



Cite this: *Phys. Chem. Chem. Phys.*,
2022, 24, 9051

Multiscale molecular modelling: from electronic structure to dynamics of nanosystems and beyond

Dennis R. Salahub

Important contemporary biological and materials problems often depend on interactions that span orders of magnitude differences in spatial and temporal dimensions. This Tutorial Review attempts to provide an introduction to such fascinating problems through a series of case studies, aimed at beginning researchers, graduate students, postdocs and more senior colleagues who are changing direction to focus on multiscale aspects of their research. The choice of specific examples is highly personal, with examples either chosen from our own work or outstanding multiscale efforts from the literature. I start with various embedding schemes, as exemplified by polarizable continuum models, 3-D RISM, molecular DFT and frozen-density embedding. Next, QM/MM (quantum mechanical/molecular mechanical) techniques are the workhorse of pm-to-nm/ps-to-ns simulations; examples are drawn from enzymes and from nanocatalysis for oil-sands upgrading. Using polarizable force-fields in the QM/MM framework represents a burgeoning subfield; with examples from ion channels and electron dynamics in molecules subject to strong external fields, probing the atto-second dynamics of the electrons with RT-TDDFT (real-time – time-dependent density functional theory) eventually coupled with nuclear motion through the Ehrenfest approximation. This is followed by a section on coarse graining, bridging dimensions from atoms to cells. The penultimate chapter gives a quick overview of multiscale approaches that extend into the meso- and macro-scales, building on atomistic and coarse-grained techniques to enter the world of materials engineering, on the one hand, and cell biology, on the other. A final chapter gives just a glimpse of the burgeoning impact of machine learning on the structure-dynamics front. I aim to capture the excitement of contemporary leading-edge breakthroughs in the description of physico-chemical systems and processes in complex environments, with only enough historical content to provide context and aid the next generation of methodological development. While I aim also for a clear description of the essence of methodological breakthroughs, equations are kept to a minimum and detailed formalism and implementation details are left to the references. My approach is very selective (case studies) rather than exhaustive. I think that these case studies should provide fodder to build as complete a reference tree on multiscale modelling as the reader may wish, through forward and backward citation analysis. I hope that my choices of cases will excite interest in newcomers and help to fuel the growth of multiscale modelling in general.

Received 27th December 2021,
Accepted 29th March 2022

DOI: 10.1039/d1cp05928a

rsc.li/pccp

1 Introduction

“What are the electrons really doing in molecules?” – Robert S. Mulliken, 1960.¹

When I was a graduate student in the ante-diluvian era, Robert Mulliken’s famous question was an inspiration, not to mention the many answers that he found through molecular-orbital theory and spectroscopy. This question still looms large,

though nowadays it is often expanded to include molecules in very complex environments, the topic of this Tutorial Review.

In what follows I will start with a fairly tight focus on electronic structure, bringing in only a few historical methods and case studies to provide the basis of contemporary developments; my main goal involves projecting the contemporary scene onto the future. Isolated molecules will play only supporting roles, though they are crucial for benchmarking methods for molecules in environments. The applications will quickly gain complexity, starting with implicit models for embedding in surrounding media, turning to explicit models for solvation and other environmental features (proteins, catalytic supports, etc.), in the QM/MM (quantum mechanical/molecular mechanical) class, often included in a molecular dynamics (MD)

Department of Chemistry, Department of Physics and Astronomy, CMS-Centre for Molecular Simulation, IQST-Institute for Quantum Science and Technology, Quantum Alberta, University of Calgary, Calgary, Alberta, T2N 1N4, Canada.
E-mail: dsalahub@ucalgary.ca

framework (QM/MM/MM). The workhorses on the MD reins of the QM/MM team use fixed-charge molecular-mechanical force-fields. I devote a section to techniques that go beyond the fixed-charge approximation to include the mutual effects of polarization of the QM and MM regions. If one aims to go beyond the nm–ns scales, eventually features of finer scales have to be integrated out. In MM force-fields, for example, the electrons themselves disappear (a sad day for card-carrying quantum chemists...) and in a section on multi-scale coarse-grained force-fields even the atoms are integrated out, in favour of “beads”. Then, at the largest scales I have chosen to discuss, electrons, atoms and even beads disappear in multiscale approaches that extend into the meso- and macro-scales, entering the world of materials engineering, on the one hand, and cell biology, on the other. Finally, the machine-learning (ML) revolution has not left multiscale molecular modelling unscathed. I give some entries to the literature; I think ML will have significant impact in the near future.

This review attempts to weave a coherent tapestry of multi-scale methods and applications following two broad cross-linked lines of enquiry (cross-linked through the common methods employed) namely, nanocatalysis and biological processes. I do this through a series of case studies, aimed at beginning researchers, graduate students, postdocs and more senior colleagues who are changing direction to focus on multiscale aspects of their research. The choice of specific examples is highly personal, with examples either chosen from our own work or outstanding multiscale efforts from the literature. I apologize to the authors of the multitude of equally, or even more, deserving cases that I could have chosen.

I aim to capture the excitement of contemporary leading-edge breakthroughs in the description of physico-chemical systems and processes in complex environments, with only enough historical content to provide context and aid the next generation of methodological development. While I aim also for a clear description of the essence of methodological breakthroughs, equations are kept to a minimum and formalism and implementation details are left to the references.



Dennis R. Salahub

Dennis Salahub, FRSC, FAAAS, Professor Emeritus of Chemistry at the University of Calgary, is also Vice-President (Research and International) Emeritus. From 1999 to 2002, he served as the Director General of the Steacie Institute for Molecular Sciences at the National Research Council of Canada in Ottawa. From 1979 to 1999, he was at the Université de Montréal. A native of Alberta, Dr Salahub's interests cover a broad range in theoretical and computational chemistry, especially Density Functional Theory and its applications. His current focus is on multi-scale modeling.

Published on 01 April 2022. Downloaded by Université Grenoble Alpes INP on 6/3/2025 9:42:06 AM.

My approach is very selective (case studies) rather than exhaustive. I think that these case studies should provide fodder to build as complete a reference tree on multiscale modelling as the reader may wish, through forward and backward citation analysis. I hope that my choices of cases will excite interest in newcomers and help to fuel the growth of multiscale modelling in general.

2 Embedding schemes

Let's start with an isolated molecule, very often treated within the (Hohenberg) Kohn–Sham density functional theory (KS-DFT).^{2,3}

The Hohenberg–Kohn (HK) theorems of 1964, showed that the ground state of a many-electron system is uniquely and variationally determined by the electron density, which is of trivial complexity, a function of three spatial variables, compared with the many-electron wave function that depends, in a highly correlated fashion, on all $3N$ electronic coordinates (plus spin). A year later, Kohn and Sham turned the HK theorems into something really useful by introducing the concept of a reference system of non-interacting electrons, constrained to have the same density as the real interacting system. The result was a set of mean-field Kohn–Sham equations that could be solved self-consistently, like the Hartree–Fock equations but, in principle, including the effects of electronic correlation. Some concerns about degenerate ground states and v-representability (whether all densities that are explored in the variational process actually correspond to physically realizable potentials) were cleared up by Levy's constrained variational approach,⁴ which should be required reading for the serious student of DFT.

The Kohn–Sham equations are:

$$\hat{h}_{\text{KS}}\phi_i(1) = \left(-\frac{1}{2}\nabla^2 + v_{\text{nuc}}(1) + v_{\text{Coul}}(1) + v_{\text{xc}}(1) \right)\phi_i(1) = \varepsilon_i\phi_i(1) \quad (1.1)$$

where the first operator in the parentheses is the kinetic energy of an electron, the second is the potential of attraction of the electron for all the nuclei in the molecule, the third is the average (mean-field) coulomb repulsion due to all the other electrons in the molecule and the fourth is the exchange–correlation potential, which contains all of the effects of exchange and correlation plus the difference between the kinetic energy of the Kohn–Sham non-interacting reference system and that of the real interacting system. The exchange–correlation potential is the functional derivative of the exchange correlation energy:

$$v_{\text{xc}} = \frac{\delta E_{\text{xc}}}{\delta \rho} \quad (1.2)$$

When a molecule is put into an environment one could, in principle, just expand the KS Hamiltonian to include the environment, but this soon meets the roadblock of computer time. Much of this review concerns various approximate additions to the KS Hamiltonian that attempt to represent the field created by the surroundings. I start, in the next six sections, with several approximations of this type.

2.1. Watson sphere

The idea of replacing an environment by an embedding potential goes back at least half a century, to the 1958 work of Watson.⁵ Using the “Whirlwind digital computer at MIT” Watson studied the O²⁻ ion with Hartree–Fock (HF) theory using an LCAO-MO (linear combination of atomic orbitals – molecular orbitals) representation for the HF orbitals. Since the outer orbitals of the doubly charged anion are unbound, in the continuum, and Watson was interested in solid-state oxides, he used a uniformly charged sphere of charge +2 to stabilize the anion and allow the self-consistent-field (SCF) iterations to converge. The relevant potential for the charged sphere is:

$$v_W(r) = -\frac{Q_W}{R_W} \quad \text{for } r \leq R_W = -\frac{Q_W}{r} \quad \text{for } r > R_W \quad (1.3)$$

where Q_W is the charge and R_W is the radius of the Watson sphere.

Case study 1: Madelung energy and watson sphere model for Li₃N

As the first case study I, rather arbitrarily, choose a 1979 study of the nitride anion, N³⁻, in Li₃N.⁶ Schwarz and Herzog showed how to optimize the radius of the Watson sphere using electrostatics, assuming +1 charges for the Li ions and requiring the electrostatic interaction energy between the N³⁻ ion and the remainder of the lattice (related to the Madelung potential) to be equal to the electrostatic interaction energy of the N³⁻ ion with the potential of the Watson sphere. The analysis yields $R_W = 1.39$ Å, in perfect agreement with a previous value obtained by comparing calculated and experimental X-Ray scattering factors.

Fig. 1 shows that the sum $v_C(r) + v_W(r)$ tends to zero at large r , the Watson potential just cancelling the Coulomb potential. Note the discontinuity in the slope of the Watson potential at the Watson radius, which is reminiscent of even more serious discontinuities of the infamous muffin-tin (MT) potential used in the Xα-Scattered Wave (Xα-SW) method⁷ that was widely used in the 1960s and 1970s and presaged the utility of DFT in chemistry. The MT potential used spherically symmetric potentials around the atoms in a molecule or solid, out to a certain radius, and a constant potential between the spheres. We used Xα-SW for many years to explore, e.g., transition-metal clusters (see, e.g. ref. 8).

2.2. Point-charge embedding

Consider now a somewhat more detailed, structured, representation of the environment that has been often used for embedded-cluster models of ionic solids, in which surrounding ions are represented only by their electrostatic Coulomb potential. Staying within DFT, an extra potential has to be added to the Kohn–Sham Hamiltonian:

$$v_{ee\text{-embedding}}(\vec{r}) = \sum_i \frac{q_i}{|\vec{R}_i - \vec{r}|} \quad (1.4)$$

where the sum is over all of the embedding charges, q_i at position \vec{R}_i . In fact this is a special case of the more general QM/MM methods treated later, in which only electrostatic interactions are retained.

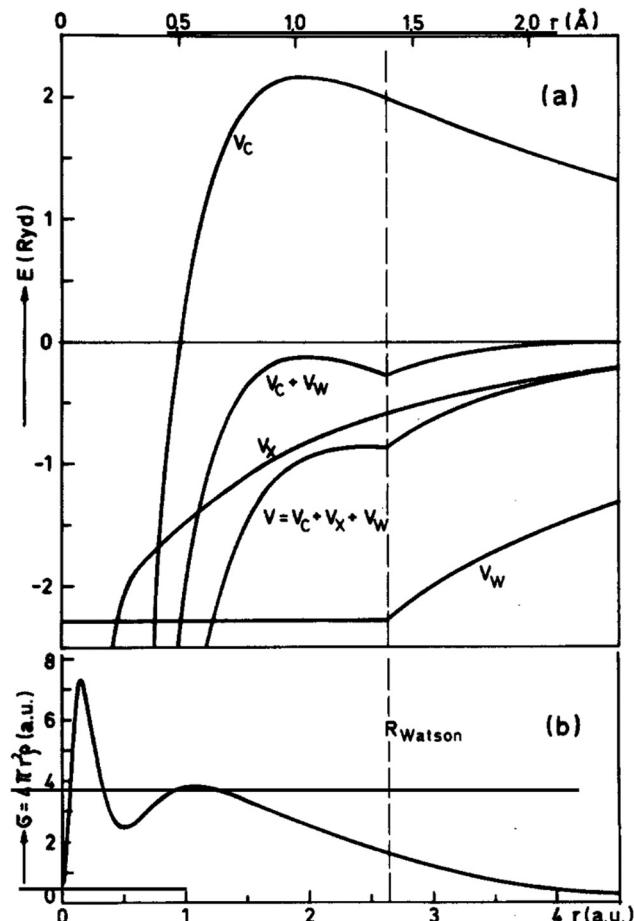


Fig. 1 The Watson sphere model for the N³⁻ ion; (a) Coulomb potential V_C , exchange potential V_x , and Watson potential V_w . (b) the radial charge density $\sigma = 4\pi r^2 \rho$. Reproduced from ref. 6 with permission from Elsevier, copyright 1979.

Case study 2: Relations between the structure and adsorption properties of zeolites

This case emphasizes that, for ionic solids like zeolites, even a very small quantum mechanical (DFT) system can provide keen insight into structure–property relationships if it is embedded in an appropriate array of point charges representing the environment and if appropriate physical interactions are accounted for (electrostatics plus induction) in the analysis.

Goursot and co-workers⁹ highlighted the strong relationship between the N₂ adsorption properties of zeolites and the type and site of accessible cations, Na⁺. One such site, the so-called “ideal” site, type III, is shown in Fig. 2a) (see ref. 9 for the definition of other sites). For example, 64 different cation sites were considered for types I and II, with 32 cations distributed among 64 type III sites. Four different framework structures were considered. For each site, a quantum-mechanical Na⁺-N₂ system was embedded in its own network of point charges arising from a full shell of 26 first-neighbor shells. Atomic charges of -1.2 for O, 1.4 for Al, 2.4 for Si and 1.0 for Na were employed, based on previous cluster calculations. Adsorption energies were calculated for the various cation sites and the

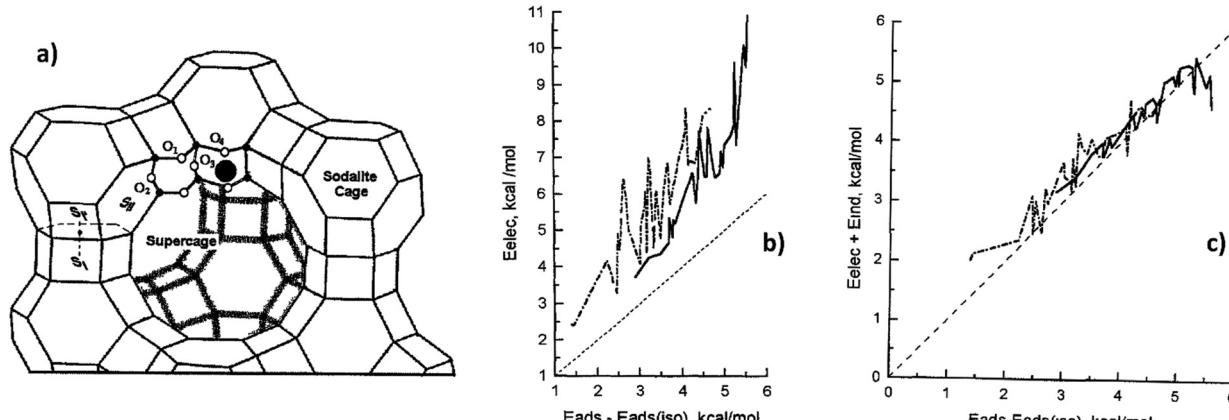


Fig. 2 (a) Faujasite type structure showing the sodalite cage with one of the cationic sites represented by the black dot. (b) Electrostatic contributions to the N₂-zeolite interaction energies compared with the calculated adsorption energies. (c) Electrostatic plus induction contributions to the N₂-zeolite interaction energies compared with the calculated adsorption energies. Reproduced from ref. 9 with permission from Elsevier, copyright 1998.

general tendencies, *e.g.*, sites II *vs.* sites III, were elucidated, showing a clear correlation between structural characteristics and adsorption strength. Delving deeper, the authors defined:

$$E_{\text{ads}} = \Delta E(\text{Na}^+ \text{N}_2) + E_{\text{int}} \quad (1.5)$$

where $\Delta E(\text{Na}^+ \text{N}_2)$ is the binding energy of N₂ to an isolated cation and E_{int} corresponds to the modification due to the external charges. They then went on to calculate the appropriate electric fields for an electrostatic analysis. Typical results in Fig. 2b) show that the straight electrostatic embedding yields a good correlation with the full calculation but induction terms (Fig. 2c)) are necessary for a more quantitative agreement.

Of course, over the intervening two decades, methodologies and computer power have improved so that “full” QM/MM calculations with force-fields that include van der Waals interactions, for example, are now routinely carried out for zeolites (*e.g.* the study of N₂O reduction by CO in Fe-BEA zeolite¹⁰). Indeed, it is now possible to include enough atoms in a DFT calculation to do meaningful cluster calculations without the need for an embedding potential (*e.g.* the study of Na-MOR and H-MOR zeolites with clusters of more than 400 atoms^{11,12}).

2.3. Polarizable continuum models (PCM)

Let us turn now to a family of models in which the atomistic structure of the surrounding environment is lost, being replaced by a polarizable continuum. Such models date back at least to the seminal 1936 work of Onsager¹³ in which a dipolar molecule was surrounded by a spherical cavity outside of which was an initially unstructured continuum, having a dielectric constant ϵ . The dipole is modified by the polarization of the dielectric that it induces, giving rise to the so-called reaction field. The paper includes an estimate of the increase in the dipole moment of a water molecule in the liquid phase relative to the gas by a factor of approximately 3/2, remarkably close to the results of contemporary simulations.

