).  $\Delta E_f$  values computed from “all-QM” calculations with varying supercell sizes are listed in **Table 3**. The converged value of  $\Delta E_f$  (w.r.t supercell size) is 8.73 eV.

$\Delta E_f$  values computed from QM/p-MM calculations listed in **Table 3**, show that  $\Delta E_f$  is nearly converged to 8.77 eV which is in excellent agreement with the “all-QM” data. However, it may





**FIGURE 4** | QM subsystems used in the four QM/MM calculations of oxygen vacancy in  $\alpha$ -Cristobalite: **(A)** 2T ( $\text{Si}_2\text{O}_7$ ), **(B)** 8T ( $\text{Si}_8\text{O}_{25}$ ), **(C)** 14T ( $\text{Si}_{14}\text{O}_{40}$ ), and **(D)** 26T ( $\text{Si}_{26}\text{O}_{67}$ ). In the vacancy calculations, the highlighted (by blue circle) O atom was removed from its lattice position. Note that MM atoms are not shown here. Atom colors: Si (yellow), O (red).

**TABLE 2** | The average value of Si–O bond length (Å), O–Si–O, and Si–O–Si angles ( $^{\circ}$ ) of  $\alpha$ -cristobalite computed from MD simulations in *NVT* ensemble at 300 K using p–MZHB, MZHB (Sahoo and Nair, 2015), QM/p–MM (14T), and QM potentials.

	“all-MM” <sup>a</sup>	“all-QM” <sup>b</sup>	QM/p-MM <sup>c</sup> (14T)	Expt.
Si–O bond length	1.61( $\pm 0.03$ )	1.63( $\pm 0.03$ )	1.62( $\pm 0.03$ )	1.60(3)
O–Si–O angle	109.4( $\pm 3.1$ )	109.2( $\pm 4.0$ )	109.4( $\pm 3.8$ )	108.2–111.4
Si–O–Si angle	147.7( $\pm 4.5$ )	142.2( $\pm 6.3$ )	143.0( $\pm 5.7$ )	146.4(9)

These results are also compared with experimental data (Downs and Palmer, 1994). For details see text.

<sup>a</sup>Using p–MZHB MM potentials.

<sup>b</sup>Using PBE density functional.

<sup>c</sup>Using PBE density functional and p–MZHB MM force–field.

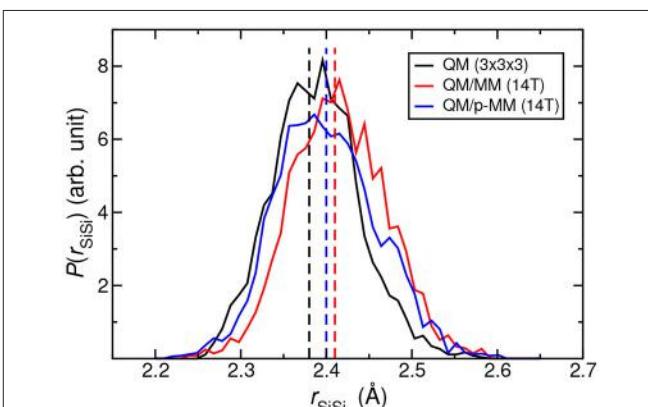
**TABLE 3** |  $\Delta E_f$  computed from the single point energy calculations using “all-QM” with PBE density functional.

Supercell	$\Delta E_f$ (eV)
$2 \times 2 \times 2$	8.68
$3 \times 3 \times 2$	8.73
$3 \times 3 \times 3$	8.73

**TABLE 4** |  $\Delta E_f$  computed from single point energy calculations using QM/p-MM potential.

Supercell	QM sites	$\Delta E_f$ (eV)
$8 \times 8 \times 8$	2T	9.43
	8T	8.76
	14T	8.78
	26T	8.77
$9 \times 9 \times 9$	8T	8.76
	14T	8.77
	26T	8.77

be noted that the values predicted by QM/p-MM calculations are sensitive to the size of the QM subsystem. The  $\Delta E_f$  value of 9.43 eV using 2T site is a very poor estimate compared with the



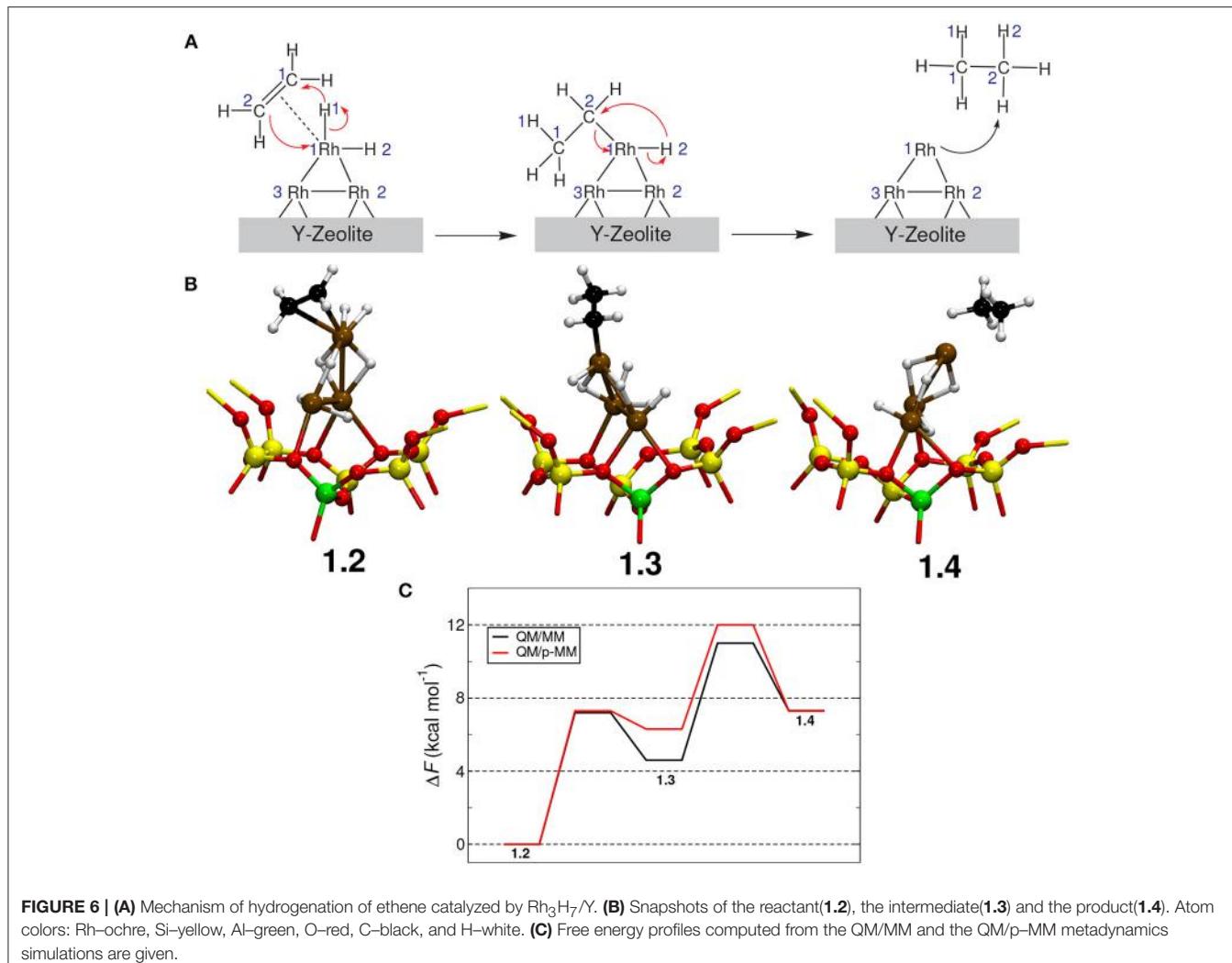
**FIGURE 5** | Probability distribution of  $r_{\text{SiSi}}$  distance obtained from the MD simulations at 300 K using “all-QM” ( $3 \times 3 \times 3$  supercell), QM/MM (14T) and QM/p-MM (14T) potentials. The vertical lines show the corresponding average values.

“all-QM” data. Clearly, the computed value converges close to “all-QM” data with increase in the size of the QM subsystem (see Table 4).

The bond length distribution of the Si–Si bond ( $r_{\text{SiSi}}$ ) at the defect site is then computed from the *NVT* trajectory using “all-QM” (using  $3 \times 3 \times 3$  supercell), QM/MM, and QM/p-MM potentials; see Figure 5. The average  $r_{\text{SiSi}}$  value using “all-QM” potential is 2.38 Å. Using the non-polarizable MZHB QM/MM simulations, we obtained a slightly increased value of 2.41 Å, while this was 2.40 Å in the polarized case. Overall, Figure 5 shows that the mean and the standard deviation of  $r_{\text{SiSi}}$  from QM/p-MM MD agree better to the “all-QM” MD results than that from the QM/MM MD using the non-polarizable MM force–field.

### 4.3. Application: Hydrogenation of Ethene Catalyzed by Rh Clusters Supported in Y-Zeolite

In order to investigate the effect of polarization of MM atoms on the free energy barriers, we revisit the study of hydrogenation of ethene catalyzed by Rh clusters supported in Y-zeolites, as



**FIGURE 6 | (A)** Mechanism of hydrogenation of ethene catalyzed by Rh<sub>3</sub>H<sub>7</sub>/Y. **(B)** Snapshots of the reactant(**1.2**), the intermediate(**1.3**) and the product(**1.4**). Atom colors: Rh—ochre, Si—yellow, Al—green, O—red, C—black, and H—white. **(C)** Free energy profiles computed from the QM/MM and the QM/p-MM metadynamics simulations are given.

in our previous study (Sahoo and Nair, 2016). Based on the experimental results from the Gates group (Liang and Gates, 2008) and the previous study (Sahoo and Nair, 2016), we choose the hydrogenated cluster Rh<sub>3</sub>H<sub>7</sub> in Y-zeolite as the catalyst model. In order to model Y-zeolite, a supercell of size 2 × 2 × 2 of pure siliceous Y-zeolite (Si<sub>1536</sub>O<sub>3072</sub>) was taken. The metal cluster, its ligands, and the metal-coordinated T25 site of the Y-zeolite were treated in the QM region (at the level of DFT/PBE+D2; Perdew et al., 1996; Grimme, 2006), while the rest of the zeolite was treated using MZHB or p-MZHB MM potentials.

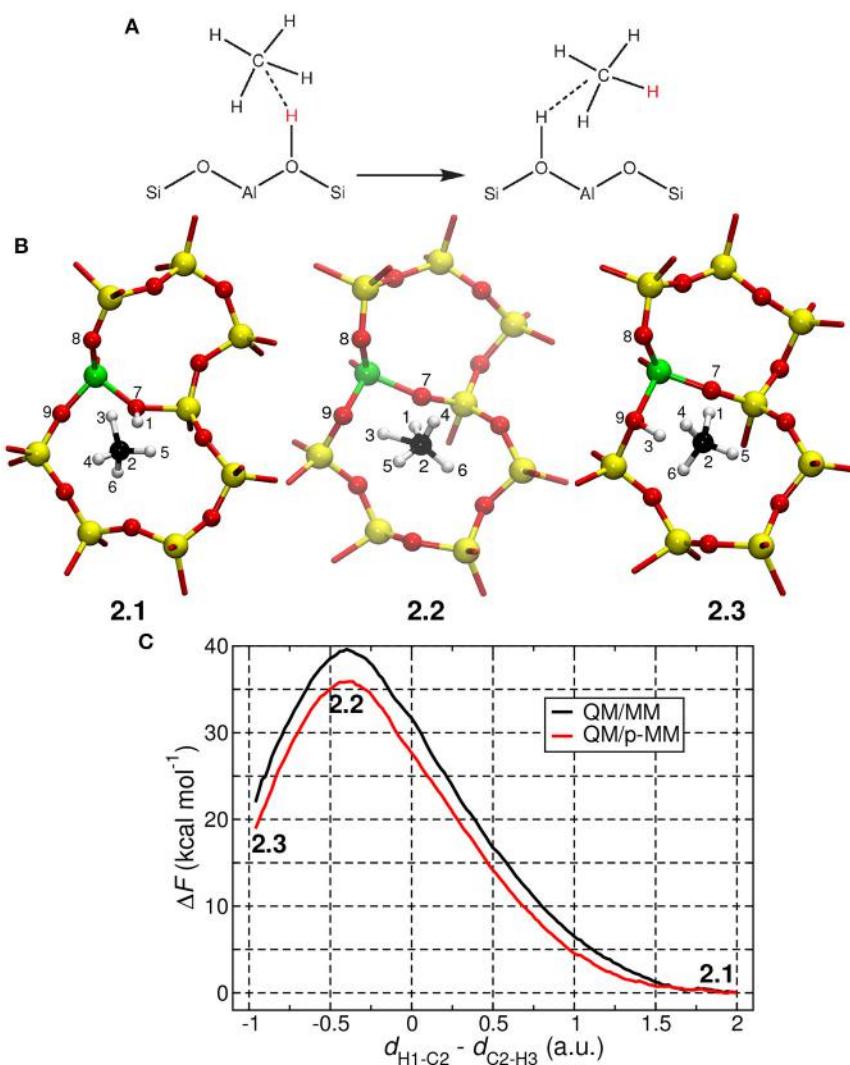
The free energy profile for ethene hydrogenation was computed employing metadynamics (Laiò and Parrinello, 2002) using the QM/MM and QM/p-MM implementations; see Figure 6 for the mechanism of the reaction studied here. For the details of the metadynamics simulation setup, see Supporting Information. In the metadynamics simulations using QM/MM and QM/p-MM methods, we observed the same reaction mechanism. Here, one of the hydrogen atoms from Rh<sub>3</sub>H<sub>7</sub> cluster moved to one of the C atoms

of ethene forming an intermediate **1.3**, which further reacted with a hydrogen atom on the Rh cluster to form ethane (**1.4**).

Interestingly, we observed differences of the order of 1 kcal mol<sup>-1</sup> only in the free energy barriers computed from QM/p-MM and QM/MM computations; see Figure 6C. The main difference is in the stability of the ethyl intermediate **1.3** and on the free energy barrier for **1.3**→**1.4**. However, the difference in free energy is only ~ 1 kcal mol<sup>-1</sup> which is close to the error associated with the metadynamics simulation. Thus, we conclude that for this specific reaction, the effect of MM polarization in the free energy estimates and on the reaction mechanism are negligible.

#### 4.4. Application: Proton Exchange Between Methane and H-ZSM-5 Zeolite

Methane activation is one of the important steps in various industrially relevant processes such as conversion of methane to higher hydrocarbons and methanol. In this process, the C–H bond, which is quite inert as evident from the high



**FIGURE 7 | (A)** Mechanism of the proton exchange reaction between a methane molecule and the H-ZSM-5 zeolite. **(B)** Snapshots of the reactant, transition state, and product states taken from QM/p-MM MD simulations. Here the atoms colors used are Si (yellow), Al (green), O (red), C (black), and H (white). **(C)** Free energy profiles constructed from TASS simulations are shown here. The free energy profiles were obtained by projecting the six-dimensional free energy surface on the CV,  $d_{\text{H}1-\text{C}2} - d_{\text{C}2-\text{H}3}$ , which is the difference in the distances H1–C2 and C2–H3.

bond formation energy ( $> 100$  kcal mol<sup>-1</sup>), is activated for further functionalization. Of great importance, protonated zeolites, particularly H-ZSM-5, is known to activate the methane C–H bonds (Arzumanov et al., 2014; Chu et al., 2016). Here we look at the following chemical reaction (Figure 7A):



One of the acidic protons in the H-ZSM-5 zeolite is exchanged with one of the protons in the methane molecule during this reaction. The mechanism for C–H activation in zeolites is previously studied in the literature (Vollmer and Truong, 2000; Zheng and Blowers, 2005; Truitt et al., 2006; Bučko et al., 2007; Gabrienko et al., 2011; Arzumanov et al., 2014; Tuma and Sauer,

2015; Chu et al., 2016). Two types of mechanisms are proposed for this reaction: (a) direct and (b) bimolecular. In the direct mechanism, hydrogen is exchanged through a carbonium ion intermediate, i.e., via the formation of the penta-coordinated carbocation. In the bimolecular mechanism, the alkane molecule dissociates one of H atoms to the zeolite framework, forming an alkoxy intermediate (Truitt et al., 2006), which subsequently undergoes reactions with other alkane molecules. Chu et al. (2016) have reported from both experimental and theoretical studies that higher alkanes react through bimolecular mechanism whereas the lower alkanes react via the direct mechanistic route. Our interest here is the direct mechanism as it involves the formation of a charged reactive intermediate and we anticipate some effect of MM polarization in the free energy estimates.

We could successfully simulate the proton exchange reaction using the Temperature Accelerated Sliced Sampling (TASS) method (Awasthi and Nair, 2017). This method allowed us to explore a high-dimensional free energy landscape composed of five collective variables (CVs). More details about the CVs used here and other technical details of the TASS simulation are given in the Supporting Information. The computed free energy profiles (projected along one of the crucial CVs) are given in **Figure 7C**. In TASS+QM/MM and TASS+QM/p-MM simulations, the hydrogen exchange reaction was found to proceed through the formation of a carbonium ion ( $\text{CH}_5^+$ ; see structure **2.2** in **Figure 7B**). The free energy barrier ( $\Delta F^\ddagger$ ) for the reaction computed from QM/MM and QM/p-MM simulations are 39.5 and 36.0 kcal mol<sup>-1</sup>, respectively. As expected, we are observing significant difference in the free energy barriers when polarized MM force-field is used. It is also noted in passing that the free energy barrier computed here are close to the potential energy barriers computed for similar reactions in zeolites in Vollmer and Truong (2000), Zheng and Blowers (2005), Bučko et al. (2007), and Tuma and Sauer (2015), which are in the range 29–38 kcal mol<sup>-1</sup>.

## 5. CONCLUSIONS

An extended Lagrangian based implementation of QM/p-MM method that allows to perform conventional Car-Parrinello MD for the QM subsystem is presented here. In particular, we have discussed a combined scheme where the extended Lagrangian dynamics of the shell (or the Drude) variables are performed together with the Car-Parrinello dynamics of the KS orbitals. Inclusion of polarization does not increase the computational cost of the QM/MM Car-Parrinello MD simulations within our approach, mainly because we use the same time step as that of the conventional Car-Parrinello MD.

We find that, invoking polarization of MM atoms in QM/MM calculations only marginally improve the predictions of equilibrium structure and dynamics of non-charged systems.

On the other hand, when the transition state is charged, free energy barriers are considerably affected on the inclusion of MM polarization. We believe that the methods and the strategies developed here for a QM/polarized-MM implementation will be useful to study more complex problems in catalysis, reactions in solid-liquid interfaces, crystallization etc.

Other than these, we also report here a new low point charge polarizable potential for silica, which is suitable for performing QM/MM calculations. The potential is shown to be performing well and is able to reproduce bulk structures of various silica polymorphs. The developed MM potential has simple and commonly used potential functions with fewer parameters, making it easy to use in various simulation packages such as GULP, DL\_POLY (Todorov et al., 2006) and LAMMPS (Plimpton, 1995).

## AUTHOR CONTRIBUTIONS

NN has planned the work and SS have executed the implementation and computations. Both authors have analyzed the results and have written the paper.

## ACKNOWLEDGMENTS

Authors acknowledge the HPC facility, IIT Kanpur. SS thanks CSIR New Delhi and IIT Kanpur for the Ph.D. fellowship. Authors are thankful to Dr. Shalini Awasthi (IIT Kanpur) for her help in setting up the metadynamics and the TASS simulations. Authors are also grateful for the valuable inputs from Prof. Alessandro Laio (SISSA, Italy).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2018.00275/full#supplementary-material>

## REFERENCES

- Arzumanov, S. S., Moroz, I. B., Freude, D., Haase, J., and Stepanov, A. G. (2014). Methane activation on in modified ZSM-5 zeolite. H/D hydrogen exchange of the alkane with Brønsted acid sites. *J. Phys. Chem. C* 118, 14427–14432. doi: 10.1021/jp5037316
- Awasthi, S., and Nair, N. N. (2017). Exploring high dimensional free energy landscapes: temperature accelerated sliced sampling. *J. Chem. Phys.* 146:094108. doi: 10.1063/1.4977704
- Bakowies, D., and Thiel, W. (1996). Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.* 100, 10580–10594.
- Boulanger, E., and Thiel, W. (2012). Solvent boundary potentials for hybrid QM/MM computations using classical Drude oscillators: a fully polarizable model. *J. Chem. Theory Comput.* 8, 4527–4538. doi: 10.1021/ct30722e
- Boulanger, E., and Thiel, W. (2014). Toward QM/MM simulation of enzymatic reactions with the Drude oscillator polarizable force field. *J. Chem. Theory Comput.* 10, 1795–1809. doi: 10.1021/ct401095k
- Bučko, T., Bencó, L., Hafner, J., and Ángyán, J. G. (2007). Proton exchange of small hydrocarbons over acidic chabazite: *Ab initio* study of entropic effects. *J. Catal.* 250, 171–183. doi: 10.1016/j.jcat.2007.05.025
- Car, R., and Parrinello, M. (1985). Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* 55, 2471–2474.
- Chu, Y., Xue, N., Xu, B., Ding, Q., Feng, Z., Zheng, A., et al. (2016). Mechanism of alkane H/D exchange over zeolite H-ZSM-5 at low temperature: a combined computational and experimental study. *Catal. Sci. Technol.* 6, 5350–5363.
- CPMD (v 13.2). CPMD Program Package. IBM Corp 1990–2018, MPI für Festkörperforschung Stuttgart 1997–2018. Available online at: <http://www.cpmd.org>
- de Leeuw, N. H., and Parker, S. C. (1998). Molecular-dynamics simulation of MgO surfaces in liquid water using a shell-model potential for water. *Phys. Rev. B* 58, 13901–13908.
- Dick, B. G., and Overhauser, A. W. (1958). Theory of the dielectric constants of alkali halide crystals. *Phys. Rev.* 112, 90–103.
- Downs, R. T., and Palmer, D. C. (1994). The pressure behavior of alpha-cristobalite. *Am. Min.* 79, 9–14.
- Gabrienko, A. A., Arzumanov, S. S., Toktarev, A. V., Freude, D., Haase, J., and Stepanov, A. G. (2011). Hydrogen H/D exchange and activation of C1n-C4

- alkanes on Ga-modified zeolite BEA studied with  $1\text{H}$  magic angle spinning nuclear magnetic resonance *in situ*. *J. Phys. Chem. C* 115, 13877–13886. doi: 10.1021/jp204398r
- Gale, J. D. (1997). GULP: a computer program for the symmetry-adapted simulation of solids. *J. Chem. Soc. Faraday Trans.* 93, 629–637.
- Geerke, D. P., Thiel, S., Thiel, W., and van Gunsteren, W. F. (2007). Combined QM/MM molecular dynamics study on a condensed-phase SN2 reaction at nitrogen: The effect of explicitly including solvent polarization. *J. Chem. Theory Comput.* 3, 1499–1509. doi: 10.1021/ct7000123
- Grimme, S. (2006). Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* 27, 1787–1799. doi: 10.1002/jcc.20495
- Illingworth, C. J., Gooding, S. R., Winn, P. J., Jones, G. A., Ferenczy, G. G., and Reynolds, C. A. (2006). Classical polarization in hybrid QM/MM methods. *J. Phys. Chem. A* 110, 6487–6497. doi: 10.1021/jp046944i
- Jorgensen, J. D. (1978). Compression mechanisms in  $\alpha$ -quartz structures-SiO2 and GeO2. *J. Appl. Phys.* 49, 5473–5478.
- Laio, A., and Parrinello, M. (2002). Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* 99, 12562–12566. doi: 10.1073/pnas.202427399
- Laio, A., VandeVondele, J., and Rothlisberger, U. (2002). A Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations. *J. Chem. Phys.* 116, 6941–6947. doi: 10.1063/1.1462041
- Lamoureux, G., and Roux, B. (2003). Modeling induced polarization with classical Drude oscillators: theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* 119, 3025–3039. doi: 10.1063/1.1589749
- Leach, A. R. (2010). *Molecular Modelling: Principle and Applications*, 2nd Edn. Harlow: Pearson.
- Lev, B., Zhang, R., de la Lande, A., Salahub, D., and Noskov, S. Y. (2010). The QM-MM interface for CHARMM-deMon. *J. Comput. Chem.* 31, 1015–1023. doi: 10.1002/jcc.21387
- Liang, A. J., and Gates, B. C. (2008). Time-resolved structural characterization of formation and break-up of rhodium clusters supported in highly dealuminated Y zeolite. *J. Phys. Chem. C* 112, 18039–18049. doi: 10.1021/jp805917g
- Lide, D. R. (2005). *Crc Handbook*. CRC Handbook of Chemistry and Physics, Internet Version 2005. Boca Raton, FL: CRC Press. Available online at: <http://www.hcpcnetbase.com>
- Lindan, P. J. D., and Gillan, M. J. (1993). Shell-model molecular dynamics simulation of superionic conduction in CaF2. *J. Phys. Condens. Matter* 5:1019.
- Loco, D., Lagardère, L., Caprasecca, S., Lipparrini, F., Mennucci, B., and Piquemal, J.-P. (2017). Hybrid QM/MM molecular dynamics with AMOEBA polarizable embedding. *J. Chem. Theory Comput.* 13, 4025–4033. doi: 10.1021/acs.jctc.7b00572
- Loco, D., Polack, É., Caprasecca, S., Lagardère, L., Lipparrini, F., Piquemal, J. P., et al. (2016). A QM/MM approach using the AMOEBA polarizable embedding: from ground state energies to electronic excitations. *J. Chem. Theory Comput.* 12, 3654–3661. doi: 10.1021/acs.jctc.6b00385
- Lu, Z., and Zhang, Y. (2008). Interfacing *ab initio* quantum mechanical method with classical Drude oscillator polarizable model for molecular dynamics simulation of chemical reactions. *J. Chem. Theory Comput.* 4, 1237–1248. doi: 10.1021/ct800116e
- Martyna, G. J., Klein, M. L., and Tuckerman, M. (1992). Nosé–hoover chains: the canonical ensemble via continuous dynamics. *J. Chem. Phys.* 97, 2635–2643.
- Marx, D., and Hutter, J. (2009). *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge: Cambridge University Press.
- Marx, D., Tuckerman, M. E., and Martyna, G. J. (1999). Quantum dynamics via adiabatic *ab initio* centroid molecular dynamics. *Comp. Phys. Commun.* 118, 166–184.
- Mitchell, P. J., and Fincham, D. (1993). Shell model simulations by adiabatic dynamics. *J. Phys.: Condens. Matter* 5:1031.
- Nasluzov, V. A., Ivanova, E. A., Shor, A. M., Vayssilov, G. N., Birkenheuer, U., and Rösch, N. (2003). Elastic polarizable environment cluster embedding approach for covalent oxides and zeolites based on a density functional method. *J. Phys. Chem. B* 107, 2228–2241. doi: 10.1021/jp026742r
- Niklasson, A. M., Steneteg, P., Odell, A., Bock, N., Challacombe, M., Tymczak, C. J., et al. (2009). Extended Lagrangian Born–Oppenheimer molecular dynamics with dissipation. *J. Chem. Phys.* 130:214109. doi: 10.1063/1.3148075
- Perdew, J. P., Burke, K., and Ernzerhof, M. (1996). Generalized gradient approximation made simple. *Phys. Rev. Lett.* 77, 3865–3868.
- Plimpton, S. (1995). Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* 117:1–19.
- Riahi, S., and Rowley, C. N. (2014). The CHARMM-Turbomole interface for efficient and accurate QM/MM molecular dynamics, free energies, and excited state properties. *J. Comput. Chem.* 35, 2076–2086. doi: 10.1002/jcc.23716
- Rowley, C. N., and Roux, B. (2012). The solvation structure of Na<sup>+</sup> and K<sup>+</sup> in liquid water determined from high level *ab initio* molecular dynamics simulations. *J. Chem. Theory Comput.* 8, 3526–3535.
- Sahoo, S. K., and Nair, N. N. (2015). A potential with low point charges for pure siliceous zeolites. *J. Comput. Chem.* 36, 1562–1567. doi: 10.1021/ct50091w
- Sahoo, S. K., and Nair, N. N. (2016). CPMD/GULP QM/MM interface for modeling periodic solids: implementation and its application in the study of Y-zeolite supported Rh<sub>n</sub> clusters. *J. Comput. Chem.* 37, 1657–1667. doi: 10.1002/jcc.23968
- Sangster, M., and Dixon, M. (1976). Interionic potentials in alkali-halides and their use in simulations of molten-salts. *Adv. Phys.* 25, 247–342.
- Schlück, T. (2010). *Molecular Modeling and Simulation: An Interdisciplinary Guide*, 2nd Edn. New York, NY: Springer.
- Sprik, M. (1991). Computer simulation of the dynamics of induced polarization fluctuations in water. *J. Phys. Chem.* 95, 2283–2291.
- Sulimov, V. B., Sushko, P. V., Edwards, A. H., Shluger, A. L., and Stoneham, A. M. (2002). Asymmetry and long-range character of lattice deformation by neutral oxygen vacancy in  $\alpha$ -quartz. *Phys. Rev. B* 66:024108. doi: 10.1103/PhysRevB.66.024108
- Todorov, I. T., Smith, W., Trachenko, K., and Dove, M. T. (2006). DL\_POLY\_3: new dimensions in molecular dynamics simulations via massive parallelism. *J. Mater. Chem.* 16, 1911–1918. doi: 10.1039/b517931a
- Truitt, M. J., Toporek, S. S., Rovira-Truitt, R., and White, J. L. (2006). Alkane C–H bond activation in zeolites: evidence for direct protium exchange. *J. Am. Chem. Soc.* 128, 1847–1852. doi: 10.1021/ja0558802
- Tuma, C., and Sauer, J. (2015). Quantum chemical *ab initio* prediction of proton exchange barriers between CH<sub>4</sub> and different H-zeolites. *J. Chem. Phys.* 143:102810. doi: 10.1063/1.4923086
- Vollmer, J. M., and Truong, T. N. (2000). Mechanisms of hydrogen exchange of methane with h-zeolite y: an *ab initio* embedded cluster study. *J. Phys. Chem. B* 104, 6308–6312. doi: 10.1021/jp0008445
- Wilson, M., and Madden, P. A. (1993). Polarization effects in ionic systems from first principles. *J. Phys. Condens. Matter* 5:2687.
- Woodcock, H. L., Hodošček, M., Gilbert, A. T. B., Gill, P. M. W., Schaefer, H. F., and Brooks, B. R. (2007). Interfacing Q-Chem and CHARMM to perform QM/MM reaction path calculations. *J. Comput. Chem.* 28, 1485–1502. doi: 10.1002/jcc.20587
- Zheng, X., and Blowers, P. (2005). A computational study of alkane hydrogen-exchange reactions on zeolites. *J. Mol. Catal. A Chem.* 242, 18–25. doi: 10.1016/j.molcata.2005.07.029

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling Editor and reviewer TH declared their involvement as co-editors in the Research Topic, and confirm the absence of any other collaboration.

Copyright © 2018 Sahoo and Nair. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Binding Mode of the Sonic Hedgehog Inhibitor Robotnikinin, a Combined Docking and QM/MM MD Study

Manuel Hitzenberger<sup>1,2</sup>, Daniela Schuster<sup>3</sup> and Thomas S. Hofer<sup>1\*</sup>

<sup>1</sup> Theoretical Chemistry Division, Institute of General, Inorganic and Theoretical Chemistry, University of Innsbruck, Innsbruck, Austria, <sup>2</sup> Department of Physics, Theoretical Biophysics (T38), Technical University of Munich, Munich, Germany,

<sup>3</sup> Pharmaceutical Chemistry, Institute of Pharmacy, University of Innsbruck, Innsbruck, Austria

## OPEN ACCESS

### Edited by:

Jean-Philip Piquemal,  
Sorbonne Universités, France

### Reviewed by:

Jitrayut Jitonnomp,  
University of Phayao, Thailand  
Albert Poater,  
University of Girona, Spain

### \*Correspondence:

Thomas S. Hofer  
t.hofer@uibk.ac.at

### Specialty section:

This article was submitted to  
Theoretical and Computational  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

Received: 25 July 2017

Accepted: 25 September 2017

Published: 23 October 2017

### Citation:

Hitzenberger M, Schuster D and  
Hofer TS (2017) The Binding Mode of  
the Sonic Hedgehog Inhibitor  
Robotnikinin, a Combined Docking  
and QM/MM MD Study.  
Front. Chem. 5:76.  
doi: 10.3389/fchem.2017.00076

Erroneous activation of the Hedgehog pathway has been linked to a great amount of cancerous diseases and therefore a large number of studies aiming at its inhibition have been carried out. One leverage point for novel therapeutic strategies targeting the proteins involved, is the prevention of complex formation between the extracellular signaling protein Sonic Hedgehog and the transmembrane protein Patched 1. In 2009 robotnikinin, a small molecule capable of binding to and inhibiting the activity of Sonic Hedgehog has been identified, however in the absence of X-ray structures of the Sonic Hedgehog-robotnikinin complex, the binding mode of this inhibitor remains unknown. In order to aid with the identification of novel Sonic Hedgehog inhibitors, the presented investigation elucidates the binding mode of robotnikinin by performing an extensive docking study, including subsequent molecular mechanical as well as quantum mechanical/molecular mechanical molecular dynamics simulations. The attained configurations enabled the identification of a number of key protein-ligand interactions, aiding complex formation and providing stabilizing contributions to the binding of the ligand. The predicted structure of the Sonic Hedgehog-robotnikinin complex is provided via a PDB file as Supplementary Material and can be used for further reference.

**Keywords:** sonic hedgehog (Shh), QM/MM, robotnikinin, sonic hedgehog inhibitor, metalloproteins, density functional theory, docking studies, molecular dynamics simulation

## 1. INTRODUCTION

The Hedgehog (Hh) family of proteins derives its name from the malformations that occur to larvae of drosophila flies upon altering of the respective gene (Varjosalo and Taipale, 2008). While for drosophila and other invertebrates only one variant of this protein is known, at least three different forms occur in vertebrates, namely Indian Hedgehog (Ihh), Desert Hedgehog (Dhh) and Sonic Hedgehog (Shh) (Ingham and McMahon, 2001; Varjosalo and Taipale, 2008). Incidentally, Dhh is more closely related to the drosophila variant of the protein, whereas Ihh and Shh share a lot of similarities, implying a more recent gene duplication event (Ingham and McMahon, 2001). The Hedgehog signaling pathway plays an important role in several crucial events during embryogenesis, including patterning of the neural tube, limb and lung development, it also steers

the segmentation of insect bodies (Ingham and McMahon, 2001; Jeong and McMahon, 2004; Varjosalo and Taipale, 2008). With the exception of adult stem cell differentiation (Palma et al., 2004) Hh signaling is mostly dormant in adults, however, aberrant activation of the pathway has been linked to a large number of cancerous diseases (Hahn et al., 1996; Goodrich et al., 1997; Berman et al., 2003; Hamed et al., 2004; Kubo et al., 2004) like for example bladder cancer, medulloblastoma, breast cancer, esophageal cancer or rhabdomyosarcoma and therefore has increasingly been targeted as a leverage point for novel anti-cancer therapies.

Shh is the most widely (Varjosalo and Taipale, 2008) expressed Hedgehog variant in vertebrates and thus most information concerning the biochemical pathways involving Hedgehog proteins has been gathered through investigations of Shh (Hwang et al., 2011). Sonic Hedgehog is synthesized inside the cell as a 45 kDa precursory protein, consisting of a 20 kDa N-terminal signaling and a C-terminal auto-catalytically active processing domain. Upon cleavage of the C-terminus, the remaining signaling peptide is modified with an N-terminal palmitic acid moiety and a C-terminal cholesterol molecule in order to mature into the morphologically active ShhN form (also referred to as Shh from this point on). The processed protein is then secreted into the extracellular matrix, where it acts as a ligand for the transmembrane protein Patched 1 (Ptc1). The binding of Shh to Ptc1 mediates the release of another transmembrane protein, Smoothened (Smo) which in turn migrates to the cell's primary cilium from where it activates the glioma-associated oncogene (Gli) transcription factors, thereby promoting the expression of Shh pathway-specific genes. An alternative binding partner on the cellular surface is the Hedgehog-interacting protein (Hhip), which is upregulated upon Shh binding and functioning as a decoy for Shh, hence acting as an antagonist for pathway activation (Ingham and McMahon, 2001; Varjosalo and Taipale, 2008; Bosanac et al., 2009). Most approaches, aiming to counteract the abnormal activation of the Hh pathway are targeting the deactivation of Smo or the transcription factors themselves (Varjosalo and Taipale, 2008; De Smaele et al., 2012). Another viable approach would be the inhibition of Shh binding to Ptc1, thereby increasing the therapeutic selectivity by minimizing the risk of unwanted deactivation of important biological pathways that are associated with Smo or Gli but independent of Shh (Rimkus et al., 2016). Known potential ligands include antibodies (Maun et al., 2010) as well as the small molecule robotnikinin (Mol. weight = 454.95g/mol; Stanton et al., 2009).

According to various investigations conducting X-ray crystallography, Shh possesses three divalent metal ions: Two Ca(II) ions, bound in loop regions by residues E90, E91, E127, D96, D130, and D132 (McLellan et al., 2008) and a Zn(II) ion, coordinated by two histidine residues (H141, H183), an aspartate (D148) and a water molecule, bridging the ion with glutamate E177 (Bishop et al., 1999, 2009; Bosanac et al., 2009). Structurally, the zinc site is analogous to those of zinc hydrolases such as thermolysin or bacterial carboxypeptidase A (Hall et al., 1995), however, extensive studies of Shh could not confirm any enzymatic activity (Fuse et al., 1999). Still, the existence of these

ions suggests that they are important for Shh to carry out its role in the pathway, hence several experimental and theoretical studies have been undertaken to uncover their influence (Bishop et al., 1999, 2009; McLellan et al., 2008; Bosanac et al., 2009; Maun et al., 2010; Hwang et al., 2011, 2013).

This work is a follow-up on a recent computational study investigating the role of the metal ions of Shh, utilizing classical molecular mechanics (MM), as well as hybrid quantum mechanical(QM)/molecular mechanical molecular dynamics (MD) simulations (Hitzenberger and Hofer, 2016). One of the findings of this study was that simple MM based approaches are not sufficient to provide an accurate model for the complex interactions present in the Zn(II) binding site. The utilized DFT BP86-D3 (Perdew, 1986; Becke, 1988; Grimme et al., 2010), triple zeta (TZ) QM/MM (Warshel and Levitt, 1976; Lyne et al., 1987, 1990; Åqvist and Warshel, 1993) link bond (Hitzenberger and Hofer, 2015) approach, however, has been shown to be able to very accurately reproduce the available experimental data.

Targeting extracellular proteins, serving as ligands of transmembrane proteins can be a very challenging task which is highlighted by the effort required to discover robotnikinin necessitating the screening of a set of 10,000 diverse compounds (Stanton et al., 2009). Rational drug design could aid in the development of novel compounds that are able to inhibit Shh. To do this, however, requires the knowledge of the Shh-robotnikinin binding mode which is still unknown since no experimental data on this complex has been published yet. For this reason a QM/MM MD-refined docking study, providing detailed and highly accurate information on the interactions of robotnikinin with Shh is presented in this work. The combination of docking, force field approaches, quantum mechanics and molecular dynamics enables an exhaustive investigation of the system. By explicitly considering the dynamical aspects of the complex at QM level, the chosen approach is able to account for small conformational adaptations concerning the binding geometry and interaction profile (De Vivo et al., 2016).

## 2. METHODOLOGY

### 2.1. Classical Simulation Setup

The starting point for the docking of robotnikinin to Shh was the equilibrated classical simulation box, used in the previous investigations (Hitzenberger and Hofer, 2016), which are themselves based on an X-ray structure (Bosanac et al., 2009) (PDB:3HO5). All histidine residues were protonated at the  $\varepsilon$ -nitrogen, with the only exception being Hid 183, which was protonated at the  $\delta$ -position to enable a binding geometry akin to the one predicted by X-ray investigations (Bosanac et al., 2009). The acidic and basic sidechains were all in the protonation state, predominately present at the physiological pH value. In order to generate a reasonable docking pose in which the ligand adequately occupies the binding groove, two iterations of docking with an intermittent MD simulation were necessary, since in the starting structures, stemming from simulations of empty Shh, the binding groove is not accessible to

the ligand in its entirety. The first cycle of docking was performed using the software package MOE (Chemical Computing Group Inc., 2016), employing the AMBER-12:EHT force field and induced fit docking. The triangle matcher method was used to place the conformers of robotnikinin in the pseudo active site, while for scoring the London dG function was utilized prior to the refinement of the pose via the force field. After that, the poses were re-scored via the GVBI/WSA function, also used for the final ranking of the docking poses. Since all known interactions of Shh with its binding partners are mediated by amino acids present in the binding groove of Sonic Hedgehog (McLellan et al., 2008; Bosanac et al., 2009; Maun et al., 2010; Hwang et al., 2011, 2013), all residues in the respective region were selected as potential receptors in the docking step of the study. A set of diverse but highly ranked structures were selected for classical MD simulations in order to generate structures for the second docking cycle. The simulations were carried out using the AMBER-12SB (Zgarbova et al., 2011) force field in order to remain consistent with the settings used for docking. All ligand interactions were described by a GAFF force field generated via Antechamber (Case et al., 2014), a program part of the AMBER14 suite. Merz-Kollmann partial charges (Singh and Kollman, 1984) were derived by performing Hartree Fock (HF) calculations with a 6-31G\* (Hariharan and Pople, 1973; Krishnam et al., 1980; Franci et al., 1982; Clark et al., 1983; Gill et al., 1992) basis set using GAUSSIAN 09 (Frisch et al., 2009), as required by the AMBER force field. The Ca(II), Zn(II) ions of Shh as well as the chloride counter ions were described by the parameters (Aqvist, 1990; Li and Merz, 2014) provided with the AMBER14 (Case et al., 2014) simulation package. The complexes were placed in periodic, cubic simulation boxes with a volume of  $\sim 540,000 \text{ \AA}^3$  and solvated in approximately 17,000 rigid TIP3P (Jorgensen et al., 1983) water molecules. The non-bonded cutoff was set to 10 Å and the long range interactions were treated by the particle mesh Ewald (PME) (Darden et al., 1993) method. In order to satisfy the requirements of the chosen NpT ensemble, temperature coupling was carried out via Langevin dynamics with a collision frequency of  $1.0 \text{ ps}^{-1}$ , the pressure was controlled by the Berendsen manostat (Berendsen et al., 1984) with a relaxation time of 2 ps. The SHAKE (Ryckaert et al., 1977) algorithm was applied to constrain all bonds involving hydrogen, enabling a time step of 2.0 fs.

After an initial energy minimization of 60,000 steps, utilizing the sander module of AMBER14 (Case et al., 2014), the systems were heated for 2 ns to the target temperature of 300 K using pmemd (Case et al., 2014) (MPI) and positional restraints to keep the protein-ligand complex fixed. Subsequently, the restraints were lifted and the systems equilibrated for 15 ns at a pressure of 1 atm, again using the MPI version of pmemd. The 100 ns production run was performed using the CUDA (Nickolls et al., 2008) implementation of pmemd, thereby considerably speeding up the process.

After the first MM MD run, the simulation in which the ligand displayed the lowest root mean square deviation was selected for the preparation of the actual simulation system by redocking the ligand using identical settings as

before. The highest scoring structure was then used for another MM MD simulation, following the same protocol as above.

## 2.2. QM/MM Setup

Choosing an appropriate QM method to describe the chemically most relevant part of the system is imperative to gain accurate and representative data from the simulation, therefore a reasonable compromise between accuracy and computational demand has to be made. DFT methods have been found to work fairly well when employed for systems containing metal ions (Kuta et al., 2006; Lepšík and Field, 2007; Hierao, 2011; Ryde and Grimme, 2011) and even though there are examples where DFT fails to deliver accurate results (Schwenk et al., 2004; Radoň and Pierloot, 2008; Yoo et al., 2009; Rowley and Roux, 2012; Gillan et al., 2016), at the moment it still represents the best tradeoff between computational cost and reliability (Senn and Thiel, 2009). Alternatives, like HF have been shown to be inadequate for such systems (Ryde and Grimme, 2011) and second-order Møller-Plesset perturbation theory (MP2), while computationally much more costly is known to occasionally perform worse than DFT (Ryde and Grimme, 2011). More sophisticated methods, such as Coupled Cluster (CC) or Configuration Interaction (CI) are too demanding to utilize them for the description of systems of the size studied in this work. Moreover, the previously published QM/MM study of this protein has shown that the used BP86 functional (Perdew, 1986; Becke, 1988), along with the cc-pVTZ (Dunning, 1989) (for C, H, N and O atoms) and def2-TZVP (Wiegend and Ahlrichs, 2005) (for Zn) basis sets and the D3 correction is able to adequately describe the system, while still being economical enough to enable acceptable trajectory lengths (Hitzenberger and Hofer, 2016). For this reason and in order to produce data that is directly comparable, the same setup has been chosen for this work. However, for one of the simulations, the double-zeta (DZ) versions of the mentioned basis sets have been used. The resolution of identity (RI) (Ren et al., 2012) approach has been employed alongside the D3 correction (Grimme et al., 2010) to speed up the calculation of the 4-center-2 electron integrals and to improve the description of dispersion effects, respectively.

The system has been partitioned into a QM and an MM zone with the focus of attention on the Zn(II) coordination site and the ligand, since the classical part of the investigation (docking and MM MD) strongly suggested that the Zn(II)-robotnikinin interaction is of great importance to the stability of the resulting complex. An energy and structure adjusted link atom approach was applied to describe all bonds penetrating the interface between the QM and the MM zone (Amara and Field, 2003; Lin and Truhlar, 2007; Hitzenberger and Hofer, 2015; Messner, 2015). In order to cleanly terminate the QM region without the introduction of artifacts stemming from the QM/MM coupling, a set of suitable link atom parameters  $\{\rho, r_0, k_L\}$  (Hitzenberger and Hofer, 2015) was derived and can be found in **Table 1**. The embedding of the QM into the MM zone was handled via the electrostatic embedding method, where the QM atoms interact with their MM counterparts via inclusion of MM partial charges into the QM Hamiltonian. While the charges of the QM atoms

**TABLE 1** | Ideal link atom parameters for RI-BP86-D3, cc-pVTZ, or cc-pVDZ with embedding charges scaled by a factor of 0.666;  $\rho$  refers to the distance ratio between  $C_{\alpha}$  and  $C_{\beta}$  on which the link atom is placed,  $r_0$  (in Å) and  $k_L$  (in kcal/mol/Å<sup>2</sup>) represent the minimum and the force constant of the harmonic energy correction potential of the link bond (Hitzenberger and Hofer, 2015), respectively.

Link Bond	Triple Zeta Basis			Double Zeta Basis		
	$\rho$	$r_0$	$k_L$	$\rho$	$r_0$	$k_L$
Hid <sup>a</sup>	0.7140	1.560	139.056	0.7226	1.564	143.739
Hie <sup>b</sup>	0.7155	1.561	145.779	0.7257	1.561	176.663
Glu	0.7241	1.547	170.580	0.7289	1.554	187.240
Asp	0.7251	1.543	145.051	0.7223	1.548	127.987
ROB 1 <sup>c</sup>	0.7363	1.507	235.182	0.7435	1.514	242.270
ROB 2 <sup>d</sup>	0.7332	1.515	240.882	0.7395	1.519	245.840

<sup>a</sup>Histidine protonated at the  $\delta$  position.

<sup>b</sup>Histidine protonated at the  $\epsilon$  position.

<sup>c</sup>First robotnikinin link bond.

<sup>d</sup>Second robotnikinin link bond.

were updated at every step and calculated via the Mulliken method (Mulliken, 1962), the MM charges used were provided by the force field. The problem with such a setup is that MM point charges used in most popular force fields and water models are usually not tailored to accurately represent the electron density of a molecule but to reproduce certain observables (such as the permittivity of water) when applied together with the other force field parameters (Senn and Thiel, 2009). In general, it can be assumed that the point charges of the force field are not compatible with the DFT-derived Mulliken charges. The oxygen atom of a TIP3P water molecule (Jorgensen et al., 1983), for example, possesses a charge of  $-0.83e$ , while for a water molecule in bulk conditions where the partial charge is calculated by the used QM setup it is only  $-0.55e$ . Similar over-polarization can be witnessed with the amino acids of the protein. In order to minimize the possibility to sample artificially strong interactions between QM and MM species at the QM/MM interface, a (QM method dependent) scaling factor for all MM charges, used for embedding, has been applied. At the same time the charges used to calculate MM/MM interactions were not altered. The applied scaling factor is in principle a simplification of an embedding scheme where all charges next to a QM atom are described by a Gaussian distribution (Amara and Field, 2003), thereby providing a distance dependent charge scaling. This Gaussian scheme was confirmed to be a very accurate embedding method when compared to more traditional techniques (Amara and Field, 2003; König et al., 2005), however, it is not compatible with many popular QM programs and strongly dependent on the proper parametrization of the blurring width. Therefore, for this study, a fixed scaling factor of 0.666 was used for all QM/MM MD simulations as this ensured that the charges of water molecules in close proximity to the QM region are scaled down to  $-0.55e$ . Additionally, test calculations showed that the mean QM/MM charge deviation of the amino acids His, Asp, and Glu (all residues in very close proximity to the QM zone are of these types) is reduced from 0.17 to 0.08 as well when the scaling factor is applied. The use of this scaling factor already

proved successful in a previous study of Shh (Hitzenberger and Hofer, 2016). All remaining interactions between the QM and MM species, like bending and torsional terms are handled via the force field. All terms, however, where the central atoms are exclusively QM species have been excluded in order to prevent extensive double counting of forces (Eurenus et al., 1996).

Altogether, four different QM/MM MD simulations have been conducted in the course of this study. **Table 2** presents an overview of the differences between the simulations. Since the MM derived data suggests that robotnikinin directly coordinates to the Zn(II) site, the QM zone for the first simulation (henceforth called “core simulation”) was chosen so that all important interactions around the ion were described by quantum mechanics. Therefore, the Zn(II) ion, the sidechains of H141, H183, D148, E177, the macrocycle, plus the chain containing the second amide function of robotnikinin (see **Figures 1A,B** for detailed information regarding QM/MM partitioning), as well as all water molecules within a radius of 5.5 Å were considered to be QM species. If robotnikinin as a whole would be included into the QM calculation then all residues in the vicinity of its phenyl and chlorophenyl rings would have to be described by QM as well because interactions between aromatics in the QM and the MM zone are very sensitive to the QM/MM potential and thus very difficult to describe correctly. Consequently, the addition of only the aromatic rings would have very likely lowered the predictive power of the simulation. The inclusion of all potential interaction partners of the rings in question would have lead to a very large QM system and thus making it impossible to sample a reasonable amount of configurations. In consequence, robotnikinin was truncated in the QM system and the substituents were described by the force field. Since the parametrization of the resulting link bonds was very thorough and the ionic site distant enough to justify an MM description of the aromatic moieties this provides an adequate compromise between effort and accuracy.

The starting point of the core QM/MM simulation was the structure resulting from the final MM MD simulation. During the first 5 picoseconds of the equilibration phase the atoms coordinating Zn(II), according to the MM simulation were restraint to the ion via harmonic bonds. In the course of this pre-equilibration process that was conducted at the target temperature of 300K, the force constants of the bonds were lowered from 500 kcal/mol/Å<sup>2</sup> to zero in 3 steps (250 kcal/mol/Å<sup>2</sup>, 100 kcal/mol/Å<sup>2</sup>, 0 kcal/mol/Å<sup>2</sup>). This was followed by 10 ps of equilibration and a 80 ps sampling phase. Since the core simulation, in contrast to the MM simulation, suggested robotnikinin forming additional hydrogen bonds with two histidine sidechains (H134 and H135) not described by QM, an additional simulation, including those residues into the QM zone has been set up in order to confirm the existence of these interactions. The starting structure for this second simulation was taken from the core simulation.

However, this enlargement of the system results in a significant increase of the computing time by a factor of approximately 2, thus a third simulation utilizing only a double zeta (DZ) basis has been conducted, reducing the time needed per simulation step by a factor of four. This simulation was

**TABLE 2 |** Overview of the conducted QM/MM simulations.

Simulation	System	QM atoms	Basis set	MD Steps/ Simulation time
#1 Core	Core	81	TZ	4,000/80 ps
#2 Extended (TZ)	Core+H134, H135	105	TZ	1,000/20 ps
#3 Extended (DZ)	Core+H134, H135	105	DZ	4,500/90 ps
#4 Empty	Empty Shh	42	DZ	6,000/120 ps

The number of QM atoms includes hydrogen atoms but does not account for water molecules. The core system contains Zn(II), robotnikinin (excluding the aromatic rings), H141, D148, E177, and H183.

used to gather additional configurations in order to improve the statistics on which the evaluated importance of the new found interactions are based. After a 10 ps equilibration phase, a 90 ps evaluation trajectory was sampled for this simulation. In order to assess the accuracy of the results obtained at the DZ level, a further simulation, this time of empty Shh was conducted. The starting point of this simulation was the equilibrated TZ simulation of empty Shh taken from the previously published study (Hitzenberger and Hofer, 2016). All settings remained the same, however the basis set was switched to the double zeta variant and the link bond parameters were adjusted accordingly. After 20 ps of equilibration, a 120 ps long sampling trajectory was produced. The process flow used in this investigation is visualized in **Figure 2**.

Calculation of MM forces, the QM/MM coupling and the MD simulation itself were all handled by the in-house-developed QMCF (Rode et al., 2006; Hofer et al., 2010) simulation package. All quantum mechanical calculations were carried out using TURBOMOLE (Turbomole, 2007) and the temperature was controlled by the Berendsen thermostat (Berendsen et al., 1984) with a relaxation time of 1.0 ps. The nonbonded interactions were calculated explicitly up to a distance of 10 Å, while long range interactions were dealt with by the reaction field method (Barker and Watts, 1973) assuming a permittivity of  $\epsilon = 78.355$ . In order to allow for a time step of 2.0 fs, the SHAKE algorithm (Ryckaert et al., 1977) was applied and the equations of motion were solved using the velocity-Verlet integrator (Swope et al., 1982).

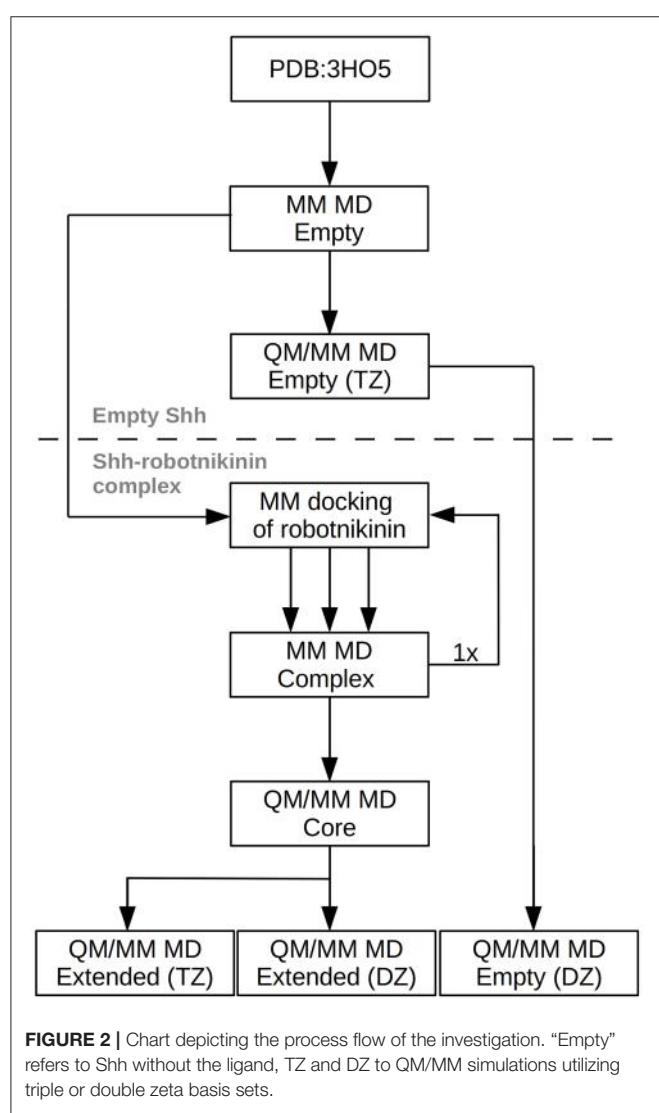
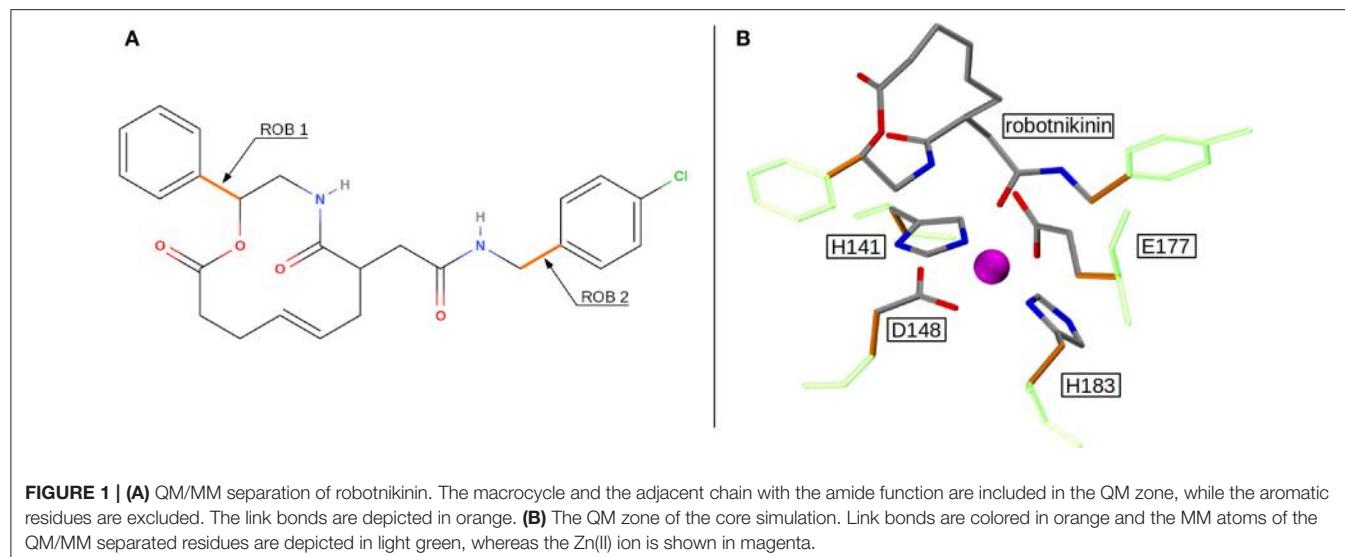
## 3. RESULTS AND DISCUSSION

### 3.1. MM MD Simulations

The binding site of Shh is located at the surface and shaped like a groove, therefore for a molecule like robotnikinin two general categories of binding poses are conceivable: one, where the chlorine atom points toward the region binding the Ca(II) ions and a second one, rotated by roughly 180° with the chlorine atom oriented in the opposite direction (see **Figures 3B,C**). During the docking phase two crucial properties were highlighted: Firstly, poses where an oxygen atom of robotnikinin coordinates to the Zn(II) ion were scored much higher than poses without direct robotnikinin-Zn(II) interactions. Nevertheless, MM MD simulations of complexes without direct Zn(II)-robotnikinin interactions have been conducted—all resulting in the disassociation of the complex. This indicates that the

interaction with the Zn(II) site is very important for the stability of the complex which is not surprising, since interactions between metal ions and polar sites are much stronger than simple hydrogen bonds. Furthermore, from experimental studies of the Shh-Hhip complex, it is well established that the Zn(II) ion is indeed accessible to ligands (Bosanac et al., 2009). The second finding was that the category containing the poses with the chlorine pointing away from the Ca(II) site is by far the prevalent one since nearly every high ranked pose that could be taken into consideration for a simulation was of that variant. Very likely, sterical effects preventing the formation of stable hydrophobic interactions of the chlorophenyl ring with its environment play an important role. These sterical clashes occur because the part of the binding groove connecting the Zn(II) and the Ca(II) sites is slightly too short to cleanly incorporate the chlorophenyl ring, which is separated from the macrocycle by 4 covalent bonds. In contrast, the phenyl ring is directly bonded to the macrocycle and thus can be easily positioned in this part of the binding site. Nevertheless, in order to make sure that the most likely binding pose is selected for the subsequent QM/MM MD simulations, the highest ranked docking structures of both robotnikinin binding pose families have been selected and MM MD simulations have been performed following the protocol from the previous section. The chlorophenyl moiety did not interact very strongly with any part of the protein when oriented in the direction of the Ca(II) ions. Instead, it was switching positions very frequently, thus leading to many configurations in which the aromatic residues of the ligand interacted with each other occasionally leading to robotnikinin dissociating from the protein. The only stable simulation of this binding mode category resulted in a robotnikinin heavy atom root mean square deviation (RMSD) shown in **Figure 3C**, which is very high compared to the RMSD derived from the simulation in which the ligand is rotated, as shown in **Figure 3D**. From this and a visual inspection of the trajectories it can be gathered that the poses from the type depicted in **Figure 3B** seem to be the prevalent ones and therefore all following simulations were based on them.

In order to elucidate the structural differences between loaded and empty Shh, heavy-atom RMSDs for all individual residues comparing each evaluation frame of loaded Shh (MM MD) to the first frame of a simulation of empty Shh (Hitzenberger and Hofer, 2016) (also MM MD) have been calculated. To eliminate high RMSDs stemming from natural residue fluctuations, per-residue RMSDs of empty Shh versus itself have been calculated as well. These “natural fluctuations” have then been subtracted from the raw data. The resulting “corrected,” color coded per-residue RMSDs are shown in a heatmap plot in **Figure 4A**, which was generated using the “Heat Mapper” tool provided with the VMD (Humphrey et al., 1996) package. The regions showing a high deviation correspond to the coil region depicted in **Figure 4B**, functioning as a lid for the binding groove. The radial distribution function (RDF) calculated from the 100 ns MM MD simulation trajectory shows a mean coordination number of 7.5 oxygen or nitrogen atoms around the Zn(II) ion (counting up to a Zn(II)-ligand distance of 3.18 Å where the RDF shows a minimum). The ion is coordinated by H141, D148 (bidentate), E177 (mono- or bidentate), H183 and the amide-oxygen atoms

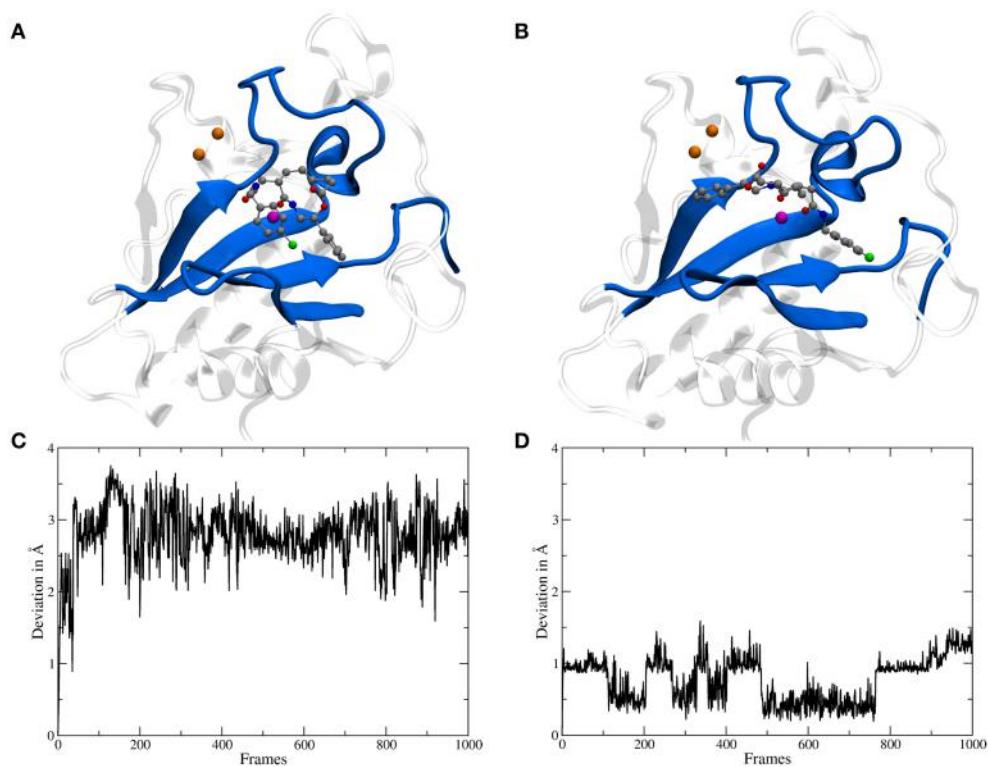


**FIGURE 2 |** Chart depicting the process flow of the investigation. “Empty” refers to Shh without the ligand, TZ and DZ to QM/MM simulations utilizing triple or double zeta basis sets.

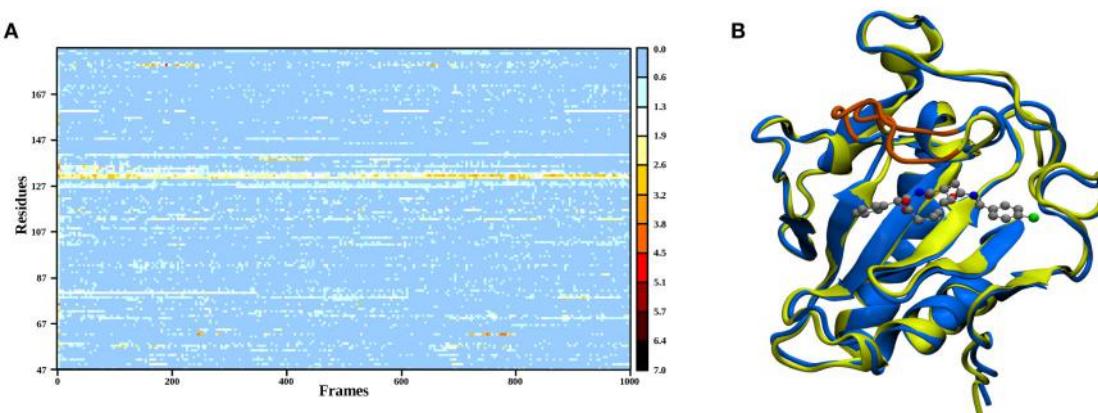
of robotnikinin. The amide function in the macrocycle shows the higher affinity and is coordinated throughout the whole simulation (see **Figure 5**). The aromatic phenyl and chlorophenyl rings interact with the nearby sidechains of T126, H181, and Y175 via hydrophobic interactions. There seem to be no interactions between the ester or amide groups of robotnikinin with any of the surrounding sidechains. Instead, the used classical model predicts the formation of a hydrogen bond between the ester function and the macrocyclic amide. If a hydrogen-donor distance of under 3.0 Å and a donor-acceptor-hydrogen angle lower than 35° are used as geometrical criteria for a hydrogen bond, then 66.1% of the 1,000 configurations chosen for evaluation feature an intramolecular hydrogen bond.

### 3.2. QM/MM MD Simulations

The core QM/MM MD simulation for which an 80 ps long evaluation trajectory has been sampled paints a different picture than the purely classical approach: Here, the Zn(II) ion is predominantly coordinated by D148 (bidentate), E177 (monodentate), H183 and the amide oxygen, part of the chain connecting the chlorophenyl ring of robotnikinin to its macrocycle. The coordination polyhedron is of quadratic pyramidal shape as can be seen in **Figure 5B** and according to an RDF (**Figure 5D**) that was calculated for the 4,000 frames of the evaluation trajectory, the mean coordination number is 5.1, when calculated up to a Zn(II) ligand distance of 2.83 Å (where the RDF reaches its first minimum), with the most likely ion-ligand distance being 2.03 Å. Incidentally, H141, part of the coordination sphere in empty Shh (Hall et al., 1995; McLellan et al., 2008; Bishop et al., 2009; Bosanac et al., 2009; Hitzenberger and Hofer, 2016) is mostly situated at distances between 3 and 4 Å of the ion, making way for robotnikinin, E177 and enabling a bidentate binding mode of D148. However, there are also configurations where it directly binds the ion thereby creating a quadratic bipyramidal coordination polyhedron. Another interesting aspect of the coordination site around the ion is



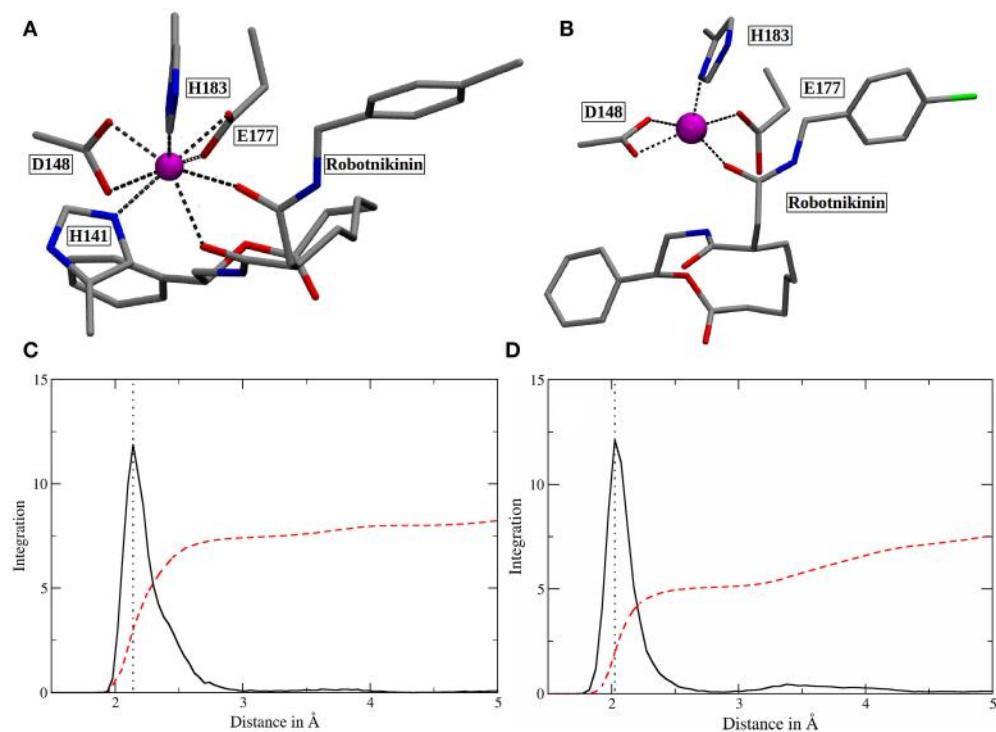
**FIGURE 3 | (A)** A snapshot from an MM MD simulation started from a docking pose with the chlorine atom of the ligand pointing toward the Ca(II) ions (depicted in orange). The snapshot shows a configuration in which the chlorophenyl ring left the binding groove and is oriented toward the ligand's second aromatic ring instead of the Ca(II) ions. This is caused by a lack of stabilizing hydrophobic interactions with the protein as well as steric clashes and indicates that the chosen starting structure is not a stable binding mode. **(B)** A snapshot from an MM simulation started from the preferred docking pose. **(C)** Heavy atom RMSD of robotnikinin, calculated from the trajectory of the simulation shown in **(A)**. **(D)** Heavy atom RMSD of robotnikinin, derived from the trajectory of the simulation shown in **(B)**. Each frame in the RMSD plots represents 100 ps.



**FIGURE 4 | (A)** Heatmap depicting the per-residue RMSDs of the classical simulations of loaded compared to empty Shh. The yellow and red regions indicate a high deviation. Note that the plot actually depicts the difference between the per-residue RMSDs of loaded versus empty Shh and the per-residue RMSDs derived from empty Shh versus itself, in order to emphasize only on the differences between the two systems. **(B)** Superposition of snapshots taken from loaded (blue) and empty (yellow) Shh simulations. The coil region spanning residues 131–136 showing the largest deviations is colored in orange.

E177, which has been shown to be mostly bridged by a water molecule in studies of empty Shh (Hitzenberger and Hofer, 2016) or X-ray derived crystal structures of Shh bound to

Hhip (McLellan et al., 2008; Bishop et al., 2009; Bosanac et al., 2009). However, the conducted QM/MM MD simulation predicts direct E177-Zn(II) coordination surely aided by the MM derived



**FIGURE 5 | (A)** Snapshot of the classical simulation, depicting 8-fold coordinated Zn(II) (magenta). **(B)** A representative snapshot from the core QM/MM simulation, showing the predominant coordination polyhedron around Zn(II). **(C,D)** Zn(II)-ligand RDFs of the classical **(C)** and the QM/MM simulation **(D)**. The RDFs are depicted in solid, black lines and their respective integrals in red. All oxygen and nitrogen atoms were considered for this plot. The dotted lines denote the peaks of the RDFs at 2.14 (MM) and 2.03 Å (QM/MM). Note that the y-axis labels refer to the integration only as the ligand atom densities are in arbitrary units.

starting structure also predicting such a coordination. In the previous study of empty Shh (Hitzenberger and Hofer, 2016), the MM MD trajectory used as starting point for the QM/MM simulation erroneously predicted E177 to directly interact with the ion during the entire simulation. However, the QM/MM model switched to the experimentally predicted coordination sphere almost instantly after heating the system. This was not the case in this QM/MM simulation, however, in the present case E177-Zn(II) binding is not implausible because the second oxygen atom of E177 forms a very strong hydrogen bond with the amide function in the macrocycle of Shh, thereby also directing the amide oxygen atom toward histidines H134 and H135 enabling potential hydrogen bonding. Therefore, the presence of E177 increases the number of possible protein-ligand interactions thus stabilizing the Sonic Hedgehog-robotnikinin complex. Although the continued presence of a water molecule in the binding site would be possible, it is however unlikely that it would aid the binding of robotnikinin as strongly as E177 does because a water molecule is neutral and possesses just one atom that can function as an electron donor, whereas E177 bridges two positively charged sites via its two spatially separated oxygen atoms. Furthermore, the elimination of strongly bound water molecules aids to ligand binding via a favorable entropy contribution (Ladbury, 1996; Li and Lazaridis, 2005; Huggins, 2015). In the previously conducted QM/MM simulation it has also been shown that while the bridged binding of E177 to Zn(II)

is absolutely predominant, there are also configurations where it binds directly (Hitzenberger and Hofer, 2016), suggesting that this form of coordination is not entirely unfeasible also without the presence of a ligand. Upon closer investigation of the E177-robotnikinin interaction it becomes clear that it is a very stable one, as the mean distance between the E177 oxygen and the hydrogen atom of the amide is 2.29 Å and applying the same H-bond criteria as before, 85.7% of all sampled configurations exhibit this particular hydrogen bond. Another argument for the exclusion of water molecules around the metal ion center is the fact that binding of robotnikinin to Shh renders Zn(II) practically inaccessible to the solvent. This is illustrated by the Zn(II)-water oxygen RDF of the core simulation: The first peak can be found at approx. 6.6 Å where integration up to this point yields 1.5 water molecules. At a separation of 5 Å integration of the RDF indicates a mean number of just 0.006 water molecules. This suggests that it is very unlikely that water re-enters the active site once it has left. If all these features are taken into consideration, it seems very plausible that robotnikinin binds to Zn(II) via a direct ionic bond, as well as indirectly via the residue E177.

The mean distances of residues H134 and H135 to the macrocyclic amide group of robotnikinin in the core QM/MM investigation are 4.41 and 2.53 Å, respectively and additionally there are some close contacts between H134 and the macrocyclic ester group (with an average distance of 4.09 Å), hinting at

the existence of favorable interactions between these histidines and the ligand. Calculating the RMSDs for just the sidechains (considering only heavy atoms) of these two residues results in a mean deviation of 0.481 Å, which compared to the RMSDs derived from the MM simulation (1.187 Å) or all sidechains in the QM/MM described protein (0.874 Å) is a very low value, further suggesting the presence of stabilizing interactions. However, if the same hydrogen bond criteria used for the MM simulations are taken into account, then in only 19.1% of all the frames H135-amide hydrogen bonds are present, while 4.8% of all sampled configurations show H134-amide H-bonding and only 43 out of 4,000 configurations fulfill the criteria for accepting the existence of a hydrogen bond between H134 and the ester functionality. Apparently, while the electrostatic interactions between the histidines and robotnikinin seem to be rather strong, the angle between donor, hydrogen atom and acceptor deviates quite far from 180° most of the time. One of the likely reasons for this behavior is the fact that the histidines in question are not part of the QM zone of this simulation, therefore all observed interactions between them and robotnikinin are somewhat error prone and thus should be closer observed by including H134 and H135 in the QM region. However, as QM calculations scale very unfavorably with system size, beside a larger triple zeta (TZ) simulation, the system was also simulated utilizing only a double zeta basis set in order to attain a larger number of configurations and thus gain data that is statistically more conclusive.

The more sophisticated TZ simulation suggests that the interaction between the histidines and the macrocycle of robotnikinin is more distinct than predicted by the core QM/MM simulation as the mean amid histidine distances are reduced to 2.77 Å (H134) and 2.08 Å (H135) in addition, also the H-bond acceptor oxygen of the ester is on average separated from the donor hydrogen of H134 by 3.94 Å and occasionally close enough for hydrogen bonding. The simulation predicts a hydrogen bond population of 80.7% for H135-amide, 40.6% for H134-amide and 1.4% for H134-ester, respectively. The mean distance between the closest E177 oxygen atom and the hydrogen of the macrocyclic amide is 2.14 Å, with a hydrogen bond abundance of 95.3%, implying the E177-robotnikinin and the H134/135-robotnikinin interactions promote each other by orienting the amide group in a favorable position. The statistically more robust (but less accurate) 90 ps DZ simulation predicts interactions that are even stronger than in the TZ case yet only by a small margin. The mean distances between H134,

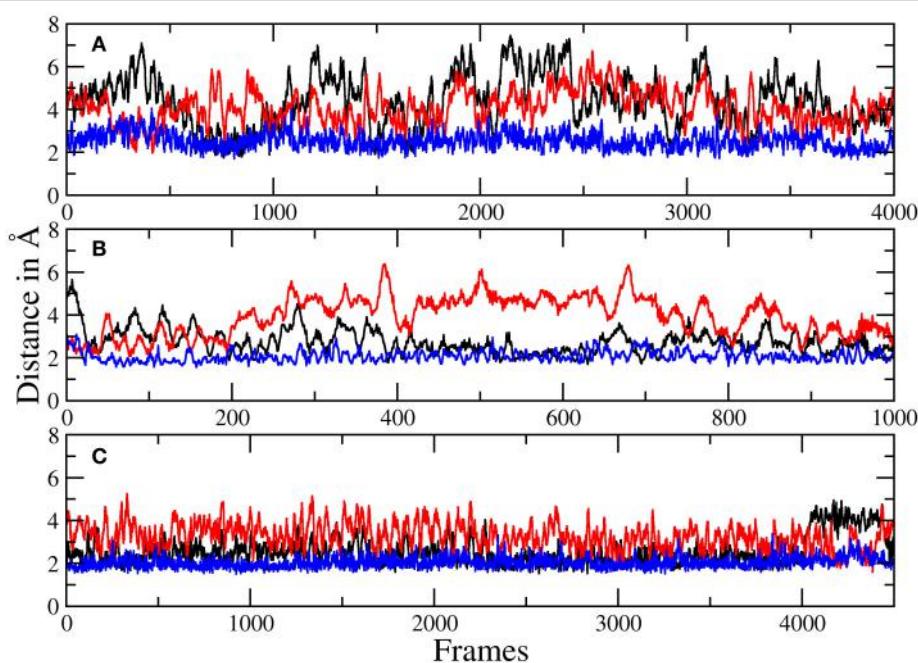
H135 and the amide are predicted to be 2.56 and 2.05 Å, respectively, with the average separation between H134 and the ester being 3.27 Å. Of the 4,500 sampled configurations, 87.5% show a hydrogen bond between H135 and the amide, 50.7% between H134 and the amide and 5.4% between H134 and the ester. The mean E177-amide distance is predicted to be 2.13 Å, with a hydrogen bond abundance of 92.9%. A summary of the discussed H-bond results regarding the interactions of H134, H135, and E177 with robotnikinin is provided in **Table 3**. To further highlight the binding motif, robotnikinin-sidechain distance plots are depicted in **Figures 6A–C**, taking only into consideration the atoms closest to each other, as these are regarded to be the mediators of the respective H-bonds. The interactions found for the aromatic rings are the same as in the MM MD simulation, which is unsurprising since in both cases they were described solely by the force field and the predominant stacking geometries are depicted in **Figure 7**.

If all these findings are taken into consideration, one can conclude that, besides the very strong Zn(II)-robotnikinin ionic bond, the most stable interaction found by the QM/MM MD simulations is the hydrogen bond between E177 and the macrocyclic amide hydrogen as this interaction has an occurrence ranging from 85.7% (core QM/MM simulation) to 92.9 and 95.3% in case of the DZ and TZ extended QM/MM simulations. The second most important hydrogen bond exists between H135 and the macrocyclic amide of robotnikinin with an abundance close to 90% in both extended simulations including H135 and H134 in the QM zone. Another frequently occurring interaction was identified between H134 and the same amide, which is present in roughly half of all sampled configurations of the extended QM/MM MD simulations. Also predicted to exist but not nearly as important as the other interactions since only witnessed in 1 to 5% of all frames is an H-bond between the ester and H134. The binding pose of robotnikinin in Shh as predicted by the extended TZ simulation is shown in **Figure 8A** and respective interactions mediating this orientation are depicted in **Figure 8B**. In order to further confirm these findings and to construct a three-dimensional map of all important interactions between robotnikinin and Sonic Hedgehog every fiftieth configuration of the extended TZ sampling trajectory was employed to create twenty different interactions models, utilizing the software tool LigandScout (version 4.09) (Wober and Langer, 2005; Wieder et al., 2017). The resulting three- and

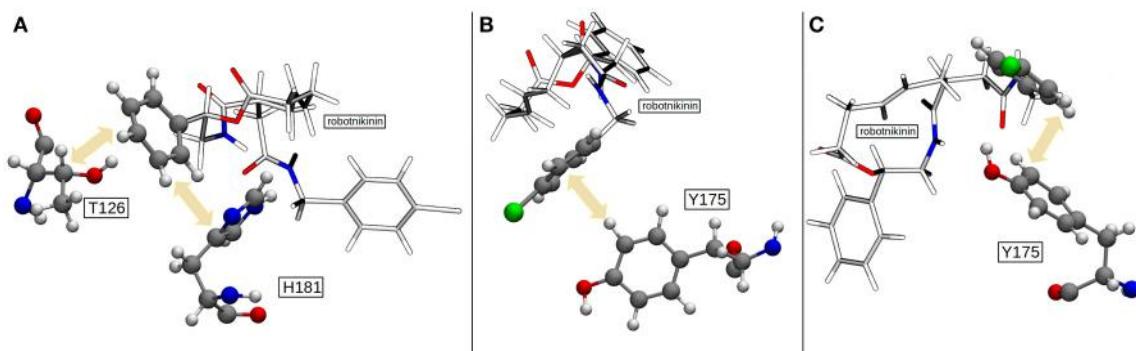
**TABLE 3 |** Distances (in Å) between residues H134, H135, E177 and the respective functional group of robotnikinin.

	Simulation #1 Core				Simulation #2 Extended (TZ)				Simulation #3 Extended (DZ)			
	Amide		Ester		Amide		Ester		Amide		Ester	
	Dist.	Occ.	Dist.	Occ.	Dist.	Occ.	Dist.	Occ.	Dist.	Occ.	Dist.	Occ.
H134	4.41	4.8	4.09	<0.1	<b>2.77</b>	<b>40.6</b>	<b>3.94</b>	<b>1.4</b>	<b>2.56</b>	<b>50.7</b>	<b>3.27</b>	<b>5.4</b>
H135	2.53	19.1	n.a.	n.a.	<b>2.08</b>	<b>80.7</b>	n.a.	n.a.	<b>2.05</b>	<b>87.5</b>	n.a.	n.a.
E177	<b>2.29</b>	<b>85.7</b>	n.a.	n.a.	<b>2.14</b>	<b>95.3</b>	n.a.	n.a.	<b>2.13</b>	<b>92.9</b>	n.a.	n.a.

The occurrences of hydrogen bond formation are given in percent. Results stemming from purely QM described interactions are printed in bold font.



**FIGURE 6** | Plots of the distances between H134 and robotnikinin's macrocyclic amide group (black), ester (red), as well as the separation of H135 and the macrocyclic amide function (blue) of robotnikinin. Each frame represents a time span of 0.02 ps. **(A)** Core simulation. **(B)** Extended simulation (TZ). **(C)** Extended simulation (DZ).

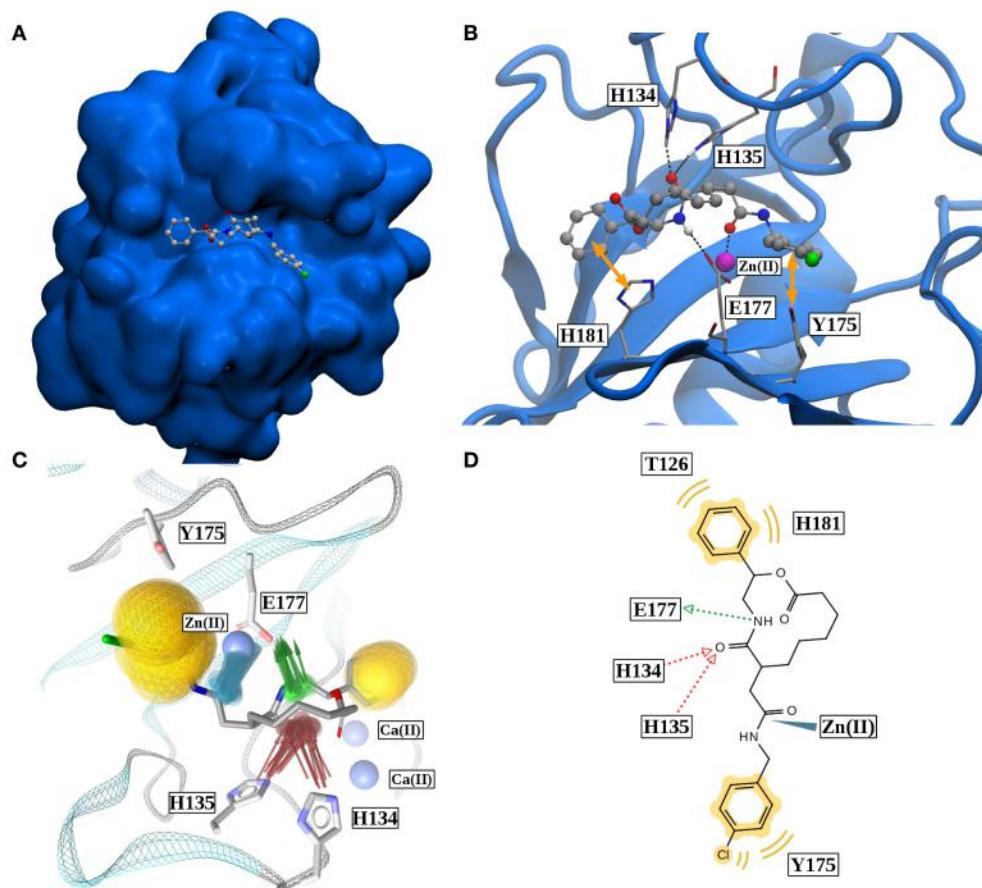


**FIGURE 7** | **(A)** The most common geometry of the hydrophobic interactions between T126, H181, and robotnikinin. **(B)** A configuration displaying T-shaped  $\pi$ - $\pi$  stacking of the aromatics belonging to robotnikinin and Y175, respectively. **(C)** The more frequent parallel-displaced stacking geometry between the chlorophenyl ring of robotnikinin and Y175 as witnessed during the QM/MM simulations.

two-dimensional interaction maps are shown in **Figures 8C,D**, confirming the observations from the distance and angle based trajectory analysis. Not shown in the pictures is residue K88, which is very close to robotnikinin's phenyl ring with a mean distance of 5.48 Å (core QM/MM simulation). It is conceivable that if the phenyl ring was modified with an additional negatively charged moiety then a hydrogen bond interaction with K88 could be achieved, further stabilizing the robotnikinin-Shh complex.

In order to assess how robust the predictions based on the DZ basis are, a QM/MM simulation of empty Shh was

conducted, utilizing the DZ basis (along with adjusted link-bond parameters) but other than that the same settings as in the previous study (Hitzenberger and Hofer, 2016). The results suggest that with a predicted Zn(II) coordination number of 5.5 (calculated up to a distance of 2.78 Å), where the ion is commonly coordinated by E177, along with H141, D148, H183 and one or two water molecules, in contrast to the results found by the TZ simulation (CN of 4.4 and very rare direct E177-Zn(II) bonding), that a double zeta basis might not be sufficient to accurately describe such a system when applied to a simulation utilizing the BP86 functional. Nevertheless, the obtained results



**FIGURE 8 | (A)** Binding pose of robotnikinin in Sonic Hedgehog (van der Waals surface). **(B)** Most important interactions between robotnikinin and Sonic Hedgehog, depicted in a snapshot of the extended TZ simulation. The Zn(II) ion is colored in magenta, H-bonds and metal-ligand interactions are represented as black, dashed lines, while hydrophobic interactions are shown as yellow arrows. With the exception of hydrogen atoms that are part of H-bonds, only heavy atoms are shown. **(C)** Superposition of 20 interaction profiles constructed from every fiftieth frame of the extended TZ sampling trajectory. Green arrows represent H-bond donors, while acceptors are colored in red, hydrophobic interactions are shown in yellow and ligand-metal interactions are depicted in blue. **(D)** Two-dimensional interaction profile of robotnikinin with Shh, adapted from the output of LigandScout. The color-coding is analogous to the one used in **(C)**.

are still much more accurate than those of the purely classical simulation, however, it is not as close to the experimental data as the TZ simulation. For this reason, all structurally relevant data for this study were taken from the TZ investigations. However, as the DZ simulation is nearly four-times faster than the TZ case and the results regarding hydrogen bond population very close to the data taken from the TZ investigations they can be viewed as a statistically more robust confirmation of the behavior witnessed in the shorter TZ simulation. The calculation of the solvent accessible surface area (SASA) of robotnikinin from the 1,000 configurations sampled in the extended TZ simulation resulted in a mean value of  $297.6 \text{ \AA}^2$  ( $\pm 0.7 \text{ \AA}^2$ ,  $P = 95\%$ ,  $SD = 12.0 \text{ \AA}^2$ ,  $n = 1,000$ ) which is very close to the  $300.5 \text{ \AA}^2$  ( $\pm 1.4 \text{ \AA}^2$ ,  $P = 95\%$ ,  $SD = 22.7 \text{ \AA}^2$ ,  $n = 1,000$ ), obtained from the classical simulation. This finding is unsurprising because no major configurational changes have been witnessed, instead the most notable differences between the QM/MM and MM structures concern areas which are buried in the Shh's binding groove in both cases.

For further reference, a representative PDB file containing the system (taken from the extended TZ simulation) is provided as Supplementary Material. In order to reduce the size of the file, all solvent molecules have been removed.

## 4. CONCLUSIONS

In the presented study, a reasonable binding mode for the small molecule inhibitor of Sonic Hedgehog, robotnikinin is suggested. The interactions, identified via a series of QM/MM MD simulations were sufficiently strong to stabilize the Shh-robotnikinin complex throughout the investigations and enabled a binding mode in which the ligand interacts with six amino acids and the Zn(II) ion present in the binding groove of Shh. The most important and stable interactions are the ionic bond between the Zn(II) ion of Sonic Hedgehog and one of the oxygen atoms of robotnikinin as well as a hydrogen bond between E177 and the macrocyclic amide group of robotnikinin,

bridging the inhibitor with the ion. Other very important contributions include hydrogen bonds between H134/H135 and the macrocyclic amide group as well as hydrophobic interactions between the aromatic rings of the ligand and the sidechains of residues T126, Y175 and H181, predominantly forming  $\pi-\pi$  interactions.

In all conducted simulations K88, also known to bind complex partners of Shh (Bosanac et al., 2009), is sufficiently close to the phenyl ring of robotnikinin in order to form an H-bond if an electron donor function were present at a suitable position, thereby potentially further improving the affinity of Shh to the small molecule inhibitor and in addition preventing the contribution of K88 to the formation of complexes with other proteins.

Even though in the core simulation the QM/MM interface very likely prevented donor-hydrogen-acceptor geometries commonly regarded as required to confirm the existence of hydrogen bonds between the classically described histidines H134/H135 and the QM described ligand, their behavior (low RMSD, close proximity to the inhibitor) was as a very strong hint for the presence of important interactions between those molecules hence justifying the utilization of the applied embedding scheme. It is, however, still very important to stress that in order to obtain reasonable structural data a careful choice of the QM zone is necessary making sure that all potential ligand protein interactions are treated at the same level of theory. However, as QM calculations scale very unfavorably with the number of atoms in the system, the size of the QM region has to be as small as possible in order to attain a sufficient number of configurations from which statistically robust data can be derived. Therefore, a reasonable compromise between method accuracy and system size is very important to sample a sufficient number of configurations that are also physically meaningful. In the presented work this could be achieved by conducting three different QM/MM simulations of the same system, differentiated by their respective basis set and system sizes.

## REFERENCES

- Aqvist, J. (1990). Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.* 94, 8021–8024. doi: 10.1021/j100384a009
- Åqvist, J., and Warshel, A. (1993). Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. *Chem. Rev.* 93:2523. doi: 10.1021/cr00023a010
- Amara, P., and Field, M. J. (2003). Evaluation of an ab initio quantum mechanical/molecular mechanical hybrid-potential link-atom method. *Theor. Chem. Acc.* 109, 43–52. doi: 10.1007/s00214-002-0413-3
- Barker, J. A., and Watts, R. O. (1973). Monte Carlo studies of the dielectric properties of water-like models. *Mol. Phys.* 26, 789–792. doi: 10.1080/00268977300102101
- Bash, P. A., Field, M. J., and Karplus, M. (1987). Free energy perturbation method for chemical reactions in the condensed phase: a dynamic approach based on a combined quantum and molecular mechanics potential. *J. Chem. Am. Soc.* 109, 8092–8094. doi: 10.1021/ja00260a028
- Becke, A. D. (1988). Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* 38:3098.
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. doi: 10.1063/1.448118

The purely classical model of the studied system, on the other hand, predicts a vastly over-coordinated Zn(II) site due to an overestimation of the ion-ligand interaction strength, which is probably also the main reason for the absence of the hydrogen bonds between robotnikinin and E177 or H134 and H135 (as they seem to promote each other, presumably due to the partial double bond character of the C-N bond in the amide). This can be viewed as a further indication that a description of the system based solely on force fields of the kind as employed in this study are not ideally suited for the physically correct description of systems involving transition metal ions. Overall, this work also highlights the capabilities of an iterative docking/(QM)MM MD cycle if used to improve the prediction power of *in silico* ligand-receptor binding studies.

## AUTHOR CONTRIBUTIONS

All simulations were executed and evaluated by MH. The simulations were performed using code written and implemented by TH who also contributed to the conceptualization of the investigation. The manuscript was drafted by MH and revised by TH and DS who was consulted due to her expertise in protein ligand docking.

## ACKNOWLEDGMENTS

The computational results presented have been achieved using the HPC infrastructure of the University of Innsbruck. The authors thank Inte:Ligand for providing LigandScout version 4.09 free of charge.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2017.00076/full#supplementary-material>

- Berman, M. D., Karhadkar, S. S., Maitra, A., de Oca, R. M., Gerstenblith, M. R., Briggs, K., et al. (2003). Widespread requirement for hedgehog ligand stimulation in growth of digestive tract tumors. *Nature* 425, 846–851. doi: 10.1038/nature01972
- Bishop, B., Aricescu, A. R., Harlos, K., O'Callaghan, C. A., Jones, E. Y. and Siebold, C. (2009). Structural insights into hedgehog ligand sequestration by the human hedgehog-interacting protein HHIP. *Nat. Struct. Biol. Mol. Biol.* 16, 698–703. doi: 10.1038/nsmb.1607
- Bosanac, I., Maun, H. R., Scales, S. J., Wen, X., Lingel, A., Bazan, J. F., et al. (2009). The structure of SHH in complex with HHIP reveals a recognition role for the SHH pseudo active site in signaling. *Nat. Struct. Biol. Mol. Biol.* 16, 691–698. doi: 10.1038/nsmb.1632
- Case, D. A., Babin, V., Berryman, J. T., Betz, R. M., Cai, Q., Cerutti, D. S., et al. (2014). *AMBER 14*. San Francisco, CA: University of California.
- Chemical Computing Group Inc. (2016). *Molecular Operating Environment (MOE); 2013.08*. Montreal, QC: Chemical Computing Group Inc.
- Clark, T., Chandrasekhar, J., and Schleyer, P. v. R. (1983). Efficient diffuse function-augmented basis sets for anion calculations. III. The 3–21+G basis set for first-row elements, Li–F. *J. Comput. Chem.* 4, 294–301. doi: 10.1002/jcc.540040303
- Darden, T., York, D., and Pedersen, L. (1993). Particle mesh ewald: An  $n \log(n)$  method for ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi: 10.1063/1.464397

- Day, E. S., Wen, D., Garber, E. A., Hong, J., Avedissian, L. S., Rayhorn, P., et al. (1999). Zinc-dependent structural stability of human sonic hedgehog. *Biochemistry* 38, 14868–14880. doi: 10.1021/bi9910068
- De Smaele, E., Ferretti, E., and Gulino, A. (2012). Vismodegib, a small-molecule inhibitor of the hedgehog pathway for the treatment of advanced cancers. *Curr. Opin. Inv. Drug* 11, 707–718.
- De Vivo, M., Masetti, M., Bottegoni, G., and Cavalli, A. (2016). Role of molecular dynamics and related methods in drug discovery. *J. Med. Chem.* 59, 4035–4061. doi: 10.1021/acs.jmedchem.5b01684
- Dunning, T. H. (1989). Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *J. Chem. Phys.* 90:1007. doi: 10.1063/1.456153
- Eurenus, K. O., Chatfield, D. C., Brooks, B. R., and Hodoscek, M. (1996). Enzyme mechanisms with hybrid quantum and molecular mechanical potentials. i. theoretical considerations. *Int. J. of Quant. Chem.* 60, 1189–1200. doi: 10.1002/(SICI)1097-461X(1996)60:6<1189::AID-QUA7>3.0.CO;2-W
- Franci, M. M., Pietro, W. J., Hehre, W. J., Binkley, J. S., Gordon, M. S., DeFrees, D. J., et al. (1982). Selfconsistent molecular orbital methods. XXIII. A polarizationtype basis set for secondrow elements. *J. Chem. Phys.* 77, 3654–3665. doi: 10.1063/1.444267
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2009). *Gaussian 09 Revision A.1*. Wallingford CT: Gaussian Inc.
- Fuse, N., Maiti, T., Wang, B., Porter, J. A., Tanaka Hall, T. M., Leahy, D. J., et al. (1999) Sonic hedgehog protein signals not as a hydrolytic enzyme but as an apparent ligand for patched. *Proc. Natl. Acad. Sci. U.S.A.* 96, 10992–10999. doi: 10.1073/pnas.96.20.10992
- Gill, P. M. W., Johnson, B. G., Pople, J. A., and Frisch, M. J. (1992). The performance of the BeckeLeeYangParr (BLYP) density functional theory with various basis sets. *Chem. Phys. Lett.* 197, 499–505. doi: 10.1016/0009-2614(92)85807-M
- Gillan, M. J., Alfè, D., and Michaelides, A. (2016). Perspective: How good is DFT for water? *J. Chem. Phys.* 144:130901. doi: 10.1063/1.4944633
- Goodrich, L. V., Milenkovic, L., Higgins, K. M., and Scott, M. P. (1997). Altered neural cell fates and medulloblastoma in mouse patched mutants. *Science* 277, 1109–1113. doi: 10.1126/science.277.5329.1109
- Grimme, S., Antony, J., Ehrlich, S., and Krieg, H. (2010). A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements h-pu. *J. Chem. Phys.* 132:154104. doi: 10.1063/1.3382344
- Hahn, H., Wicking, C., Zaphiropoulos, P. G., Gailani, M. R., Shanley, S., Chidambaram, A., et al. (1996). Mutations of the human homolog of drosophila patched in the nevoid basal cell carcinoma syndrome. *Cell* 85, 841–851. doi: 10.1016/S0092-8674(00)81268-4
- Hall, T. M., Porter, J. A., Beachy, P. A., and Leahy, D. J. (1995). A potential catalytic site revealed by the 1.7 Å crystal structure of the amino-terminal signalling domain of sonic hedgehog. *Nature* 378, 212–216. doi: 10.1038/378212a0
- Hamed, S., LaRue, H., Hovington, H., Girard, J., Jeannotte, L., Latulippe, E., et al. (2004). Accelerated induction of bladder cancer in patched heterozygous mutant mice. *Cancer Res.* 64, 1938–1942. doi: 10.1158/0008-5472.CAN-04-0416
- Hariharan, P. C., and Pople, J. A. (1973). The influence of polarization functions on molecular orbital hydrogenation energies. *Theoret. Chim. Acta* 28, 213–222. doi: 10.1007/BF00533485
- Hierao, H. (2011). Which DFT functional performs well in the calculation of methylcobalamin? Comparison of the B3LYP and BP86 functionals and evaluation of the impact of empirical dispersion correction. *J. Phys. Chem. A* 115, 9308–9313. doi: 10.1021/jp2052807
- Hitzenberger, M., and Hofer, T. S. (2015). Probing the range of applicability of structure- and energy-adjusted QM/MM link bonds. *J. Comput. Chem.* 39, 1929–1939. doi: 10.1002/jcc.24032
- Hitzenberger, M., and Hofer, T. S. (2016). The influence of metal-ion binding on the structure and surface composition of sonic hedgehog: a combined classical and hybrid QM/MM MD study. *Phys. Chem. Chem. Phys.* 18, 22254–22265. doi: 10.1039/C6CP03960J
- Hofer, T. S., Pribil, A. B., Randolph, B. R., and Rode, B. M. (2010). Ab initio quantum mechanical charge field molecular dynamics nonparametrized first-principle approach to liquids and solutions. *Adv. Quant. Chem.* 59, 213–246. doi: 10.1016/S0065-3276(10)59007-5
- Huggins, D. J. (2015). Quantifying the entropy of binding for water molecules in protein cavities by computing correlations. *Biophys. J.* 108, 928–936. doi: 10.1016/j.biophys.2014.12.035
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graphics* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5
- Hwang, S., Thangapandian, S., and Lee, K. W. (2013). Molecular dynamics simulations of sonic hedgehog-receptor and inhibitor complexes and their applications for potential anticancer agent discovery. *PLoS ONE* 8:e68271. doi: 10.1371/journal.pone.0068271
- Hwang, S., Thangapandian, S., Lee, Y., Sakkiah, S., John, S. and Lee, K. W. (2011). Discovery and evaluation of potential sonic hedgehog signaling pathway inhibitors using pharmacophore modeling and molecular dynamics simulations. *J. Bioinform. Comput. Biol.* 9, 15–35. doi: 10.1142/S0219720011005732
- Ingham, P. W., and McMahon, A. P. (2001). Hedgehog signaling in animal development: paradigms and principles. *Genes Dev.* 15, 3059–3087. doi: 10.1101/gad.938601
- Jeong, J., and McMahon, A. P. (2004). Growth and pattern of the mammalian neural tube are governed by partially overlapping feedback activities of the hedgehog antagonists patched 1 and HHIP1. *Development* 132, 143–154. doi: 10.1242/dev.01566
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869
- Krishnam, R., Binkley, J. S., Seeger, R., and Pople, J. A. (1980). Selfconsistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* 72, 650–654. doi: 10.1063/1.438955
- Kubo, M., Nakamura, M., Tasaki, A., Yamanaka, N., Nakashima, H., Nomura, M., et al. (2004). Hedgehog signaling pathway is a new therapeutic target for patients with breast cancer. *Cancer Res.* 64, 6071–6074. doi: 10.1158/0008-5472.CAN-04-0416
- Kuta, J., Patchkovskii, S., Zgierski, M. Z., and Kozlowski, P. M. (2006). Performance of DFT in modeling electronic and structural properties of cobalamins. *J. Comput. Chem.* 27, 1429–1437. doi: 10.1002/jcc.20454
- König, P. H., Hoffmann, M., Frauenheim, T. H., and Cui, Q. (2005). A critical evaluation of different QM/MM frontier treatments with SCC-DFTB as the QM method. *J. Phys. Chem. B* 109, 9082–9095. doi: 10.1021/jp0442347
- Ladbury, J. E. (1996). Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem. Biol.* 3, 973–980. doi: 10.1016/S1074-5521(96)90164-7
- Lepšík, M., and Field, M. J. (2007). Binding of calcium and other metal ions to the ef-hand loops of calmodulin studied by quantum chemical calculations and molecular dynamics simulations. *J. Phys. Chem. B* 111, 10012–10022. doi: 10.1021/jp0716583
- Li, Z., and Lazaridis, Z. (2005). The effect of water displacement on binding thermodynamics: concanavalin a. *J. Phys. Chem. B* 109, 662–670. doi: 10.1021/jp0477912
- Li, P., and Merz, K. M. (2014). Taking into account the ion-induced dipole interaction in the nonbonded model of ions. *J. Comput. Chem.* 10, 289–297. doi: 10.1021/ci400751u
- Lin, H., and Truhlar, D. G. (2007). QM/MM: what have we learned, where are we, and where do we go from here. *Theor. Chem. Acc.* 117, 185–199. doi: 10.1007/s00214-006-0143-z
- Lyne, P. D., Hodoscek, M., and Karplus, M. (1990). A hybrid QM-MM potential employing Hartree-Fock or density functional methods in the quantum region. *J. Phys. Chem. A* 103:3462. doi: 10.1021/jp982115j
- Maun, H. R. Wen, X., de Savage, F. J., Lazarus, R. A., Scales, S. J., and Hymowitz, S. G. (2010). Hedgehog pathway antagonist 5E1 binds hedgehog at the pseudo-active site. *J. Biol. Chem.* 285, 26570–26580. doi: 10.1074/jbc.M110.112284
- McLellan, J. S., Zheng, X., Hauk, G., Ghirlando, R., Beachy, P. A., and Leahy, D. J. (2008). The mode of hedgehog binding to IHOG homologues is not conserved across different phyla. *Nature* 455, 979–984. doi: 10.1038/nature07358
- Messner, C. B., Bonn, G. K., and Hofer, T. S. (2015). QM/MM MD simulations of la(iii)-phosphopeptide complexes. *Mol. BioSyst.* 11, 232–238. doi: 10.1039/C4MB00424H
- Mulliken, R. S. (1962). Criteria for the construction of good selfconsistentfield molecular orbital wave functions, and the significance of lcaomo population analysis. *J. Chem. Phys.* 36, 3428–3439. doi: 10.1063/1.1732476

- Nickolls, J., Buck, I., Garland, M., and Skadron, K. (2008). Scalable parallel programming with cuda. *ACM Queue* 6, 40–53. doi: 10.1145/1365490.1365500
- Palma, V., Lim, D. A., Dahmane, N., Sanchez, P., Brionne, T. C., Herzberg, C. D. et al. (2004). Sonic hedgehog controls stem cell behavior in the postnatal and adult brain. *Development* 132, 335–344. doi: 10.1242/dev.01567
- Perdew, J. P. (1986). Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* 33, 8822–8824. doi: 10.1103/PhysRevB.33.8822
- Radoń, M., and Pierloot, K. (2008). Binding of CO, NO, and O<sub>2</sub> to heme by density functional and multireference ab initio calculations. *J. Phys. Chem. A* 112, 11824–11832. doi: 10.1021/jp806075b
- Ren, X., Rinke, P., Blum, V., Wieferink, J., Tkatchenko, A., Sanfilippo, A., et al. (2012). Resolution-of-identity approach to HartreeFock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions. *New J. Phys.* 14:053020. doi: 10.1088/1367-2630/14/5/053020
- Rimkus, T. K., Carpenter, R. L., Qasem, S., Chan, M., and Lo, H. (2016). Targeting the sonic hedgehog signaling pathway: Review of smoothed and GLI inhibitors. *Cancers* 8, 22–45. doi: 10.3390/cancers8020022
- Rode, B. M., Hofer, T. S., Randolph, B. R., Schwenk, C., Xenides, D., and Vchirawongkwin, V. (2006). Ab initio quantum mechanical charge field (QMCF) molecular dynamics: a QM/MM MD procedure for accurate simulations of ions and complexes. *Theor. Chem. Acc.* 115, 77–85. doi: 10.1007/s00214-005-0049-1
- Rowley, C. N., and Roux, B. (2012). The solvation structure of Na<sup>+</sup> and K<sup>+</sup> in liquid water determined from high level ab initio molecular dynamics simulations. *J. Chem. Theor. Comput.* 8, 3526–3535. doi: 10.1021/ct300091w
- Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23, 327–341. doi: 10.1016/0021-9991(77)90098-5
- Ryde, U., and Grimme, S. (2011). Does DFT-D estimate accurate energies for the binding of ligands to metal complexes? *Dalton Trans.* 40, 11176–11183. doi: 10.1039/c1dt10867k
- Schwenk, C. F., Hofer, T. S., and Rode, B. M. (2004). Structure breaking effect of hydrated Cs<sup>+</sup>. *J. Phys. Chem. A* 108, 1509–1514. doi: 10.1021/jp037179v
- Senn, H. M., and Thiel, W. (2009). QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.* 48, 1198–1229. doi: 10.1002/anie.200802019
- Singh, U. C., and Kollman, P. A. (1984). An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* 5, 129–145. doi: 10.1002/jcc.540050204
- Stanton, B. Z., Peng, L. F., Maloof, N., Nakai, K., Wang, X., Duffner, J. L., et al. (2009). A small molecule that binds hedgehog and blocks its signaling in human cells. *Nat. Chem. Biol.* 5, 154–156. doi: 10.1038/ncembio.142
- Swope, W. C., Andersen, H. C., Berens, P. H., and Wilson, K. R. (1982). A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: application to small water clusters. *J. Chem. Phys.* 76, 637–649. doi: 10.1063/1.442716
- Turbomole (2007). *TURBOMOLE V6.5 2013, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH*. Available online at: <http://www.turbomole.com>
- Varjosalo, M., and Taipale, J. (2008). Hedgehog: functions and mechanisms. *Genes Dev.* 22, 2454–2472. doi: 10.1101/gad.1693608
- Warshel, A., and Levitt, M. (1976). Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbenium ion in the reaction of lysozyme. *J. Mol. Biol.* 103:227. doi: 10.1016/0022-2836(76)90311-9
- Wieder, M., Garon, A., Perricone, U., Boresch, S., Seidel, T., Almerico, A. M., et al. (2017). Common hits approach: Combining pharmacophore modeling and molecular dynamics simulations. *J. Chem. Inf. Model.* 57, 365–385. doi: 10.1021/acs.jcim.6b00674
- Wiegend, F., and Ahlrichs, R. (2005). Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: design and assessment of accuracy. *Phys. Chem. Chem. Phys.* 7, 3297–3305. doi: 10.1039/b508541a
- Wober, G., and Langer, T. (2005). Ligandscout: 3-d pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* 45, 160–169. doi: 10.1021/ci049885e
- Yoo, S., Zeng, X. C., and Xantheas, S. S. (2009). On the phase diagram of water with density functional theory potentials: The melting temperature of ice Ih with the perdewburkeerzherhof and beckleeyangparr functionals. *J. Chem. Phys.* 130:221102. doi: 10.1063/1.3153871
- Zgarbova, M., Otyepka, M., Sponer, J., Mladek, A., Banas, P., Cheatham, T. E., et al. (2011). Refinement of the cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theor. Comput.* 7, 2886–2902. doi: 10.1021/ct200162x

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Hitzenberger, Schuster and Hofer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# How Are Substrate Binding and Catalysis Affected by Mutating Glu<sub>127</sub> and Arg<sub>161</sub> in Prolyl-4-hydroxylase? A QM/MM and MD Study

Amy Timmins and Sam P. de Visser\*

School of Chemical Engineering and Analytical Science, Manchester Institute of Biotechnology, The University of Manchester, Manchester, United Kingdom

Prolyl-4-hydroxylase is a vital enzyme for human physiology involved in the biosynthesis of 4-hydroxyproline, an essential component for collagen formation. The enzyme performs a unique stereo- and regioselective hydroxylation at the C<sup>4</sup> position of proline despite the fact that the C<sup>5</sup> hydrogen atoms should be thermodynamically easier to abstract. To gain insight into the mechanism and find the origin of this regioselectivity, we have done a quantum mechanics/molecular mechanics (QM/MM) study on wildtype and mutant structures. In a previous study (Timmins et al., 2017) we identified several active site residues critical for substrate binding and positioning. In particular, the Glu<sub>127</sub> and Arg<sub>161</sub> were shown to form multiple hydrogen bonding and ion-dipole interactions with substrate and could thereby affect the regio- and stereoselectivity of the reaction. In this work, we decided to test that hypothesis and report a QM/MM and molecular dynamics (MD) study on prolyl-4-hydroxylase and several active site mutants where Glu<sub>127</sub> or Arg<sub>161</sub> are mutated for Asp, Gln, or Lys. Thus, the R161D and R161Q mutants give very high barriers for hydrogen atom abstraction from any proline C–H bond and therefore will be inactive. The R161K mutant, by contrast, sees the regio- and stereoselectivity of the reaction change but still is expected to hydroxylate proline at room temperature. By contrast, the Glu<sub>127</sub> mutants E127D and E127Q show possible changes in regioselectivity with the former being more probable to react compared to the latter.

## OPEN ACCESS

**Edited by:**

Nino Russo,  
Università della Calabria, Italy

**Reviewed by:**

Tiziana Marino,  
University of Calabria, Italy  
Xavier Assfeld,  
Université de Lorraine, France

**\*Correspondence:**

Sam P. de Visser  
sam.devisser@manchester.ac.uk

**Specialty section:**

This article was submitted to  
Theoretical and Computational  
Chemistry,  
a section of the journal  
*Frontiers in Chemistry*

**Received:** 13 September 2017

**Accepted:** 24 October 2017

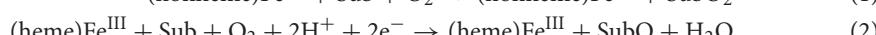
**Published:** 09 November 2017

**Citation:**

Timmins A and de Visser SP (2017)  
How Are Substrate Binding and  
Catalysis Affected by Mutating Glu<sub>127</sub>  
and Arg<sub>161</sub> in Prolyl-4-hydroxylase? A  
QM/MM and MD Study.  
*Front. Chem.* 5:94.  
doi: 10.3389/fchem.2017.00094

## INTRODUCTION

Metalloenzymes play vital roles in nature and are involved in biosynthesis as well as biodegradation of compounds (Solomon et al., 2000; Costas et al., 2004; Abu-Omar et al., 2005; Kryatov et al., 2005; Bruijnincx et al., 2008; Kadish et al., 2010). Due to its large natural abundance often metalloenzymes contain one or more iron centers; however, in this work we will restrict ourselves to mononuclear iron enzymes only and particularly those that utilize molecular oxygen. In general, iron containing dioxygenases and monooxygenases use one molecule of molecular oxygen in their catalytic cycle and either transfer both oxygen atoms to substrate(s) or a single one with a water molecule as by-product (Sub = substrate), Equations 1, 2.

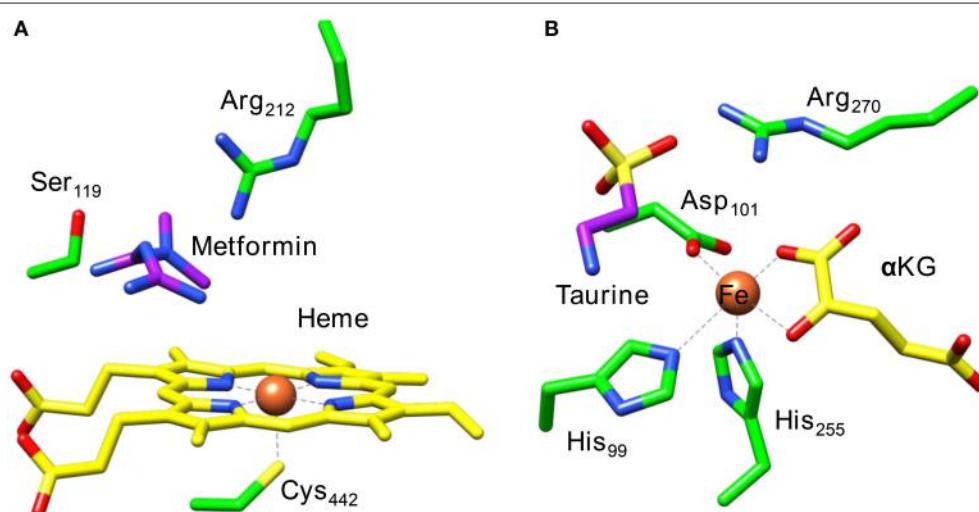


Thus, heme monooxygenases, like the cytochromes P450, react as monooxygenases and proceed through a catalytic cycle starting from an iron(III)-heme resting state with a protein cysteinate and a water molecule in the fifth and sixth iron ligand positions, respectively (Meunier et al., 2004; Ortiz de Montellano, 2004, 2010; Denisov et al., 2005; Kadish et al., 2010). The water molecule is released after substrate binding, which triggers a spin state change from low-spin to high-spin and enables molecular oxygen binding to the iron center. The iron-superoxide is subsequently reduced and protonated to form an iron(III)-hydroperoxo(heme) complex, also called Compound 0 (Meunier et al., 2004; Denisov et al., 2005; Shaik et al., 2005; Ortiz de Montellano, 2010). A final protonation step gives water and an iron(IV)-oxo(heme cation radical) species called Compound I (CpdI) (de Visser et al., 2003; Rittle and Green, 2010; de Visser and Kumar, 2011). Now, CpdI is the active species of P450 enzymes and reacts with substrates through oxygen atom transfer and hence converts aliphatic groups to alcohols (Ogliaro et al., 2000b; de Visser et al., 2004; Ji et al., 2015), C = C double bonds to epoxides (de Visser et al., 2001, 2002b; Sainna et al., 2015), sulfides to sulfoxides (Sharma et al., 2003; Kumar et al., 2005b, 2011a), and arenes to phenols (de Visser and Shaik, 2003; de Visser, 2006c; Cantú Reinhard et al., 2016b). The mechanism of these reactions has been established with computational modeling including density functional theory (de Visser, 2012; Blomberg et al., 2014). In recent years full calculations on enzymatic structures were done and identified the effect of the protein, substrate orientation and hydrogen bonding interactions on the kinetics and thermodynamics of the reaction and the product distributions of P450 catalyzed reaction mechanisms. Thus, it was shown with small model complexes that hydrogen bond donations toward the axial thiolate ligand affected the electron affinity of this cysteinate residue, which led to a push-effect of electrons to the heme that influenced its redox potential and hence catalytic potential (Ogliaro et al.,

2000a; de Visser et al., 2002a; Schöneboom et al., 2002). In the heme enzyme cytochrome *c* peroxidase, a model that included a cation binding site reproduced the experimentally characterized electronic configuration, and highlighted the importance of long-range electrostatic effects in enzyme models (de Visser, 2005). Because of these long-range effects, more and more computational studies are done using full enzymatic systems using the Quantum Mechanics/Molecular Mechanics (QM/MM) technique.

**Figure 1** displays the active site structures of **(Figure 1A)** cytochrome P450 and **(Figure 1B)** taurine/α-ketoglutarate dioxygenase (TauD) as a structural comparison (O'Brien et al., 2003; Guo and Sevrioukova, 2017). Thus, the P450s are heme enzymes, where the heme is linked to the protein backbone through an interaction of the metal with a cysteinate residue (the axial ligand). On the distal site of the heme the substrate binds, which is the drug molecule metformin in the 5G5J protein databank (pdb) file. The distal site of the heme has several hydrogen bonding and polar residues, such as Ser<sub>119</sub> and Arg<sub>212</sub>, the former has been proposed to be involved in the proton relay mechanisms during the catalytic cycle (Kumar et al., 2005a), whereas the latter holds the substrate through a salt bridge into position.

A second iron enzyme class that utilizes molecular oxygen is the nonheme iron dioxygenases (Bugg, 2001; Ryle and Hausinger, 2002; Solomon et al., 2013). These dioxygenases are found in all forms of life and are involved in the biosynthesis of antibiotics (Choroba et al., 2000; Higgins et al., 2005; Siitonnen et al., 2016), DNA and RNA repair enzymes (O'Brien, 2006; Yi et al., 2009), as well as the metabolism of toxic natural compounds such as cysteine (Stipanuk, 2004; Straganz and Nidetzky, 2006; de Visser, 2009; Buongiorno and Straganz, 2013). These enzymes are structurally very different from the heme monooxygenases as they link the iron atom to the protein with only amino acid side chains such as His, Asp, or Glu residues. Usually, the nonheme



**FIGURE 1** | Extracts from the crystal structure coordinates of cytochrome P450 **(A)** and taurine/α-ketoglutarate dioxygenase **(B)** as taken from the 5G5J and 1OS7 pdb files.

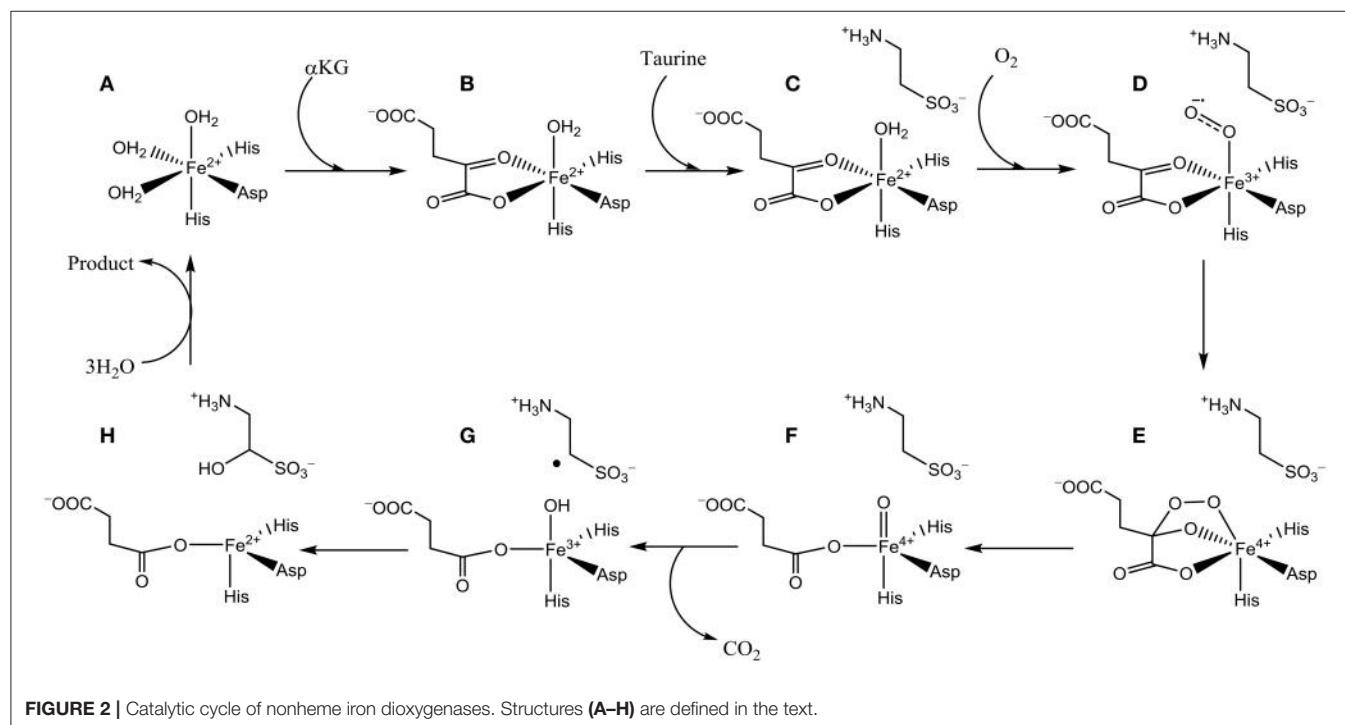
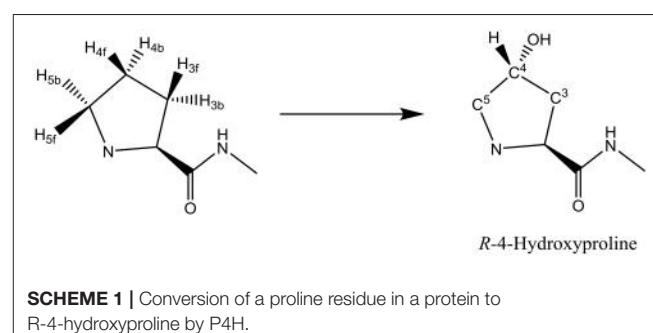
iron dioxygenases contain a facial triad of amino acid ligands with two histidine and one carboxylate group, i.e., 2-His/1-Asp, Glu (Que, 2000; Kovaleva and Lipscomb, 2008). As an example of a dioxygenase with these ligand features we show in **Figure 1B** the active site of taurine/α-ketoglutarate dioxygenase (TauD). TauD is a dioxygenase involved in the metabolism of cysteine, whereby it converts taurine to hydroxy-taurine. The 1OS7 pdb file (O'Brien et al., 2003) is a substrate and α-ketoglutarate (α-KG) bound structure of TauD with the iron bound to the protein through the side chains of residues His<sub>99</sub>, Asp<sub>101</sub> and His<sub>255</sub>. Substrate taurine is located nearby the metal and is held in position through a salt bridge with residue Arg<sub>270</sub>. Co-substrate α-KG is bound to the metal as a bidentate ligand through the keto and acid groups.

The catalytic cycle of TauD has been established through a combination of experimental and computational studies (Borowski et al., 2004; Bollinger et al., 2005; de Visser, 2006a,b, 2007; Godfrey et al., 2008). **Figure 2** schematically depicts the catalytic cycle of TauD specifically and starts from the resting state structure where iron is bound to the 2-His/1-Asp ligand system and the other ligand positions of the metal are occupied by three water molecules (structure A). When α-KG enters the pocket two water molecules are displaced and replaced by the keto and acid groups of α-KG (structure B). In the next step, substrate taurine binds, which displaces the last water molecule from iron (structure C) and is replaced by molecular oxygen that binds as an iron(III)-superoxide (structure D). Subsequently, the superoxide group attacks the α-keto position of α-KG to form a bicyclic ring-structure (structure E). In the next step, the dioxygen bond breaks to form a peracid succinate with the release of CO<sub>2</sub>. Finally, the peracid bond breaks and splits into an

iron(IV)-oxo species and succinate (structure F). Iron(IV)-oxo is known to be a powerful oxidant that abstracts a hydrogen atom from taurine to give an iron(III)-hydroxo group (structure G) and the hydroxyl radical is then rebound to form hydroxy-taurine as product (structure H). Products hydroxy-taurine and succinate are released from the iron center and their positions are replaced by water molecules to bring the catalytic cycle back into the resting state.

Another nonheme iron dioxygenase with a catalytic cycle similar to TauD is prolyl-4-hydroxylase (P4H), which regio- and stereospecifically hydroxylates a proline residue in a protein to R-4-hydroxyproline, **Scheme 1**. Product R-4-hydroxyproline is a common amino acid in animals and plants and has functions in collagen, where it enables crosslinking between individual strands. In addition, it is relevant to the synthesis of the hypoxia induced factor in animals (McDonough et al., 2006).

A range of biochemical and spectroscopic studies on P4H established key details of the catalytic cycle. Thus, reactions



of P4H with (de Visser et al., 2001)  $^{18}\text{O}_2$  provided evidence of the transfer of one atom of molecular oxygen to proline (Myllyharju and Kivirikko, 1997). Low-temperature Mössbauer, electron paramagnetic resonance (EPR) and UV-Vis absorption spectroscopic studies characterized several intermediates in the catalytic cycle, including the iron(IV)-oxo species (Hoffart et al., 2006). It was shown that the iron(IV)-oxo species has a quintet spin ground state and reacts with the substrate through a rate-determining hydrogen atom abstraction. In particular, rate constants for the reaction with taurine and taurine-d<sub>2</sub> gave a large kinetic isotope effect. To confirm the reaction mechanism computational studies on the catalytic cycle of P4H were performed: One study using an active site model complex (Karamzadeh et al., 2010) and another using the full enzyme structure with QM/MM (Timmins et al., 2017). These studies established the technical details of the catalytic cycle and confirmed the mechanism shown above in **Figure 2**. Furthermore, key functions of several amino acids were identified related to substrate positioning and product release as will be described in more detail later.

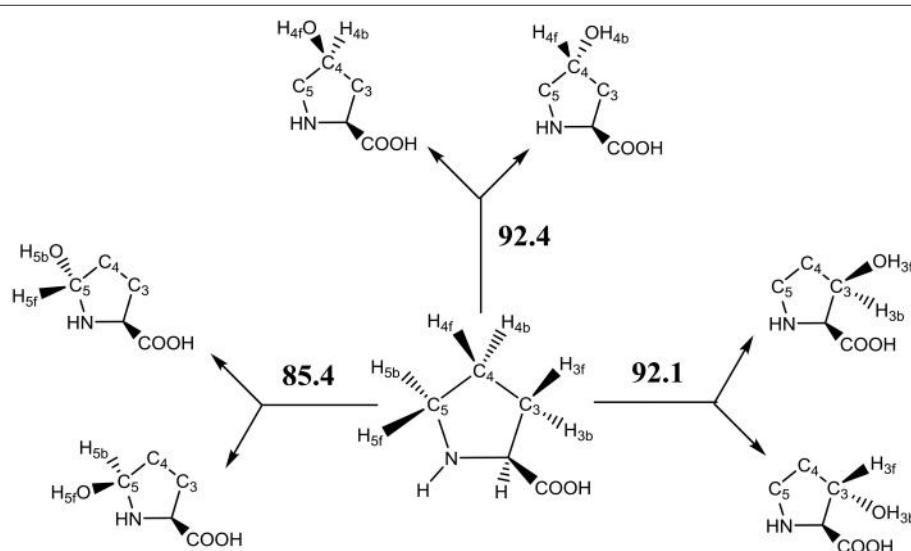
The key step in the catalytic cycle of P4H is the hydrogen atom abstraction of substrate by the iron(IV)-oxo intermediate (Hoffart et al., 2006), which was shown to be rate-determining. In principle, substrate proline has six aliphatic C-H bonds at positions C<sup>3</sup>, C<sup>4</sup>, and C<sup>5</sup> that could lead to six different product isomers, **Scheme 2**. We label the two hydrogen atoms on C<sup>3</sup>, C<sup>4</sup>, and C<sup>5</sup> as front (f) or back (b). Also shown in **Scheme 2** are bond dissociation free energies (BDFE) of each of these C-H bonds as calculated at UB3LYP/6-311+G\* as the difference in free energy of proline and the sum of a hydrogen atom and [proline - H<sup>•</sup>]. As can be seen, the C-H bond strength at the C<sup>3</sup> and C<sup>4</sup> positions in proline are comparable, while the one at the C<sup>5</sup> position is much weaker in energy. Therefore, in the gas-phase proline hydroxylation should happen at the C<sup>5</sup> position as it is

the weakest bond to break rather than at the thermodynamically unfavorable C<sup>4</sup> position. How P4H prevents hydroxylation of the weaker C<sup>5</sup> position in favor of hydroxylation at the C<sup>4</sup> position is the topic of this paper.

In addition to studies on wildtype (WT) P4H, we looked at the structure and catalytic properties of two active site mutants where the Glu<sub>127</sub> and Arg<sub>161</sub> residues were mutated to alternative groups. A previous study (Timmins et al., 2017) identified these two amino acids as key for substrate positioning in the substrate binding pocket and hence mutating them to a different amino acid should have a considerable effect.

## METHODS

The calculations presented in this work follow previously described and benchmarked methods using QM/MM (Porro et al., 2009; Kumar et al., 2011b; Quesne et al., 2014, 2016a; Faponle et al., 2017; Li et al., 2017). Specifically, our previous work on the mechanism of the possible reaction channels of cytochrome P450 decarboxylase leading to decarboxylation of fatty acids or hydroxylation of fatty acids, predicted the correct regioselectivity of the reaction as compared to experiment and reproduced experimentally determined kinetic isotope effects (Faponle et al., 2016). Furthermore, QM/MM studies on 1-H-3-hydroxy-4-oxoquininaline 2,4-dioxygenase focused on the rate-determining step of the co-factor independent reaction of substrate with molecular oxygen and predicted a rate constant in good agreement with experiment and explained how this enzyme functions without a metal cofactor present (Hernández-Ortega et al., 2014, 2015). Very recently, we used QM/MM modeling to predict spectroscopic fingerprints of short-lived catalytic cycle intermediates and used this on cysteine dioxygenase enzymes. Calculated UV-Vis absorption spectra and Mössbauer



**SCHEME 2** | Aliphatic C-H bonds of proline and definition of atom labels and their bond dissociation free energies (BDFE in kcal mol⁻¹).

and EPR parameters enabled the experimental characterization of a short-lived oxygen-bound intermediate (Fellner et al., 2016; Tchesnokov et al., 2016).

## Model Set-Up

Our QM/MM starting point structures were set-up using previously described methods and procedures (Quesne et al., 2016a; Timmins et al., 2017), which we will summarize briefly here. The crystal structure coordinates from the 3GZE pdb file was used as a starting point for all models (Koski et al., 2009). The 3GZE pdb file represents a resting state P4H structure with  $Zn^{2+}$ , pyridine-dicarboxylate co-substrate mimic and the (Ser-Pro)<sub>5</sub> peptide chain bound. The active site zinc(II)-water(pyridine-dicarboxylate) was replaced with iron(IV)-oxo(succinate) manually with an Fe-O distance of 1.63 Å: a typical distance found for analogous nonheme iron(IV)-oxo complexes in enzymes and model complexes (de Visser, 2006a,b,d, 2007; Godfrey et al., 2008; Quesne et al., 2016b; Cantú Reinhard and de Visser, 2017). The short peptide chain (Ser-Pro)<sub>5</sub> we retained in the model as it has its proline residue tightly packed nearby the iron(IV)-oxo group.

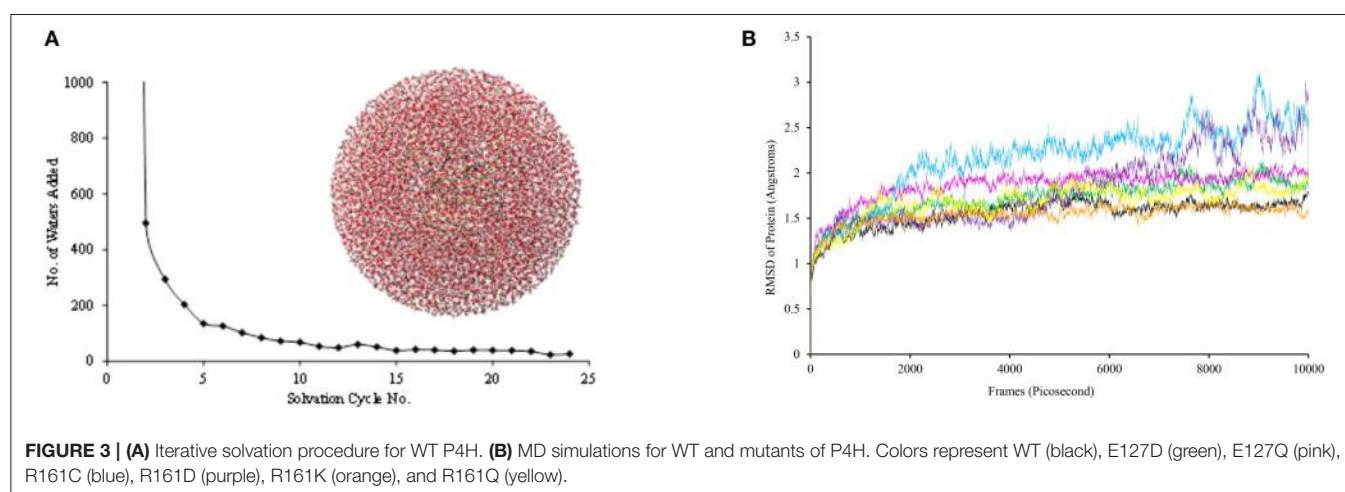
Subsequently, hydrogen atoms were added to the protein structure using the pdbtopqr program package assuming a pH = 7 (Dolinsky et al., 2007). Thus, all acid residues, i.e., Glu and Asp, were deprotonated whereas the basic residues, i.e., Arg and Lys, were protonated. The protonation state of each individual histidine residue was decided upon visual inspection of its local environment (donating/accepting hydrogen bonds) and we chose to assign all as singly protonated. Thereafter, the protein structure was solvated in a sphere with radius of 40 Å and energy minimized with the Charmm forcefield (Brooks et al., 1983). The solvation procedure was repeated a number of times until a situation was reached (Figure 3), whereby <20 water molecules were added to the chemical system. The saturated structure was then minimized without geometric constraints and heated to a temperature of 298 K. Finally, a full molecular dynamics (MD) simulation was run for 10 ns. The full set-up procedures were repeated for the mutant structures, whereby one amino acid was manually replaced.

As follows from the MD simulations shown in Figure 3B, all converge well within 10 ns. For each of the structures, we started QM/MM calculations using the snapshots taken after 5 ns ( $Sn_{5ns}$ ).

## QM/MM Procedures

Density functional theory (DFT) methods were used to describe the QM region of the QM/MM calculation. In particular, we used the unrestricted hybrid density functional method B3LYP (Lee et al., 1988; Becke, 1993) in all cases as recent benchmark studies from our group showed this procedure to give rate constants in very good agreement with experiment (Cantú Reinhard et al., 2016a). In particular, QM methods with dispersion included were shown to underestimate free energies of activation considerably (Cantú Reinhard et al., 2016a). Furthermore, B3LYP was shown previously to predict regioselectivities and bifurcation pathways well as compared to experiment (Kumar et al., 2004; Barman et al., 2016; Brazzolotto et al., 2017). Also, DFT calculated free energies of activation were shown to match experimentally determined ones of biomimetic model complexes containing iron and manganese very well and reproduced Hammett trends (Vardhaman et al., 2011, 2013; Kumar et al., 2014; Yang et al., 2016). Here, DFT calculations were run in Turbomole (Ahlrichs et al., 1989), and the MM ones in DL-Poly with the Charmm forcefield (Smith and Forester, 1996). The ChemShell software package (Sherwood et al., 2003) interfaced Turbomole and DL-Poly and was used to obtain QM/MM energies and derivatives. The link-atom approach was used to describe atoms on the border between the QM and MM regions and essentially replaced a covalent bond with a C-H bond (Bakowies and Thiel, 1996). All calculations use electronic embedding of the charges of the MM region included into the QM Hamiltonian.

Geometry optimizations and reaction coordinate scans were done with an SV(P) basis set on all atoms: basis set BSI (Schafer et al., 1992). Reaction coordinate scans were run with one degree of freedom fixed and explored the potential energy surface between reactants, intermediates and products. The maxima of these scans were used as starting points for transition state



**FIGURE 3 | (A)** Iterative solvation procedure for WT P4H. **(B)** MD simulations for WT and mutants of P4H. Colors represent WT (black), E127D (green), E127Q (pink), R161C (blue), R161D (purple), R161K (orange), and R161Q (yellow).

searches. The energies of the stationary points were improved by running a single point calculation with an all-electron Wachters-type basis set on iron and def2-TZVP on the rest of the atoms: basis set BSII (Wachters, 1970).

## QM Region

As the active site region of the protein contains many hydrogen bonding and  $\pi$ -stacking interactions we considered two QM regions: a minimal QM region A and an expanded QM region AB, see **Figure 4**. Thus, the minimal QM region A contains the iron(IV)-oxo group and its direct ligands (His<sub>143</sub>, His<sub>227</sub>, Asp<sub>145</sub>, and succinate) as well as the proline ring of the peptide substrate. The larger QM region AB was expanded with the indole ring of Trp<sub>243</sub>, the phenol group of Tyr<sub>140</sub> and three water molecules.

## RESULTS AND DISCUSSION

### P4H WT Structure

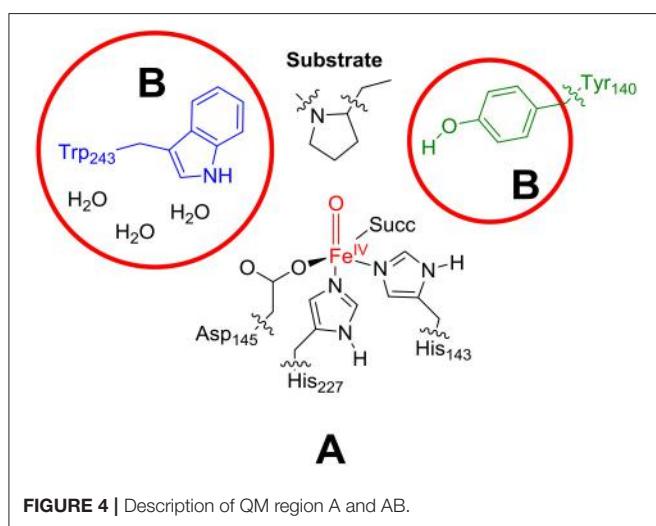
The iron(IV)-oxo species (structure G in the catalytic cycle of **Figure 2**) was fully optimized with QM/MM methods using B3LYP/BSI and QM region A and AB, see **Figure 5**. In agreement with experimental EPR and Mössbauer spectroscopic studies (Hoffart et al., 2006) the quintet spin state is the ground state, while we located the triplet and singlet spin states higher in energy by 16.0 and 33.1 kcal mol<sup>-1</sup>, respectively. Geometrically, the iron(IV)-oxo species is bound to the protein with two Fe-N<sub>His</sub> interactions of 2.06 and 2.08 Å, which is typical for metal-histidine interactions in proteins (de Visser et al., 2009). The carboxylate group of Asp<sub>145</sub> binds as a monodentate ligand at a distance of 2.01 Å, whereas the succinate carboxylate group binds as a bidentate ligand with distances of 2.20 and 2.30 Å. Again, these distances match previous calculations on similar complexes nicely (Pratter et al., 2013). Substrate proline is not bound directly to the iron center but its transferring C-H<sup>4b</sup> hydrogen atom is found at a distance of 2.86 Å from the oxo group and hence is positioned in the ideal orientation for oxidation. The triplet and singlet spin states give analogous ligand distances but are

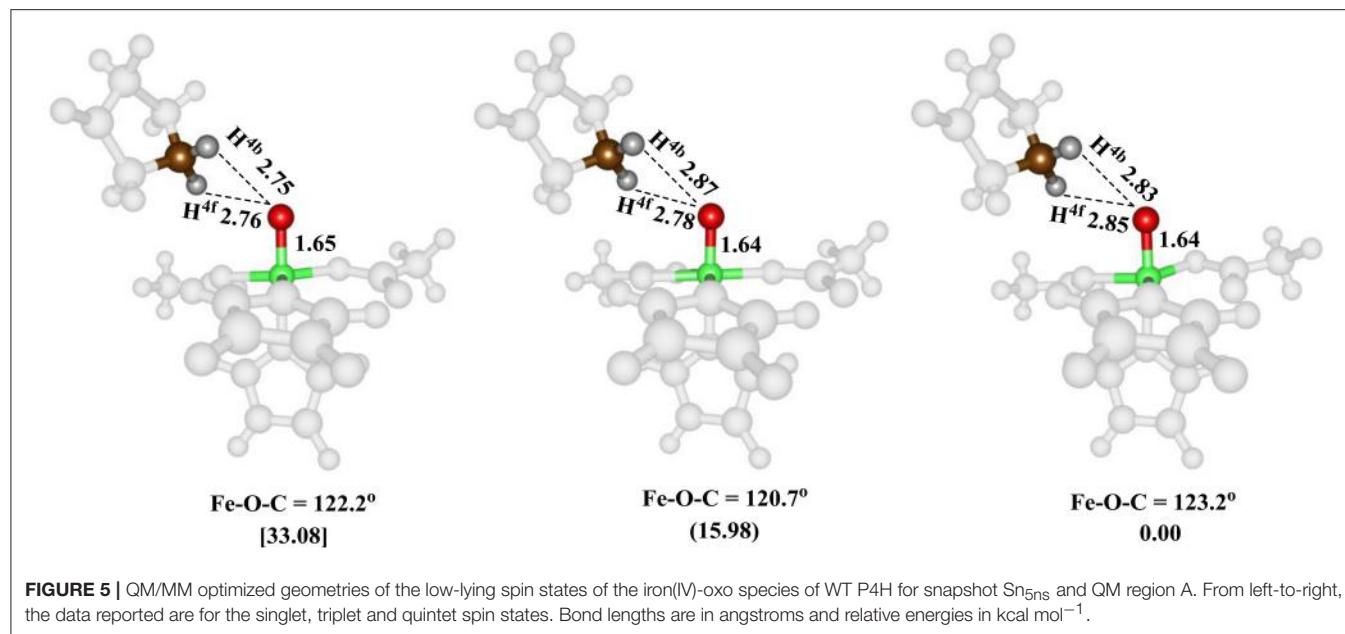
distinguished by their differences in iron(IV)-oxo bond length due to differences in molecular orbital occupation.

High-lying occupied and low-lying virtual orbitals of the iron(IV)-oxo species are shown in **Figure 6**, which gives the molecular z-axis along the Fe-O bond. The lowest lying orbital is the  $\pi^*_{xy}$  orbital that represents the interactions of the metal 3d<sub>xy</sub> orbital with the equatorial ligands, namely His<sub>143</sub>, Asp<sub>145</sub> and succinate (Succ). A bit higher in energy are the two orthogonal  $\pi^*$  orbitals along the Fe-O bond that correspond to the mixing of the 3d<sub>xz</sub> on iron with 2p<sub>x</sub> on oxygen, i.e.,  $\pi^*_{xz}$ , and the 3d<sub>yz</sub> on iron with the 2p<sub>y</sub> on oxygen, i.e.,  $\pi^*_{yz}$ . Higher in energy still are two  $\sigma$ -type orbital interactions. The first one along the z-axis for the mixing of the metal 3d<sub>z2</sub> with a 2p<sub>z</sub> on oxygen:  $\sigma^*_{z2}$ . The second one is located in the xy-plane and results from the interaction of the 3d<sub>x2-y2</sub> orbital on iron with orbitals on the ligands:  $\sigma^*_{x2-y2}$ . The quintet spin state has orbital occupation  $\pi^*_{xy} \pi^*_{xz} \pi^*_{yz} \sigma^*_{x2-y2} \sigma^*_{z2}$ , while the triplet spin state has configuration  $\pi^*_{xy} \pi^*_{xz} \pi^*_{yz} \sigma^*_{x2-y2} \sigma^*_{z2}$ . Generally, enzymatic iron(IV)-oxo species tend to have a quintet spin ground state (Latifi et al., 2009), while most synthetic biomimetic models have a triplet spin ground state (de Visser et al., 2012). Computational studies showed that this is the result of differences in coordination system, where biomimetic models are often in octahedral coordination, whereas enzymatic structures have the iron(IV)-oxo in pentacoordination (Latifi et al., 2013).

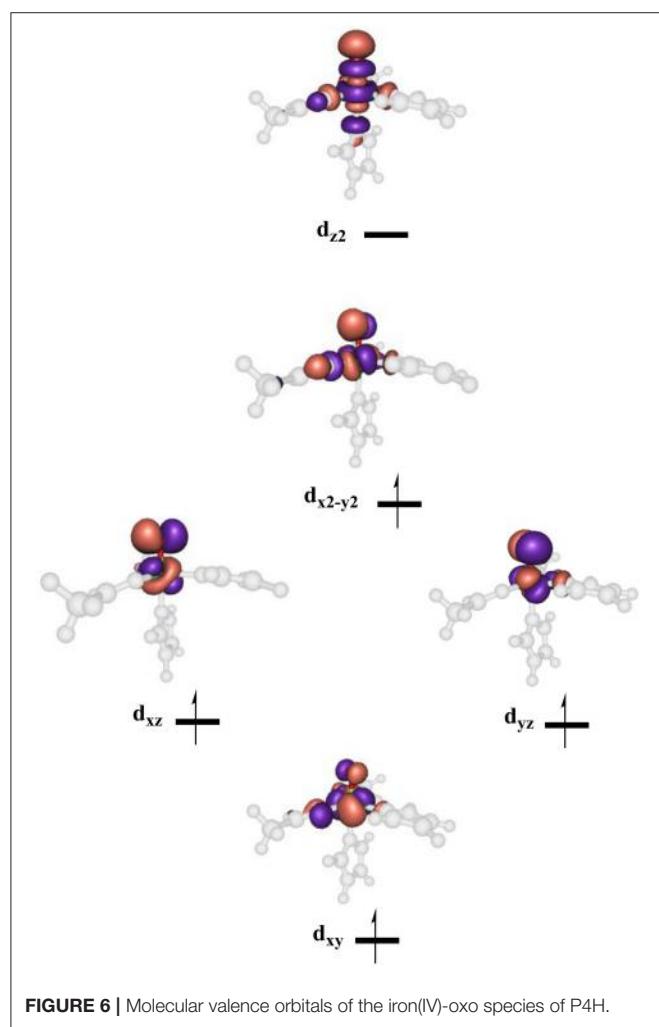
### P4H Hydroxylation Mechanisms

Subsequently, we investigated proline hydroxylation at the C<sup>3</sup>, C<sup>4</sup> and C<sup>5</sup> position of P4H and considered both hydrogen atoms at each of these positions. All reactions were found to be stepwise with an initial hydrogen atom abstraction via transition state TS<sub>HA</sub> to form a radical intermediate I<sub>H</sub>. A radical rebound step via TS<sub>reb</sub> then produced the alcohol product complexes P. In all cases the rebound step was small and the hydrogen atom abstraction was rate-determining. The QM/MM calculated energy landscapes are given in **Figure 7** for hydrogen atom abstraction from C<sup>4f</sup>, C<sup>4b</sup>, C<sup>5f</sup>, C<sup>5b</sup>, C<sup>3f</sup>, and C<sup>3b</sup>. The lowest energy barrier height is the one for C<sup>4b</sup> and after rebound will give the R-4-hydroxyproline product complex, which is the experimentally determined stereo- and regioselective product found. Although the difference in relative energies is small, the calculations predict the correct regio- and stereoselectivity. Slightly higher in energy (1 kcal mol<sup>-1</sup> using basis set BSII and 2.9 kcal mol<sup>-1</sup> using basis set BSI) we find the pathway for hydrogen atom abstraction from C<sup>5b</sup>, the thermodynamically more favorable pathway. Indeed the radical intermediate for C<sup>5b</sup> is much lower in energy than the one for C<sup>4b</sup> in agreement with the thermodynamic predictions. About 5.5 kcal mol<sup>-1</sup> higher in energy than <sup>5</sup>TS<sub>HA,C4b</sub> is the barrier <sup>5</sup>TS<sub>HA,C4f</sub>, which implies that there are strong energetic differences between hydrogen atom abstraction of the two hydrogen atoms on carbon center C<sup>4</sup>. The three barriers for hydrogen atom abstraction from C<sup>5f</sup>, C<sup>3f</sup>, and C<sup>3b</sup> are all well higher in energy than <sup>5</sup>TS<sub>HA,C4b</sub> by at least 15 kcal mol<sup>-1</sup> and hence will play little role of importance. As the singlet and triplet spin states were already considerably higher in energy at the reactant stage, they remain well higher





**FIGURE 5** | QM/MM optimized geometries of the low-lying spin states of the iron(IV)-oxo species of WT P4H for snapshot Sn<sub>5</sub>ns and QM region A. From left-to-right, the data reported are for the singlet, triplet and quintet spin states. Bond lengths are in angstroms and relative energies in kcal mol<sup>-1</sup>.