For more complex solutes a spherical cavity is clearly not ideal; an important landmark is represented by the 1981 paper of Miertus, Scrocco and Tomasi¹⁴ who worked out the details of

the reaction field for a general molecule represented by overlapping atomic spheres, surrounded by a solvent-excluded surface (SES) (the “Connolly surface”,^{15–17}) constructed by rolling a probe sphere representing a solvent molecule over the surface formed from the union of the atomic spheres (Fig. 3). An excellent presentation of the details may be found in the extensive review of continuum solvation models by Tomasi, Mennucci and Cammi.¹⁸

Following,¹⁸ if we assume for the moment that the solute density $\rho_M = \rho_M^e + \rho_M^n$ (superscripts e and n refer to electrons and nuclei) is wholly contained within the SES the electrostatics have to take account of the polarization of the continuum by ρ_M which in turn modifies ρ_M so that a self-consistent iterative process is necessary to eventually define the solvent reaction (embedding) potential which has to be added to the quantum-mechanical Hamiltonian of the solute. The electrostatic problem involves solution of the Poisson equation:

$$\begin{aligned} -\nabla^2 v(\vec{r}) &= 4\pi\rho_M(\vec{r}) && \text{within the cavity} \\ -\epsilon\nabla^2 v(\vec{r}) &= 0 && \text{outside the cavity} \end{aligned} \quad (1.6)$$

v is the sum of the electrostatic potential generated by the charge distribution and the reaction potential generated by the polarization of the continuum:

$$v(\vec{r}) = v_M(\vec{r}) + v_R(\vec{r}) \quad (1.7)$$

In the most common version of the method (ASC – apparent surface charge) an auxiliary charge $\sigma(\vec{s})$ spread over the cavity surface is determined by the electrostatics and boundary conditions.

$$v_\sigma(\vec{r}) = \int_{\Gamma} \frac{\sigma(\vec{s})}{|\vec{r} - \vec{s}|} d^2s \quad (1.8)$$

So, the entire continuum is replaced by the auxiliary charge on the cavity surface Γ which is typically approximated in terms of a set of finite elements (tesserae) of area A_k small enough to consider the charge constant in each of them. Thus, the integral

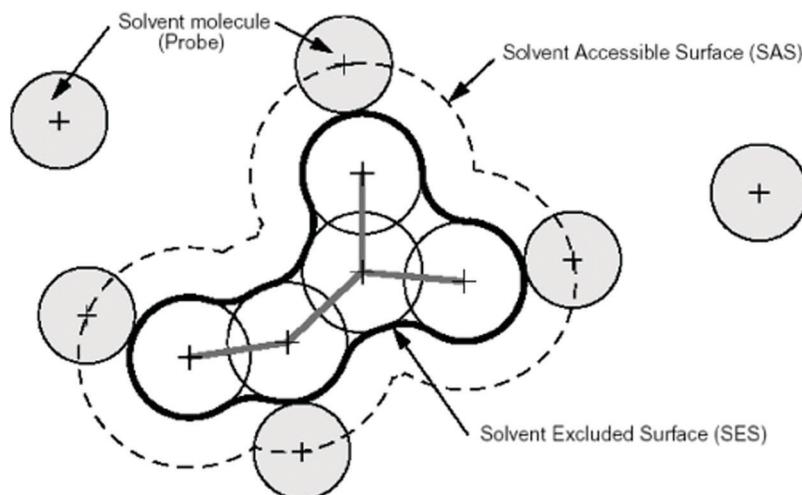


Fig. 3 Solvent-accessible surface (SAS) traced out by the center of the probe representing a solvent molecule. The solvent-excluded surface (SES) is the topological boundary of the union of all possible probes that do not overlap with the molecule. Reprinted with permission from ref. 18. Copyright 2005 American Chemical Society.

is converted to a sum:

$$v_\sigma(\vec{r}) \approx \sum_k \frac{\sigma(\vec{s}_k) A_k}{|\vec{r} - \vec{s}_k|} = \sum_k \frac{q_k}{|\vec{r} - \vec{s}_k|} \quad (1.9)$$

essentially reducing the potential to a special case of point-charge embedding.

The conductor-like screening model (COSMO)¹⁹ is a popular variant of PCM in which the dielectric constant is set to ∞ , corresponding to a conductor, and then the ideal unscreened charge density, σ^* , corresponding to $\epsilon = \infty$, is scaled by a function of ϵ

$$\sigma(s) = f(\epsilon)\sigma^*(s) \quad (1.10)$$

The scaling function, f , is determined empirically by comparing unscaled COSMO energies with correct solute–solvent energies, leading to the formula:

$$f(\epsilon) = \frac{\epsilon - 1}{\epsilon + k} \quad (1.11)$$

where k is small (see ref. 18 for references for the choice of k). An important development, COSMO-RS²⁰ (for “real solvents”) brings in nonlinear effects and is widely used to calculate standard thermodynamic data for fluids, vapor pressures, activity coefficients, excess properties, etc.

A further family of variants of PCM comes from the Minnesota group, SMx²¹ (Solvent Model x , where x is a version number), which fall into the set of generalised-Born²² approximation methods based on discrete atomic charges and the popular SMD²³ (Solvent Model Density) which uses the entire solute density. SMD lies within the IEF (integral equation formalism)-PCM class of methods²⁴ and refs. therein.

Case study 3: Insight into an enzymatic mechanism via a cluster/SMD model

In this case study, Sheng and Himo²⁵ use a ~300 atom cluster model with B3LYP-D3(BJ) and the SMD solvation model to examine a recent proposal for the mechanism of action of

3-methylglutaconyl CoA decarboxylase, an important step in the biosynthesis of isovaleryl-coenzyme A in myxobacteria. This is just a recent example of many such cluster studies spearheaded by Siegbahn, Fahmi, Blomberg and co-workers.²⁶

Fig. 4 presents two competing mechanisms for a key step in the mechanism. In the initial proposal Fig. 4(a) the reaction proceeds by a two-step direct decarboxylation involving an enolate intermediate. However, mutations of the cysteine residue did not reduce the activity of the enzyme, indicating that perhaps another mechanism was operative, the intramolecular decarboxylation shown in Fig. 4(b), which does not involve directly any residue of the enzyme. Such a proposal would make the enzyme unique.

Sheng and Himo mapped out the energy profiles for both mechanisms using a ~300 atom cluster model. Peripheral atoms are held fixed in the optimizations and the whole cluster is embedded in an SMD solvent model with $\epsilon = 4$.

Comparing the energy profiles in Fig. 5 and 6, it was shown that the intramolecular mechanism leads to prohibitively high activation energies whereas the originally proposed mechanism has a surmountable barrier that is consistent with experiment. As for the proton source in mutants that remove the cysteine residues, the authors found, thorough further cluster calculations, that a glutamate residue could serve as an alternate proton source.

A nice piece of detective work, using cluster models for enzymes.

2.4. 3-D RISM Integral-equation theory of liquids

The PCM methods treated in the last section demand a lot of the solute, forcing it, as best it can, to use the dielectric medium to reflect its properties without even a hint of its molecular structure, just a single dielectric constant, ϵ . A more structured solvent can be introduced through the long-established integral-equation theory (IET) of liquids (see e.g. ref. 27–29 for textbooks on the theory of liquids). Here I will focus on the so-called reference interaction site model (RISM);

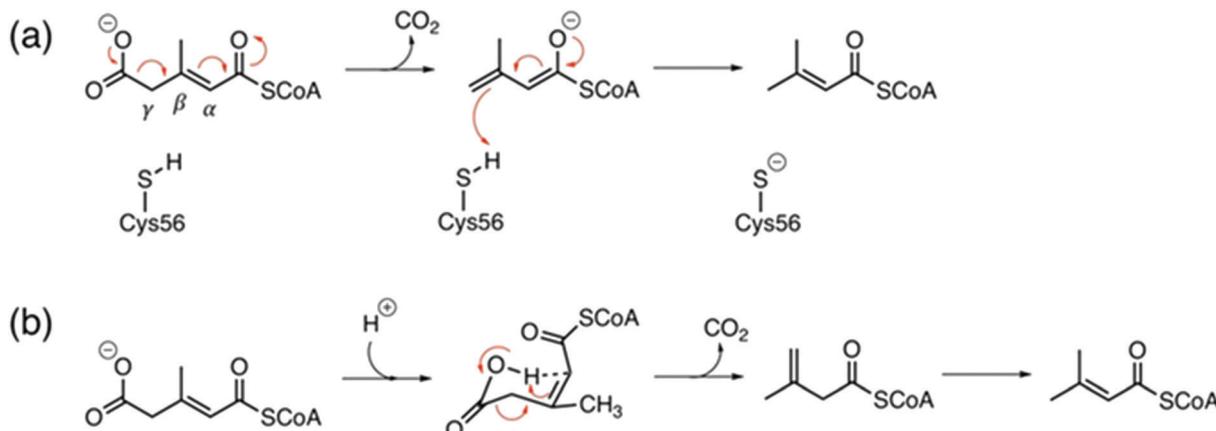


Fig. 4 Proposed mechanisms for AibA/AibB: (a) direct decarboxylation, (b) intramolecular decarboxylation. Reproduced with permission from ref. 25 which is an open access article under the terms of the creative commons license.

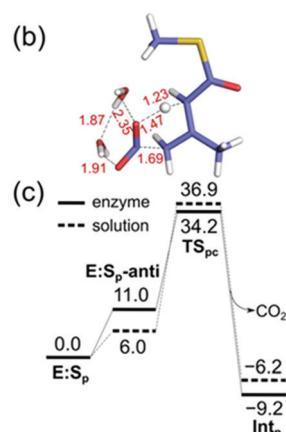


Fig. 5 (b) Optimized structure of the pericyclic transition state in AibA/AibB, in solution, and (c) calculated energy profiles. Reproduced with permission from ref. 25 which is an open access article under the terms of the creative commons license.

for recent reviews and perspectives see ref. 26 and 30. I will only sketch out some of the key concepts and equations, following closely the excellent presentation of Ratkova, Palmer and Fedorov.²⁶ This will take us from the Ornstein–Zernike (OZ) equation³¹ for simple atomic liquids, with an integral equation that has to be solved recursively, requiring a closure relationship to break the chain of equations. Extension to molecular solvents leads to 6-dimensional equations depending on solvent–molecule positions and orientations. Necessary simplification is gained by introducing atomic sites, eventually leading to the powerful 3-D RISM method that is the focus of attention here.

The starting point is the OZ equation that gives a total correlation function between a pair of spherical particles 1 and 2 in a homogeneous liquid of density ρ :

$$h(r_{12}) = c(r_{12}) + \rho \int c(r_{13})h(r_{23})dr_3 \quad (1.12)$$

where c is the direct correlation function between 1 and 2

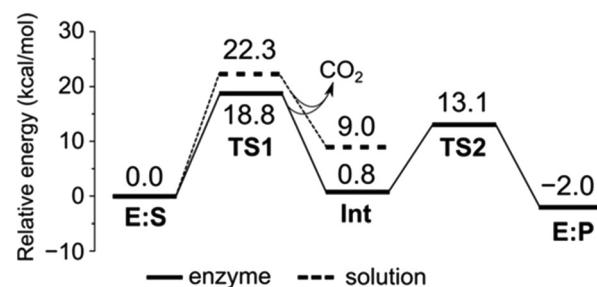


Fig. 6 Calculated energy profiles for the direct decarboxylation mechanism in AibA/AibB and in solution. Reproduced with permission from ref. 25 which is an open access article under the terms of the creative commons license.

modulated by interactions with other particles brought in through the integral. The pair correlation function, $g(r) = h(r) + 1$, gives the probability of finding a particle at a distance r from another particle.

Since h appears on both sides of eqn (12) one can try to solve for it recursively:

$$\begin{aligned} h(r_{12}) &= c(r_{12}) + \rho \int c(r_{13})c(r_{23})dr_3 \\ &\quad + \rho^2 \int \int c(r_{13})c(r_{34})c(r_{42})dr_3dr_4 + \dots \end{aligned} \quad (1.13)$$

The total correlation function, h , is a sum of the direct correlation function, $c(r_{12})$, and an indirect correlation function ($\gamma(r_{12}) = h(r_{12}) - c(r_{12})$) that is propagated through the recursion by an increasing number of intermediate particles, an infinite chain that eventually will have to be broken (closure). Typical forms of total and direct correlation functions are shown in Fig. 7.

The peaks in h correspond to solvation shells while c dies off asymptotically as $C(r) \rightarrow -\beta u(r)$ and $r \rightarrow \infty$ where $u(r)$ is the pair interaction potential and $\beta = 1/k_B T$.

Since the OZ equation contains two unknowns, h and c , another equation is needed, the closure relation that couples h

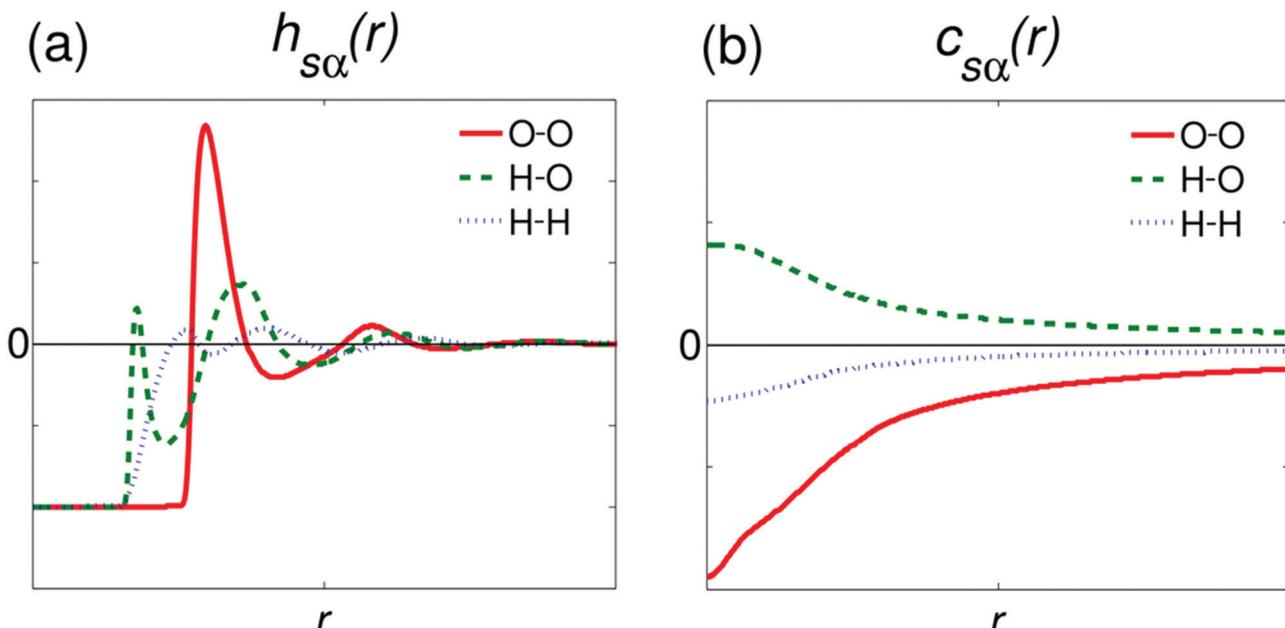


Fig. 7 Typical form of the (a) total and (b) direct site–site correlation functions for different components of a liquid (bulk water). Reprinted with permission from ref. 26 Copyright 2015 American Chemical Society.

and c to the pair potential $u(r)$:

$$h(r) + 1 = \exp(-\beta u(r) + \gamma(r) + B(r)) \quad (1.14)$$

The bridge function B is a functional of the indirect correlation function ($B(r) = B[\gamma(r)]$). Like the exchange–correlation functional in DFT, B is unknown, so approximations are sought.³² For example, neglecting B yields the hypernetted chain (HNC), the mean spherical approximation (MSA) sets $B[\gamma(r)] = \ln(1 + \gamma(r) - \beta u(r)) - \gamma(r) + \beta u(r)$, and combining these two leads to the widely used Kovalenko–Hirata³³ closure:

$$h(r) + 1 = \begin{cases} \exp(-\beta u(r) + \gamma(r)) & \text{if } h(r) \leq 0 \\ 1 - \beta u(r) + \gamma(r) & \text{if } h(r) > 0 \end{cases} \quad (1.15)$$

Until here, consideration has been limited to atomic liquids. As outlined in ref. 29 and 32 the molecular OZ (MOZ) equation is:

$$\begin{aligned} h(r_{12}, \Theta_1, \Theta_2) = & c(r_{12}, \Theta_1, \Theta_2) \\ & + \frac{\rho}{Z} \int c(r_{13}, \Theta_1, \Theta_3) h(r_{32}, \Theta_3, \Theta_2) dr_3 d\Theta_3 \end{aligned} \quad (1.16)$$

where r_{12} is the displacement of the two particles and the Θ represent Euler angles giving the orientation of the molecule. Z is equal to 4π for linear molecules and $8\pi^2$ for nonlinear molecules. Because of the high-dimensionality of the MOZ equation it is difficult to solve, so the workhorses are found in approximations like RISM (see references in ref. 26 for progress with the high-dimension equations, including early work on the statistical-mechanical classical DFT theory, “molecular DFT” to which I will return in the next Section.)

In their seminal work, Chandler and Andersen³⁴ introduced the concept of interaction sites within a molecular component of a liquid, bringing molecular size and shape into the theory

and providing an avenue to reduce the dimensionality of the MOZ equation. Fig. 8 captures the essence of both the 1D and 3D versions of RISM.

In 1D-RISM, the MOZ equation is approximated by a set of one-dimensional equations, assuming spherical symmetry of the interaction sites; hence, the correlation functions only depend on inter-site distances (see ref. 26 for details on solute–solvent and solvent–solvent interactions). The main limitation of 1D-RISM is that it does not properly take into account the spatial correlations of the solvent density around the solute. A more accurate treatment, 3D-RISM, was therefore proposed by Beglov and Roux,^{35,36} reducing the problem to a set of 3D integral equations by partially integrating over the Euler angles. Now N 3-D equations have to be solved, where N is the number of sites in a solvent molecule, leading to the computationally efficient methods used contemporaneously.

In 3D-RISM, the solvent site – solute correlations are the central variables, with total, $h_\alpha(r)$, and direct, $c_\alpha(r)$, correlation functions. The 3D-RISM equation reads:

$$h_\alpha(\mathbf{r}) = \sum_{\xi=1}^N \int_{R^3} c_\xi(\mathbf{r} - \mathbf{r}') \chi_{\xi\alpha}(|\mathbf{r}'|) d\mathbf{r}' \quad \alpha = 1, \dots, N \quad (1.17)$$

where $\chi_{\xi\alpha}(r)$ is the bulk solvent susceptibility function, ξ and α run over the N sites in a solvent molecule. N closure relations are needed:

$$h_\alpha(\mathbf{r}) = \exp(-\beta u_\alpha(\mathbf{r}) + \gamma_\alpha(\mathbf{r}) + B_\alpha(\mathbf{r})) - 1 \quad \alpha = 1, \dots, N \quad (1.18)$$

I will leave the numerical grid-based solution of the 3D-RISM equations to the references in, *e.g.*, ref. 26 and 30 and return to one of the main themes of this review, namely, the potential representing the solvent in a 3D-RISM approach to embedding a quantum-mechanical system, using KS-DFT as the prototype.³⁷

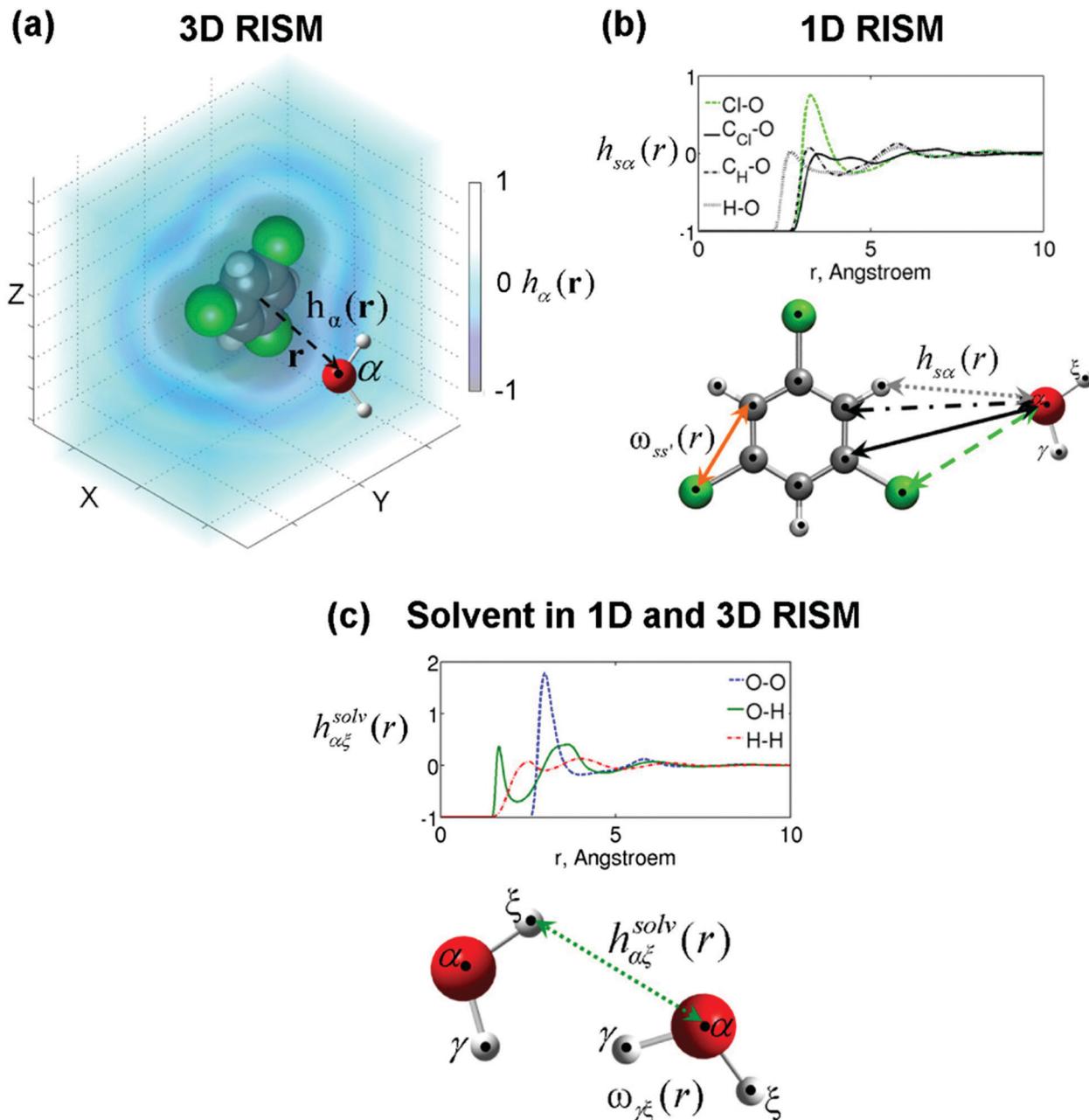


Fig. 8 1D- and 3D-RISM correlation functions. (a) Intermolecular solute–solvent correlation function computed by 3D-RISM for a model solute. (b) Spherically symmetric correlation functions in 1D-RISM. (c) Solvent-solvent correlations in both 1D- and 3D-RISM methods. Reprinted with permission from ref. 26. Copyright 2015 American Chemical Society.