**FIGURE 6** | Molecular valence orbitals of the iron(IV)-oxo species of P4H.

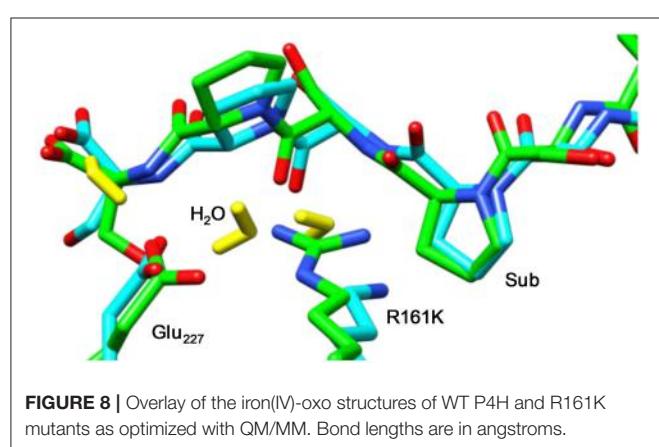
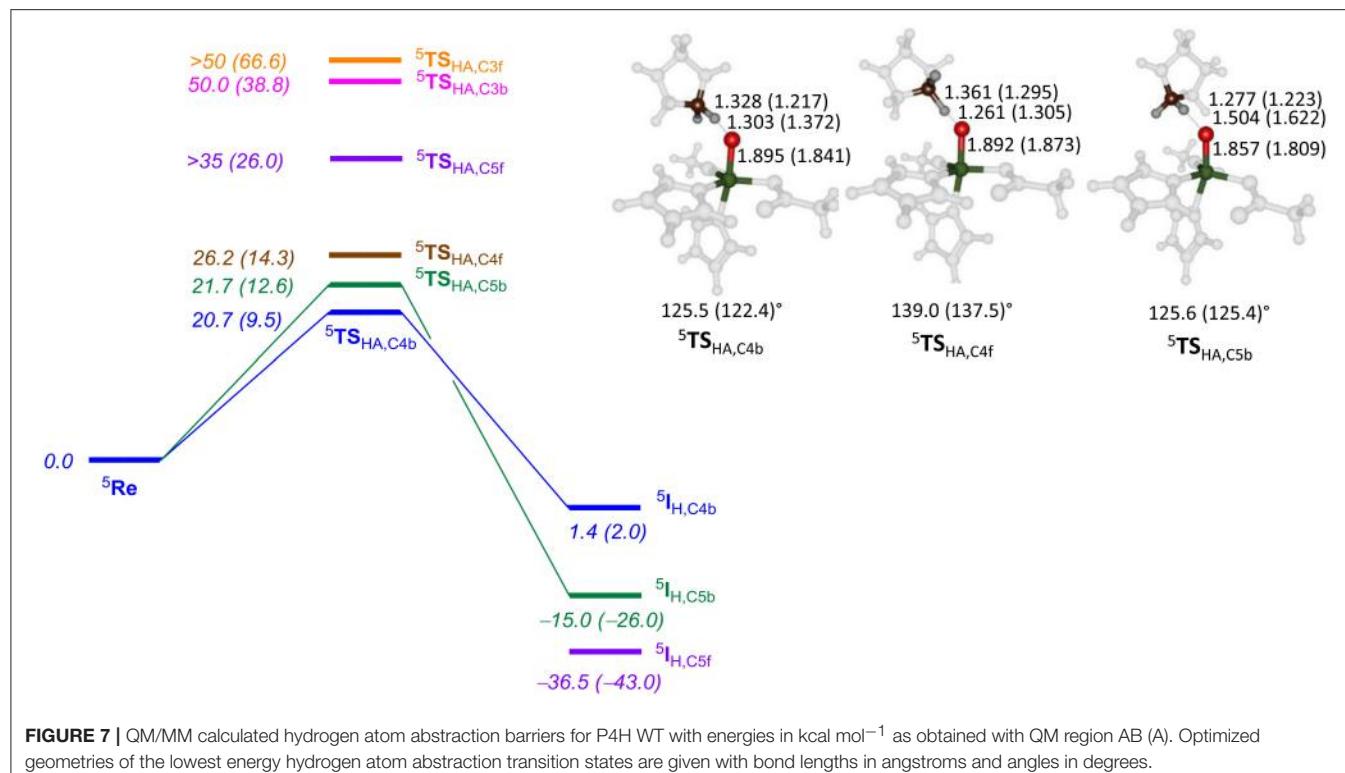
in the hydrogen atom abstraction transition states as well. The triplet and singlet spin barriers for abstracting the C<sup>4b</sup> hydrogen atoms are 34.0 and 40.2 kcal mol<sup>-1</sup> at UB3LYP/BSI in QM/MM. As such the reactivity takes place on a single spin state only, the quintet spin state, and other spin states play no role in the rate-determining pathway.

The optimized geometries of <sup>5</sup>TS<sub>HA,C4b</sub>, <sup>5</sup>TS<sub>HA,C4f</sub>, and <sup>5</sup>TS<sub>HA,C5b</sub> (right-hand-side of Figure 7) give insight into their energetic ordering and relative energies. Thus, in <sup>5</sup>TS<sub>HA,C4b</sub> the transferring hydrogen atom is almost midway in between donor and acceptor atom and the Fe-O-C<sup>4b</sup> angle is about 125°. In <sup>5</sup>TS<sub>HA,C4f</sub>, by contrast, the substrate is oriented along a much larger angle of 139.0°. The <sup>5</sup>TS<sub>HA,C5b</sub> structure, on the other hand, has the transferring hydrogen atom at a relatively large distance from the accepting oxygen atom and hence is destabilized considerably.

### P4H Mutations of Arg<sub>161</sub>

To find out the effect of substrate positioning and catalytic turnover of active site mutations, we investigated several P4H models, where amino acids were replaced. Thus, in our previous studies (Timmins et al., 2017) we implicated an important role of Arg<sub>161</sub> and Glu<sub>127</sub> through hydrogen bonding interactions. In this section we will look into the structure and catalytic activity of P4H mutants with Arg<sub>161</sub> replaced by either Asp, Gln, or Lys. These changes could be dramatic as the Arg<sub>161</sub>Asp (R161D) mutation will replace a positively charged residue with a negatively one. Similarly, the Arg<sub>161</sub>Gln (R161Q) mutation changes a cationic residue into a neutral one.

Figure 8 displays an overlay of the structures of the iron(IV)-oxo species for WT and R161K mutation after a full QM/MM geometry optimization, where the positively charged Arg residue is replaced by the positively charged Lys amino acid. As can be seen the mutation displaces the salt bridge between Arg<sub>161</sub> and



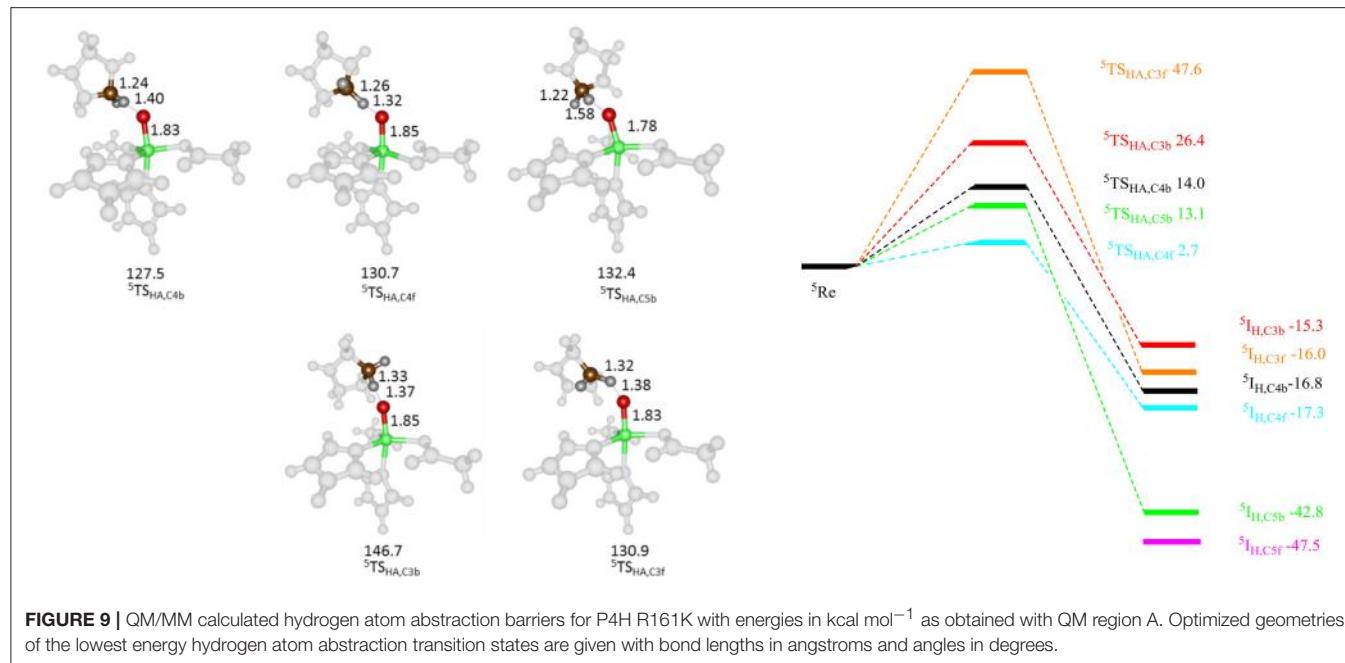
Glu<sub>127</sub>, which bends outward. The space provided by the Glu<sub>127</sub> migration is filled up with extra water molecules. However, the removal of the strong hydrogen bonding interactions of the Arg<sub>161</sub>-Glu<sub>127</sub> couple toward the substrate has a major effect on the stability of the substrate and its positioning. Thus, substrate is less tightly bound in the R161K mutant than in WT and hence its regio- and stereoselective substrate activation may be affected.

Subsequently, we studied the hydrogen atom abstraction mechanisms of the R161K mutant from the C<sup>3</sup>, C<sup>4</sup>, and C<sup>5</sup> positions of proline for the back and front protons. **Figure 9** displays relative energies and optimized geometries of selected hydrogen atom abstraction transition states for the R161K mutant. The hydrogen atom abstraction barrier from the C<sup>4b</sup>

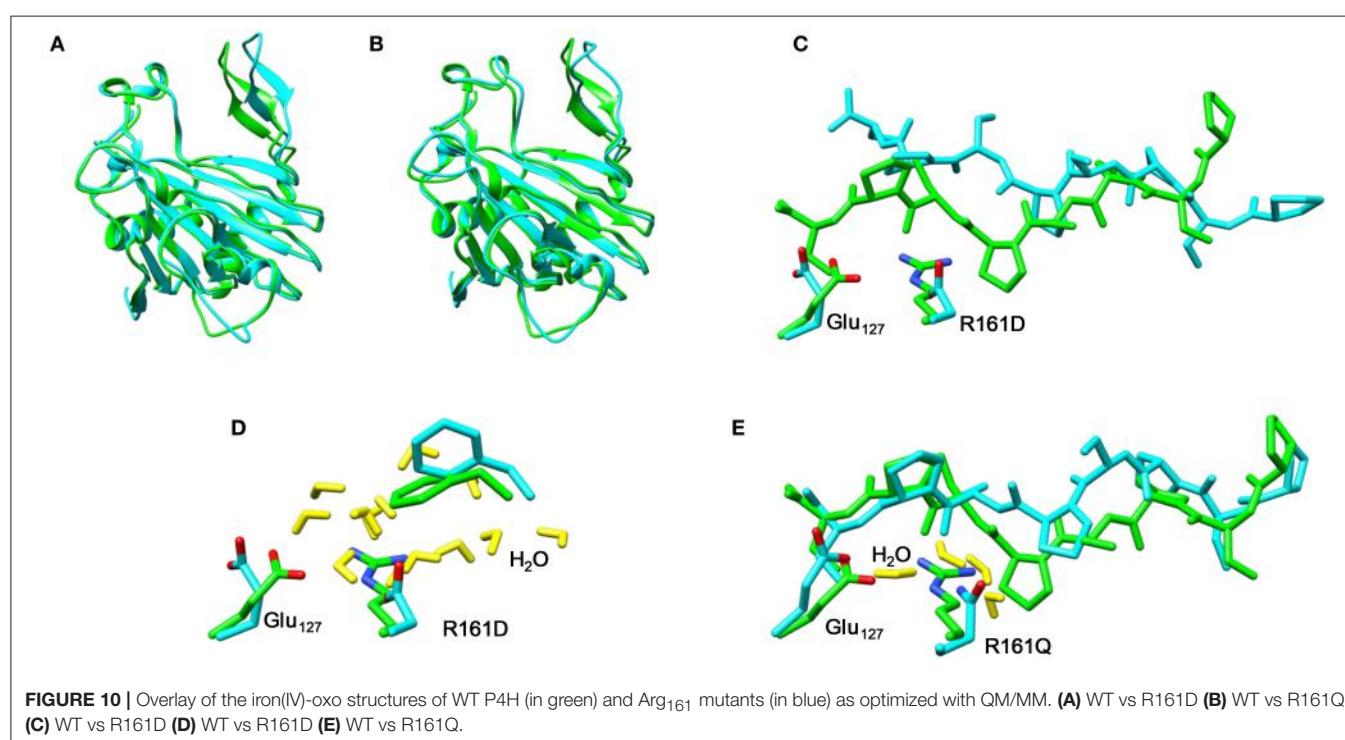
position (<sup>5</sup>TS<sub>R161K,C4b</sub>) has an energy of 14.0 kcal mol<sup>-1</sup> (UB3LYP/BSI), which is almost identical to the one observed for WT of 14.3 kcal mol<sup>-1</sup>. Indeed, the optimized geometries are very similar: C-H and O-H distances are found of 1.24 and 1.40 Å for R161K, whereas they are 1.22 and 1.37 Å, respectively, for WT (**Figure 7** above). However, a much lower transition state is found for activation of the C<sup>4f</sup> position of only 2.7 kcal mol<sup>-1</sup>. Therefore, the R161K mutation will not affect the catalytic performance of the enzyme: It should react faster than WT, but will give a reversal of stereochemistry and predominantly produce the S-4-hydroxyproline product instead.

Comparison of <sup>5</sup>TS<sub>C4b</sub> transition state structures in R161K to WT shows that the positions of the active site components change little, however, the Fe-O-C<sup>4</sup> angle and N<sub>His</sub>-Fe-O-C<sup>4</sup> dihedral angle are different, namely 137.5° and -73.4° for WT and 130.7° and -67.5° for R161K, respectively. Therefore, this mutation allows for the orientation of the substrate relative to the iron(IV)-oxo to change in such a way that the C<sup>4f</sup> position is now accessible to it. Hydrogen atom abstraction from the C<sup>3</sup> position is seen to be lowered as compared to WT but is significantly higher in energy than the barrier <sup>5</sup>TS<sub>R161K,C4f</sub>. The barrier for C<sup>5b</sup> hydrogen atom abstraction is similar to that for C<sup>4b</sup> but now slightly lower in energy.

Thereafter, we studied the R161Q and R161D mutants and **Figure 10** gives structures of the iron(IV)-oxo species as compared to WT. As can be seen both mutations have a dramatic effect on substrate binding and positioning as a result of changes in the hydrogen bonding network between Glu<sub>127</sub>, R161D and surrounding residues located in the  $\beta_{II}$ - $\beta_{III}$  and  $\beta_3$ - $\beta_4$  loops. We



**FIGURE 9 |** QM/MM calculated hydrogen atom abstraction barriers for P4H R161K with energies in  $\text{kcal mol}^{-1}$  as obtained with QM region A. Optimized geometries of the lowest energy hydrogen atom abstraction transition states are given with bond lengths in angstroms and angles in degrees.



**FIGURE 10 |** Overlay of the iron(IV)-oxo structures of WT P4H (in green) and Arg<sub>161</sub> mutants (in blue) as optimized with QM/MM. (A) WT vs R161D (B) WT vs R161Q (C) WT vs R161D (D) WT vs R161D (E) WT vs R161Q.

then attempted to abstract hydrogen atoms from proline by the R161Q and R161D mutants. **Table 1** gives data with calculated barrier heights for several hydrogen atoms of proline. None of these barrier heights, however, is low enough in energy to make them accessible at room temperature. Therefore, the R161Q and R161D mutants will be catalytically inactive. As a result, the Arg<sub>161</sub> residue has a critical function in P4H enzymes in positioning the substrate in the correct orientation. This is done

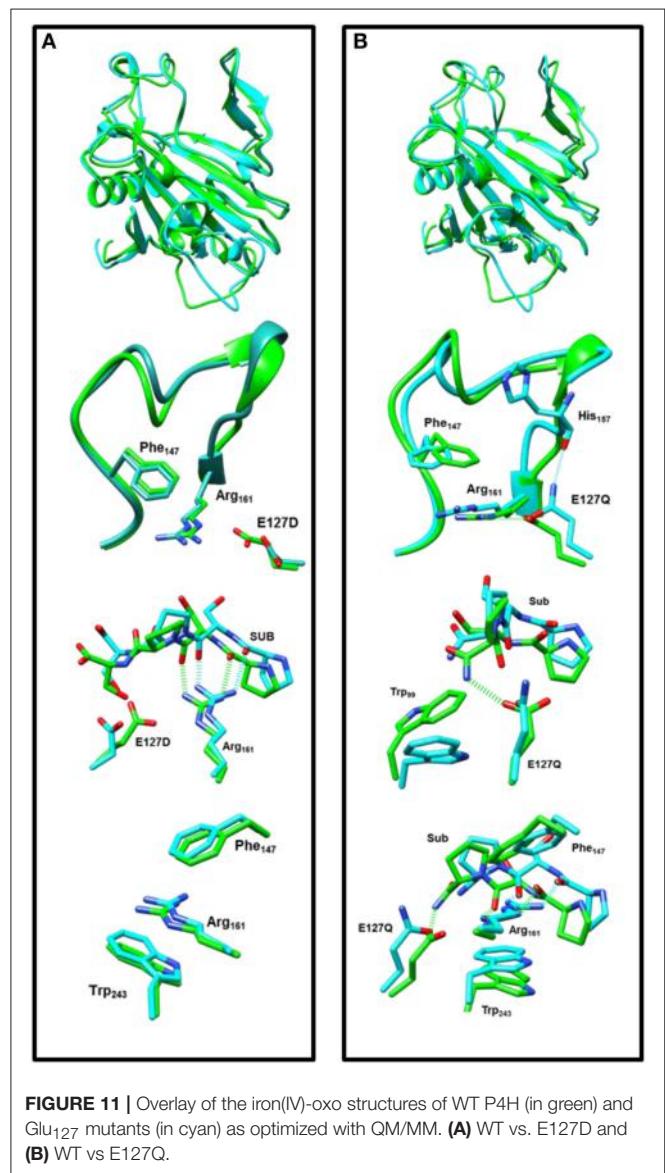
in conjunction with the Glu<sub>127</sub> residue that hydrogen bonds the protein loop of the substrate and makes sure it can approach the iron(IV)-oxo species.

### P4H Mutations of Glu<sub>127</sub>

In a final set of calculations we investigated P4H mutants where Glu<sub>127</sub> is replaced by either Asp or Gln. **Figure 11** displays the QM/MM optimized iron(IV)-oxo species of WT version E127D

**TABLE 1** | Calculated barrier heights (kcal mol<sup>-1</sup>) for several hydrogen atom transfers from the substrate proline residue to the iron(IV)-oxo oxidant in prolyl-4-hydroxylase.

	<sup>5</sup> TS <sub>HA,C4b</sub>	<sup>5</sup> TS <sub>HA,C4f</sub>	<sup>5</sup> TS <sub>HA,C5b</sub>	<sup>5</sup> TS <sub>HA,C5f</sub>	<sup>5</sup> TS <sub>HA,C3b</sub>	<sup>5</sup> TS <sub>HA,C3f</sub>
WT	9.8	14.8	12.7	27.6	38.4	66.6
R161D	60.4	60.8	45.4	68.4	85.3	73.8
R161Q	46.1	48.3	N/A	N/A	N/A	62.8

**FIGURE 11** | Overlay of the iron(IV)-oxo structures of WT P4H (in green) and Glu<sub>127</sub> mutants (in cyan) as optimized with QM/MM. **(A)** WT vs. E127D and **(B)** WT vs E127Q.

and E127Q mutants. In E127D the hydrogen bond between Asp<sub>127</sub> and Arg<sub>161</sub> is broken and as a result as Asp<sub>127</sub> swings out, to subsequently effecting the positions of surrounding residues and the substrate. Interestingly, the hydrogen bond between Arg<sub>161</sub> and the substrate is maintained; however, the position of this residue corresponds to the change in the substrate position. It suggests that the function of Glu<sub>127</sub> is to anchor Arg<sub>161</sub> in a fixed position, which is essential for proper substrate positioning.

**TABLE 2** | Calculated barrier heights (kcal mol<sup>-1</sup>) for several hydrogen atom transfers from the substrate proline residue to the iron(IV)-oxo oxidant in prolyl-4-hydroxylase.

	<sup>5</sup> TS <sub>HA,C4b</sub>	<sup>5</sup> TS <sub>HA,C4f</sub>	<sup>5</sup> TS <sub>HA,C5b</sub>	<sup>5</sup> TS <sub>HA,C5f</sub>	<sup>5</sup> TS <sub>HA,C3b</sub>	<sup>5</sup> TS <sub>HA,C3f</sub>
WT	9.8	14.8	12.7	27.6	38.4	66.6
E127D	11.6	20.8	12.5	43.1	30.0	44.1
E127Q	25.0	N/A	20.1	41.8	42.5	55.6

In E127Q, there are notable changes around the active site with, e.g., Tyr<sub>140</sub> rotating out of its WT position breaking its hydrogen bond to the iron(IV)-oxo which has previously been shown to be important for correct substrate positioning and release. Previous research (Koski et al., 2009) has suggested that any disturbance of this “conformational switch” would result in the inactivation of the enzyme as shown by its mutation to alanine in experiment (Koski et al., 2009). The position of Trp<sub>243</sub> is also altered affecting substrate positioning even more, and consequently the E127Q mutation is likely to lead to an inactive form of the enzyme. To test the hypothesis, we explored hydrogen atom abstraction from various positions of proline residue which are given in **Table 2**.

In E127D, no longer is hydrogen atom abstraction (HAT) from C<sup>4b</sup> the favored pathway, but now, HAT from C<sup>5b</sup> is within 1 kcal mol<sup>-1</sup> and hence the two pathways are competitive. Considering the energies of the intermediate radical species in the mutant, it becomes clear that the C<sup>5b</sup> structure (-31.8 kcal mol<sup>-1</sup>) product will be the major product over the C<sup>4b</sup> (-2.9 kcal mol<sup>-1</sup>), therefore the E127D mutation leads to a change in the regioselectivity of the reaction. This is similar to what has been seen in previous DFT calculations which showed that if given the choice HAT from the C<sup>5</sup> position will always be favored over the C<sup>4</sup> position as the BDE at the former position is weaker compared to the latter (Karamzadeh et al., 2010). Comparison of the <sup>5</sup>TS<sub>C5b</sub> transition state structures in WT and E127D reveal similar geometric parameters, however, the Fe-O-C<sup>5</sup> angle and N<sub>His</sub>-Fe-O-C<sup>5</sup> dihedral angle are different, namely 125.4° and -77.8° for WT and 132.1° and -88.2° for E127D, respectively. As such, the orientation of the substrate relative to the iron(IV)-oxo has changed, allowing the substrate in E127D to adopt a more favorable position for HAT from the C<sup>5b</sup> position as compared to WT. In E127Q, HAT from the C<sup>5b</sup> position is favored compared to that from C<sup>4b</sup>, as shown also for E127D. However, as the transition state barrier is 20.1 kcal mol<sup>-1</sup> for HAT from the C<sup>4b</sup> position, the mutant will result in much slower reactivity as compared to WT.

In conclusion, previous QM/MM studies on P4H have elucidated the reasoning behind its observed regioselectivity and stereoselectivity in the WT through various mutations to the protein. The results of that study highlighted the role of Glu<sub>127</sub> and Arg<sub>161</sub> in substrate positioning and suggested how mutating those residues could alter the regioselectivity and stereoselectivity with minimal influence on the enzymes stability and catalytic ability, an important consideration for future biotechnological applications. This study has revealed that mutations E127D and R161K are possible mutation candidates which will result in a change in the regioselectivity and stereoselectivity. Additionally,

the work highlights the importance in conserving the charge in the substrate binding residues of the enzyme around the substrate cavity to ensure interactions between the substrate and protein are maintained when an amino acid is mutated.

## CONCLUSION

Here we describe a detailed computational study into the activity of prolyl-4-hydroxylase enzymes and several Glu<sub>127</sub> and Arg<sub>161</sub> mutants. In particular, a comprehensive QM/MM study is presented, whereby we investigated hydrogen atom abstraction channels of each pair of hydrogen atoms bound to C<sup>3</sup>, C<sup>4</sup>, and C<sup>5</sup> of the proline residue of the substrate. Studies on WT predict the experimentally observed product distributions and give regio- and enantioselective R-4-hydroxyproline as a product. Analysis of the structure and electronic configurations show the regioselectivity to be guided by substrate positioning and

hydrogen bonding interactions. Mutations of Glu<sub>127</sub> and Arg<sub>161</sub> have major effects and lead to inactivity of the protein in several cases, in particular, when an anionic residue is replaced by a cationic one. Only in the case of the E127D mutant significant activity remains although competitive C<sup>4b</sup> and C<sup>5b</sup> hydroxylation is predicted.

## AUTHOR CONTRIBUTIONS

AT and SdV designed and developed the project. AT performed the calculations. AT and SdV wrote the paper.

## FUNDING

The work was funded through a studentship from the Biotechnology and Biological Sciences Research Council (BBSRC) under grant code BB/J014478/1.

## REFERENCES

- Abu-Omar, M. M., Loaiza, A., and Hontzeas, N. (2005). Reaction mechanisms of mononuclear non-heme iron oxygenases. *Chem. Rev.* 105, 2227–2252. doi: 10.1021/cr0406530
- Ahrlrichs, R., Bär, M., Häser, M., Horn, H., and Kölmel, C. (1989). Electronic structure calculations on workstation computers: the program system Turbomole. *Chem. Phys. Lett.* 162, 165–169. doi: 10.1016/0009-2614(89)85118-8
- Bakowies, D., and Thiel, W. (1996). Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.* 100, 10580–10594. doi: 10.1021/jp9536514
- Barman, P., Upadhyay, P., Faponle, A. S., Kumar, J., Nag, S. S., Kumar, D., et al. (2016). Deformylation reaction by a nonheme manganese(III)-peroxo complex via initial hydrogen atom abstraction. *Angew. Chem. Int. Ed.* 55, 11091–11095. doi: 10.1002/anie.201604412
- Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98, 5648–5652. doi: 10.1063/1.464913
- Blomberg, M. R. A., Borowski, T., Himo, F., Liao, R.-Z., and Siegbahn, P. E. M. (2014). Quantum chemical studies of mechanisms for metalloenzymes. *Chem. Rev.* 114, 3601–3658. doi: 10.1021/cr400388t
- Bollinger, J. M. Jr., Price, J. C., Hoffart, L. M., Barr, E. W., and Krebs, C. (2005). Mechanism of taurine:  $\alpha$ -ketoglutarate dioxygenase (TauD) from *Escherichia coli*. *Eur. J. Inorg. Chem.* 4245–4254. doi: 10.1002/ejic.200500476
- Borowski, T., Bassan, A., and Siegbahn, P. E. M. (2004). Mechanism of Dioxygen activation in 2-oxoglutarate-dependent enzymes: a hybrid DFT Study. *Chem. Eur. J.* 10, 1031–1041. doi: 10.1002/chem.200305306
- Brazzolotto, D., Cantú Reinhard, F. G., Smith-Jones, J., Retegan, M., Amidani, L., Faponle, A. S., et al. (2017). A high-valent non heme  $\mu$ -oxo Mn<sup>IV</sup> dimer generated from a thiolate-bound Mn<sup>II</sup> complex and O<sub>2</sub>. *Angew. Chem. Int. Ed.* 56, 8211–8215. doi: 10.1002/anie.201703215
- Brooks, B. R., Brucolieri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4, 187–217. doi: 10.1002/jcc.540040211
- Bruijnincx, P. C. A., van Koten, G., and Klein Gebbink, R. J. M. (2008). Mononuclear non-heme iron enzymes with the 2-His-1-carboxylate facial triad: recent developments in enzymology and modeling studies. *Chem. Soc. Rev.* 37, 2716–2744. doi: 10.1039/b707179p
- Bugg, T. D. H. (2001). Oxygenases: mechanisms and structural motifs for O(2) activation. *Curr. Opin. Chem. Biol.* 5, 550–555. doi: 10.1016/S1367-5931(00)00236-2
- Buongiorno, D., and Straganz, G. D. (2013). Structure and function of atypically coordinated enzymatic mononuclear non-heme-Fe(II) centers. *Coord. Chem. Rev.* 257, 541–563. doi: 10.1016/j.ccr.2012.04.028
- Cantú Reinhard, F. G., and de Visser, S. P. (2017). Oxygen atom transfer using an iron(IV)-oxo embedded in a tetracyclic N-heterocyclic carbene system: how does the reactivity compare to Cytochrome P450 Compound? *Chem. Eur. J.* 23, 2935–2944. doi: 10.1002/chem.201605505
- Cantú Reinhard, F. G., Faponle, A. S., and de Visser, S. P. (2016a). Substrate sulfoxidation by an iron(IV)-oxo complex: benchmarking computationally calculated barrier heights to experiment. *J. Phys. Chem. A* 120, 9805–9814. doi: 10.1021/acs.jpca.6b09765
- Cantú Reinhard, F. G., Sainna, M. A., Upadhyay, P., Balan, G. A., Kumar, D., Fornarini, S., et al. (2016b). A systematic account on aromatic hydroxylation by a cytochrome P450 model Compound I: a low-pressure mass spectrometry and computational study. *Chem. Eur. J.* 22, 18608–18619. doi: 10.1002/chem.201604361
- Choroba, O. W., Williams, D. H., and Spencer, J. B. (2000). Biosynthesis of the vancomycin group of antibiotics: involvement of an unusual dioxygenase in the pathway to (s)-4-hydroxyphenylglycine. *J. Am. Chem. Soc.* 122, 5389–5390. doi: 10.1021/ja000076v
- Costas, M., Mehn, M. P., Jensen, M. P., and Que, L. Jr. (2004). Dioxygen activation at mononuclear nonheme iron active sites: enzymes, models, and intermediates. *Chem. Rev.* 104, 939–986. doi: 10.1021/cr020628n
- Denisov, I. G., Makris, T. M., Sligar, S. G., and Schlichting, I. (2005). Structure and chemistry of cytochrome P450. *Chem. Rev.* 105, 2253–2277. doi: 10.1021/cr0307143
- de Visser, S. P. (2005). What affects the quartet-doublet energy splitting in peroxidase enzymes? *J. Phys. Chem. A* 109, 11050–11057. doi: 10.1021/jp053873u
- de Visser, S. P. (2006a). Differences in and comparison of the catalytic properties of heme and non-heme enzymes with a central oxo-iron group. *Angew. Chem. Int. Ed.* 45, 1790–1793. doi: 10.1002/anie.200503841
- de Visser, S. P. (2006b). Propene activation by the oxo-iron active species of taurine/ $\alpha$ -ketoglutarate dioxygenase (TauD) enzyme. How does the catalysis compare to heme-enzymes? *J. Am. Chem. Soc.* 128, 9813–9824. doi: 10.1021/ja061581g
- de Visser, S. P. (2006c). Substitution of hydrogen by deuterium changes the regioselectivity of ethylbenzene hydroxylation by an oxo-iron-porphyrin catalyst. *Chem. Eur. J.* 12, 8168–8177. doi: 10.1002/chem.200600376
- de Visser, S. P. (2006d). What factors influence the ratio of C–H hydroxylation versus C=C epoxidation by a nonheme cytochrome P450 biomimetic? *J. Am. Chem. Soc.* 128, 15809–15818. doi: 10.1021/ja065365j

- de Visser, S. P. (2007). Can the peroxosuccinate complex in the catalytic cycle of taurine/α-ketoglutarate dioxygenase (TauD) act as an alternative oxidant? *Chem. Commun.* 171–173. doi: 10.1039/B611273K
- de Visser, S. P. (2009). Elucidating enzyme mechanism and intrinsic chemical properties of short-lived intermediates in the catalytic cycles of cysteine dioxygenase and taurine/α-ketoglutarate dioxygenase. *Coord. Chem. Rev.* 293, 754–768. doi: 10.1016/j.ccr.2008.05.001
- de Visser, S. P. (2012). Predictive studies of oxygen atom transfer reactions by Compound I of cytochrome P450: aliphatic and aromatic hydroxylation, epoxidation and sulfoxidation. *Adv. Inorg. Chem.* 64, 1–31. doi: 10.1016/B978-0-12-396462-5.00001-5
- de Visser, S. P., and Kumar, D. (2011). *Iron-Containing Enzymes: Versatile Catalysts of Hydroxylation Reaction in Nature*. Cambridge: RSC Publishing.
- de Visser, S. P., Kumar, D., Cohen, S., Shacham, R., and Shaik, S. (2004). A predictive pattern of computed barriers for C–H hydroxylation by Compound I of cytochrome P450. *J. Am. Chem. Soc.* 126, 8362–8363. doi: 10.1021/ja048528h
- de Visser, S. P., Ogliaro, F., Harris, N., and Shaik, S. (2001). Multi-state epoxidation of ethene by cytochrome P450: a quantum chemical study. *J. Am. Chem. Soc.* 123, 3037–3047. doi: 10.1021/ja003544+
- de Visser, S. P., Ogliaro, F., Sharma, P. K., and Shaik, S. (2002a). Hydrogen bonding modulates the selectivity of enzymatic oxidation by P450: a chameleon oxidant behavior of Compound I. *Angew. Chem. Int. Ed.* 41, 1947–1951.
- de Visser, S. P., Ogliaro, F., Sharma, P. K., and Shaik, S. (2002b). What factors affect the regioselectivity of oxidation by cytochrome P450? A DFT study of allylic hydroxylation and double bond epoxidation in a model reaction. *J. Am. Chem. Soc.* 124, 11809–11826. doi: 10.1021/ja026872d
- de Visser, S. P., Rohde, J.-U., Lee, Y.-M., Cho, J., and Nam, W. (2012). Intrinsic properties and reactivities of mononuclear nonheme iron–oxygen complexes bearing the tetramethylcyclam ligand. *Coord. Chem. Rev.* 257, 381–393. doi: 10.1016/j.ccr.2012.06.002
- de Visser, S. P., and Shaik, S. (2003). A proton-shuttle mechanism mediated by the porphyrin in benzene hydroxylation by cytochrome P450 enzymes. *J. Am. Chem. Soc.* 125, 7413–7424. doi: 10.1021/ja034142f
- de Visser, S. P., Shaik, S., Sharma, P. K., Kumar, D., and Thiel, W. (2003). Active species of horseradish peroxidase (HRP) and cytochrome P450: two electronic chameleons. *J. Am. Chem. Soc.* 125, 15779–15788. doi: 10.1021/ja0380906
- de Visser, S. P., Tahsin, L., and Nam, W. (2009). How does the axial ligand of cytochrome P450 biomimetics influence the regioselectivity of aliphatic versus aromatic hydroxylation? *Chem. Eur. J.* 15, 5577–5587. doi: 10.1002/chem.200802234
- Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., et al. (2007). PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucl. Acids Res.* 35, 522–525. doi: 10.1093/nar/gkm276
- Faponle, A. S., Quesne, M. G., and de Visser, S. P. (2016). Origin of the regioselective fatty acid hydroxylation versus decarboxylation by a cytochrome P450 peroxxygenase: what drives the reaction to biofuel production? *Chem. Eur. J.* 22, 5478–5483. doi: 10.1002/chem.201600739
- Faponle, A. S., Seebeck, F. P., and de Visser, S. P. (2017). Sulfoxide synthase versus cysteine dioxygenase reactivity in a nonheme iron enzyme. *J. Am. Chem. Soc.* 139, 9259–9270. doi: 10.1021/jacs.7b04251
- Fellner, M., Siakkou, E., Faponle, A. S., Tchesnokov, E. P., de Visser, S. P., Wilbanks, S. M., et al. (2016). Influence of cysteine 164 on active site structure in rat cysteine dioxygenase. *J. Biol. Inorg. Chem.* 21, 501–510. doi: 10.1007/s00775-016-1360-0
- Godfrey, E., Porro, C. S., and de Visser, S. P. (2008). Comparative quantum mechanics / molecular mechanics (QM/MM) and density functional theory calculations on the oxo-iron species of taurine/α-ketoglutarate dioxygenase. *J. Phys. Chem. A.* 112, 2464–2468. doi: 10.1021/jp710999v
- Guo, Z., and Sevrioukova, I. (2017). Crystal structure of human Cyp3A4 bound to metformin. doi: 10.2210/pdb5g5j/pdb
- Hernández-Ortega, A., Quesne, M. G., Bui, S., Heuts, D. P. H. M., Steiner, R. A., Heyes, D. J., et al. (2014). Origin of the proton-transfer step in the cofactor-free 1-H-3-hydroxy-4-oxoquinaldine 2,4-dioxygenase: effect of the basicity of an active site His residue. *J. Biol. Chem.* 289, 8620–8632. doi: 10.1074/jbc.M113.543033
- Hernández-Ortega, A., Quesne, M. G., Bui, S., Heyes, D. J., Steiner, R. A., Scrutton, N. S., et al. (2015). Catalytic mechanism of cofactor-free dioxygenases and how they circumvent spin-forbidden oxygenation of their substrates. *J. Am. Chem. Soc.* 137, 7474–7487. doi: 10.1021/jacs.5b03836
- Higgins, L. J., Yan, F., Liu, P., Liu, H.-W., and Drennan, C. L. (2005). Structural insight into antibiotic fosfomycin biosynthesis by a mononuclear iron enzyme. *Nature* 437, 838–844. doi: 10.1038/nature03924
- Hoffart, L. M., Barr, E. W., Guyer, R. B., Bollinger Jr, J. M., and Krebs, C. (2006). Direct spectroscopic detection of a C–H-cleaving high-spin Fe(IV) complex in a prolyl-4-hydroxylase. *Proc. Natl. Acad. Sci. U.S.A.* 103, 14738–14743. doi: 10.1073/pnas.0604005103
- Ji, L., Faponle, A. S., Quesne, M. G., Sainna, M. A., Zhang, J., Franke, A., et al. (2015). Drug metabolism by cytochrome P450 enzymes: what distinguishes the pathways leading to substrate hydroxylation over desaturation? *Chem. Eur. J.* 21, 9083–9092. doi: 10.1002/chem.201500329
- Kadish, K. M., Smith, K. M., and Guillard, R. (2010). *Handbook of Porphyrin Science*. Hackensack, NJ: World Scientific Publishing Co.
- Karamzadeh, B., Kumar, D., Sastry, G. N., and de Visser, S. P. (2010). Steric factors override thermodynamic driving force in regioselectivity of proline hydroxylation by prolyl-4-hydroxylase enzymes. *J. Phys. Chem. A* 114, 13234–13243. doi: 10.1021/jp1089855
- Koski, M. K., Hieto, R., Hirsila, M., Ronka, A., Myllyharju, J., and Wierenga, R. K. (2009). The crystal structure of an algal prolyl 4-hydroxylase complexed with a proline-rich peptide reveals a novel buried tripeptide binding motif. *J. Biol. Chem.* 284, 25290–25301. doi: 10.1074/jbc.M109.014050
- Kovaleva, E. G., and Lipscomb, J. D. (2008). Versatility of biological non-heme Fe(II) centers in oxygen activation reactions. *Nat. Chem. Biol.* 4, 186–193. doi: 10.1038/nchembio.71
- Kryatov, S. V., Rybak-Akimova, E. V., and Schindler, S. (2005). Kinetics and mechanisms of formation and reactivity of non-heme iron oxygen intermediates. *Chem. Rev.* 105, 2175–2226. doi: 10.1021/cr030709z
- Kumar, D., de Visser, S. P., Sharma, P. K., Cohen, S., and Shaik, S. (2004). Radical clock substrates, their C–H hydroxylation mechanism by cytochrome P450 and other reactivity patterns: what does theory reveal about the clocks' behavior? *J. Am. Chem. Soc.* 126, 1907–1920. doi: 10.1021/ja039439s
- Kumar, D., de Visser, S. P., Sharma, P. K., Hirao, H., and Shaik, S. (2005b). Sulfoxidation mechanisms catalyzed by cytochrome P450 and horseradish peroxidase models: spin selection induced by the ligand. *Biochemistry* 44, 8148–8158. doi: 10.1021/bi050348c
- Kumar, D., Hirao, H., de Visser, S. P., Zheng, J., Wang, D., Thiel, W., et al. (2005a). New features in the catalytic cycle of cytochrome P450 during the formation of Compound I from Compound 0. *J. Phys. Chem. B* 109, 19946–19951. doi: 10.1021/jp054754h
- Kumar, D., Sastry, G. N., and de Visser, S. P. (2011a). Effect of the axial ligand on substrate sulfoxidation mediated by iron(IV)-oxo porphyrin cation radical oxidants. *Chem. Eur. J.* 17, 6196–6205. doi: 10.1002/chem.201003187
- Kumar, D., Thiel, W., and de Visser, S. P. (2011b). Theoretical study on the mechanism of the oxygen activation process in cysteine dioxygenase enzymes. *J. Am. Chem. Soc.* 133, 3869–3882. doi: 10.1021/ja107514f
- Kumar, S., Faponle, A. S., Barman, P., Vardhaman, A. K., Sastri, C. V., Kumar, D., et al. (2014). Long-range electron transfer triggers mechanistic differences between iron(IV)-oxo and iron(IV)-imido oxidants. *J. Am. Chem. Soc.* 136, 17102–17115. doi: 10.1021/ja508403w
- Latifi, R., Bagherzadeh, M., and de Visser, S. P. (2009). Origin of the correlation of the rate constant of substrate hydroxylation by nonheme iron(IV)-oxo complexes with the bond-dissociation energy of the C–H bond of the substrate. *Chem. Eur. J.* 15, 6651–6662. doi: 10.1002/chem.200900211
- Latifi, R., Sainna, M. A., Rybak-Akimova, E. V., and de Visser, S. P. (2013). Does hydrogen bonding-donation to manganese(IV)-oxo and iron(IV)-oxo oxidants affect the oxygen atom transfer ability? A computational study. *Chem. Eur. J.* 19, 4058–4068. doi: 10.1002/chem.201202811
- Lee, C., Yang, W., and Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* 37, 785–789. doi: 10.1103/PhysRevB.37.785
- Li, X.-X., Postils, V., Sun, W., Faponle, A. S., Solà, M., Wang, Y., et al. (2017). Reactivity patterns of (protonated) Compound II and Compound I of Cytochrome P450: which is the better oxidant? *Chem. Eur. J.* 23, 6406–6418. doi: 10.1002/chem.201700363

- McDonough, M. A., Li, V., Flashman, E., Chowdhury, R., Mohr, C., Lienard, B. M., et al. (2006). Cellular oxygen sensing: crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2). *Proc. Natl. Acad. Sci. U.S.A.* 103, 9814–9819. doi: 10.1073/pnas.0601283103
- Meunier, B., de Visser, S. P., and Shaik, S. (2004). Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes. *Chem. Rev.* 104, 3947–3980. doi: 10.1021/cr020443g
- Myllyharju, J., and Kivirikko, K. I. (1997). Characterization of the iron- and 2-oxoglutarate-binding sites of human prolyl 4-hydroxylase. *EMBO J.* 16, 1173–1180. doi: 10.1093/emboj/16.6.1173
- O'Brien, J. R., Schuller, D. J., Yang, V. S., Dillard, B. D., and Lanzilotta, W. N. (2003). Substrate-induced conformational changes in *Escherichia coli* taurine/alpha-ketoglutarate dioxygenase and insight into the oligomeric structure. *Biochemistry* 42, 5547–5554. doi: 10.1021/bi0341096
- O'Brien, P. J. (2006). Catalytic promiscuity and the divergent evolution of DNA repair enzymes. *Chem. Rev.* 106, 720–752. doi: 10.1021/cr040481v
- Ogliaro, F., Cohen, S., de Visser, S. P., and Shaik, S. (2000a). Medium polarization and hydrogen bonding effects on Compound I of cytochrome P450: what kind of a radical is it really? *J. Am. Chem. Soc.* 122, 12892–12893. doi: 10.1021/ja005619f
- Ogliaro, F., Harris, N., Cohen, S., Filatov, M., de Visser, S. P., and Shaik, S. (2000b). A model “rebound” mechanism of hydroxylation by cytochrome P450: stepwise and effectively concerted pathways, and their reactivity patterns. *J. Am. Chem. Soc.* 122, 8977–8989. doi: 10.1021/ja991878x
- Ortiz de Montellano, P. R. (2004). *Cytochrome P450: Structure, Mechanism and Biochemistry*, 3rd Edn., New York, NY: Kluwer Academic/Plenum Publishers.
- Ortiz de Montellano, P. R. (2010). Hydrocarbon hydroxylation by cytochrome P450 enzymes. *Chem. Rev.* 110, 932–948. doi: 10.1021/cr9002193
- Porro, C. S., Sutcliffe, M. J., and de Visser, S. P. (2009). Quantum mechanics/molecular mechanics studies on the sulfoxidation of dimethyl sulfide by Compound I and Compound 0 of Cytochrome P450: which is the better oxidant? *J. Phys. Chem. A* 113, 11635–11642. doi: 10.1021/jp9023926
- Pratter, S. M., Konstantinovics, C., DiGiuro, C. L. M., Leitner, E., Kumar, D., de Visser, S. P., et al. (2013). Inversion of enantio-selectivity of a mononuclear non-heme iron(II)-dependent hydroxylase by tuning the interplay of metal-center geometry and protein structure. *Angew. Chem. Int. Ed.* 52, 9677–9681. doi: 10.1002/anie.201304633
- Que, L. Jr. (2000). One motif — many different reactions. *Nat. Struct. Biol.* 7, 182–184. doi: 10.1038/73270
- Quesne, M. G., Borowski, T., and de Visser, S. P. (2016a). Quantum mechanics/molecular mechanics modelling of enzymatic processes: caveats and breakthroughs. *Chem. Eur. J.* 22, 2562–2581. doi: 10.1002/chem.201503802
- Quesne, M. G., Latifi, R., Gonzalez-Ovalle, L. E., Kumar, D., and de Visser, S. P. (2014). Quantum mechanics/molecular mechanics study on the oxygen binding and substrate hydroxylation step in AlkB repair enzymes. *Chem. Eur. J.* 20, 435–446. doi: 10.1002/chem.201303282
- Quesne, M. G., Senthilnathan, D., Singh, D., Kumar, D., Maldivi, P., Sorokin, A. B., et al. (2016b). Origin of the enhanced reactivity of  $\mu$ -nitrido-bridged diiron(IV)-oxo porphyrinoid complexes over cytochrome P450 Compound I. *ACS Catal.* 6, 2230–2243. doi: 10.1021/acscatal.5b02720
- Rittle, J., and Green, M. T. (2010). Cytochrome P450 compound I: capture, characterization, and C-H bond activation kinetics. *Science* 330, 933–937. doi: 10.1126/science.1193478
- Ryle, M. J., and Hausinger, R. P. (2002). Non-heme iron oxygenases. *Curr. Opin. Chem. Biol.* 6, 193–201. doi: 10.1016/S1367-5931(02)00302-2
- Sainna, M. A., Kumar, S., Kumar, D., Formarini, S., Crestoni, M. E., and de Visser, S. P. (2015). A comprehensive test set of epoxidation rate constants by iron(IV)-oxo porphyrin complexes. *Chem. Sci.* 6, 1516–1529. doi: 10.1039/C4SC02717E
- Schafer, A., Horn, H., and Ahlrichs, R. (1992). Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.* 97, 2571–2577. doi: 10.1063/1.463096
- Schöneboom, J. C., Lin, H., Reuter, N., Thiel, W., Cohen, S., Ogliaro, F., et al. (2002). The elusive oxidant species of cytochrome P450 enzymes: characterization by combined quantum mechanical/molecular mechanical (QM/MM) calculations. *J. Am. Chem. Soc.* 124, 8142–8151. doi: 10.1021/ja026279w
- Shaik, S., Kumar, D., de Visser, S. P., Altun, A., and Thiel, W. (2005). Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. *Chem. Rev.* 105, 2279–2328. doi: 10.1021/cr030722j
- Sharma, P. K., de Visser, S. P., and Shaik, S. (2003). Can a single oxidant with two spin states masquerade as two different oxidants? A study of the sulfoxidation mechanism by cytochrome P450. *J. Am. Chem. Soc.* 125, 8698–8699. doi: 10.1021/ja035135u
- Sherwood, P., de Vries, A. H., Guest, M. F., Schreckenbach, G., Catlow, C. R. A., French, S. A., et al. (2003). QUASI: a general purpose implementation of the QM/MM approach and its application to problems in catalysis. *J. Mol. Struct.* 632, 1–28. doi: 10.1016/S0166-1280(03)00285-9
- Siitonen, V., Selvaraj, B., Niiranen, L., Lindqvist, Y., Schneider, G., and Metsä-Ketelä, M. (2016). Divergent non-heme iron enzymes in the nogalamycin biosynthetic pathway. *Proc. Natl. Acad. Sci. U.S.A.* 113, 5251–5256. doi: 10.1073/pnas.1525034113
- Smith, W., and Forester, T. R. (1996). DL\_POLY\_2.0: a general-purpose parallel molecular dynamics simulation package. *J. Mol. Graphics* 14, 136–141. doi: 10.1016/S0263-7855(96)00043-4
- Solomon, E. I., Brunold, T. C., Davis, M. I., Kemsley, J. N., Lee, S.-K., Lehnert, N., et al. (2000). Geometric and electronic structure/function correlations in non-heme iron enzymes. *Chem. Rev.* 100, 235–349. doi: 10.1021/cr9900275
- Solomon, E. I., Light, K. M., Liu, L. V., Srnec, M., and Wong, S. D. (2013). Geometric and electronic structure contributions to function in non-heme iron enzymes. *Acc. Chem. Res.* 46, 2725–2739. doi: 10.1021/ar400149m
- Stipanuk, M. H. (2004). Sulfur amino acid metabolism: pathways for production and removal of homocysteine and cysteine. *Annu. Rev. Nutr.* 24, 539–577. doi: 10.1146/annurev.nutr.24.012003.132418
- Straganz, G. D., and Nidetzky, B. (2006). Variations of the 2-His-1-carboxylate Theme in Mononuclear Non-Heme FeII Oxygenases. *Chembiochem* 7, 1536–1548. doi: 10.1002/cbic.200600152
- Tchesnokov, E. P., Faponle, A. S., Davies, C. G., Quesne, M. G., Turner, R., Fellner, M., et al. (2016). An iron-oxygen intermediate formed during the catalytic cycle of cysteine dioxygenase. *Chem. Commun.* 52, 8814–8817. doi: 10.1039/C6CC03904A
- Timmins, A., Saint-André, M., and de Visser, S. P. (2017). Understanding how prolyl-4-hydroxylase structure steers a ferryl oxidant toward scission of a strong C-H bond. *J. Am. Chem. Soc.* 139, 9855–9866. doi: 10.1021/jacs.7b02839
- Vardhaman, A. K., Barman, P., Kumar, S., Sastri, C. V., Kumar, D., and de Visser, S. P. (2013). Comparison of the reactivity of nonheme iron(IV)-oxo versus iron(IV)-imido complexes: which is the better oxidant? *Angew. Chem. Int. Ed.* 52, 12288–12292. doi: 10.1002/anie.201305370
- Vardhaman, A. K., Sastri, C. V., Kumar, D., and de Visser, S. P. (2011). Nonheme ferric hydroperoxo intermediates are efficient oxidants of bromide oxidation. *Chem. Commun.* 47, 11044–11046. doi: 10.1039/c1cc13775a
- Wachters, A. J. H. (1970). Gaussian basis set for molecular wavefunctions containing third-row atoms. *J. Chem. Phys.* 52, 1033–1036. doi: 10.1063/1.1673095
- Yang, T., Quesne, M. G., Neu, H. M., Cantí Reinhard, F. G., Goldberg, D. P., and de Visser, S. P. (2016). Singlet versus triplet reactivity in an Mn(V)-Oxo species: testing theoretical predictions against experimental evidence. *J. Am. Chem. Soc.* 138, 12375–12386. doi: 10.1021/jacs.6b05027
- Yi, C., Yang, C. G., and He, C. (2009). A non-heme iron-mediated chemical demethylation in DNA and RNA. *Acc. Chem. Res.* 42, 519–529. doi: 10.1021/ar800178j
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- The reviewer, TM, and handling Editor declared their shared affiliation.
- Copyright © 2017 Timmins and de Visser. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Chloride Ion Transport by the *E. coli* CLC $\text{Cl}^-/\text{H}^+$ Antiporter: A Combined Quantum-Mechanical and Molecular-Mechanical Study

Chun-Hung Wang, Adam W. Duster, Baris O. Aydintug, MacKenzie G. Zarecki and Hai Lin\*

Department of Chemistry, University of Colorado Denver, Denver, CO, United States

## OPEN ACCESS

### Edited by:

Thomas S. Hofer,  
University of Innsbruck, Austria

### Reviewed by:

Xavier Assfeld,  
Université de Lorraine, France  
Gerardo Andres Cisneros,  
University of North Texas,  
United States

### \*Correspondence:

Hai Lin  
hai.lin@ucdenver.edu

### Specialty section:

This article was submitted to  
Theoretical and Computational  
Chemistry,  
a section of the journal  
*Frontiers in Chemistry*

Received: 23 January 2018

Accepted: 26 February 2018

Published: 13 March 2018

### Citation:

Wang C-H, Duster AW, Aydintug BO, Zarecki MG and Lin H (2018) Chloride Ion Transport by the *E. coli* CLC  $\text{Cl}^-/\text{H}^+$  Antiporter: A Combined Quantum-Mechanical and Molecular-Mechanical Study. *Front. Chem.* 6:62.  
doi: 10.3389/fchem.2018.00062

We performed steered molecular dynamics (SMD) and umbrella sampling simulations of  $\text{Cl}^-$  ion migration through the transmembrane domain of a prototypical *E. coli* CLC  $\text{Cl}^-/\text{H}^+$  antiporter by employing combined quantum-mechanical (QM) and molecular-mechanical (MM) calculations. The SMD simulations revealed interesting conformational changes of the protein. While no large-amplitude motions of the protein were observed during pore opening, the side chain rotation of the protonated external gating residue Glu148 was found to be critical for full access of the channel entrance by  $\text{Cl}^-$ . Moving the anion into the external binding site ( $S_{\text{ext}}$ ) induced small-amplitude shifting of the protein backbone at the N-terminal end of helix F. As  $\text{Cl}^-$  traveled through the pore, rigid-body swinging motions of helix R separated it from helix D. Helix R returned to its original position once  $\text{Cl}^-$  exited the channel. Population analysis based on polarized wavefunction from QM/MM calculations discovered significant (up to 20%) charge loss for  $\text{Cl}^-$  along the ion translocation pathway inside the pore. The delocalized charge was redistributed onto the pore residues, especially the functional groups containing  $\pi$  bonds (e.g., the Tyr445 side chain), while the charges of the H atoms coordinating  $\text{Cl}^-$  changed almost negligibly. Potentials of mean force computed from umbrella sampling at the QM/MM and MM levels both displayed barriers at the same locations near the pore entrance and exit. However, the QM/MM PMF showed higher barriers ( $\sim 10$  kcal/mol) than the MM PMF ( $\sim 2$  kcal/mol). Binding energy calculations indicated that the interactions between  $\text{Cl}^-$  and certain pore residues were overestimated by the semi-empirical PM3 Hamiltonian and underestimated by the CHARMM36 force fields, both of which were employed in the umbrella sampling simulations. In particular, CHARMM36 underestimated binding interactions for the functional groups containing  $\pi$  bonds, missing the stabilizations of the  $\text{Cl}^-$  ion due to electron delocalization. The results suggested that it is important to explore these quantum effects for accurate descriptions of the  $\text{Cl}^-$  transport.

**Keywords:** QM/MM, CLC, chloride transport, electron delocalization, conformational change, umbrella sampling, steered molecular dynamics, potential of mean force

## INTRODUCTION

CLC transport proteins are a ubiquitous group of  $\text{Cl}^-$  ion channels and  $\text{Cl}^-/\text{H}^+$  antiporters that can be found in eukaryotes and bacteria. They are associated with many critical physiological and cellular processes such as aiding extreme acid-resistance response, cell-volume regulation, and muscle contraction (Maduke et al., 2000; Dutzler, 2006; Accardi and Picollo, 2010; Jentsch, 2015). Mutations in CLC proteins cause inherited diseases in humans, such as myotonia congenita, Dent's disease, Bartter's syndrome, osteopetrosis, and idiopathic epilepsy (Jentsch, 2008). CLC transport proteins assemble and function as homodimers, of which each monomer subunit contains an independent ion translocation pore (Chen, 2005). The architecture of the transmembrane domain is highly conserved across the CLC family (Estévez et al., 2003; Lin and Chen, 2003; Chen, 2005; Engh and Maduke, 2005). While CLC channels translocate  $\text{Cl}^-$  ions passively, CLC antiporters mediate the *coupled* influx of  $\text{Cl}^-$  and the efflux of  $\text{H}^+$  with a  $2\text{Cl}^-/1\text{H}^+$  ratio (Accardi and Miller, 2004; Picollo and Pusch, 2005; Scheel et al., 2005; Matulef and Maduke, 2007; Miller and Nguitragool, 2009; Feng et al., 2012; Accardi, 2015).

A prototype CLC antiporter is EcCLC from *Escherichia coli*, which has been subjected to extensive structural and functional studies (Dutzler et al., 2002, 2003; Accardi and Miller, 2004; Accardi et al., 2005, 2006; Nguitragool and Miller, 2006; Walden et al., 2007; Jayaram et al., 2008; Elvington et al., 2009; Lim and Miller, 2009; Miller and Nguitragool, 2009; Picollo et al., 2009, 2012; Robertson et al., 2010; Lim et al., 2012; Vien et al., 2017). Each subunit of EcCLC contains 18  $\alpha$ -helices, which form the ion-permeation path. In each subunit, there are three binding locations: the intracellular binding site  $S_{\text{int}}$ , the central binding site  $S_{\text{cen}}$ , and the extracellular binding site  $S_{\text{ext}}$ . In  $S_{\text{int}}$ , the  $\text{Cl}^-$  ion is coordinated by the backbone amine groups from Gly106 and Ser107 and may be partially hydrated (Dutzler et al., 2003; Gouaux and MacKinnon, 2005). In  $S_{\text{cen}}$ , the  $\text{Cl}^-$  ion is coordinated by the side chain hydroxyl groups of Ser107 in helix D and of Tyr445 in helix R as well as the backbone amine groups from Ile356 and Phe357 of helix N (Dutzler et al., 2002, 2003). In the wild-type protein crystal structure,  $S_{\text{ext}}$  is occupied by the side chain of a highly conserved residue Glu148 (Dutzler et al., 2003). However, mutation of Glu148 to a charge-neutral residue results in the trapping of one  $\text{Cl}^-$  ion in  $S_{\text{ext}}$ , where the ion coordinates with the backbone amine groups of Arg147 to Gly149 and Ile356 to Ala358 (Dutzler et al., 2003). This distinct arrangement, where the anions are extensively coordinated by the backbone amide group, is known as the “broken-helix” architecture (Dutzler et al., 2002, 2003; Feng et al., 2010). Interestingly, the charge-neutralization mutation at Glu148 converts EcCLC to a  $\text{Cl}^-$  channel (Dutzler et al., 2003; Accardi and Miller, 2004) and similar phenomena have been observed for mammalian CLC-4 and CLC-5 proteins (Zdebik et al., 2008). It has been suggested that the wild-type crystal structure represents the closed state and the E148Q mutation structure, the open state (Dutzler et al., 2003).

The structural information provides an important starting point for comprehending the ion binding and permeation. A

number of computer modeling and simulations of CLC transport proteins have been carried out to study the mechanisms of ion transfer (Bostick and Berkowitz, 2004; Cohen and Schulter, 2004; Corry et al., 2004; Faraldo-Gómez and Roux, 2004; Miloshevsky and Jordan, 2004; Yin et al., 2004; Bisset et al., 2005; Gervasio et al., 2006; Cheng et al., 2007; Engh et al., 2007a,b; Kuang et al., 2007; Wang and Voth, 2009; Coalson and Cheng, 2010, 2011; Ko and Jo, 2010a,b; Miloshevsky et al., 2010; Kieseritzky and Knapp, 2011; Smith and Lin, 2011; Zhang and Voth, 2011; Cheng and Coalson, 2012; Picollo et al., 2012; Yu et al., 2012; Church et al., 2013; Nieto-Delgado et al., 2013; Han et al., 2014; Pezeshki et al., 2014; Chen and Beck, 2016; Jiang et al., 2016; Khantwal et al., 2016; Lee et al., 2016a,b,c; Chenal and Gunner, 2017). Regarding  $\text{Cl}^-$  transport, all studies agree that the protonation of the external gating residue Glu148 is key to antiporter activation and that Ser107 and Tyr445 form an internal gate. Moreover, the reported free-energy barriers for  $\text{Cl}^-$  translocation are relatively low, ranging from 3 to 8 kcal/mol (Cohen and Schulter, 2004; Bisset et al., 2005; Gervasio et al., 2006; Ko and Jo, 2010a; Cheng and Coalson, 2012). However, some issues are still controversial, such as whether the crystal structures of E148A and E148Q mutants correspond to the truly open state (Dutzler et al., 2003; Miloshevsky and Jordan, 2004) and how protonation of E148 is coupled to  $\text{Cl}^-$  binding. (Dutzler et al., 2003; Bostick and Berkowitz, 2004; Miloshevsky and Jordan, 2004; Yin et al., 2004; Gervasio et al., 2006; Ko and Jo, 2010b) The disagreements are partly due to differences in the employed methodology, constructed models, and adopted parameters. While these calculations shed light on the operating mechanisms of CLC transport, complementing experimental measurements, many molecular details are yet to be elucidated.

One long-standing puzzle concerns the extent of protein conformational change associated with  $\text{Cl}^-$  ions passage. The crystal structures of the protein are highly similar in both the closed and open states (Dutzler et al., 2003). This observation prompted people to propose that only local side chain motions are involved in the EcCLC operation. This mechanism differs fundamentally from other antiporters where significant motions of helices or even domains are necessary (Feng et al., 2010, 2012). On the other hand, recent experiments on mutants that are restrained through cross linking between selected helices suggested that rigid-body movements of certain helices play a role (Basilio et al., 2014; Khantwal et al., 2016). Molecular dynamics (MD) simulations so far have not observed significant global conformational changes of the protein; but this could be because global rearrangements can occur in time scales longer than what the simulations have covered. On the other hand, normal-mode vibrational analysis suggested that large-amplitude swinging of helices A and R may be one such conformational change (Miloshevsky et al., 2010). However, it is possible that local and global conformational changes both contribute to some extent.

Carrying a significant charge,  $\text{Cl}^-$  can strongly polarize nearby solvent molecules and pore residues. Moreover, under certain circumstances, the excess charge of  $\text{Cl}^-$  can easily delocalize to the ion's surroundings. Previous quantum-mechanical (QM) calculations using truncated models of EcCLC have indeed

revealed substantial mutual polarization and partial charge transfer for the  $\text{Cl}^-$  ions in the binding sites (Smith and Lin, 2011; Church et al., 2013; Nieto-Delgado et al., 2013). The most prominent manifestations of polarization and charge transfer were found among the  $\pi$  orbitals of the nearby protein atoms, e.g., the atoms of the backbone peptide links and of the side chains of Glu148 and Tyr445. Furthermore, energy decomposition analysis confirmed that the quantum effects of electron delocalization over  $\text{Cl}^-$  and the  $\pi$ -orbitals atoms contributed significantly to the stabilization of the  $\text{Cl}^-$  ions in the biding sites (Church et al., 2013). On the other hand, it has been demonstrated that these induction effects can impact the ion binding and translocation in cation channels (Compont et al., 2004; Allen et al., 2006; Huetz et al., 2006; Bucher et al., 2007, 2010; Dudev and Lim, 2009; Illingworth and Domene, 2009; Bostick and Brooks, 2010; Illingworth et al., 2010; Varma and Rempe, 2010; Roux et al., 2011; Wang et al., 2012). Do the extensive charge redistributions of  $\text{Cl}^-$  affect its transport in EcCLC? It will be interesting to find out.

In this study, we investigate possible protein conformational changes and explore how electron delocalization affects the ion's expedition through EcCLC. We carry out dynamics simulations at both the molecular mechanical (MM) and the combined quantum-mechanical/molecular-mechanical (QM/MM) levels (Warshel and Levitt, 1976; Field et al., 1990; Gao, 1996; Zhang et al., 1999; Rode et al., 2006; Lin and Truhlar, 2007; Senn and Thiel, 2007; van der Kamp and Mulholland, 2013; Pezeshki and Lin, 2015; Duster et al., 2017). We perform steered molecular dynamics (SMD) simulations (Izrailev et al., 1998) to escort a  $\text{Cl}^-$  ion through the pore. We examine the changes in pore size and protein conformations as the ion moves through and estimate the associated potential of mean force (PMF) using umbrella sampling (Torrie and Valleau, 1977; Roux, 1995). By comparing and combing the MM and QM/MM results, this study will deepen our understanding of the  $\text{Cl}^-$  translocation process operated by EcCLC.

## COMPUTATIONAL DETAILS

### Model Preparation

The model system was constructed based on the crystal structure of the wild-type (WT) protein's transmembrane domain (PDB code: 1OTS) (Dutzler et al., 2003). Because the two subunits each function independently, only one subunit (chain A) was used to build the model system. The protonation states of the residues were set according to an earlier Poisson-Boltzmann calculation (Faraldo-Gómez and Roux, 2004) except for the external gate Glu148, which was protonated. The protonated Glu148 was manually rotated outwards to open the pore. Therefore, we were simulating the open state of the protein. The protein with the two bound  $\text{Cl}^-$  ions in the  $S_{\text{int}}$  and  $S_{\text{cen}}$  sites was embedded in a two-layer 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) lipid bilayer. The protein-membrane complex was then solvated by adding water molecules on both sides. The thickness of the slab of water at either side was about 20 Å. We then replaced 109 randomly selected water molecules by 58  $\text{K}^+$  and 51  $\text{Cl}^-$  ions to neutralize the total

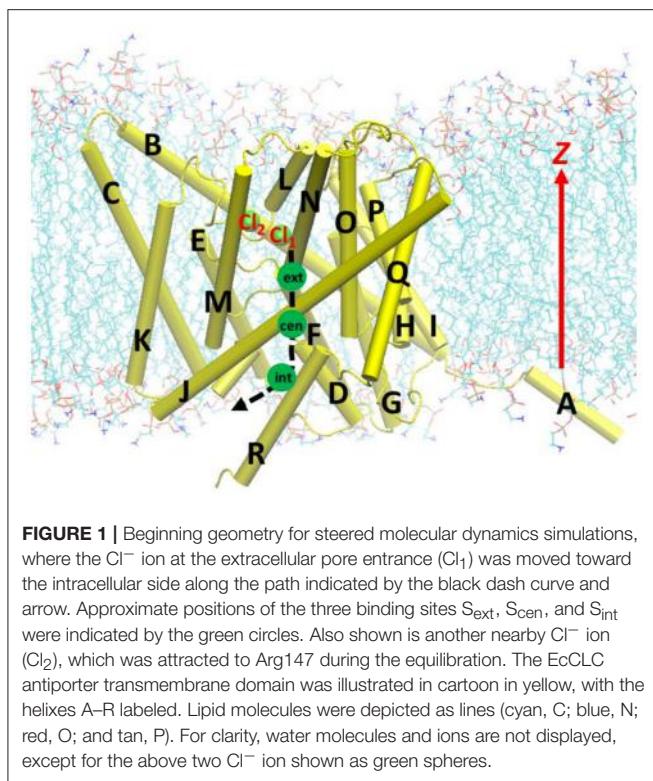
charge and to achieve an approximate 0.15 M physiological salt concentration for the solution. The primary cell of the model system consists of 6823 protein atoms, 18308 water molecules, 60  $\text{Cl}^-$  ions, 51  $\text{K}^+$  ions, and 292 POPE molecules. The protein, lipid, and ions were described by the CHARMM36 force fields (MacKerell et al., 1998; Mackerell et al., 2004; Klauda et al., 2010; Best et al., 2012), and water by the TIP3P model (Jorgensen et al., 1983).

The system was equilibrated under the  $NpT$  ensemble at  $p = 1$  bar and  $T = 310$  K for 10 ns, followed by the  $NVT$  ensemble at  $T = 310$  K for 45 ns, with final unit cell dimensions of  $101.3 \times 99.0 \times 92.9$  Å<sup>3</sup>. Temperature was controlled through Langevin dynamics (Phillips et al., 2005) where the temperature dampening coefficient was set to 1.0 ps<sup>-1</sup>, and pressure through Langevin piston Nosé-Hoover method (Martyna et al., 1994; Feller et al., 1995) with a barostat oscillation period of 175 fs and a damping time of 150 fs. Periodic boundary conditions were employed, and long-range electrostatic interactions were computed by particle mesh Ewald (PME) method (Darden et al., 1993; Essmann et al., 1995). A 14.0 Å cutoff was used for nonbonded interactions, with smoothing switch at 13.0 Å and pair lists cutoff at 16.0 Å. The SHAKE algorithm (Ryckaert et al., 1977) was used to make waters rigid as well as to constrain all bonds between hydrogen and heavy atoms. A time step of 2 fs was used. The equilibrations were performed by using the NAMD (Phillips et al., 2005) program version 2.10. The saved trajectories were visually inspected using the program VMD (Humphrey et al., 1996).

### Steered Molecular Dynamics

The  $\text{Cl}^-$  transport by EcCLC is complicated due to the coupling with the  $\text{H}^+$  migration, and a  $2\text{Cl}^-/1\text{H}^+$  ratio has been established (Accardi and Miller, 2004; Picollo and Pusch, 2005; Scheel et al., 2005; Matulef and Maduke, 2007; Miller and Nguiragool, 2009; Feng et al., 2012; Accardi, 2015). However, exactly how the two  $\text{Cl}^-$  ions were transferred in every cycle is still under debate. As such, we only considered the uncoupled translocation of one  $\text{Cl}^-$  ion to simplify the analysis, which suffices the purpose of this work.

The parameters and program used in the model system equilibration were used in the constant-velocity SMD (Izrailev et al., 1998) simulations, except those indicated below. Before the SMD simulations, the  $\text{Cl}^-$  ion at  $S_{\text{cen}}$  in the equilibrated model was replaced and fixed at the extracellular side at  $z = 0$ , and the model system was re-equilibrated for 2 ns. The resulting geometry, where the  $\text{Cl}^-$  ion stayed just outside the extracellular pore entrance, served as the starting geometry for the SMD simulation (see Figure 1). For the SMD simulations, the  $\text{Cl}^-$  ion was dragged through the pore toward the intracellular side from  $z = 0.0$  to  $-20.0$  Å. Three steering speeds were tested: 2, 5, and 10 Å/ns. (As a comparison, the slowest speed applied in a previous study, Ko and Jo, 2010a was 10 Å/ns). The steering forces were applied along the  $-z$  direction, and the spring force constants were set to 10 kcal/mol/Å<sup>2</sup>. The trajectories were saved every 100 steps. The  $\text{C}\alpha$  atoms of randomly selected protein residues Gly141, Thr226, Ala325, and Leu421, which belongs to helices E, I, L, and Q, respectively, were restrained to their initial positions

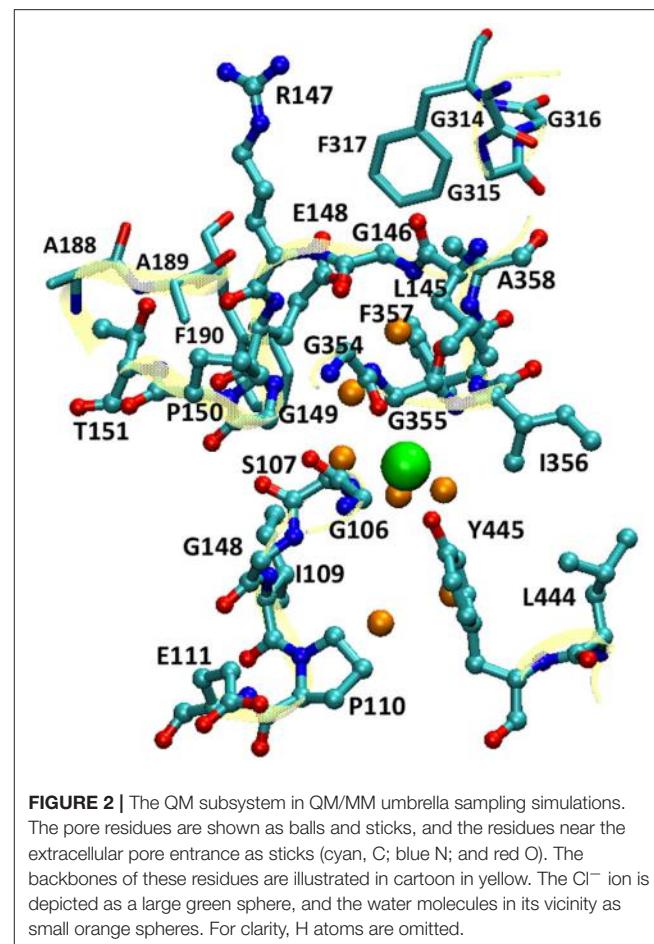


by harmonic potentials with force constants of 1.0 kcal/mol/Å<sup>2</sup> to prevent the protein from translating with the  $\text{Cl}^-$  ion. The simulations were conducted with constant volume and constant temperature. The time step was set to 1.0 fs, as the SHAKE algorithm was not used (the same applied to umbrella sampling below).

## Umbrella Sampling

The starting geometries for umbrella sampling calculations were extracted from the trajectory of the SMD simulations at the slowest speed of 2 Å/ns. The  $z$  component of the  $\text{Cl}^-$  ion's coordinate was chosen to be the reaction coordinate. In total, there were 41 sampling windows covering  $0 > z > -20.0$  Å, with equal spacing of 0.5 Å and each window containing the  $\text{Cl}^-$  ion initially located approximately at the window center. The simulations were done at both the QM/MM and MM levels. The PMF was obtained using the weighted histogram analysis method (WHAM) (Kumar et al., 1992).

For the QM/MM umbrella sampling, the QM subsystem covered the entire pore section that the  $\text{Cl}^-$  ion traverses. This included the  $\text{Cl}^-$  ion, the pore residues (Gly106 to Pro110, Glu111 backbone, Leu145 to Pro150, Thr151 backbone, Gly354 to Ala358, Leu444, and Tyr445), selected residues near the extracellular pore entrance (Ala188 to Phe190 and Gly314 to Gly317), and 7 water molecules in the vicinity of the  $\text{Cl}^-$  ion (Figure 2). For simplicity, the mechanical-embedding scheme with H as link-atom was adopted (Bakowies and Thiel, 1996; Lin and Truhlar, 2005). For the first and last residues in a list of consecutive residues in the QM subsystem, the amine



and carbonyl groups were replaced by H atoms, respectively. When the QM/MM boundary separated backbone and side chain, H atoms were used to cap the  $\text{C}\alpha$  (or  $\text{C}\beta$ ) atom. To prevent MM water molecules from diffusing into the QM region, the MM water molecules adjacent to the QM subsystem (within 10.0 Å from the QM subsystem) were restrained by imposing harmonic potentials to their O atoms with force constants of 1.0 kcal/mol/Å<sup>2</sup>. The QM subsystem thus covered the entire pore section that the  $\text{Cl}^-$  ion passed through. For computational efficiency, the PM3 (Stewart, 1989a,b) semi-empirical Hamiltonian was chosen to model the QM subsystem. Higher-level electronic-structure methods are more accurate, but they were not employed for umbrella sampling due to their much higher computational costs. Periodic boundary conditions with a minimum-image convention were employed, where the cutoffs for electrostatic and van der Waals interactions were set to 14.0 Å with smoothing switches at 13.0 Å. For each window, equilibration of 10 ps and production of 200 ps were performed. A Nose-Hoover thermostat (Hoover, 1985; Kreis et al., 2016) with a coupling constant of 4.0 fs was applied to control the temperature. The force constants of the bias harmonic potential were set to 14 kcal/mol/Å<sup>2</sup>. The QM/MM simulations were done using a local version of the QM/MM program (Lin et al., 2017), which called the MNDO program (Thiel, 2005) for QM

calculations and NAMD for MM calculations, synthesized the QM and MM gradients, and propagated the trajectory.

For the MM umbrella sampling, longer simulations could be afforded, and 0.5 ns equilibration followed by 2.0 ns production run was performed for each window. The force constants of the bias harmonic potential were set to 5 kcal/mol/Å<sup>2</sup>. The same setups and program used in the model system equilibration were again used here, unless otherwise noted.

## Pore Radius Calculations

The pore radius at the Cl<sup>-</sup> ion locations were calculated for representative geometries extracted from umbrella sampling simulations. This analysis was carried out for both the MM and QM/MM simulations. In either case, 7 snapshots were extracted from the production phase of the trajectory for each of the 41 windows. The pore radii were computed utilizing the program HOLE (Smart et al., 1996), where a Monte Carlo-based search algorithm was employed to identify the best route for a sphere with a given radius to pass through the protein. The “hard-core” radii in Turano et al. (1992) was adopted for the calculations, which implicitly account for the thermal motions of atoms. The results of the 7 representative geometries were then averaged for a given window.

## Electron Delocalization

To characterize the Cl<sup>-</sup> electron delocalization during the ion translocation along the pore, electrostatic-embedding QM/MM single-point calculations were carried out on one representative geometry for each window of QM/MM umbrella sampling. The QM subsystem was the same as the one in the umbrella sampling except that the selected residues near the pore entrance (Ala188 to Phe190 and Gly314 to Gly317) were excluded; these residues were treated by MM due to computational cost considerations. We employed the charge-redistribution scheme for the boundary treatment (Lin and Truhlar, 2005). The B3LYP (Becke, 1988, 1993; Lee et al., 1988) functional with the 6-31+G(d) basis set (Ditchfield et al., 1971; Hehre et al., 1972; Franci et al., 1982; Clark et al., 1983) were used for the QM description, and the MM force fields were the same as those in the umbrella sampling. All MM atoms within 12.0 Å from the QM subsystem were included as background point charges in the embedded-QM calculations. The calculations were performed using the local version of QM/MM, which called Gaussian09 (Frisch et al., 2010) for the QM calculations. The Löwdin charges (Löwdin, 1950; Wang et al., 2013) and the natural charges based on Natural Bond Orbital analysis (Foster and Weinhold, 1980; Reed et al., 1985) were computed using the polarized electron density of the QM subsystem. We tested and found that further inclusion of more MM background point charges in the embedded-QM Hamiltonian did not noticeably change the atomic charges for the QM atoms.

## RESULTS

### Equilibrated Model System

During equilibration, the protonated Glu148 side chain, which had been manually rotated out of S<sub>ext</sub>, moved toward S<sub>ext</sub> and

partially obstructed the pore entrance, but it did not enter the pore to reclaim the binding site. Located at the transition area between the intracellular aqueous solution and the pore, the Cl<sup>-</sup> ion at S<sub>int</sub> drifted away during the equilibration. This is in line with the weak binding affinity of Cl<sup>-</sup> at this site (Lobet and Dutzler, 2006), suggesting that S<sub>int</sub> is less critical than S<sub>ext</sub> and S<sub>cen</sub> to the Cl<sup>-</sup> transport. Attracted by the positively charged Arg147, Cl<sup>-</sup> ions from the bulk often diffused to near the pore entrance on the extracellular side and stayed there for up to 1 ns. This additional weak binding site has been reported in previous computational studies, although its exact location varied in the literature (Bostick and Berkowitz, 2004; Cohen and Schulter, 2004; Faraldo-Gómez and Roux, 2004; Yin et al., 2004; Gervasio et al., 2006; Smith and Lin, 2011; Church et al., 2013). It is quite possible that this additional binding site plays a role in the Cl<sup>-</sup> ion recruitment.

### Conformational Changes Along Cl<sup>-</sup> Translocation

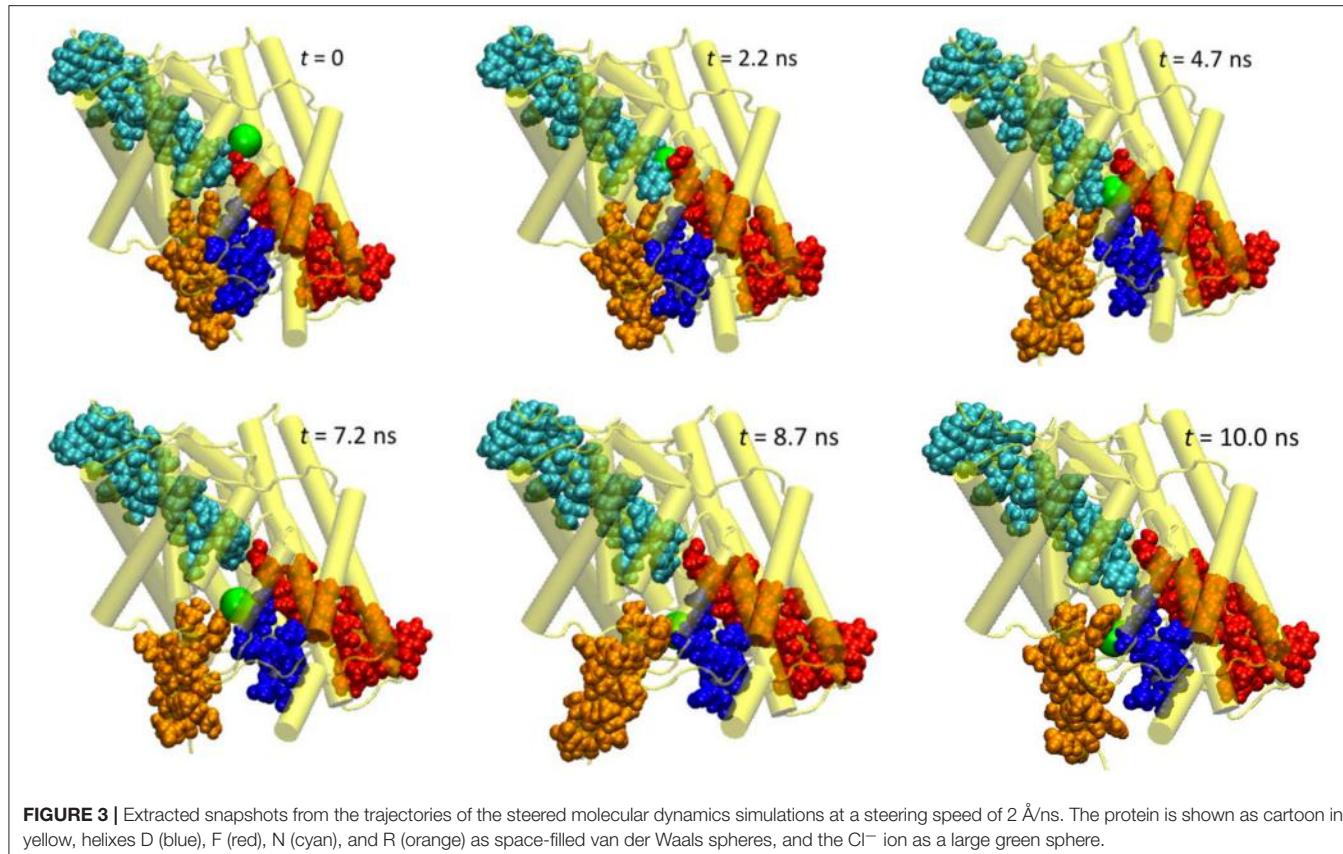
Our analysis on the protein conformational changes will put emphasis on the SMD trajectories, where a complete and continuous voyage was simulated for the Cl<sup>-</sup> ion through the pore.

#### Global Conformational Changes of the Protein

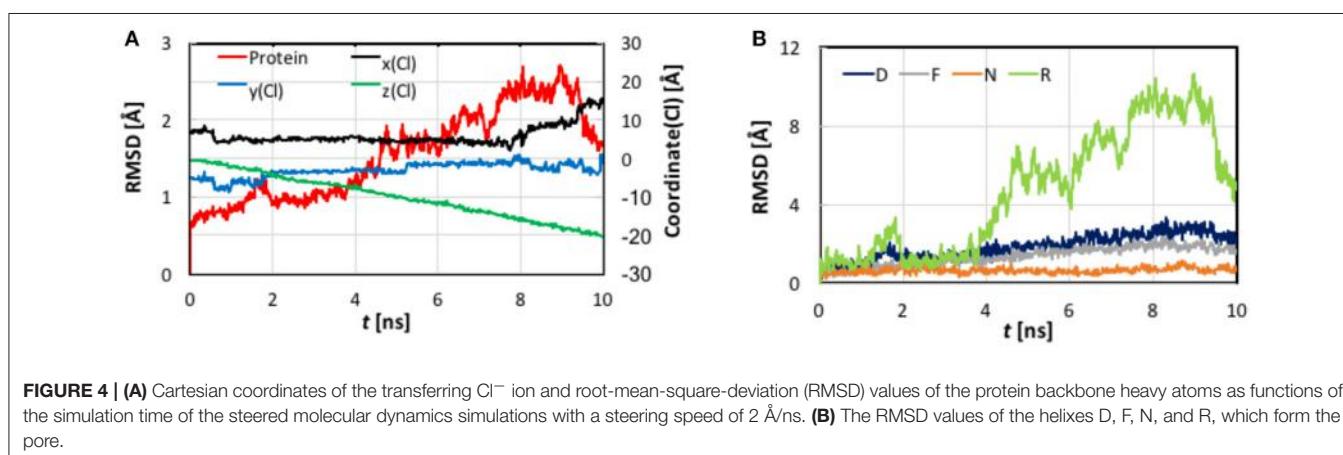
Generally speaking, the SMD trajectories of the three tested steering speeds showed rather similar changes in the protein conformation, but the variations became more prominent as the steering speed decreased. The overall conformational changes can be quantified by root-mean-square-deviation (RMSD) values of the protein backbones, which are plotted as functions of the simulation progress (Figure S1). All trajectories show a trend of increasing RMSD over time. While the trend was moderate at the speeds of 5 and 10 Å/ns, it was considerably enhanced at 2 Å/ns. While this could be incidental, we suspect that it is at least partly due to the longer relaxation time permitted for the model system at a slower steering speed. As such, we will present and discuss the data obtained with the slowest speed of 2 Å/ns, while bringing in the results of 5 and 10 Å/ns only when necessary.

First, let us look at the global backbone movements induced by the Cl<sup>-</sup> passage. Figure 3 shows the orientations of the four helices (D, F, N, and R) that define the pore in selected snapshots extracted from the SMD trajectory. In Figure 4, we display, vs. the simulation time *t*, the Cartesian coordinates of the Cl<sup>-</sup> ion as well as the backbone RMSD values of the protein and of these four helices. (See Figure S2 for the backbone RMSD values of the complete set of all helices).

As evident from Figure 4, the *z* coordinate of the Cl<sup>-</sup> ion varied linearly as a result of constant-velocity steering. The journey of the ion can be divided into three stages. In the first 2 ns, which we call the “entrance” stage, the ion wandered around in the extracellular vestibule just outside the entrance (*z* ~ -4 Å) before finding its way into the pore. In the second “central” stage of *t* = 2–8 ns, the ion walked essentially straight down the path, as manifested by the approximately constant *x* and *y* coordinates. During this section, the ion traveled through S<sub>ext</sub> at *z* ~ -6 Å and S<sub>cen</sub> at *z* ~ -10 Å. The ion’s arrival at the kink (*z* ~ -16 Å)



**FIGURE 3 |** Extracted snapshots from the trajectories of the steered molecular dynamics simulations at a steering speed of  $2 \text{ \AA/ns}$ . The protein is shown as cartoon in yellow, helices D (blue), F (red), N (cyan), and R (orange) as space-filled van der Waals spheres, and the  $\text{Cl}^-$  ion as a large green sphere.



**FIGURE 4 | (A)** Cartesian coordinates of the transferring  $\text{Cl}^-$  ion and root-mean-square-deviation (RMSD) values of the protein backbone heavy atoms as functions of the simulation time of the steered molecular dynamics simulations with a steering speed of  $2 \text{ \AA/ns}$ . **(B)** The RMSD values of the helices D, F, N, and R, which form the pore.

at  $t = \sim 8 \text{ ns}$  marked the beginning of the last “exit” stage. The deviation in the  $y$  coordinate indicated that the pore is curved along this part. On its way out, the ion passed  $S_{\text{int}}$  ( $z \sim -18 \text{ \AA}$ ) at  $t = \sim 9 \text{ ns}$ .

The protein backbone RMSD increased slowly in the first 4 ns, implying that there were relatively small changes in the protein conformation during the entrance and early-central stages. The RMSD then grew rapidly in the next 4 ns from  $\sim 1 \text{ \AA}$  to  $\sim 2.5 \text{ \AA}$ , remained there for  $\sim 1 \text{ ns}$ , and dropped sharply down to  $1.5 \text{ \AA}$  in the last 1 ns. The substantial rise and fall in the RMSD in 4–10 ns

indicated significant conformational changes of the protein over this period. The conclusion is exemplified by the representative snapshots in Figure 3, where large-amplitude rigid-body displacements of part of the protein can be seen during this time. Three helices were identified in the large-amplitude rigid-body displacements induced by the  $\text{Cl}^-$  penetration: helix R most prominently and helices D and F moderately, all of which are critical to the ion passage (Figure 4B).

- *Helix R:* This helix, which contains the internal gating residue Tyr445, exhibited marked swinging motions similar to a door

opening and closing. This motion corresponds with a normal mode presented in an earlier study (Miloshevsky et al., 2010). It appeared that the movement of Tyr445 propagated through the helix, giving rise to the observed big changes in the backbone RMSD. Separating helices R and D, the outward swinging reached its maximum at  $\sim 8.7$  ns. After that, the ion, which had been following the tilted path since  $t \sim 7.9$  ns, exited the pore, and helices R and D returned to their original positions (though not completely restored in the next 2 ns). Because helix R extrudes into the intracellular solution, it is less restrained than the other helices by the rest of the protein or by the membrane, thus possessing larger flexibility. However, one limitation of the current model is that it contains only one subunit. In the dimer structure, helix R may interact with helix A from the other subunit. Moreover, the C-terminal domain, which was not included in this model, is connected to helix R and may also reduce its mobility. Therefore, the motion amplitude of helix R was likely exaggerated here.

- **Helix D:** Another important residue along the  $\text{Cl}^-$  pathway is Ser107, which is situated at the beginning of helix D and, together with Y445, contributes to  $\text{Cl}^-$  binding at  $S_{\text{cen}}$ . The movement of the Ser107 also propagated through helix D and contributed to changes in the backbone RMSD. However, the shorter helix D is embedded deeper into the protein matrix and thus constrained more than helix R. Unsurprisingly, helix D moved only moderately.
- **Helix F:** The  $S_{\text{ext}}$  site is located between the ends of helices F and N, where the amine groups of the backbone extensively coordinate ion. A bottleneck exists between  $S_{\text{ext}}$  and the pore entrance at the extracellular vestibule. However, it seemed that neither helix F nor N underwent significant global rigid-body movements when the  $\text{Cl}^-$  ion passed the bottleneck, implying that the adaptations to accommodate the ion's passage were primarily local. Similar conclusions were drawn in an earlier SMD simulations at a higher steering speed of 10 Å/ns (Ko and Jo, 2010a). During the late central and exit stages of the  $\text{Cl}^-$  excursion ( $t > 4$  ns), helix F displayed swinging motions, albeit only to a modest extent, probably because it was pushed by the nearby helix D in response to the internal gate's opening/closing.

We note that helix A backbone RMSD also exhibited significant changes. Similar to helix R, helix A was exposed to the solvent and thus was quite mobile. However, the trend of its backbone RMSD change was not consistent across the three simulations with various steering speeds (Figure S3). Because helix A resides far from the rest of the protein in the current model, its motions were largely independent of the excursion of the  $\text{Cl}^-$  ion in this study.

A small “bump” in the backbone RMSD of the protein from  $t = 1-2$  ns was predominantly due to the small changes in the backbone RMSD of helices D and R. Inspection of the trajectory unveiled that this was caused by the highly flexible Phe357 side chain, the fluctuation of which disturbed Tyr445 and, in turn, Ser107. However, this happened when the  $\text{Cl}^-$  ion was still in the extracellular vestibule, and Tyr445 had roughly returned to its initial position by  $t = 2$  ns when the ion began entering the

pore. Moreover, such a bump was not observed in the simulations of 5 or 10 Å/ns speed. Therefore, this bump seemed irrelevant or insignificant to  $\text{Cl}^-$  transport.

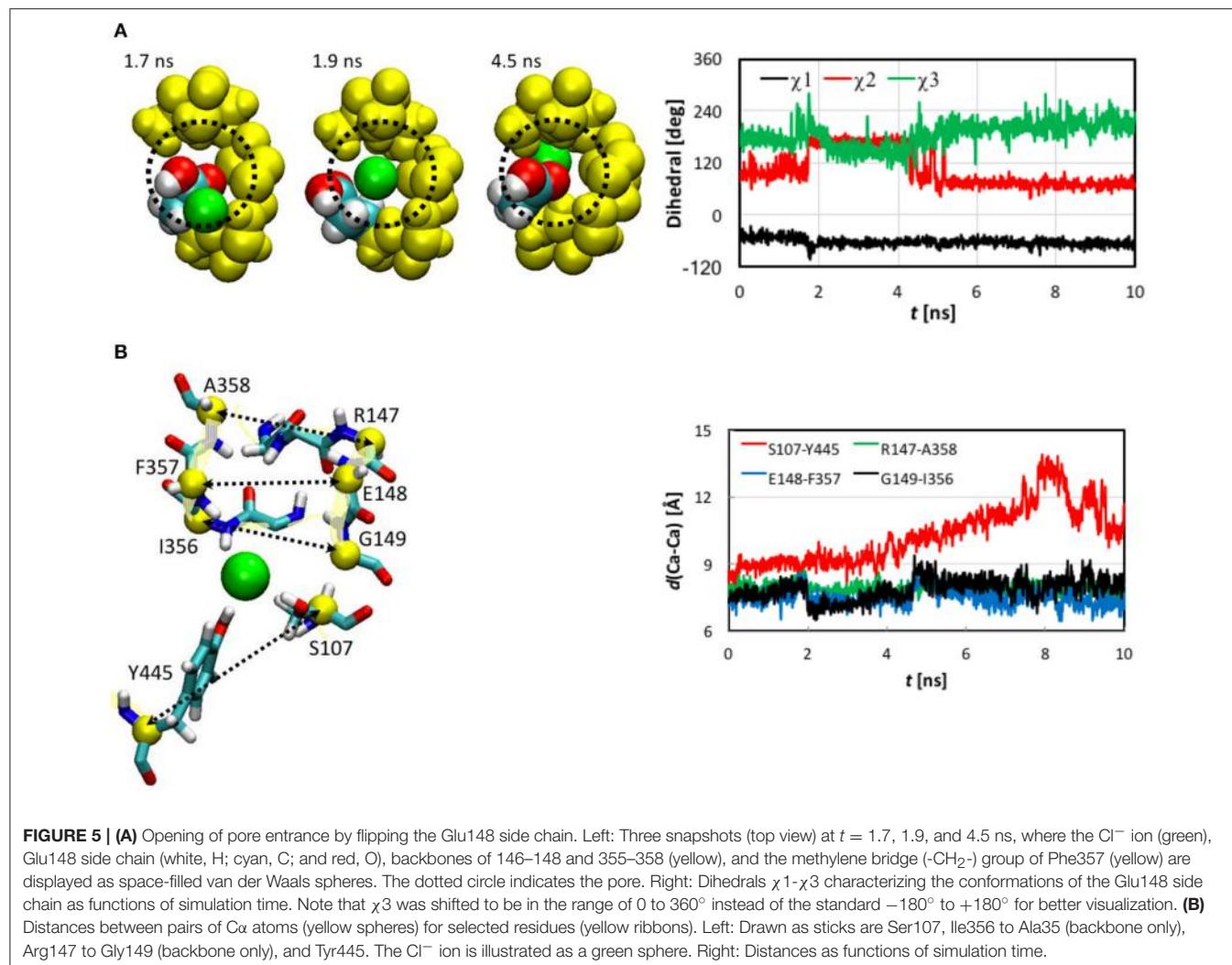
The protein conformations featured in the SMD trajectories were preserved in the trajectories generated by both the QM/MM and MM umbrella sampling simulations. Because the MM umbrella sampling allowed even longer equilibration (2.5 ns for each window), the observation of similar conformations in both the SMD and umbrella sampling simulations supports the hypothesis that other major conformational changes besides the helix R movement, if any, should occur at longer time scales.

## Local Conformational Changes of Pore Residues

Next, we examined the conformational changes of individual residues. The first thing we looked at was how the  $\text{Cl}^-$  ion entered the pore. In the entrance stage of its voyage, the  $\text{Cl}^-$  ion was accompanied by the side chain carbonyl group of Glu148 as well as by the backbone amine groups of Arg147 and Glu148 (Figure S5). Inspection of the trajectory discovered that at  $t \sim 1.8$  ns the side chain carboxyl group of Glu148, which had obstructed the pore entrance, flipped, revealing the pore for the  $\text{Cl}^-$  ion to access. The side chain returned to its initial orientation after the ion passed through  $S_{\text{ext}}$ . The conformational change of Glu148 side chain was characterized by a  $60^\circ$  rotation of dihedral  $\chi_2$ , while  $\chi_1$  and  $\chi_3$  remained largely unchanged (Figure 5A).

The  $\text{Cl}^-$  ion in  $S_{\text{ext}}$  was held by the backbone amine groups of Gly147 to Gly149 and Ile356 to Ala358 (Figure 5B). The Gly146 backbone also stayed close, but the minimum distance from its amine H to the  $\text{Cl}^-$  was 3.3 Å, much longer than the other residue backbones considered here, for which the distances could reach 2.1 Å or shorter. We thus conclude that this interaction was not so critical as the other residue backbones. As the  $\text{Cl}^-$  ion entered  $S_{\text{ext}}$ , the backbones of Glu148 and Gly149 from helix F shifted closer to  $\text{Cl}^-$  to better solvate the ion. In contrast, the backbones of Ile356 to Ala358 from helix N displaced less significantly, which can again be attributed to the tighter embrace of the N helix by the protein matrix. These local conformational changes were well captured by the variations in the backbone RMSD of these residues (Figure S4). Rapid increases can be seen in the backbone RMSD of Glu148 and Gly149 at  $\sim 2.0$  ns. (Note that the side chain RMSD of Glu148 increased  $\sim 0.2$  ns earlier than the rise in the backbone RMSD of Glu148 due to pore opening). The backbone RMSD values of Glu148 and Gly149 remained high for  $\sim 2.4$  ns before appreciably reducing. This period of  $\sim 2.4$  ns corresponded to the ion cruising through  $S_{\text{ext}}$ . Interestingly, although the Phe357 side chain was highly mobile all the time, its activities did not seem very relevant.

The shifts of Glu148 and Gly149 backbones were also reflected by the distances between the  $\text{C}\alpha$  atoms of the involved residues (Figure 5B), where distances reduced by  $\sim 1$  Å were observed for the Glu148-Phe357 and Gly149-Ile356 pairs near  $t = 2.0$  ns and remained low for  $\sim 2.4$  ns. However, we note that, unlike the flipping of Glu148 side chain, the shifts of Glu148 and Gly149 backbones were not a prerequisite of the  $\text{Cl}^-$  ion passage through  $S_{\text{ext}}$ . Rather, they were a consequence. In other words, the ion could still have cruised through without the movement of the



**FIGURE 5 | (A)** Opening of pore entrance by flipping the Glu148 side chain. Left: Three snapshots (top view) at  $t = 1.7$ ,  $1.9$ , and  $4.5$  ns, where the  $\text{Cl}^-$  ion (green), Glu148 side chain (white, H; cyan, C; and red, O), backbones of 146–148 and 355–358 (yellow), and the methylene bridge ( $-\text{CH}_2-$ ) group of Phe357 (yellow) are displayed as space-filled van der Waals spheres. The dotted circle indicates the pore. Right: Dihedrals  $\chi_1$ – $\chi_3$  characterizing the conformations of the Glu148 side chain as functions of simulation time. Note that  $\chi_3$  was shifted to be in the range of  $0$  to  $360^\circ$  instead of the standard  $-180^\circ$  to  $+180^\circ$  for better visualization. **(B)** Distances between pairs of  $\text{C}\alpha$  atoms (yellow spheres) for selected residues (yellow ribbons). Left: Drawn as sticks are Ser107, Ile356 to Ala35 (backbone only), Arg147 to Gly149 (backbone only), and Tyr445. The  $\text{Cl}^-$  ion is illustrated as a green sphere. Right: Distances as functions of simulation time.

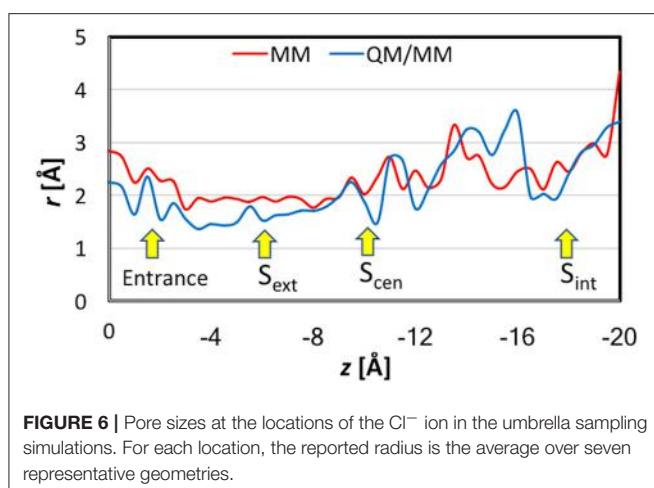
Glu148 and Gly149 backbones, because the pore sizes were large enough in this section. In fact, it has been known that even bigger anions such as  $\text{Br}^-$  can also be transported rather efficiently (the relative permeability  $P_{\text{Br}}^-/P_{\text{Cl}}^- = 0.7$ ; White and Miller, 1981; Gouaux and MacKinnon, 2005). The lack of high selectivity between  $\text{Cl}^-$  and  $\text{Br}^-$  is probably not problematic physiologically because of the much lower physiological abundance of  $\text{Br}^-$  than  $\text{Cl}^-$  ions (Gouaux and MacKinnon, 2005).

The Ser107 and Tyr445 side chains played a central role in the late central and exit stages of the  $\text{Cl}^-$  ion's expedition. The side chain hydroxyl groups of Ser107 and Tyr445 accommodated the  $\text{Cl}^-$  ion as soon as it left  $\text{S}_{\text{ext}}$  ( $t \sim 4.4$  ns) and escorted it all the way through its exit from the channel. Maintaining the ion coordination by these two residues drove the movement of helices D and R and led to the temporary separation of the two helices. Consequently, the distance between the  $\text{C}\alpha$  atoms of Ser107 and Tyr445 increased until the ion exited from the pore (Figure 5B). Interestingly, we did not observe much rotation of the Ser107 side chain after 4.4 ns, during which  $\chi_1$  was mostly constant (Figure S6). The attraction by the  $\text{Cl}^-$  ion might have

somewhat reduced the mobility of this side chain. The Tyr445 side chain  $\chi_1$  and  $\chi_2$  did not really change throughout the entire journey. The side chain hydroxyl groups of both residues remained quite flexible, though.

### Pore Sizes

Figure 6 displays the pore sizes at the locations of the  $\text{Cl}^-$  ion in the umbrella sampling simulations. For each location, the radius was averaged over 7 representative geometries. Please note that this is different from the pore sizes along the pathway for a given snapshot of the protein. The pore size plot here reflects the binding environment of the  $\text{Cl}^-$  ion as it travels down the channel. Nevertheless, the overall shapes of the MM and QM/MM pore-size curves agree reasonably well with those reported previously (Bostick and Berkowitz, 2004; Corry et al., 2004; Yin et al., 2004; Kuang et al., 2008; Ko and Jo, 2010a; Khantwal et al., 2016). The MM pore radius was in general larger by up to  $0.5$  Å than the QM/MM radius, especially at the binding sites.



**FIGURE 6 |** Pore sizes at the locations of the  $\text{Cl}^-$  ion in the umbrella sampling simulations. For each location, the reported radius is the average over seven representative geometries.

## Cl<sup>-</sup> ion Charge Redistribution

In previous studies, (Smith and Lin, 2011; Church et al., 2013; Nieto-Delgado et al., 2013) it was found that the charge carried by the  $\text{Cl}^-$  ion is delocalized to its solvation shell when it is bound at  $S_{\text{ext}}$  and  $S_{\text{cen}}$ . In the present work, we examined the trend along the entire pore. The QM charges were computed using the electrostatic-embedding QM/MM scheme on representative geometries from umbrella sampling simulations. Compared with our early truncated-QM model calculations for the pore in the gas phase (Smith and Lin, 2011; Church et al., 2013) here the electrostatic-embedding QM/MM calculations more realistically incorporated the polarization effects due to the protein/solvent environment. The results for the  $\text{Cl}^-$  ion and for selected atoms and functional groups belonged to three residue cohorts (described later in this section) are depicted in Figure S7 as functions of the reaction coordination  $z$ . The total charges of each cohort are displayed in Figure 7. To better assist visualization, we have divided the reaction pathway into 10 sections (each of 2 Å long) and collected the average charges over each section. Importantly, both the natural and Löwdin charges gave qualitatively similar descriptions, although the Löwdin charges probably exaggerated the extents of charge variation. Therefore, our analysis will focus on the natural charges.

The functional groups in a given cohort were those who critically participated in the  $\text{Cl}^-$  coordination at the corresponding stage of ion transport.

- **Entrance cohort:** side chain carboxyl group of Glu148 and backbones of Arg147 and Glu148
- **Central cohort:** side chain hydroxyl group of Ser107, side chain of Tyr445, and backbones of Arg147-Gly149 and Ile356-Ala358
- **Exit cohort:** side chain hydroxyl group of Ser107 and side chain of Tyr445

Note that there are overlaps in the assignment of the functional groups, because one functional group might participate in the  $\text{Cl}^-$  ion solvation in more than one stage. For example, the entire exit cohort was part of the central cohort. Because our earlier QM analyses had discovered that the transferred charge would

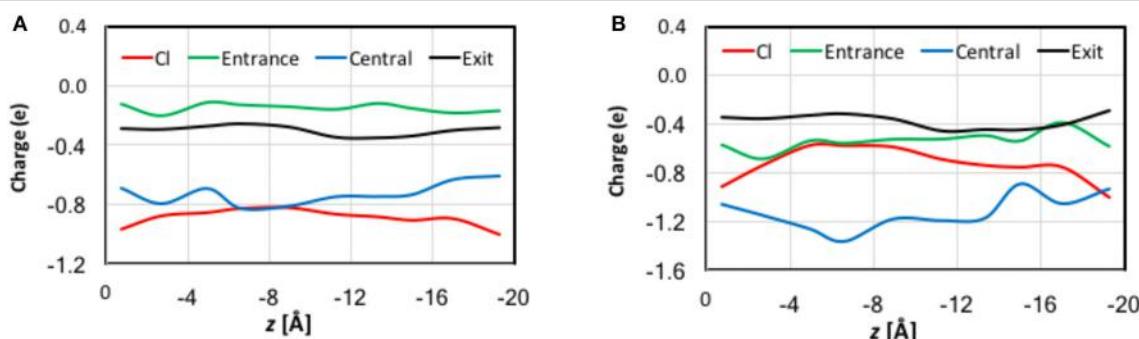
be delocalized over all nearby atoms with  $\pi$  bonds (Church et al., 2013) we treated the entire backbone of a residue rather than just the amine group as one unit in the present analysis. The same was applied to the ring in the Tyr445 side chain.

There are significant charge redistributions between  $\text{Cl}^-$  and the coordinating residues along the ion transport pathway. In the external vestibules, if the  $\text{Cl}^-$  ion had no contact with any cohort residues, it retained a charge of nearly  $-1.0$  e, despite being coordinated by water molecules. As soon as it interacted with the entrance cohort, notable partial charge transfer was observed. The amount of charge loss increased as the ion advanced through the pore and interacted with more residues. The maximal charge deduction was  $\sim 20\%$  and occurred when the ion traveled from  $S_{\text{ext}}$  ( $-6$  Å) to  $S_{\text{cen}}$  ( $-10$  Å). This is not surprising, because the central cohort has the largest number of coordinating functional groups and caused most of the significant charge redistribution. Correspondingly, a dip appeared near  $z = -8$  Å in the total-charge curve of the central cohort. Finally, the exit cohort withdrew some of the charge density from  $\text{Cl}^-$  as the ion approached it. The charge transfer to the exit cohort was the greatest from  $S_{\text{cen}}$  to the kink ( $-10 > z > -16$  Å). After leaving the kink, the ion gradually regained its charge on the way out the channel. Interestingly, the H atoms that directly coordinated the  $\text{Cl}^-$  ion displayed almost negligible changes in their charges.

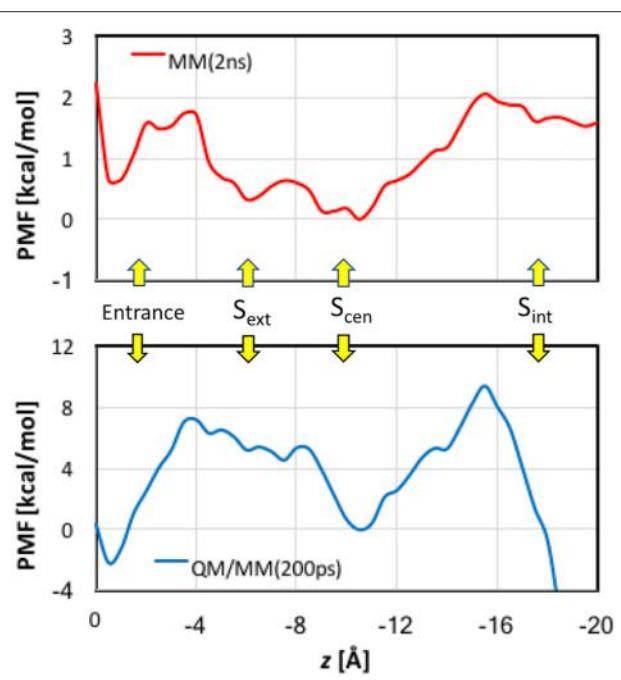
## Potential of Mean Force Through Umbrella Sampling

When analyzing the one-dimensional PMFs, we focused on the data in the range of  $z > -16$  Å, because the pore is wide and considerably tilted after  $z < -16$  Å, for which multi-dimensional PMFs shall provide better descriptions of the ion dynamics. First, we checked the convergence of the MM PMF with respect to the sampling time per window (Figure S8A). The MM PMF converged to within 0.2 kcal/mol from 1.5 ns/window to 2.0 ns/window, suggesting that the sampling time length was adequate. The final MM PMF is shown in Figure 8 as the red curve. The lowest free energy in the entire PMF corresponded to  $S_{\text{cen}}$ , but  $S_{\text{ext}}$  is higher by only  $\sim 0.5$  kcal/mol. The well at the entrance was roughly the same as  $S_{\text{ext}}$ . The PMF possesses a barrier of  $\sim 1.5$  kcal/mol between the pore entrance and  $S_{\text{ext}}$ . This was primarily due to the need to rotate the Glu148 side chain that blocked the channel so that the pore could be accessed. Because our model was constructed in the open state instead of the closed state, this barrier did not account for the changes needed to prepare the protein in the open state. Going from  $S_{\text{ext}}$  to  $S_{\text{cen}}$  experienced only a low barrier of  $\sim 0.3$  kcal/mol. A barrier of  $\sim 2$  kcal/mol must be overcome when the ion departed from the channel. This was mainly caused by the internal gate opening by Tyr445 and Ser107, the movements of which propagated to the backbones of helices R and D.

The convergence of QM/MM PMF was less satisfactory because of the shorter sampling time (Figure S8B). Although the QM/MM PMF is probably not fully converged, the shape of the curve was distinct. The final QM/MM PMF of 200 ps sampling time is also displayed in Figure 8 as the blue curve. The QM/MM PMF agrees with the MM PMF very well in the locations of the



**FIGURE 7** | QM charges for the  $\text{Cl}^-$  ion and the total charges for the entrance, central, and exit cohorts of residue functional groups against the reaction coordinate  $z$  for representative geometries in QM/MM umbrella sampling simulations. Note that to assist better visualization, we have divided the reaction pathway into 10 sections (each of 2 Å long) and collected the average charges over each section. See the text for the cohort definitions. **(A)** Natural charges. **(B)** Löwdin charges.



**FIGURE 8** | Potentials of mean force for  $\text{Cl}^-$  ion passage obtained by umbrella sampling simulations. Note the different scales in the MM and QM/MM plots.

wells and barriers. Although  $S_{\text{cen}}$  still corresponded to the lowest free energy when the ion was in the pore ( $-4 > z > -16$  Å), it was higher by  $\sim 2$  kcal/mol in energy than the entrance, and  $S_{\text{ext}}$  is  $\sim 5$  kcal/mol above  $S_{\text{cen}}$ . The QM/MM PMF exhibited markedly higher barriers for  $\text{Cl}^-$  entering and exiting the pore (8–10 kcal/mol), and conversely a small ( $\sim 0.5$  kcal/mol) barrier from  $S_{\text{ext}}$  to  $S_{\text{cen}}$ .

## DISCUSSION

### External Gating by Glu148

As revealed from our analysis, rotation of Glu148 side chain was necessary to make the pore entrance fully exposed to the

$\text{Cl}^-$  ion. The average  $\chi_2$  values were  $104^\circ$ ,  $167^\circ$ , and  $78^\circ$  for 0–1.8, 1.8–4.4, and 4.4–10 ns, respectively. The corresponding values are  $53^\circ$ ,  $64^\circ$ , and  $64^\circ$  in  $\chi_1$  and  $179^\circ$ ,  $154^\circ$ , and  $201^\circ$  in  $\chi_3$ , respectively (note that  $\chi_3$  has been offset to be in the range of  $0$ – $360^\circ$ ). The  $\sim 60^\circ$  change in  $\chi_2$  upon pore opening was much larger than the variations of  $\sim 10^\circ$  in  $\chi_1$  and  $\sim 25^\circ$  in  $\chi_3$ . The first and third average values of  $\chi_2$  (and of  $\chi_3$ ) are similar, signifying the side chain returning to partially block the pore entrance after departure of the  $\text{Cl}^-$ . We suspect that their existing small difference of  $\sim 20^\circ$  was due to the presence of  $\text{Cl}^-$  at the entrance during the first 1.8 ns.

The pattern of dihedral evolution seems in line with what have been seen in the experimental structures of EcCLC. In the crystal structure of the wild-type (WT) protein, which is in the closed state,  $\chi_1$ – $\chi_3$  (averaged over chains A and B) are  $-25^\circ$ ,  $70^\circ$ , and  $-107^\circ$ , respectively. For the E148Q mutant, which serves as a mimic of the open state, the average values of  $\chi_1$ – $\chi_3$  are  $-36^\circ$ ,  $-57^\circ$ , and  $-84^\circ$ , respectively. Thus, the changes are substantial ( $\sim 130^\circ$ ) in  $\chi_2$  but insignificant in  $\chi_1$  ( $\sim 10^\circ$ ) and  $\chi_3$  ( $\sim 20^\circ$ ). It is encouraging that both the computations and experiments support the gating role of Glu148 side chain.

Our observation of the side chain rotation as the external gating mechanism shares some similarities with what was found for a eukaryotic *Cyanidioschyzon merolae* CLC (CmCLC) transporter in a recent computational study (Cheng and Coalson, 2012). In that paper, constant-force steering was applied to guide  $\text{Cl}^-$  ions through the pore. No large-amplitude motions of the helices were found. The side chain rotation of Glu210 (Glu148 equivalent in CmCLC) was characterized by significant changes in both  $\chi_1$  and  $\chi_2$  ( $>100^\circ$ ). In contrast, only changes in  $\chi_2$  was notable in the present study.

We note that an earlier paper (Bisset et al., 2005) based on the simulations of a pure  $\text{Cl}^-$  channel CLC-0 suggested an alternative gating mechanism of Glu148: The Glu148 side chain was pushed back by the incoming  $\text{Cl}^-$  ions into a more central position and pressed against the channel wall, such that it did not block the travel of  $\text{Cl}^-$  ions. Such a conformation of the external gating residue was also detected in a recent experimental structure of CmCLC transporter (Feng et al., 2010). The discrepancies between this earlier and our current simulations may arise from

the different proteins and/or the different starting geometries of Glu148 (the protonated Glu148 might stay initially deeper in the pore in the previous computation). On the other hand, it is certainly possible for the Glu148 side chain to move in either direction to escape from  $S_{ext}$ . Therefore, while not observed in our simulations, we cannot rule out this alternative mechanism for pore opening. However, because EcCLC is an antiporter, the release of  $H^+$  to the extracellular side will probably require Glu148 be exposed to the extracellular solution sometime during a transport cycle.

### Sext-Free vs. Sext-Occupied Open States

It is intriguing to see that the presence of the  $Cl^-$  ion in  $S_{ext}$  induced subtle local movements of Glu148 and Gly149 backbones. The approach of helix F to helix N can be attributed to the negatively charge of  $Cl^-$ , which “glued” the two helix N-terminal ends. Note that these N-terminal ends are partially positively charged. The departure of the ion from  $S_{ext}$  thus widened  $S_{ext}$  again. These open-state geometries with and without the anion in  $S_{ext}$  can be labeled as  $S_{ext}$ -free and  $S_{ext}$ -occupied, respectively. Because the deprotonated carboxyl groups of Glu148 also carries a negative charge, the close-state geometry and the  $S_{ext}$ -occupied open-state geometry should bear similarities in the  $S_{ext}$  region. This is indeed the case for the crystal structures (Dutzler et al., 2002, 2003). As can be seen in **Table 1**, the distances between  $C\alpha$  atom pairs for Arg147-Ala358, Glu148-Phe357, and Gly149-Ile356 (**Figure 5B**), which characterize the size of  $S_{ext}$ , are rather similar across the three crystal structures. Averaged over the three pairs and over chains A and B, the distances are 7.0, 6.9, and 7.0 Å for WT, E148A, and E148Q, respectively. In view of this, the “true” open state could be the  $S_{ext}$ -free one, for which E148Q without an anion at  $S_{ext}$  will be a realistic mimic.

We note that the distances in our SMD simulations are consistently larger than those from the crystal structure. Even for those of 2.0–4.4 ns corresponding to the  $S_{ext}$ -occupied state, the

distance average over these three pairs is 7.5 Å, ~0.5 Å wider than the crystal structures. The average distance over the  $S_{ext}$ -free state is 7.8 Å. Currently we do not know exactly what caused the difference, although the different environments in which the protein stays may have contributed to the disparity—the protein is more tightly packed in the crystal form but probably more expanded when embedded in membrane.

### Quantum Effects of Electron Delocalization on $Cl^-$ Translocation

Comparisons between the MM and QM/MM results obtained in this study provide an opportunity to gauge the quantum effects of electron delocalization on  $Cl^-$  transport by EcCLC. First, we point out that the MM (2 kcal/mol) and QM/MM (10 kcal/mol) barriers estimated here for  $Cl^-$  translocation are close to those (3–8 kcal/mol) obtained in previous computational studies (Cohen and Schulten, 2004; Gervasio et al., 2006; Ko and Jo, 2010a; Cheng and Coalson, 2012) again suggesting that the slow transport of  $Cl^-$  ions ( $\sim 10^3$  s $^{-1}$ ) by EcCLC (Walden et al., 2007) arises from other factors not considered here. The same locations of the barriers and wells in both MM and QM/MM PMFs implied that electron delocalization was fine tuning rather than dramatically changing the  $Cl^-$  transport mechanism. However, the QM/MM barriers are much higher than the MM barriers, suggesting stronger binding, especially at the entrance and  $S_{cen}$ , at the QM/MM level than at the MM level.

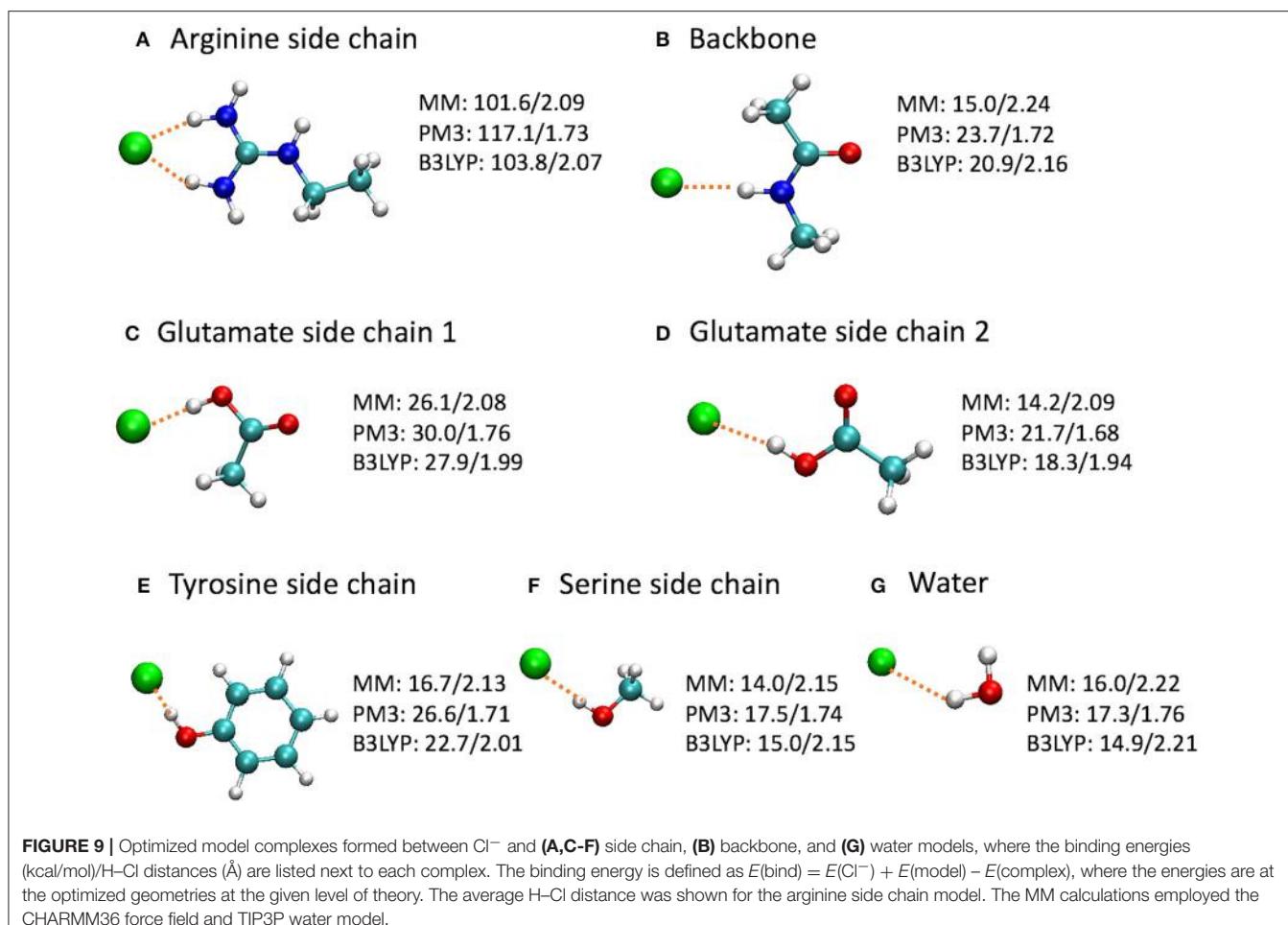
To gain a deeper insight on the differences, we computed the binding energies for a few model complexes formed between  $Cl^-$  and residue backbone or side chains at the CHARMM36, PM3, and B3LYP/6-31+G(d) levels (**Figure 9**). Note that the Glu– $Cl^-$  complex has two different conformations, both of which had been observed for Glu148 in the umbrella sampling trajectories. The positively charged arginine side chain was also included because of the strong attraction between the Arg147 side chain and the  $Cl^-$  in the entrance, although the Arg147 side chain is not part of the entrance cohort. Among the three levels of theory, the B3LYP calculations are the most accurate and serve as the reference for comparisons. Overall, the QM and MM optimized geometries were similar. However, in the PM3 calculations, the binding energies were overestimated by 7–17% and the bond lengths were shorter by 10–20% than the reference calculations. More specifically, the errors were 13 kcal/mol for the arginine side chain and 2–4 kcal/mol for the other models. In contrast, the H– $Cl$  distances by CHARMM36 calculations agreed almost perfectly with B3LYP for water, serine side chain, and arginine side chain, but were somewhat too long (by ca. 4–8%) for the backbone and the side chains of glutamate and tyrosine. Not surprisingly, the CHARMM36 performance in energy seemed mixed. While performing remarkably well (within 2 kcal/mol, or within 7%) for water, serine side chain, and the first (*anti*) conformation of glutamate side chain, CHARMM36 noticeably underestimated the binding energies for the backbone, tyrosine side chain, and the second (*syn*) conformation of glutamate side chain by 4–6 kcal/mol, or 23–28%.

Interestingly, the model complexes on which CHARMM36 performed poorly contain  $\pi$  bonds. As our earlier study revealed,

**TABLE 1** | Average distances (in Å) between pairs of  $C\alpha$  atoms for selected residues: Arg147-Ala358, Glu148-Phe357, Gly149-Ile356, and Ser107-Tyr445.

	Arg147-Ala358	Glu148-Phe357	Gly149-Ile356	Ser107-Tyr445
SMD	0–2.0 ns	8.05	7.58	7.75
	2.0–4.4 ns	7.83	7.37	7.34
	4.4–10.0 ns	7.90	7.54	8.12
Exp	WT (chain A)	7.34	6.49	7.22
	WT (chain B)	7.38	6.53	7.24
	E148A (chain A)	7.06	6.22	7.50
	E148A (chain B)	7.08	6.06	7.62
	E148Q (chain A)	7.21	6.59	7.06
	E148Q (chain B)	7.28	6.53	7.11

In addition to the SMD data, the experimental data are also provided for wild-type (WT, PDB code 1OTS) and two mutants E148A (PDB code 1OTT) and E148Q (PDB code 1OTU) (Dutzler et al., 2003).



these  $\pi$  bonds stabilize the  $\text{Cl}^-$  ion substantially through mutual polarization and partial charge transfer (Church et al., 2013). The MM calculation failed to account for these quantum effects of electron delocalization in anion- $\pi$  interactions. On the other hand, MM describes classical electrostatic interaction very well. Therefore, CHARMM36 significantly outperformed PM3 for the arginine side chain, which, although contains  $\pi$  bonds, interacts with  $\text{Cl}^-$  predominantly through electrostatic attractions. This also explains the different performances by MM for the two conformations of glutamate side chain. As the  $\text{Cl}^-$  ion stands closer to the  $\text{CH}_3$  group in the first conformation, electrostatic interactions contribute a larger share in the total binding energies. This was confirmed by natural orbital analysis, which revealed less electron delocalization in the first conformation than in the second: the  $\text{Cl}^-$  ion lost 0.107 e in the first conformation but 0.140 e in the second conformation. As expected, CHARMM36 reproduced the interaction energy of the first conformation more accurately.

One may argue that the MM parameters were optimized for condensed-phase simulations, while these model complexes are gas-phase calculations. This is certainly true. However, this means that MM typically *overestimates* the binding energies for gas-phase models. In contrast, MM underestimates the binding

affinities here. Thus, these inaccurate MM energies are most likely caused by the missing stabilizations due to quantum electron delocalization.

While the overestimation of binding energies at the PM3 level likely increased QM/MM barriers, the underestimation of binding energy at the CHARMM36 level probably lowered MM barriers. In particular, the significantly overestimated attraction by Arg147 side chain near the entrance at the PM3 level contributed to an artificially deep well in  $0 > z > 2 \text{ \AA}$  in the QM/MM PMF. The dip at  $S_{\text{cen}}$ , however, might be described more realistically in the QM/MM PMF curve because of the substantial under-binding with the Tyr445 side chain by CHARMM36. It is a bit puzzling that  $S_{\text{ext}}$ , where the  $\text{Cl}^-$  ion was extensively coordinated by backbone amine groups, was  $\sim 5$  kcal/mol higher than  $S_{\text{cen}}$  in the QM/MM PMF, despite the over-binding with the backbone in the PM3 calculations. We do not know the reason, but it could be due to that the QM/MM PMF had not fully converged, as the barrier between  $S_{\text{ext}}$  and  $S_{\text{cen}}$  seemed to keep increasing with extending simulation time (Figure S8).

The differences between CHARMM36 and PM3 bindings also manifest in the pore-size plots (Figure 6). The stronger PM3 binding led to generally smaller pore sizes in QM/MM, especially

in the binding sites. Interestingly, while the MM pore radii were  $\sim 1.8$  Å or larger, which correspond to the generally accepted  $\text{Cl}^-$  ionic radius of 1.81 Å (Shannon, 1976) the QM/MM radii could go down to as small as  $\sim 1.4$  Å. This smaller size indicated that the anion's size was effectively reduced as it lost substantial amount of its electron density. For reference, we note that the radius of the charge-neutral Cl atom was determined to be 0.99 Å (Pyykkö and Atsumi, 2009).

## SUMMARY

We have carried out SMD simulations to escort a  $\text{Cl}^-$  ion through the pore of EcCLC to identify the accompanied local and global conformational changes of the protein. We observed that the side chain rotation of the external gating residue Glu148 was essential for the full access of the pore entrance by  $\text{Cl}^-$ . Occupation of  $S_{\text{ext}}$  by the anion induced small but non-negligible shifting of the Glu148 and Gly149 backbones, which came closer to  $\text{Cl}^-$  for better ion solvation. This local conformational change implied subtle differences between the  $S_{\text{ext}}$ -free and  $S_{\text{ext}}$ -occupied open states. The ion's passing through the internal gating residue Tyr445 prompted helix R, which extrudes into the intracellular solution, to swing away from helix D. Helix R returned to its initial position once the ion exited sideways from the pore. There was substantial electron delocalization of  $\text{Cl}^-$  to its surroundings, especially to the  $\pi$ -bonds (e.g., the Tyr445 side chain), along the ion's journey through the channel. Umbrella sampling simulations at both the MM and QM/MM levels were performed to quantify the free-energy profile associated with  $\text{Cl}^-$  transport. The results identified major barriers for  $\text{Cl}^-$  moving into  $S_{\text{ext}}$  from the extracellular vestibule and exiting the pore from  $S_{\text{cen}}$ . The barrier heights were  $\sim 2$  kcal/mol in MM and  $\sim 10$  kcal/mol in QM/MM, respectively. The higher barriers in the QM/MM than MM free-energy landscapes were attributed to the differences in the interaction energies between  $\text{Cl}^-$  and nearby residues, which were overestimated by PM3 but underestimated by CHARMM36. The QM/MM and MM results here probably provide the upper and lower bounds for the barrier heights. The weaker binding interactions predicted by MM was primarily caused by missing stabilization from electron delocalization between  $\text{Cl}^-$  and functional groups containing  $\pi$  bonds. The findings here suggest that quantum effects of electron delocalization may be more important than previously considered and should be taken into account if more accurate descriptions are desired for  $\text{Cl}^-$  transport proteins.

However, some of the above conclusions must be taken with caution, considering the intrinsic limitations in the model and methods employed this study. The current model system consisted of only one independent subunit of the transmembrane domain, while EcCLC contains a C-terminal domain and usually forms dimers. The amplitude of motion for helix R may be decreased in the presence of the C-terminal domain and the other subunit. The slowest steering speed in the SMD simulations was 2 Å/ns, which, although quite slow, might still be too fast

for near-equilibrium steering. The QM/MM umbrella sampling was conducted employing the semi-empirical PM3 Hamiltonian with relatively short (200 ps per window) sampling time due to computational cost considerations. More advanced electronic-structure theory and longer simulations times are certainly desired for more accurate calculations. Classical polarizable force fields that explicitly include many-body contributions offer another choice (Kaminski et al., 2004; Patel and Brooks, 2004; Patel et al., 2004; Jorgensen, 2007; Warshel et al., 2007; Xie and Gao, 2007; Cisneros et al., 2014, 2016; Vanommeslaeghe and MacKerell, 2015; Albaugh et al., 2016). Although polarizable force fields are more expensive than standard non-polarizable force fields, they are more efficient than QM calculations. The disadvantage is that force fields based on classical electrostatics do not explicitly account for quantum effects such as polarization-exchange coupling and charge transfer (Illingworth and Domene, 2009). The simulations translocated only one  $\text{Cl}^-$  ion, leaving out the scenarios of two or more  $\text{Cl}^-$  ion moving together. The existence and movements of other  $\text{Cl}^-$  ions in the channel will likely impact the binding and translocation of a given  $\text{Cl}^-$  ion. The possible coupling between  $\text{Cl}^-$  ion transport and  $\text{H}^+$  delivery was also missed. Finally, the model we constructed was an "open-state" model, so the conversion between the closed and open states has not been visited. Future studies will need to address these issues.

## AUTHOR CONTRIBUTIONS

HL designed the project. AD carried out the code implementation. C-HW, BA, MZ, and AD performed the calculations. C-HW, AD, BA, and HL analyzed the results. C-HW, AD, and HL wrote the manuscript, which is then revised by all authors, who have given approval to the final version of the manuscript.

## FUNDING

This work is supported by the National Science Foundation (CHE-1564349), Camille and Henry Dreyfus Foundation (TH-14-028), NVIDIA Corporation, and the University of Colorado Denver Undergraduate Research Opportunity Program. This work used XSEDE under grant CHE-140070, supported by NSF grant number ACI-1053575, and NERSC under grant m2495.

## ACKNOWLEDGMENTS

We thank Christal Davis and Christina Garza for insightful discussion.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2018.00062/full#supplementary-material>

## REFERENCES

- Accardi, A. (2015). Structure and gating of CLC channels and exchangers. *J. Physiol.* 593, 4129–4138. doi: 10.1113/jp270575
- Accardi, A., Lobet, S., Williams, C., Miller, C., and Dutzler, R. (2006). Synergism between halide binding and proton transport in a CLC-type exchanger. *J. Mol. Biol.* 362, 691–699. doi: 10.1016/j.jmb.2006.07.081
- Accardi, A., and Miller, C. (2004). Secondary active transport mediated by a prokaryotic homologue of CLC  $\text{Cl}^-$  channels. *Nature* 427, 803–807. doi: 10.1038/nature02314
- Accardi, A., and Picollo, A. (2010). CLC channels and transporters: proteins with borderline personalities. *Biochim. Biophys. Acta* 1798, 1457–1464. doi: 10.1016/j.bbamem.2010.02.022
- Accardi, A., Walden, M., Nguitragool, W., Jayaram, H., Williams, C., and Miller, C. (2005). Separate ion pathways in a  $\text{Cl}^-/\text{H}^+$  exchanger. *J. Gen. Physiol.* 126, 563–570. doi: 10.1085/jgp.200509417
- Albaugh, A., Boateng, H. A., Bradshaw, R. T., Demerdash, O. N., Dziedzic, J., Mao, Y., et al. (2016). Advanced potential energy surfaces for molecular simulation. *J. Phys. Chem. B* 120, 9811–9832. doi: 10.1021/acs.jpcb.6b06414
- Allen, T. W., Andersen, O. S., and Roux, B. (2006). Ion permeation through a narrow channel: using gramicidin to ascertain all-atom molecular dynamics potential of mean force methodology and biomolecular force fields. *Biophys. J.* 90, 3447–3468. doi: 10.1529/biophysj.105.077073
- Bakowies, D., and Thiel, W. (1996). Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.* 100, 10580–10594. doi: 10.1021/jp9536514
- Basilio, D., Noack, K., Picollo, A., and Accardi, A. (2014). Conformational changes required for  $\text{H}^+/\text{Cl}^-$  exchange mediated by a CLC transporter. *Nat. Struct. Mol. Biol.* 21, 456–464. doi: 10.1038/nsmb.2814
- Becke, A. D. (1988). Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* 38, 3098–3100. doi: 10.1103/PhysRevA.38.3098
- Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98, 5648–5652. doi: 10.1063/1.464913
- Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., et al. (2012). Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone  $\varphi$ ,  $\psi$  and side-chain  $\chi_1$  and  $\chi_2$  dihedral angles. *J. Chem. Theory Comput.* 8, 3257–3273. doi: 10.1021/ct300400x
- Bisset, D., Corry, B., and Chung, S.-H. (2005). The fast gating mechanism in CLC-0 channels. *Biophys. J.* 89, 179–186. doi: 10.1529/biophysj.104.053447
- Bostick, D. L., and Berkowitz, M. L. (2004). Exterior site occupancy infers chloride-induced proton gating in a prokaryotic homolog of the CLC chloride channel. *Biophys. J.* 87, 1686–1696. doi: 10.1529/biophysj.104.042465
- Bostick, D. L., and Brooks, C. L. (2010). Selective complexation of  $\text{K}^+$  and  $\text{Na}^+$  in simple polarizable ion-ligating systems. *J. Am. Chem. Soc.* 132, 13185–13187. doi: 10.1021/ja106197e
- Bucher, D., Guidoni, L., Carloni, P., and Rothlisberger, U. (2010). Coordination numbers of  $\text{K}^+$  and  $\text{Na}^+$  ions inside the selectivity filter of the KcsA potassium channel: Insights from first-principles molecular dynamics. *Biophys. J.* 98, L47–L49. doi: 10.1016/j.bpj.2010.01.064
- Bucher, D., Guidoni, L., and Rothlisberger, U. (2007). The protonation state of the Glu-71/Asp-80 residues in the KcsA potassium channel: a first-principles QM/MM molecular dynamics study. *Biophys. J.* 93, 2315–2324. doi: 10.1529/biophysj.106.102509
- Chen, T.-Y. (2005). Structure and function of CLC channels. *Annu. Rev. Physiol.* 67, 809–839. doi: 10.1146/annurev.physiol.67.032003.153012
- Chen, Z., and Beck, T. L. (2016). Free energies of ion binding in the bacterial CLC-Ec1 chloride transporter with implications for the transport mechanism and selectivity. *J. Phys. Chem. B* 120, 3129–3139. doi: 10.1021/acs.jpcb.6b01150
- Chenal, C., and Gunner, M. R. (2017). Two  $\text{Cl}^-$  ions and a glu compete for a helix cage in the CLC proton/ $\text{Cl}^-$  antiporter. *Biophys. J.* 113, 1025–1036. doi: 10.1016/j.bpj.2017.07.025
- Cheng, M. H., and Coalson, R. D. (2012). Molecular dynamics investigation of  $\text{Cl}^-$  and water transport through a eukaryotic CLC transporter. *Biophys. J.* 102, 1363–1371. doi: 10.1016/j.bpj.2012.01.056
- Cheng, M. H., Mamonov, A. B., Dukes, J. W., and Coalson, R. D. (2007). Modeling the fast gating mechanism in the CLC-0 chloride channel. *J. Phys. Chem. B* 111, 5956–5965. doi: 10.1021/jp063993h
- Church, J., Pezeshki, S., Davis, C., and Lin, H. (2013). Charge transfer and polarization for chloride ions bound in CLC transport proteins: natural bond orbital and energy decomposition analyses. *J. Phys. Chem. B* 117, 16029–16043. doi: 10.1021/jp409306x
- Cisneros, G. A., Karttunen, M., Ren, P., and Sagui, C. (2014). Classical electrostatics for biomolecular simulations. *Chem. Rev.* 114, 779–814. doi: 10.1021/cr300461d
- Cisneros, G. A., Wikfeldt, K. T., Ojamäe, L., Lu, J., Xu, Y., Torabifard, H., et al. (2016). Modeling molecular interactions in water: from pairwise to many-body potential energy functions. *Chem. Rev.* 116, 7501–7528. doi: 10.1021/acs.chemrev.5b00644
- Clark, T., Chandrasekhar, J., Spitznagel, G. W., and Schleyer, P. V. R. (1983). Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li-F. *J. Comput. Chem.* 4, 294–301. doi: 10.1002/jcc.540040303
- Coalson, R. D., and Cheng, M. H. (2010). Discrete-state representation of ion permeation coupled to fast gating in a model of CLC chloride channels: comparison to multi-ion continuous space brownian dynamics simulations. *J. Phys. Chem. B* 114, 1424–1433. doi: 10.1021/jp907965b
- Coalson, R. D., and Cheng, M. H. (2011). Discrete-state representation of ion permeation coupled to fast gating in a model of CLC-chloride channels: analytic estimation of the state-to-state rate constants. *J. Phys. Chem. A* 115, 9633–9642. doi: 10.1021/jp200749s
- Cohen, J., and Schulten, K. (2004). Mechanism of anionic conduction across CLC. *Biophys. J.* 86, 836–845. doi: 10.1016/S0006-3495(04)74159-4
- Compoint, M., Ramseyer, C., and Huetz, P. (2004). Ab initio investigation of the atomic charges in the KcsA channel selectivity filter. *Chem. Phys. Lett.* 397, 510–515. doi: 10.1016/j.cplett.2004.09.033
- Corry, B., O'Mara, M., and Chung, S.-H. (2004). Conduction mechanisms of chloride ions in CLC-type channels. *Biophys. J.* 86, 846–860. doi: 10.1016/S0006-3495(04)74160-0
- Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: an  $N\log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi: 10.1063/1.464397
- Ditchfield, R., Hehre, W. J., and Pople, J. A. (1971). Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules. *J. Chem. Phys.* 54, 724–728.
- Dudek, T., and Lim, C. (2009). Determinants of  $\text{K}^+$  vs.  $\text{Na}^+$  selectivity in potassium channels. *J. Am. Chem. Soc.* 131, 8092–8101. doi: 10.1021/ja900168k
- Duster, A. W., Wang, C.-H., Garza, C. M., Miller, D. E., and Lin, H. (2017). Adaptive quantum/molecular mechanics: what have we learned, where are we, and where do we go from here? *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 7:e1310. doi: 10.1002/wcms.1310
- Dutzler, R. (2006). The CLC family of chloride channels and transporters. *Curr. Opin. Struct. Biol.* 16, 439–446. doi: 10.1016/j.sbi.2006.06.002
- Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T., and MacKinnon, R. (2002). X-ray structure of a CLC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* 415, 287–294. doi: 10.1038/415287a
- Dutzler, R., Campbell, E. B., and MacKinnon, R. (2003). Gating the selectivity filter in CLC chloride channels. *Science* 300, 108–112. doi: 10.1126/science.1082708
- Elvington, S. M., Liu, C. W., and Maduke, M. C. (2009). Substrate-driven conformational changes in CLC-ec1 observed by fluorine NMR. *EMBO J.* 28, 3090–3102. doi: 10.1038/embj.2009.259
- Engh, A. M., Faraldo-Gómez, J. D., and Maduke, M. (2007a). The mechanism of fast-gate opening in CLC-0. *J. Gen. Physiol.* 130, 335–349. doi: 10.1085/jgp.200709759
- Engh, A. M., Faraldo-Gómez, J. D., and Maduke, M. (2007b). The role of a conserved lysine in chloride- and voltage-dependent CLC-0 fast gating. *J. Gen. Physiol.* 130, 351–363. doi: 10.1085/jgp.200709760
- Engh, A. M., and Maduke, M. (2005). Cysteine accessibility in CLC-0 supports conservation of the CLC intracellular vestibule. *J. Gen. Physiol.* 125, 601–617. doi: 10.1085/jgp.200509258
- Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *J. Chem. Phys.* 103, 8577–8593. doi: 10.1063/1.470117
- Estévez, R., Schroeder, B. C., Accardi, A., Jentsch, T. J., and Pusch, M. (2003). Conservation of chloride channel structure revealed by an inhibitor binding site in CLC-1. *Neuron* 38, 47–59. doi: 10.1016/S0896-6273(03)00168-5

- Faraldo-Gómez, J. D., and Roux, B. (2004). Electrostatics of ion stabilization in a CLC chloride channel homologue from *Escherichia coli*. *J. Mol. Biol.* 339, 981–1000. doi: 10.1016/j.jmb.2004.04.023
- Feller, S. E., Zhang, Y., Pastor, R. W., and Brooks, B. R. (1995). Constant pressure molecular dynamics simulation: the langevin piston method. *J. Chem. Phys.* 103, 4613–4621. doi: 10.1063/1.470648
- Feng, L., Campbell, E. B., Hsiung, Y., and MacKinnon, R. (2010). Structure of a eukaryotic CLC transporter defines an intermediate state in the transport cycle. *Science* 330, 635–641. doi: 10.1126/science.1195230
- Feng, L., Campbell, E. B., and MacKinnon, R. (2012). Molecular mechanism of proton transport in CLC  $\text{Cl}^-/\text{H}^+$  exchange transporters. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11699–11704. doi: 10.1073/pnas.1205764109
- Field, M. J., Bash, P. A., and Karplus, M. (1990). A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* 11, 700–733. doi: 10.1002/jcc.540110605
- Foster, J. P., and Weinhold, F. (1980). Natural hybrid orbitals. *J. Am. Chem. Soc.* 102, 7211–7218. doi: 10.1021/ja00544a007
- Franci, M. M., Pietro, W. J., Hehre, W. J., Binkley, J. S., DeFrees, D. J., Pople, J. A., et al. (1982). Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* 77, 3654–3665.
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2010). *Gaussian09*. Version, B.01. Wallingford, CT: Gaussian, Inc.
- Gao, J. (1996). Methods and applications of combined quantum mechanical and molecular mechanical potentials. *Rev. Comput. Chem.* 7, 119–185. doi: 10.1002/9780470125847.ch3
- Gervasio, F. L., Parrinello, M., Ceccarelli, M., and Klein, M. L. (2006). Exploring the gating mechanism in the CLC chloride channel via metadynamics. *J. Mol. Biol.* 361, 390–398. doi: 10.1016/j.jmb.2006.06.034
- Gouaux, E., and MacKinnon, R. (2005). Principles of selective ion transport in channels and pumps. *Science* 310, 1461–1465. doi: 10.1126/science.1113666
- Han, W., Cheng, R. C., Maduke, M. C., and Tajkhorshid, E. (2014). Water access points and hydration pathways in CLC  $\text{H}^+/\text{Cl}^-$  transporters. *Proc. Natl. Acad. Sci. U.S.A.* 111, 1819–1824. doi: 10.1073/pnas.1317890111
- Hehre, W. J., Ditchfield, R., and Pople, J. A. (1972). Self-consistent molecular orbital methods. XII. Further extensions of Gaussian-type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.* 56, 2257–2261. doi: 10.1063/1.1674902
- Hoover, W. G. (1985). Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* 31, 1695–1697. doi: 10.1103/PhysRevA.31.1695
- Huetz, P., Boiteux, C., Compain, M., Ramseyer, C., and Girardet, C. (2006). Incidence of partial charges on ion selectivity in potassium channels. *J. Chem. Phys.* 124:044703. doi: 10.1063/1.2159483
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graphics.* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5
- Illingworth, C. J., and Domene, C. (2009). Many-body effects and simulations of potassium channels. *Proc. R. Soc. A.* 465, 1701–1716. doi: 10.1098/rspa.2009.0014
- Illingworth, C. J. R., Furini, S., and Domene, C. (2010). Computational studies on polarization effects and selectivity in  $\text{K}^+$  channels. *J. Chem. Theory Comput.* 6, 3780–3792. doi: 10.1021/ct100276c
- Izrailev, S., Stepaniants, S., Isralewitz, B., Kosztin, D., Lu, H., Molnar, F., et al. (1998). “Steered molecular dynamics,” in *Computational Molecular Dynamics: Challenges, Methods, Ideas*, eds P. Deufelhard, J. Hermans, B. Leimkuhler, A. E. Mark, S. Reich, and R. D. Skeel. (Berlin: Springer-Verlag), 39–65.
- Jayaram, H., Accardi, A., Wu, F., Williams, C., and Miller, C. (2008). Ion permeation through a  $\text{Cl}^-$ -selective channel designed from a CLC  $\text{Cl}^-/\text{H}^+$  exchanger. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11194–11199. doi: 10.1073/pnas.0804503105
- Jentsch, T. J. (2008). CLC chloride channels and transporters: from genes to protein structure, pathology and physiology. *Crit. Rev. Biochem. Mol. Biol.* 43, 3–36. doi: 10.1080/10409230701829110
- Jentsch, T. J. (2015). Discovery of CLC transport proteins: cloning, structure, function and pathophysiology. *J. Physiol.* 593, 4091–4109. doi: 10.1113/JPhys20043
- Jiang, T., Han, W., Maduke, M., and Tajkhorshid, E. (2016). Molecular basis for differential anion binding and proton coupling in the  $\text{Cl}^-/\text{H}^+$  exchanger CLC-ec1. *J. Am. Chem. Soc.* 138, 3066–3075. doi: 10.1021/jacs.5b12062
- Jorgensen, W. L. (2007). Special issue on polarization. *J. Chem. Theory Comput.* 3, 1877–1877. doi: 10.1021/ct700252g
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869
- Kaminski, G. A., Stern, H. A., Berne, B. J., and Friesner, R. A. (2004). Development of an accurate and robust polarizable molecular mechanics force field from ab initio quantum chemistry. *J. Phys. Chem. A* 108, 621–627. doi: 10.1021/jp0301103
- Khantwal, C. M., Abraham, S. J., Han, W., Jiang, T., Chavan, T. S., Cheng, R. C., et al. (2016). Revealing an outward-facing open conformational state in a CLC  $\text{Cl}^-/\text{H}^+$  exchange transporter. *Elife* 5:e11189. doi: 10.7554/eLife.11189
- Kieseritzky, G., and Knapp, E.-W. (2011). Charge transport in the CLC-type chloride-proton anti-porter from *Escherichia coli*. *J. Biol. Chem.* 286, 2976–2986. doi: 10.1074/jbc.M110.163246
- Klauda, J. B., Venable, R. M., Freites, J. A., O’Connor, J. W., Tobias, D. J., Mondragon-Ramirez, C., et al. (2010). Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* 114, 7830–7843. doi: 10.1021/jp101759q
- Ko, Y. J., and Jo, W. H. (2010a). Chloride ion conduction without water coordination in the pore of CLC protein. *J. Comput. Chem.* 31, 603–611. doi: 10.1002/jcc.21432
- Ko, Y. J., and Jo, W. H. (2010b). Secondary water pore formation for proton transport in a CLC exchanger revealed by an atomistic molecular-dynamics simulation. *Biophys. J.* 98, 2163–2169. doi: 10.1016/j.bpj.2010.01.043
- Kreis, K., Tuckerman, M. E., Donadio, D., Kremer, K., and Potestio, R. (2016). From classical to quantum and back: a Hamiltonian scheme for adaptive multiresolution classical/path-integral simulations. *J. Chem. Theory Comput.* 12, 3030–3039. doi: 10.1021/acs.jctc.6b00242
- Kuang, Z. F., Liu, A. P., and Beck, T. L. (2008). TransPath: a computational method for locating ion transit pathways through membrane proteins. *Proteins* 71, 1349–1359. doi: 10.1002/prot.21822
- Kuang, Z., Mahankali, U., and Beck, T. L. (2007). Proton pathways and  $\text{H}^+/\text{Cl}^-$  stoichiometry in bacterial chloride transporters. *Proteins* 68, 26–33. doi: 10.1002/prot.21441
- Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., and Kollman, P. A. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. I. the method. *J. Comput. Chem.* 13, 1011–1021. doi: 10.1002/jcc.540130812
- Lee, C., Yang, W., and Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B Condens. Matter* 37, 785–789. doi: 10.1103/PhysRevB.37.785
- Lee, S., Liang, R., Voth, G. A., and Swanson, J. M. J. (2016a). Computationally efficient multiscale reactive molecular dynamics to describe amino acid deprotonation in proteins. *J. Chem. Theory Comput.* 12, 879–891. doi: 10.1021/acs.jctc.5b01109
- Lee, S., Mayes, H. B., Swanson, J. M. J., and Voth, G. A. (2016b). The origin of coupled chloride and proton transport in a  $\text{Cl}^-/\text{H}^+$  antiporter. *J. Am. Chem. Soc.* 138, 14923–14930. doi: 10.1021/jacs.6b06683
- Lee, S., Swanson, J. M. J., and Voth, G. A. (2016c). Multiscale simulations reveal key aspects of the proton transport mechanism in the CLC-ec1 antiporter. *Biophys. J.* 110, 1334–1345. doi: 10.1016/j.bpj.2016.02.014
- Lim, H.-H., and Miller, C. (2009). Intracellular proton-transfer mutants in a CLC  $\text{Cl}^-/\text{H}^+$  exchanger. *J. Gen. Physiol.* 133, 131–138. doi: 10.1085/jgp.200810112
- Lim, H.-H., Shane, T., and Miller, C. (2012). Intracellular proton access in a  $\text{Cl}^-/\text{H}^+$  antiporter. *PLoS Biol.* 10:e1001441. doi: 10.1371/journal.pbio.1001441
- Lin, C.-W., and Chen, T.-Y. (2003). Probing the pore of CLC-0 by substituted cysteine accessibility method using methane thiosulfonate reagents. *J. Gen. Physiol.* 122, 147–159. doi: 10.1085/jgp.200308845
- Lin, H., and Truhlar, D. G. (2005). Redistributed charge and dipole schemes for combined quantum mechanical and molecular mechanical calculations. *J. Phys. Chem. A* 109, 3991–4004. doi: 10.1021/jp0446332
- Lin, H., and Truhlar, D. G. (2007). QM/MM: What have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* 117, 185–199. doi: 10.1007/s00214-006-0143-z
- Lin, H., Zhang, Y., Pezeshki, S., Duster, A., and Truhlar, D. G. (2017). QMMM. Version 2.2.8.CO. Minneapolis MN: University of Minnesota.