Following³⁷ seek the solvent potential that enters the Kohn-Sham equations:

$$\begin{aligned} \hat{h}_{\text{KS}}\phi_i(1) &= \left(-\frac{1}{2}\nabla_1^2 + v_{\text{nuc}}(1) + v_{\text{Coul}}(1) \right. \\ &\quad \left. + v_{\text{xc}}(1) + v_{\text{solv}}(1)\right)\phi_i(1) = \varepsilon_i\phi_i(1) \end{aligned} \quad (1.19)$$

Start with the Helmholtz free energy of the total system:

$$A[n_e(\mathbf{r}), \{\rho_\gamma(\mathbf{r})\}] = E_{\text{solute}}[n_e(\mathbf{r})] + \Delta\mu_{\text{solv}}[n_e(\mathbf{r}), \{\rho_\gamma(\mathbf{r})\}] \quad (1.20)$$

where E_{solute} is the usual Kohn-Sham energy of the solute and μ_{solv} is the excess chemical potential of solvation, $n_e(\mathbf{r})$ is the electron density and $\rho_\gamma(\mathbf{r})$ is the classical density distribution of an interaction site in a solvent molecule. Minimizing A , subject to normalization constraints, yields the various potentials in eqn (19), including

$$\nu_{\text{solv}}(\mathbf{r}) = \frac{\delta\Delta\mu_{\text{solv}}[n_e(\mathbf{r}), \{\rho_\gamma(\mathbf{r})\}]}{\delta n_e(\mathbf{r})} \quad (1.21)$$

$\Delta\mu$ is sought within the 3D-RISM-KH approximation that

involves the solvent–solute potential through the correlation functions, yielding

$$\nu_{\text{solv}}(\mathbf{r}) = \frac{\delta \Delta \mu_{\text{solv}}}{\delta n_e(\mathbf{r})} = \rho \sum_{\gamma} \int d\mathbf{r}' h_{\gamma}(\mathbf{r}) \nu_{\gamma}^{\text{ps}}(|\mathbf{r} - \mathbf{r}'|) \quad (1.22)$$

where ν_{γ}^{ps} is the contribution of interaction site γ to a pseudo-potential for a solvent molecule acting on the electron density. ν_{γ}^{ps} contains a short-range Lennard-Jones potential plus the long-range interaction of the solvent interaction sites with the solute nuclei and electrons (see ref. 37 for details).

Case study 4: Solvent structure around an ion in an ionic liquid by KS-DFT/3D-RISM-KH

To illustrate the fact that 3D-RISM-KH performs well even for ionic solutes, I choose a study of the electronic structure of the methyl-methylimidazolium ion in bulk [mmim][Cl].³⁸ The solvation structure (Fig. 9) shows Cl^- tightly solvating the cation, with other cations forming a looser structure. There is remarkable agreement with explicit solvent calculations (CPMD). For example, the dipole moment, increases from 2.10D (isolated cation) to 2.59 for both KS-DFT/3D-RISM-KH and CPMD, the coordination number is 7.4 anions around the cation vs. 7.5 from CPMD, and, in both solvent models, adjacent cations stack with the central CR hydrogens of neighbouring sites pointing in opposite directions. Some differences involving the details of hydrogen bonding are noted in ref. 38. The calculated total free energy of solvation for both ions, $-71.2 \text{ kcal mol}^{-1}$ is within the range of 70–100 kcal mol⁻¹ estimated for ionic liquids.

Clearly, in this example, the intricate solvent–solute and solvent–solvent interactions are reasonably well reflected in KS-DFT/3D-RISM-KH. Remarkable, considering the modest computational cost compared with explicit-solvent dynamics.

2.5. Molecular DFT

I turn now to “the other DFT”, the statistical mechanical classical (Molecular DFT – MDFT) theory of solvation coupled

with electronic DFT for the solute.³⁹ The roots of Molecular DFT can be traced back to Mermin's⁴⁰ 1965 finite-temperature extension of the Hohenberg–Kohn theorems (which was still quantum mechanical) as incorporated into the statistical-mechanical theory of classical fluids, admirably exposed, e.g., in Evans' 1979 review.⁴¹ For DFT aficionados I also note that an alternative derivation of the central variational principle was formulated⁴² using Levy's constrained search method.⁴

Here, I will content myself with a sketch of the method and key equations following ref. 39 and a simple proof-of-principle case study.

Electronic DFT describes the solute while the solvent is described with a classical grand-canonical ensemble; in the most recent version both densities are optimized self-consistently, accounting for their mutual polarization. The full density of the solute is used, without resorting to fitting to point charges. Like in the 3D-RISM approach, stat-mech obviates the need for explicit sampling, affording a computational advantage with respect to MD or MC while retaining molecular structure in the solvent, as opposed to continuum solvent models (CSM) such as PCM.

Following ref. 39, in MDFT the solvent molecules are represented by a classical force-field, say, SPC/E for water, with rigid molecules so that \mathbf{r} and Ω describe the solvent coordinates that enter the density $\rho(\mathbf{r}, \Omega)$. Now, from the perspective of the solvent, for any perturbation (e.g. injecting a solute) there exists a unique functional F of the solvent density ρ ⁴¹ which reaches a minimum equal to the solvation free energy (SFE) at the equilibrium density ρ_{eq} . F may be written as

$$F[\rho] = F_{\text{id}}[\rho] + F_{\text{ext}}[\rho] + F_{\text{exc}}[\rho] \quad (1.23)$$

where F_{id} is the entropy of the non-interacting fluid, F_{ext} is the external free-energy, due to the solute, and F_{exc} , the excess functional, is due to solvent–solute interactions. For the latter, e.g., the hypernetted chain closure can be employed. The solute–solvent interaction, F_{ext} , which will eventually yield the perturbation of the

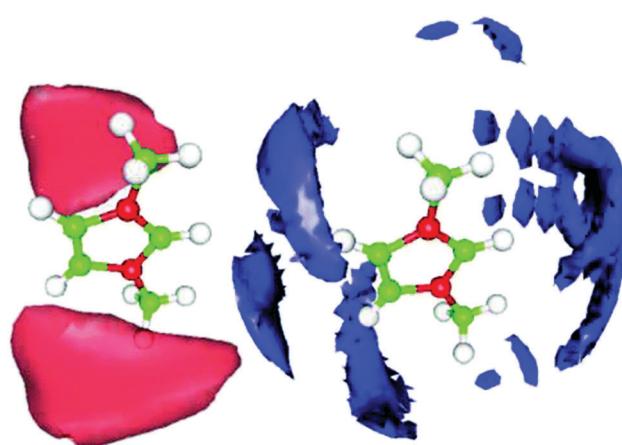
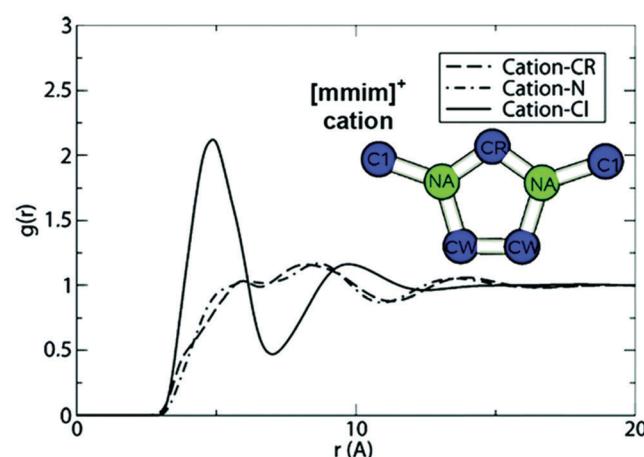


Fig. 9 Solvation structure of the methyl–methyl imidazolium ion in bulk liquid of [mmim][Cl] at $T = 400 \text{ K}$, obtained by the self-consistent field KS/DFT/3D-RISM-KH multiscale theory of electronic structure in solution. Left part: Isosurface of the nitrogen of $[\text{mmim}]^+$ cations at $g_N(r) > 2$ (blue) and of Cl^- anions at $g_C(r) > 5$ (red). Right part: Radial distribution functions of Cl^- anions (solid line), cation CR site (dashed line), and cation Na^+ site (dash-dotted line) around the center of mass of the $[\text{mmim}]^+$ cation in IL. Reproduced from ref. 30 with permission from the PCCP Owner Societies, copyright 2018.



KS potential due to the solvent, is

$$F_{\text{ext}}[\rho] = \iint \rho(\mathbf{r}, \Omega) V_{\text{ext}}(\mathbf{r}, \Omega) d\mathbf{r} d\Omega \quad (1.24)$$

where V_{ext} is the external energy density, which involves electronic DFT, eDFT (for which I write the density as ρ_e in this section). The electrostatic solute–solvent interaction is

$$E_{\text{ES}}[\rho, \rho_e] = \iint \frac{\sigma_V(\mathbf{r}) \sigma_U(\mathbf{r}')}{4\pi\epsilon_0 |\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \quad (1.25)$$

where the solute charge density is

$$\sigma_U(\mathbf{r}) = \sum_i Z_i \delta(\mathbf{r} - \mathbf{r}_i) - \rho_e(\mathbf{r}) \quad (1.26)$$

and the solvent charge density is

$$\sigma_V(\mathbf{r}) = \iint \rho(\mathbf{r}', \Omega) \sigma(\mathbf{r} - \mathbf{r}', \Omega) d\mathbf{r}' d\Omega \quad (1.27)$$

Similarly to QM/MM schemes (see below), short-range repulsion and dispersion interactions are treated with Lennard-Jones potentials on the nuclei of the solute. Finally, the free-energy difference for the solute on going from the gas phase to solution is approximated by the $T = 0$ energy terms.

$$\Delta F_{\text{QM}}[\rho_e] \approx \Delta E_{\text{QM}}[\rho_e] = E_e[\rho_e] - E_e[\rho_e^{\text{vac}}] \quad (1.28)$$

to yield the solvation free energy that can be computed by minimizing the functional

$$F[\rho_e, \rho] = \Delta E_{\text{QM}}[\rho_e] + F_{\text{id}}[\rho] + F_{\text{ext}}[\rho_e, \rho] + F_{\text{exc}}[\rho] \quad (1.29)$$

Rather than the joint minimization, for practical calculations, a step-wise self-consistent protocol was used (see ref. 39 for details).

Case study 5: Benchmark comparison of eDFT/MDFT with PCM

Fig. 10 compares the solvent charge densities calculated with MDFT with those from a continuum solvent model (CSM). The much richer detail of the former is apparent to the eye. See ref. 39 for details and another promising example, an S_N2

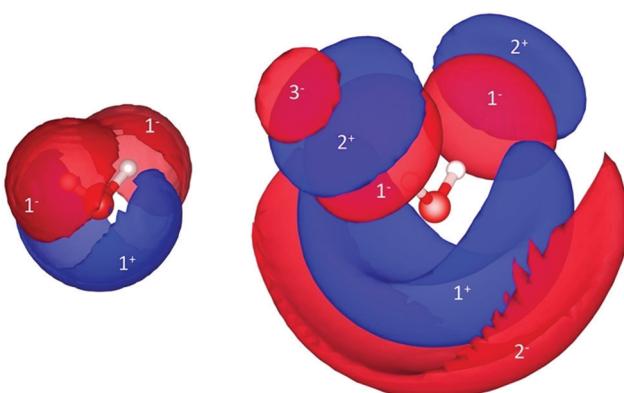


Fig. 10 Isosurfaces of solvent charge densities: positive surfaces are displayed in blue and negative ones in red. The left figure has been obtained with CSM. The right one has been obtained with MDFT. Reprinted with permission from ref. 39. Copyright 2020 American Chemical Society.

reaction between chloromethane and chloride in water. Though the number of applications of MDFT is still quite small, it will be interesting to monitor its evolution, hopefully, including detailed comparisons with 3D-RISM for both accuracy and computational speed.

2.6. Frozen-density embedding

Another way to put structure into the surrounding medium uses (electronic) DFT in a frozen-density embedding theory (FDET).^{43,44} This approach, originally formulated to embed DFT solutes (sometimes known as KSCED – Kohn–Sham Constrained Electron Density^{45–47}) has been generalized and developed over the years, notably by Wesolowski and coworkers, for the embedding of solutes described at a variety of quantum-mechanical levels.^{43,44} Here I will follow the general Hohenberg-Kohn approach of ref. 43 and 44; the case study will involve KS-DFT.

In FDET, one starts with a composite system, AB, with N_{AB} electrons in an external potential v_{AB} for which the density of B, $\rho_B(\vec{r})$, is frozen, yielding a constraint on the total density, $\rho(\vec{r}) = \rho_A(\vec{r}) + \rho_B(\vec{r})$, $\forall \vec{r} \rho(\vec{r}) \geq \rho_B(\vec{r})$. The Hohenberg-Kohn minimization:

$$E_0 = \min_{\int \rho(\vec{r}) d\vec{r} = N_{\text{AB}}} E_{v_{\text{AB}}}^{\text{HK}}[\rho] = E_{v_{\text{AB}}}^{\text{HK}}[\rho_0^{\text{AB}}] \quad (1.30)$$

is now:

$$E^{\text{FDET}}[\rho_B] = \min_{\int \rho(\vec{r}) d\vec{r} = N_{\text{AB}}} E_{v_{\text{AB}}}^{\text{HK}}[\rho] = E_{v_{\text{AB}}}^{\text{HK}}[\rho_{\text{AB}}^{\text{FDET}}] \quad (1.31)$$

$$\forall \vec{r} \rho(\vec{r}) \geq \rho_B(\vec{r})$$

where the second condition cannot, in general, be satisfied without knowing the exact density of the composite system; hence FDET is an approximate formulation, like other embedding schemes. Since ρ_B is frozen, the equivalent form of eqn (31)

$$E^{\text{FDET}}[\rho_B] = \min_{\int \rho_A(\vec{r}) d\vec{r} = N_A} E_{v_{\text{AB}}}^{\text{HK}}[\rho_A + \rho_B] = E_{v_{\text{AB}}}^{\text{HK}}[\rho_A^{\text{opt}} + \rho_B] \quad (1.32)$$

$$\forall \vec{r} \rho_A(\vec{r}) \geq 0$$

brings out better the operational variation of the solute, subject to the constraints imposed by the solvent.

In order to perform the search for the optimal ρ_A one can introduce an auxiliary, embedded, wave function, Ψ_A^{emb} , yielding the variational equation

$$\frac{\delta E_{\text{AB}}^{\text{EWF}}[\Psi_A^{\text{emb}}, \rho_B]}{\delta \Psi_A^{\text{emb}}} - \lambda \Psi_A^{\text{emb}} = 0 \quad (1.33)$$

where the Lagrange multiplier ensures normalization of the embedded wavefunction.

Now, partitioning the external potential into two components, $v_{\text{AB}}(\vec{r}) = v_A(\vec{r}) + v_B(\vec{r})$ defines a system of N_A electrons in its external potential $v_A(\vec{r})$. Introducing Kohn–Sham exchange–correlation and kinetic-energy functionals allows

the FDET total energy functional to be written:

$$\begin{aligned} E_{AB}^{\text{EWF}}[\Psi^A, \rho_B] = & \langle \Psi^A | \hat{H} | \Psi^A \rangle + \Delta F^{\text{SC}}[\rho_A] \\ & + \int \rho_A(\vec{r}) v_B(\vec{r}') d\vec{r}' + \int \frac{\rho_A(\vec{r}) \rho_B(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r}' d\vec{r}'' \\ & + T_s^{\text{nad}}[\rho_A, \rho_B] + E_{xc}^{\text{nad}}[\rho_A, \rho_B] \\ & + E_{v_B}^{\text{HK}}[\rho_B] + \int \rho_B(\vec{r}') v_A(\vec{r}') d\vec{r}' \end{aligned} \quad (1.34)$$

where the non-additive kinetic and exchange-correlation terms are defined as the differences between the corresponding functionals evaluated with the total density *vs.* the sum of the functionals of the two separate components. See ref. 43 for a discussion of the term ΔF^{SC} for various choices of the quantum chemical method used for the solute. For the case of KS-DFT it is:

$$\begin{aligned} \Delta F^{\text{SC(KS)}}[\rho_A] = & \frac{1}{2} \iint \frac{\rho(\vec{r}) \rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' + E_{xc}[\rho_A] \\ & - \int \rho_A(\vec{r}) (v_{\text{eff}}^{\text{KS}}[\rho_A](\vec{r}) - v_A(\vec{r})) d\vec{r} \end{aligned} \quad (1.35)$$

The all-important embedding potential, the quantity that has been tracked so far in this review, is given by

$$\begin{aligned} v_{\text{emb}}[\rho_A, \rho_B, v_B](\vec{r}) = & v_B(\vec{r}) + \int \frac{\rho_B(\vec{r}')}{|\vec{r}' - \vec{r}|} d\vec{r}' + \frac{\delta T_s^{\text{nad}}[\rho_A, \rho_B]}{\delta \rho_A(\vec{r})} \\ & + \frac{\delta E_{xc}^{\text{nad}}[\rho_A, \rho_B]}{\delta \rho_A(\vec{r})} + \frac{\delta \Delta F^{\text{SC}}[\rho_A]}{\delta \rho_A(\vec{r})} \end{aligned} \quad (1.36)$$

To go further, to applications, one needs explicit approximations for the non-additive functionals and a recipe for ρ_B . I leave most of the discussion of these aspects to the review literature^{43,44} but, to give the reader a sense of what is involved, give simple examples now.

For the case of homogeneous gases, the Thomas-Fermi^{49,50} treatment is one starting point:

$$T_s^{\text{nad(TF)}}[\rho_A, \rho_B] = C_{\text{TF}} \int ((\rho_A + \rho_B)^{5/3} - \rho_A^{5/3} - \rho_B^{5/3}) d\vec{r} \quad (1.37)$$

The Dirac exchange functional of the homogeneous electron gas leads to a similar equation with exponents of 4/3. Of course, this diverges for uniform densities, but it does provide a simple approximation for molecular densities. Exact properties of the nonadditive functionals that can help guide the search for better approximations are discussed in Section 2.2.6 of ref. 43 and Section 4.1.1 gives expressions for the GGA level.

As for the generation of $\rho_B(\vec{r})$ six possibilities are mentioned in ref. 43

1. Use the density of a quantum-mechanical method for the isolated subsystem B.
2. Superposed densities of fragments.
3. Optimized by freeze-and-thaw from subsystem DFT.

4. Simplified techniques for polarization.

5. Average from statistical ensembles (see the case study below).

6. Use a buffer zone with capping atoms if covalent bonds link the two subsystems.

Case study 6: Combining FDET with 3D-RISM-KH to model solvatochromism.