- Lobet, S., and Dutzler, R. (2006). Ion-binding properties of the CLC chloride selectivity filter. *EMBO J.* 25, 24–33. doi: 10.1038/sj.emboj.7600909
- Löwdin, P.-O. (1950). On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals. *J. Chem. Phys.* 18, 365–375. doi: 10.1063/1.1747632
- MacKerell, A. D. Jr., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102, 3586–3616. doi: 10.1021/jp973084f
- Mackerell, A. D., Feig, M., and Brooks, C. L. (2004). Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25, 1400–1415. doi: 10.1002/jcc.20065
- Maduke, M., Miller, C., and Mindell, J. A. (2000). A decade of CLC chloride channels: structure, mechanism, and many unsettled questions. *Annu. Rev. Biophys. Biomol. Struct.* 29, 411–438. doi: 10.1146/annurev.biophys.29.1.411
- Martyna, G. J., Tobias, D. J., and Klein, M. L. (1994). Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* 101, 4177–4189. doi: 10.1063/1.467468
- Matulef, K., and Maduke, M. (2007). The CLC ‘chloride channel’ family: revelations from prokaryotes. *Mol. Membr. Biol.* 24, 342–350. doi: 10.1080/09687680701413874
- Miller, C., and Nguitragool, W. (2009). A provisional transport mechanism for a chloride channel-type  $\text{Cl}^-/\text{H}^+$  exchanger. *Phil. Trans. R. Soc. B* 364, 175–180. doi: 10.1098/rstb.2008.0138
- Miloshevsky, G. V., Hassanein, A., and Jordan, P. C. (2010). Antiport mechanism for  $\text{Cl}^-/\text{H}^+$  in CLC-ecl from normal-mode analysis. *Biophys. J.* 98, 999–1008. doi: 10.1016/j.bpj.2009.11.035
- Miloshevsky, G. V., and Jordan, P. C. (2004). Anion pathway and potential energy profiles along curvilinear bacterial CLC  $\text{Cl}^-$  pores: electrostatic effects of charged residues. *Biophys. J.* 86, 825–835. doi: 10.1016/S0006-3495(04)74158-2
- Nguitragool, W., and Miller, C. (2006). Uncoupling of a CLC  $\text{Cl}^-/\text{H}^+$  exchange transporter by polyatomic anions. *J. Mol. Biol.* 362, 682–690. doi: 10.1016/j.jmb.2006.07.006
- Nieto-Delgado, P. G., Arreola, J., and Guirado-López, R. A. (2013). Atomic charges of  $\text{Cl}^-$  ions confined in a model *Escherichia coli* CLC- $\text{Cl}^-/\text{H}^+$  ion exchanger: a density functional theory study. *Mol. Phys.* 111, 3218–3233. doi: 10.1080/00268976.2013.776709
- Patel, S., and Brooks, C. L. III. (2004). CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J. Comput. Chem.* 25, 1–15. doi: 10.1002/jcc.10355
- Patel, S., Mackerell, A. D. Jr., and Brooks, C. L. III. (2004). CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Chem.* 25, 1504–1514. doi: 10.1002/jcc.20077
- Pezeshki, S., Davis, C., Heyden, A., and Lin, H. (2014). Adaptive-partitioning QM/MM dynamics simulations: 3. Solvent molecules entering and leaving protein binding sites. *J. Chem. Theory Comput.* 10, 4765–4776. doi: 10.1021/ct500553x
- Pezeshki, S., and Lin, H. (2015). Recent developments in QM/MM methods towards open-boundary multi-scale simulations. *Mol. Simulat.* 41, 168–189. doi: 10.1080/08927022.2014.911870
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781–1802. doi: 10.1002/jcc.20289
- Picollo, A., Malvezzi, M., Houtman, J. C. D., and Accardi, A. (2009). Basis of substrate binding and conservation of selectivity in the CLC family of channels and transporters. *Nat. Struct. Mol. Biol.* 16, 1294–1301. doi: 10.1038/nsmb.1704
- Picollo, A., and Pusch, M. (2005). Chloride/proton antiporter activity of mammalian CLC proteins CLC-4 and CLC-5. *Nature* 436, 420–423. doi: 10.1038/nature03720
- Picollo, A., Xu, Y., Johner, N., Bernèche, S., and Accardi, A. (2012). Synergistic substrate binding determines the stoichiometry of transport of a prokaryotic  $\text{H}^+/\text{Cl}^-$  exchanger. *Nat. Struct. Mol. Biol.* 19, 525–531. doi: 10.1038/nsmb.2277
- Pyykkö, P., and Atsumi, M. (2009). Molecular single-bond covalent radii for elements 1–118. *Chem. Eur. J.* 15, 186–197. doi: 10.1002/chem.200800987
- Reed, A. E., Weinstock, R. B., and Weinhold, F. (1985). Natural population analysis. *J. Chem. Phys.* 83, 735–746. doi: 10.1063/1.449486
- Robertson, J. L., Kolmakova-Partensky, L., and Miller, C. (2010). Design, function and structure of a monomeric CLC transporter. *Nature* 468, 844–847. doi: 10.1038/nature09556
- Rode, B. M., Hofer, T. S., Randolph, B. R., Schwenk, C. F., Xenides, D., and Vchirawongkwin, V. (2006). Ab initio quantum mechanical charge field (QMCF) molecular dynamics: a QM/MM – MD procedure for accurate simulations of ions and complexes. *Theor. Chem. Acc.* 115, 77–85. doi: 10.1007/s00214-005-0049-1
- Roux, B. (1995). The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* 91, 275–282. doi: 10.1016/0010-4655(95)00053-I
- Roux, B., Bernèche, S., Egwolff, B., Lev, B., Noskov, S. Y., Rowley, C. N., et al. (2011). Ion selectivity in channels and transporters. *J. Gen. Physiol.* 137, 415–426. doi: 10.1085/jgp.201010577
- Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* 23, 327–341. doi: 10.1016/0021-9991(77)90098-5
- Scheel, O., Zdebik, A. A., Lourdel, S., and Jentsch, T. J. (2005). Voltage-dependent electrogenic chloride/proton exchange by endosomal CLC proteins. *Nature* 436, 424–427. doi: 10.1038/nature03860
- Senn, H. M., and Thiel, W. (2007). QM/MM methods for biological systems. *Top. Curr. Chem.* 268, 173–290. doi: 10.1007/128\_2006\_084
- Shannon, R. (1976). Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallographica Sect. A* 32, 751–767. doi: 10.1107/S0567739476001551
- Smart, O. S., Neduvilil, J. G., Wang, X., Wallace, B. A., and Sansom, M. S. P. (1996). HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graphics* 14, 354–360. doi: 10.1016/S0263-7855(97)00009-X
- Smith, M., and Lin, H. (2011). Charge delocalization upon chloride ion binding in CLC chloride ion channels/transporters. *Chem. Phys. Lett.* 502, 112–117. doi: 10.1016/j.cplett.2010.12.031
- Stewart, J. P. (1989a). Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.* 10, 209–220. doi: 10.1002/jcc.540100208
- Stewart, J. P. (1989b). Optimization of parameters for semiempirical methods II. Applications. *J. Comput. Chem.* 10, 221–264 doi: 10.1002/jcc.540100209
- Thiel, W. (2005). MNDO2005. Version, 7.0. Mühheim an der Ruhr: Max-Planck-Institut für Kohlenforschung.
- Torrie, G. M., and Valleau, J. P. (1977). Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J. Comput. Phys.* 23, 187–199. doi: 10.1016/0021-9991(77)90121-8
- Turano, B., Pear, M., and Busath, D. (1992). Gramicidin channel selectivity. Molecular mechanics calculations for formamidinium, guanidinium, and acetamidinium. *Biophys. J.* 63, 152–161. doi: 10.1016/S0006-3495(92)81574-6
- van der Kamp, M. W., and Mulholland, A. J. (2013). Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry* 52, 2708–2728. doi: 10.1021/bi400215w
- Vanommeslaeghe, K., and Mackerell, A. D. (2015). CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochim. Biophys. Acta* 1850, 861–871. doi: 10.1016/j.bbagen.2014.08.004
- Varma, S., and Rempe, S. B. (2010). Multibody effects in ion binding and selectivity. *Biophys. J.* 99, 3394–3401. doi: 10.1016/j.bpj.2010.09.019
- Vien, M., Basilio, D., Leisle, L., and Accardi, A. (2017). Probing the conformation of a conserved glutamic acid within the  $\text{Cl}^-$  pathway of a CLC  $\text{H}^+/\text{Cl}^-$  exchanger. *J. Gen. Physiol.* 149, 523–529. doi: 10.1085/jgp.201611682
- Walden, M., Accardi, A., Wu, F., Xu, C., Williams, C., and Miller, C. (2007). Uncoupling and turnover in a  $\text{Cl}^-/\text{H}^+$  exchange transporter. *J. Gen. Physiol.* 129, 317–329. doi: 10.1085/jgp.200709756
- Wang, D., and Voth, G. A. (2009). Proton transport pathway in the CLC  $\text{Cl}^-/\text{H}^+$  antiporter. *Biophys. J.* 97, 121–131. doi: 10.1016/j.bpj.2009.04.038
- Wang, J., Wu, Y., Ma, C., Fiorin, G., Wang, J., Pinto, L. H., et al. (2013). Structure and inhibition of the drug-resistant S31N mutant of the M2 ion channel of influenza A virus. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1315–1320. doi: 10.1073/pnas.1216526110
- Wang, S., Orabi, E. A., Baday, S., Bernèche, S., and Lamoureux, G. (2012). Ammonium transporters achieve charge transfer by fragmenting their substrate. *J. Am. Chem. Soc.* 134, 10419–10427. doi: 10.1021/ja300129x

- Warshel, A., Kato, M., and Pisliakov, A. V. (2007). Polarizable force fields: history, test cases, and prospects. *J. Chem. Theory Comput.* 3, 2034–2045. doi: 10.1021/ct700127w
- Warshel, A., and Levitt, M. (1976). Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* 103, 227–249. doi: 10.1016/0022-2836(76)90311-9
- White, M. M., and Miller, C. (1981). Probes of the conduction process of a voltage-gated  $\text{Cl}^-$  channel from *Torpedo* electroplax. *J. Gen. Physiol.* 78, 1–18. doi: 10.1085/jgp.78.1.1
- Xie, W., and Gao, J. (2007). Design of a next generation force field: the X-POL potential. *J. Chem. Theory Comput.* 3, 1890–1900. doi: 10.1021/ct700167b
- Yin, J., Kuang, Z., Mahankali, U., and Beck, T. L. (2004). Ion transit pathways and gating in ClC chloride channels. *Proteins* 57, 414–421. doi: 10.1002/prot.20208
- Yu, T., Wang, X.-Q., Sang, J.-P., Pan, C.-X., Zou, X.-W., Chen, T.-Y., et al. (2012). Influences of mutations on the electrostatic binding free energies of chloride ions in *Escherichia coli* ClC. *J. Phys. Chem. B* 116, 6431–6438. doi: 10.1021/jp300430f
- Zdebik, A. A., Zifarelli, G., Bergsdorf, E.-Y., Soliani, P., Scheel, O., Jentsch, T. J., et al. (2008). Determinants of anion-proton coupling in mammalian endosomal CLC proteins. *J. Biol. Chem.* 283, 4219–4227. doi: 10.1074/jbc.M708368200
- Zhang, Y., and Voth G. A. (2011). The coupled proton transport in the ClC-ec1  $\text{Cl}^-/\text{H}^+$  antiporter. *Biophys. J.* 101, L47–L49. doi: 10.1016/j.bpj.2011.10.021
- Zhang, Y., Lee, T.-S., and Yang, W. (1999). A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *J. Chem. Phys.* 110, 46–54. doi: 10.1063/1.478083

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Wang, Duster, Aydintug, Zarecki and Lin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Modeling Chemical Reactions by QM/MM Calculations: The Case of the Tautomerization in Fireflies Bioluminescent Systems

Romain Beraud-Pache, Cristina Garcia-Iriepea and Isabelle Navizet\*

Université Paris-Est, Laboratoire Modélisation et Simulation Multi Echelle, MSME, UMR 8208 CNRS, UPEM, Marne-la-Vallée, France

## OPEN ACCESS

### Edited by:

Sam P. De Visser,  
University of Manchester,  
United Kingdom

### Reviewed by:

José Pedro Cerón-Carrasco,  
Universidad Católica San Antonio de  
Murcia, Spain  
Xavier Assfeld,  
Université de Lorraine, France

### \*Correspondence:

Isabelle Navizet  
isabelle.navizet@u-pem.fr

### Specialty section:

This article was submitted to  
Theoretical and Computational  
Chemistry,  
a section of the journal  
*Frontiers in Chemistry*

Received: 31 January 2018

Accepted: 29 March 2018

Published: 17 April 2018

### Citation:

Beraud-Pache R, Garcia-Iriepea C and  
Navizet I (2018) Modeling Chemical  
Reactions by QM/MM Calculations:  
The Case of the Tautomerization in  
Fireflies Bioluminescent Systems.  
*Front. Chem.* 6:116.  
doi: 10.3389/fchem.2018.00116

In less than half a century, the hybrid QM/MM method has become one of the most used technique to model molecules embedded in a complex environment. A well-known application of the QM/MM method is for biological systems. Nowadays, one can understand how enzymatic reactions work or compute spectroscopic properties, like the wavelength of emission. Here, we have tackled the issue of modeling chemical reactions inside proteins. We have studied a bioluminescent system, fireflies, and deciphered if a keto-enol tautomerization is possible inside the protein. The two tautomers are candidates to be the emissive molecule of the bioluminescence but no outcome has been reached. One hypothesis is to consider a possible keto-enol tautomerization to treat this issue, as it has been already observed in water. A joint approach combining extensive MD simulations as well as computation of key intermediates like TS using QM/MM calculations is presented in this publication. We also emphasize the procedure and difficulties met during this approach in order to give a guide for this kind of chemical reactions using QM/MM methods.

**Keywords:** oxyluciferin, TD-DFT, molecular dynamics, QM/MM, keto-enol tautomerization, emission spectra, bioluminescence

## INTRODUCTION

Chemical reactions are ubiquitous in biology, controlling a wide variety of fundamental biological processes such as photosynthesis, the process of vision, bioluminescence, among others (Hales, 1976; Adam, 1982; Metzler and Metzler, 2001; Palczewski, 2012). In this regard, a full understanding of the underlined chemical mechanisms is crucial. Only then, the factors that govern the specificity and efficiency of these biological processes can be determined.

In order to get insight into the mechanism and intermediates, some experimental methodologies can be used such as spectroscopic techniques, mutagenesis experiments and kinetic evaluations (Meister, 1983; Zscherp and Barth, 2001; Frey and Hegeman, 2007). However, in general many chemical questions remain unanswered from an experimental point of view and hence, the modeling of the system and the mechanism by computational chemistry is needed (Becker et al., 2001; Mulholland, 2005; van der Kamp et al., 2008).

Regarding to the modeling, quantum mechanics/molecular mechanics (QM/MM) methods are the state-of-the-art computational techniques to study chemical reactions and to compute electronic properties in complex environments as biomolecular systems (Senn and Thiel, 2007, 2009; Acevedo and Jorgensen, 2010; van der Kamp and Mulholland, 2013; Sousa et al., 2017). In QM/MM methods, the chemically active site (i.e., where the chemical reaction takes place or the molecule whose properties are going to be calculated) is treated at the QM level whereas the protein surroundings and the explicit solvent molecules are treated at the MM level. Many possibilities arise depending on the QM and MM methods used and on the QM/MM interface (Senn and Thiel, 2009). Up to now, QM/MM methods have been applied to get insight into different biological issues: (i) enzymatic reaction mechanisms (Friesner and Guallar, 2005; Senn and Thiel, 2007; Acevedo and Jorgensen, 2010), (ii) the calculation of spectroscopic properties (Gascón et al., 2005; Sabin et al., 2010; Gattuso et al., 2017), (iii) the investigation of electronically excited states (Navizet et al., 2010; Sabin et al., 2010; García-Iriepa et al., 2016; Gozem et al., 2017) or (iv) the calculation of pKa values (Jensen et al., 2005; Riccardi et al., 2006). It should be stressed that through QM/MM methods, transition state structures, intermediates and activation energies can be computed, being central for reactivity (Gao et al., 2006; Ramos and Fernandes, 2008; Lonsdale et al., 2012). For this reason, QM/MM methods have become in the last decades an essential tool to get insight into the mechanism of biochemical reactions and hence, to confirm or discard different mechanistic proposals which cannot be elucidated from experimental data.

The multiple strengths of QM/MM methods are that large molecules can be modeled including explicitly the entire system in the calculations, they balance the simulation cost and accuracy and they complement experimental data. On the contrary the main limitation of them is the system setup prior the QM/MM calculations, as there are many possible conformations that the macromolecule can assume, also considering the presence of solvent. However, available experimental data, such as crystallographic structures or physiological measurements, ease the setup preparation.

In this work we present an application of QM/MM methods to study the bioluminescent system of fireflies, which basis are still not fully understood. In particular, the chemical nature of the emissive species, so-called oxyluciferin, is still under debate as it can assume six different chemical forms due to triple equilibrium of phenol and/or enol deprotonation together with a keto-enol tautomerization (Hirano et al., 2009; Navizet et al.,

2010; Hosseinkhani, 2011). The aim of this work is to elucidate if keto-enol tautomerization (**Scheme 1**) of phenolate-oxyluciferin (OxyLH<sup>-</sup>) is feasible, both in the ground and in the excited state, inside the active site of the fireflies' luciferase enzyme. For this aim, MD simulations have been performed to sample different possible conformations combined with QM/MM calculations to find the transition states and qualitatively evaluate the energy barriers.

## METHODS

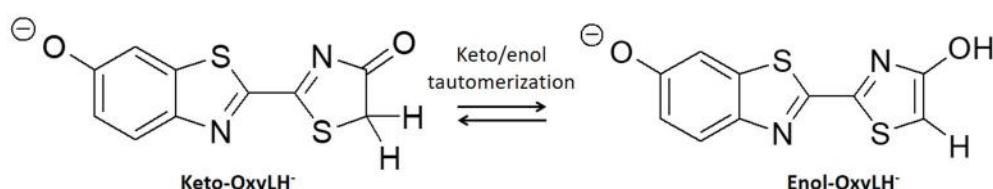
### Model Setup

The crystal structure used in the present publication comes from a north American firefly chemically engineered luciferase, obtained by the group of Prof. Branchini (Branchini et al., 2011; Sundlov et al., 2012) (PDB 4G37). It has been designed to reproduce the conformation of the protein when the dioxygen binds to an intermediate of the bioluminescent reaction. This conformation was obtained thanks to a disulfide bridge between two residues, Isoleucine 108 and Tyrosine 447. Moreover, it contains the 5'-O-[N-(dehydroluciferyl)- sulfamoyl]adenosine (DLSA) (**Scheme 2A**) substrate.

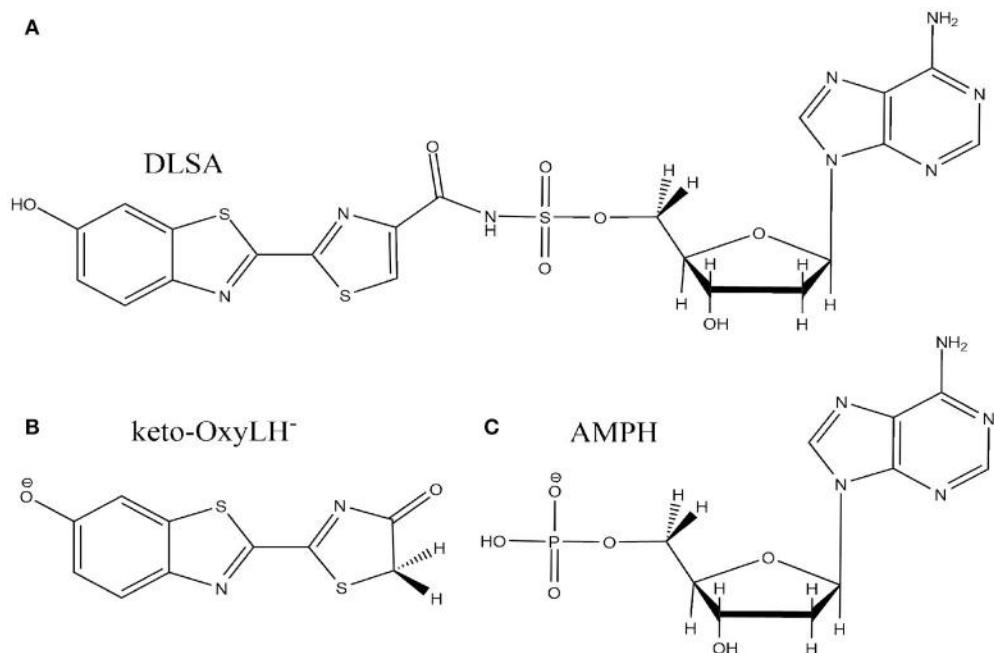
The missing loops in the crystallographic structure 4G37.pdb downloaded from the RSCB PDB website were added with Disgro program (Tang et al., 2014). Then, the DLSA residue was replaced by protonated adenosine monophosphate AMPH, charged -1, and the phenolate-keto form of oxyluciferin, keto-OxyLH<sup>-</sup> (**Schemes 2B,C**). The two SO<sub>4</sub><sup>2-</sup> groups and the disulfide bridging residue were removed. The two SO<sub>4</sub><sup>2-</sup> groups are used as precipitant for the protein (and are not key to model the protein), and were deleted. The S-S cross-linked molecule has also been removed because no accurate parameters are available. No structural changes between the domains of the luciferase were observed during the MD. During the simulations, the protein conformation remains similar to the one of the crystallographic structure. Finally, and in order to neutralize the system, we protonated some histidine residues using the H++ program (Anandakrishnan et al., 2012), keeping unchanged the ones close to the protein active site. In this case, the following histidines were protonated (76, 171, 310, 332, 419, 461 and 489), each one yielding a +1 charge.

### MD Simulations

Classical dynamics simulations were made with Amber14 program (Case et al., 2014). The AMBER99ff force field was used to model the residues of the protein. The parameters used for



**SCHEME 1** | Keto-enol tautomerization of phenolate-oxyluciferin (OxyLH<sup>-</sup>) under study.



**SCHEME 2** | Representation of the DLSA molecule (A), the phenolate-keto oxyluciferin form (keto-OxyLH<sup>-</sup>) (B) and AMPH (C).

both OxyLH<sup>-</sup> forms (keto or enol) and AMPH were designed by our group (Navizet et al., 2010, 2011, 2013). The model was solvated with TIP3P (Jorgensen et al., 1983) water molecules within a cube box, ensuring a solvent shell of at least 15 Å around the solute. The resulting system contained roughly 28,000 water molecules and 90,000 atoms in total. The system was heated from 100 to 300 K in 20 ps. Then, under NPT conditions with T = 300 K and P = 1 atm, 21 molecular dynamics of 10 ns using periodic boundary conditions were performed with a 2 fs time step. During these simulations, pressure and temperature were maintained using the Berendsen algorithm and SHAKE constraints were applied to all bonds involving hydrogen atoms (Ryckaert et al., 1977).

### QM/MM Setup

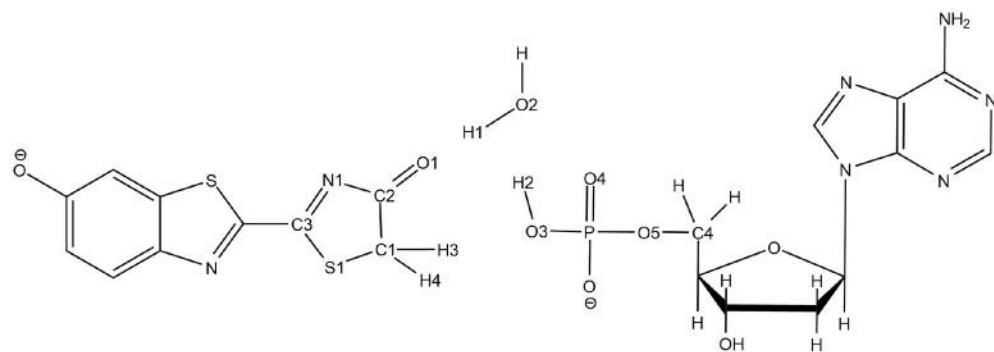
The QM/MM calculations have been carried out using a QM/MM coupling scheme (Ferré and Ángyán, 2002) between Gaussian09 (G09 D.01) (Frisch et al., 2010) and Tinker (Ponder, 2004). In particular, the interaction between the QM charge density (electrons and nuclei) and the external electrostatic potential of the MM part was computed by the electrostatic potential fitted (ESPF) method (Ferré and Ángyán, 2002). The microiterations technique (Melaccio et al., 2011) was used to converge the MM subsystem geometry for every QM minimization step. The QM part is composed of the OxyLH<sup>-</sup>, either in the keto or in the enol form, the AMPH and the water molecule n°559 (hereafter named Wat1), placed between the oxyluciferin and AMPH. The rest of the system, that is the protein and the water molecules, are included in the MM part.

QM/MM calculations were used for searching of the transition states (TS) in the protein, defining energetic profiles bridging the

TS to the keto and enol forms, and calculating the electronic transitions between the first singlet excited state (S1) and the ground state (GS). For TS and energetic profiles search, unrestricted with broken-symmetry and restricted DFT and TD-DFT calculations were performed using the M06-2X functional (Zhao and Truhlar, 2008) which includes a dispersive term and provides accurate results when studying chemical reactions (Chéron et al., 2012). Unrestricted and restricted calculations gave similar results in terms of energy and geometries, of about 0.12 eV for the transition state (TS), and about 0.15 eV for energetics profiles. Therefore restricted results are presented in this publication. TD-DFT calculations were performed using 3 roots. The emission energy (T<sub>e</sub>) between the first singlet excited state (S1) and the ground state (GS), from geometries obtained in the S1 state, were computed using both the M06-2X and the B3LYP (Stephens et al., 1994) functionals. The B3LYP functional is known to give emission energy values close to experiment for fireflies (Beraud-Pache and Navizet, 2016). Finally, the basis set used is the 6-311G(2d,p) as used in previous publications (Laurent et al., 2015; Beraud-Pache and Navizet, 2016).

### Transition State (TS) Search and Energetic Profiles

Once we have selected the MD snapshot for the QM/MM calculations, we can study the keto-enol tautomerization mechanism. Oxyluciferin, AMPH and Wat1 are included in the QM part (Scheme 3) whereas the protein and the other water molecules are treated at the MM level. To study the desired reaction, first the TS connecting the keto and enol forms of oxyluciferin has to be located. However, searching



**SCHEME 3** | Representation of the QM part extracted from the snapshot and atoms numbering of the relevant atoms.

the TS structure inside the protein by QM/MM methods is a complicated task due to both the complexity of the system and the non-availability of certain optimization algorithms [e.g., the Synchronous Transit-Guided Quasi-Newton (STQN) algorithm (Peng and Schlegel, 1993)] in G09/Tinker.

To solve this issue, the QM part selected for QM/MM calculations has been extracted (**Scheme 3**) and then we have first computed the TS in the GS, without including the protein surrounding but considering water as an implicit solvent by a Polarizable Continuum Model (PCM) (Cancès et al., 1997; Mennucci et al., 1997) of water. Then, the STQN algorithm is used, for which both the keto and the enol structures have to be provided. In our case, the keto form is the one extracted from the selected MD snapshot whereas the enol one is built based on the keto form, just by moving manually H1 to O1, H2 to O2, and H3 to O3 (see **Scheme 3** for atoms numbering).

To refine the TS structure obtained with the STQN algorithm, a Berny optimization has been performed. During the Berny optimization, all the non-involved atoms in the tautomerization reaction are blocked (see Supplementary Figure 1). In detail, it corresponds to the benzothiazole cycle of  $\text{OxyLH}^-$ , the ribose and the adenosine part of AMPH. A frequency calculation was then performed to validate the found structure as TS, named  $\text{TS}_{\text{GS}}^{\text{PCM}}$  (see Note 1 in Supplementary Material). Then, the TS in the first excited state, named  $\text{TS}_{\text{S1}}^{\text{PCM}}$  was computed while considering 3 excited electronic states, starting from the  $\text{TS}_{\text{GS}}^{\text{PCM}}$  structure.

The keto-enol tautomerization was then studied inside the protein using QM/MM calculations.  $\text{TS}_{\text{GS}}^{\text{PCM}}$  structure was put in the MD snapshot to replace the QM part coordinates. Then, a Berny optimization has been performed to obtain the transition state in the ground state in the protein,  $\text{TS}_{\text{GS}}^{\text{prot}}$ . Then search of TS in the first excited state in protein starting from  $\text{TS}_{\text{GS}}^{\text{prot}}$  was performed to obtain  $\text{TS}_{\text{S1}}^{\text{prot}}$ .

To further validate the TS structures, intrinsic reaction coordinate (IRC) calculations have been performed finding that these structures actually connect the keto and the enol forms of oxyluciferin. The IRC has been done for both PCM and QM/MM calculations using the algorithm local quadratic approximation (LQA) (Page and McIver, 1988; Page et al., 1990)

and a step of 0.01 Bohr (see Note 2 in Supplementary Material and Supplementary Figure 2).

In the IRC graphs shown in the rest of the publication, the point at the reaction coordinate (RC) equal 0 corresponds to the energy of the relaxed TS structure. Indeed, because some of the atoms are frozen during the Berny optimization, the first step of the IRC shows a big decrease of the energy corresponding to the removing of the constraints of all the atoms of the QM part. However, no geometrical differences are observed between the TS geometry and the relaxed one.

We have also performed a QM/MM dihedral scan of the phosphate group of the AMPH substrate inside the protein at the GS to link the keto- $\text{OxyLH}^-$  structure form the MD and the one obtained after the IRC search. This scan has been performed along the O4-P-O5-C4 atoms in GS using 10 steps of  $-10^\circ$  while the ribose, the adenosine and the  $\text{OxyLH}^-$  were kept frozen.

## RESULTS AND DISCUSSION

In the present paper, we aim to study the keto-enol tautomerization of oxyluciferin inside the luciferase protein both in the ground state (GS) and in the first singlet excited state (S1). In fireflies, it is well accepted that the final product of the biochemical reaction corresponds to the keto form of oxyluciferin (Liu et al., 2009). However, some experimental measurements (Naumov et al., 2009; Snellenburg et al., 2016) have deciphered the presence of the enol form inside the protein after the bioluminescence emission. The keto-enol tautomerization of oxyluciferin takes place inside the protein. By computing both the GS and S1 energetic profiles of the tautomerization reaction, we can also provide insights to explain if the reaction takes place before or after the emission. The first step is to describe correctly the proteic environment (this has been done by constructing some models from classical molecular dynamics (MD) simulations). We then have to find the geometry of the TSs joining the keto and the enol forms of the oxyluciferin in the proteic environment. As the search of TSs inside the protein, using QM/MM calculations, is complicated we have initially pictured the TSs using a simpler environment, implicit solvation with PCM. Once TSs were located inside the protein,

energetic profiles connecting the keto and enol forms were computed.

## Classical MD Simulation

First, the system was built starting from the 4G37 luciferase structure as detailed in the Methods section. In this case, the phenolate keto form of oxyluciferin and the AMPH were selected for study.

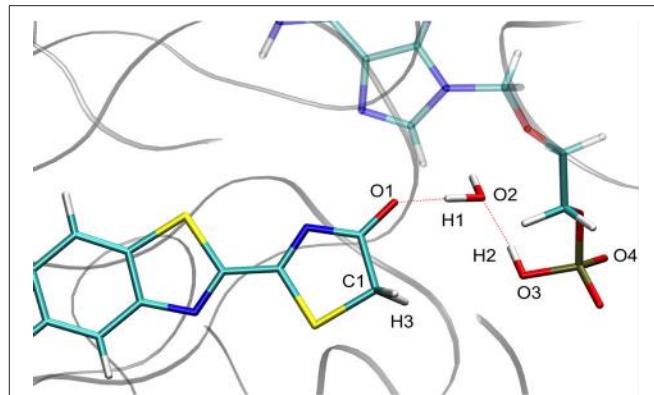
As demonstrated in previous studies (Navizet et al., 2010; Beraud-Pache and Navizet, 2016), the use of MD simulations is mandatory to investigate the luciferase-oxyluciferin system. MD simulations permit to equilibrate the system, especially important when water molecules are near or in the enzymatic cavity and could lead to a hydrogen-bond network between the substrates and residues of the protein. 21 MD simulations each lasting 10 ns, were performed and thereafter numbered from MD1 to MD21. In all the simulations, we have observed the displacement of water molecules inside the cavity of luciferase. Particularly, in 5 out of the 21 MD simulations, a water molecule (thereafter called Wat1) positions between the keto moiety of oxyluciferin and AMPH. This conformation is stable along the simulation time as a hydrogen-bond network is created involving Wat1, oxyluciferin, AMPH and other protein residues, like LYS 443. By analyzing the MD simulations, we have observed that Wat1 comes close to oxyluciferin at different simulation times depending on the dynamic. For example, Wat1 comes close to oxyluciferin at around 1 ns in MD1, whereas it takes more than 6 ns to reach this position in MD17.

The position of Wat1 suggests that this water molecule could be involved in the keto-enol tautomerization mechanism of oxyluciferin inside the protein. This way, the tautomerization could take place by the displacement of 3 protons: one from oxyluciferin, one from Wat1 and one from AMPH. We therefore decided to study the feasibility of this mechanism both in the GS and in S1 inside the protein.

For calculation purposes, we selected one representative snapshot corresponding to the lowest structure in energy found along the MD (snapshot taken at 3.708 ns from MD1). In this snapshot a water molecule Wat1 was found between oxyluciferin and AMPH. By analyzing in more detail the structure of this snapshot (Figure 1), we observe that the nearby residues do not share any hydrogen-bond with the substrates or Wat1. Only the hydrogen-bond network between oxyluciferin, AMPH and Wat1 is present. Thus, the choice of this snapshot and hence of this hydrogen-bond network was motivated to simplify the system, preventing the inclusion of more residues in the QM part, which would have increased the computational cost.

## Description of the TS in Implicit Solvent

The optimized TS structure in the GS computed in PCM ( $TS_{GS}^{PCM}$ ) corresponds to a 3 protons transfer reaction between 4 centers: O1, O2, O3 and C1 (see Supplementary Figure 3). Moreover, the  $TS_{GS}^{PCM}$  is structurally more similar to the keto form of oxyluciferin than to the enol form. For instance, H1 is closer to O2 of Wat1 (0.98 Å) than to O1 of oxyluciferin and H2 is closer to O3 of AMPH (1.03 Å) than to O2 of Wat1.



**FIGURE 1** | Representation of the MD snapshot selected as starting structure for QM/MM calculations. The oxyluciferin is represented in its keto form.

However, for the TS optimized in S1 in PCM ( $TS_{S1}^{PCM}$ ), the position of O4 leads to a 5 centers TS, including O1, O2, O3, O4, and C1. It allows a bigger flexibility for the atoms and for the proton transfer. In  $TS_{S1}^{PCM}$  (see Supplementary Figure 4), H1 is closer to O1 (1.01 Å) and H3 is still located between C1 and O3. The notable difference with the  $TS_{GS}^{PCM}$  is the presence of O4, another oxygen of the AMPH substrate, which is involved in the tautomerization reaction. Here the H2 is not located between O2 and O3 as in  $TS_{GS}^{PCM}$  but between O2 and O4. During the optimization of  $TS_{S1}^{PCM}$ , the phosphate group of the AMPH is rotated about 30°, bringing the O4 close to the water molecule. Some details about the IRC calculations with implicit solvent can be found in note 3 and Supplementary Figure 5.

In conclusion, when computing the TS in PCM both in the GS and in S1 ( $TS_{GS}^{PCM}$  and  $TS_{S1}^{PCM}$ ), two different TS geometries are obtained. The tautomerization mechanism concerning  $TS_{GS}^{PCM}$  is a 4 centers reaction while for  $TS_{S1}^{PCM}$ , 5 centers are involved. This study shows that a TS connecting the keto and enol forms of oxyluciferin can be obtained with PCM. The computed geometry of  $TS_{GS}^{PCM}$  with PCM has served as starting guess for the calculation of the TS in GS inside the protein at the QM/MM level, which is the main goal of the present study.

## Keto-enol Tautomerization Inside the Protein

We have decided to use  $TS_{GS}^{PCM}$ , a 4 centers TS, as starting structure for the QM/MM calculations inside the protein. Indeed, in all the computed MD Wat1 adopts a position that suggests a 4 centers TS, as seen in Scheme 3 and Figure 1.

The TS optimization done inside the protein in the GS gives some unexpected results. First, two different TS geometries can be obtained, both corresponding to the keto-enol tautomerization. In the first one, the position of the 3 protons suggests that the TS is derived from the enol tautomer of the oxyluciferin. It is thus named  $TS_{GS\_enol}^{prot}$ . In the same way, the second TS, named  $TS_{GS\_keto}^{prot}$ , shows protons close to the position adopted in the keto form. It should be remarked here that the  $TS_{GS\_enol}^{prot}$  lies *ca.* 34 kcal/mol above in energy compared to  $TS_{GS\_keto}^{prot}$  in the GS.

The second unexpected result concerns the geometries of all computed TSs. When using QM/MM calculations inside the protein, only a 5 centers tautomerization is observed in the GS even when starting with the 4 centers  $TS_{GS}^{PCM}$ . Within the protein active site, the substrates prefer to fill more space when performing the keto-enol tautomerization. All other attempts to compute a 4 centers TS in the GS failed, giving in all cases a 5 centers one (see **Scheme 4**).

Finally, the TSs in S1 have been located starting from both the structure of  $TS_{GS\_keto}^{prot}$  and of  $TS_{GS\_enol}^{prot}$ . As in the GS, two different TS structures have been found in S1, one structurally closer to the keto form,  $TS_{S1\_keto}^{prot}$ , and the other one closer to the enol form,  $TS_{S1\_enol}^{prot}$ . The  $TS_{S1\_keto}^{prot}$  is also lower in energy compared to  $TS_{S1\_enol}^{prot}$ . These differences are detailed in the Keto-enol tautomerization subsection.

## Geometrical Description of the TSs

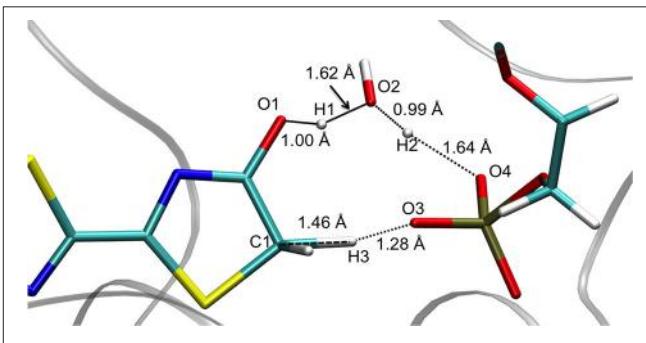
The  $TS_{enol}^{prot}$  structures obtained in both GS and S1 are quite similar. They can be described by the transfer of 3 protons (H1, H2, and H3) between 5 centers (O1, O2, O3, O4, and C1) (see **Figure 2**). In particular, H1 is placed between O1 (oxyluciferin) and O2 (Wat1) but it is closer to O1 (1.00 Å) and so, resembles the enol form. Moreover, H2 is located close to O2 (Wat1) (0.99 Å) while H3 is situated between C1 and O3 (d<sub>H3-C1</sub> = 1.46 Å and d<sub>H3-O3</sub> = 1.28 Å).

Concerning the  $TS_{keto}^{prot}$ , the structures computed in the GS and in S1 are also quite similar, corresponding to the transfer of 3 protons involving 5 centers (see **Figure 3**). However, in these cases H1 is closer to O2 (Wat1) (1.03 Å) than to O1 (oxyluciferin) (1.48 Å). Moreover, H2 is located close to O4 (AMPH) (1.04 Å) and H3 is nearly bonded to O3 (AMPH) (1.11 Å).

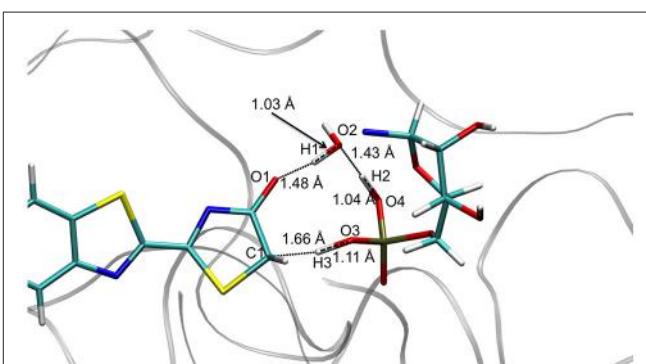
## Analysis of the Energetic Profiles of the Two TSs

After the optimization of the TSs, we computed the IRCs to check if they really connect the keto and enol forms of oxyluciferin both in the GS and in S1. The results presented in **Figures 4, 5** correspond to a superposition of the IRC computed in the GS and in S1 for a better comparison. However, reader has to have in

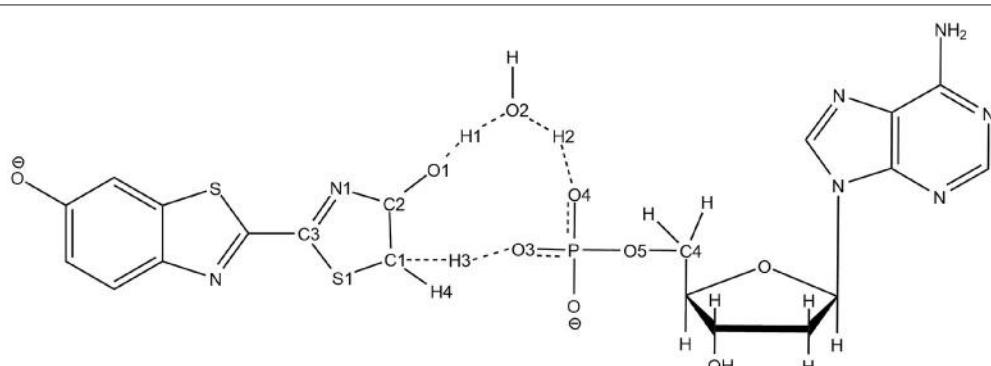
mind that the reaction coordinates of the IRCs in the GS and in S1 are not the same. Moreover, the TS structure has been defined as the 0 value of the reaction coordinate. When moving to the right toward positive values of the reaction coordinate, the keto tautomer is formed. Similarly, negative values of the reaction coordinate correspond to the formation of the enol form. The energy of the keto tautomer in the GS was considered as the reference.



**FIGURE 2** | Structure of  $TS_{S1\_enol}^{prot}$  computed at the QM/MM level.



**FIGURE 3** | Structure of  $TS_{S1\_keto}^{prot}$  computed at the QM/MM level.



**SCHEME 4** | Representation of the 5 centers mechanism for keto-enol tautomerization inside the protein.

## Energetic Profile of the TS\_enol

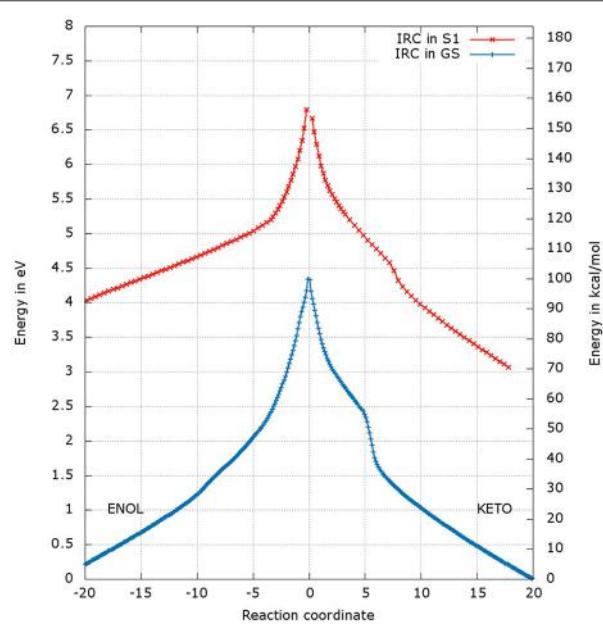
The shapes of the energetic profiles starting from  $TS_{GS\_enol}^{\text{prot}}$  and  $TS_{S1\_enol}^{\text{prot}}$  in the GS and in S1, respectively, are similar (see **Figure 4**). Moreover, it is observed that, the keto form is more

stable than the enol form of about 5 kcal/mol in the GS and 11 kcal/mol in S1.

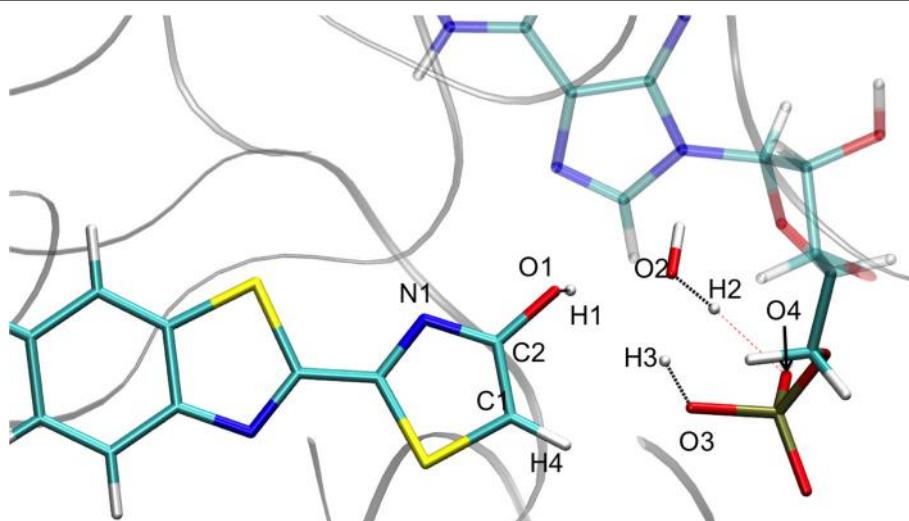
When following the energetic curve in the direction of the enol tautomer (i.e., negative values of reaction coordinate RC), we observed a steep decrease of the energy followed by a smaller slope. The steep gradient corresponds mainly to the movement of the proton H3, as the protons H1 and H2 are already in the final position of the enol form in the  $TS_{enol}^{\text{prot}}$ , both in the GS and in S1. Thus, at about RC equal  $-4$ , the enol tautomer is already formed. From this point, the system stabilization is due to a rearrangement of oxyluciferin and AMPH leading to the final enol structure (**Figure 5**). Finally, the energy barrier for tautomerization starting from the enol form found is 100 kcal/mol for the GS and 66 kcal/mol for the S1.

On the side of the curve between the TS and the keto tautomer (i.e., positive values of reaction coordinate), the profile can also be divided into two steps. In the GS curve, between RC 0 and RC 5, the proton H3 is moving in direction of C1. The shoulder at RC equal 5 corresponds to the break of the bonds O1-H1 and O2-H2. From this point, the system minimization can be described by the movement of these two protons H1 and H2. The final structure is represented in **Figure 6**. The energetic barrier for tautomerization starting from the keto form is of 105 kcal/mol for the GS and 77 kcal/mol for the S1.

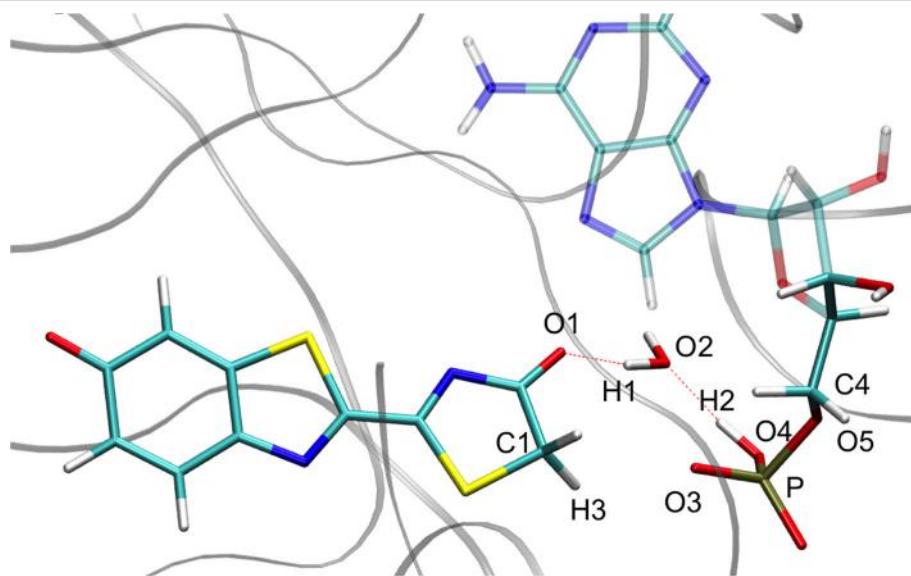
Finally, in order to verify that the structures obtained at the end of the IRC calculations correspond to the keto and enol tautomers (**Figures 5, 6**), we have computed their electronic transition energies ( $T_e$ ) between S1 and the GS (respectively  $T_{e\_keto\_IRC}^{\text{prot}}$  and  $T_{e\_enol\_IRC}^{\text{prot}}$ ). The electronic transition between S0 and S1 corresponds to a  $\pi-\pi^*$  transition with a small charge transfer between the benzothiazole ring and the thiazolone ring for both tautomers. Then, the computed energies have been compared to reference values ( $T_{e\_keto\_ref}^{\text{prot}}$  and  $T_{e\_enol\_ref}^{\text{prot}}$ ) obtained



**FIGURE 4** | Energetic profile of the keto-enol tautomerization starting from  $TS_{GS\_enol}^{\text{prot}}$  and  $TS_{S1\_enol}^{\text{prot}}$  at the reaction coordinate RC = 0. Superposition of GS in blue and S1 profiles in red. Positive RC values lead to the keto form while the negative ones to the enol form. The point at 0 eV corresponds to the lowest point of the energy profile that is to the keto form in the GS. The graph is a superposition of two graphs: reaction coordinates of the IRCs in the GS and in S1 are not the same.



**FIGURE 5** | Representation of the enol tautomer obtained from the IRC calculation starting from  $TS_{S1\_enol}^{\text{prot}}$  in S1 inside the protein. The structure is similar to the one starting from the  $TS_{S1\_keto}^{\text{prot}}$ .



**FIGURE 6 |** Representation of the keto tautomer obtained from the IRC calculation starting from  $TS_{S1\_enol}^{\text{prot}}$  in S1 inside the protein. The structure is similar to the one starting from the  $TS_{S1\_keto}^{\text{prot}}$ .

directly from the MD snapshots of these oxyluciferin forms (see Note 4 in Supplementary Material).  $T_e$  have been computed with both, the M06-2X and the B3LYP functionals, the last one giving emission energy values closer to experiment (Beraud-Pache and Navizet, 2016) (see Table 1). The computed electronic transition energies of the keto and enol forms obtained by the IRC show good agreement with reference values. The difference of energy is less than 0.1 eV, which corresponds to the DFT level of incertitude.

### Energetic Profile of the TS\_keto

The energetic profiles found in the GS and in S1 starting from  $TS_{GS\_keto}^{\text{prot}}$  and  $TS_{S1\_keto}^{\text{prot}}$ , respectively, are similar (see Figure 7). The keto tautomer is again more stable than the enol one, of about 7 kcal/mol in the GS and 15 kcal/mol in S1. Concerning the formation of the enol form (negative values of RC), H3 moves first, corresponding to the shoulder observed at RC -7 in the GS and RC -9 in S1. Then, the bonds H1-O2 and H2-O3 break and the two protons H1 and H2 move toward O1 and O2 respectively, leading to the enol tautomer. The computed energetic barriers for the tautomerization starting from the enol form are 58 kcal/mol in the GS and 51 kcal/mol in S1.

For the formation of the keto form, positive RC values, a steep gradient is observed that corresponds to the movement of H3. The calculated energetic barriers from the keto form to the TS are 65 kcal/mol in the GS and 67 kcal/mol in S1.

Finally, the  $T_e$  have also been computed for the obtained structures reached by the IRC (see Figures 5, 6), starting from  $TS_{S1\_keto}^{\text{prot}}$  (see Table 2). The electronic transition between S0 and S1 corresponds to a  $\pi-\pi^*$  transition with a small charge transfer between the benzothiazole ring and the thiazolone ring for both tautomers. The computed values are again in good agreement

**TABLE 1 |** Electronic transition energies between S1 and GS ( $T_e$ ) in the protein at the TD-DFT/MM level for the resulting structures of the IRC, starting with  $TS_{S1\_enol}^{\text{prot}}$ .

Structure obtained in S1	M06-2X $T_e$ eV (nm)	B3LYP $T_e$ eV (nm)	
keto-OxyLH <sup>-</sup>	Last point of the profile: $T_{e\_keto\_IRC}^{\text{prot}}$	2.65 (472)	2.34 (529)
	$T_{e\_keto\_ref}^{\text{prot}}$	2.55 (486) <sup>a</sup>	2.24 (555) <sup>a</sup>
enol-OxyLH <sup>-</sup>	Last point of the profile: $T_{e\_enol\_IRC}^{\text{prot}}$	2.83 (439)	2.52 (492)
	$T_{e\_enol\_ref}^{\text{prot}}$	2.86 (434) <sup>a</sup>	2.47 (502) <sup>a</sup>

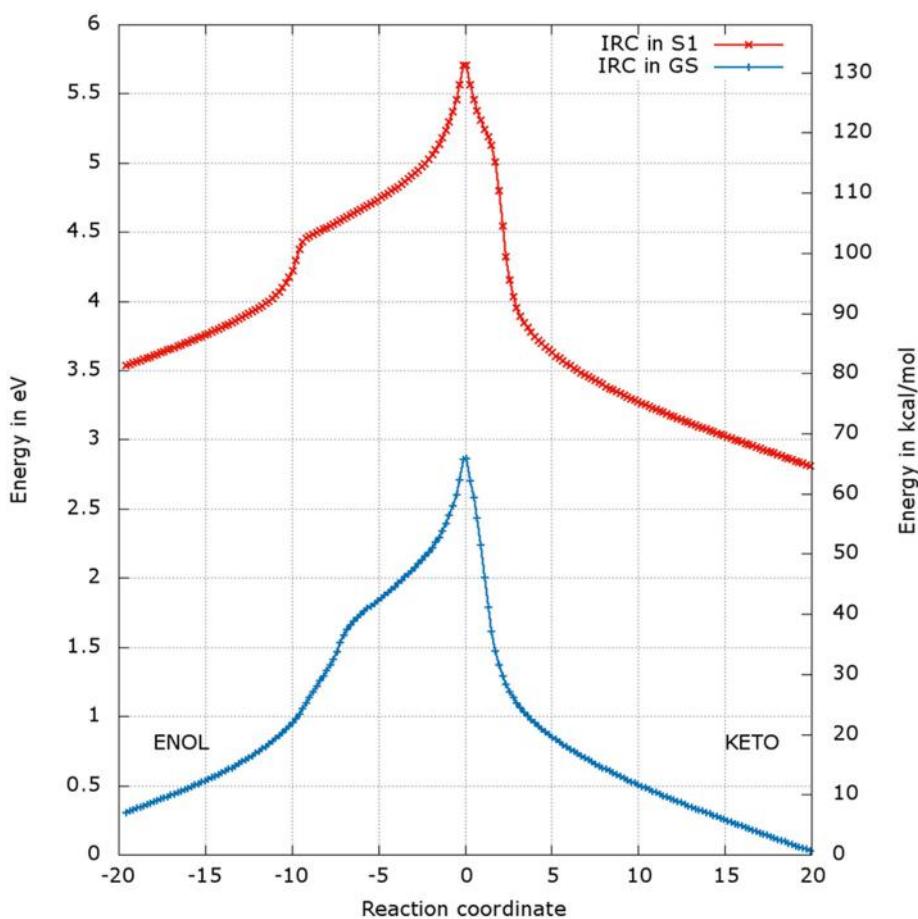
<sup>a</sup>Data in S1.

with previous results (see Note 4 in Supplementary Material), being the energy differences between 0.09 and 0.13 eV.

### Rotation

During the study of the tautomerization inside the protein, we find that the reaction involves 5 centers. Thus, the keto form from the MD snapshot (Figures 1, 8a) does not match with the one obtained by the IRC calculations in the GS starting from both, the  $TS_{GS\_keto}^{\text{prot}}$  or the  $TS_{GS\_enol}^{\text{prot}}$  (Figures 6, 8b). Indeed, by construction of the AMPH classical structure, the proton H2 is bound to the oxygen O3 in all MD. In the 5 centers reaction inside the protein leading to the keto form, the O4 from the AMPH retrieve the proton H2. The resulting keto system shows therefore a protonated AMPH on its O4 oxygen.

We have performed a dihedral scan to see the energy needed to allow the rotation of the phosphate group. We have selected the last point from the IRC calculation and



**FIGURE 7** | Energetic profile of the keto-enol tautomerization starting from the  $TS_{GS\_keto}^{\text{prot}}$  and  $TS_{S1\_keto}^{\text{prot}}$  at the reaction coordinate  $RC = 0$ . Superposition of GS in blue and S1 profiles in red. Positive RC values lead to the keto form while negative ones to the enol form. The point at 0 eV corresponds to the lowest point of the energy profile that is to the keto form in the GS. The graph is a superposition of two graphs: reaction coordinates of the IRCs in the GS and in S1 are not the same.

**TABLE 2** | Electronic transition energies between S1 and GS ( $T_e$ ) in the protein at the TD-DFT/MM level for the resulting structures of the IRC, starting with  $TS_{S1\_keto}^{\text{prot}}$ .

Structure obtained in S1	M06-2X $T_e$ eV (nm)	B3LYP $T_e$ eV (nm)
keto-OxyLH <sup>-</sup>	Last point of the profile: $T_{e\_keto\_IRC}^{\text{prot}}$	2.64 (470)      2.36 (526)
	$T_{e\_keto\_ref}^{\text{prot}}$	2.55 (486) <sup>a</sup> 2.24 (555) <sup>a</sup>
enol-OxyLH <sup>-</sup>	Last point of the profile: $T_{e\_enol\_IRC}^{\text{prot}}$	2.73 (451)      2.59 (479)
	$T_{e\_enol\_ref}^{\text{prot}}$	2.86 (434) <sup>a</sup> 2.47 (502) <sup>a</sup>

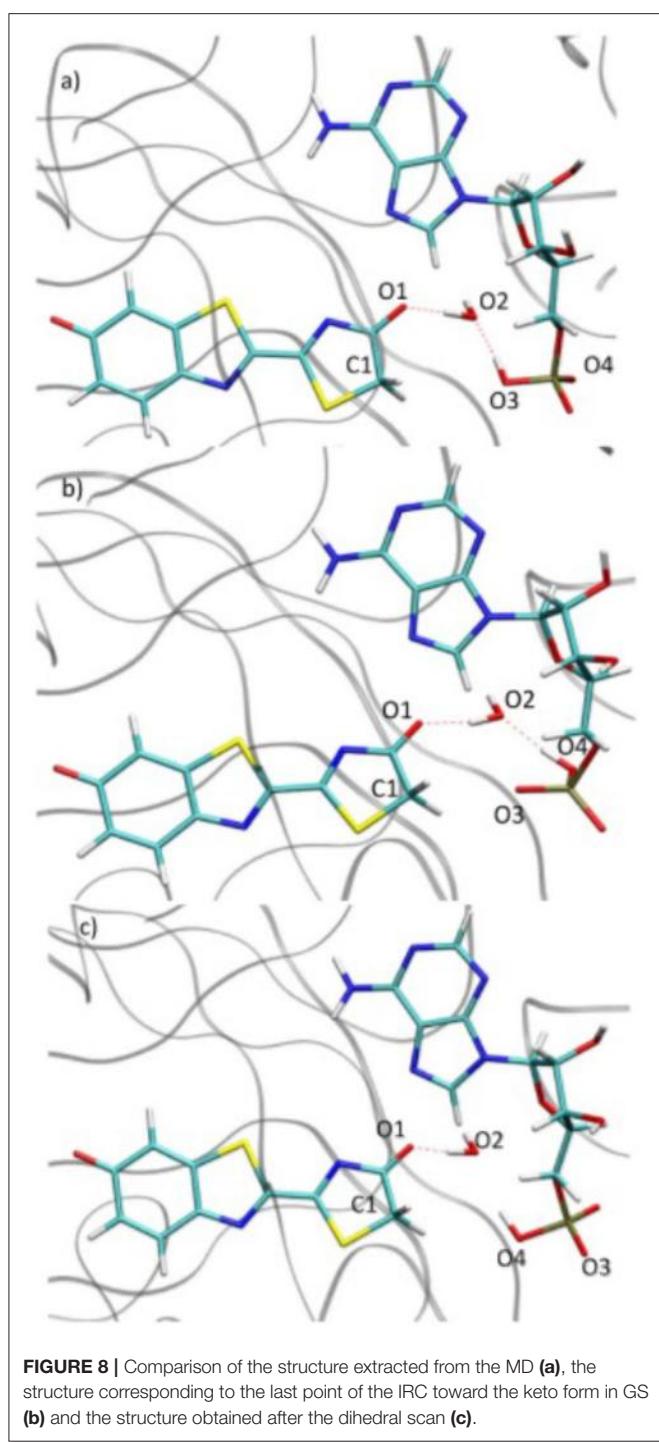
<sup>a</sup>Data in SI.

performed a scan in GS around the dihedral O4-P-O5-C4 inside the protein. The same scan at the S1 level would be much longer to obtain and we hypothesize that the results would not be very different from the ones in GS. When looking at the energetic profile, the energy reaches a minimum when the hydroxyl group O4-H2 has the same position as

O3-H2 observed during the MD (see Figure 8). However, the difference of stability and the rotation barrier between the two structures is about 5 kcal/mol, which is rather small (see Figure 9).

### Keto-enol Tautomerization

We have found that using QM/MM calculations different TSs,  $TS_{keto}^{\text{prot}}$  and  $TS_{enol}^{\text{prot}}$ , both in the GS and in S1, could lead to the keto-enol tautomerization. Besides, both TSs correspond to a 5 centers mechanism, involving two oxygen atoms of the phosphate group of the AMPH. In the  $TS_{keto}^{\text{prot}}$ , the oxyluciferin is close to the keto conformation while it looks more like the enol form in the  $TS_{enol}^{\text{prot}}$ . The energy of the  $TS_{keto}^{\text{prot}}$  is always lower than the one of  $TS_{enol}^{\text{prot}}$ , both in GS and in S1. The energy difference between the  $TS_{keto}^{\text{prot}}$  and the  $TS_{enol}^{\text{prot}}$  is quite high, about 34 kcal/mol in GS and 26 kcal/mol in S1. In all calculations we have done, the keto-OxyLH<sup>-</sup> is always more stable than the enol form, therefore it is not surprising to find  $TS_{keto}^{\text{prot}}$  more stable than  $TS_{enol}^{\text{prot}}$ .



**FIGURE 8 |** Comparison of the structure extracted from the MD (a), the structure corresponding to the last point of the IRC toward the keto form in GS (b) and the structure obtained after the dihedral scan (c).

This difference of energy also plays a major role in the barriers heights. The ones computed for the  $\text{TSs}_{\text{keto}}^{\text{prot}}$  are significantly lower than the ones of  $\text{TSs}_{\text{enol}}^{\text{prot}}$ , both in GS and in S1. However, the computed values remain high (the lowest computed one is *ca.* 51 kcal/mol, see Table 3) compared to previous studies of keto-enol tautomerization (*ca.* 40 kcal/mol) (Cucinotta et al., 2006; Alagona and Ghio, 2008).

The presence of the  $\text{TSs}_{\text{keto}}^{\text{prot}}$  and  $\text{TSs}_{\text{enol}}^{\text{prot}}$  also raises the question of a preferable pathway. Indeed, one possible hypothesis is that, because the  $\text{TSs}_{\text{keto}}^{\text{prot}}$  structures are closer to the keto-OxyLH<sup>-</sup> geometry, the tautomerization reaction from the keto tautomer to the enol tautomer might go through the  $\text{TSs}_{\text{keto}}^{\text{prot}}$ . The reverse reaction, enol to keto tautomerization, should also go through the  $\text{TSs}_{\text{keto}}^{\text{prot}}$  as they are lower in energy but because the  $\text{TSs}_{\text{enol}}^{\text{prot}}$  protons' arrangement is close to the reactant enol, the path through these TSs might also be considered.

From our calculations and with the hypothesis we have taken, the energy barriers calculated in the protein are very high and show that the tautomerization would not be easy inside the protein both in GS and S1. Experimental results show that after the reaction, a mixing keto and enol forms have been detected inside the protein (Naumov et al., 2009). However, other experimental results show that the keto-OxyLH<sup>-</sup> tautomer is the only bioluminescence emitter in fireflies, as one other recent study also shows (Pirrung et al., 2017). From all these results, we can deduce that the tautomerization is most probably difficult in S1 and, for it to happen in GS after light emission, the protein environment should change (for example, movement of the C-term and rearrangement of the H-bonding network). This is still to be proved with further calculations, especially when using other hypothesis like the protonation state of AMP, using more snapshot from the MD, or using a bigger QM region. Moreover, some other details can be taken into account to refine the model. In the chosen snapshot, the QM part does not exhibit strong interaction with the protein. As we already mentioned before, some water molecules or residues can form hydrogen bonds with the concern atoms of the tautomerization reaction.

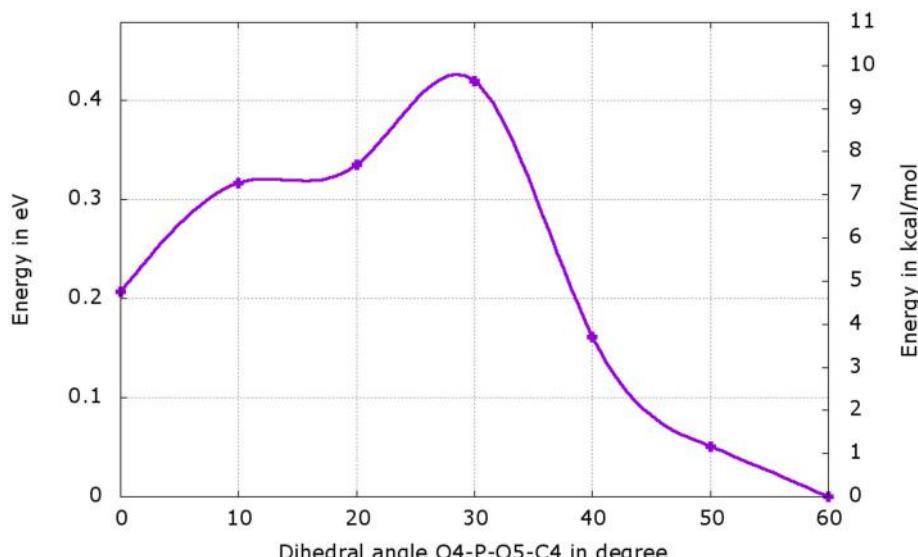
The use of QM/MM calculations has provided a better model regarding the keto-enol tautomerization compared to implicit solvent model. The main finding concerns the characterization of the TSs. Inside the protein, the concerted displacement of 3 protons is described. The 5 centers TS geometry shows that the active site of the protein is quite flexible and can sustain a complex chemical reaction.

## CONCLUSION

In this publication we demonstrate the possibility to explore chemical reactions using QM/MM calculations by the study of keto-enol tautomerization of the emitter of the bioluminescence in fireflies.

Extensive MD calculations show a recurrence of the presence of a water molecule between the oxyluciferin and AMPH, which allows 3 protons transfer during tautomerization. Preliminary QM calculations in PCM are necessary to guess the TSs as models for further calculation at the QM/MM level.

The use of QM/MM calculations to study the chemical reaction unveils some unexpected results. First, the reaction is possible in protein when 5 chemical centers are involved, in contrary to the PCM study, where only a structure compatible with the 4 centers reaction is observed in the GS. Secondly, we have found two different TSs that can carry out the tautomerization reaction. These TSs reflect a tautomer of the



**FIGURE 9** | Energetic profile of the scan around the O4-P-O5-C4 dihedral. The structure with dihedral at 0° corresponds to the last point of the IRC starting from both, the  $TS_{GS\_ketol}^{prot}$  or the  $TS_{GS\_enol}^{prot}$ , (Figure 8b). The point with dihedral at 60° (Figure 8c) corresponds to a structure similar to the starting structure extracted from the MD (Figure 8a).

**TABLE 3** | Energy barriers in kcal/mol of keto-enol and enol-keto tautomerization paths, in the GS and in S1, through  $TS_{keto}^{prot}$  and  $TS_{enol}^{prot}$ .

	Through $TS_{keto}^{prot}$		Through $TS_{enol}^{prot}$	
	GS	S1	GS	S1
Keto to enol path	65	67	105	77
Enol to keto path	58	51	100	66

emitter, one is similar to the keto form while the other is close to the enol one. In addition, these TSs have distinct energies, the  $TS_{keto}^{prot}$  are the most stable ones in both the GS and S1.