Innovative work by Wesolowski, Kovalenko and coworkers⁵¹ brought together the statistical-mechanical 3D-RISM-KH treatment to define the solvent density and FDET to handle the embedding. Shedge and Wesolowski⁴⁸ applied the technique to absorption and emission spectra of a coumarin in several solvents, using a simple (spherically symmetric) model of atom densities, the probabilities for finding the atoms at particular positions being defined self-consistently by 3D-RISM-KH and FDET.

Fig. 11 shows the embedding potential for the coumarin in water and in cyclohexane. In both solvents one observes a “cavity” (not as an explicit part of the computational setup, as in PCM, but occurring “naturally” from self-consistent 3D-RISM/KH/FDET). For cyclohexane, the cavity is “hard”, reflecting the hydrophobic nature of the solvent, whereas, for water, contour lines extend into the solute, reflecting a hydrogen-bonding network around the carbonyl group at the lower right. See ref. 48 for a discussion of how the embedding potentials affect the HOMO and LUMO, yielding solvatochromic shifts. A nice union of two methods, both based on good physics!

3 QM/MM and QM/MM/MD

With apologies for the “folkloric” appearance of the graphics, Fig. 12 shows an old friend, a slide I have used in many presentations over the years to illustrate hybrid quantum mechanical/molecular mechanical (QM/MM) modelling. In the top-left corner I describe a general approach to modelling challenging processes in complex environments, which involves both reductionist and integrative elements (break the system up and put it back together, making methods that work on different scales work together (either consecutively or concurrently). The right-hand figure shows the active site of an enzyme (HIV-1 protease) which would be described with a “high-level” QM method (say DFT-B3LYP, or, perhaps, even better methods...) embedded in lower-level, QM’ (say, DFTB (density functional tight binding) or an SQM (semi-empirical quantum method like PM7) and MM (say, one of the usual biomolecular force fields), representing the environment. For student audiences, referring to the bottom-left part of the slide, I always like to emphasize that some of the richest problems of science may be found in the intersections of disciplines (quantum chemistry, dynamics, stat-mech and meso-to-macro scale methods like finite elements, computational fluid dynamics and kinetic Monte Carlo).

QM/MM has its roots in the 1970s with the seminal works of Warshel and Karplus⁵² and Warshel and Levitt.⁵³ The field has been reviewed many times over^{54–65} so I will only highlight here

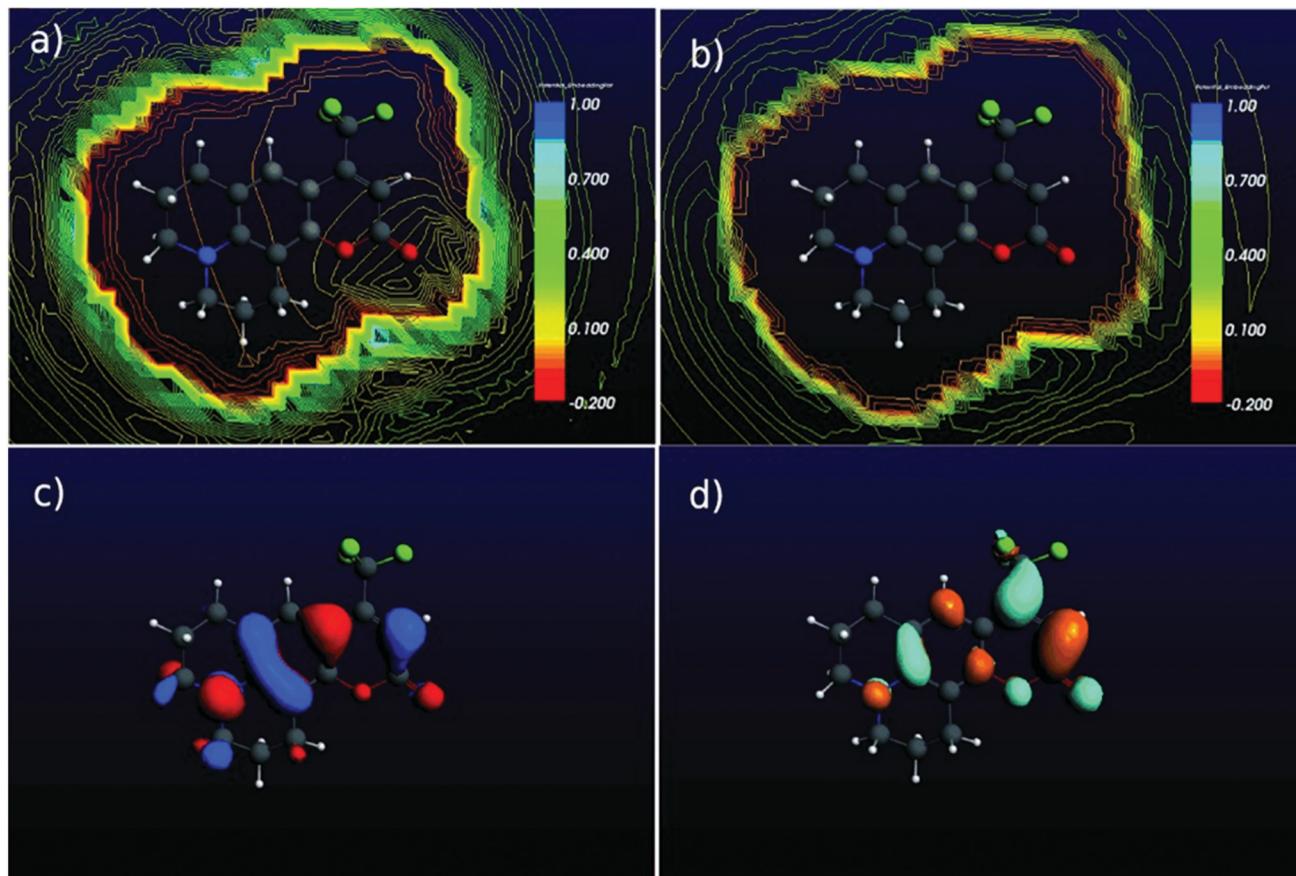


Fig. 11 Embedding potential for the ground state of a coumarin in (a) water and (b) cyclohexane. The lower panels show the frontier orbitals: (c) HOMO and (d) LUMO. Reproduced from ref. 48 with permission from Wiley-VCH GmbH, copyright 2014.

a few of the general features of the main flavors of QM/MM and then propose a computational protocol for modelling enzymes, updated from our recent review⁶⁵ and provided, especially for newcomers, to give a feeling for the types of, hopefully, expert choices for computational set-ups that will have to be made. This is followed with a case study chosen from the field of nanocatalysis (for oil sands upgrading).

The usual model in QM/MM calculations is atomistic, requiring a choice for which atoms are in the QM and which in the MM region. Then a choice has to be made for either an additive or a subtractive (interpolation) approach for the total energy. In the additive scheme, with electrostatic embedding (like in Section 2.2 for point charges), the total energy may be written (see ref. 63 for details on three implementations involving deMon2k) as

$$E = E^{\text{QM}} + E^{\text{QMMM}} + E^{\text{MM}} \quad (1.38)$$

where the QM energy contains two parts, the usual SCF electronic energy, including the embedding potential, and the augmented nuclear repulsion energy,

$$E^{\text{NN}} = \sum_{A < B}^{\text{QM}} \frac{Z_A Z_B}{|\vec{A} - \vec{B}|} + \sum_{A}^{\text{QM}} \sum_{D}^{\text{MM}} \frac{Z_A Q_D}{|\vec{A} - \vec{D}|} \quad (1.39)$$

E^{QMMM} is the mechanical interaction energy, which is often

expressed as a Lennard-Jones potential

$$E^{\text{QMMM}} = \sum_{A}^{\text{QM}} \sum_{D}^{\text{MM}} \varepsilon_{AD} \left[\left(\frac{R_{AD}}{|\vec{A} - \vec{D}|} \right)^{12} - 2 \left(\frac{R_{AD}}{|\vec{A} - \vec{D}|} \right)^6 \right] \quad (1.40)$$

E^{MM} is the molecular mechanical energy, which depends on the force-field chosen, typically

$$E^{\text{MM}} = E^{\text{bond}} + E^{\text{bend}} + E^{\text{tors}} + E^{\text{vdW}} + E^{\text{QQ}} \quad (1.41)$$

with bond length, bending angles, torsional angles, van der Waals and electrostatic contributions.

If covalent bonds join the QM and MM regions then various schemes, such as link atoms, are employed to “smooth the rupture”; I leave details to the review literature and references therein. The additive schemes represent natural extensions of the usual SCF (or other QM techniques) to include the electrostatic embedding potential of the environment along with mechanical embedding.

The subtractive QM/MM schemes are, in essence, extrapolations in which a part of the total system is treated with a high-level of theory (usually QM) and the total system is treated at a low-level of theory (usually MM). In the original formulation of the ONIOM (Our own N-layered Integrated molecular Orbital and Molecular mechanics) method⁶⁰ only mechanical embedding was

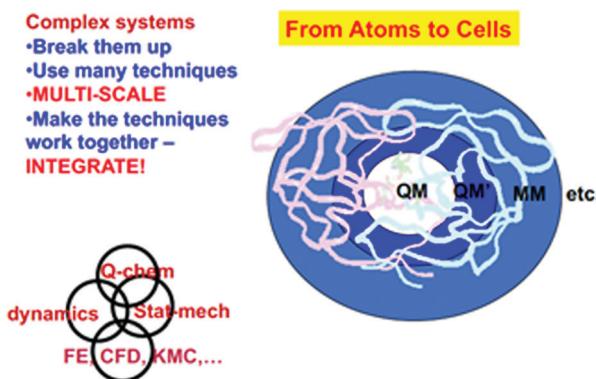


Fig. 12 A slide used often by the author over the last few decades to illustrate multi-scale modelling based on QM/QM'/MM. The QM region represents the active site of an enzyme (HIV-1 protease); embedded in lower-level, QM' and MM, representations of the environment. See the text for further explanation.

considered; more recent extensions include the electrostatic field of the environment. The subtractive energy is

$$E_{\text{QM/MM}} = E_{\text{QM}}^{\text{QM}} + E_{\text{QM/MM}}^{\text{MM}} - E_{\text{QM}}^{\text{MM}} \quad (1.42)$$

where $E_{\text{QM}}^{\text{QM}}$ and $E_{\text{QM}}^{\text{MM}}$ are the energy of the QM region computed at the QM (high) and MM (low) level, respectively. $E_{\text{QM/MM}}^{\text{MM}}$ is the energy of the whole system calculated with MM. Note that, in contrast to the additive schemes, this last term requires force-field parameters for the QM region. There are pros and cons of both additive and subtractive schemes which are well debated in the review literature.

The various QM/MM energy expressions provide approximations to the potential energy surfaces of embedded quantum-chemical systems. *Per se*, they operate at 0K so that statistical mechanics must be brought into play to calculate the entropic terms that, along with the PES, define the free energy. If the portions of the PES that are thermally accessible remain harmonic, then the “usual” rigid-rotor/harmonic oscillator partition functions can be used. However, the general case requires anharmonic portions of the PES and two or more catchment regions (*e.g.* reactants, transition states and products) to be explored. The “sampling problem” joins the “PES accuracy” problem to enrich the lives of multiscale modelers.

There is a rich literature on enhanced sampling techniques, designed to prevent MD trajectories from getting trapped in local minima with high barriers that prevent escape to other catchment regions. I will be content here with a few entries into the field, following a recent concise overview.⁶⁶ There are two broad categories of enhanced sampling techniques, those that require the definition of reaction coordinates or collective variables to guide the sampling and those that do not.

In the first category (*e.g.* umbrella sampling,⁶⁷ metadynamics⁶⁸) a reaction coordinate(s) (collective variable(s) (CV), order parameter(s)) is(are) chosen at the outset and a biasing potential is added. For the case of metadynamics, following,⁶⁶ a small number of CVs, $\mathbf{s}(\mathbf{R})$, are defined to represent the slow

motions. The equilibrium distribution and the free energy are given by:

$$p_0(\mathbf{s}) = \int d\mathbf{R} \delta[\mathbf{s} - \mathbf{s}(\mathbf{R})] p_0(\mathbf{R}) = \langle \delta[\mathbf{s} - \mathbf{s}(\mathbf{R})] \rangle \quad (1.43)$$

$$F(\mathbf{s}) = -\frac{1}{\beta} \log[p_0(\mathbf{s})] \quad (1.44)$$

$$p_0(\mathbf{R}) = \frac{e^{-\beta U(\mathbf{R})}}{\int d\mathbf{R} e^{-\beta U(\mathbf{R})}} \quad (1.45)$$

A biasing potential is added to help overcome the barriers separating different catchment regions and the free energy becomes

$$F(\mathbf{s}) = -\frac{1}{\beta} \log[p(\mathbf{s})] - V(\mathbf{s}) \quad (1.46)$$

where $p(\mathbf{s})$ is the sampled distribution. In metadynamics the history-dependent bias potential gradually accumulates repulsive Gaussians to induce the dynamics to surmount barriers:

$$V(\mathbf{s}, t) = \int_0^t d\tau G[\mathbf{s}, \mathbf{s}(\tau)] \quad (1.47)$$

$$G(\mathbf{s}, \mathbf{s}') = \sum_i^d \omega \exp[-(s_i - s'_i)^2 / 2\sigma_i^2] \quad (1.48)$$

where d is the number of CVs each with its own height, ω , and width, σ .

As time goes on, in an ideal metadynamics run, all of the catchment regions fill up and the distribution flattens, so that the free energy is simply the negative of V . In practice, of course, convergence is never ideal and various schemes, such as well-tempered metadynamics⁶⁹ have been devised.

As a final point to this brief overview on CVs I note that there is a burgeoning field of machine-learning techniques designed to facilitate the identification of CVs or avoid them all together (*e.g.* ref. 70–75), making exquisite use of the correlations inherent in the dynamics. Such methods will undoubtedly occupy a growing space in multiscale modelling.

As an example for methods that do not require CVs, consider replica-exchange MD (REMD) (aka parallel tempering).^{66,76,77} REMD runs several parallel simulations, replicas, at different temperatures, with a Monte Carlo weight determining if coordinates of the replicas are swapped. The high-T replicas allow higher-energy portions of the PES to be explored. The acceptance probability for swaps is:

$$P_{\text{ex}} = \min \left\{ 1, \frac{e^{-\left[\frac{U_i(\mathbf{R}_j)}{k_B T_i} + \frac{U_j(\mathbf{R}_i)}{k_B T_j} \right]}}{e^{-\left[\frac{U_i(\mathbf{R}_i)}{k_B T_i} + \frac{U_j(\mathbf{R}_j)}{k_B T_j} \right]}} \right\} \quad (1.49)$$

The calculation of P_{ex} is inexpensive because the potential is already available. See, *e.g.*, ref. 66 and refs. therein for a review of variants and the critical issue of selecting the exchange attempt frequency.

For both categories, CVs or no CVs, machine learning will undoubtedly play a growing role in speeding up the calculations or, indeed, in improving accuracy through schemes like Δ -ML which learns the difference between a lower-level and a higher-level method, allowing the sampling advantage of the former along with the accuracy of the (learned) latter (see *e.g.* ref. 78 and 79).

Case study 7: Proposed protocol for QM/MM/MD of enzymes

In our previous review⁶⁵ we ended with a proposed computational protocol for the multiscale modelling of enzymes. I repeat it here, with just one addition in point 28 (I view this as a living document that will have to be further revised, amended and extended as the field evolves). I hope this protocol will give newcomers a good idea of some of the choices to be made, and potential pitfalls:

1. Choose a relevant biological process and system. What are the biological/biochemical/biophysical questions that are to be addressed? This choice will of necessity involve only a small part of biological reality. We are in a reductionist mode here; the question of integration into a systems framework will, for the most part, be left to other publications.

2. Choose an appropriate active component of the biological system (in our case, an enzyme active site).

3. Consider a proposed mechanism and build a model, using crystallographic data if available, from the active site outwards. Choose important protein residues, substrates, intermediates, and products, water molecules involved in the mechanism, *etc.*, knowing that the validity of the results will depend critically on these choices. Crystallographic data with high resolutions (of at least 2 Å) might be essential for the choice of starting structures (sometimes Boltzmann weighted), but the true active site does not always coincide with the binding sites from X-ray crystal structures. Protein crystallography is powerful, but not omnipotent.

4. In the case of a “simple” QM-cluster model, it is necessary to fix some of the peripheral atoms at their crystallographic positions. The current consensus is to fix three atoms at each boundary point. Fixing both the hydrogen atoms and the closest heavy atom at the boundary improves the results for QM-cluster optimization.

5. Also in the case of a small QM-cluster model, include the effects of the surrounding protein and solvent through the use of a PCM. Small models are sensitive to the value of the dielectric constant; results converge as the cluster size increases. Choose the value of the dielectric constant (often chosen to be 4; some workers test whether 4 and 80 yield different results, using this as an indicator for whether the cluster size has to be increased).

6. Decide on the QM method to use (*ab initio*, DFT, or SQM methods; dispersion corrections are needed for high accuracy) along with the computational parameters (spin states, basis sets, auxiliary basis sets, SCF, and geometry convergence criteria *etc.*). Remember that the B3LYP functional is often adequate but it is not omnipotent. If in any doubt, check the sensitivity to the functional and calibrate against *ab-initio* methods (like LCCSD(T)) if possible.

7. Choose a software package (or write one...).

8. Design the study to check that the QM-cluster model is large enough either through cluster-convergence tests,

identification of residues with significant electrostatic effects that are important for the reaction energies (using, *e.g.*, the charge deletion analysis method) or, if that is not possible, through well-reasoned choices of the residues to include, ideally calibrated against experimental data. These days, 300 or so QM atoms are becoming routine. Larger QM-cluster (more than 600 atoms) models can be considered, however nonpolar solvation, electrostatic solvation, dispersion, the surroundings, and SIE effects may be crucial.

9. Decide whether the methodology is accurate enough to allow the use of calculated energies for all steps of the mechanism or if the use of some empirical data (*e.g.*, experimental redox potentials or pKa values) is preferred/necessary.

10. A small QM-cluster model may provide valid insight into non-metalloenzymes, but caution is advised. The rigidity of the active site can be caused by either noncovalent interactions of its organic building blocks or metal coordination.

11. A QM-cluster model may give good agreement with the experimental activation energy but bringing in the environment could destroy that agreement. Always be on guard about getting the “right answer for the wrong reason” when comparing with a single experimental datum.

12. It is safer to extend the model, most often using hybrid QM/MM methods to incorporate the effects of the surrounding protein and solvent.

13. QM/MM treatments show significantly faster convergence than QM-cluster; NMR chemical shifts provide a stringent test; mechanical embedding can lead to slow convergence relative to electrostatic embedding; a charge-shift analysis can indicate the minimum number of protein residues for large-QM calculations.

14. The role of water may be crucial. Adding explicit water molecules may turn out to be the key to valid mechanistic insight.

15. SQM methods (including DFTB) have to be carefully parameterized (often against DFT results); tests should be performed against a wide range of chemical properties. Solvent boundary potential methods might be helpful in large enzyme systems, and the reaction energetics can be sensitive to the choice of boundary potential, contrary to vibrational frequencies.

16. The ONIOM (QM/QM') approach is a very useful method to investigate geometries. In some cases, the activation barriers and reaction energies are not sensitive to the choice of functional, but in other cases they are. Caveat: ONIOM (QM/QM') fails for cases in which the QM and the QM' approaches do not describe the same electronic state during the reaction. The accuracy of a QM' method has to be tested before it can be used for a new system. SQM methods may result in a wrong reaction pathway even though activation barriers agree with experiment.

17. Enzymatic electron-transfer reactions require specialized techniques based on the nonadiabatic 2-state MT (Marcus theory) and its extensions. If you want to treat electron transfer in enzymes, you should learn MT and its various extensions.

18. Decide whether entropic and dynamic effects are likely to be important (for example if charged or gas species are released to the solvent) and, if so, decide on whether a QM or QM/MM approach (calculating the partition function within a harmonic-

oscillator approximation) may be used or whether a MD-based approach (*e.g.*, FEP theory, umbrella sampling, or metadynamics) should be used to properly sample phase space. Look for similar enzymes and reactions in the literature as a guide. If none are available, exercise caution in drawing conclusions about the role of entropy; ignore it at your peril.

19. Be aware of the local minima problem; be prepared to use many snapshots or to calculate free energies.

20. For a QM/MM/MD approach design a protocol (pre-equilibration with a classical force-field, substrate docking, boundary conditions, number and length of sampling “windows”, *etc.*). QM/MM/MD may be the only option for flexible coordination shells.

21. For reactions involving hydrogen transfer – atoms, protons, hydrides – be aware that NQEs (zero-point energies, tunnelling) may be crucial, for example in the calculation of KIEs.

22. Proton-coupled electron transfer reactions require specialized techniques based on extensions of MT that include nonadiabatic transitions between reactant and product electron-proton vibronic states.

23. DFTB methods can be successful for QM/MM/MD simulations. In some cases, good correlation has been reached with either the experimental or DFT results while in others various discrepancies have been found. Although a good qualitative account of enzymatic reactions may be expected with DFTB, caution is still needed for quantitative energetics.

24. Decide whether it is possible to reach sufficient sampling using a QM or QM/MM Hamiltonian. Reference potential methods (*e.g.* EVB) are useful alternatives to traditional DFT, SQM, and *ab initio* methodologies; however, the reference and target potentials must be similar. Another alternative is the QTCP method (sampling with classical MD and doing a limited number of QM/MM calculations). However, the results are sensitive to the functional and the solvation model so that due caution is needed.

25. Consider an integrative approach. Often the problem of interest can be divided into parts where large conformational

changes are expected (and the classical MD is able to answer questions), while the most crucial steps are modeled by QM/MM/MD. Metadynamics can be used to enhance the sampling when necessary.

26. Pay attention to the quality of the force-field in QM/MM and QM/MM/MD computations. Polarizable force-fields are required, for example, to capture the dynamics of water in the first coordination shell of cations; effects are dramatic for divalent cations.

27. In QM/MM/MD simulations, one of the most delicate choices is that of a suitable reaction coordinate. This remains a prime challenge to the field and one where new approaches would be very welcome.

28. Examine whether the recent machine-learning techniques could be of help in accelerating conformational searches and/or improving accuracy.

29. For all approaches, analyze the results in terms of structural, energetic, and dynamic aspects and using the interpretational tools for DFT or other QM methods as appropriate.

30. Let the literature be your guide (an afternoon in the (online) library may save many months of wasted effort at the computer).

31. Get more computer power...

Case study 8: Molybdenum carbide nanocatalysts at work in the *in situ* environment: a DFTB/MM/MD study

Turning from enzymes to another flavor of catalysis, consider now transition-metal-containing nanocatalysts used in the upgrading of Alberta’s oil sands. In a 2015 study⁸⁰ Liu used a newly-parameterized DFTB/MM/MD approach in which a 1.2 nm nanoparticle of Mo₂C (MCNP) with a benzene molecule being hydrogenated, treated with DFTB, was embedded in an MM region of 100 other benzene molecules representing a hydrocarbon environment (Fig. 13).

QM/MM molecular dynamics used umbrella sampling⁶⁷ with a “natural” reaction coordinate, the C–H distance for

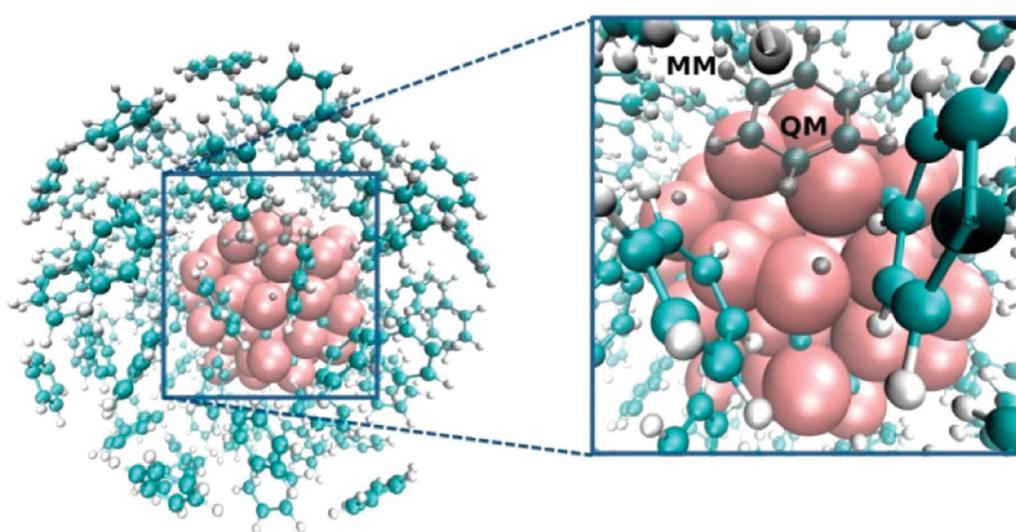


Fig. 13 QM/MM model of a 1.2 nm MCNP (Mo atoms in red and C atoms in cyan) with adsorbed benzene (in black) and the two dissociated H atoms (in black) embedded in the model aromatic solvent. Adapted with permission from ref. 80. Copyright 2015 American Chemical Society.

the hydrogen being added to one of the six carbon atoms of benzene. With the relatively inexpensive DFTB for the QM part, it was possible to map out the entire catalytic cycle, starting with benzene and ending with cyclohexane. The free-energy profiles shown in Fig. 14 (for the second hydrogenation step) compare the umbrella sampling (classical-mechanical motion of the nuclei, fully anharmonic) with the “quantum chemistry” approach (quantum vibrations in the harmonic-oscillator approximation). The differences between the two approaches are substantial, reflecting the “fluidity” of the system; movies (supporting information of ref. 80) show the startlingly high-amplitude of motion of the MCNP at working temperature (673 K). (They really are worth a view!)

Although new ground was broken with this study, one lingering doubt remained, namely the validity of the reaction coordinate, the C–H bond. Validating (or not) this choice may be possible in future work using some of the ML techniques mentioned above^{70–75} (see also Section 7).

4 QM/MMPol

The vast majority of QM/MM calculations use fixed-charge force-fields, with the electrostatic potential from the MM region entering the QM Hamiltonian but the MM part remaining immutable to events in the QM region. This often misses important mutual interactions of the QM and MM regions, leading to a growing number of methodological studies and applications that include polarization of the MM region, often iterating to a self-consistent treatment with mutual polarization of both regions. Indeed, as often pointed out in review articles,

the original QM/MM paper⁵³ proposed polarizable MM. Here I will review briefly two of the most common approaches, Drude oscillators and induced dipoles.

4.1. Drude Oscillators

The Drude model has its roots in antiquity, the turn-of-the-last-century electron theory of metals;⁸¹ a contemporary depiction is shown in Fig. 15.

In the Drude model, as described in many recent articles and reviews,^{82–92} each polarizable (non-hydrogen) atom is joined by an auxiliary charged particle held at a distance d by a harmonic spring, allowing the atom to respond to an external electric field by an induced dipole moment

$$\vec{\mu} = \frac{q_D^2}{K_D} \vec{E} \quad (1.50)$$

The charge on the Drude particle is determined by the atomic polarizability:

$$\alpha = \frac{q_D^2}{K_D} \quad (1.51)$$

To include short-range, *e.g.* 1–2 and 1–3, interactions without a “polarization catastrophe”, screening of the Coulomb interaction is introduced by a Thole damping function:⁹³

$$S_{ij}(r_{ij}) = 1 - \left[\left(1 + \frac{(t_i + t_j)r_{ij}}{2(\alpha_i\alpha_j)^{1/6}} \right) \right] \exp \left[\frac{-(t_i + t_j)r_{ij}}{(\alpha_i\alpha_j)^{1/6}} \right] \quad (1.52)$$

The Drude particles enter the dynamics either through an SCF approach in which their positions are optimized at each time step or, more often, through an extended Lagrangian

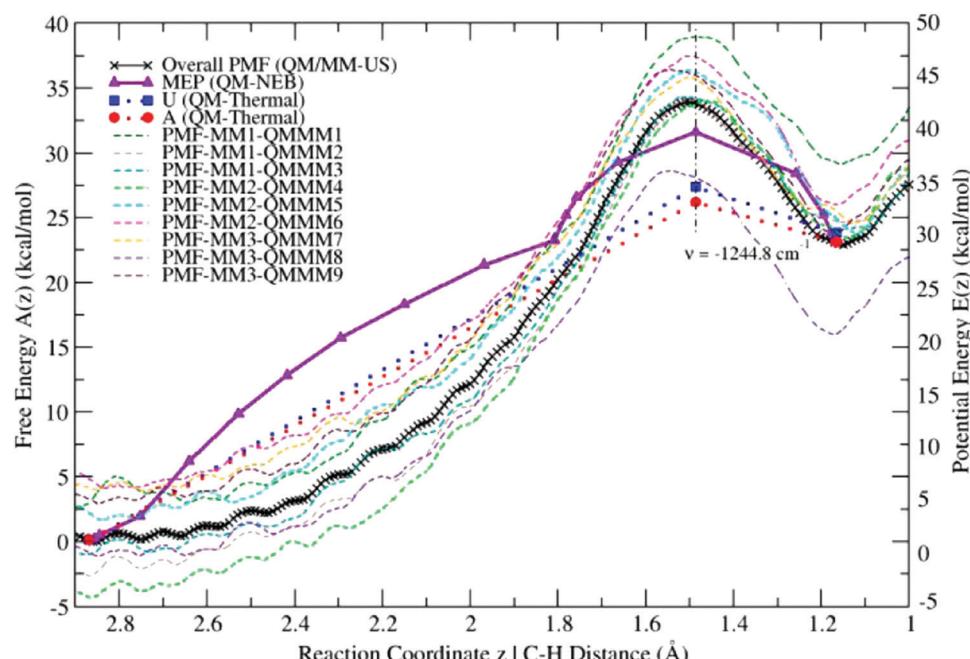


Fig. 14 Calculated free energy profiles of the second hydrogenation reactions from different starting configurations. “MM” means the MM configuration of the solvent benzene molecules, “QMMM” means the overall configuration of the whole system. Reprinted with permission from ref. 80. Copyright 2015 American Chemical Society.

scheme in which the Drude particles are assigned a small mass and a separate thermostat.⁹⁴

Parameterization of the Drude FF has been an on-going endeavour over the last two decades, with proteins, DNA, lipids (ref. 91 and refs. therein) and protein-ion interactions^{83–85,89,90} representing major targets. For the latter, although it is clear that Drude represents a great improvement over fixed-charge, additive FFs, the frontier now addresses questions about divalent ions and the possible importance of charge-transfer terms that go beyond the Drude model.⁹⁰

While most of the work on the Drude model has been on the standalone FF, there have been a few QM/MMDrude studies.

The formalism was mapped out quite nicely by Lu and Zhang.⁸² The total QM/MMDrude energy is

$$E_{\text{tot}} = E_{\text{qm}} + E_{\text{qm/mm}} + E_{\text{mm}} + E_{\text{Drude-ele}} + E_{\text{self}} \quad (1.53)$$

where $E_{\text{Drude-ele}}$ is the electrostatic energy involving the Drude particles:

$$\begin{aligned} E_{\text{Drude-ele}} &= \sum_{\alpha}^{\text{MM}} \sum_{\beta'}^{\text{MM}'} \frac{q_{\alpha} q_{\beta'}}{r_{\alpha \beta'}} + \sum_{\alpha'}^{\text{MM}'} \sum_{\beta' > \alpha'}^{\text{MM}'} \frac{q_{\alpha'} q_{\beta'}}{r_{\alpha \beta'}} + \\ &\sum_A^{\text{QM}} \sum_{\alpha}^{\text{MM}'} \frac{Q_A q_{\alpha'}}{r_{A \alpha'}} + \sum_{\mu, \nu}^{\text{AO}} D_{\mu \nu} \left\langle \mu \left| - \sum_i^{\text{electron}} \sum_{\alpha'}^{\text{MM}'} \frac{q_{\alpha'}}{r_{i \alpha'}} \right| \nu \right\rangle \quad (1.54) \\ &= \sum_{\alpha'}^{\text{MM}'} q_{\alpha'} \left(\varphi_{\alpha'}^{\text{MM}} + \frac{1}{2} \varphi_{\alpha'}^{\text{MM}'} + \varphi_{\alpha'}^{\text{Nuc}} + \varphi_{\alpha'}^{\text{electron}} \right) \end{aligned}$$

and

$$E_{\text{self}} = \frac{1}{2} \sum_{\alpha'}^{\text{MM}'} K_{\alpha'} d_{\alpha'}^2 \quad (1.55)$$

where a prime indicates the Drude particle sites. The four terms in eqn (54) represent the electrostatic potential at the Drude site arising from the classical point charges, the charges of Drude particles, the QM nuclei and the QM electrons, respectively. Direct SCF and microiterative schemes were examined

for a number of model systems (QM water- Drude water, the S_N2 methyl exchange in methyl-Cl-Cl⁻, the intramolecular proton transfer in glycine).

König *et al.*⁸⁸ compared QM/MM calculations with and without Drude oscillators for the hydration free energies of twelve simple molecules covering a range of hydrophilicity/phobicity and using several QM methods ranging from Hartree-Fock to GGA and hybrid density functional approximations, to MP2 and semiempirical methods. The effect of MM polarization is only moderate compared with the well-developed CHARMM fixed-charge FF for this application.

Case study 9: Effect of Drude-model polarization on the QM/MM free-energy profile for an enzymatic reaction

I choose for a case study involving Drude, the 2017 work of Ganguly, Boulanger and Thiel⁸⁷ who examined the effect of MM polarization for two catalytic reactions, (i) the Claisen rearrangement in chorismate mutase and (ii) the hydroxylation reaction in p-hydroxybenzoate hydroxylase. For reaction (i) the effects are quite modest (as expected for a reaction involving negligible charge transfer), whereas they are substantial for reaction ii, as shown in Fig. 16. The activation free energy increases from 13.7 kcal mol⁻¹ to 18.2 kcal mol⁻¹ and there are significant differences in the geometries.

Since the DFT/Drude calculations are about 7 times more expensive than calculations with the additive FF, the authors conclude with the thought that a moderate extension of the QM region of a QM/MM-ADD method might be competitive and that “further validation work is required to establish the best QM/MM-based procedure for handling polarization effects in enzymatic reactions.”.

4.2. Induced dipoles

A good entry into the literature of polarizable force-fields that use calculations of electric fields and induced multipole moments is the recent Account by Loco *et al.*⁹⁵ which includes key references to the development of the AMOEBA FF^{96–98} in the QM/MMpol context.⁹⁹ While AMOEBA includes response up

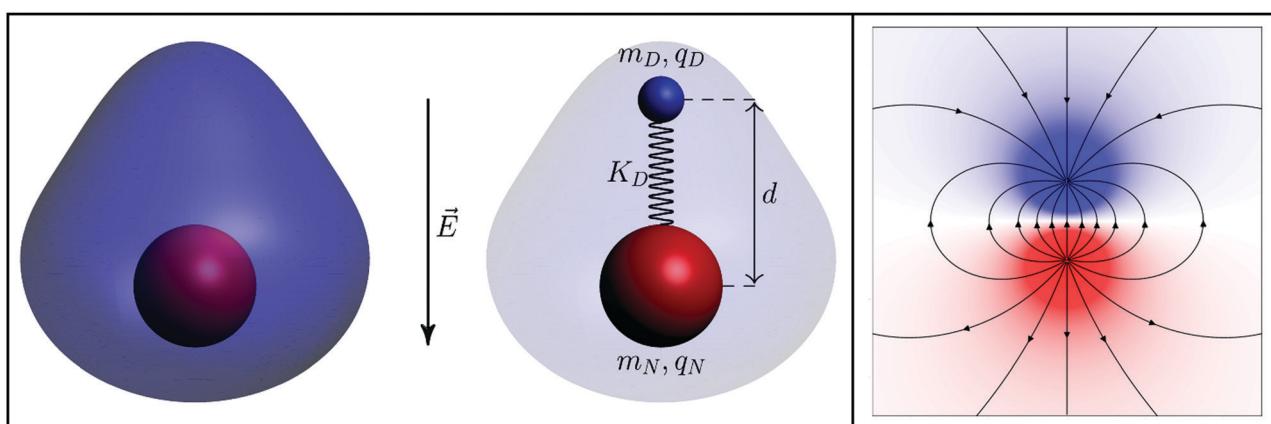


Fig. 15 Schematic of a Drude polarizable atom and resulting FF. Left: Conceptual depiction of electron density around an atom polarized by an external field, \vec{E} . Middle: Drude model of the same atom with Drude particle in blue. The Drude particle has a mass of m_D and charge of q_D , whereas the parent nucleus has a mass of m_N and a charge of q_N . The distance d is controlled by a spring with the force constant K_D . Right: A physical dipole with field lines representing potential gradients. Reproduced from ref. 90 with permission of AIP Publishing, copyright 2020.

to quadrupoles, for simplicity, I will limit the following to the dipole case, as described in ref. 100 from which I also derive Case Study 10 which couples RT-TDDFT with a polarizable environment (a similar method was also reported by Mennucci *et al.*¹⁰¹).