The computed barriers are quite high or even impassable. It is thus nowadays complicated to think that the keto-enol tautomerization can take place inside the protein in the excited state before emitting. This gives another proof of the role of the keto-OxyLH<sup>-</sup> as the main emitter of the bioluminescence.

The results presented here can be improved in different ways. It is possible to take into account several snapshots and improve the general picture of the system. For this purpose, we are presently studying the emission spectra simulation of different analogs of the oxyluciferin using a statistical number of snapshots. Moreover, some residues can have an effect on

the barrier height. The residue LYS 443, which is important in fireflies seems a good participant. Other methods can be used and have already prove efficiency, like QM/MM dynamics or meta-dynamics (Cucinotta et al., 2006).

We think that this protocol can be applied to other biological systems, like DNA (Cerón-Carrasco and Jacquemin, 2013), and bring new insights in modeling chemical reactions.

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

## ACKNOWLEDGMENTS

All authors are grateful to the French Agence Nationale de la Recherche (grant ANR-BIOLUM ANR-16-CE29-0013). Cristina García-Iriepa acknowledge Fundación Ramón Areces for a postdoctoral fellowship.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2018.00116/full#supplementary-material>

## REFERENCES

- Acevedo, O., and Jorgensen, W. L. (2010). Advances in Quantum and Molecular Mechanical (QM/MM) simulations for organic and enzymatic reactions. *Acc. Chem. Res.* 43, 142–151. doi: 10.1021/ar900171c

- Adam, W. (1982). *Chemical and Biological Generation of Excited States*. New York, NY: Elsevier Science.  
 Alagona, G., and Ghio, C. (2008). Keto-enol tautomerism in linear and cyclic  $\beta$ -diketones: a DFT study in vacuo and in solution. *Int. J. Quant. Chem.* 108, 1840–1855. doi: 10.1002/qua.21613

- Anandakrishnan, R., Aguilar, B., and Onufriev, A. V. (2012). H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* 40, W537–W541. doi: 10.1093/nar/gks375
- Becker, O. M., MacKerell, A. D., Roux, B., and Watanabe, M. (2001). *Computational Biochemistry and Biophysics*. New York, NY; Basel: CRC Press.
- Beraud-Pache, R., and Navizet, I. (2016). QM/MM calculations on a newly synthesised oxyluciferin substrate: new insights into the conformational effect. *Phys. Chem. Chem. Phys.* 18, 27460–27467. doi: 10.1039/C6CP02585D
- Branchini, B. R., Rosenberg, J. C., Fontaine, D. M., Southworth, T. L., Behney, C. E., and Uzasci, L. (2011). Bioluminescence is produced from a trapped firefly luciferin conformation predicted by the domain alternation mechanism. *J. Am. Chem. Soc.* 133, 11088–11091. doi: 10.1021/ja2041496
- Cancès, E., Mennucci, B., and Tomasi, J. (1997). A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. *J. Chem. Phys.* 107, 3032–3041. doi: 10.1063/1.474659
- Case, D. A., Babin, V., Berryman, J., Betz, R. M., Cai, Q., Cerutti, D. S., et al. (2014). *Amber* 14. San Francisco, CA: University of California.
- Cerón-Carrasco, J. P., and Jacquemin, D. (2013). Electric field induced DNA damage: an open door for selective mutations. *Chem. Commun. (Camb.)* 49, 7578–7580. doi: 10.1039/c3cc42593b
- Chéron, N., Jacquemin, D., and Fleurat-Lessard, P. (2012). A qualitative failure of B3LYP for textbook organic reactions. *Phys. Chem. Chem. Phys.* 14, 7170–7175. doi: 10.1039/c2cp40438a
- Cucinotta, C. S., Ruini, A., Catellani, A., and Stirling, A. (2006). Ab initio molecular dynamics study of the keto-enol tautomerism of acetone in solution. *Chemphyschem* 7, 1229–1234. doi: 10.1002/cphc.200600007
- Ferré, N., and Ángyan, J. G. (2002). Approximate electrostatic interaction operator for QM/MM calculations. *Chem. Phys. Lett.* 356, 331–339. doi: 10.1016/S0009-2614(02)00343-3
- Frey, P. A., and Hegeman, A. D. (2007). *Enzymatic Reaction Mechanisms*. Oxford: Oxford University Press.
- Friesner, R. A., and Gualar, V. (2005). Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. *Annu. Rev. Phys. Chem.* 56, 389–427. doi: 10.1146/annurev.physchem.55.091602.094410
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2010). *Gaussian09 Revision, D.01*. Wallingford, CT: Gaussian Inc.
- Gao, J., Ma, S., Major, D. T., Nam, K., Pu, J., and Truhlar, D. G. (2006). Mechanisms and free energies of enzymatic reactions. *Chem. Rev.* 106, 3188–3209. doi: 10.1021/cr050293k
- García-Iriepa, C., Gueye, M., Léonard, J., Martínez-López, D., Campos, P. J., Frutos, L. M., et al. (2016). A biomimetic molecular switch at work: coupling photoisomerization dynamics to peptide structural rearrangement. *Phys. Chem. Chem. Phys.* 18, 6742–6753. doi: 10.1039/C5CP07599H
- Gascón J. A., Sproviero E. M., and Batista, V. S. (2005). QM/MM Study of the NMR spectroscopy of the retinyl chromophore in visual rhodopsin. *J. Chem. Theory Comput.* 1, 674–85. doi: 10.1021/ct0500850
- Gattuso, H., García-Iriepa, C., Sampedro, D., Monari, A., and Marazzi, M. (2017). Simulating the electronic circular dichroism spectra of photoreversible peptide conformations. *J. Chem. Theory Comput.* 13, 3290–3296. doi: 10.1021/acs.jctc.7b00163
- Gozem, S., Luk, H. L., Schapiro, I., and Olivucci, M. (2017). Theory and simulation of the ultrafast double-bond isomerization of biological chromophores. *Chem. Rev.* 117, 13502–13565. doi: 10.1021/acs.chemrev.7b00177
- Hales, B. J. (1976). Bioenergetics of Photosynthesis. *Photochem. Photobiol.* 23, 145–145. doi: 10.1111/j.1751-1097.1976.tb06788.x
- Hirano, T., Hasumi, Y., Ohtsuka, K., Maki, S., Niwa, H., Yamaji, M., et al. (2009). Spectroscopic studies of the light-color modulation mechanism of firefly (beetle) bioluminescence. *J. Am. Chem. Soc.* 131, 2385–2396. doi: 10.1021/ja808836b
- Hosseinkhani, S. (2011). Molecular enigma of multicolor bioluminescence of firefly luciferase. *Cell. Mol. Life Sci.* 68, 1167–1182. doi: 10.1007/s00018-010-0607-0
- Jensen, J. H., Li, H., Robertson, A. D., and Molina, P. A. (2005). Prediction and Rationalization of Protein pKa Values Using QM and QM/MM Methods. *J. Phys. Chem. A* 109, 6634–6643. doi: 10.1021/jp051922x
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869
- Laurent, A. D., Blondel, A., and Jacquemin, D. (2015). Choosing an atomic basis set for TD-DFT, SOPPA, ADC(2), CIS(D), CC2 and EOM-CCSD calculations of low-lying excited states of organic dyes. *Theor. Chem. Acc.* 134:76. doi: 10.1007/s00214-015-1676-9
- Liu, F., Liu, Y., De Vico, L., and Lindh, R. (2009). A CASSCF/CASPT2 approach to the decomposition of thiazole-substituted dioxetanone: Substitution effects and charge-transfer induced electron excitation. *Chem. Phys. Lett.* 484, 69–75. doi: 10.1016/j.cplett.2009.11.009
- Lonsdale, R., Harvey, J. N., and Mulholland, A. J. (2012). A practical guide to modelling enzyme-catalysed reactions. *Chem. Soc. Rev.* 41, 3025–3038. doi: 10.1039/c2cs15297e
- Meister, A. (1983). *Advances in Enzymology : and Related Areas of Molecular Biology*. Vol. 54. New York, NY; Chichester; Brisbane; Toronto; Singapore: Wiley.
- Melaccio, F., Olivucci, M., Lindh, R., and Ferré, N. (2011). Unique QM/MM potential energy surface exploration using microiterations. *Int. J. Quantum Chem.* 111, 3339–3346. doi: 10.1002/qua.23067
- Mennucci, B., Cancès, E., and Tomasi, J. (1997). Evaluation of solvent effects in isotropic and anisotropic dielectrics and in ionic solutions with a unified integral equation method: theoretical bases, computational implementation, and numerical applications. *J. Phys. Chem. B* 101, 10506–10517. doi: 10.1021/jp971959k
- Metzler, D. E., and Metzler, C. M. (2001). *Biochemistry : The Chemical Reactions of Living Cells*. San Diego, CA: Harcourt/Academic Press.
- Mulholland, A. J. (2005). Modelling enzyme reaction mechanisms, specificity and catalysis. *Drug Discov. Today* 10, 1393–1402. doi: 10.1016/S1359-6446(05)03611-1
- Naumov, P., Ozawa, Y., Ohkubo, K., and Fukuzumi, S. (2009). Structure and spectroscopy of oxyluciferin, the light emitter of the firefly bioluminescence. *J. Am. Chem. Soc.* 131, 11590–11605. doi: 10.1021/ja904309q
- Navizet, I., Liu, Y. J., Ferré, N., Roca-Sanjuán, D., and Lindh, R. (2011). The chemistry of bioluminescence: An analysis of chemical functionalities. *Chemphyschem* 12, 3065–3076. doi: 10.1002/cphc.201100504
- Navizet, I., Liu, Y.-J., Ferré, N., Roca-Sanjuán, D., and Lindh, R. (2010). Color-Tuning mechanism of firefly investigated by multi-configurational perturbation method. *J. Am. Chem. Soc.* 132, 706–712. doi: 10.1021/ja908051h
- Navizet, I., Roca-Sanjuán, D., Yue, L., Liu, Y. J., Ferré, N., and Lindh, R. (2013). Are the bio- and chemiluminescence states of the firefly oxyluciferin the same as the fluorescence state? *Photochem. Photobiol.* 89, 319–325. doi: 10.1111/php.12007
- Page, M., Doubleday, C., and McIver, J. W. (1990). Following steepest descent reaction paths. The use of higher energy derivatives with ab initio electronic structure methods. *J. Chem. Phys.* 93, 5634–5642. doi: 10.1063/1.459634
- Page, M., and McIver, J. W. (1988). On evaluating the reaction path Hamiltonian. *J. Chem. Phys.* 88, 922–935. doi: 10.1063/1.454172
- Palczewski, K. (2012). Chemistry and biology of vision. *J. Biol. Chem.* 287, 1612–1619. doi: 10.1074/jbc.R111.301150
- Peng, C., and Schlegel, H. B. (1993). Combining synchronous transit and quasi-newton methods to find transition states. *Isr. J. Chem.* 33, 449–454. doi: 10.1002/ijch.199300051
- Pirrung, M. C., Dorsey, A., Howitt, N., De, and Liao, J. (2017).  $\beta$ -Deuterium Isotope Effects on Firefly Luciferase Bioluminescence. *ChemistryOpen* 6, 697–700. doi: 10.1002/open.201700136
- Ponder, J. W. (2004). *TINKER, Software Tools for Molecular Design, Version 6.3*.
- Ramos, M. J., and Fernandes, P. A. (2008). Computational enzymatic catalysis. *Acc. Chem. Res.* 41, 689–698. doi: 10.1021/ar7001045
- Riccardi, D., Schaefer, P., Yang, Y., Yu, H., Ghosh, N., Prat-Resina, X., et al. (2006). Development of Effective Quantum Mechanical/Molecular Mechanical (QM/MM) methods for complex biological processes. *J. Phys. Chem. B* 110, 6458–6469. doi: 10.1021/jp0563610
- Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977). Numerical integration of the Cartesian equations of motion of a system with

- constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23, 321–341. doi: 10.1016/0021-9991(77)90098-5
- Sabin, J. R., Brändas, E., and Canuto, S. (2010). *Advances in Quantum Chemistry. Combining Quantum Mechanics and Molecular Mechanics*. San Diego, CA: Academic Press.
- Senn, H. M., and Thiel, W. (2007). QM/MM studies of enzymes. *Curr. Opin. Chem. Biol.* 11, 182–187. doi: 10.1016/j.cbpa.2007.01.684
- Senn, H. M., and Thiel, W. (2009). QM/MM Methods for Biomolecular Systems. *Angew. Chemie Int. Ed.* 48, 1198–1229. doi: 10.1002/anie.200802019
- Snellenburg, J. J., Laptenok, S. P., DeSa, R. J., Naumov, P., and Solntsev, K. M. (2016). Excited-State Dynamics of Oxyluciferin in Firefly Luciferase. *J. Am. Chem. Soc.* 138, 16252–16258. doi: 10.1021/jacs.6b05078
- Sousa, S. F., Ribeiro, A. J. M., Neves, R. P. P., Brás, N. F., Cerqueira, N. M. F. S. A., Fernandes, P. A., et al. (2017). Application of quantum mechanics/molecular mechanics methods in the study of enzymatic reaction mechanisms. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 7:e1281. doi: 10.1002/wcms.1281
- Stephens, P. J., Devlin, F. J., Chabalowski, C. F., and Frisch, M. J. (1994). Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* 98, 11623–11627. doi: 10.1021/j100096a001
- Sundlov, J. A., Fontaine, D. M., Southworth, T. L., Branchini, B. R., and Gulick, A. M. (2012). Crystal structure of firefly luciferase in a second catalytic conformation supports a domain alternation mechanism. *Biochemistry* 51, 6493–6495. doi: 10.1021/bi300934s
- Tang, K., Zhang, J., and Liang, J. (2014). Fast protein loop sampling and structure prediction using distance-guided sequential chain-growth monte carlo method. *PLoS Comput. Biol.* 10:e1003539. doi: 10.1371/journal.pcbi.1003539
- van der Kamp, M. W., and Mulholland, A. J. (2013). Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry* 52, 2708–2728. doi: 10.1021/bi400215w
- van der Kamp, M. W., Shaw, K. E., Woods, C. J., and Mulholland, A. J. (2008). Biomolecular simulation and modelling: status, progress and prospects. *J. R. Soc. Interface* 5 (Suppl 3), S173–S190. doi: 10.1098/rsif.2008.0105. focus
- Zhao, Y., and Truhlar, D. G. (2008). The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other function. *Theor. Chem. Acc.* 120, 215–241. doi: 10.1007/s00214-007-0310-x
- Zscherp, C., and Barth, A. (2001). Reaction-Induced infrared difference spectroscopy for the study of protein reaction mechanisms. *Biochemistry* 40, 1875–1883. doi: 10.1021/bi002567y

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Beraud-Pache, Garcia-Iriegua and Navizet. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Exploring the Interaction Mechanism Between Cyclopeptide DC3 and Androgen Receptor Using Molecular Dynamics Simulations and Free Energy Calculations

Huimin Zhang<sup>1†</sup>, Tianqing Song<sup>1†</sup>, Yizhao Yang<sup>2</sup>, Chenggong Fu<sup>1</sup> and Jiazhong Li<sup>1\*</sup>

<sup>1</sup> School of Pharmacy, Lanzhou University, Lanzhou, China, <sup>2</sup> College of Chemistry and Chemical Engineering, Lanzhou University, Lanzhou, China

## OPEN ACCESS

### Edited by:

Sam P. De Visser,  
University of Manchester,  
United Kingdom

### Reviewed by:

Gerardo Andres Cisneros,  
University of North Texas,  
United States  
Jitrayut Jittonnom,  
University of Phayao, Thailand

### \*Correspondence:

Jiazhong Li  
lijiazhong@lzu.edu.cn

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Theoretical and Computational  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

Received: 29 January 2018

Accepted: 30 March 2018

Published: 19 April 2018

### Citation:

Zhang H, Song T, Yang Y, Fu C and Li J (2018) Exploring the Interaction Mechanism Between Cyclopeptide DC3 and Androgen Receptor Using Molecular Dynamics Simulations and Free Energy Calculations. *Front. Chem.* 6:119.  
doi: 10.3389/fchem.2018.00119

Androgen receptor (AR) is a key target in the discovery of anti-PCa (Prostate Cancer) drugs. Recently, a novel cyclopeptide *Diffusa Cyclotide-3* (DC3), isolated from *Hedyotis diffusa*, has been experimentally demonstrated to inhibit the survival and growth of LNCap cells, which typically express T877A-mutated AR, the most frequently detected point mutation of AR in castration-resistant prostate cancer (CRPC). But the interaction mechanism between DC3 and AR is not clear. Here in this study we aim to explore the possible binding mode of DC3 to T877A-mutated AR from molecular perspective. Firstly, homology modeling was employed to construct the three-dimensional structure of the cyclopeptide DC3 using 2kux.1.A as the template. Then molecular docking, molecular dynamics (MD) simulations, and molecular mechanics/generalized Born surface area (MM-GBSA) methods were performed to determine the bind site and explore the detailed interaction mechanism of DC3-AR complex. The obtained results suggested that the site formed by H11, loop888-893, and H12 (site 2) was the most possible position of DC3 binding to AR. Besides, hydrogen bonds, hydrophobic, and electrostatic interactions play dominant roles in the recognition and combination of DC3-AR complex. The essential residues dominant in each interaction were specifically revealed. This work facilitates our understanding of the interaction mechanism of DC3 binding to AR at the molecular level and contributes to the rational cyclopeptide drug design for prostate cancer.

**Keywords:** Cyclopeptide DC3, androgen receptor, protein drug interaction, homology modeling, molecular docking, molecular dynamics simulations

## INTRODUCTION

Prostate cancer (PCa) has become the second frequently diagnosed cancer in men throughout the world (American Cancer Society, 2015). Prostate, lung and bronchus, colorectal cancers accounts for about 44% of all cancer cases in men, with PCa alone accounting for 1 in 5 new diagnoses (Siegel et al., 2016). PCa is especially common in economically developed countries and regions like Northern and Western Europe, Northern America, and Oceania (American Cancer Society, 2015). In America, prostate cancer is the most common cancer and was predicted as the leading cause of male cancer-related death over the next decade (Siegel et al., 2016). In those less developed

countries, the incidence rate of prostate cancer is increasing with stable or increasing mortality trend in recent years (Center et al., 2012).

Androgen receptor (AR) (NR3C4, nuclear receptor subfamily 3, group C, gene 4), a member of steroid hormone group of nuclear receptor superfamily, plays an essential role in the development and proliferation of prostate cancer (Tsai and O'Malley, 1994; Mangelsdorf et al., 1995; Nuclear Receptors Nomenclature Committee, 1999). The survival and growth of PCa cells are dependent on the androgenic stimulation through AR. Firstly, 5 $\alpha$ -dihydrotestosterone (DHT) binds to AR to promote the association of AR co-regulators. Then the activated AR migrates into nucleus and regulates the expression of target genes in prostate cells (Heinlein and Chang, 2004).

Clinically, PCa is commonly treated by AR pathway perturbation, such as androgen suppression via surgical or chemical castration [gonadotropin-releasing hormone (GnRH) analogs] means (Palmbos and Hussain, 2013). AR antagonist drugs, such as flutamide, nilutamide, bicalutamide, and enzalutamide, take effects by suppressing the action of androgens via competing for AR binding sites (Yamamoto et al., 2012). These androgen blockade therapies are initially effective, however, a considerable population of patients ultimately develop as castration-resistant prostate cancer (CRPC) after prolonged use of an AR antagonist (Schröder, 2008; Yamaoka et al., 2010). AR mutation is one of the leading causes of antiandrogens resistance (Tan et al., 2015). These mutated ARs bind to other steroid hormones and induce the activation of AR transcriptional activity in response to antiandrogens, which results in the PCa growth (Tan et al., 2015). In this case, it shows far-reaching significance to seek and explore novel anti-CRPC drugs targeting gene mutational AR.

DC3 (*Diffusa Cyclotide-3*) is a novel cyclopeptide isolated from the traditional Chinese Medicine (TCM) *Hedyotis diffusa* (Hu et al., 2015), which has been widely used for the treatment of various cancers and tumors, including prostate cancer, in China with a long history (Lin et al., 2010, 2011; Liu et al., 2010; Lee et al., 2011). It has been experimentally detected that DC3 expresses potent cytotoxicity against LNCaP cells and inhibits the cell migration and invasion. Besides, it can significantly inhibit the development of tumor in weight and size in the mouse xenograft model. All these findings lead to the conclusion that DC3 has evident anti-PCa effects both *in vitro* and *in vivo* (Hu et al., 2015). Moreover, in the DC3 sensitivity experiments on three types of human prostate cancer cells, androgen dependent LNCaP cell lines showed obvious higher sensitivity to DC3 comparing to androgen independent PC3 and DU145 cell lines. Besides, LNCaP cell lines typically express T877A-mutated AR, which is the most frequently detected point mutation in CRPC (Veldscholte et al., 1992; Zhou et al., 2010; Yamada et al., 2013). All these evidences suggest that DC3 is a potential candidate binding to T877A AR.

However, the interaction mechanisms between DC3 and T877A-mutated AR are not clear. Therefore, it will be constructive and profoundly significant to launch the mechanism-relevant research. Fortunately, the amino acid sequence of DC3 has been experimentally determined (Hu

et al., 2015), which makes it possible to explore the interaction mechanism at the molecular level. Referencing to the published papers, the current research could consist of three parts: (1) constructing the three-dimensional structure of the cyclopeptide (Jitonnom et al., 2012), in our case is DC3; (2) determining the binding site and binding pose of cyclopeptide to protein (Punkvarg et al., 2015); (3) investigating the detailed interaction mechanism between cyclopeptide and protein (Liu et al., 2015; Hitzenberger et al., 2017).

In our study, first of all, homology modeling technology was conducted to construct the three-dimensional structure of cyclopeptide DC3 based on its amino acid sequence. Then molecular docking, all-atom molecular dynamics (MD) simulations and molecular mechanics/generalized Born surface area (MM/GBSA) methods and various MD trajectory analysis methods were combined to explore the most possible binding site of DC3 to AR, investigate the key residues dominant in the binding process, and elucidate the detailed interaction mechanism. The results are expected to reveal the interaction mechanism of DC3-AR complex, promote the development of DC3 and correlative cyclopeptide AR antagonist, which will contribute to the rational drug design for prostate cancer.

## METHODS

### Homology Modeling of Cyclopeptide DC3

Homology modeling is a common technique to construct three-dimensional structure from amino acid sequence using homologous proteins with known structure as templates (Topham et al., 1990; Bordoli et al., 2009; Wang Z. et al., 2015). As amino acid sequence of DC3 was confirmed by Edman degradation and gene cloning (Hu et al., 2015), homology modeling was adopted here to build the 3D structure of DC3 using SWISS-MODEL (Arnold et al., 2006; Guex et al., 2009; Kiefer et al., 2009; Biasini et al., 2014; Bienert et al., 2016). Here, the SWISS-MODEL Template Library (SMTL) is searched both with BLAST and HHblits to identify templates and target-template alignments (Arnold et al., 2006). Then the template was selected based on various criteria such as sequence similarity, sequence identity, coverage, the global quality estimation score (GMQE) and so on.

### Molecular Docking Analysis of DC3 to AR

Molecular docking (Benkert et al., 2011; Meng et al., 2011; Yuriev et al., 2015) was used to analyze the possible binding site and preferred orientation of DC3 into androgen receptor by simulating combining conformation and computing binding affinity. Here the crystal structure coordinates of the T877A-mutated AR LBD was obtained from the RCSB Protein Data Bank (<http://www.rcsb.org/pdb>; PDB ID: 4OHA). The missing loop regions were refined by Discovery Studio 2.5. (Accelrys Inc. CA, 2009). Molecular docking process was carried out by using ZDOCK module. The rigid-body protein-protein docking program ZDOCK uses the Fast Fourier Transform algorithm to enable an efficient global docking search on a 3D grid, and utilizes a combination of shape complementarity, electrostatics and statistical potential terms for scoring. Finally,

two simple scoring functions—ZRank Score and ZDock Score, pose amount of each cluster, and the rationality of binding mode were taken into consideration to evaluate the docking results.

## Molecular Dynamics Simulations

Molecular dynamics (MD) simulations were operated through Amber12 package (Case et al., 2012). All the simulations are under the circumstance of ff99SB force field (Hornak et al., 2006) and periodic boundary condition. Firstly, six chloride counterions were added to each system to maintain the electro-neutrality. Then all studied systems were, respectively, immersed into a cubic box of TIP3P (van der Spoel and van Maaren, 2006) water with edge of the box at least 10 Å distant from the complex. Energy minimization was carried out in three stages with different harmonic restraint: all atoms constrained by 5.0 kcal·mol<sup>-1</sup>·Å<sup>-2</sup>, only receptor backbone atoms constrained by 3.0 kcal·mol<sup>-1</sup>·Å<sup>-2</sup> and without any restraint. Each minimization was executed for 5,000 steps, in which the first 2,500 steps were calculated by the steepest descent method while the subsequent 2,500 steps were executed by conjugated gradient method. These systems were heated up to 310.0 K in the NVT ensemble for 100 ps with the receptor backbone atoms constrained by 5.0 kcal·mol<sup>-1</sup>·Å<sup>-2</sup>. And then, a total of 1.5 ns equilibration of each system was performed in NPT ensemble, where the former 800 ps were divided into four stages and the restraints applied to these stages were in a descending order (4.0, 3.0, 2.0, 1.0 kcal·mol<sup>-1</sup>·Å<sup>-2</sup> respectively), the latter 700 ps were carried out without any restraint. Minimization and heat, as well as equilibration were executed in the Sander program. Finally, a 150 ns production of MD simulations of each system was performed in the PMEMD program at 310.0 K, 1 atm in the NPT ensemble without any restraint. The Langevin thermostat was used to control the temperature and the Berendsen barostat was used for constant pressure simulation. The time step was set as 2 fs, and the coordinates of trajectories were recorded every 2 ps. During this simulation, the SHAKE algorithm (Ryckaert et al., 1977) was employed to constrain the bond lengths involving hydrogen, the Particle Mesh Ewald (PME) (Darden et al., 1993; Fischer et al., 2015) was adopted to calculate of electrostatic interaction with a 10 Å non-bonded cutoff.

## Free Energy Calculations

To investigate the interaction of DC3-AR systems from the energetic perspective, the binding free energy calculations based on the trajectories of MD simulations were performed by MM-GBSA method (Hou et al., 2011a,b; Xu et al., 2013; Sun et al., 2014a,b; Chen et al., 2016). The binding free energies  $\Delta G_{\text{bind}}$  was calculated as following equation:

$$\Delta G_{\text{bind}} = \Delta H - T\Delta S = \Delta E_{\text{gas}} + \Delta G_{\text{sol}} - T\Delta S$$

where  $\Delta H$  represents enthalpy contribution, which is composed of enthalpy changes in gas-phase ( $E_{\text{gas}}$ ) and solvent-phase ( $\Delta G_{\text{sol}}$ ).  $-T\Delta S$  represents entropy contribution. Entropic calculation is time-consuming, and its value will fluctuate if a

small quantity of snapshots were adopted (Hou et al., 2011a; Wang Q. et al., 2015). In this study it was omitted.  $E_{\text{gas}}$  was considered as the sum of internal interaction ( $\Delta E_{\text{int}}$ ) from bonds, angles, and torsions, van der Waals ( $\Delta E_{\text{vdw}}$ ) and electrostatic energies ( $\Delta E_{\text{ele}}$ ) as follow:

$$\Delta E_{\text{gas}} = \Delta E_{\text{int}} + \Delta E_{\text{ele}} + \Delta E_{\text{vdw}}$$

$\Delta G_{\text{sol}}$  can be decomposed into the polar and nonpolar contributions as follow:

$$\Delta G_{\text{sol}} = \Delta G_{\text{GB}} + \Delta G_{\text{SA}}$$

Here,  $\Delta G_{\text{GB}}$  represents the polar solvation contribution, which is calculated by solving GB equation (Kollman et al., 2000; Onufriev et al., 2004).  $\Delta G_{\text{SA}}$ , estimated by the solvent accessible surface area, represents the nonpolar solvation contribution.

To further explore the detailed interaction information of DC3-AR complex, free energy decomposition was performed by using MM-GBSA method to identify the key residues responsible for binding energy. The contribution of each residue was calculated without considering the contribution of entropies. The contribution is defined as the sum of van der Waals contribution ( $\Delta E_{\text{vdW}}$ ), electrostatic contribution ( $\Delta E_{\text{ele}}$ ), polar solvation contribution ( $\Delta G_{\text{GB}}$ ), and nonpolar solvation contribution ( $\Delta G_{\text{SA}}$ ):

$$\Delta G_{\text{residue}} = \Delta E_{\text{vdW}} + \Delta E_{\text{ele}} + \Delta G_{\text{GB}} + \Delta G_{\text{SA}}$$

Snapshots, used for both binding free energy and free energy decomposition calculations, were extracted from the last 50 ns of MD trajectories at intervals of 2 ps.

## MD Trajectory Analysis

### Hydrogen Bond Analysis

The number of formed hydrogen bonds vs. simulation time was calculated to detect the system stability during the process of simulation. Here, the hydrogen bond criteria was set as the distance of acceptor-donor <0.35 nm and the angle >120° (Fu et al., 2013). 0.35 nm is a common choice of hydrogen bond distance in literature (Liu et al., 2014; Wang Q. et al., 2015). The frames adopted for this calculation were extracted from the whole 150 ns MD trajectories at intervals of 2 ps. Besides, in order to determine exactly how hydrogen bonds play dominant roles in maintaining system stability in the last 50 ns MD simulations, the occupations of hydrogen bonds formed in this period were calculated as following equation (Liu et al., 2014):

$$P_{\text{hbond}} = \frac{N_{\text{exist}}}{N_{\text{total}}} \times 100\%$$

Where  $N_{\text{exist}}$  is the number of frames which formed targeted hydrogen bond, and  $N_{\text{total}}$  is the total number of frames. The occupations are varied from 1 to 100%, and a higher percentage represents a more stable-existed hydrogen bond.

## Dynamic Cross Correlation Matrix

The dynamic cross-correlation matrix (DCCM) analysis of the  $\text{C}\alpha$  atoms during the last 50 ns of the first parallel trajectory was performed to explore the correlated motion between residues of DC3-AR complex. The cross correlation matrix  $C_{ij}$ , which reflects the displacements of the  $\text{C}\alpha$  atoms relative to average positions, was determined by following equation (Lange et al., 2005; Ghosh and Vishveshwara, 2007):

$$C_{ij} = \frac{\langle \Delta R_i \cdot \Delta R_j \rangle}{\sqrt{\langle \Delta R_i \cdot \Delta R_i \rangle \langle \Delta R_j \cdot \Delta R_j \rangle}}$$

Where  $\Delta R_i$  and  $\Delta R_j$  represent displacements of atom  $i$  and  $j$ , respectively. The value of  $C_{ij}$  fluctuated from  $-1$  to  $1$ , the positive value indicates a correlated motion between the residue  $i$  and residue  $j$ , while negative values indicates an anti-correlated motion.

## Clustering Analysis

Clustering analysis was conducted by using the MMTSB toolset in Amber 12 to determine the representative structure of DC3-AR complex during the last 50 ns MD simulations. Firstly, the similar conformations of DC3-AR complex generated from the trajectory were classified into one cluster, and the most populated cluster has maximum number of conformations. Centroids of the generated clusters were then calculated and generated. Subsequently, RMSD of each structure was calculated with respect to specific centroid. Ultimately, the structure with lowest RMSD to cluster centroid from the most populated cluster was defined as the representative structure of DC3-AR. After that, the representative structure was adopted to generate the hydrophobic and electrostatic interaction surface of DC3-AR complex by using UCSF Chimera package (Pettersen et al., 2004; <http://www.cgl.ucsf.edu/chimera>).

## Dynamical Correlation Network

The cross-correlation matrix  $C_{ij}$  was also employed to build the dynamical correlation network to intuitively exhibit the correlated motion between residues in different protein domains. The  $\text{C}\alpha$  atom of each residue was defined as a “node,” and “edge” is the connection of each pair of nodes if the residue pairs interact with each other (Liu et al., 2014). The edges were computed as the following equation:

$$d_{ij} = -\log(|C_{ij}|)$$

Here, each edge has a specific contribution to the movement of complex, the motion of residue  $i$  can be used to predict the motion direction of residue  $j$ . If  $|i - j| <= 10$ , cross-correlation between  $i$  and  $j$  are ignored to remove the correlations due to special closeness. Besides,  $-0.3 \leq C_{ij} \leq 0.3$  were also deleted to make network plot more concise. Network View plugin in visual Molecular Dynamics 1.9.2 (VMD) (Humphrey et al., 1996; Hsin et al., 2008) was used to visualize the interaction network.

## RESULTS AND DISCUSSION

### Homology Modeling of Cyclopeptide DC3

Through templates searching by SWISS-MODEL, 28 qualified templates for DC3 sequence were found. According to the selection criteria and the basic structure characters of cyclopeptide (Craik et al., 1999; Sze et al., 2009) 2kux.1.A (Plan et al., 2010) was selected as the final template to construct the three-dimensional structure of DC3. As shown in **Figure 1A**, the amino acid sequence identity between template and target is 56.67%, sequence similarity is 0.52, Global Model Quality Estimation (GMQE) value is 0.94, and the target sequence is all covered. The sequence alignment and structure comparison of target and template were shown in **Figure 1B**, from where highly similarity can be easily observed. Besides, the Z-score information and predicted local similarity of each residue to target were shown in **Figures 1C,D**, respectively. All these information reflects the reliability of the constructed model. As the AR residues were numbered from 671 to 919, here, the number of DC3-residues was defined from 641 to 670 for convenience.

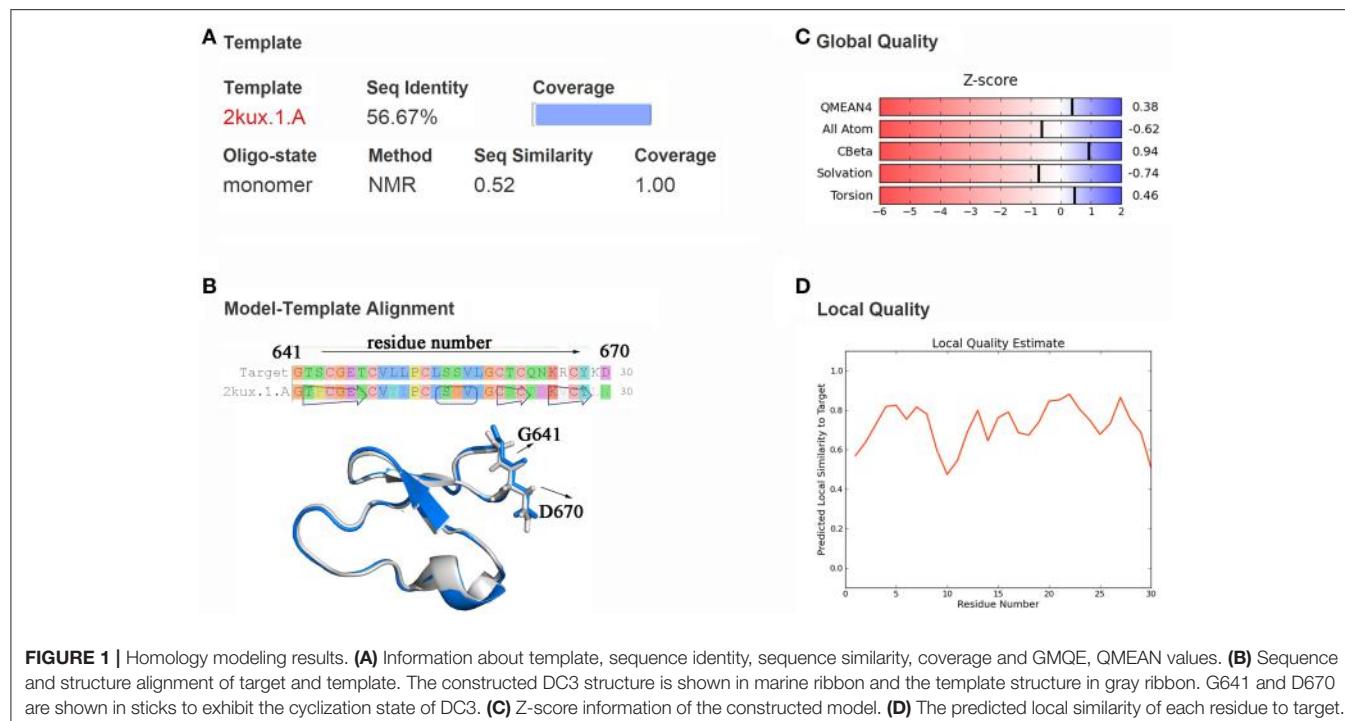
### Binding Site Exploration

#### Molecular Docking Analysis

Molecular docking was then performed to explore the possible binding site and binding mechanism of DC3-AR complex. As a result, 2,000 poses of 60 clusters were generated by ZDOCK module. The pose amount of each cluster, ZRank Score, ZDock Score, and the rationality of binding mode were combined to assess the poses. The ZRank Score represents the extent of energy contribution to the system when a ligand binds to a receptor. The ZDock Score is calculated based on the shape matching degree of receptor and ligand, and a higher score represents a better pose. Furthermore, the electrostatic interaction, Van der Waals’ force, and desolvation energy were also taken into consideration. A lower score represents a better pose. Based on these criteria, top four possible binding sites of DC3 to AR were selected. As shown in **Figures 2A,B**, the top four clusters of DC3-ARs located in Site 1, Site 2, Site 4, and Site 3, respectively. The score of binding to Site 2 was higher than other possible binding sites although the pose number of this cluster is relatively small. However, it should be noted that scoring functions do not always yield the best predictions of binding affinity (Ramírez and Caballero, 2016). To further confirm the binding affinity prediction, Molecular dynamics (MD) simulations were subsequently performed to obtain more conformational sampling of these four systems. Four poses (**Table 1**, **Figures 2C–F**) of DC3-AR complex were determined as the initial structures, named as system 1, system 2, system 3, and system 4, to perform MD simulations, which were selected from the top four possible binding sites with good ZRank Score, fine ZDock Score, and rational conformations.

### Root Mean Square Deviation

One hundred and fifty nano seconds MD simulations were calculated on the four DC3-AR complexes systems acquired from molecular docking were then performed respectively. To obtain reliable and repeatable results, three parallel MD



**FIGURE 1 |** Homology modeling results. **(A)** Information about template, sequence identity, sequence similarity, coverage and GMQE, QMEAN values. **(B)** Sequence and structure alignment of target and template. The constructed DC3 structure is shown in marine ribbon and the template structure in gray ribbon. G641 and D670 are shown in sticks to exhibit the cyclization state of DC3. **(C)** Z-score information of the constructed model. **(D)** The predicted local similarity of each residue to target.

simulations processes were executed on each system. Then root mean square deviation (RMSD) values of DC3-AR complexes backbone atoms were calculated relative to the initial structures to monitor the stability and overall convergence of each system during the simulation process. As shown in **Figure 3**, all systems experienced various degrees of fluctuations at first, but gradually tended to converge. It can be seen that, the first and third trajectories of system 4 and all parallel trajectories of other three systems reached equilibrium in the last 50 ns, which were qualified for subsequent analyses of the dynamic behavior. However, the second parallel trajectory of system 4 experienced great structure changes at about 50 ns, which suggested the relatively poor stability of system 4. Considering the abnormality, this trajectory was eliminated in the succeeding binding free energy analysis which was carried out by averaging the values of parallel trajectories.

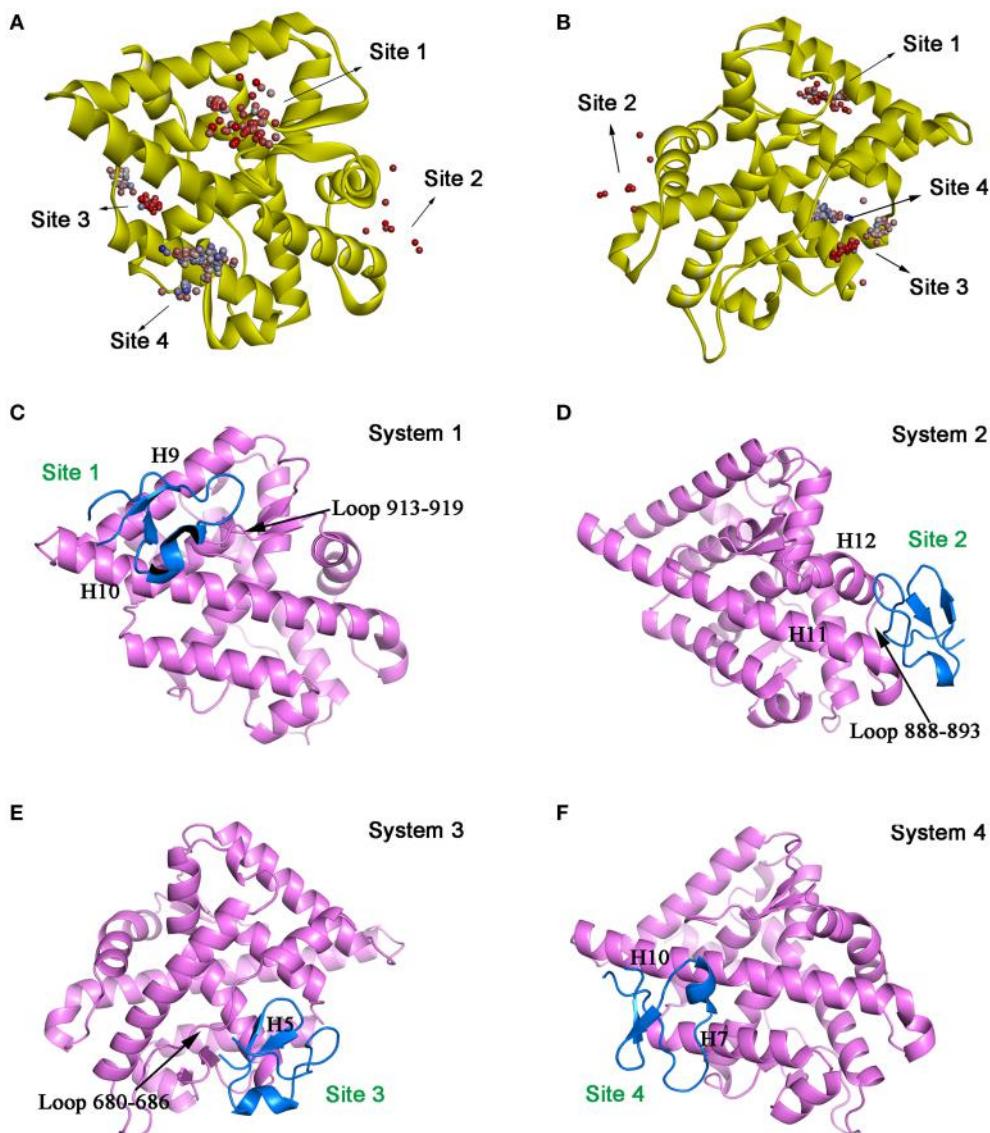
### Root Mean Square Fluctuation

The root mean square fluctuation (RMSF) reveals the fluctuation of certain residues during simulation process around its average position, which is also a tool to assess the dynamics stability of system. Here, RMSF values of  $\text{C}\alpha$  atoms in the last 50 ns were calculated by employing the first parallel trajectory of each system. RMSF of DC3 residues were shown in **Figure 4A** to explore the stability of the DC3. It can be seen that, different from the great fluctuation in other systems, DC3 residues in system 2 experienced minor motions. It demonstrated that DC3 in system 2 showed obvious superiority in the stability. RMSF of AR residues in these four systems were compared with apo-AR system (made up by AR only) to determine whether the binding of DC3 affects the stability of AR. As shown in **Figure 4B**,

the overall RMSF of system 2 is lower than apo-AR system, especially the residues 840-870 (corresponding to H9, loop 843-849, H10), 880-905 (corresponding to H10, H11, loop 888-893, H12). Moreover, only in apo-AR system and system 2, the RMSF values of all residues were under 10 Å. Whereas, residues in other systems showed apparently larger conformational changes comparing to apo-AR system. These results demonstrated that the combination of DC3 to AR in site 2 (corresponding to Helix 11, loop 888-893, Helix 12) could stabilize androgen receptor. However, DC3 combination in other sites could visibly reduce AR stability. These RMSF analyses indicated that site 2 is the most possible site of DC3 binding to AR.

### Interaction Energetic Features

In order to explore the interaction energetic features of DC3-AR complexes, MM-GBSA method was employed to calculate the binding free energies of each system. The average binding free energies and detailed energetic contribution components of the last 50 ns of parallel trajectories were calculated and shown in **Table 2**. It can be seen that the free energy of system 2 ( $-40.94$  kcal/mol) is apparently lower than system 1 ( $-33.49$  kcal/mol), system 3 ( $-18.99$  kcal/mol), and system 4 ( $-19.73$  kcal/mol). It demonstrated that DC3 showed a higher binding affinity to AR in system 2 comparing to other systems, which indicated that DC3 has a great tendency to bind to AR in site 2 and system 2 is more likely to remain stable. This result conforms to the conclusion obtained from the previous RMSF analysis. Moreover, details of the dominant components driving DC3 to bind to AR can be acquired by dissecting the binding free energy into contributing components. Here, the electrostatic interaction ( $\Delta E_{\text{ele}}$ ) in system 2 ( $-232.64$  kcal/mol) can be found to make a



**FIGURE 2 |** The top four possible binding modes of DC3-AR complex. **(A,B)** AR is shown in yellow ribbon, and each dot represent a DC3 conformation. The color of dot from red to blue indicates the ZRank Score from high to low. **(C-F)** Four initial structures of DC3-AR complex selected from the top four possible binding sites for MD simulations. DC3 is shown in marine and AR is shown in violet.

great contribution to the low binding free energy of the whole system, which reflects that significant electrostatic interactions may exist between DC3-AR complex and contribute greatly to the system stability.

#### Dynamic Cross-Correlation Matrices Analysis

The dynamic cross-correlation matrices (DCCM) analysis was further analyzed (Figure 5) to investigate the correlated conformational motions of DC3-AR complexes. Here, highly positive regions (colored by red and yellow) are associated with strong correlated motions (residue pairs move in the same direction), while negative regions (colored by blue) are linked with strong anti-correlation movements (residue pairs move in

the opposite direction). Inspecting the DC3 domains of the four systems, it can be observed that relatively stronger correlations exist between DC3 residues in system 2. Moreover, comparing to other systems, obviously more correlated and anti-correlated motions between DC3 residues and AR residues can be found in system 2. These obvious differences indicated that there were more and stronger cross-correlation motions between residues in system 2, demonstrating more intense interaction and better stability of this DC3-AR complex.

#### Hydrogen Bonds Analysis

The stronger cross-correlation between DC3 and AR residues found in system 2 might also due to the formation of hydrogen

bonds during MD simulations. Hydrogen bonds, as critical indicators of nonbonding interactions, play vital roles in the protein-ligand recognition process (Ramírez and Caballero, 2016). During this MD simulation, the number of hydrogen bonds formed between DC3 and AR vs. simulation time was calculated and plotted in **Figure 6**. Though we set 0.35 nm as the hydrogen bond criteria in this study, the distance we calculated for hydrogen bond were almost all around 3 Angstrom, long-distance hydrogen bond do not exist. As shown in this figure, hydrogen bond interaction patterns formed in system 2 remained constant during the entire simulation time. While in other systems, hydrogen bonds were unstable and most of them disappeared in about 90 ns. Even in the last 50 ns of simulation, the amount of hydrogen bonds still fluctuated a lot. This result reflects the obvious stability of system 2, which further proves that DC3 tends to bind to AR at site 2 as previous RMSF and binding free energy analyses demonstrated. Besides, the

intermolecular hydrogen bonds formed in the last 50 ns MD simulations with occupation more than 10% were listed in **Table 3**. It can be clearly observed that much more hydrogen bonds are stably formed in system 2. On one hand, there are 13 hydrogen bonds occupied more than 10% in system 2, while only 4 hydrogen bonds in system 1, one hydrogen bond in system 3, and even no one in system 4. On the other hand, the highest occupation of hydrogen bonds are 23.51 and 12.55% in system 1 and system 3, respectively, while in system 2 hydrogen bonds formed between K669-H885, D890-R666, S900-T647, and D890-E643 occupied 85.23, 74.54, 72.49, and 64.38%, respectively.

Based on these data, it can be concluded that site 2 (H11, loop888-893, H12) is the most possible site of DC3 binding to AR complex.

## Exploration of the Binding Mechanism Between DC3 and AR Site 2

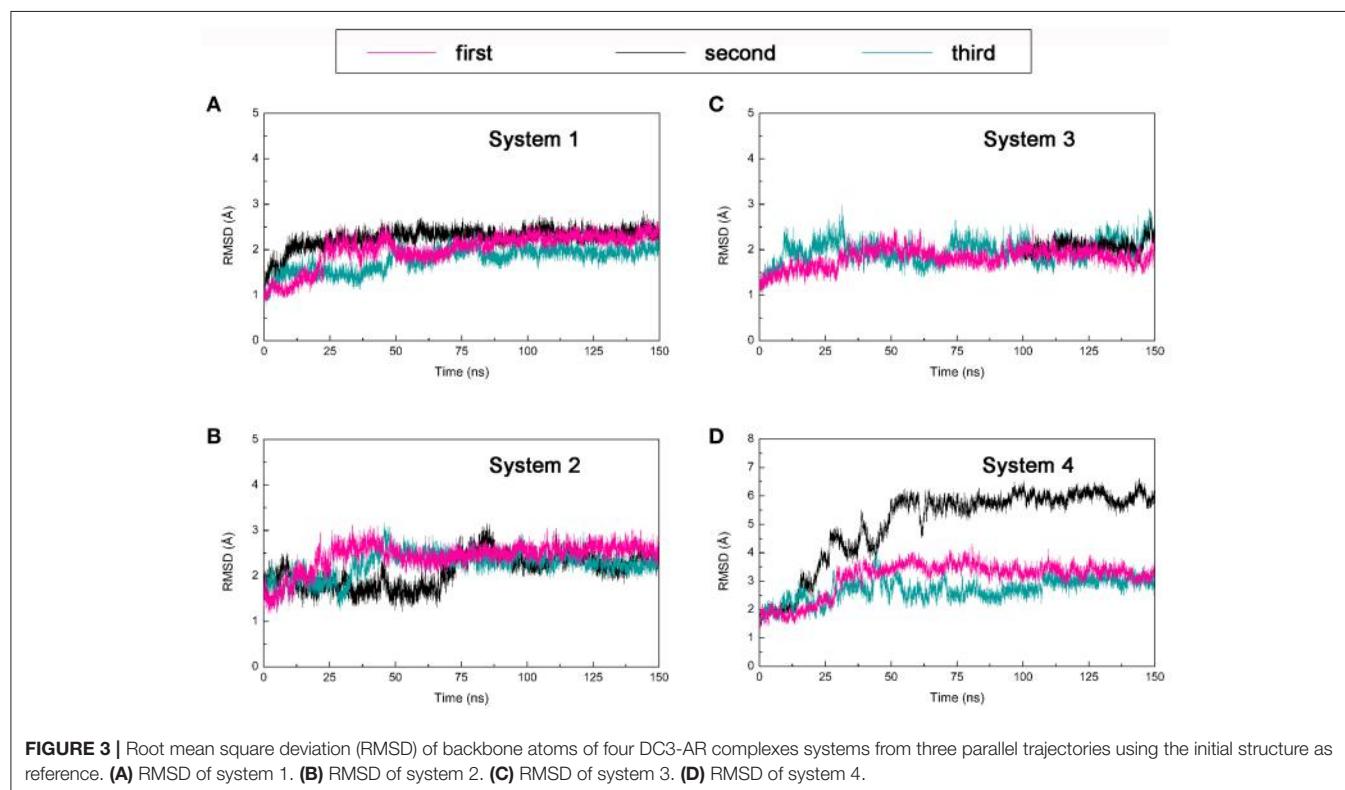
To fully explore the binding modes and interaction mechanisms of DC3 and AR, system 2 was further studied to reveal the complicated binding mechanism of DC3-AR complex.

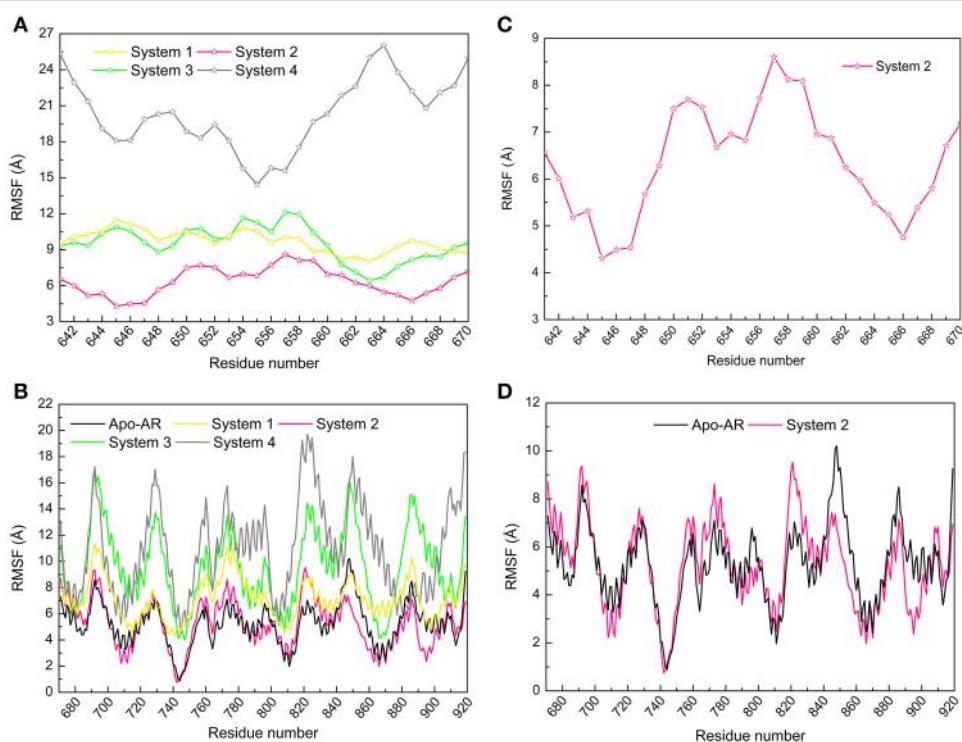
### Root Mean Square Fluctuation Analysis

According to the RMSF values of DC3 residues and AR residues shown in **Figures 4C,D**, respectively, it is easy to determine the key residues dominant in the binding process of DC3-AR complex. In DC3, residues 645-647 and 666 with low RMSF values experienced minor fluctuation, which indicated these residues were relatively more stable during the simulation process. Comparing the AR residues in system 2 to apo-AR

**TABLE 1** | The selected top four possible binding sites of DC3-AR complex and the corresponding poses scoring.

System	ZRank score (kcal/mol)	ZDock score	Binding site
1	-85.450	11.66	Site1: Helix 9, Helix 10, Loop 913-919
2	-97.309	10.92	Site2: Helix 11, Loop 888-893, Helix 12
3	-92.708	13.08	Site3: Loop 680-686, Helix 5
4	-78.121	12.00	Site4: Helix 7, Helix 10





**FIGURE 4** | Root mean square fluctuation (RMSF) values of C $\alpha$  atoms calculated by the first parallel trajectory of studied systems in the last 50 ns MD simulation. **(A)** RMSF values of DC3 residues in four systems. **(B)** RMSF values of DC3 residues in system 2. **(C)** RMSF values of AR residues in four systems and apo-AR system. **(D)** RMSF values of AR residues in system 2 and apo-AR system.

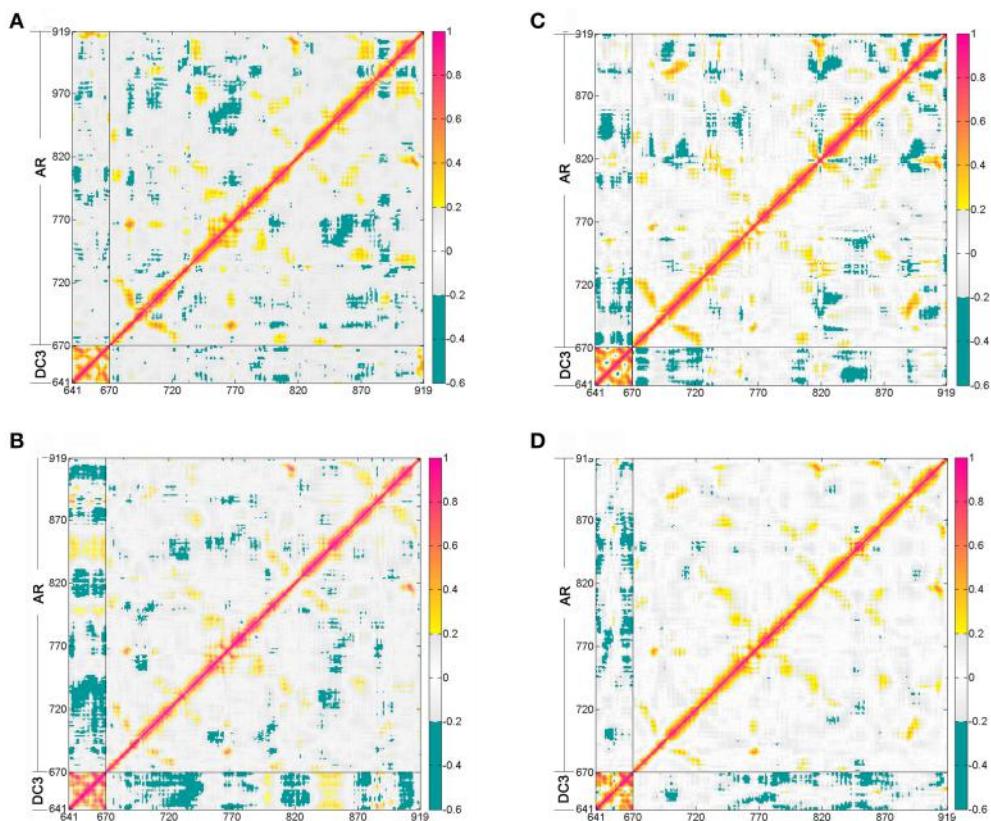
**TABLE 2** | Binding free energy and the detailed energetic contribution components of four systems of DC3-AR complex averaged by the last 50 ns of parallel trajectories (kcal/mol).

Contributions	System 1	System 2	System 3	System 4
$E_{vdW}$	-70.98(2.97)	-50.84(0.06)	-44.89(1.50)	-44.18(4.74)
$E_{ele}$	-22.58(19.09)	-232.64(25.84)	-13.28(8.82)	-36.53(10.14)
$G_{GB}$	69.34(18.17)	250.09(26.80)	44.79(7.60)	67.15(6.13)
$G_{SA}$	-9.27(0.53)	-7.55(0.23)	-5.61(0.28)	-6.16(0.82)
$E_{gas}$	-93.56(18.46)	-283.48(25.78)	-58.17(8.69)	-80.71(5.40)
$E_{solv}$	60.06(17.70)	242.54(26.57)	39.18(7.74)	60.99(6.95)
$\Delta G_{bind}$	-33.49(7.10)	-40.94(0.79)	-18.99(2.18)	-19.73(1.56)

system, it can be observed that the RMSF values of residues 840-870 (correspond to H9, loop 843-849, H10), 790-800 (correspond to H7, loop 797-800) in system 2 were lower than apo-AR, which revealed the definite role of these residues in maintaining the system stability. Moreover, residues of binding site 2 (residues 880-903, correspond to H11, loop 888-893, H12) also exhibited obvious lower fluctuation comparing to apo-AR system. It can be observed that the binding site has become one of the most stable regions in system 2. These results not only indicated the dominant role of these residues, but also suggested that specific interactions must have been formed between residues 880-903 and DC3 residues, which then constrained the mobility of them and made the whole system stable.

### Clustering Analysis

The representative conformation of DC3-AR complex during MD simulation was extracted by clustering analysis. The first parallel trajectory of system 2 was grouped into 4 clusters based on the conformational similarity. The most populated cluster contained 14068 frames, which accounted for 56.27% of all frames extracted from the last 50 ns MD simulation. Then the conformation with least RMSD value in most populated cluster was defined as the representative structure. Here, the representative structure of DC3-AR complex extracted from MD simulations trajectory and the initial structure acquired from molecular docking were plotted in **Figures 7A,B**, respectively, to exhibit the interaction of key residues visually. From these



**FIGURE 5 |** Dynamic cross-correlation matrices (DCCM) of the C $\alpha$  atoms around their mean positions during the last 50 ns trajectory from the first parallel trajectory. The degrees of correlation and anti-correlation correspond to the color bar. **(A)** System 1. **(B)** System 2. **(C)** System 3. **(D)** System 4.

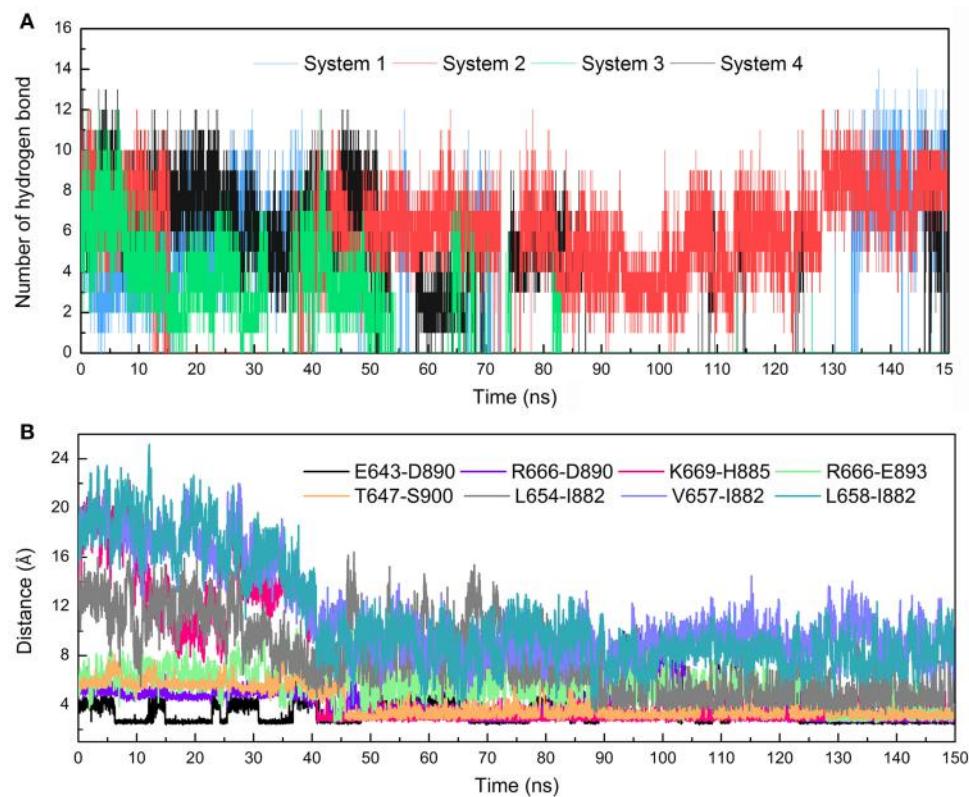
two figures, it can be seen that residues pairs K669-H885, R666-D890, S643-D890, and R666-E893 were apparently drawn near to each other during the MD simulations. All these indicated that some significant interaction forces might formed between these residues, which further stabilized the complex as RMSF data verified. Similarly, the closeness of residues L654-I882, V657-I882, L658-I882 could also be easily observed in MD simulation through **Figures 7C,D**. As residues L654, V657, L658, and I882 are nonpolar amino acid, it suggests specific hydrophobic interactions may formed, which needs to be further validated.

### Interaction Surface Exploration and Free Energy Decomposition

To figure out the binding mechanism between the key residues in the binding process of DC3-AR complex, hydrophobic and electrostatic interaction surfaces of representative structure were generated. The hydrophobic interaction surface was plotted in **Figure 8A**, where dodger blue represents hydrophobic minimum, gray depicts the hydrophobicity of 0, and orange represents the largest hydrophobicity. It can be observed that the binding site of AR indeed exists strong hydrophobic interactions with DC3, which promoted the identification and combination of ligand and receptor to a certain extent.

Furthermore, two highly hydrophobic interaction domains (deep orange) formed by V649, L650, L651-V901, and L654, V657, L658-L880, L881, I882, respectively, can be found, which played dominant roles in the development of hydrophobic interaction.

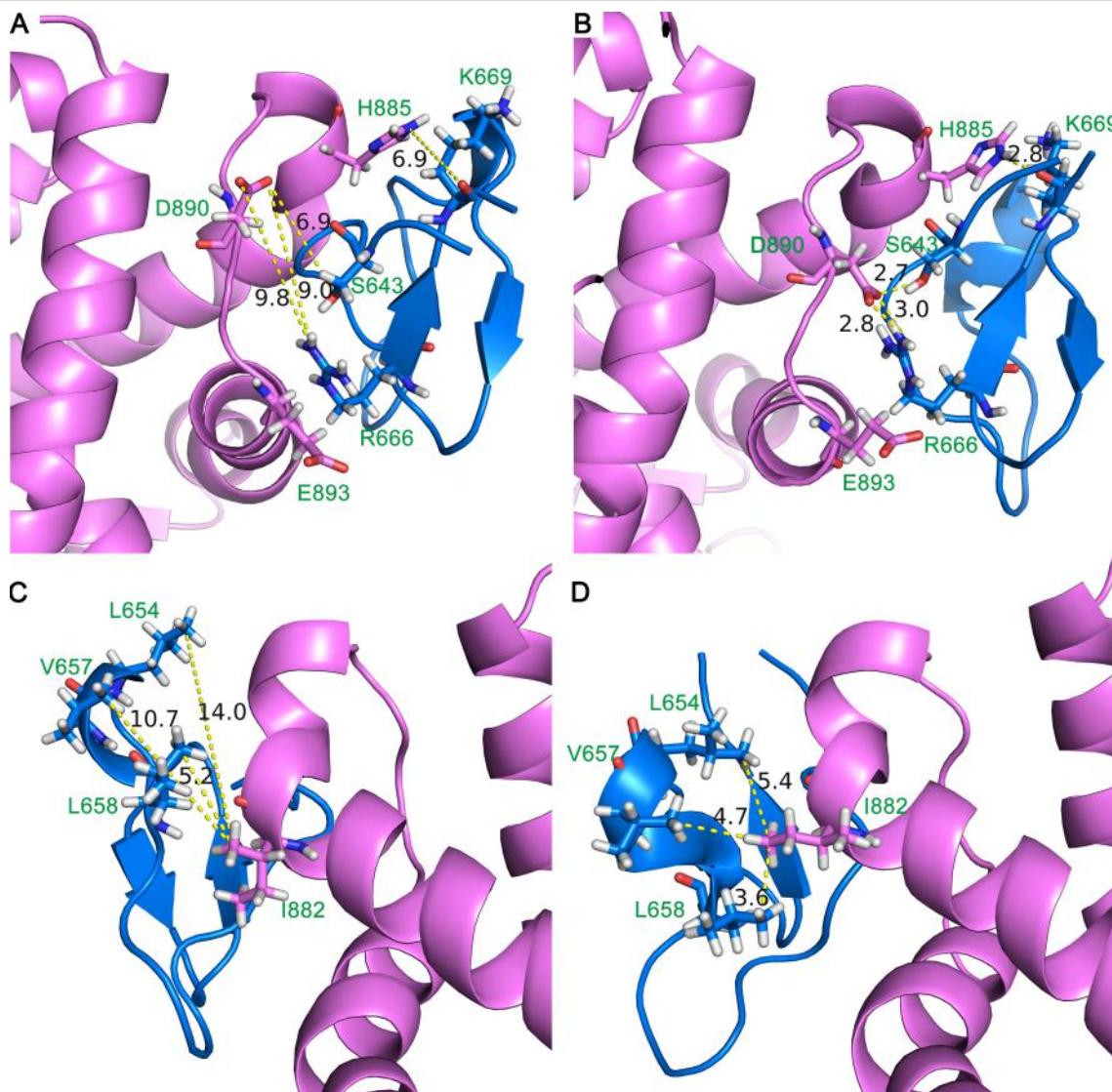
The electrostatic interaction surface was shown in **Figure 8B**. It can be seen that most DC3 residues carry positive charge, which come into being a positively charged surface (blue) in the interface. Whereas, a certain number of negatively charged residues (red) existed in binding site of AR. These residues with opposite charges in the interface attracted each other, and made great contribution to the binding process. To deeply investigate the energetic contribution, especially the electrostatic contribution of key residues, free-energy decomposition was performed based on the last 50 ns MD simulations of system 2. The energy contributions of DC3 and AR residues were shown in **Figures 9A,B**. The electrostatic contributions of DC3 and AR residues were depicted in **Figures 9C,D** respectively. From the energy contribution it can be seen that residue R666 made incomparable contribution to free-energy, which demonstrated that interactions existed between R666 and AR residues played essential roles in DC3-AR binding process. Meanwhile residues H885, D890, and S900 dominated the energetic contribution. These results proved that residues of both DC3 and AR in



**FIGURE 6 |** Hydrogen bonds and Distance analyses results. **(A)** Number of hydrogen bonds formed between DC3 and AR vs. simulation time calculated from the first parallel MD simulation of studied systems. **(B)** Distance of key residue pairs vs. simulation time calculated from the first parallel MD simulation of system 2.

**TABLE 3 |** The hydrogen bonds formed between DC3 and AR with occupation more than 10% in the last 50 ns MD simulations of four systems.

Complex	Acceptor	Donor	Frames	Frac	AvgDist	AvgAng
System 1	F916@O	C662@N	5878	0.2351	2.8803	161.0701
	S853@OG	D670@N	4556	0.1822	3.0546	153.5131
	S656@OG	T918@N	3965	0.1586	3.1798	154.4460
	C653@O	T918@OG1	3145	0.1258	2.7764	161.6225
System 2	K669@O	H885@NE2	20880	0.8352	2.9836	148.9214
	D890@OD1	R666@NH2	18635	0.7454	2.8981	155.1367
	S900@OG	T647@N	18123	0.7249	3.1593	156.3738
	D890@OD2	R666@NH1	16246	0.6498	2.8350	159.6511
	D890@OD2	S643@OG	16094	0.6438	2.6662	164.7336
	D890@OD2	R666@NH2	15971	0.6388	3.1114	140.8035
	E893@OE2	R666@NE	8377	0.3351	3.0023	148.8850
	E893@OE1	R666@NE	8179	0.3272	3.0252	148.1640
	D890@OD1	S643@OG	6708	0.2683	2.7365	161.1178
	E893@OE1	R666@NH2	5978	0.2391	2.8769	153.9622
	F891@O	R666@NH2	5776	0.2310	2.8978	140.2469
	E893@OE2	R666@NH2	5165	0.2066	2.8858	153.5793
	D890@OD1	R666@NH1	3571	0.1428	3.2415	137.6629
System 3	T755@O	C653@N	3137	0.1255	2.9959	158.6020
System 4	-	-	-	-	-	-



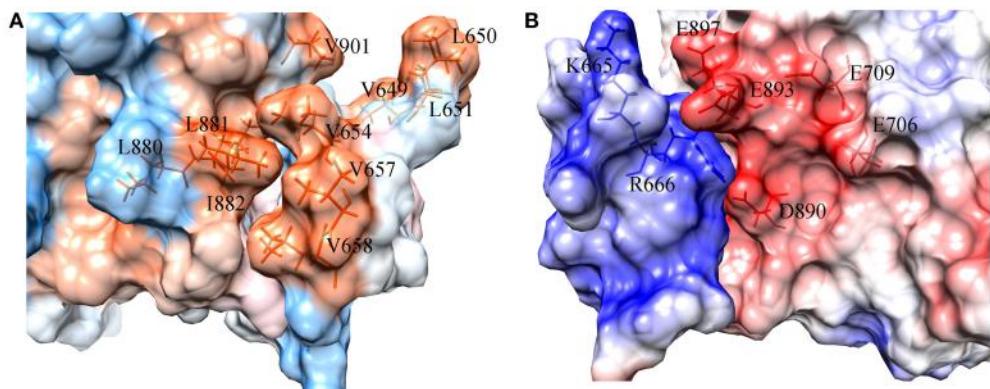
**FIGURE 7 |** Initial structure (**A,C**) and representative structure (**B,D**) of the binding site of DC3-AR complex. DC3 is shown in marine and AR is shown in violet. Key residues are shown in stick, and yellow dashed lines represent the distance of specific residue pairs.

binding site do make contributions in the decrease of free energy. Combined with the hydrogen bond analyzed before, it comes to a conclusion that hydrogen bonds formed between K669-H885, D890-R666, S900-T647, and D890-E643 are especially critical components to the interaction between DC3 and AR. Based on all the results above, we can reach a conclusion that K665, R666 of DC3 and E706, E709, D890, E893, E897 of AR ultimately make great contributions to the binding process.