In the point-charge dipole model, each polarizable MM site is provided with an induced dipole, $\vec{\mu}_i$, that is determined by the electric field \vec{F}_i coming from the QM system plus all of the other MM charges and dipoles, plus any external electric fields, the induced dipoles being controlled by atomic polarizabilities:

$$\vec{\mu}_i = \alpha_i \vec{F}_i = \alpha_i (\vec{F}_i^0 + \vec{F}_i^{\text{ind}} + \vec{F}_i^{\text{QM}} + \vec{F}_i^{\text{ext}}) \quad (1.56)$$

$$\vec{F}_i^0 = \sum_{j \in \text{MM} \neq i} \frac{q_j}{r_{ij}^3} \vec{r}_{ij} \quad (1.57)$$

$$\vec{F}_i^{\text{ind}} = - \sum_{j \in \text{MM} \neq i} T_{ij} \mu_j \quad (1.58)$$

$$T_{ij} = \frac{1}{r_{ij}^5} I - \frac{3}{r_{ij}^5} \begin{vmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{vmatrix} \quad (1.59)$$

$$\begin{aligned} \vec{F}_i^{\text{QM}} &= \vec{F}_i^{\text{ZQM}} + \vec{F}_i^{\rho} \\ &= \sum_{k \in \text{QM}} \frac{Z_k}{r_{ik}^3} \vec{r}_{ik} - \int \frac{\rho(\vec{r})}{|\vec{r}_i - \vec{r}|^3} (\vec{r}_i - \vec{r}) d\vec{r} \end{aligned} \quad (1.60)$$

where \vec{F}_i^0 is the electric field created by the other MM permanent charges, \vec{F}_i^{ind} comes from the other MM induced dipoles,

\vec{F}_i^{QM} from the QM region, that is from the atomic nuclear charges, \vec{F}_i^{ZQM} , and the electron density, \vec{F}_i^{ρ} . T_{ij} is the dipole-dipole interaction tensor and I is the identity matrix. The total induction energy contains three terms that reflect the interaction between the MM induced dipoles with i) the MM permanent charges, ii) the atomic nuclei of the QM atoms and iii) the electron cloud:

$$E_{\text{tot}}^{\text{ind}} = E_{\mu-\text{qMM}}^{\text{ind}} + E_{\mu-\text{ZQM}}^{\text{ind}} + E_{\mu-\rho}^{\text{ind}} \quad (1.61)$$

$$E_{\mu-\text{qMM}}^{\text{ind}} = -\frac{1}{2} \sum_{i \in \text{MM}} \vec{\mu}_i \cdot \vec{F}_i^0 \quad (1.62)$$

$$E_{\mu-\text{ZQM}}^{\text{ind}} = -\frac{1}{2} \sum_{i \in \text{MM}} \vec{\mu}_i \cdot \vec{F}_i^{\text{ZQM}} \quad (1.63)$$

$$\begin{aligned} E_{\mu-\rho}^{\text{ind}} &= \sum_{i \in \text{MM}} \int \frac{\rho(\vec{r})}{|\vec{r}_i - \vec{r}|^3} \vec{\mu}_i \cdot d\vec{r} \\ &= \sum_{i \in \text{MM}} \sum_{\sigma, \tau} P_{\sigma\tau} \left\langle \sigma \left| \frac{\vec{\mu}_i}{|\vec{r}_i - \vec{r}|^3} \right| \tau \right\rangle \end{aligned} \quad (1.64)$$

Since the induced dipoles depend on the density, and *vice versa*, an iterative self-consistent field type of approach is needed. In the deMon2k implementation, in order to take advantage of the sophisticated treatment of point-charge embedding,¹⁰² dipoles are represented by two charges of opposite sign separated by a small distance (0.5 Bohr). As was the

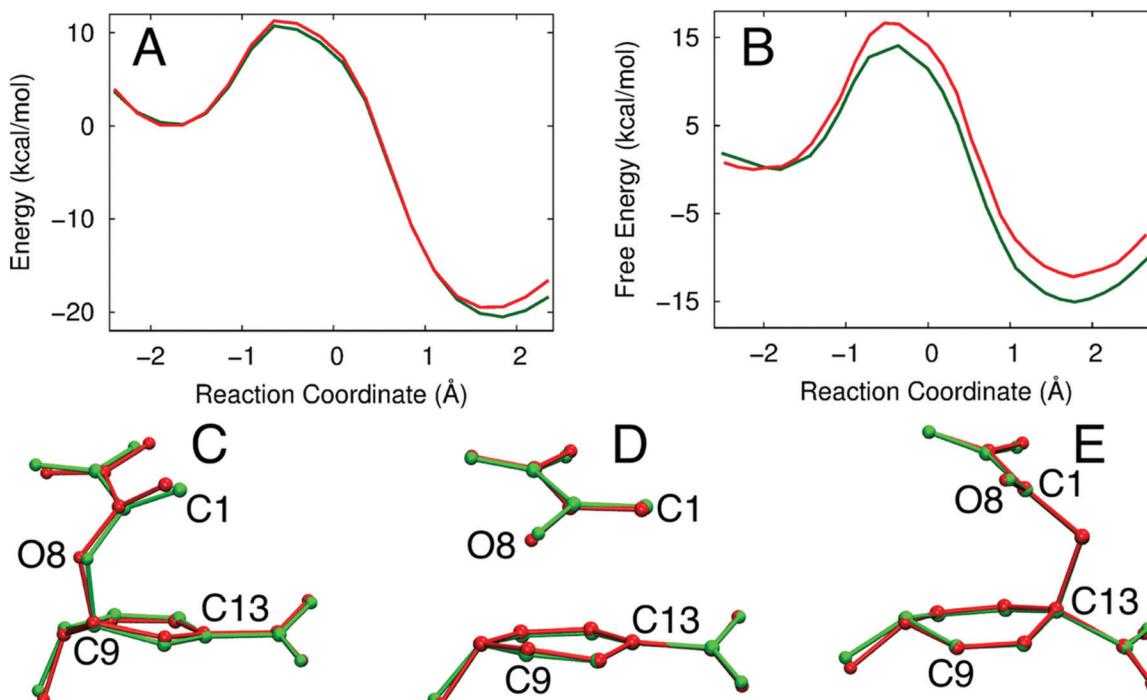


Fig. 16 Investigation of the PHBH-catalyzed hydroxylation reaction using the QM/MM-DO and QM/MM-ADD models. Potential energy profiles (A) and free-energy profiles (B) obtained using QM/MM-DO (red) and QM/MM-ADD (green). The potential energy profiles correspond to a particular snapshot. Superimposition of the average reactant (C), transition state (D) and product (E) geometries sampled in the QM/MM-DO (red) and QM/MM-ADD (green) umbrella-sampling simulations. The QM region was described at the B3LYP/def2-SVP level of theory. Reprinted with permission from ref. 87. Copyright 2017 American Chemical Society.

case for the Drude polarizable FF, care must be taken to avoid the “polarization catastrophe” by introducing Thole factors (see ref. 100 for details).

Case study 10: Electron dynamics in a polarizable environment

Wu *et al.*¹⁰⁰ aim to describe atto-second phenomena by introducing a RT-TDDFT methodology into deMon2k to propagate solutions of the time-dependent Kohn–Sham equation:

$$i\frac{\partial\psi_i(t)}{\partial t} = K[\rho(t)]\psi_i(t) \quad (1.65)$$

where K is the time-dependent Kohn–Sham operator which is a functional of the density at a given time. In the QM/MMpol context, K also contains applied external fields and contributions from the MM charges and induced dipoles, with matrix elements:

$$K_{\sigma\tau}(t) = K_{\sigma\tau}^{\text{isol}}(t) + K_{\sigma\tau}^{\text{app}}(t) + K_{\sigma\tau}^{\text{Q}\rho}(t) + K_{\sigma\tau}^{\mu\rho}(t) \quad (1.66)$$

The adiabatic approximation is made, considering only the spatial dependence of the xc potential, neglecting temporal non-locality (memory effects).¹⁰³ Eqn (65) is discretized and propagated using a Magnus propagator;¹⁰⁴ the details are left to the references.

Fig. 17 depicts a biomolecule, methionine enkephalin solvated by 4030 polarizable water molecules. For analysis, the solvent is separated into concentric shells of 3 Å thickness.

Fig. 18 illustrates the response of the solvent following a delta-function “kick” excitation, followed over 3000 as. The left-hand figures show the change in dipole moment for increasing intensities of the applied electric field, from very weak to very strong (in this fixed-nuclei model dissociation is prevented). The right-hand figures show the autocorrelation function for the first 400 as. At weak field the solvent is progressively less perturbed as the distance from the molecule increases. At the highest field the solvent molecules are essentially in lock step with the molecule, the curves being superimposed out to 15 Å.

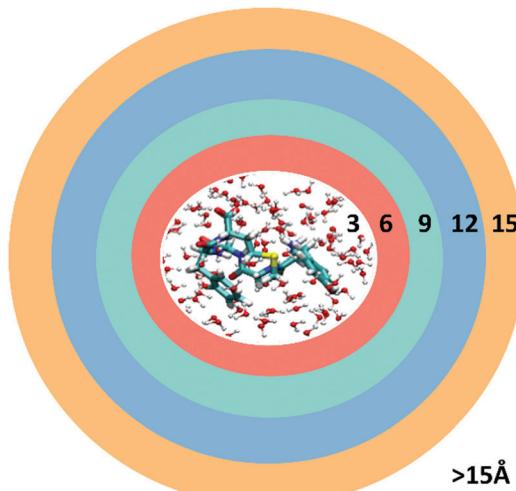


Fig. 17 Methionine enkephalin solvated by 4030 polarizable water molecules. For analysis the solvent is separated into concentric shells of 3 Å thickness.

See ref. 100 for a more detailed analysis in terms of the various sources of electric field.

The work has been extended to include the effects of retardation,¹⁰⁵ Ehrenfest dynamics⁶⁴ and a variety of applications on the effects of ionizing radiation.^{106,107}

5. Coarse-graining

As one strives to extend the spatial and temporal dimensions of MM or MMPol models, to include more and more atoms for longer and longer simulations, the limits of computational resources are reached. Depending on the property under study, it may be interesting to integrate out some of the fine, atomistic, structural features and formulate a model in which atoms are grouped into coarse-grained (CG) “beads”. The MARTINI¹⁰⁸ model is a popular CG-FF, originally formulated for lipids but now extended to proteins and carbohydrates as well (see ref. 109 for the most recent version). Here I give only a brief description of vanilla MARTINI (perish the thought...), comment briefly on recent developments in CG-FF, then a case study in the QM/CN-FF space.

Fig. 19 shows the basic idea of coarse graining; in most cases four atoms are lumped together into a bead. Of course, this removes finer details, usually higher frequency vibrations, resulting in a considerable decrease in the number of degrees of freedom, allowing more atoms and longer simulations for given computational resources.

The force-field itself is built on interaction sites of various types, polar (P), nonpolar (N), apolar (C) and charged (Q). Each bead type has several sub-types that allow a range of parameters. For example, hydrogen-bond donors and acceptors are identified, 5 degrees of polarity are defined, *etc*. The non-bonded interactions are represented by the familiar Lennard-Jones and coulomb terms:

$$U_{\text{LJ}}(r) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^6 \right] \quad (1.67)$$

σ_{ij} represent the closest distance of approach and the same size, 0.47 nm, is chosen for most pairs, with some exceptions for rings, charged beads *etc*. The well-depths depend on the types and sub-types of beads that are interacting. The large (18×18) Table 1 in ref. 108 gives the values of ϵ_{ij} with 10 possible values, ranging from the most polar (5.6 kJ mol^{-1}) to weak apolar (2.0 kJ mol^{-1}), with hydrogen-bonds in nonpolar beads in between, *etc*. Charged groups interact via a screened Coulomb potential:

$$U_{\text{el}} = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r} \quad (1.68)$$

with dielectric constant 15. Bonded interactions are represented by harmonic potentials in both bond distances and angles.

There is a vast literature of applications of CG FFs to biochemical and materials problems (see, *e.g.* ref. 110–115). Fig. 20 depicts the kind of system that can be treated with MARTINI these days. It compares the structure of dimers of transmembrane protein domains (GpA) calculated with

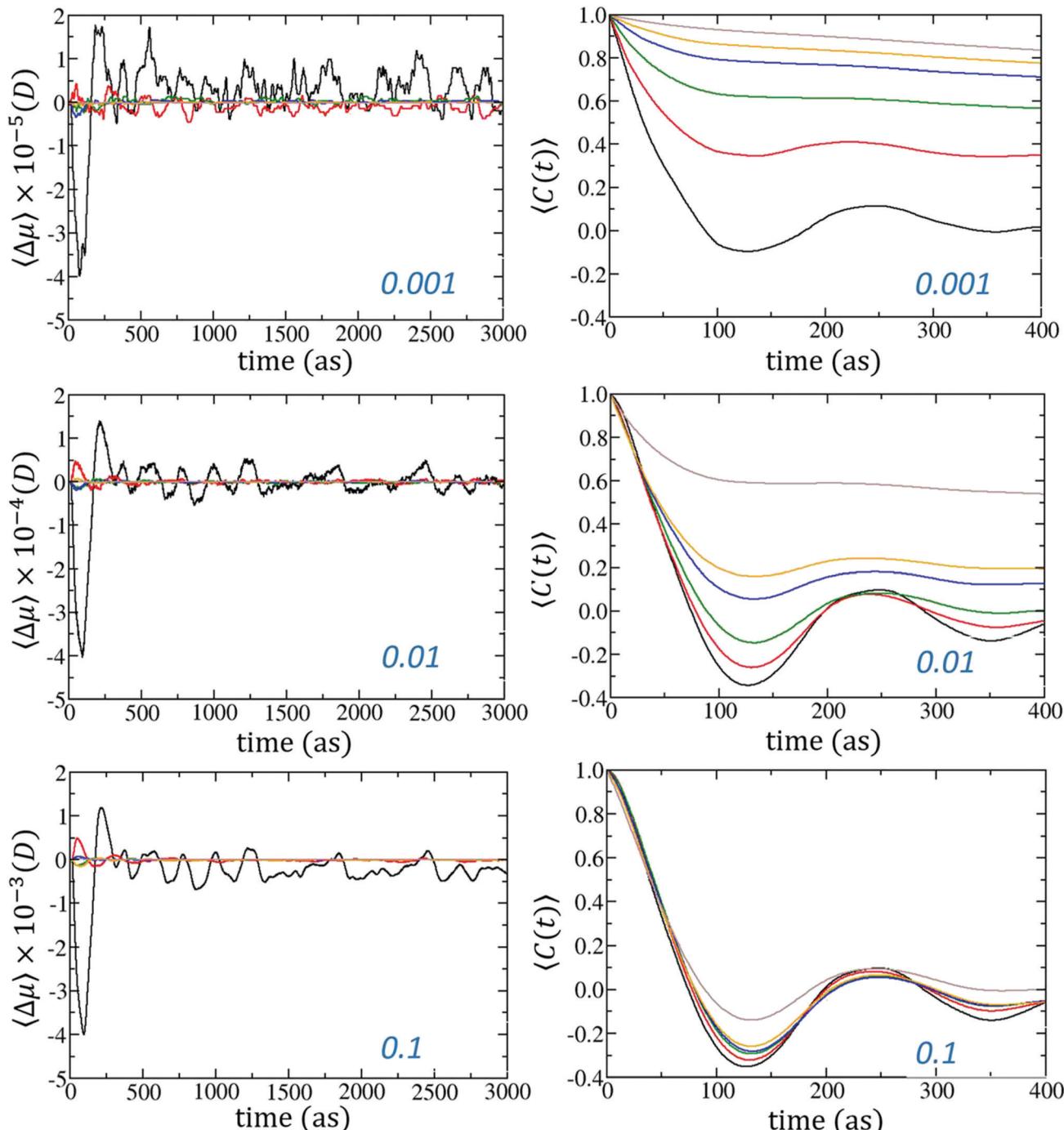


Fig. 18 Evolution of properties of the solvated methionine enkephalin as a function of perturbing electric field strength. Left-hand figures, changes in average solvent dipole moment; right-hand figures, autocorrelation function for shells of solvent (see Fig. 17 for color code). Adapted from ref. 100 with permission. Copyright 2017 American Chemical Society.

MARTINI 2 (left) and MARTINI 3 (centre) with the NMR structure (right). Clearly progress has been made. Has the MARTINI hangover raised in ref. 116 for MARTINI 2 been ameliorated by MARTINI 3? Undoubtedly further improvements, particularly for water¹⁰⁹ would be welcome.

Case study 11: QM/MM-CG

For the coarse-graining case study I select a recent paper¹¹⁷ from the Voth group which combines a CG-FF with a DFT

(DFTB) quantum region, aiming to reduce the overhead associated with electrostatic interactions in atomistic QM/MM techniques. The proof-of-principles examples in this paper are liquid carbon tetrachloride (one QM CCl_4 molecule in a CG solvent) and the reaction of *tert*-butyl hypochlorite (TBHC) with the benzyl radical in a CG CCl_4 solvent.

Fig. 21 shows the workflow, implemented in CP2K,¹¹⁸ which involves mapping of forces between atomistic and coarse-

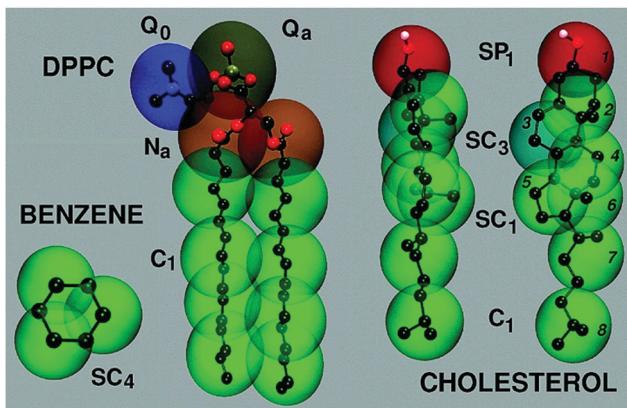


Fig. 19 Mapping between the chemical structure and the coarse-grained model for DPPC, cholesterol, and benzene. The coarse-grained bead types which determine their relative hydrophilicity are indicated. The prefix "S" denotes a special class of CG sites introduced to model rings. Reprinted with permission from ref. 108. Copyright 2007 American Chemical Society.

grained representations (a) and showing the QM/CG representation for CCl_4 in CCl_4 .

Fig. 22 shows the model system for the reaction of TBHC with the benzyl radical in CCl_4 . The reaction free energy is calculated with DFTB-QM/CG-MM using umbrella sampling with the reaction collective variable chosen to be r_{AC} - r_{AB} , where A, B and C are chlorine, oxygen and carbon atoms, respectively, B and C belong to TBHC and benzyl (Fig. 22).

The reaction free energy profiles for QM/MM and QM/CG-MM are shown in Fig. 23. The sampling is done with DFTB for the QM part while the potential of mean force is calculated with DFT (see ref. 117). The CG method accurately reproduces the overall free energy of the reaction, with deviations up to 1–2 kcal mol^{−1} at the peak. The authors cite various factors for this rather small discrepancy: (1) the choice of DFTB for parameterization, (2) neglect of electrostatic interactions, (3) CG basis set incompleteness, (4) neglected CG potential variations along the reaction coordinate and (5) too short QM/MM MD trajectories.

In terms of computational times, relative to atomistic MM, some information is given for CCl_4 in CCl_4 which improves by

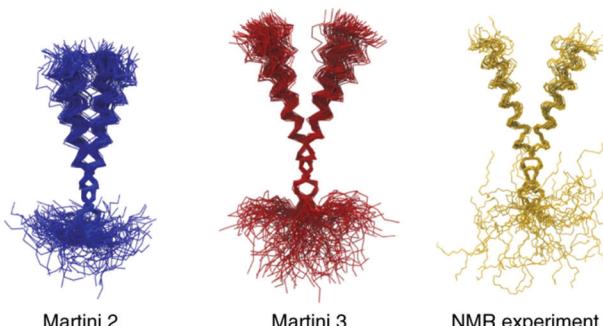


Fig. 20 Structure of dimers of transmembrane protein domains (GpA) calculated with MARTINI 2 (left) and MARTINI 3 (centre) with the NMR structure (right). Reprinted with permission from Springer Nature from ref. 109 COPYRIGHT (2021).

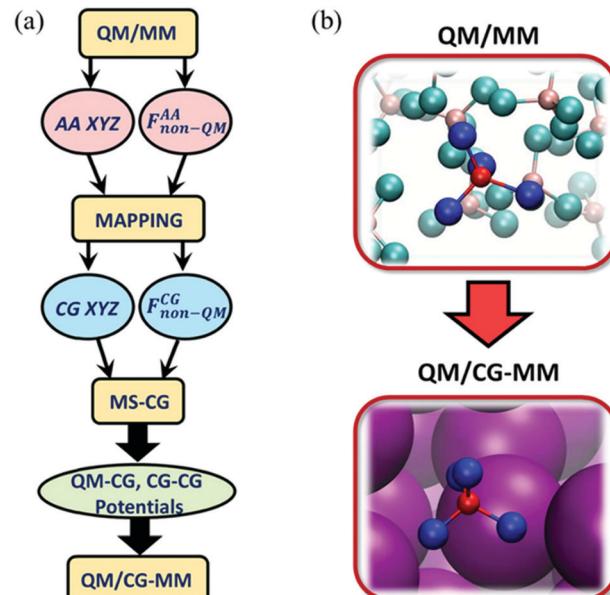


Fig. 21 (a) Information workflow for parameterization of QM/CG-MM. (b) Transitioning from the QM/MM to the QM/CG-MM for CCl_4 in CCl_4 . Reprinted with permission from ref. 117 Copyright 2021 American Chemical Society.

an insignificant amount (7%) for the small model chosen for the study. Expanding the model to 5 QM molecules and 67 495 MM atoms shows better improvement (77%). While this is not

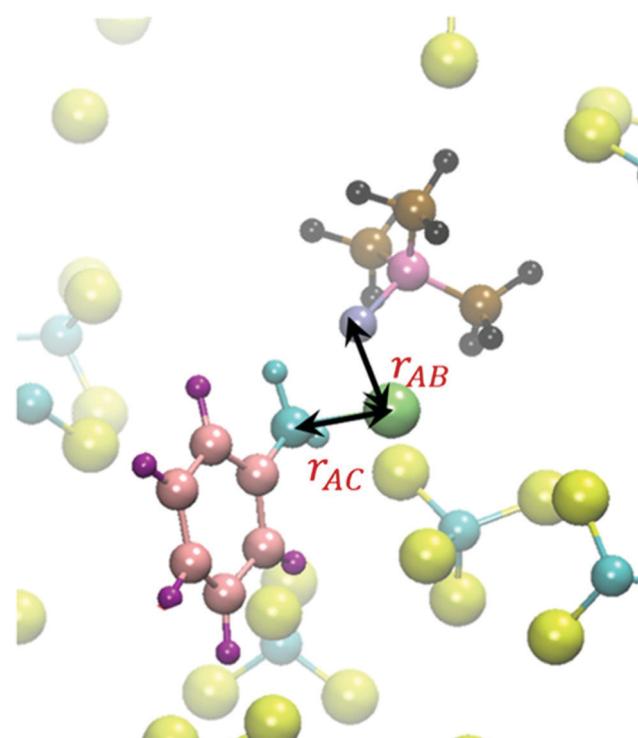


Fig. 22 Reactive model system consisting of a benzyl radical and a TBHC species. Reprinted with permission from ref. 117 Copyright 2021 American Chemical Society.

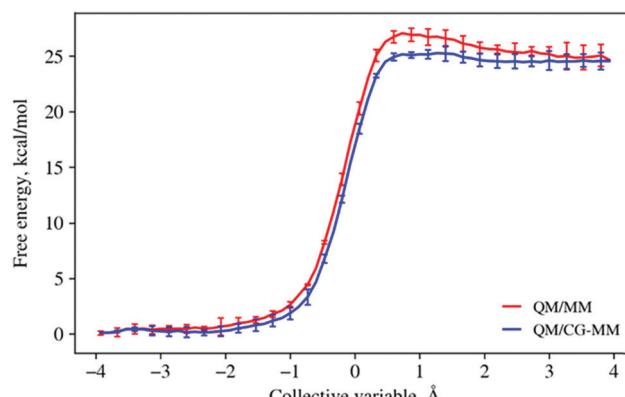


Fig. 23 Free energy profile for the TBHC reaction with the benzyl radical along the reaction coordinate defined in Fig. 22. Energies are computed with DZVP/DFT-QM/MM (red) and DZVP/DFT-QM/CG-MM (blue). The QM/MM profile is averaged over three independent umbrella sampling/WHAM calculations. 95% confidence intervals are shown. Reprinted with permission from ref. 117 Copyright 2021 American Chemical Society.

very significant *per se*, it does indicate the right trend, which would undoubtedly lead to much greater savings for more realistic models. Future developments should be very interesting.

6 Meso- and macro-scale models for materials engineering or cell biology

Having carried the ball from small molecules embedded in dielectric continua, through more and more complex molecules and environments, the ball may now be handed off to researchers who inhabit a larger- and longer-multiscale world. The goal is to provide entries into the literature for systems and properties that require spatial dimensions beyond nanometers and time dimensions beyond milliseconds. I follow the bidirectional theme of materials design, on the one hand, and biology, on the other.

6.1. Materials engineering

An excellent overview¹¹⁹ of the state of the art in multiscale materials modelling as of 2020 gives a structured glimpse of 14 research areas with succinct statements of (i) status (ii) current and future challenges and (iii) advances in science and technology to meet the challenges.

Although some chapters are reasonably familiar territory to this author (*e.g.* Ch. 6 on potentials, including machine learning, aspects of Ch. 14 which follows the cooling of steel and seeks the links in the chain from DFT to force-fields to phase-fields, aspects of Ch. 15 on the required cyberinfrastructure, data bases and the need for new workflow protocols and software), some of the concepts, though undoubtedly well known to a sub-set of materials science modelers, were new to me, such as the phase-field approach that links microstructures to macro materials properties, or the acronym UQ (uncertainty quantification) which reflects the crucial task of knowing the error bars to associate with simulations. Overall, the fifteen chapters convey a picture of research at the very frontier of a still-emerging field,

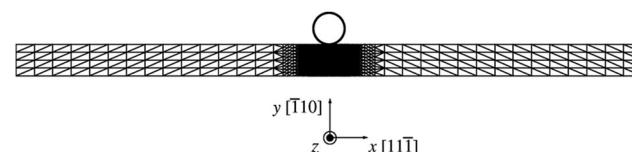
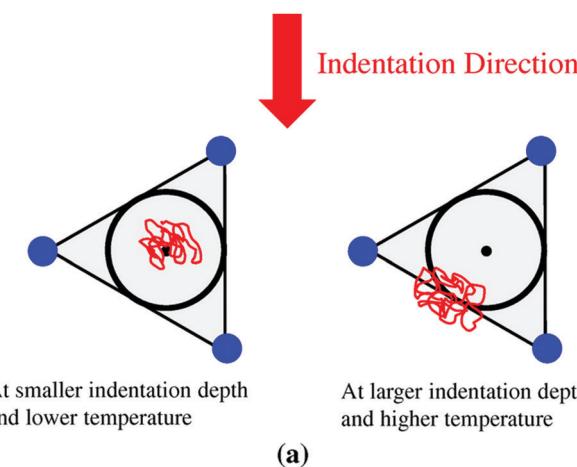


Fig. 24 The hyper-QC model for nanoindentation consisting of a cylindrical indenter and a Ni fcc thin film. The region beneath the indenter is fully refined down to the atomic-scale with a coarser mesh used in the surrounding continuum region. Reproduced from ref. 120 with permission from Taylor & Francis Ltd, <http://www.tandfonline.com>. Copyright 2017.

with many partial successes to report but, for the most part, still a work in progress, the holy grail of fully integrated, hierarchical or concurrent, scale-bridging being on the horizon but a lot of work remaining before total success may be claimed. An ideal field for younger researchers! It is recommended reading for those wishing to enter the mesoscopic world of materials modelling.



At smaller indentation depth
and lower temperature

At larger indentation depth
and higher temperature

(a)

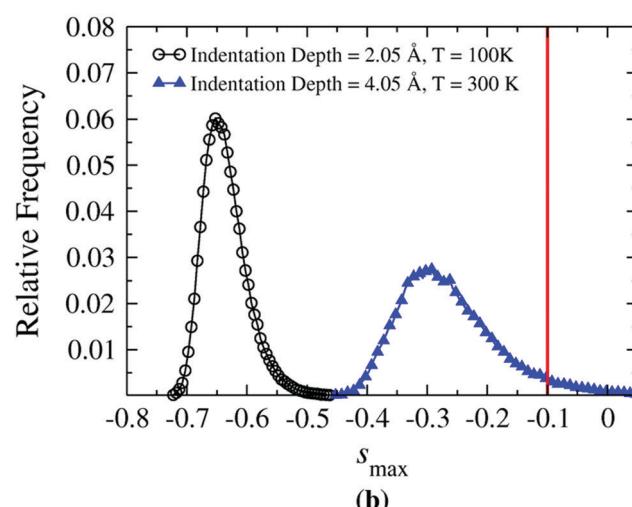


Fig. 25 (a) The region visited by an atom in a slip triangle; and (b) the probability distributions at an indenter velocity of $v = 10-4 \text{ m s}^{-1}$ for two cases: (1) an indentation depth of 2.05 Å at $T = 100 \text{ K}$, and (2) an indentation depth of 4.05 Å at $T = 300 \text{ K}$. Reproduced from ref. 120 with permission from Taylor & Francis Ltd, <http://www.tandfonline.com>, copyright 2017.

Case study 12: Accelerated quasi-continuum: nanoindentation

Of the myriad of possible case studies, I choose a work on nanoindentation¹²⁰ because it bridges spatial dimensions and accelerates time. I will leave the details to the interested reader; suffice it to say that the quasi-continuum (QC) method contains both atomistic and continuum regions (represented with finite elements) and that hyper-QC brings in time acceleration through a bias potential, along the lines of, say, metadynamics.

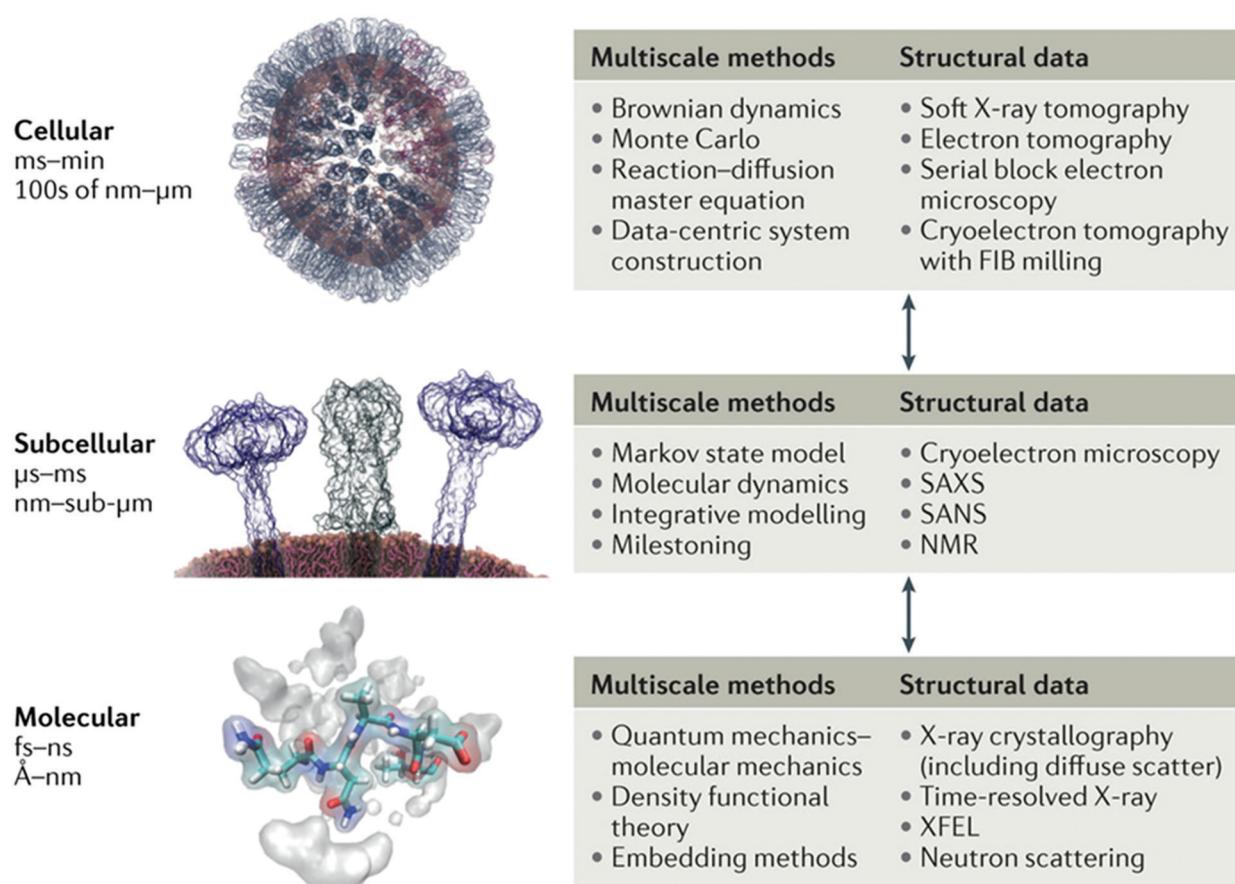
For the nanoindentation study, Fig. 24 shows the model, a 2000 Å × 100 Å fcc Ni thin film and a cylindrical indenter. The film contains 79 461 atoms, but only 4779 atoms in the atomistic region. The continuum region is coarse-grained using an FEM mesh with 1356 nodes. There are a total of 6135 “repatoms” (less than a 10th of the number of atoms in the system).

Fig. 25 shows some of the results, the amount of slip for different indentations, along with intriguing trajectories showing dislocation. The speedup relative to fully atomistic simulations is estimated to be a factor of 1000 at 300 K and 10 000 at 200 K.

6.2 Cell biology

I often say “proteins are nano; cells are micro”; in this section I will focus on multiscale methods that extend into the cellular regime (while recognizing that another school of thought attempts to reach length scales of microns and beyond, and time scales of milliseconds and beyond, in atomistic simulations, through a combination of specialized hardware and software¹²¹). I choose three recent papers as entries into the literature.

Amaro and Mulholland¹²² focus on drug discovery and discuss scales starting at the cellular level, then cascading down to subcellular, molecular and atomistic and, finally to electronic structure. Indeed Fig. 26 shows electronic clouds under the Molecular banner. Each broad category is accompanied by some of the multiscale methods used and, importantly, the modern structural methods that can be interpreted through multiscale simulations. At the top end, Brownian Dynamics is a rough biological equivalent to finite-element methods in engineering.



Nature Reviews | Chemistry

Fig. 26 Multiscale structure-based and physics-based methods bridging from atoms to cells. Emerging multiscale computational methods coupled with increasingly accurate structural data on biological and chemical systems enables the development of highly detailed and predictive models of drug action across spatial scales ranging from angstroms to microns, and temporal scales ranging from femtoseconds to minutes. Such approaches can be gainfully used to address a number of outstanding challenges in drug discovery and design (TABLE 1). FIB, focused ion beam; SAXS, small angle X-ray scattering; SANS, small angle neutron scattering; XFEL, X-ray free-electron laser. Reprinted with permission from Nature Springer from ref. 122 COPYRIGHT (2018).

Feig and Sugita review whole-cell models and simulations in molecular detail,¹²³ addressing the challenge that “a complete link between molecular structure and biological function requires the integration of structure and conformational dynamics at the atomistic level with dynamics and interactions at the cellular level under realistic biological conditions”. The focus is on physical models, as opposed to kinetic networks, the hope being that the kinetics will eventually follow from the bottom up. Fig. 27 will hopefully whet the reader’s appetite to explore further.

The final entry-point into the literature, a paper from Schulten and coworkers,¹²⁴ “Atoms to Phenotypes: Molecular Design Principles of Cellular Energy Metabolism” is also the source of Case Study 13.

Case study 13: A photosynthetic chromatophore at work, converting light energy into ATP

Fig. 28 portrays the vast scope of the environmental factors influencing the function of a photosynthetic chromatophore. Account must be taken of light absorption, redox reactions, molecular diffusion, within a multi-protein-lipid complex that

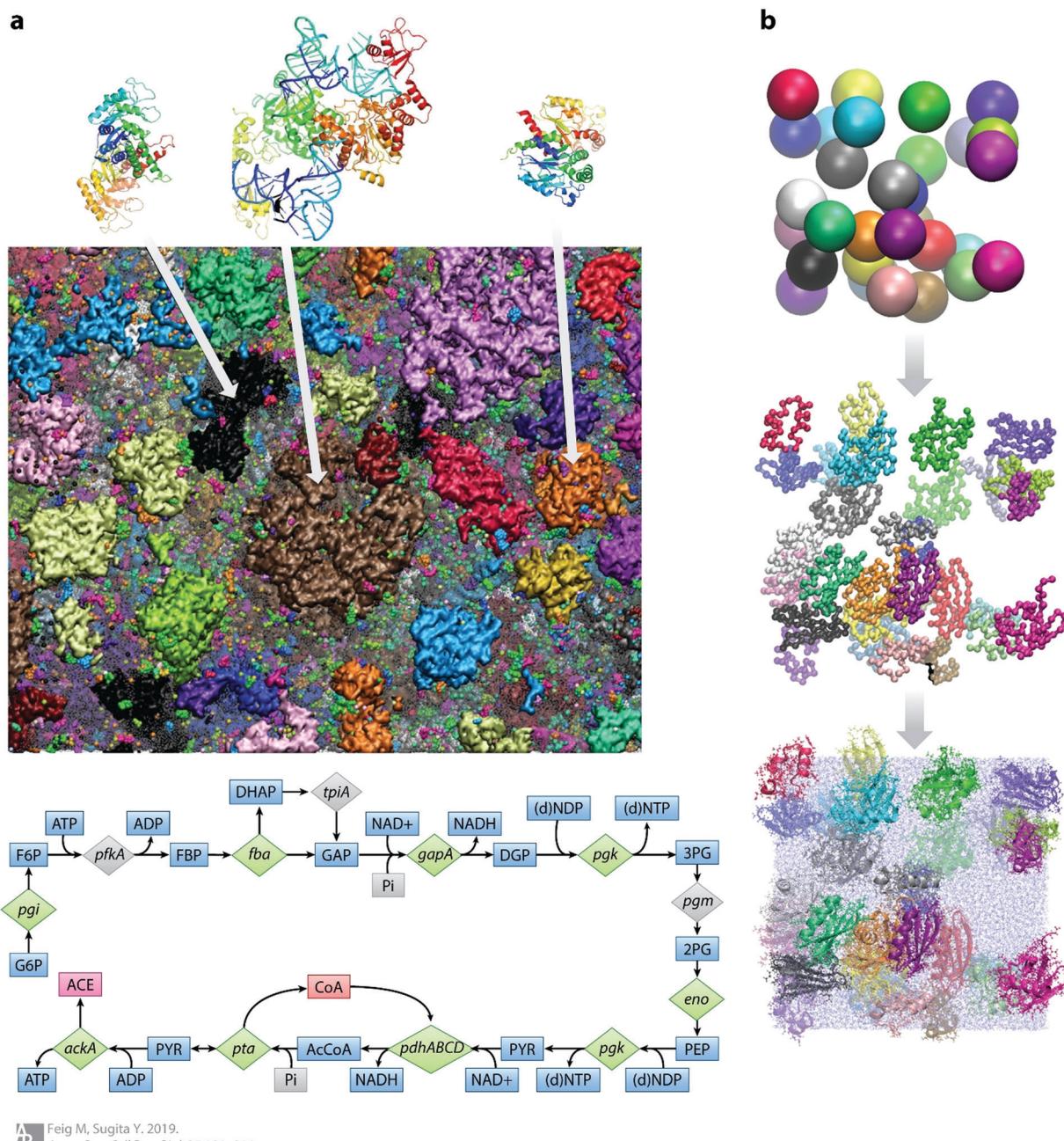


Fig. 27 (a) Construction of cellular systems from atomistic structures of individual molecules based on biochemical pathway reconstruction for the cytoplasm of *Mycoplasma genitalium*. (b) Multiscale assembly protocol from spherical models to a fully solvated atomistic system as described in ref. 125. Reprinted with permission from Annual Reviews Inc. from ref. 123 Copyright (2019).