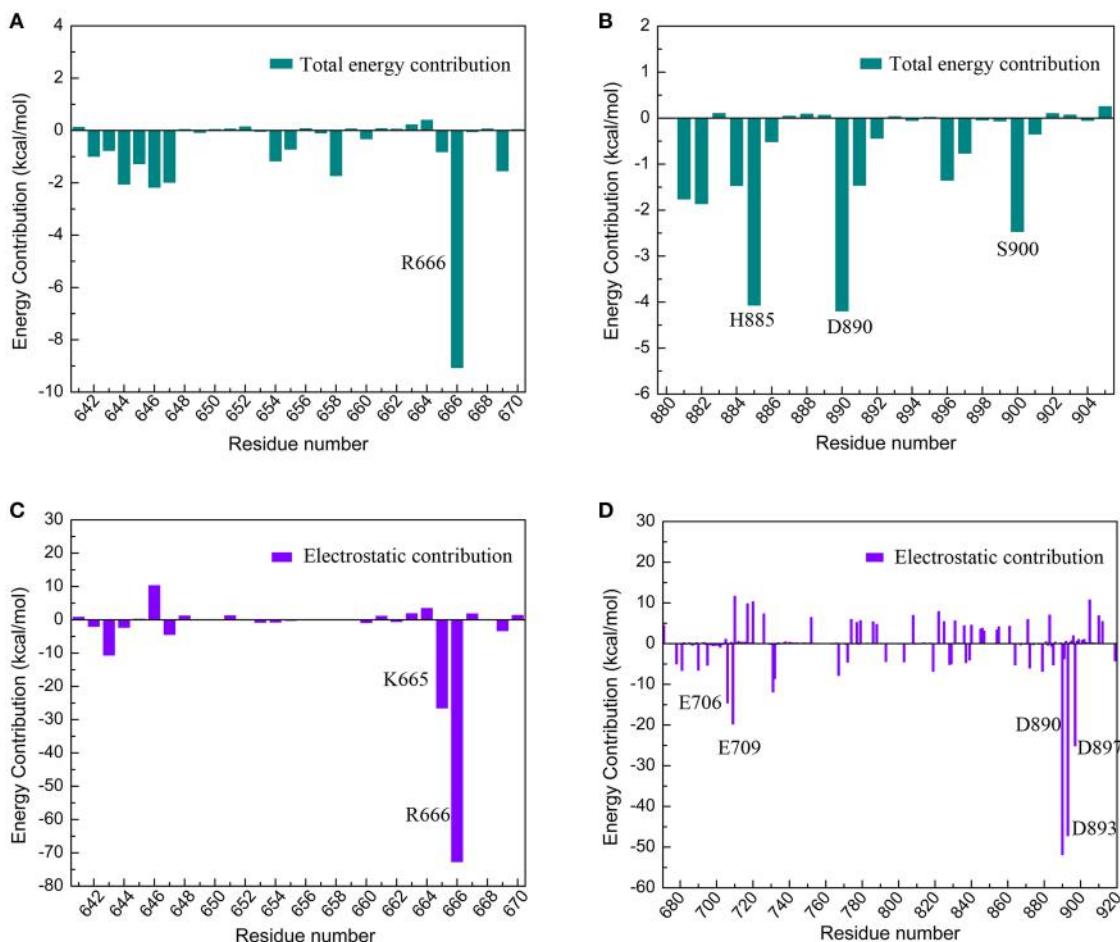
### Distance Analysis

To further validate the interaction formed between key residues and investigate the formation process, the distances of key residue pairs mentioned above vs. simulation time were calculated and plotted in **Figure 6B**. It can be firstly observed

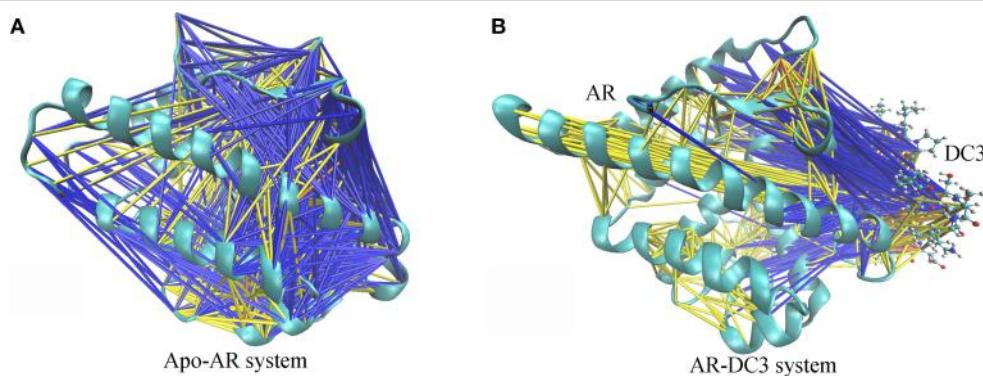
that the distance of residue pairs K669-H885, L654-I882, V657-I882, and L658-I882 experienced an obvious decrease in about 40 ns and maintained stable in the later simulation. Significant conformational changes of binding site could be speculated based on this crucial distance variation. This result revalidated the formation of hydrogen bond between K669-H885, and hydrophobic interaction between L645-I882, V657-I882, L658-I882. In addition, residue pairs E643-D890, T647-S900, R666-D890, and R666-E893 kept highly close (about 5 Å) and remain stable throughout the simulation process, and some of them even reached about 2.5 Å in the last 50 ns. Besides, from the hydrogen bonds figure shown in **Figure 6A**, it can also be seen that certain hydrogen bonds formed in about 40 ns. All these results proved the stable existence of hydrogen bonds and electrostatic interaction during the whole MD simulations.



**FIGURE 8 |** Surface of hydrophobic interaction **(A)** and electrostatic interaction **(B)** in the binding site of DC3-AR complex. **(A)** Dodger blue represents hydrophobic minimum, orange represents hydrophobic maximum, key residues are indicated as orange sticks. **(B)** Positively charged domain colored by blue, negatively charged domain colored by red, key residues are indicated as blue and red sticks.



**FIGURE 9 |** Free-energy decomposition of DC3-AR complex of the first parallel MD simulation in the last 50 ns MD simulation. **(A)** Energy contribution of DC3 residues. **(B)** Energy contribution of AR residues in the binding site. **(C)** Electrostatic contribution of DC3 residues. **(D)** Electrostatic contribution of AR residues.



**FIGURE 10 |** Cross-correlation networks of apo-AR system **(A)** and DC3-AR system **(B)**. AR is shown in cyan cartoon and DC3 is shown as CPK model. Cross-correlation of residue pairs are distinguished by colors. Correlation are indicated by yellow ( $0.3 \leq C_{ij} < 0.5$ ), orange ( $0.5 \leq C_{ij} < 0.7$ ) and red ( $C_{ij} \geq 0.7$ ). Anti-correlation are indicated by blue ( $-0.5 < C_{ij} \leq -0.3$ ) and black ( $-0.7 < C_{ij} \leq -0.5$ ).

### Cross-Correlation Networks

To characterize and intuitively exhibit the underlying dynamical cross-correlations among different parts of DC3-AR complex, the overall cross-correlation networks of system 2 and apo-AR system were constructed. As shown in **Figure 10A**, masses of correlation and anti-correlation both widely and simultaneously existed in apo-AR system, which made it hard to identify the specific cross-correlation pattern, in other words, the interaction between different protein domains was mixed and disorderly. However, when DC3 binding to AR (**Figure 10B**), all anti-correlation among different AR regions decreased or disappeared, which indicated the improvements of system stability. Besides, organized anti-correlation developed between DC3 and AR regions (loop 687-690, H3, loop 722-725, loop 727-730, H4, loop 822-825, H9, H11, loop 888-893, H12). It reflected the opposite movement tendency between these regions and DC3, namely, they moved close to each other along with the simulation. Moreover, distinct correlation also formed between DC3 and some AR regions (H9, loop 843-849, H10). Combining previous study that residues in these regions had low RMSF values, it can be concluded that this correlation patterns decreased residues fluctuation and enhanced the system stability.

### CONCLUSION

In this work, the three-dimensional structure of cyclopeptide DC3 was firstly constructed by homology modeling technology using 2kux.1.A as template. Then molecular docking was carried out to predict possible binding site and preferred orientation of DC3 into AR. Finally, four systems with best docking score from top four clusters were selected to perform 150 ns all-atom molecular dynamics (MD) simulations. The MM/GBSA method and a series of MD trajectory analyses were subsequently conducted. The analyses of RMSF, binding free energy, DCCM and hydrogen bonds indicated that DC3 showed a higher binding affinity to AR in site 2 (corresponding to H10, H11, loop888-893, H12) and this system showed obvious superiority in stability comparing to other systems. Besides, much more intermolecular

hydrogen bonds were constantly formed in system 2 with high occupation. Stronger cross-correlation among DC3 residues and stronger anti-correlation between DC3 and AR residues also existed here. These results suggest that DC3 is most likely to bind to AR in site 2 encompassed by H10, H11, loop888-893, and H12. Subsequently, combining further analysis of free-energy decomposition, interaction surface, distance, and cross-correlation network, it can be observed that hydrogen bonds, hydrophobic, and electrostatic interactions play dominant roles in the recognition and combination of DC3-AR complex. For hydrogen bonds, it frequently existed between K669-H885, D890-R666, S900-T647, and D890-E643. Besides, K665, R666 of DC3, and E706, E709, D890, E893, E897 of AR made great contributions to electrostatic interaction values. V649, V650, V651-V801, and L654, V657, L658-L880, L881 play essential parts of hydrophobic interaction. These results elucidated the detail interaction mechanism of DC3-AR complex and the key residues dominated in specific interaction. These findings will significantly facilitate our understanding of action mode of DC3 to AR at the molecular level, and contribute to the future rational cyclopeptide drug design for prostate cancer.

### AUTHOR CONTRIBUTIONS

JL: conceived and coordinated the study; HZ and TS: did the homology modeling, molecular docking, molecular dynamics simulations, and they wrote the paper; YY and CF: helped to do molecular simulations. All authors analyzed the results and approved the final version of the manuscript.

### ACKNOWLEDGMENTS

This research is supported by the Fundamental Research Funds for the Central Universities (lzujbky-2017-203). We would like to thank Prof. Yuguang Mu and Dr. Liangzhen Zheng from School of Biological Sciences of Nanyang Technological University for helping us to treat the data for the figure of cross-correlation network analysis.

## REFERENCES

- Accelrys Inc. CA (2009). *Discovery Studio Version 2.5*.
- American Cancer Society (2015). *Global Cancer Facts and Figures, 3rd Edn.* Atlanta: American Cancer Society.
- Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22, 195–201. doi: 10.1093/bioinformatics/bti770
- Benkert, P., Biasini, M., and Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 27, 343–350. doi: 10.1093/bioinformatics/btq662
- Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., et al. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 42, W252–W258. doi: 10.1093/nar/gku340
- Bienert, S., Waterhouse, A., de Beer, T. A., Tauriello, G., Studer, G., Bordoli, L., et al. (2016). The SWISS-MODEL repository-new features and functionality. *Nucleic Acids Res.* 45, D313–D319. doi: 10.1093/nar/gkw1132
- Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., and Schwede, T. (2009). Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* 4, 1–13. doi: 10.1038/nprot.2008.197
- Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C. L., Wang, J., Duke, R. E., et al. (2012). *AMBER 12*. San Francisco, CA: University of California.
- Center, M. M., Jemal, A., Lortet-Tieulent, J., Ward, E., Ferlay, J., Brawley, O., et al. (2012). International variation in prostate cancer incidence and mortality rates. *Eur. Urol.* 61, 1079–1092. doi: 10.1016/j.eururo.2012.02.054
- Chen, F., Liu, H., Sun, H., Pan, P., Li, Y., Li, D., et al. (2016). Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein-protein binding free energies and re-rank binding poses generated by protein-protein docking. *Phys. Chem. Chem. Phys.* 18, 22129–22139. doi: 10.1039/C6CP03670H
- Craik, D. J., Daly, N. L., Bond, T., and Waine, C. (1999). Plant cyclotides: a unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. *J. Mol. Biol.* 294, 1327–1336. doi: 10.1006/jmbi.1999.3383
- Darden, T., York, D., Pedersen, L., Darden, T. A., York, D. M., Pedersen, L. G., et al. (1993). Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi: 10.1063/1.464397
- Fischer, N. M., van Maaren, P. J., Ditz, J. C., Yildirim, A., and van der Spoel, D. (2015). Properties of organic liquids when simulated with long-range Lennard-Jones interactions. *J. Chem. Theory Comput.* 11, 2938–2944. doi: 10.1021/acs.jctc.5b00190
- Fu, T., Wu, X., Xiu, Z., Wang, J., Yin, L., and Li, G. (2013). Understanding the molecular mechanism of binding modes of aurora a inhibitors by long time scale Gpu dynamics. *J. Theor. Comput. Chem.* 12:1341003. doi: 10.1142/S0219633613410034
- Ghosh, A., and Vishveshwara, S. (2007). A study of communication pathways in methionyl- tRNA synthetase by molecular dynamics simulations and structure network analysis. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15711–15716. doi: 10.1073/pnas.0704459104
- Guex, N., Peitsch, M. C., and Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* 30 (Suppl. 1), S162–S173. doi: 10.1002/elps.200900140
- Heinlein, C. A., and Chang, C. (2004). Androgen receptor in prostate cancer. *Endocr. Rev.* 25, 276–308. doi: 10.1210/er.2002-0032
- Hitzenberger, M., Schuster, D., and Hofer, T. S. (2017). The binding mode of the sonic hedgehog inhibitor robotnikinin, a combined docking and QM/MM MD study. *Front. Chem.* 5:76. doi: 10.3389/fchem.2017.00076
- Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65, 712–725. doi: 10.1002/prot.21123
- Hou, T., Wang, J., Li, Y., and Wang, W. (2011a). Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* 51, 69–82. doi: 10.1021/ci100275a
- Hou, T., Wang, J., Li, Y., and Wang, W. (2011b). Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. *J. Comput. Chem.* 32, 866–877. doi: 10.1002/jcc.21666
- Hsin, J., Arkhipov, A., Yin, Y., Stone, J. E., and Schulten, K. (2008). Using VMD: an introductory tutorial. *Curr. Protoc. Bioinformatics* Chapter 5, Unit 5.7. doi: 10.1002/0471250953.bi0507s24
- Hu, E., Wang, D., Chen, J., and Tao, X. (2015). Novel cyclotides from *Hedyotisdiffusa* induce apoptosis and inhibit proliferation and migration of prostate cancer cells. *Int. J. Clin. Exp. Med.* 8, 4059–4065.
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5
- Jitonnom, J., Lomthaisong, K., and Lee, V. S. (2012). Computational design of peptide inhibitor based on modifications of proregion from *Plutellaxylostella* midgut trypsin. *Chem. Biol. Drug Des.* 79, 583–593. doi: 10.1111/j.1747-0285.2011.01312.x
- Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., and Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.* 37, D387–D392. doi: 10.1093/nar/gkn750
- Kollman, P. A., Massoval, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., et al. (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33, 889–897. doi: 10.1021/ar000033j
- Lange, O. F., Grubmüller, H., and de Groot, B. L. (2005). Molecular dynamics simulations of protein G challenge NMR-derived correlated backbone motions. *Angew. Chem. Int. Ed. Engl.* 44, 3394–3399. doi: 10.1002/anie.200462957
- Lee, H. Z., Bau, D. T., Kuo, C. L., Tsai, R. Y., Chen, Y. C., and Chang, Y. H. (2011). Clarification of the phenotypic characteristics and anti-tumor activity of *Hedyotisdiffusa*. *Am. J. Chin. Med.* 39, 201–213. doi: 10.1142/S0192415X11008750
- Lin, C. C., Kuo, C. L., Lee, M. H., Hsu, S. C., Huang, A. C., Tang, N. Y., et al. (2011). Extract of *Hedyotisdiffusa* willd influences murine leukemia WEHI-3 Cells *in vivo* as well as promoting T- and B-Cell proliferation in leukemic mice. *In Vivo* 25, 633–640.
- Lin, J., Chen, Y., Wei, L., Chen, X., Xu, W., Hong, Z., et al. (2010). *Hedyotis Diffusa* willd extract induces apoptosis via activation of the mitochondrion-dependent pathway in human colon carcinoma cells. *Int. J. Oncol.* 37, 1331–1338.
- Liu, H., An, X., Li, S., Wang, Y., Li, J., and Liu, H. (2015). Interaction mechanism exploration of R-bicalutamide/S-1 with WT/W741L AR using molecular dynamics simulations. *Mol. Biosyst.* 11, 3347–3354. doi: 10.1039/C5MB00499C
- Liu, M., Wang, L., Sun, X., and Zhao, X. (2014). Investigating the impact of Asp181 point mutations on interactions between PTP1B and phosphotyrosine substrate. *Sci. Rep.* 4:5095. doi: 10.1038/srep05095
- Liu, Z., Liu, M., Liu, M., and Li, J. (2010). Methylanthraquinone from *Hedyotisdiffusa* willd induces  $\text{Ca}^{2+}$ -mediated apoptosis in human breast cancer cells. *Toxicol. In Vitro* 24, 142–147. doi: 10.1016/j.tiv.2009.08.002
- Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schütz, G., Umesono, K., et al. (1995). The nuclear receptor superfamily: the second decade. *Cell* 83, 835–839. doi: 10.1016/0092-8674(95)90199-X
- Meng, X. Y., Zhang, H. X., Mezei, M., and Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Curr. Comput. Aided Drug Des.* 7, 146–157. doi: 10.2174/157340911795677602
- Nuclear Receptors Nomenclature Committee (1999). A unified nomenclature system for the nuclear receptor superfamily. *Cell* 97, 161–163. doi: 10.1016/S0092-8674(00)80726-6
- Onufriev, A., Bashford, D., and Case, D. A. (2004). Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* 55, 383–394. doi: 10.1002/prot.20033
- Palmbos, P. L., and Hussain, M. (2013). Non-castrate metastatic prostate cancer: have the treatment options changed? *Semin. Oncol.* 40, 337–346. doi: 10.1053/j.seminoncol.2013.04.007
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. doi: 10.1002/jcc.20084
- Plan, M. R., Rosengren, K. J., Sando, L., Daly, N. L., and Craik, D. J. (2010). Structural and biochemical characteristics of the cyclotides from *Oldenlandiaaffinis*. *Biopolymers* 94, 647–658. doi: 10.1002/bip.21409
- Punkvong, A., Kamsri, P., Saparpakorn, P., Hannongbua, S., Wolschann, P., Irle, S., et al. (2015). Key structures and interactions for binding of mycobacterium

- tuberculosis protein kinase b inhibitors from molecular dynamics simulation. *Chem. Biol. Drug Des.* 86, 91–101. doi: 10.1111/cbdd.12465
- Ramírez, D., and Caballero, J. (2016). Is it reliable to use common molecular docking methods for comparing the binding affinities of enantiomer pairs for their protein target? *Int. J. Mol. Sci.* 17:E525. doi: 10.3390/ijms17040525
- Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977). Numerical integration of a system with constraints: of the cartesian equations of motion molecular dynamics of n-alkanes. *J. Comput. Phys.* 23, 327–341. doi: 10.1016/0021-9991(77)90098-5
- Schröder, F. H. (2008). Progress in understanding androgen-independent prostate cancer (AIPC): a review of potential endocrine-mediated mechanisms. *Eur. Urol.* 53, 1129–1137. doi: 10.1016/j.eururo.2008.01.049
- Siegel, R. L., Miller, K. D., and Jemal, A. (2016). Cancer statistics, 2016. *CA Cancer J. Clin.* 66, 7–30. doi: 10.3322/caac.21332
- Sun, H., Li, Y., Shen, M., Tian, S., Xu, L., Pan, P., et al. (2014a). Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring. *Phys. Chem. Chem. Phys.* 16, 22035–22045. doi: 10.1039/C4CP03179B
- Sun, H., Li, Y., Tian, S., Xu, L., and Hou, T. (2014b). Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. *Phys. Chem. Chem. Phys.* 16, 16719–16729. doi: 10.1039/C4CP01388C
- Sze, S. K., Wang, W., Meng, W., Yuan, R., Guo, T., Zhu, Y., et al. (2009). Elucidating the structure of cyclotides by partial acid hydrolysis and LC-MS/MS analysis. *Anal. Chem.* 81, 1079–1088. doi: 10.1021/ac802175r
- Tan, M. H., Li, J., Xu, H. E., Melcher, K., and Yong, E. L. (2015). Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacol. Sin.* 36, 3–23. doi: 10.1038/aps.2014.18
- Topham, C. M., Thomas, P., Overington, J. P., Johnson, M. S., Eisenmenger, F., and Blundell, T. L. (1990). An assessment of COMPOSER: a rule-based approach to modelling protein structure. *Biochem. Soc. Symp.* 57, 1–9.
- Tsai, M. J., and O’Malley, B. W. (1994). Molecular mechanisms of action of steroid/thyroid receptor superfamily members. *Annu. Rev. Biochem.* 63, 451–486. doi: 10.1146/annurev.bi.63.070194.002315
- van der Spoel, D., and van Maaren, P. J. (2006). The origin of layer structure artifacts in simulations of liquid water. *J. Chem. Theory Comput.* 2, 1–11. doi: 10.1021/ct0502256
- Veldscholte, J., Berrevoets, C. A., Brinkmann, A. O., Grootegoed, J. A., and Mulder, E. (1992). Anti-androgens and the mutated androgen receptor of Incap cells: differential effects on binding affinity, heat-shock protein interaction, and transcription activation. *Biochemistry* 31, 2393–2399. doi: 10.1021/bi00123a026
- Wang, Q., Ning, L., Niu, Y., Liu, H., and Yao, X. (2015). Molecular mechanism of the inhibition and remodeling of human islet amyloid polypeptide (hIAPP(1–37)) oligomer by resveratrol from molecular dynamics simulation. *J. Phys. Chem. B* 119, 15–24. doi: 10.1021/jp507529f
- Wang, Z., Chen, G., Chen, L., Liu, X., Fu, W., Zhang, Y., et al. (2015). Insights into the binding mode of curcumin to MD-2: studies from molecular docking, molecular dynamics simulations and experimental assessments. *Mol. Biosyst.* 11, 1933–1938. doi: 10.1039/C5MB00085H
- Xu, L., Sun, H., Li, Y., Wang, J., and Hou, T. (2013). Assessing the performance of mm/pbsa and mm/gbsa methods. 3. the impact of force fields and ligand charge models. *J. Phys. Chem. B* 117, 8408–8421. doi: 10.1021/jp404160y
- Yamada, A., Fujii, S., Mori, S., and Kagechika, H. (2013). Design and synthesis of 4-(4-Benzoylaminophenoxy)phenol derivatives as androgen receptor antagonists. *ACS Med. Chem. Lett.* 4, 937–941. doi: 10.1021/ml4001744
- Yamamoto, S., Matsunaga, N., Hitaka, T., Yamada, M., Hara, T., Miyazaki, J., et al. (2012). Design, synthesis, and biological evaluation of 4-phenylpyrrole derivatives as novel androgen receptor antagonists. *Bioorg. Med. Chem.* 20, 422–434. doi: 10.1016/j.bmc.2011.10.067
- Yamaoka, M., Hara, T., and Kusaka, M. (2010). Overcoming persistent dependency on androgen signaling after progression to castration-resistant prostate cancer. *Clin. Cancer Res.* 16, 4319–4324. doi: 10.1158/1078-0432.CCR-10-0255
- Yuriev, E., Holien, J., and Ramsland, P. A. (2015). Improvements, trends, and new ideas in molecular docking: 2012–2013 in review. *J. Mol. Recognit.* 28, 581–604. doi: 10.1002/jmr.2471
- Zhou, J., Liu, B., Geng, G., and Wu, J. H. (2010). Study of the impact of the T877A mutation on ligand-induced helix-12 positioning of the androgen receptor resulted in design and synthesis of novel antiandrogens. *Proteins* 78, 623–637. doi: 10.1002/prot.22592

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Zhang, Song, Yang, Fu and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Conceptual DFT Study of the Local Chemical Reactivity of the Colored BISARG Melanoidin and Its Protonated Derivative

Juan Frau<sup>1</sup> and Daniel Glossman-Mitnik<sup>1,2\*</sup>

<sup>1</sup> Departament de Química, Universitat de les Illes Balears, Palma de Mallorca, Spain, <sup>2</sup> Laboratorio Virtual NANOCOSMOS, Centro de Investigación en Materiales Avanzados, Departamento de Medio Ambiente y Energía, Chihuahua, Mexico

## OPEN ACCESS

### Edited by:

Sam P. De Visser,  
University of Manchester,  
United Kingdom

### Reviewed by:

Carles Curutchet,  
Universitat de Barcelona, Spain  
Mark Earl Casida,  
Université Grenoble Alpes, France

### \*Correspondence:

Daniel Glossman-Mitnik  
daniel.glossman@cimav.edu.mx

### Specialty section:

This article was submitted to  
Theoretical and Computational  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

Received: 14 February 2018

Accepted: 09 April 2018

Published: 01 May 2018

### Citation:

Frau J and Glossman-Mitnik D (2018)  
Conceptual DFT Study of the Local  
Chemical Reactivity of the Colored  
BISARG Melanoidin and Its  
Protonated Derivative.  
Front. Chem. 6:136.  
doi: 10.3389/fchem.2018.00136

This computational study assessed eight fixed RSH (range-separated hybrid) density functionals that include CAM-B3LYP, LC- $\omega$ PBE, M11, MN12SX, N12SX,  $\omega$ B97,  $\omega$ B97X, and  $\omega$ B97XD related to the Def2TZVP basis sets together with the SMD solvation model in the calculation the molecular structure and reactivity properties of the BISARG intermediate melanoidin pigment (5-(2-(E)-(Z)-5-[(2-furyl)methylidene]-3-(4-acetylamino-4-carboxybutyl)-2-imino-1,3-dihydroimidazol-4-ylideneamino(E)-4-[(2-furyl)methylidene]-5-oxo-1H-imidazol-1-yl)-2-acetylaminovaleric acid) and its protonated derivative, BISARG(p). The chemical reactivity descriptors for the systems were calculated via the Conceptual Density Functional Theory. The choice of active sites applicable to nucleophilic, electrophilic as well as radical attacks were made by linking them with Fukui functions indices, electrophilic and nucleophilic Parr functions, and the condensed Dual Descriptor  $\Delta f(r)$ . The study found the MN12SX and N12SX density functionals to be the most appropriate in predicting the chemical reactivity of the molecular systems under study starting from the knowledge of the HOMO, LUMO, and HOMO-LUMO gap energies.

**Keywords:** BISARG, conceptual DFT, chemical reactivity, dual descriptor, Parr functions

## 1. INTRODUCTION

Visual color in processed foods is largely due to colored products of Maillard or nonenzymic browning reactions. In spite of the longstanding aesthetic and practical interest in Maillard derived food coloring materials, relatively little is known about the chemical structures responsible for visual color (Rizzi, 1997). These chemical structures are known as Colored Maillard Reaction Products and can be isolated at intermediate stages during the melanoidin formation process.

Besides their interest as dye molecules which may be useful as food additives, but also as dyes for dye-sensitized solar cells (DSSC), these compounds have also antioxidant capabilities. Thus, they are amenable to be studied by analyzing their molecular reactivity properties.

One of these isolated molecules is named by the acronym BISARG and together with its protonated derivative, BISARG(p) have been experimentally studied as a part of a work related to the formation of melanoidins (Hofmann, 1998) and we believe that it could be of interest to study their molecular reactivity by using the ideas of Conceptual DFT, in the same way of our previous works (Alvarado-González et al., 2013; Cervantes-Navarro and Glossman-Mitnik, 2013; Glossman-Mitnik, 2013a,b; Martínez-Araya et al., 2013a,b; Salgado-Morán et al., 2013; Glossman-Mitnik, 2014a,b,c,d; Martínez-Araya and Glossman-Mitnik, 2015;

Martínez-Araya et al., 2015; Soto-Rojo et al., 2015; Frau et al., 2016a,b,c; Mendoza-Huizar et al., 2016; Frau et al., 2017a,b,c,d,e; Frau and Grossman-Mitnik, 2017a,b,c,d,e,f,g; Sastre et al., 2017).

The interest in using range-separated (RS) exchange correlation functionals in KS DFT is on the rise (Gledhill et al., 2016). The functionals tend to partition the  $r_{12}^{-1}$  operator and exchange them into long- and short-range parts, whose range separation parameter  $\omega$  controls the rate of attaining the long-range behavior. It is possible to fix the value of  $\omega$ . The value can also be nonempirically “tuned” through a system-by-system mechanism that minimizes some tuning norms. The basis of the optimal tuning approach is the knowledge that the energy that the HOMO should have,  $\epsilon_H(N)$ , in exact KS as well as generalized KS theory for an  $N$  electron system, ought to be exactly  $-IP(N)$ . Hence, IP represents the vertical ionization potential that is calculated by considering a particular functional energy difference  $E(N-1) - E(N)$ . If approximate functionals are used, it is possible to have considerable differences between  $\epsilon_H(N)$ , and  $-IP(N)$ . Optimal tuning constitutes determining a system-specific range-separation parameter  $\omega$  non-empirically in an RSE functional. Optionally, it also implied that several other parameters including  $\epsilon_H(N) = -IP(N)$  are satisfied optimally (Jacquemin et al., 2014). Even though no equivalency exists to match this prescription of electron affinity (EA) coupled with LUMO in the case of neutral species, it is possible to say that  $\epsilon_H(N+1) = -EA(N)$ , that is, the electron affinity of the neutral system is equal to minus the HOMO energy of the anion (SOMO), which facilitates the finding of an optimized value of  $\omega$ , and is then optimized to establish both properties simultaneously. Some concerns have been raised during the preparation of this paper regarding the validity of the ionization potential theorem (IP) within the context of Generalized Kohn-Sham (GKS) theory. However, it must be stressed that Baer et al. (2010) and more recently Baerends et al. (2013) and Karolewski et al. (2013) have given arguments that the same criterion applies in GKS theories and with hybrid and range-separated hybrid functionals. This will make it easy to predict the Conceptual DFT descriptors. In the past, the simultaneous prescription has been referred to as the “KID procedure” (for Koopmans in DFT), courtesy of the analogy it shares with the Koopmans’ theorem within the Hartree-Fock theory. This SOMO energy will not be, in general, equal to the LUMO of the neutral, but if the difference between them, which we have called  $\Delta SL$ , is small enough to be considered negligible for predictions of the Conceptual DFT descriptors, then the practical KID procedure will have a computational support.

This implies that the appropriateness of a particular density functional in making predictions of the Conceptual DFT descriptors directly by relying on the properties that the neutral molecule can be easily estimated. It only requires one to check the way that it has followed the KID procedure. Nevertheless, tune-optimization depends on the system and must be performed for each molecule one at a time. Therefore, examining the various density functionals exhibiting significant accuracy across various types of databases in physics, chemistry, and where the  $\omega$  value is fixed will determine how they perform the practical technique.

Thus, in this computational study we will assess eight density functionals in calculating the molecular properties and structure of the BISARG intermediate melanoidin pigment and its protonated derivative, BISARG(p). Following the same ideas of previous works, we will consider fixed RSH functional instead of the optimally-tuned RSH density functionals that have attained great success and have also supported the validity of the IP theorem in the context of the GKS theory (Stein et al., 2009a,b; Karolewski et al., 2011; Kuritz et al., 2011; Refaeli-Abramson et al., 2011; Foster and Wong, 2012; Koppen et al., 2012; Kronik et al., 2012; Phillips et al., 2012a,b; Karolewski et al., 2013; Moore and Autschbach, 2013; Egger et al., 2014; Foster et al., 2014; Jacquemin et al., 2014; Niskanen and Hukka, 2014; Sun and Autschbach, 2014; Manna et al., 2015; Lima et al., 2016; Pereira et al., 2017).

## 2. THEORETICAL BACKGROUND

The theoretical background of this study is similar to the previous conducted research presented complete purposes, because this research is a component of a major project that it is in progress. If we consider the KID procedure mentioned in the Introduction together with a finite difference approximation, then the global reactivity descriptors can be written as:

Electronegativity	$\chi = -\frac{1}{2}(I + A) \approx \frac{1}{2}(\epsilon_L + \epsilon_H)$	Parr and Yang (1984)
Global Hardness	$\eta = (I - A) \approx (\epsilon_L - \epsilon_H)$	Parr and Yang (1984)
Electrophilicity	$\omega = \frac{\mu^2}{2\eta} = \frac{(I+A)^2}{4(I-A)} \approx \frac{(\epsilon_L + \epsilon_H)^2}{4(\epsilon_L - \epsilon_H)}$	Parr et al. (1999)
Electrodonating Power	$\omega^- = \frac{(3I+A)^2}{16(I-A)} \approx \frac{(3\epsilon_H + \epsilon_L)^2}{16\eta}$	Gázquez et al. (2007)
Electroaccepting Power	$\omega^+ = \frac{(I+3A)^2}{16(I-A)} \approx \frac{(\epsilon_H + 3\epsilon_L)^2}{16\eta}$	Gázquez et al. (2007)
Net electrophilicity	$\Delta\omega^\pm = \omega^+ - (-\omega^-) = \omega^+ + \omega^-$	Chattaraj et al. (2009)

where  $\epsilon_H$  and  $\epsilon_L$  are the energies of the highest occupied and the lowest unoccupied molecular orbitals (HOMO and LUMO), respectively.

Applying the same ideas, the definitions for the local reactivity descriptors are:

Nucleophilic Fukui Function	$f^+(\mathbf{r}) = \rho_{N+1}(\mathbf{r}) - \rho_N(\mathbf{r})$	Parr and Yang (1984)
Electrophilic Fukui Function	$f^-(\mathbf{r}) = \rho_N(\mathbf{r}) - \rho_{N-1}(\mathbf{r})$	Parr and Yang (1984)
Dual Descriptor	$\Delta f(\mathbf{r}) = \left( \frac{\partial f(\mathbf{r})}{\partial N} \right)_{V(\mathbf{r})}$	Morell et al. (2005, 2006)
Nucleophilic Parr Function	$P^-(\mathbf{r}) = \rho_S^{FC}(\mathbf{r})$	Domingo et al. (2013)
Electrophilic Parr Function	$P^+(\mathbf{r}) = \rho_S^{FA}(\mathbf{r})$	Domingo et al. (2013)

where  $\rho_{N+1}(\mathbf{r})$ ,  $\rho_N(\mathbf{r})$ , and  $\rho_{N-1}(\mathbf{r})$  are the electronic densities at point  $\mathbf{r}$  for the system with  $N + 1$ ,  $N$ , and  $N - 1$  electrons,

respectively, and  $\rho_s^{rc}(\mathbf{r})$  and  $\rho_s^{ra}(\mathbf{r})$  are related to the atomic spin density (ASD) at the  $\mathbf{r}$  atom of the radical cation or anion of a given molecule, respectively (Domingo et al., 2016).

### 3. SETTINGS AND COMPUTATIONAL METHODS

Following the lines of our previous works, the computational studies were performed with the Gaussian 09 (Frisch et al., 2018) series of programs with density functional methods as implemented in the computational package. The basis set used in this work was Def2SVP for geometry optimization and frequencies, while Def2TZVP was considered for the calculation of the electronic properties (Weigend and Ahlrichs, 2005; Weigend, 2006). All the calculations were performed in the presence of water as the solvent by doing Integral Equation Formalism-Polarized Continuum Model (IEF-PCM) computations according to the SMD solvation model (Marenich et al., 2009).

For the calculation of the molecular structure and properties of the studied systems, we have chosen eight density functionals which are known to consistently provide satisfactory results for several structural and thermodynamic properties:

CAM-B3LYP	Long-range-corrected B3LYP by the CAM method	Yanai et al. (2004)
LC- $\omega$ PBE	Long-range-corrected $\omega$ PBE density functional	Henderson et al. (2009)
M11	Range-separated hybrid meta-GGA	Peverati and Truhlar (2011)
MN12SX	Range-separated hybrid nonseparable meta-NGA	Peverati and Truhlar (2012)
N12SX	Range-separated hybrid NGA	Peverati and Truhlar (2012)
$\omega$ B97	Long-range corrected density functional	Chai and Head-Gordon (2008b)
$\omega$ B97X	Long-range corrected density functional	Chai and Head-Gordon (2008b)
$\omega$ B97XD	$\omega$ B97X version including empirical dispersion	Chai and Head-Gordon (2008a)

In these functionals, GGA stands for generalized gradient approximation (in which the density functional depends on the up and down spin densities and their reduced gradient) and NGA stands for nonseparable gradient approximation (in which the density functional depends on the up/down spin densities and their reduced gradient, and also adopts a nonseparable form).

### 4. RESULTS AND DISCUSSION

The three-dimentional molecular structure of the BISARG system was built with the aid of molecular graphics program starting from structure presented in the original article (Hofmann, 1998). Starting from this, the molecular structure of its protonated derivative, BISARG(p) was built with the aid of a chemical visualization software. The pre-optimization of the systems was done using random sampling that involved

molecular mechanics techniques and inclusion of the various torsional angles via the general MMFF94 force field (Halgren, 1996a,b,c, 1999; Halgren and Nachbar, 1996) through the Marvin View 17.15 program that constitutes an advanced chemical viewer suited to multiple and single chemical queries, structures and reactions (<https://www.chemaxon.com>). Afterwards, the structures that the resultant lower-energy conformers assumed for both molecules were reoptimized using the eight density functionals mentioned in the previous section together with the Def2SVP basis set as well as the SMD solvation model using water as the solvent.

The analysis of the results obtained in the study aimed at verifying that the KID procedure was fulfilled. On doing it previously, several descriptors associated with the results that HOMO and LUMO calculations obtained are related with results obtained using the vertical I and A following the  $\Delta$ SCF procedure. A link exists between the three main descriptors and the simplest conformity to the Koopmans' theorem by linking  $\epsilon_H$  with  $-I$ ,  $\epsilon_L$  with  $-A$ , and their behavior in describing the HOMO-LUMO gap as  $J_I = |\epsilon_H + E_{gs}(N-1) - E_{gs}(N)|$ ,  $J_A = |\epsilon_L + E_{gs}(N) - E_{gs}(N+1)|$ , and  $J_{HL} = \sqrt{J_I^2 + J_A^2}$ . Notably, the  $J_A$  descriptor consists of an approximation that remains valid only when the HOMO that a radical anion has (the SOMO) shares similarity with the LUMO that the neutral system has. Consequently, we decided to design another descriptor  $\Delta$ SL (the difference between the SOMO and LUMO energies), to guide in verifying how the approximation is accurate.

The results of the calculation of the electronic energies of the neutral, positive and negative molecular systems (in au) of BISARG and BISARG(p), the HOMO, LUMO, and SOMO orbital energies (in eV),  $J_I$ ,  $J_A$ ,  $J_{HL}$ , and  $\Delta$ SL descriptors calculated with the eight density functionals and the Def2TZVP basis set using water as solvent simulated with the SMD parametrization of the IEF-PCM model are presented in **Tables 1, 2**.

As presented in previous works, we considered four other descriptors that analyze how well the studied density functionals are useful for the prediction of the electronegativity  $\chi$ , the global hardness  $\eta$ , and the global electrophilicity  $\omega$ , and for a combination of these Conceptual DFT descriptors, considering only the energies of the HOMO and LUMO or the vertical I and A:  $J_\chi = |\chi - \chi_K|$ ,  $J_\eta = |\eta - \eta_K|$ ,  $J_\omega = |\omega - \omega_K|$ , and  $J_{CDFT} = \sqrt{J_\chi^2 + J_\eta^2 + J_\omega^2}$ , where CDFT stands for Conceptual DFT. The underscript K stands for the descriptor calculated by applying the KID procedure. The results of the calculations of  $J_\chi$ ,  $J_\eta$ ,  $J_\omega$ , and  $J_{CDFT}$  for the low-energy conformers of BISARG and BISARG(p) in water are displayed in **Tables 3, 4**, respectively.

As **Tables 1–4** provide, the KID procedure applies accurately from MN12SX and N12SX density functionals that are range-separated hybrid meta-NGA as well as range-separated hybrid NGA density functionals respectively. In fact, the values of  $J_I$ ,  $J_A$ , and  $J_{HL}$  are actually not zero. Nevertheless, the results tend to be impressive especially for the MN12SX density functional. As well, the  $\Delta$ SL descriptor reaches the minimum values when MN12SX and N12SX density functionals are used in the calculations. This implies that there are sufficient justifications to assume that the

**TABLE 1** | Electronic energies of the neutral, positive and negative molecular systems (in au) of the BISARG molecule, the HOMO, LUMO, and SOMO orbital energies (in eV),  $J_A$ ,  $J_{HL}$ , and  $\Delta SL$  descriptors (also in eV) calculated with the eight RSH density functionals and the Def2TZVP basis set using water as solvent simulated with the SMD parametrization of the IEF-PCM model.

	<b>E<sub>0</sub></b>	<b>E<sub>+</sub></b>	<b>E<sub>-</sub></b>	<b>HOMO</b>	<b>LUMO</b>	<b>SOMO</b>	<b>J<sub>I</sub></b>	<b>J<sub>A</sub></b>	<b>J<sub>HL</sub></b>	<b>ΔSL</b>
CAM-B3LYP	-2276.0764	-2275.8795	-2276.1832	-6.6124	-1.6898	-4.1310	1.2522	1.2168	1.7458	2.4412
LC- $\omega$ BPE	-2275.6818	-2275.4754	-2275.7958	-7.8736	-0.9566	-5.2792	2.2572	2.1454	3.1139	4.3226
M11	-2275.9123	-2275.7056	-2276.0238	-7.6575	-1.1204	-4.9627	2.0321	1.9146	2.7920	3.8422
MN12SX	-2275.1376	-2274.9395	-2275.2504	-5.3883	-3.0639	-3.0873	0.0022	0.0082	0.0084	0.0144
N12SX	-2276.0990	-2275.9088	-2276.2081	-5.1959	-2.9278	-3.0064	0.0218	0.0433	0.0484	0.0787
$\omega$ B97	-2276.7888	-2276.5871	-2276.8951	-7.7315	-0.7982	-5.0476	2.2420	2.0947	3.0685	4.2494
$\omega$ B97X	-2276.5890	-2276.3879	-2276.6957	-7.5543	-0.9354	-4.9069	2.0819	1.9682	2.8652	3.9715
$\omega$ B97XD	-2276.4388	-2276.2394	-2276.5463	-7.2139	-1.1972	-4.6660	1.7858	1.7271	2.4842	3.4688

**TABLE 2** | Electronic energies of the neutral, positive and negative molecular systems (in au) of the protonated BISARG(p) molecule, the HOMO, LUMO, and SOMO orbital energies (in eV),  $J_A$ ,  $J_{HL}$ , and  $\Delta SL$  descriptors (also in eV) calculated with the eight RSH density functionals and the Def2TZVP basis set using water as solvent simulated with the SMD parametrization of the IEF-PCM model.

	<b>E<sub>0</sub></b>	<b>E<sub>+</sub></b>	<b>E<sub>-</sub></b>	<b>HOMO</b>	<b>LUMO</b>	<b>SOMO</b>	<b>J<sub>I</sub></b>	<b>J<sub>A</sub></b>	<b>J<sub>HL</sub></b>	<b>ΔSL</b>
CAM-B3LYP	-2276.5424	-2276.3458	-2276.6659	-6.6364	-2.0844	-4.6328	1.2862	1.2775	1.8128	2.5484
LC- $\omega$ BPE	-2276.1485	-2275.9438	-2276.2828	-7.8698	-1.3822	-5.8926	2.2969	2.2703	3.2296	4.5103
M11	-2276.3780	-2276.1724	-2276.5077	-7.6670	-1.5308	-5.5304	2.0719	2.0011	2.8804	3.9995
MN12SX	-2275.6003	-2275.4000	-2275.7267	-5.4495	-3.4299	-3.4481	0.0014	0.0114	0.0114	0.0182
N12SX	-2276.5694	-2276.3768	-2276.6925	-5.2653	-3.2944	-3.3910	0.0237	0.0558	0.0604	0.0966
$\omega$ B97	-2277.2550	-2277.0549	-2277.3813	-7.7220	-1.2241	-5.6655	2.2784	2.2150	3.1776	4.4314
$\omega$ B97X	-2277.0557	-2276.8559	-2277.1814	-7.5546	-1.3488	-5.4980	2.1179	2.0727	2.9634	4.1492
$\omega$ B97XD	-2276.9093	-2276.7104	-2277.0339	-7.2329	-1.5926	-5.1894	1.8210	1.7976	2.5588	3.5967

**TABLE 3** |  $J_\chi$ ,  $J_\eta$ ,  $J_\omega$ , and  $J_{CDFT}$  (in eV) of the BISARG intermediate melanoidin pigment.

	<b>J<sub>χ</sub></b>	<b>J<sub>η</sub></b>	<b>J<sub>ω</sub></b>	<b>J<sub>CDFT</sub></b>
CAM-B3LYP	0.0185	2.4706	1.7322	3.0174
LC- $\omega$ BPE	0.0552	4.4023	2.3704	5.0002
M11	0.0591	3.9462	2.1447	4.4917
MN12SX	0.0045	0.0030	0.0131	0.0142
N12SX	0.0109	0.0610	0.1205	0.1355
$\omega$ B97	0.0738	4.3370	2.0710	4.8067
$\omega$ B97X	0.0565	4.0499	2.0530	4.5409
$\omega$ B97XD	0.0294	3.5156	2.0168	4.0531

**TABLE 4** |  $J_\chi$ ,  $J_\eta$ ,  $J_\omega$ , and  $J_{CDFT}$  (in eV) of the protonated BISARG(p) intermediate melanoidin pigment.

	<b>J<sub>χ</sub></b>	<b>J<sub>η</sub></b>	<b>J<sub>ω</sub></b>	<b>J<sub>CDFT</sub></b>
CAM-B3LYP	0.0046	2.5626	2.6800	3.7080
LC- $\omega$ BPE	0.0131	4.5716	3.9039	6.0117
M11	0.0363	4.0705	3.3156	5.2501
MN12SX	0.0059	0.0084	0.0334	0.0349
N12SX	0.0161	0.0795	0.2316	0.2454
$\omega$ B97	0.0316	4.4894	3.3713	5.6144
$\omega$ B97X	0.0225	4.1892	3.2674	5.3128
$\omega$ B97XD	0.0107	3.6182	3.0654	4.7422

LUMO of the neutral approximates the electron affinity. The same density functionals follow the KID procedure in the rest of the descriptors such as  $J_\chi$ ,  $J_\eta$ ,  $J_\omega$ , and  $J_{CDFT}$ .

Having verified that the MN12SX/Def2TZVP model chemistry is a good choice for the calculation of the global reactivity descriptors, we now present the optimized molecular structures of BISARG and BISARG(p) in water in Supplementary Figures 1, 2. Meanwhile, the calculated bond lengths and bond angles for both cases are shown in Supplementary Tables 1–4.

As a summary of the previous results, the global reactivity descriptors for the BISARG and BISARG(p) molecules calculated

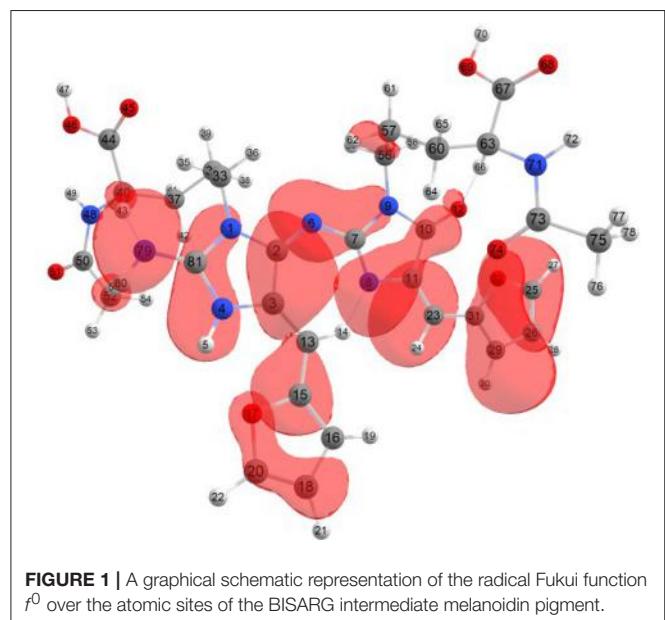
with the MN12SX/Def2TZVP model chemistry in water are presented in Table 5.

The calculations of the condensed Fukui functions and dual descriptor are done by using the Chemcraft molecular analysis program to extract the Mulliken and NPA atomic charges (Zhurko and Zhurko, 2012) beginning with single-point energy calculations involving the MN12SX density functional that uses the Def2TZVP basis set in line with the SMD solvation model, and water utilized as the solvent.

Considering the potential application the studied molecules as antioxidants, it is of interest to get insight into the active sites for radical attack. Graphical representations of the radical Fukui

**TABLE 5 |** Global reactivity descriptors for the BISARG intermediate melanoidin pigment and its protonated derivative BISARGd(p) calculated with the MN12SX density functional.

	Electronegativity ( $\chi$ )	Chemical hardness ( $\eta$ )	Electrophilicity ( $\omega$ )
BISARG	4.2253	2.3240	3.8410
BISARG(p)	4.4389	2.0193	4.8790
Electrodonating Power ( $\omega^-$ )	Electroaccepting Power ( $\omega^+$ )	Net electrophilicity ( $\Delta\omega^\pm$ )	
BISARG	5.8449	4.4318	10.2767
BISARG(p)	7.1172	5.6637	12.7809

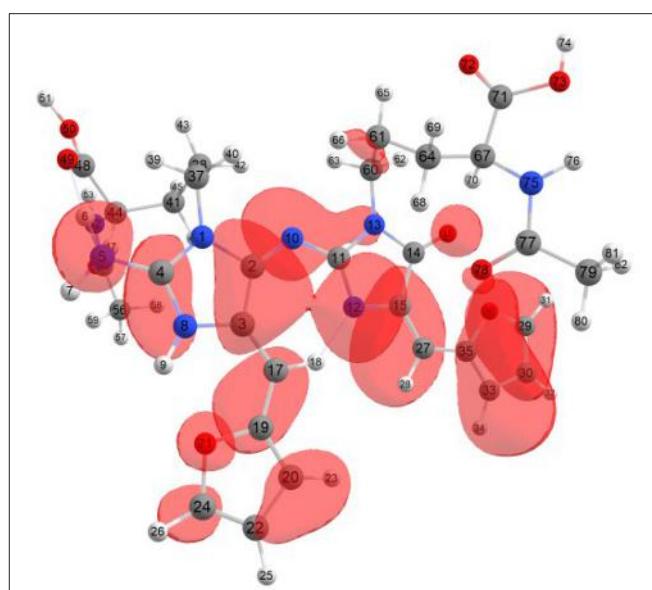


**FIGURE 1 |** A graphical schematic representation of the radical Fukui function  $f^0$  over the atomic sites of the BISARG intermediate melanoidin pigment.

function  $f^0$  calculated with the MN12SX/Def2TZVP model chemistry for both systems in water are presented in **Figures 1, 2**.

The condensed electrophilic and nucleophilic Parr functions  $P_k^+$  and  $P_k^-$  over the atoms of the BISARG and BISARG(p) molecules in water have been calculated by extracting the Mulliken and Hirshfeld (or CM5) atomic charges using the Chemcraft molecular analysis program (Zhurko and Zhurko, 2012) starting from single-point energy calculations of the ionic species with the MN12SX density functional using the Def2TZVP basis set in the presence of the solvents according to the SMD solvation model.

The results for the condensed dual descriptor calculated with Mulliken atomic charges  $\Delta f_k$  (M), with NPA atomic charges  $\Delta f_k$  (N), the electrophilic and nucleophilic Parr functions with Mulliken atomic spin densities  $P_k^+$  (M) and  $P_k^-$  (M), and the electrophilic and nucleophilic Parr functions with Hirshfeld (or CM5) atomic spin densities  $P_k^+$  (H) and  $P_k^-$  (H) are displayed in **Tables 6, 7** for the BISARG and BISARG(p) molecules in water, respectively, while **Figures 3, 4** show schematic representations of the molecules with the numbering of the most important reactive sites according to the results in **Tables 6, 7**.



**FIGURE 2 |** A graphical schematic representation of the radical Fukui function  $f^0$  over the atomic sites of the protonated BISARG(p) intermediate melanoidin pigment.

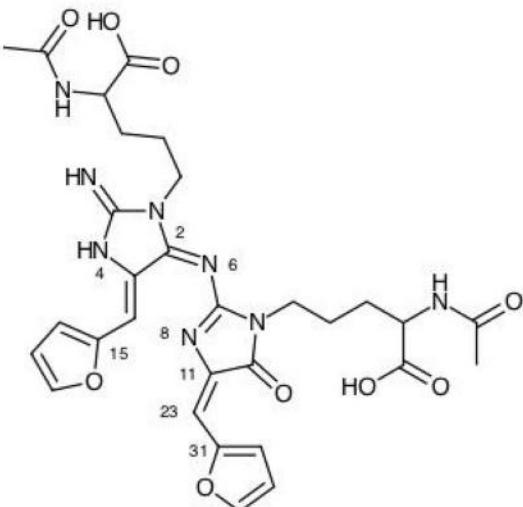
**TABLE 6 |** The condensed dual descriptor calculated with Mulliken atomic charges  $\Delta f_k$  (M), and with NPA atomic charges  $\Delta f_k$  (N), the electrophilic and nucleophilic Parr functions with Mulliken atomic spin densities  $P_k^+$  (M) and  $P_k^-$  (M), and the electrophilic and nucleophilic Parr functions with Hirshfeld (or CM5) atomic spin densities  $P_k^+$  (H) and  $P_k^-$  (H) for the BISARG melanoidin molecule. Hydrogens and atomic sites where the absolute value of the dual descriptor is lower than 1 are not shown.

Atom	$\Delta f_k$ (M)	$\Delta f_k$ (N)	$P_k^+$ (M)	$P_k^-$ (M)	$P_k^+$ (H)	$P_k^-$ (H)
1N	1.95	0.17	0.0265	0.0052	0.0355	0.0013
2C	12.26	9.75	0.2046	-0.0477	0.1282	-0.0128
3C	-2.86	-3.50	-0.0495	0.0454	0.0107	0.0382
4N	-4.84	-3.07	-0.0093	0.0664	-0.0078	0.0502
6N	-4.90	-4.93	-0.0386	0.1255	-0.0014	0.0842
7C	4.31	3.52	0.0990	-0.0287	0.0724	0.0118
8N	-7.69	-5.30	0.0121	0.1827	0.0144	0.1373
10C	3.75	3.52	0.0606	-0.0093	0.0446	0.0055
11C	-4.19	-3.69	-0.0307	0.0601	0.0164	0.0672
13C	3.64	3.40	0.1828	0.0879	0.1057	0.0693
15C	-3.10	-0.39	-0.0496	0.0160	-0.0010	0.0274
18C	-1.20	-0.94	-0.0285	-0.0154	-0.0028	0.0082
20C	-2.43	-1.71	0.0724	0.1095	0.0502	0.0801
23C	8.06	7.25	0.2653	0.0842	0.1563	0.0720
25C	-3.21	-1.99	0.1020	0.1482	0.0709	0.1094
26C	-1.89	-1.47	-0.0388	-0.0136	-0.0033	0.0150
31C	-4.63	-0.94	-0.0663	0.0355	0.0001	0.0422
32O	1.16	1.07	0.0184	-0.0118	0.0194	0.0009
79N	2.58	1.75	0.0473	-0.0153	0.0355	-0.0113
81C	1.28	0.19	0.0000	-0.0043	0.0127	0.0010

From the results for the local reactivity descriptors in **Table 6**, it can be concluded that C2 and C23 will be the preferred sites for a nucleophilic attack and that these atoms will act as electrophilic

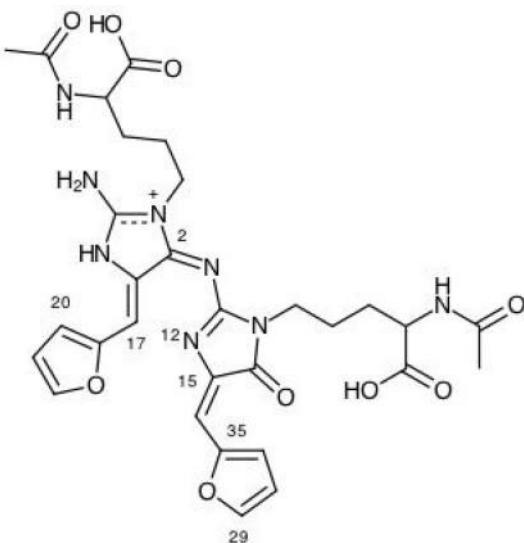
**TABLE 7 |** The condensed dual descriptor calculated with Mulliken atomic charges  $\Delta f_k$  (M), and with NPA atomic charges  $\Delta f_k$  (N), the electrophilic and nucleophilic Parr functions with Mulliken atomic spin densities  $P_k^+$  (M) and  $P_k^-$  (M), and the electrophilic and nucleophilic Parr functions with Hirshfeld (or CM5) atomic spin densities  $P_k^+$  (H) and  $P_k^-$  (H) for the BISARG(p) melanoidin molecule. Hydrogens and atomic sites where the absolute value of the dual descriptor is lower than 1 are not shown.

Atom	$\Delta f_k$ (M)	$\Delta f_k$ (N)	$P_k^+$ (M)	$P_k^-$ (M)	$P_k^+$ (H)	$P_k^-$ (H)
1N	-0.28	-1.09	-0.0083	0.0179	0.0158	0.0112
2C	12.50	8.97	0.2317	-0.0230	0.1412	-0.0021
4C	3.40	2.52	0.0402	-0.0084	0.0341	-0.0032
5N	2.40	2.14	0.0260	-0.0076	0.0234	-0.0062
8N	-1.30	-0.84	-0.0074	0.0218	-0.0056	0.0161
12N	-6.95	-3.44	0.0425	0.1817	0.0308	0.1445
15C	-10.14	-7.30	-0.0588	0.1222	-0.0089	0.1049
16O	-3.36	-4.03	0.0136	0.0797	0.0125	0.0662
17C	13.43	8.74	0.3116	0.0363	0.1795	0.0297
19C	-1.32	1.08	-0.0883	0.0093	-0.0042	0.0134
20C	6.25	4.80	0.1683	0.0409	0.1008	0.0299
21O	1.44	1.41	0.0184	-0.0043	0.0200	0.0002
24C	3.85	2.95	0.1294	0.0546	0.0900	0.0401
27C	2.39	4.54	0.1996	0.0718	0.1140	0.0759
29C	-7.81	-5.59	0.0807	0.1977	0.0566	0.1467
30C	-2.76	-2.17	-0.0284	-0.0127	-0.0017	0.0222
33C	-3.67	-2.45	0.1007	0.1274	0.0610	0.0975
35C	-7.01	-2.75	-0.0547	0.0663	-0.0039	0.0638



**FIGURE 3 |** A graphical schematic representation of the BISARG intermediate melanoidin pigment with the numbering of the most important reactive sites.

species in a chemical reaction. In turn, it can be appreciated that N8 will be prone to electrophilic attacks and that this atomic site will act as a nucleophilic species in chemical reactions that involve the BISARG molecule in water. In turn, for the case of the BISARG(p) molecule in water, C2 and C17 will be the preferred



**FIGURE 4 |** A graphical schematic representation of the protonated BISARG(d) intermediate melanoidin pigment with the numbering of the most important reactive sites.

## 5. CONCLUSIONS

Eight fixed RSH density functionals, including CAM-B3LYP, LC- $\omega$ PBE, M11, MN12SX, N12SX,  $\omega$ B97,  $\omega$ B97X, and  $\omega$ B97XD, were examined to determine whether they fulfill the empirical KID procedure so as to provide computational support for this common practice. The assessment was conducted by comparing the values from HOMO and LUMO calculations to those generated by the  $\Delta$ SCF technique for the BISARG molecule and its protonated derivative, BISARG(p). BISARG and BISARG(p) are intermediate melanoidin pigments that are of academic and industrial interest. The study has observed that the range-separated and hybrid meta-NGA density functionals tend to be the most suited in meeting this goal. Thus, they can be suitable alternatives to density functionals where the behavior of them is optimally tuned using a gap-fitting procedure. They also exhibit the desirable prospect of benefiting future studies aimed at understanding the chemical reactivity of colored melanoidins with larger molecular weights when reducing sugars react with proteins and peptides.

It is not the goal of Computational Chemistry to perform studies to reproduce known experimental results except in the case that they can be used for the calibration of a particular technique. Instead, it can be useful to predict in advance the structural and chemical reactivity characteristics of new or unknown molecular systems whose properties have not been reported and as guide for future research. As far as we know, there are no reports in the literature about the chemical reactivity properties for the molecular systems considered in this work and it is not possible to perform any kind of comparisons.

However, the present study shows that with an adequate choice of the model chemistry we have been able to predict the sites of interaction of the BISARG and BISARG(p) molecules with impressive accuracy starting from the knowledge of the HOMO, LUMO, and HOMO-LUMO gap energies of the studied systems. This involves having DFT-based reactivity descriptors, including Fukui functions, Parr functions, and Dual Descriptor calculations. In conclusion, the Conceptual DFT descriptors are useful in characterizing and describing the preferred reactive sites and in comprehensively explaining the reactivity of the molecules.

## AUTHOR CONTRIBUTIONS

DG-M conceived and designed the research and headed, wrote and revised the manuscript, while JF contributed to the writing and the revision of the article.

## REFERENCES

- Alvarado-González, M., Flores-Holguín, N., and Glossman-Mitnik, D. (2013). Computational nanochemistry study of the molecular structure and properties of the chlorophyll a molecule. *Int. J. Photoener.* 2013:8. doi: 10.1155/2013/424620
- Baer, R., Livshits, E., and Salzner, U. (2010). Tuned range-separated hybrids in density functional theory. *Annu. Rev. Phys. Chem.* 61, 85–109. doi: 10.1146/annurev.physchem.012809.103321
- Baerends, E. J., Gritsenko, O. V., and van Meer, R. (2013). The Kohn-Sham gap, the fundamental gap and the optical gap: the physical meaning of occupied and virtual Kohn-Sham orbital energies. *Phys. Chem. Chem. Phys.* 15, 16408–16425. doi: 10.1039/c3cp52547c
- Cervantes-Navarro, F., and Glossman-Mitnik, D. (2013). Density functional study of the effects of substituents on the chemical reactivity of the Indigo molecule. *J. Theory Comput. Chem.* 12:1350013. doi: 10.1142/S0219633613500132
- Chai, J., and Head-Gordon, M. (2008a). Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* 10, 6615–6620. doi: 10.1039/b810189b
- Chai, J., and Head-Gordon, M. (2008b). Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* 128:084106. doi: 10.1063/1.2834918
- Chattaraj, P., Chakraborty, A., and Giri, S. (2009). Net electrophilicity. *J. Phys. Chem. A* 113, 10068–10074. doi: 10.1021/jp904674x
- Domingo, L. R., Pérez, P., and Sáez, J. (2013). Understanding the local reactivity in polar organic reactions through electrophilic and nucleophilic parr functions. *RSC Adv.* 3, 1486–1494. doi: 10.1039/C2RA22886F
- Domingo, L. R., Ríos-Gutiérrez, M., and Pérez, P. (2016). Applications of the conceptual density functional theory indices to organic chemistry reactivity. *Molecules* 21:e748. doi: 10.3390/molecules21060748
- Egger, D. A., Weissman, S., Refaelly-Abramson, S., Sharifzadeh, S., Dauth, M., Baer, R., et al. (2014). Outer-valence electron spectra of prototypical aromatic heterocycles from an optimally tuned range-separated hybrid functional. *J. Chem. Theory Comput.* 10, 1934–1952. doi: 10.1021/ct400956h
- Foster, M. E., Azoulay, J. D., Wong, B. M., and Allendorf, M. D. (2014). Novel metal-organic framework linkers for light harvesting applications. *Chem. Sci.* 5, 2081–2090. doi: 10.1021/jp0622381
- Foster, M. E., and Wong, B. M. (2012). Nonempirically tuned range-separated DFT accurately predicts both fundamental and excitation gaps in DNA and RNA nucleobases. *J. Chem. Theory Comput.* 8, 2682–2687. doi: 10.1021/ct300420f
- Frau, J., and Glossman-Mitnik, D. (2017a). A comparative study of the glycation power of simple carbohydrates in the maillard reaction by means of conceptual DFT descriptors. *Br. J. Appl. Sci. Technol.* 21:32795. doi: 10.9734/BJAST/2017/32795
- Frau, J., and Glossman-Mitnik, D. (2017b). A conceptual DFT study of the molecular properties of glycation carbonyl compounds. *Chem. Cent. J.* 11:8. doi: 10.1186/s13065-017-0239-7
- Frau, J., and Glossman-Mitnik, D. (2017c). Chemical reactivity theory study of advanced glycation endproduct inhibitors. *Molecules* 22:226. doi: 10.3390/molecules22020226
- Frau, J., and Glossman-Mitnik, D. (2017d). Computational prediction of the reactivity sites of Alzheimer amyloid  $\beta$ -peptides  $\text{A}\beta40$  and  $\text{A}\beta42$ . *ChemXpress* 10:120.
- Frau, J., and Glossman-Mitnik, D. (2017e). Conceptual DFT descriptors of amino acids with potential corrosion inhibition properties calculated with the latest Minnesota density functionals. *Front. Chem.* 5:16. doi: 10.3389/fchem.2017.00016
- Frau, J., and Glossman-Mitnik, D. (2017f). Molecular modeling study of the structures, properties and glycation power of some reducing disaccharides. *MOJ Drug Design Dev. Ther.* 1:00003. doi: 10.15406/mojddt.2017.01.0003
- Frau, J., and Glossman-Mitnik, D. (2017g). Pyridoxamine derivatives as non enzymatic glycation inhibitors: the conceptual DFT viewpoint. *Res. J. Life Sci. Bioinform. Pharm. Chem. Sci.* 2, 103–122. doi: 10.26479/2017.0206.09
- Frau, J., Hernández-Haro, N., and Glossman-Mitnik, D. (2017a). Computational prediction of the pKas of small peptides through conceptual DFT descriptors. *Chem. Phys. Lett.* 671, 138–141. doi: 10.1016/j.cplett.2017.01.038
- Frau, J., Muñoz, F., and Glossman-Mitnik, D. (2016a). A molecular electron density theory study of the chemical reactivity of cis- and trans-resveratrol. *Molecules* 21:1650. doi: 10.3390/molecules21121650
- Frau, J., Muñoz, F., and Glossman-Mitnik, D. (2016b). A theoretical study of the chemical reactivity of neohesperidin dihydrochalcone through conceptual DFT descriptors. *SDRP J. Comput. Chem. Mol. Model.* 1:3. doi: 10.25177/JCCMM.1.2.3
- Frau, J., Muñoz, F., and Glossman-Mitnik, D. (2016c). Validation of the Koopmans' theorem by means of the calculation of the conceptual DFT descriptors of three fluorescent DNA staining dyes. *Chem. Inform.* 2:7. doi: 10.21767/2470-6937.100021
- Frau, J., Muñoz, F., and Glossman-Mitnik, D. (2017b). A comparison of the Minnesota family of density functionals for the calculation of conceptual DFT descriptors: citrus flavonoids as a test case. *Res. J. Chem. Sci.* 7, 46–58.
- Frau, J., Muñoz, F., and Glossman-Mitnik, D. (2017c). A conceptual DFT study of the chemical reactivity of magnesium octaethylporphyrin (MgOEP) as predicted by the Minnesota family of density functionals. *Química Nova* 40, 402–406. doi: 10.21577/0100-4042.20170004
- Frau, J., Muñoz, F., and Glossman-Mitnik, D. (2017d). Application of DFT concepts to the study of the chemical reactivity of some resveratrol derivatives through the assessment of the validity of the "Koopmans in DFT" (KID) procedure. *J. Theor. Comput. Chem.* 16:1750006. doi: 10.1142/S0219633617500067

## ACKNOWLEDGMENTS

This work has been partially supported by CIMAV, SC, and Consejo Nacional de Ciencia y Tecnología (CONACYT, Mexico) through Grant 219566-2014 for Basic Science Research. DG-M conducted this work while a Visiting Lecturer at the University of the Balearic Islands from which support is gratefully acknowledged. This work was cofunded by the Ministerio de Economía y Competitividad (MINECO) and the European Fund for Regional Development (FEDER) (CTQ2014-55835-R).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2018.00136/full#supplementary-material>

- Frau, J., Ramis, R., and Glossman-Mitnik, D. (2017e). Computational prediction of the preferred glycation sites of model helical peptides derived from Human Serum Albumin (HSA) and Lysozyme Helix 4 (LH4). *Theor. Chem. Acc.* 136:39. doi: 10.1007/s00214-017-2070-6
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2018). *Gaussian 09 Revision D.01*. Wallingford, CT: Gaussian Inc.
- Gázquez, J., Cedillo, A., and Vela, A. (2007). Electrodonating and electroaccepting powers. *J. Phys. Chem. A* 111, 1966–1970. doi: 10.1021/jp065459f
- Gledhill, J. D., De Proft, F., and Tozer, D. J. (2016). Range-separation parameter in tuned exchange-correlation functionals: successive ionizations and the Fukui function. *J. Chem. Theory Comput.* 12, 4879–4884. doi: 10.1021/acs.jctc.6b00709
- Glossman-Mitnik, D. (2013a). A comparison of the chemical reactivity of naringenin calculated with M06 family of density functionals. *Chem. Cent. J.* 7, 155–161. doi: 10.1186/1752-153X-7-155
- Glossman-Mitnik, D. (2013b). Computational nanochemistry study of the chemical reactivity properties of the rhodamine B molecule. *Proc. Comput. Sci.* 18, 816–825. doi: 10.1016/j.procs.2013.05.246
- Glossman-Mitnik, D. (2014a). Chemical reactivity theory within DFT applied to the study of the prunin flavonoid. *Eur. Int. J. Sci. Techn.* 3, 195–207.
- Glossman-Mitnik, D. (2014b). Computational chemistry of natural products: a comparison of the chemical reactivity of isonaringin calculated with the M06 family of density functionals. *J. Mol. Model.* 20:2316. doi: 10.1007/s00894-014-2316-3
- Glossman-Mitnik, D. (2014c). “Computational nanochemistry report of the molecular structure, spectra and chemical reactivity properties of pheophorbide A,” in *Design and Applications of Nanomaterials for Sensors*, ed. J. M. Seminario (Dordrecht: Springer Science + Business Media), 359–402.
- Glossman-Mitnik, D. (2014d). “Computational nanochemistry study of the molecular structure, spectra and chemical reactivity properties of the BFPF green fluorescent protein chromophore,” in *Biosensors Nanotechnology*, eds A. Tiwari and A. P. Turner (Hoboken, NJ: John Wiley & Sons), 201–238.
- Halgren, T. A. (1996a). Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* 17, 490–519.
- Halgren, T. A. (1996b). Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* 17, 520–552.
- Halgren, T. A. (1996c). Merck molecular force Field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. *J. Comput. Chem.* 17, 616–641.
- Halgren, T. A. (1999). MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* 20, 720–729.
- Halgren, T. A., and Nachbar, R. B. (1996). Merck molecular force field. IV. Conformational energies and geometries for MMFF94. *J. Comput. Chem.* 17, 587–615.
- Henderson, T. M., Izmaylov, A. F., Scalmani, G., and Scuseria, G. E. (2009). Can short-range hybrids describe long-range-dependent properties? *J. Chem. Phys.* 131:044108. doi: 10.1063/1.3185673
- Hofmann, T. (1998). 4-Alkylidene-2-imino-5-[4-alkylidene-5-oxo-1,3-imidazol-2-inyl]aza-methylidene-1,3-imidazolidine: a novel colored substructure in melanoidins formed by maillard reactions of bound Arginine with glyoxal and furan-2-carboxaldehyde. *J. Agric. Food Chem.* 46, 3896–3901.
- Jacquemin, D., Moore, B., Planchat, A., Adamo, C., and Autschbach, J. (2014). Performance of an optimally tuned range-separated hybrid functional for 0–0 electronic excitation energies. *J. Chem. Theory Comput.* 10, 1677–1685. doi: 10.1021/ct5000617
- Karolewski, A., Kronik, L., and Kümmel, S. (2013). Using optimally tuned range separated hybrid functionals in ground-state calculations: consequences and caveats. *J. Chem. Phys.* 138:204115. doi: 10.1063/1.4807325
- Karolewski, A., Stein, T., Baer, R., and Kümmel, S. (2011). Communication: tailoring the optical gap in light-harvesting molecules. *J. Chem. Phys.* 134, 151101–151105. doi: 10.1063/1.3581788
- Koppen, J. V., Hapka, M., Szczepiak, M. M., and Chalasinski, G. (2012). Optical absorption spectra of gold clusters Au(n) (n = 4, 6, 8, 12, 20) from long-range corrected functionals with optimal tuning. *J. Chem. Phys.* 137:114302. doi: 10.1063/1.4752433
- Kronik, L., Stein, T., Refaelly-Abramson, S., and Baer, R. (2012). Excitation gaps of finite-sized systems from optimally tuned range-separated hybrid functionals. *J. Chem. Theory Comput.* 8, 1515–1531. doi: 10.1021/ct2009363
- Kuritz, N., Stein, T., Baer, R., and Kronik, L. (2011). Charge-transfer-like  $\pi \rightarrow \pi^*$  excitations in time-dependent density functional theory: a conundrum and its solution. *J. Chem. Theory Comput.* 7, 2408–2415. doi: 10.1021/ct2002804
- Lima, I. T., Prado, A. d. S., Martins, J. B. L., de Oliveira Neto, P. H., Ceschin, A. M., da Cunha, W. F., et al. (2016). Improving the description of the optical properties of carotenoids by tuning the long-range corrected functionals. *J. Phys. Chem. A* 120, 4944–4950. doi: 10.1021/acs.jpca.5b12570
- Manna, A. K., Lee, M. H., McMahon, K. L., and Dunietz, B. D. (2015). Calculating high energy charge transfer states using optimally tuned range-separated hybrid functionals. *J. Chem. Theory Comput.* 11, 1110–1117. doi: 10.1021/ct501018n
- Marenich, A., Cramer, C., and Truhlar, D. (2009). Universal solvation model based on solute electron density and a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* 113, 6378–6396. doi: 10.1021/jp810292n
- Martínez-Araya, J. I., and Glossman-Mitnik, D. (2015). The substituent effect from the perspective of local hyper-softness. An example applied on normeloxicam, meloxicam and 4-meloxicam: non-steroidal anti-inflammatory drugs. *Chem. Phys. Lett.* 618, 162–167. doi: 10.1016/j.cplett.2014.11.004
- Martínez-Araya, J. I., Grand, A., and Glossman-Mitnik, D. (2015). Towards the rationalization of catalytic activity values by means of local hyper-softness on the catalytic site: a criticism about the use of net electric charges. *Phys. Chem. Chem. Phys.* 17, 29764–29775. doi: 10.1039/c5cp03822g
- Martínez-Araya, J. I., Salgado-Morán, G., and Glossman-Mitnik, D. (2013a). Computational nanochemistry report on the oxicams - conceptual DFT and chemical reactivity. *J. Phys. Chem. B* 117, 6639–6651. doi: 10.1021/jp400241q
- Martínez-Araya, J. I., Salgado-Morán, G., and Glossman-Mitnik, D. (2013b). Computational nutraceutics: chemical reactivity properties of the flavonoid naringin by means of conceptual DFT. *J. Chem.* 2013:8. doi: 10.1155/2013/850297
- Mendoza-Huizar, L. H., Salgado-Morán, G., Ramírez-Tagle, R., and Glossman-Mitnik, D. (2016). A theoretical quantum study of the intramolecular interactions and chemical reactivity of polymorphs A and B of famotidine in the gas, DMSO, and aqueous phases. *Comput. Theor. Chem.* 1075, 54–62. doi: 10.1016/j.comptc.2015.11.007
- Moore, B. II., and Autschbach, J. (2013). Longest-wavelength electronic excitations of linear cyanines: the role of electron delocalization and of approximations in time-dependent density functional theory. *J. Chem. Theory Comput.* 9, 4991–5003. doi: 10.1021/ct400649r
- Morell, C., Grand, A., and Toro-Labbé, A. (2005). New dual descriptor for chemical reactivity. *J. Phys. Chem. A* 109, 205–212. doi: 10.1021/jp046577a
- Morell, C., Grand, A., and Toro-Labbé, A. (2006). Theoretical support for using the  $\Delta f(r)$  descriptor. *Chem. Phys. Lett.* 425, 342–346. doi: 10.1016/j.cplett.2006.05.003
- Niskanen, M., and Hukka, T. I. (2014). Modeling of photoactive conjugated donor-acceptor copolymers: the effect of the exact HF exchange in DFT functionals on geometries and gap energies of oligomer and periodic models. *Phys. Chem. Chem. Phys.* 16, 13294–13305. doi: 10.1039/C4CP01165A
- Parr, R., Szentpaly, L., and Liu, S. (1999). Electrophilicity index. *J. Am. Chem. Soc.* 121, 1922–1924.
- Parr, R., and Yang, W. (1984). Density functional approach to the frontier-electron theory of chemical reactivity. *J. Am. Chem. Soc.* 106, 4049–4050.
- Pereira, T. L., Leal, L. A., da Cunha, W. F., Timóteo de Sousa Júnior, R., Ribeiro Junior, L. A., et al. (2017). Optimally tuned functionals improving the description of optical and electronic properties of the phthalocyanine molecule. *J. Mol. Model.* 23:71. doi: 10.1007/s00894-017-3246-7
- Peverati, R., and Truhlar, D. G. (2011). Improving the accuracy of hybrid meta-GGA density functionals by range separation. *J. Phys. Chem. Lett.* 2, 2810–2817. doi: 10.1021/jz201170d
- Peverati, R., and Truhlar, D. G. (2012). Screened-exchange density functionals with broad accuracy for chemistry and solid-state physics. *Phys. Chem. Chem. Phys.* 14, 16187–16191. doi: 10.1039/C2CP42576A
- Phillips, H., Geva, E., and Dunietz, B. D. (2012a). Calculating off-site excitations in symmetric donor-acceptor systems via time-dependent density functional theory with range-separated density functionals. *J. Chem. Theory Comput.* 8, 2661–2668. doi: 10.1021/ct300318g

- Phillips, H., Zheng, S., Hyla, A., Laine, R., Goodson, T. III, Geva, E., et al. (2012b). Ab initio calculation of the electronic absorption of functionalized octahedral silsesquioxanes via time-dependent density functional theory with range-separated hybrid functionals. *J. Phys. Chem. A* 116, 1137–1145. doi: 10.1021/jp208316t
- Refaely-Abramson, S., Baer, R., and Kronik, L. (2011). Fundamental and excitation gaps in molecules of relevance for organic photovoltaics from an optimally tuned range-separated hybrid functional. *Phys. Rev. B* 84:075144–8. doi: 10.1103/PhysRevB.84.075144
- Rizzi, G. P. (1997). Chemical structure of colored maillard reaction products. *Food Rev. Int.* 13, 1–28.
- Salgado-Morán, G., Ruiz-Nieto, S., Gerli-Candia, L., Flores-Holguín, N., Favila-Pérez, A., and Glossman-Mitnik, D. (2013). Computational nanochemistry study of the molecular structure and properties of ethambutol. *J. Mol. Model.* 19, 3507–3515. doi: 10.1007/s00894-013-1884-y
- Sastre, S., Frau, J., and Glossman-Mitnik, D. (2017). Computational prediction of the protonation sites of Ac-Lys-(Ala)<sub>n</sub>-Lys-NH<sub>2</sub> peptides through conceptual DFT and MEDT descriptors. *Molecules* 22:458. doi: 10.3390/molecules22030458
- Soto-Rojo, R., Baldenebro-López, J., and Glossman-Mitnik, D. (2015). Study of chemical reactivity in relation to experimental parameters of efficiency in coumarin derivatives for dye sensitized solar cells using DFT. *Phys. Chem. Chem. Phys.* 17, 14122–14129. doi: 10.1039/C5CP01387A
- Stein, T., Kronik, L., and Baer, R. (2009a). Prediction of charge-transfer excitations in coumarin-based dyes using a range-separated functional tuned from first principles. *J. Chem. Phys.* 131:244119. doi: 10.1063/1.3269029
- Stein, T., Kronik, L., and Baer, R. (2009b). Reliable prediction of charge transfer excitations in molecular complexes using time-dependent density functional theory. *J. Am. Chem. Soc.* 131, 2818–2820. doi: 10.1021/ja8087482
- Sun, H., and Autschbach, J. (2014). Electronic energy gaps for  $\pi$ -conjugated oligomers and polymers calculated with density functional theory. *J. Chem. Theory Comput.* 10, 1035–1047. doi: 10.1021/ct4009975
- Weigend, F. (2006). Accurate coulomb-fitting basis sets for H to R. *Phys. Chem. Chem. Phys.* 8, 1057–1065. doi: 10.1039/b515623h
- Weigend, F., and Ahlrichs, R. (2005). Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy. *Phys. Chem. Chem. Phys.* 7, 3297–3305. doi: 10.1039/B508541A
- Yanai, T., Tew, D. P., and Handy, N. C. (2004). A new hybrid exchange-correlation functional using the coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* 393, 51–57. doi: 10.1016/j.cplett.2004.06.011
- Zhurko, G., and Zhurko, D. (2012). *Chemcraft Program Revision 1.6*. Grigoriy A. Zhurko.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Frau and Glossman-Mitnik. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# QM/MM Investigation of the Role of a Second Coordination Shell Arginine in [NiFe]-Hydrogenases

Andrés M. Escoria and Matthias Stein\*

Molecular Simulations and Design Group, Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany

## OPEN ACCESS

### Edited by:

Hans Martin Senn,  
University of Glasgow,  
United Kingdom

### Reviewed by:

Etienne Derat,  
Université Pierre et Marie Curie,  
France  
Giampaolo Barone,  
Università degli Studi di Palermo, Italy

### \*Correspondence:

Matthias Stein  
matthias.stein@  
mpi-magdeburg.mpg.de

### Specialty section:

This article was submitted to  
Theoretical and Computational  
Chemistry, a section of the journal  
Frontiers in Chemistry

Received: 15 March 2018

Accepted: 23 April 2018

Published: 15 May 2018

### Citation:

Escoria AM and Stein M (2018)  
QM/MM Investigation of the Role of a  
Second Coordination Shell Arginine in  
[NiFe]-Hydrogenases.  
Front. Chem. 6:164.  
doi: 10.3389/fchem.2018.00164

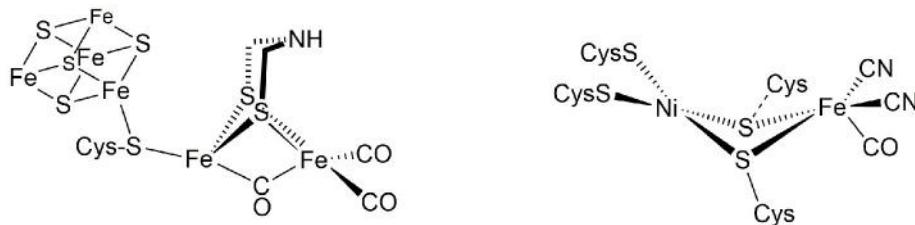
[NiFe]-hydrogenases are highly efficient catalysts for the heterolytic splitting of molecular hydrogen ( $H_2$ ). The heterobimetallic cysteine-coordinated active site of these enzymes is covered by a highly conserved arginine residue, whose role in the reaction is not fully resolved yet. The structural and catalytic role of this arginine is investigated here using QM/MM calculations with various exchange-correlation functionals. All of them give a very consistent picture of the thermodynamics of  $H_2$  oxidation. The concept of the presence of a neutral arginine and its direct involvement as a Frustrated Lewis Pair (FLP) in the reaction is critically evaluated. The arginine, however, would exist in its standard protonation state and perform a critical role in positioning and slightly polarizing the substrate  $H_2$ . It is not directly involved in the heterolytic processing of  $H_2$  but guides its approach and reduces its flexibility during binding. Upon substitution of the positively charged arginine by a charge-conserving lysine residue, the  $H_2$  binding position remains unaffected. However, critical hydrogen bonding interactions with nearby aspartate residues are lost. In addition, the  $H_2$  polarization is unfavorable and the reduced side-chain volume may negatively affect the kinetics of the catalytic process.

**Keywords:** hydrogen conversion, enzyme, QM/MM, amino acid substitution, catalysis

## INTRODUCTION

Hydrogen ( $H_2$ ) is among the most important energy carriers in a post-fossil era. The generation of  $H_2$  as a biofuel from sustainable sources is a versatile alternative to the standard generation process from electrolysis of water which requires elevated temperature and expensive catalyst metals (Holladay et al., 2009; Rodionova et al., 2017). Enzymes from bacteria and microalgae are able to perform the same catalysis at room temperature and standard pressure in the absence of a precious noble metal, and can also catalyze the reverse reaction, the heterolytic cleavage of  $H_2$ . Biological  $H_2$  conversion has attracted much interest owing to its potential application in a post-carbon based scenario employing  $H_2$  as an energy storage compound and as a transportable fuel itself (Cammack et al., 2001).

These enzymes, the hydrogenases, are classified according to their active site composition as [NiFe]- and [FeFe]-hydrogenases (see Figure 1). The active site of [FeFe]-hydrogenases consists of a  $\mu$ -carbonyl bridged iron-iron cluster with two additional terminal CO ligands. A bridging azadiothiolate ligand acts as an intermediate proton acceptor during formation of  $H_2$ . The “H-cluster” is connected to a [4Fe4S]-cluster via a bridging cysteine amino acid. The azadithiolate nitrogen in [FeFe]-hydrogenase enzymes acts as an initial site of protonation before the product molecule hydrogen is formed upon reacting with a terminal Fe-bound hydride.



**FIGURE 1 |** Active sites of [FeFe]- and [NiFe]-hydrogenase enzymes.

Meanwhile, the active site of [NiFe]-hydrogenases involves two terminal and two bridging cysteine residues and three diatomic inorganic ligands at the iron atom (one carbonyl and two cyanides). The [NiFe]-hydrogenases have a bias toward the heterolytic splitting of  $H_2$  into protons and electrons. The oxidation of  $H_2$  makes them useable in a fuel cell (“microbial fuel cell”).

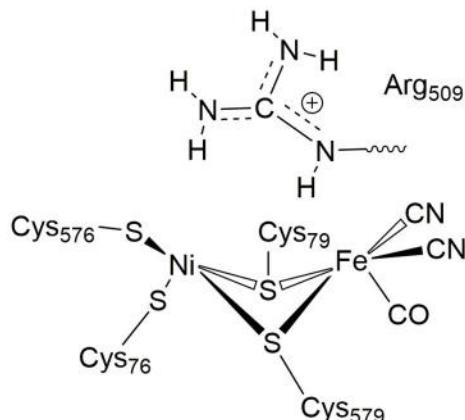
$H_2$  can be used in a fuel cell to generate electric power from its oxidation and the reduction of  $O_2$  to give water. The absence of noble metals and operation conditions at room temperature make [NiFe]-hydrogenase enzymes a system of interest to scientist and engineers, and may act as an inspiration to develop novel bio-inspired catalysts (Du et al., 2007; Cracknell et al., 2008; Santoro et al., 2017). [NiFe]-hydrogenase adsorbed on a pyrolytic graphite electrode catalyzes  $H_2$  oxidation at a diffusion-controlled rate matching that achieved by platinum (Jones et al., 2002).

In [NiFe]-hydrogenases, the “as-isolated” oxidized state contains a hydroxide anion ( $OH^-$ ) binding between the Ni(III) and Fe(II) ions. During the process of  $H_2$  activation, nickel shuttles between Ni(III) and Ni(II) oxidation states whereas Fe remains redox-inactive in a 2+ state of oxidation. The catalytic reaction intermediate, Ni-C, is a Ni(III) Fe(II) species with a  $\mu$ -bridging hydride, but the exact site that acts as the proton acceptor has not been resolved yet. QM and QM/MM calculations have favored one of the terminal cysteines to be the site of protonation (Niu et al., 1999; Lill and Siegbahn, 2009; Hu et al., 2013; Dong and Ryde, 2016; Dong et al., 2018). The ultra-resolution X-ray structure of the fully reduced state of the enzyme, Ni-R, indeed enables to reveal a hydride in the bridging position and one of the terminal cysteines protonated (Ogata et al., 2015b). Recently, however, a non-coordinating amino acid residue was identified to play a major role in  $H_2$  activation by *E. coli* Hyd-1 (Evans et al., 2016). Substitution of a strictly conserved arginine residue (R509)  $\sim 4.4$  Å above the active site nickel (see Figure 2) by a charge-conserving lysine led to a  $>100$ -fold lower activity in comparison to the wildtype enzyme. This led to the hypothesis of the arginine guanidine group acting as the general base in  $H_2$  activation (Carr et al., 2016). This would require R509 to at least be fractionally deprotonated and neutral, in order to be able to play a functional role similar to that of a frustrated Lewis pair (FLP) (Stephan and Erker, 2010).

At neutral pH, only lysine, arginine and sometimes histidine possess sidechains with a positive charge. The pKa-value describes the pH-value at which deprotonated and protonated

forms are in equilibrium and for arginine, a pKa-value of 12 is usually given in textbooks (Hunter and Borsook, 1924; Berg et al., 2002). At pH  $< 12$ , the guanidine nitrogen atom becomes protonated and a positive charge is delocalized via the nitrogen atoms N $\eta$ 1, N $\eta$ 2, and N $\epsilon$  (Figure 2). The protein environment can lead to local deviations of the pKa-values of the amino acid side chains due to strong electrostatic interactions with other fully or partially charged groups as well as the polarity or dielectric constants of the medium that surrounds them. pKa-values of catalytic amino acids in or near the active sites of enzymes may be significantly perturbed by more than 2 units due to structural details and the energetics of the reactions that they catalyze (Harris and Turner, 2002). For arginine, however, no detectable shifts in pKa-values were ever reported and, for example, all buried 25 arginine residues in the staphylococcus nuclease remained in the charged state (Harms et al., 2011). Significant perturbations of pKa-values of arginine residues were only found from free energy perturbation calculations from MD trajectories of an arginine residue in a highly hydrophobic membrane environment (Yoo and Cui, 2008). Only when positioned close to the center of the bulk lipid membrane, an effective pKa-value of 7.7 could be obtained. Thus, significant populations of both the protonated and the neutral forms are only possible near the center of the strongly hydrophobic environment. The protein environment surrounding the conserved arginine in [NiFe]-hydrogenases is however far from being hydrophobic and strong (negative) electrostatics from aspartate residues dominate instead.

In this work, we investigate the possible involvement of both neutral and positively charged R509 in the heterolytic splitting of  $H_2$  by *E. coli* Hyd-1, using QM/MM calculations with various sizes of QM regions. The charge distribution and the energetics of protonation of R509 is almost identical to that of a free arginine residue. Energetically, a simultaneous protonation of R509 and a proton transfer to a nearby aspartate residue is not favorable and the terminal cysteine residue C576 is the preferred proton acceptor. In the R509K mutant enzyme, structural parameters and the charge distribution are only affected to a minor degree. When substrate  $H_2$  binds to the wildtype enzyme, it is slightly polarized by R509, which facilitates the heterolytic splitting and the proton transfer to the nearby terminal cysteine. This effect is absent in the mutant enzyme. Moreover, R509 is expected to exhibit very low flexibility, as it forms strong hydrogen bonds with surrounding aspartate residues. Thus, we can suggest a dual role of the arginine in the second coordination



**FIGURE 2 |** Positioning of a strictly conserved second coordination shell arginine amino acid residue (Arg509) above the active site of [NiFe]-hydrogenases (using the amino acid numbering of the [NiFe]-hydrogenase from *E. coli*).

shell of the hydrogenase enzyme: (i) an electronic function by polarizing the substrate  $H_2$ , and (ii) a structural role by strong electrostatic interactions with negatively charged aspartate amino acid residues and displaying reduced conformational flexibility. By contrast, K509 is not able to form such interactions and it is thus expected to display a higher flexibility than R509, which may kinetically hinder hydrogen conversion in the mutant enzyme.

## COMPUTATIONAL DETAILS

The initial coordinates of the wildtype and R509K variant of *E. coli* Hyd-1 (hereafter referred to as EH1 and K-EH1) were taken from the crystal structures 5A4M and 4UE3, respectively (Evans et al., 2016). In these structures, the active site is in the oxidized “ready” Ni-B state. The oxygen atom coordinated between the Ni and Fe ions (in both protein structures) as well as oxidized C576 (in 5A4M) were thus removed to meet the functional form of the active site in the Ni-SIA state (see Figure 2). The PDB2PQR suite of programs was used to check the orientation of the side chains of Asn, Gln, and His. We used the PROPKA module of PDB2PQR to assign the protonation states of titratable residues (Dolinsky et al., 2004, 2007; Li et al., 2005; Bas et al., 2008). Both subunits (L: large subunit, and S: small subunit) of the enzymes were taken into account during the PDB2PQR procedure. After protonation, the heteroatoms Ni and Fe (including its CO and  $CN^-$  ligands) ions of the active site of the L subunit as well as the [FeS]-clusters of the S subunit were inserted at their respective crystal-structure positions. Then, the protonation states and side chain orientations were carefully checked by visual inspection, and potential shifts of the PROPKA-calculated pKa-values due to the presence of the [NiFe] and [FeS] structural motifs were qualitatively assessed for a final assignment of the protonation states of the surrounding residues. Thereafter, the S subunit was removed, while the L subunit was retained for subsequent QM/MM calculations. Moreover, all

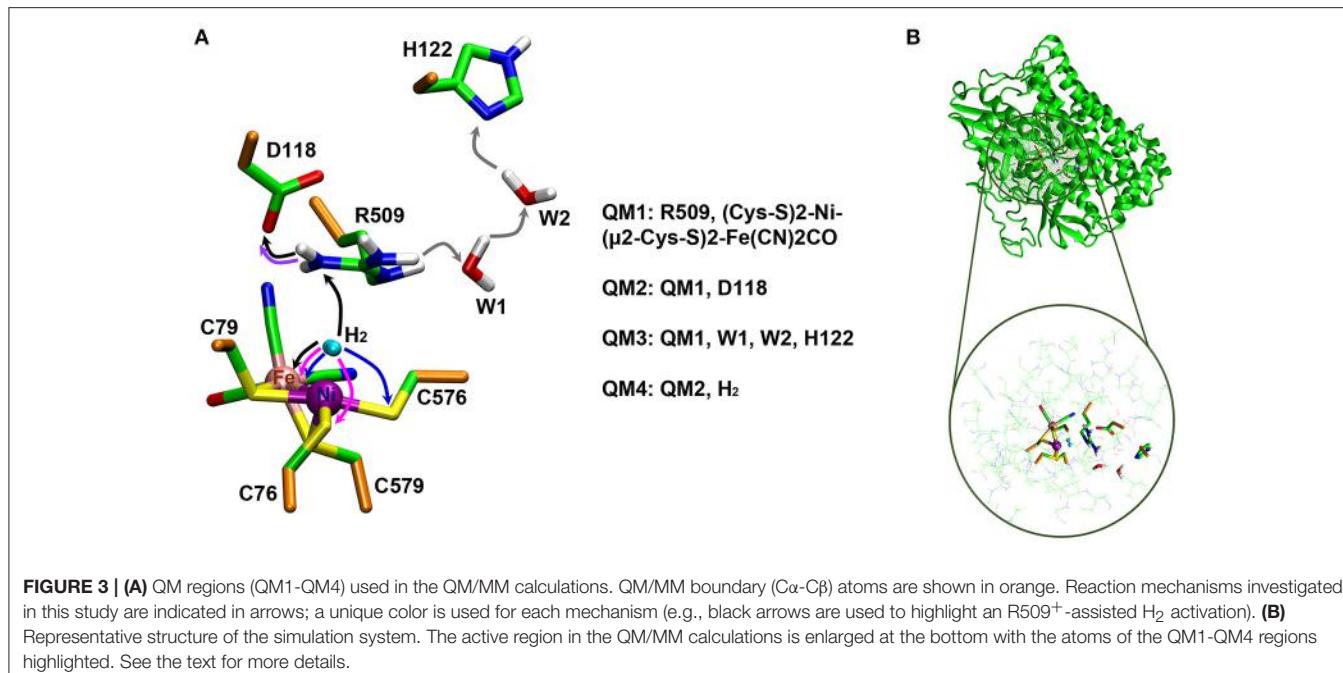
crystal waters of the latter were deleted except for the active-site water molecules (EH1<sub>L</sub>: 9; K-EH1<sub>L</sub>: 10) (Evans et al., 2016). The protonation states of all EH1<sub>L</sub> and K-EH1<sub>L</sub> residues were identical. All acidic residues were negatively charged except for D67, D350, D574, and E73, which were protonated. All lysine and arginine residues were positively charged. Histidine residues were either singly protonated at N<sub>e</sub> (H30, 83, 117, 119, 122, 189, 220, 229, 351, 364, 457, and 514), singly protonated at N<sub>8</sub> (H421, 571, and 582), or doubly protonated (H205). All cysteine residues coordinating to metals were deprotonated (C76, C79, C576, and C579).

EH1<sub>L</sub> and K-EH1<sub>L</sub> were used as starting structures in subsequent QM/MM calculations with different QM regions as described below.

## QM/MM Investigation of the R509 Protonation State and Its Involvement in Proton Transfer

Two QM/MM optimizations of EH1<sub>L</sub> were carried out with R509 being either in neutral (R509<sup>0</sup>) or protonated (R509<sup>+</sup>) form. Among the nitrogen atoms of the guanidinium group of R509, the N<sub>η1</sub> atom is proposed to be the one potentially involved in  $H_2$  activation since it is closer to the NiFe active site (Evans et al., 2016). This nitrogen atom was therefore chosen as deprotonation target to give R509<sup>0</sup>. The QM region (hereafter referred to as QM1 region) consisted of the side chain of R509<sup>+/0</sup> as well as the Ni and Fe ions with their first coordination sphere ligands (CO, two  $CN^-$  groups, and the side chains (thiolate groups) of the four metal-coordinating cysteine residues) (see Figure 3A). The rest of the system (including the active-site water molecules) was treated at the MM level. The total charge of the QM1 region was  $-1$  for R509<sup>+</sup> and  $-2$  for R509<sup>0</sup>. All atoms within 6 Å of the QM1 region were unconstrained during QM/MM optimization whereas the positions of the more distant atoms were kept fixed (see Figure 3B). The QM/MM calculations were performed with the ChemShell<sup>1</sup> package (Sherwood et al., 2003; Metz et al., 2014) (version 3.7). The TURBOMOLE (Ahlrichs et al., 2011) (version 6.6) and DL POLY (Smith and Forester, 1996) (version 4.08) packages were used as QM and MM interfaces, respectively. The DL-FIND optimiser module of ChemShell was used for the optimizations (Kästner et al., 2009). The electrostatic interaction between the QM1 region and the surrounding partial charges was treated using the electrostatic embedding scheme with charge shift correction (Bakowies and Thiel, 1996; de Vries et al., 1999). Hydrogen link atoms were used to saturate the valencies at the covalent bonds crossing the QM/MM boundary (see Figure 3A; Sherwood et al., 1997). DFT was used to describe the QM1 region while the MM region was described by the CHARMM27 force field (MacKerell et al., 1998; Mackerell et al., 2004). Geometries were optimized using BP86 (Slater, 1951; Vosko et al., 1980; Perdew, 1986a,b; Becke, 1988) as the DFT functional with the def2-TZVP (Weigend and Ahlrichs, 2005) basis set. The calculations were sped up by using the resolution-of-identity (RI) approximation (Eichkorn et al., 1995, 1997). Single-point

<sup>1</sup>ChemShell, a Computational Chemistry Shell, see [www.chemshell.org](http://www.chemshell.org)



**TABLE 1** | Definition of structural parameters of  $\text{Arg}509$  coordination in the vicinity of the active site of the *E. coli* [NiFe]-hydrogenase.

	Distances / $\text{\AA}$	X-ray	QM/MM $\text{Arg}509^+$ QM1; QM2	QM/MM $\text{Arg}509^0$ QM1; QM2
$\text{Ni} \dots \text{N}\eta 1$	4.44		4.74; 4.48	4.21; 4.48
$\text{Ni} \dots \text{N}\eta 2$	5.45		5.71; 5.69	5.37; 5.59
$\text{Ni} \dots \text{N}\epsilon$	4.89		4.84; 4.93	4.66; 4.76
$\text{Fe} \dots \text{N}\eta 1$	4.40		4.66; 4.57	4.28; 4.48
$\text{Fe} \dots \text{N}\eta 2$	6.27		6.46; 6.48	6.25; 6.35
$\text{Fe} \dots \text{N}\epsilon$	5.02		5.10; 5.14	4.97; 4.98
$\text{CN} \dots \text{N}\eta 1$	3.22		2.90; 2.93	3.16; 3.11
$\text{CN} \dots \text{N}\eta 2$	5.43		5.15; 5.15	5.36; 5.31

QM/MM optimization results vs. X-ray structural data (PDB code 5A4M).

energy calculations at the BP86-D3 level were performed to check the influence of dispersion (Slater, 1951; Vosko et al., 1980; Perdew, 1986a,b; Becke, 1988; Grimme, 2006). For further validation of consistency, single-point calculations were also carried out using the density functionals TPSSH (Tao et al., 2003) and B3LYP (Slater, 1951; Vosko et al., 1980; Becke, 1988, 1993) with dispersion corrections (Grimme, 2006). From the QM/MM energies of  $\text{EH1}_L\text{-R}509^0$  and  $\text{EH1}_L\text{-R}509^+$ , the energy for the deprotonation of R509 was calculated.

QM/MM calculations were also performed to investigate the ability of  $\text{R}509^+$  to transfer a proton to residues in its close surrounding, and to get insight into potential deprotonation mechanisms leading to formation of  $\text{R}509^0$ . Two different mechanisms were considered: (i) a proton transfer from the  $\text{R}509^+:\text{N}\eta 1$  atom to  $\text{D}118:\text{O}\delta$ , and (ii) a water-mediated proton transfer from  $\text{R}509^+:\text{N}\eta 2$  to  $\text{H}122:\text{N}\delta$ . The reactant ( $\text{R}509^+$ ) and expected product for such reaction mechanisms were optimized to calculate the respective reaction energies. The calculations were carried out using the same QM/MM setup described above.

The only difference was in the QM region. For the proton transfer to  $\text{D}118$ , the side chain of the latter was also included as part of the QM region, whereas for the water-mediated proton transfer the side chain of  $\text{H}122$  and two active-site water molecules were included in the QM region. The resulting QM regions are referred to as QM2 (with a charge of  $-2$ ) and QM3 (with a charge of  $-1$ ), respectively (see Figure 3A). Initially, the reactant state was optimized. Then the optimized reactant structures were used as starting points to manually build the products (by relocation of the respective protons) and optimize them.

## QM/MM Investigation of the Thermodynamics of $\text{H}_2$ Activation

We also performed QM/MM calculations to compute the thermodynamic profiles of the  $\text{EH1}_L$  catalyzed  $\text{H}_2$  oxidation with both  $\text{R}509^0$  and  $\text{R}509^+$  acting as a base in the reaction. The potential ability of  $\text{R}509^+$  to mediate  $\text{H}_2$  dissociation was considered to be assisted by a strong  $\text{R}509^+:\text{N}\eta 1\text{-D}118:\text{O}\delta$  hydrogen bond, which could make  $\text{R}509^+:\text{N}\eta 1$  slightly

nucleophilic and thus enable a double proton transfer ( $\text{H}^+ \rightarrow \text{R509}^+ \rightarrow \text{D118}$ ) reaction. For comparison, we computed the reaction energies with C76 and C576 as the  $\text{H}_2$  proton acceptors, for both  $\text{EH1}_L$  and  $\text{K-EH1}_L$ . The reactant  $[(\text{K})\text{EH1}_L \cdot \text{H}_2]$  complexes were optimized first. The latter were built by manual docking of  $\text{H}_2$  into the  $\text{EH1}_L$  and  $\text{K-EH1}_L$  active sites;  $\text{H}_2$  was positioned between Ni and either R509 or K509 and then fully optimized without any additional constraints. Again, the optimized reactants served as starting points to build the products, which were then also fully optimized. This time the QM/MM optimizations were carried out using a QM region (referred to as QM4) which includes the QM2 components as well as the  $\text{H}_2$  atoms (see **Figure 3A**). All geometry optimizations were carried out considering both Ni and Fe to be in a closed-shell low-spin singlet state. Though the spin state of  $\text{Ni}^{2+}$  is a controversial topic, recent computational and experimental studies on [NiFe]-hydrogenases using advanced and accurate methods (e.g., coupled cluster calculations and subatomic resolution protein crystallography) support the singlet state to be preferred (Bruschi et al., 2014; Delcey et al., 2014; Ogata et al., 2015a,b, 2016; Dong et al., 2017). Moreover, the protonation site of the enzyme during  $\text{H}_2$  splitting as identified by subatomic resolution X-ray crystallography (Ogata et al., 2015b), was perfectly reproduced by computational calculations carried out with Ni in a single state (Dong and Ryde, 2016). For further validation in this regard, we have carried out single point calculations of the optimized geometries with Ni in a triplet state. Unless mentioned otherwise all energies reported in this manuscript are given with respect to the reactant complex and correspond to the computed energies for Ni in a singlet state which is shown to be the ground state (see below).

Atomic charge distributions in the QM region were calculated by the Mulliken population analysis approach in order to allow to make general statements about trend rather than absolute numbers (Mulliken, 1955a,b).

## RESULTS

### Structural Parameters and Charge Distributions

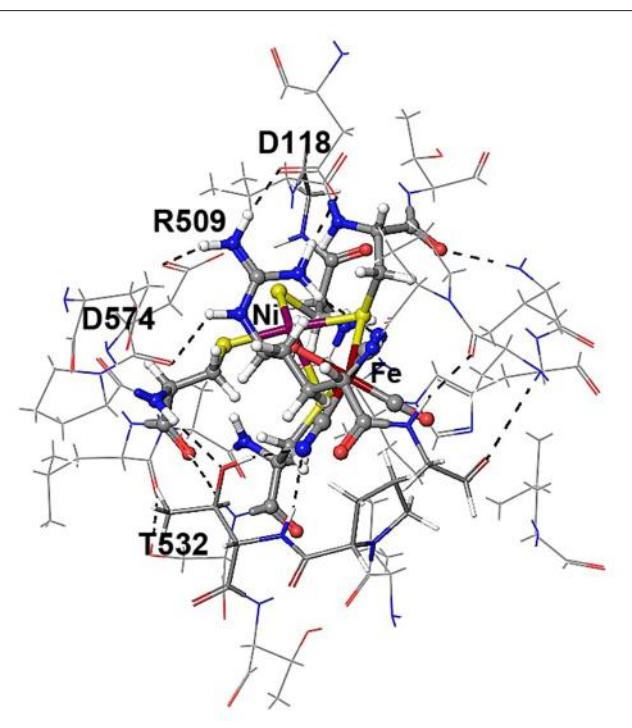
**Table 1** compares the structural parameters of the relevant interactions of R509 with the active site as obtained from the QM/MM calculations using the QM regions 1 and 2 (see above) to those obtained from X-ray crystallography. The BP86 functional was shown to give reliable structural parameters in good agreement with experiment for the active site of [NiFe]-hydrogenases (Stein et al., 2001; Stein and Lubitz, 2004) and other transition metal complexes (Bühl and Kabrede, 2006). Overall, the QM(BP86)/MM(CHARMM) calculations reproduce well the interatomic distances corresponding to the non-covalent interactions between R509 and the Ni-Fe catalytic core. This is more difficult to achieve compared to covalent bonding, since structural flexibility and weak packing interactions must be treated equally well.

The  $\text{N}_{\text{H}1}$  amine group forms a hydrogen bond with one of the cyanide ligands of the Fe atom ( $\text{CN} \cdots \text{N}_{\text{H}1}$  distance

$\sim 3 \text{ \AA}$ ) as well as strong electrostatic interactions with the negatively charged aspartate residue D118 (see **Figure 4**). When D118 is incorporated into the QM region (QM2), the structural parameters are in better agreement with experiment. This shows that an appropriate choice of the QM size is critical for an accurate description of long range interactions in an enzyme.

When R509 is deprotonated, there are only minor structural differences to be seen (**Table 1**). The variations are within the accuracy of the computational method and give no additional information as to the protonation state of R509 in the crystal structure.

**Table 2** provides the partial charges of the R509 atoms as obtained from the QM/MM calculations using the Mulliken population analysis approach (Mulliken, 1955a,b) and the QM regions 1 and 2, and compares them to those calculated for a free arginine. The formation of a delocalized double bond (a partial charge character) makes the central  $\text{C}_{\text{H}}$  atom positively charged with 0.45 in a free arginine residue and 0.33 and 0.37 in QM regions 1 and 2, respectively. The  $\text{N}_{\text{H}1}$  and  $\text{N}_{\text{H}2}$  atoms are overall chemically equivalent in the free arginine with charges of  $-0.41$  and  $-0.42$ . In the calculation with the QM1 region, charges of  $-0.5$  and  $-0.53$  indicate a stronger polarization due to



**FIGURE 4** | Details of hydrogen bonding interactions of the active site of *E. coli* Hyd-1 with surrounding amino acid residues of the large subunit with a focus on the arginine residue 509. The  $\text{N}_{\text{H}1}$  hydrogen atoms form hydrogen bonds with aspartate 118 and one of the cyanide ligands. The  $\text{N}_{\text{H}2}$  hydrogen atoms make interactions with the aspartate residues 118 and 574. The terminal cyanide ligands are hydrogen bound to R509 side chain  $\text{NH}_2$  and main chain NH protons; the second cyanide group accepts hydrogen bonds from the OH side chain and the NH main chain protons of Thr532.

**TABLE 2** | Charge distributions of a free arginine residue and the residue Arg509 of the *E. coli* [NiFe]-hydrogenase in their standard and neutral protonation states.

	Atom	Free Arg <sup>+</sup> /Arg <sup>0</sup>	QM/MM Arg509 <sup>+</sup> /Arg509 <sup>0</sup> QM1; QM2
	N <sub>η</sub> 1	-0.413/-0.415	-0.498/-0.441; -0.400/-0.433
	N <sub>η</sub> 2	-0.422/-0.473	-0.527/-0.623; -0.462/-0.560
	N <sub>ε</sub>	-0.224/-0.285	-0.318/-0.328; -0.313/-0.328
	C <sub>ζ</sub>	0.447/0.247	0.333/0.289; 0.370/0.258

**TABLE 3** | Calculated deprotonation energies of R509 of *E. coli* Hyd-1 using QM/MM calculations in kcal/mol.

Level of Theory	Deprotonation Energy
BP86	154.1
BP86-D3	156.3
B3LYP-D3	155.5
TPSSH-D3	158

interactions with surrounding residues. When D118 is explicitly included in the QM region (QM2), charges of -0.40 and -0.46 are obtained. This shows that the explicit QM electrostatic interaction of residue D118 leads to a slightly chemical non-equivalence of the N<sub>η</sub> atoms of R509. The N<sub>ε</sub> is significantly less negatively charged (-0.22 in free arginine, -0.32 in QM1, and -0.31 in QM2) than the other nitrogen atoms. In the neutral form R509<sup>0</sup>, C<sub>ζ</sub> becomes significantly less positively charged (+0.25) in free arginine but the effect is less pronounced when embedded in the protein (+0.29 and +0.26 in the different QM regions). Upon deprotonation of N<sub>η</sub>1, its charge changes only marginally both in the free residue (by 0.01) and the one in the protein (by 0.03–0.06). For N<sub>η</sub>2, on the other hand, the negative charge increases by 0.05 in the free residue and 0.09–0.10 in the protein. Thus, N<sub>η</sub>2 becomes more nucleophilic in neutral arginine. In the enzyme, this nitrogen atom is however at a too large distance from Ni (>6 Å) so that a direct involvement in hydrogen activation is not feasible.

In conclusion, we cannot report a significant perturbation of the charge distribution in R509 with respect to a free arginine residue. We therefore do not expect this residue to display a significantly perturbed pKa-value.

On the other hand, we calculate the reaction energy for the deprotonation of R509 from the computed QM/MM energies of EH1<sub>L</sub>-R509<sup>+</sup> and EH1<sub>L</sub>-R509<sup>0</sup>. We assume that the free proton is able to diffuse out of the protein (EH1<sub>L</sub>-R509<sup>+</sup> ↔ EH1<sub>L</sub>-R509<sup>0</sup> + H<sup>+</sup>) by considering the free energy of solvation of H<sup>+</sup> in water [ $\Delta G_{\text{solv}}(H^+)$  = -265.9 kcal/mol; (Tawa et al., 1998; Tissandier et al., 1998; Topol et al., 2000; Kelly et al., 2007; Rebollar-Zepeda and Galano, 2012)]. The results are shown in **Table 3**, where a positive deprotonation energy indicates the thermodynamic preference for protonated R509<sup>+</sup>. All DFT calculations give consistent results for the thermodynamic equilibrium between R509<sup>+</sup> and R509<sup>0</sup>. The differences between different functionals are within 4 kcal/mol. The protonated form R509<sup>+</sup> in the protein is energetically favored by 154–158 kcal/mol. This large deprotonation energy shows that in the [NiFe]-hydrogenase the

arginine residue R509 close to the active site is predominantly in its protonated form. Meanwhile, attempts to calculate the energy difference between R509<sup>+</sup> and R509<sup>0</sup> with either D118 or H122 as a proton acceptor (see **Figure 3**) were unsuccessful; reversion to the zwitterionic state of R509 occurred immediately in the QM/MM optimizations involving R509<sup>0</sup>.

It should be noted that we are not attempting to calculate standard and perturbed pKa-values here, since it requires of a more robust computational approach to be implemented, including e.g., high level electronic structure methods, explicit solvent coordination with a certain number of solvent molecules plus continuum solvation, consideration of entropic contributions, and conformational sampling (Ghosh and Cui, 2008; Rebollar-Zepeda and Galano, 2012; Uddin et al., 2013). It can only be stated that in *E. coli* Hyd-1, the amount of neutral R509 is negligible and the thermodynamic equilibrium is far toward a positively charged residue in its standard protonation state. Thus, the direct involvement of the neutral form of R509 as a strong FLP in H<sub>2</sub> oxidation appears impossible.

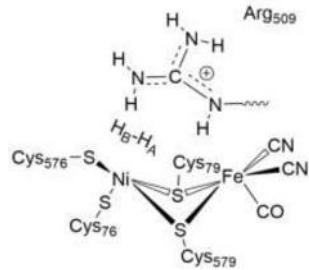
## The Substrate Bound Complex

Ni-SIa is the catalytically active species which performs hydrogen oxidation. In Ni-SIa, H<sub>2</sub> approaches the Ni site where it is heterolytically cleaved. The hydride occupies the  $\mu$ -bridging position between the Ni and Fe atoms and one residue in the vicinity must act as a proton acceptor.

**Table 4** gives structural data for the H<sub>2</sub> Ni-SIa complexes corresponding to the EH1<sub>L</sub>-R509<sup>+</sup> wildtype enzyme and the R509K variant. Attempts to compute structural parameters and reaction energies for EH1<sub>L</sub>-R509<sup>0</sup> were unsuccessful, since the optimization of the respective reactant complex evolved spontaneously toward the product with a  $\mu$ -hydride and a protonated arginine. This reinforces our conclusion on the preference of R509 to be in the protonated state.

In the reactant complex (RC<sub>EH1</sub>), H<sub>2</sub> is located above and very close to the Ni ion (at 1.6 Å), whereas the distance between the Fe ion and H<sub>2</sub> is longer (2.5 Å). This is in agreement with previous studies on [NiFe] hydrogenases which show that H<sub>2</sub> prefers to bind to Ni, rather than to Fe (Ogata et al., 2002; Dong et al., 2017). Upon H<sub>2</sub> binding, the relevant interatomic distances between the active site and R509 are overall unchanged compared to the Ni-SIa state (see **Tables 1, 4**). Moreover, structural parameters at the reactant complex are similar for the wildtype and the mutant enzyme. Nickel-nitrogen and iron-nitrogen distances as well as the interactions of H<sub>2</sub> with nickel, iron, or cysteine residues do not change overall. This indicates that upon the arginine-to-lysine mutation, the active center

**TABLE 4 |** Structural data of the H<sub>2</sub> Ni-Sla complexes of the wildtype *E. coli* Hyd-1 and the R509K variant optimized using the larger QM2 region.

	Arg <sub>509</sub>	Distances/Å	Wildtype R509 <sup>+</sup>	Mutant R509K
				
		Ni ... N <sub>η</sub> 1	4.4	4.5 <sup>a</sup>
		Ni ... N <sub>η</sub> 2	5.5	b
		Ni ... N <sub>ε</sub>	4.7	b
		Fe ... N <sub>η</sub> 1	4.6	4.9 <sup>a</sup>
		Fe ... N <sub>η</sub> 2	6.4	b
		Fe ... N <sub>ε</sub>	5.1	b
		CN ... N <sub>η</sub> 1	2.9	3.0 <sup>a</sup>
<b>H<sub>2</sub>-BINDING</b>				
		Ni ... H <sub>A</sub>	1.6	1.6
		Ni ... H <sub>B</sub>	1.6	1.6
		Fe ... H <sub>A</sub>	2.5	2.4
		Fe ... H <sub>B</sub>	3.2	3.1
		N <sub>η</sub> 1 ... H <sub>A</sub>	2.9	3.1 <sup>a</sup>
		N <sub>η</sub> 1 ... H <sub>B</sub>	2.9	3.0 <sup>a</sup>
		C76:S ... H <sub>A</sub>	3.6	3.6
		C76:S ... H <sub>B</sub>	3.1	3.1
		C576:S ... H <sub>A</sub>	2.9	2.9
		C576:S ... H <sub>B</sub>	2.7	2.7

<sup>a</sup>Distances to the N<sub>η</sub> atom of lysine. <sup>b</sup>Not applicable.

remains fully assembled and structurally intact to perform the catalytic hydrogen oxidation. This is in agreement with structural and spectroscopic investigations (Evans et al., 2016).

The electronic structure also only changes slightly in the R509K mutant (Table 5). The N<sub>η</sub>1 atom becomes less negative (−0.31) and thus less nucleophilic and will be then a weaker proton acceptor when H<sub>2</sub> is heterolytically split. Atomic charges at Ni, Fe and all other active site atoms remain overall unchanged (see Table 5). What becomes apparent by analyzing the partial charges is the fact that in the wildtype enzyme both hydrogen atoms of H<sub>2</sub> are slightly positively polarized (+0.07). In the lysine mutant, however, the hydrogen atoms become less and oppositely charged, with the hydrogen atom pointing toward the bridging position positive (H<sub>A</sub>, +0.02) and the distal hydrogen between lysine and cysteine negative (H<sub>B</sub>, −0.02). Since H<sub>A</sub> will become the bridging hydride, H<sub>B</sub> ought to be accepted by a (negatively charged) proton acceptor. This indicates that the introduction of a lysine residue does not structurally impair the catalytic function, but it reduces the proton affinity of a potential proton acceptor nitrogen atom and at the same time leads to a partial negative charge on the putative protonic species.

## The Thermodynamics of H<sub>2</sub> Oxidation

The QM/MM energies calculated for the binding and heterolytic splitting of H<sub>2</sub> by EH1<sub>L</sub>-R509<sup>+</sup> and K-EH1 are shown in Table 6 for a series of different functionals. Since we could not obtain a stationary intermediate for the EH1<sub>L</sub>-R509<sup>0</sup> H<sub>2</sub> complex, those energies cannot be reported. All DFT calculations give a very consistent picture of the energetics of H<sub>2</sub> binding and splitting.

**TABLE 5 |** Charge distributions in the H<sub>2</sub> Ni-Sla complexes of the wildtype *E. coli* Hyd-1 and the R509K mutant optimized using the larger QM2 region.

	Arg <sub>509</sub>	Atom	Wildtype R509 <sup>+</sup>	Mutant R509K
		N <sub>η</sub> 1	−0.402	−0.308 <sup>a</sup>
		N <sub>η</sub> 2	−0.466	b
		N <sub>ε</sub>	−0.296	b
		Ni	−0.251	−0.225
		Fe	0.147	0.156
		C76:S	−0.488	−0.475
		C576:S	−0.041	−0.041
		H <sub>A</sub>	0.069	0.020
		H <sub>B</sub>	0.065	−0.018

<sup>a</sup>N<sub>η</sub> atom of lysine (K509). <sup>b</sup>Not applicable.

**TABLE 6 |** Substrate binding energies and thermodynamics of heterolytic H<sub>2</sub> splitting (in kcal/mol) in the wildtype and R509K mutant [NiFe]-hydrogenase from *E. coli*.

	BP86	BP86-D3	B3LYP-D3	TPSSH-D3
<b>Wildtype</b>				
H <sub>2</sub> binding	−2.8	−9.4	−4.4	−8.4
<b>HETEROlytic H<sub>2</sub> SPLITTING. HA-Ni-μH<sup>−</sup>-FE</b>				
<b>Proton acceptor A</b>				
Arginine 509 <sup>+</sup>	19.3	16.2	13.9	15.1
Cysteine 76	−3.7	−2.2	−9.9	−7.2
Cysteine 576	−8.4	−7.0	−15.3	−12.2
<b>K509 Mutant</b>				
H <sub>2</sub> binding	−1.6	−7.5	−2.7	−6.7
<b>HETEROlytic H<sub>2</sub> SPLITTING. HA-Ni-μH<sup>−</sup>-FE</b>				
<b>Proton acceptor A</b>				
Cysteine 76	−11.9	−11.8	−19.1	−16.2
Cysteine 576	−10.1	−9.4	−17.3	−14.4

The hydride occupies the μ-bridging position between the Ni...Fe atoms.

This provides a reliable insight into the thermodynamics of H<sub>2</sub> oxidation and an estimate of the uncertainty of the computed energies.

As can be seen in Table 6, our calculations indicate that a potential heterolytic splitting of H<sub>2</sub> involving participation of R509<sup>+</sup> as a proton acceptor (H<sub>2</sub> → R509<sup>+</sup> → D118) is thermodynamically unfavorable by 13.9–19.3 kcal/mol. By contrast, C76 and C576 are found to be able to act as proton acceptors and mediate this process favorably, the reaction being exothermic by 3.7–9.9 and 7.0–15.3 kcal/mol, respectively. At all levels of theory evaluated protonation of C576 is thermodynamically more favorable than that of C76; the reaction energy associated with C576 is 4.7–5.4 kcal/mol lower in comparison to C76. This is in line with previous computational and X-ray crystallography studies on [NiFe]-hydrogenase from *Desulfovibrio vulgaris* Miyazaki F, which show that protonation of C546 (the equivalent for C576 in EH1) is preferred over the

other coordinating cysteines (Ogata et al., 2015b; Dong and Ryde, 2016).

The structures of the optimized  $\text{H}_2\ldots\text{Ni-SI}_a$  complexes are shown in **Figure 5**. Apart from a different positioning of the  $\text{H}_2:\text{H}_\text{B}$  atom (see **Figure 5**), there are only a few structural differences between the optimized product complexes. In  $\text{PC76}_{\text{EH1}}$  (protonated cysteine C76), the orientation of the side chain of the residue E28 (located in the MM region) is different and is better stabilized by the surrounding residues (via hydrogen bonds) in comparison to both  $\text{PC576}_{\text{EH1}}$  (protonated cysteine C576) and  $\text{PC509}_{\text{EH1}}$  (product complex for a proton transfer to  $\text{R509}^+$ ). This is also true when comparing  $\text{PC76}_{\text{EH1}}$  and  $\text{RC}_{\text{EH1}}$ . Therefore, the hydrogen bonding interactions of E28 in  $\text{PC76}_{\text{EH1}}$  are considered to be important for the exothermic formation of this product, which is supported by the lower value of the MM energy component with respect to that for  $\text{RC}_{\text{EH1}}$  (see Supplementary Material for a detailed analysis of the QM and MM energy contributions to the QM/MM energies). (Senn and Thiel, 2009; Escoria et al., 2017) Meanwhile,  $\text{PC509}_{\text{EH1}}$  differs from  $\text{PC76}_{\text{EH1}}$  and  $\text{PC576}_{\text{EH1}}$  regarding geometry and orientation of the side chain of  $\text{R509}^+$ . The characteristic planar geometry of the guanidinium group is distorted in  $\text{PC509}_{\text{EH1}}$ . In addition, the hydrogen bond interactions with the surrounding aspartate residues (D118 and D574) are overall weaker. Together these terms may contribute significantly to the endothermic formation of  $\text{PC509}_{\text{EH1}}$ , as given by the higher value of the QM energy component in comparison to  $\text{PC76}_{\text{EH1}}$  and  $\text{PC576}_{\text{EH1}}$  (see 'Supplementary Material').

According to these results, R509 is not expected to be directly involved in the reaction mechanism of the  $\text{H}_2$  activation by EH1. Instead, we propose this residue to be important for  $\text{H}_2$  activation by guiding its access to the active site, promoting its binding to nickel and facilitating its polarization. Our conclusions are based on the QM/MM calculations with  $\text{K-EH1}_\text{L}$ . As can be seen in **Table 6**,  $\text{H}_2$  activation by  $\text{K-EH1}_\text{L}$  is

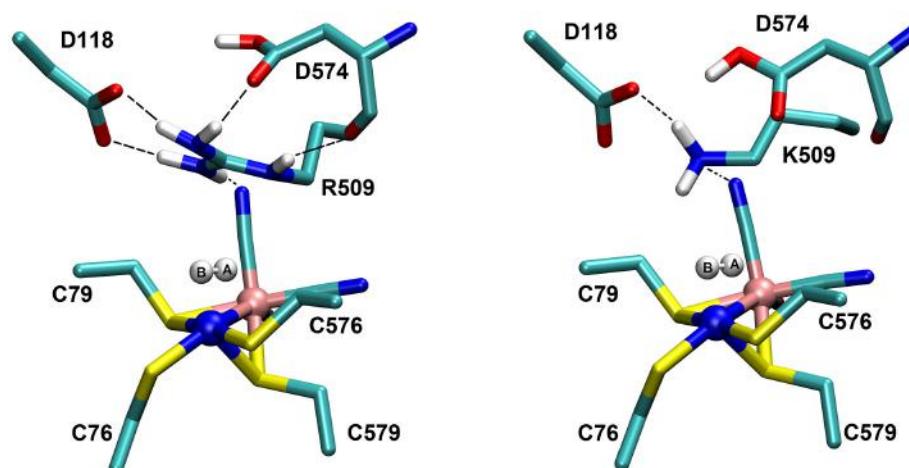
also thermodynamically feasible with either C76 or C576 acting as a base. The computed reaction energies also suggest the process to be thermodynamically comparable with respect to the EH1 wildtype enzyme. This shows that a mere thermodynamic argumentation cannot explain the high activity of EH1 and the  $>100$ -fold reduction in  $\text{K-EH1}_\text{L}$ .

Also, the binding energy of  $\text{H}_2$  is favored by 1.2 kcal/mol in the wildtype EH1. As described above, the charge analysis showed that the  $\text{H}_2:\text{H}_\text{B}$  atom is more polarized in  $\text{RC}_{\text{EH1}}$  than in  $\text{RC}_{\text{K-EH1}}$ , with an atomic charge value of 0.07 and  $-0.02$ , respectively (**Table 5**). Considering this and the similar negative charge of the sulfur atoms of both C76 and C576, the  $\text{H}_2$  splitting is expected to be kinetically favored (i.e. with a lower energy barrier) in EH1.

The most important structural differences between EH1 and  $\text{K-EH1}$  are found to be in the immediate vicinity of the R509 and K509 residues. The former is strongly stabilized by the surrounding aspartate residues through hydrogen bond interactions (**Figure 5**), which hold R509 in place and may make this residue more rigid in comparison to lysine. The smaller spatial extension and a potential higher degree of flexibility of the side chain of K509 may account for a weaker binding of  $\text{H}_2$  as well as a correct positioning and polarization of the latter in the wildtype enzyme. The investigation of the effect of the flexibility of the side chain on the kinetics of the  $\text{H}_2$  splitting will require extensive QM/MM MD simulations with a sufficient degree of conformational sampling.

All energies discussed up to this point were obtained with Ni in a low-spin singlet state ( $S = 0$ ).

The spin state of the EPR-silent  $\text{Ni}^{2+}$  intermediate state is still a controversial issue. Recent computational studies on [NiFe]-hydrogenases using state-of-the-art methods (e.g., coupled cluster calculations and DMRG) support the singlet state to be preferred over the triplet in a cluster model of the active site (Dong et al., 2017). BP86 gave a low spin Ni(II) state for



**FIGURE 5 |** Details of the substrate  $\text{H}_2$  binding in the Ni-SI<sub>a</sub> state of the catalytic site of the [NiFe]-hydrogenase. **(Left)**: In the wildtype *E. coli* Hyd-1, the arginine residue 509 makes strong hydrogen bonding interactions with the aspartates 118 and 574. **(Right)**: In the R509K mutant, explicit hydrogen bond formation is not given and the electrostatic interactions with the aspartate residues are less pronounced.

NiSI (Stein et al., 2001; Stein and Lubitz, 2004) and was shown to be close to the DMRG results in terms of spin state splitting energies. Meanwhile, B3LYP gave reasonable thermodynamics for H<sub>2</sub> splitting but did not perform well for triplet vs. singlet Ni(II) spin state energies (Dong et al., 2017). With the B3LYP functional, a high spin Ni(II) was found to be the ground state in an earlier study (Pardo et al., 2006). For Ni(II) tetrathiolate complexes, the singlet-triplet energy splitting is very sensitive to the amount of exact Hartree-Fock exchange. A reduction to 0.15 in B3LYP\* gave an improved description of the relative spin state ordering for Ni(II)S4 model complexes and [NiFe]-hydrogenase active site models (Bruschi et al., 2004). On the other hand, the TPSSH functional with 0.1 of exact exchange gave reliable structural parameters and bond energies for a set of 80 transition-metal-containing complexes. Furthermore, TPSSH provided reliable energies when tested against typical bioinorganic reactions including spin inversion and electron affinity in iron–sulfur clusters, and breaking or formation of bonds in iron proteins and cobalamins (Jensen, 2008). Thus, we have additionally carried out QM/MM calculations for the thermodynamics of H<sub>2</sub> splitting with Ni(II) in a triplet state (S = 1).

We compare the relative spin ordering of singlet and triplet spin states for the BP86, B3LYP, and TPSSH functionals. As shown in **Table 7**, the results from all functionals are absolutely consistent and indicate that the reactivity on the singlet state spin surface is favored over the triplet state surface for both the wildtype and the mutant enzyme, by 13–23 kcal/mol.

## CONCLUSIONS

Hydrogen oxidation by *E. coli* Hyd-1 was investigated by QM/MM calculations. Substitution of a highly conserved arginine amino acid residue by a charge conserving lysine does not affect the structural parameters and the electronic structure of the active site. The active site is fully assembled and pre-formed for catalysis. The introduction of lysine, however, leads to an unfavorable polarization of the substrate H<sub>2</sub> and makes proton transfer to a negatively charged terminal cysteine kinetically impaired. This explains the reduction of activity by

**TABLE 7 |** Singlet-triplet spin state splitting energies (in kcal/mol) from QM/MM calculations.

	BP86	BP86-D3	B3LYP-D3	TPSSH-D3
<b>Wildtype</b>				
RC	22.7	22.7	20.9	21.6
PC76	19.2	19.1	15.4	15.5
PC576	18.7	18.7	13.2	13.7
<b>K509 Mutant</b>				
RC	22.8	22.8	21.1	21.8
PC76	20.6	20.5	16.6	16.8
PC576	20.3	20.3	15.0	15.6

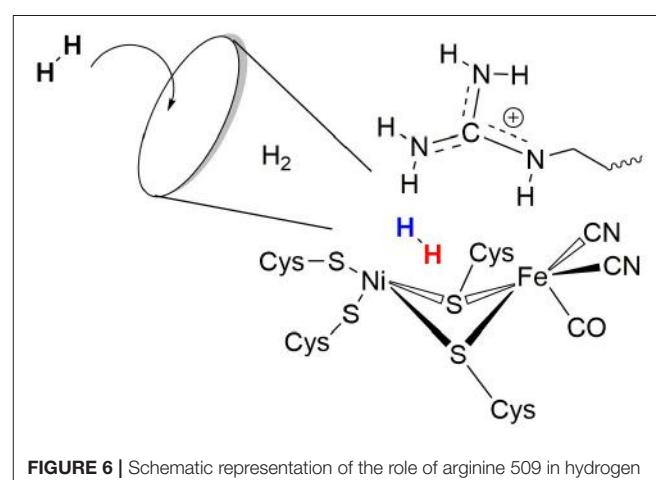
A positive energy indicates the singlet state of Ni(II) to be the ground state.  
RC: reactant complex (H<sub>2</sub>@Ni-Sla complex). PC76: protonated cysteine C76. PC576: protonated cysteine C576.

two orders of magnitude in the K-EH1 mutant enzyme. It was initially suggested that a neutral arginine R509<sup>0</sup> might directly be involved in catalysis and act as a FLP for proton acceptance.

FLPs were identified to be highly effective in activating a variety of small molecules and prompted strong interest in their investigation, e.g., the activation of molecular hydrogen in the absence of a transition metal catalyst by a FLP was also reported (Stephan and Erker, 2010). The concept of FLPs has also been applied to the design of model systems for the active sites of the transition metal-containing hydrogenases. DuBois, Bullock, and co-workers (Raugei et al., 2012) developed enzyme model systems that combine a metal center with non-coordinating amine donor ligands. These pendant, neutral amine groups act in concert with the Ni center to give rise to electrochemical H<sub>2</sub> oxidation and the authors directly note the analogy to FLPs.

In the [NiFe]-hydrogenase enzyme, we cannot verify the existence of a neutral arginine amino acid residue close to the active site. This would be possible only if the pKa-value of that residue was strongly perturbed by the interactions with the protein environment. According to our findings, the charge distribution of R509 in the enzyme is very close to that of a free arginine amino acid residue and the deprotonation energy too high to enable generation of a neutral arginine. In biological catalysis, it is the positively charged side-chain guanidinium group that is often utilized as an electrophilic catalyst with a very high pKa-value ( $\geq 12$ ). In the heterolytic cleavage of H<sub>2</sub>, a hydride occupies the  $\mu$ -bridging position between the Ni and Fe atoms. Rather, a nucleophilic proton acceptor is necessary to take up the product proton. The positively charged R509 residue can still facilitate H<sub>2</sub> splitting via polarization of the latter due to interactions with the partially negatively charged (nucleophilic) N $\gamma$ 1 atom.

The role of the conserved arginine in hydrogenases is thus three-fold: (i) strong electrostatic interactions with nearby aspartate amino acid residues enable an easy H<sub>2</sub> access to the Ni atom with an access channel radius of  $\sim 4\text{\AA}$  (see **Figure 6**); (ii) the arginine assists the positioning and polarization of H<sub>2</sub> to



**FIGURE 6 |** Schematic representation of the role of arginine 509 in hydrogen binding and oxidation by [NiFe]-hydrogenases.

enable a swift proton transfer to one of the terminal cysteines; and (iii) the strong electrostatic interactions with the protein environment keep the arginine in a rigid position and obstruct any conformational changes which otherwise might impede catalysis (see **Figure 5**).

## AUTHOR CONTRIBUTIONS

MS: designed and initiated the project; AE: performed the calculations; MS and AE: analyzed and interpreted the data, wrote the manuscript, and approved the final version.

## REFERENCES

- Ahrlrichs, R., Bar, M., Baron, H.-P., Bauernschmitt, S., Bocker, S., Ehrig, M., et al. (2011). *TURBOMOLE V6.3 2011*, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007. Available online at: <http://www.turbomole.com>
- Bakowies, D., and Thiel, W. (1996). Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.* 100, 10580–10594. doi: 10.1021/jp9536514
- Bas, D. C., Rogers, D. M., and Jensen, J. H. (2008). Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins* 73, 765–783. doi: 10.1002/prot.22102
- Becke, A. D. (1988). Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* 38, 3098–3100. doi: 10.1103/PhysRevA.38.3098
- Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98, 5648–5652. doi: 10.1063/1.464913
- Berg, J. M., Tymoczko, J. L., and Stryer, L. (2002). *Biochemistry, 5th Edn.* New York, NY: W H Freeman.
- Bruschi, M., Gioia, L. D., Zampella, G., Reiher, M., Fantucci, P., and Stein, M. (2004). A theoretical study of spin states in Ni-S4 complexes and models of the [NiFe] hydrogenase active site. *J. Biol. Inorg. Chem.* 9, 873–884. doi: 10.1007/s00775-004-0588-2
- Bruschi, M., Tiberti, M., Guerra, A., and De Gioia, L. (2014). Disclosure of key stereoelectronic factors for efficient H2 binding and cleavage in the active site of [NiFe]-hydrogenases. *J. Am. Chem. Soc.* 136, 1803–1814. doi: 10.1021/ja408511y
- Bühl, M., and Kabrede, H. (2006). Geometries of transition-metal complexes from density-functional theory. *J. Chem. Theory Comput.* 2, 1282–1290. doi: 10.1021/ct6001187
- Cammack, R., Frey, M., and Robson, R. (2001). *Hydrogen as a Fuel: Learning from Nature*. London: Taylor & Francis
- Carr, S. B., Evans, R. M., Brooke, E. J., Wehlin, S. A. M., Nomerotskaia, E., Sargent, F., et al. (2016). Hydrogen activation by [NiFe]-hydrogenases. *Biochem. Soc. Trans.* 44, 863–868. doi: 10.1042/BST20160031
- Cracknell, J. A., Vincent, K. A., and Armstrong, F. A. (2008). Enzymes as working or inspirational electrocatalysts for fuel cells and electrolysis. *Chem. Rev.* 108, 2439–2461. doi: 10.1021/cr0608639
- Delcay, M. G., Pierloot, K., Phung, Q. M., Vancoillie, S., Lindh, R., and Ryde, U. (2014). Accurate calculations of geometries and singlet-triplet energy differences for active-site models of [NiFe] hydrogenase. *Phys. Chem. Chem. Phys.* 16, 7927–7938. doi: 10.1039/C4CP00253A
- de Vries, A. H., Sherwood, P., Collins, S. J., Rigby, A. M., Rigo, M., and Kramer, G. J. (1999). Zeolite structure and reactivity by combined quantum-chemical-classical calculations. *J. Phys. Chem. B* 103, 6133–6141. doi: 10.1021/jp9913012
- Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., et al. (2007). PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* 35, W522–W525. doi: 10.1093/nar/gkm276
- Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* 32, W665–W667. doi: 10.1093/nar/gkh381
- Dong, G., Phung, Q. M., Hallaert, S. D., Pierloot, K., and Ryde, U. (2017). H2 binding to the active site of [NiFe] hydrogenase studied by multiconfigurational and coupled-cluster methods. *Phys. Chem. Chem. Phys.* 19, 10590–10601. doi: 10.1039/C7CP01331K
- Dong, G., and Ryde, U. (2016). Protonation states of intermediates in the reaction mechanism of [NiFe] hydrogenase studied by computational methods. *J. Biol. Inorg. Chem.* 21, 383–394. doi: 10.1007/s00775-016-1348-9
- Dong, G., Ryde, U., Jensen, H. J. A., and Hedegård, E. D. (2018). Exploration of H2 binding to the [NiFe]-hydrogenase active site with multiconfigurational density functional theory. *Phys. Chem. Chem. Phys.* 20, 794–801. doi: 10.1039/C7CP06767D
- Du, Z., Li, H., and Gu, T. (2007). A state of the art review on microbial fuel cells: a promising technology for wastewater treatment and bioenergy. *Biotechnol. Adv.* 25, 464–482. doi: 10.1016/j.biotechadv.2007.05.004
- Eichkorn, K., Treutler, O., Öhm, H., Häser, M., and Ahrlrichs, R. (1995). Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.* 240, 283–290. doi: 10.1016/0009-2614(95)00621-A
- Eichkorn, K., Weigend, F., Treutler, O., and Ahrlrichs, R. (1997). Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials. *Theor. Chem. Acc.* 97, 119–124. doi: 10.1007/s002140050244
- Escoria, A. M., Sen, K., Daza, M. C., Doerr, M., and Thiel, W. (2017). Quantum mechanics/molecular mechanics insights into the enantioselectivity of the O-Acetylation of (R,S)-Propranolol catalyzed by candida antarctica lipase B. *ACS Catal.* 7, 115–127. doi: 10.1021/acscatal.6b02310
- Evans, R. M., Brooke, E. J., Wehlin, S. A. M., Nomerotskaia, E., Sargent, F., Carr, S. B., et al. (2016). Mechanism of hydrogen activation by [NiFe] hydrogenases. *Nat. Chem. Biol.* 12, 46–50. doi: 10.1038/nchembio.1976
- Ghosh, N., and Cui, Q. (2008). pKa of residue 66 in Staphylococal nuclease. I. Insights from QM/MM simulations with conventional sampling. *J. Phys. Chem. B* 112, 8387–8397. doi: 10.1021/jp800168z
- Grimme, S. (2006). Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* 27, 1787–1799. doi: 10.1002/jcc.20495
- Harms, M. J., Schlessman, J. L., Sue, G. R., and Bertrand García-Moreno, E., (2011). Arginine residues at internal positions in a protein are always charged. *Proc. Natl. Acad. Sci. U.S.A.* 108, 18954–18959. doi: 10.1073/pnas.1104808108
- Harris, T. K., and Turner, G. J. (2002). Structural basis of perturbed pKa values of catalytic groups in enzyme active sites. *IUBMB Life* 53, 85–98. doi: 10.1080/15216540211468
- Holladay, J. D., Hu, J., King, D. L., and Wang, Y. (2009). An overview of hydrogen production technologies. *Catal. Today* 139, 244–260. doi: 10.1016/j.cattod.2008.08.039
- Hu, L., Söderhjelm, P., and Ryde, U. (2013). Accurate reaction energies in proteins obtained by combining QM/MM and large QM calculations. *J. Chem. Theory Comput.* 9, 640–649. doi: 10.1021/ct3005003
- Hunter, A., and Borsook, H. (1924). The dissociation constants of arginine. *Biochem. J.* 18, 883–890. doi: 10.1042/bj0180883
- Jensen, K. P. (2008). Bioinorganic chemistry modeled with the TPSSh density functional. *Inorg. Chem.* 47, 10357–10365. doi: 10.1021/ic800841t

## FUNDING

Financial support by the Max Planck Society for the Advancement of Science and the EU COST Action CM1305 ECOSTBio is gratefully acknowledged.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2018.00164/full#supplementary-material>

- Jones, A. K., Sillery, E., Albracht, S. P. J., and Armstrong, F. A. (2002). Direct comparison of the electrocatalytic oxidation of hydrogen by an enzyme and a platinum catalyst. *Chem. Commun.* 866–867. doi: 10.1039/b201337a
- Kästner, J., Carr, J. M., Keal, T. W., Thiel, W., Wander, A., and Sherwood, P. (2009). DL-FIND: an open-source geometry optimizer for atomistic simulations. *J. Phys. Chem. A* 113, 11856–11865. doi: 10.1021/jp9028968
- Kelly, C. P., Cramer, C. J., and Truhlar, D. G. (2007). Single-ion solvation free energies and the normal hydrogen electrode potential in methanol, acetonitrile, and dimethyl sulfoxide. *J. Phys. Chem. B* 111, 408–422. doi: 10.1021/jp065403l
- Li, H., Robertson, A. D., and Jensen, J. H. (2005). Very fast empirical prediction and rationalization of protein pKa values. *Protein. Struct. Funct. Bioinformatics* 61, 704–721. doi: 10.1002/prot.20660
- Lill, S. O. N., and Siegbahn, P. E. M. (2009). An autocatalytic mechanism for NiFe-Hydrogenase: reduction to Ni(I) followed by oxidative addition. *Biochem. Mosc.* 48, 1056–1066. doi: 10.1021/bi081218n
- MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102, 3586–3616. doi: 10.1021/jp973084f
- Mackerell, A. D., Feig, M., and Brooks, C. L. (2004). Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25, 1400–1415. doi: 10.1002/jcc.20065
- Metz, S., Kästner, J., Sokol, A. A., Keal, T. W., and Sherwood, P. (2014). ChemShell—a modular software package for QM/MM simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 4, 101–110. doi: 10.1002/wcms.1163
- Mulliken, R. S. (1955a). Electronic Population Analysis on LCAO-MO molecular wave functions. I. *J. Chem. Phys.* 23, 1833–1840.
- Mulliken, R. S. (1955b). Electronic population analysis on LCAO-MO molecular wave functions. II. overlap populations, bond orders, and covalent bond energies. *J. Chem. Phys.* 23, 1841–1846.
- Niu, S., Thomson, L. M., and Hall, M. B. (1999). Theoretical characterization of the reaction intermediates in a model of the nickel–iron hydrogenase of *Desulfovibrio gigas*. *J. Am. Chem. Soc.* 121, 4000–4007. doi: 10.1021/ja983469r
- Ogata, H., Krämer, T., Wang, H., Schilter, D., Pelmenschikov, V., and van Gastel, M., et al. (2015a). Hydride bridge in [NiFe]-hydrogenase observed by nuclear resonance vibrational spectroscopy. *Nat. Commun.* 6:7890. doi: 10.1038/ncomms8890
- Ogata, H., Lubitz, W., and Higuchi, Y. (2016). Structure and function of [NiFe] hydrogenases. *J. Biochem. (Tokyo)* 160, 251–258. doi: 10.1093/jb/mvw048
- Ogata, H., Mizoguchi, Y., Mizuno, N., Miki, K., Adachi, S., Yasuoka, N., et al. (2002). Structural studies of the carbon monoxide complex of [NiFe]hydrogenase from *Desulfovibrio vulgaris* Miyazaki F: Suggestion for the initial activation site for dihydrogen. *J. Am. Chem. Soc.* 124, 11628–11635. doi: 10.1021/ja012645k
- Ogata, H., Nishikawa, K., and Lubitz, W. (2015b). Hydrogens detected by subatomic resolution protein crystallography in a [NiFe] hydrogenase. *Nature* 520, 571–574. doi: 10.1038/nature14110
- Pardo, A., Lacey, A. L. D., Fernández, V. M., Fan, H.-J., Fan, Y., and Hall, M. B. (2006). Density functional study of the catalytic cycle of nickel–iron [NiFe] hydrogenases and the involvement of high-spin nickel(II). *J. Biol. Inorg. Chem.* 11, 286–306. doi: 10.1007/s00775-005-0076-3
- Perdew, J. P. (1986a). Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* 33, 8822–8824.
- Perdew, J. P. (1986b). Erratum: density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* 34, 7406–7406.
- Raugei, S., Chen, S., Ho, M.-H., Ginovska-Pangovska, B., Rousseau, J., Dupuis, M., et al. (2012). The role of pendant amines in the breaking and forming of molecular hydrogen catalyzed by Nickel complexes. *Chem. Eur. J.* 18, 6493–6506. doi: 10.1002/chem.201103346
- Rebollar-Zepeda, A. M., and Galano, A. (2012). First principles calculations of pKa values of amines in aqueous solution: application to neurotransmitters. *Int. J. Quantum Chem.* 112, 3449–3460. doi: 10.1002/qua.24048
- Rodionova, M. V., Poudyal, R. S., Tiwari, I., Voloshin, R. A., Zharmukhamedov, S. K., Nam, H. G., et al. (2017). Biofuel production: challenges and opportunities. *Int. J. Hydrog. Energy* 42, 8450–8461. doi: 10.1016/j.ijhydene.2016.11.125
- Santoro, C., Arbizzani, C., Erable, B., and Ieropoulos, I. (2017). Microbial fuel cells: from fundamentals to applications. A review. *J. Power Sources* 356, 225–244. doi: 10.1016/j.jpowsour.2017.03.109
- Senn, H. M., and Thiel, W. (2009). QM/MM methods for biomolecular systems. *Angew. Chem. Int. Edn.* 48, 1198–1229. doi: 10.1002/anie.200802019
- Sherwood, P., de Vries, A. H., Guest, M. F., Schreckenbach, G., Catlow, C. R. A., French, S. A., et al. (2003). QUASI: a general purpose implementation of the QM/MM approach and its application to problems in catalysis. *J. Mol. Struct. Theochem* 632, 1–28. doi: 10.1016/S0166-1280(03)00285-9
- Sherwood, P., Vries, A. H., de, Collins, S. J., Greatbanks, S. P., Burton, N. A., Vincent, M. A., et al. (1997). Computer simulation of zeolite structure and reactivity using embedded cluster methods. *Faraday Discuss.* 106, 79–92. doi: 10.1039/a701790a
- Slater, J. C. (1951). A simplification of the hartree-fock method. *Phys. Rev.* 81, 385–390. doi: 10.1103/PhysRev.81.385
- Smith, W., and Forester, T. R. (1996). DL\_POLY\_2.0: a general-purpose parallel molecular dynamics simulation package. *J. Mol. Graph.* 14, 136–141. doi: 10.1016/S0263-7855(96)00043-4
- Stein, M., and Lubitz, W. (2004). Relativistic DFT calculation of the reaction cycle intermediates of [NiFe] hydrogenase: a contribution to understanding the enzymatic mechanism. *J. Inorg. Biochem.* 98, 862–877. doi: 10.1016/j.jinorgbio.2004.03.002
- Stein, M., van Lenthe, E., Baerends, E. J., and Lubitz, W. (2001). Relativistic DFT calculations of the paramagnetic intermediates of [NiFe] hydrogenase. Implications for the enzymatic mechanism. *J. Am. Chem. Soc.* 123, 5839–5840. doi: 10.1021/ja005808y
- Stephan, D. W., and Erker, G. (2010). Frustrated lewis pairs: metal-free hydrogen activation and more. *Angew. Chem. Int. Ed.* 49, 46–76. doi: 10.1002/anie.200903708
- Tao, J., Perdew, J. P., Staroverov, V. N., and Scuseria, G. E. (2003). Climbing the density functional ladder: nonempirical meta-generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.* 91, 146401. doi: 10.1103/PhysRevLett.91.146401
- Tawa, G. J., Topol, I. A., Burt, S. K., Caldwell, R. A., and Rashin, A. A. (1998). Calculation of the aqueous solvation free energy of the proton. *J. Chem. Phys.* 109, 4852–4863. doi: 10.1063/1.477096
- Tissandier, M. D., Cowen, K. A., Feng, W. Y., Gundlach, E., Cohen, M. H., Earhart, A. D., et al. (1998). The proton's absolute aqueous enthalpy and gibbs free energy of solvation from cluster-ion solvation data. *J. Phys. Chem. A* 102, 7787–7794.
- Topol, I. A., Tawa, G. J., Caldwell, R. A., Eissenstat, M. A., and Burt, S. K. (2000). Acidity of organic molecules in the gas phase and in aqueous solvent. *J. Phys. Chem. A* 104, 9619–9624. doi: 10.1021/jp001938h
- Uddin, N., Choi, T. H., and Choi, C. H. (2013). Direct absolute pKa predictions and proton transfer mechanisms of small molecules in aqueous solution by QM/MM-MD. *J. Phys. Chem. B* 117, 6269–6275. doi: 10.1021/jp400180x
- Vosko, S. H., Wilk, L., and Nusair, M. (1980). Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* 58, 1200–1211. doi: 10.1139/p80-159
- Weigend, F., and Ahlrichs, R. (2005). Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy. *Phys. Chem. Chem. Phys.* 7, 3297–3305. doi: 10.1039/b508541a
- Yoo, J., and Cui, Q. (2008). Does arginine remain protonated in the lipid membrane? Insights from microscopic pKa calculations. *Biophys. J.* 94, L61–L63. doi: 10.1529/biophysj.107.122945
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2018 Escoria and Stein. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read for greatest visibility and readership



## FAST PUBLICATION

Around 90 days from submission to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative, and constructive peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers acknowledged by name on published articles

### Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

Visit us: [www.frontiersin.org](http://www.frontiersin.org)

Contact us: [info@frontiersin.org](mailto:info@frontiersin.org) | +41 21 510 17 00



## REPRODUCIBILITY OF RESEARCH

Support open data and methods to enhance research reproducibility



## DIGITAL PUBLISHING

Articles designed for optimal readership across devices



FOLLOW US  
[@frontiersin](https://twitter.com/frontiersin)



## IMPACT METRICS

Advanced article metrics track visibility across digital media



## EXTENSIVE PROMOTION

Marketing and promotion of impactful research



## LOOP RESEARCH NETWORK

Our network increases your article's readership