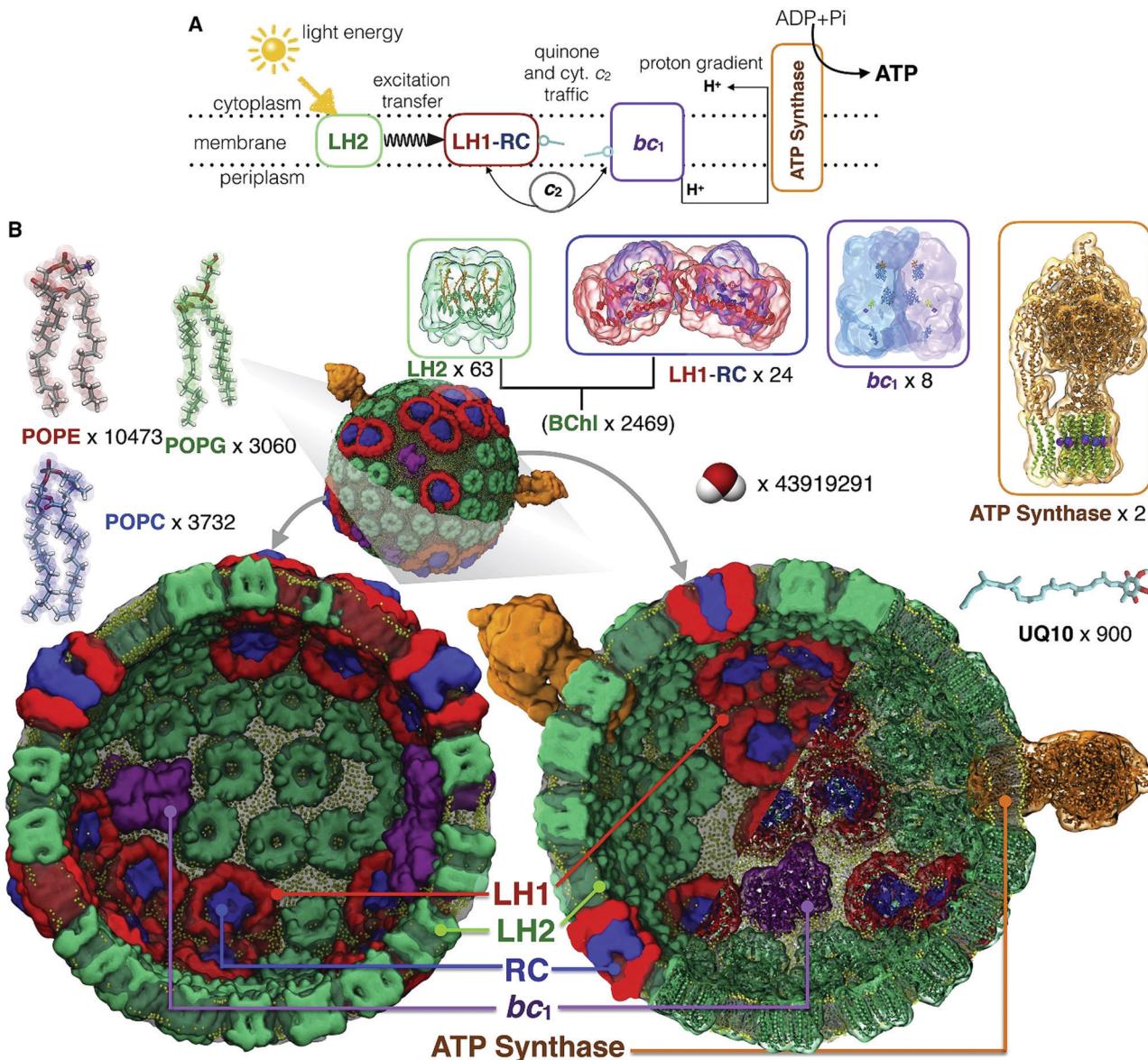


Fig. 28 Physical model of a chromatophore. (A) Photosynthetic energy conversion in a purple bacterial chromatophore. Light energy absorbed by light-harvesting LH2 and LH1 complexes is channeled to the reaction center (RC), where a charge separation is stabilized as a reduced quinone, and the electron hole is filled by reduced cyt *c*2. Subsequent turnovers of the bc1 complex generate a proton-motive force that drives ATP synthesis. (B) All-atom model of the chromatophore from the purple bacterium *Rba. sphaeroides*. Illustrated using two halves of a transverse section, the final configuration of the all-atom model was obtained after 0.5 ms of explicit solvent all-atom equilibrium MD simulation. The model features 82 bioenergetic complexes (63 light-harvesting LH2 complexes [green], 11 dimeric and 2 monomeric RC-LH1 complexes [LH1, red; RC, blue], 4 bc1 dimers [magenta], and 2 ATP synthases [orange]), together with 4011 light-absorbing antenna molecules (2469 BChls and 1542 carotenoids) embedded in a membrane of approximately 17 200 lipid molecules (lipid phosphate is indicated in yellow). The vesicle is embedded in a rectangular volume of explicit water (shown in SI) and 4624 sodium ions to ensure neutrality. The simulation system consists of 136 833 034 atoms and measures 112 3 112 3 112 nm. Reprinted with permission from Elsevier from ref. 124. Copyright (2015).

requires more than 136 million atoms for an atomistic description. The methods involved in this multiscale initiative are summarized in Fig. 29 (Table 1 from ref. 124); they range from semi-empirical electronic structure theory, to atomistic and coarse-grained molecular dynamics, to umbrella sampling and Brownian Dynamics, with the final step to the phenotype scale being left to a kinetic analysis using the results from the underlying simulation. While one might quibble about any number of details, orchestrating such

a ballet is a daunting challenge and I take my hat off to the late Klaus Schulten for this landmark performance. Bravo Maestro!

7 Machine-learning acceleration

I will end this Perspective with some entry points into the booming literature on machine-learning applications. Anything

Table 1. Table Summarizing All Simulations Employed in the Current Study

Simulation Type	Simulation System	System Size	Number of Runs	Simulation Length	Cumulative Time
MD	chromatophore vesicle	136 M	1	0.5 μ s	0.5 μ s
ZNDO/S-CIS	63 LH2 rings (vesicle)	63 \times 1M	200 snapshots	50 ps	10 ns
ZNDO/S-CIS	1 LH2 rings (flat)	0.2 M	10,000 snapshots	1 ps	10 ns
MARTINI	1 bc ₁ complex	0.06 M	1	40 μ s	40 μ s
BD	500 copies of cyt. c ₂	N/A	6	10 μ s	30 ms
	5,000 quinone molecules	N/A	1	2.8 ms	2.8 ms
Bias exchange umbrella sampling	1 \times bc ₁ ox-c ₂ red	0.5 M	34 windows	100 ns	3.4 μ s
	1 \times bc ₁ red-c ₂ ox	0.5 M	34 windows	100 ns	3.4 μ s
	1 \times RCox-c ₂ red	0.5 M	34 windows	100 ns	3.4 μ s
	1 \times RCred-c ₂ ox	0.5 M	34 windows	100 ns	3.4 μ s
	1 \times bc ₁ ox-c ₂ red (20 mM)	0.5 M	34 windows	100 ns	3.4 μ s

Fig. 29 (Table 1 from ref. 124) Summary of all simulations employed in the study. Reprinted with permission from Elsevier from ref. 124 with permission. COPYRIGHT (2015)

approaching completeness would be a futile endeavor; new applications and new review papers are sprouting up like mushrooms. So, I will be content with a few general reviews and some papers that I found particularly interesting or promising. The focus will be on methods that retain a structural/dynamic model where the role of ML is to accelerate calculations/simulations rather than the, equally meritorious, approach of data base searches, for example in the area of drug design (where we have made some recent contributions^{126–128}). Below, I will follow the bi-thematic approach adopted so far in this Tutorial Review with sights firmly set on complex materials and biological systems.

Like many of us, only two or three years ago I had no experience and only a very little general knowledge of artificial intelligence (AI) or ML, so finding some good textbooks was paramount. Thanks to James Lewis for recommending Murphy's text¹²⁹ on a probabilistic approach to ML and the hands-on ML guide from Géron.¹³⁰

To start a brief parade of reviews, perspective articles or special issues, I choose three on the theme of ML for chemical discovery,^{131–133} followed by a virtual issue on advancing physical chemistry with ML,¹³⁴ a review on ML and the physical sciences,¹³⁵ another on ML meeting chemical physics,¹³⁶ one focusing on quantum chemistry,¹³⁷ two on ML-forcefields^{138,139} and two on molecular simulation.^{140,141} Following up by searching for any of these authors should be richly rewarding.

7.1 Materials

With apologies to the many meritorious authors who are working on ML applications to materials discovery, characterization, and optimization, I will be content with naming a few leaders in the field along with a reference(s) for each: Aspuru-Guzik,¹⁴² Ceriotti,¹⁴³ Csanyi,^{144–146} Lookman,¹⁴⁷ Rinke,¹⁴⁸ Marques,¹⁴⁹ Smit,¹⁵⁰ Bligaard¹⁵¹ and, last but definitely not least, Kulik.¹⁵²

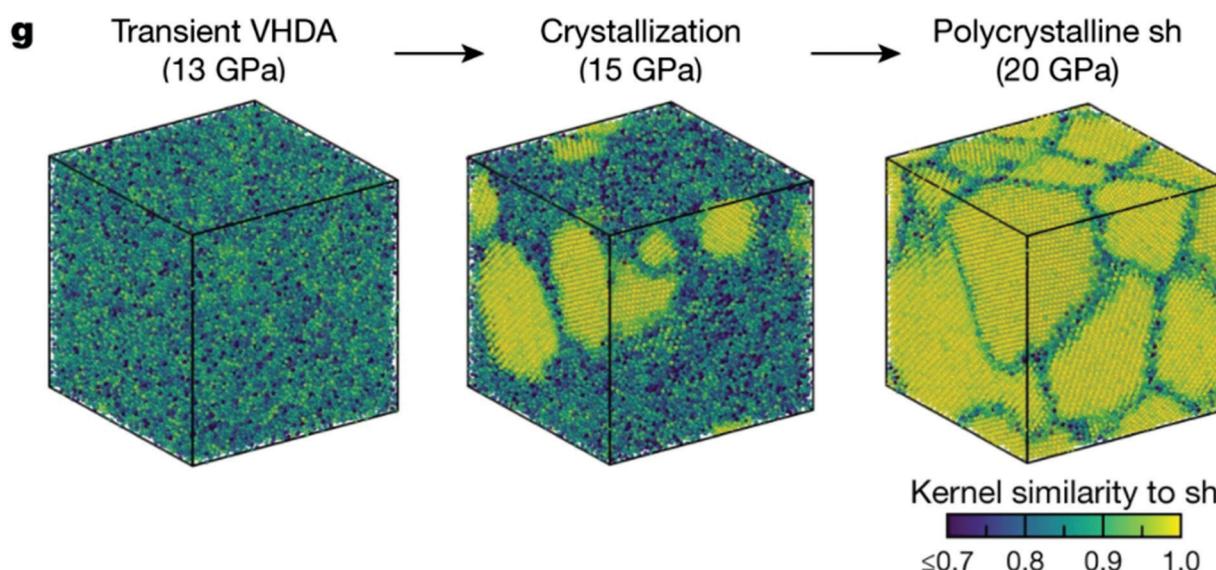
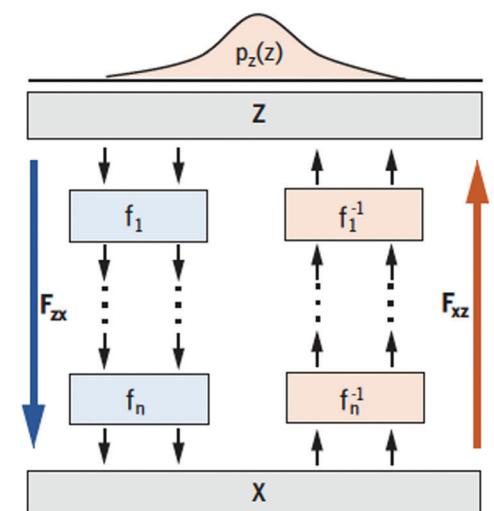


Fig. 30 Details of the crystallization of amorphous Si. Adapted with permission from Nature Springer from ref. 157 COPYRIGHT (2021).

1 Sample Gaussian distribution



2 Generate distribution

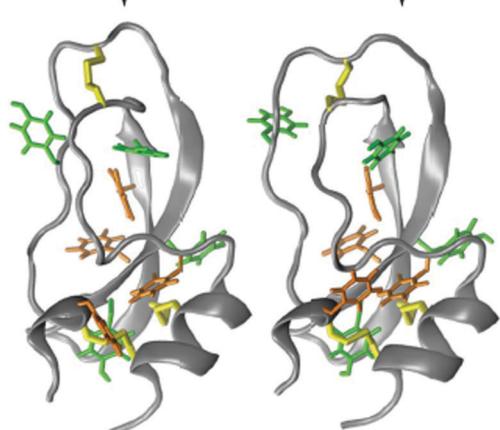
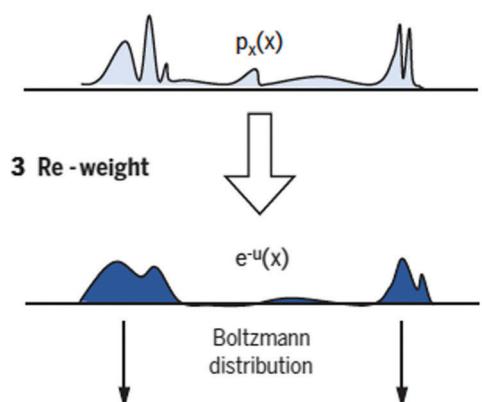


Fig. 31 Boltzmann generators overcome sampling problems between long-lived states. The Boltzmann generator works as follows: 1. We sample from a simple (e.g., Gaussian) distribution. 2. An invertible deep neural network is trained to transform this simple distribution to a distribution $p_x(x)$ that is similar to the desired Boltzmann distribution of the system of interest. 3. To compute thermodynamics quantities, the samples are reweighted to the Boltzmann distribution using statistical mechanics methods. Reproduced from ref. 74 with permission from the American Association for the Advancement of Science, copyright 2019.

Again, forward and backward citation searches should complete the picture with as much detail as desired.

As for our own work, initial forays have been aimed at the global optimization of nanoparticle structures, with a focus on active learning methods, using small amounts of data to begin with and building the dataset progressively guided by Gaussian processes or neural networks.^{153–156} Our plan is to extend this work in the next phase to finding transition states and then to full-fledged ML-accelerated dynamics for free-energy simulations. Although ML is often viewed as “interpolation” amongst sometimes huge datasets, our initial experience with the active learning protocols indicates some success at “extrapolation” combining local searches (exploitation) with more wide-ranging extrapolated exploration of potential energy surfaces. To be continued...

Case study 14: Origins of structural and electronic transitions in disordered silicon

For the ML-materials case study, I choose the recent paper by Deringer *et al.*¹⁵⁷ who have met the challenge of describing phase transitions in disordered materials using machine-learned potentials trained on quantum mechanical calculations. 100 000 atoms (10 nanometers) are followed through liquid-amorphous and amorphous-amorphous phase transitions revealing features that agree with experiment but had been beyond the capability of previous simulations. Fig. 30 shows the progression from 13 GPa to 15, then 20 revealing unprecedented details of the crystallization process. Details of the electronic structure have also been exposed. Potentially, a landmark paper in materials simulation!

7.2 Biomolecular systems

For bio-molecular applications using ML, some of the (rather arbitrarily chosen) parade leaders in the quest for efficient free-energy simulations are: Noé, Tkatchenko, Müller, Clementi,¹⁴⁰ Aspuru-Guzik,¹⁵⁸ Schöberl,⁷¹ Parrinello,⁷⁰ Ferguson,⁷² Tuckerman¹⁵⁹ and Noé (again).^{74,160}

Case study 15: Boltzmann generators: sampling equilibrium states of many-body systems with deep learning

Amongst the plethora of noteworthy applications of ML to dynamics, I choose Noé *et al.*'s⁷⁴ Boltzmann generators for this final case study. As the sketch shown in Fig. 31 shows, Boltzmann generators use an invertible neural network to transform a simple prior distribution to a distribution that resembles the desired Boltzmann distribution. Typically, relatively short simulations (compared with free sampling MD) are all that is needed in each of the relevant catchment regions. Several model examples are given, then an application to a small protein (BPTI). Speedups of many orders of magnitude are projected for even higher dimensional systems. Although the Boltzmann generators do not require pre-defined reaction coordinates, for at least some applications, direct paths in the lower-dimensional latent space that interpolate between two minima do result in back-mapped paths that have low energies and high probabilities, providing good candidates for order parameters/reactive coordinates/collective variables. Required reading for students of ML-accelerated free-energy simulations.

8 Conclusion

The above ramble covers more than six decades of progress towards the multiscale modelling of bio-physico-chemical processes in complex environments. I hope it will provide some perspective on the past and exhilarating hope for the future. I will be especially happy if newcomers to the field find something intriguing that they can latch onto and explore further. It's an exciting time in a still burgeoning field.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

I am grateful to the dozens of students, postdocs and collaborators with whom I have had the great fortune to work over the decades. That this is a solo Tutorial Review reflects only my desire to spare them the burden of critique so that they could concentrate on moving the field forward.

References

- 1 R. S. Mulliken, Presented in part at the American Chemical Society, California section, G. N. Lewis Award Lecture., 1960.
- 2 W. Kohn, *Phys. Rev. B: Solid State*, 1964, **136**, B864–B871.
- 3 W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133–A1138.
- 4 M. Levy, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1982, **26**, 1200–1208.
- 5 R. E. Watson, *Phys. Rev.*, 1958, **111**, 1108–1110.
- 6 K. Schwarz and P. Herzig, *Solid State Commun.*, 1979, **31**, 825–827.
- 7 K. H. Johnson, *J. Chem. Phys.*, 1966, **45**, 3085–3095.
- 8 D. R. Salahub and R. P. Messmer, *Surf. Sci.*, 1981, **106**, 415–421.
- 9 A. Arbužnikov, V. Vasilyev and A. Goursot, *Surf. Sci.*, 1998, **397**, 395–405.
- 10 G. Fischer, A. Goursot, B. Coq, G. Delahay and S. Pal, *Chem. Phys. Chem.*, 2006, **7**, 1795–1801.
- 11 V. D. Dominguez-Soria, P. Calaminici and A. Goursot, *J. Chem. Phys.*, 2007, **127**, 154710.
- 12 V. D. Dominguez-Soria, P. Calaminici and A. Goursot, *J. Phys. Chem. C*, 2011, **115**, 6508–6512.
- 13 L. Onsager, *J. Am. Chem. Soc.*, 1936, **58**, 1486–1493.
- 14 S. Miertus, E. Scrocco and J. Tomasi, *Chem. Phys.*, 1981, **55**, 117–129.
- 15 M. L. Connolly, *Science*, 1983, **221**, 709–713.
- 16 M. L. Connolly, *J. Appl. Crystallogr.*, 1983, **16**, 548–558.
- 17 M. L. Connolly, *J. Mol. Graphics*, 1993, **11**, 139–143.
- 18 J. Tomasi, B. Mennucci and R. Cammi, *Chem. Rev.*, 2005, **105**, 2999–3093.
- 19 A. Klamt and G. Schuurmann, *J. Chem. Soc., Perkin Trans. 2*, 1993, 799–805, DOI: [10.1039/p29930000799](https://doi.org/10.1039/p29930000799).
- 20 A. Klamt, *J. Phys. Chem.*, 1995, **99**, 2224–2235.
- 21 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 4538–4543.
- 22 M. Born, *Z. Phys.*, 1920, **1**, 45–48.
- 23 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378–6396.
- 24 J. Tomasi, B. Mennucci and E. Cancès, *J. Mol. Struct.: THEOCHEM*, 1999, **464**, 211–226.
- 25 X. Sheng and F. Himo, *Angew. Chem., Int. Ed.*, 2020, **59**, 22973–22977, DOI: [10.1002/anie.202008919](https://doi.org/10.1002/anie.202008919).
- 26 E. L. Ratkova, D. S. Palmer and M. V. Fedorov, *Chem. Rev.*, 2015, **115**, 6312–6356.
- 27 A. Ben-Naim, *Molecular Theory of Solutions*, Oxford University Press, New York, 2006.
- 28 J.-L. Barrat and J.-P. Hansen, *Basic concepts for simple and complex liquids*, Cambridge University Press, New York, 2003.
- 29 J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, Elsevier Academic Press, Amsterdam, 4th edn, 2000.
- 30 A. Kovalenko and S. Gusarov, *Phys. Chem. Chem. Phys.*, 2018, **20**, 2947–2969.
- 31 L. S. Ornstein and F. Zernike, *Proc. K. Ned. Akad. Wet.*, 1914, **17**, 793–806.
- 32 Molecular theory of solvation, Kluwer Academic Publishers, Dordrecht, 2003.
- 33 A. Kovalenko and F. Hirata, *J. Chem. Phys.*, 1999, **110**, 10095–10112.
- 34 D. Chandler and H. C. Andersen, *J. Chem. Phys.*, 1972, **57**, 1930–1937.
- 35 D. Beglov and B. Roux, *J. Chem. Phys.*, 1996, **104**, 8678–8689.
- 36 D. Beglov and B. Roux, *J. Phys. Chem. B*, 1997, **101**, 7821–7826.
- 37 S. Gusarov, T. Ziegler and A. Kovalenko, *J. Phys. Chem. A*, 2006, **110**, 6083–6090.
- 38 M. Malvaldi, S. Bruzzone, C. Chiappe, S. Gusarov and A. Kovalenko, *J. Phys. Chem. B*, 2009, **113**, 3536–3542.
- 39 G. Jeanmairet, M. Levesque and D. Borgis, *J. Chem. Theory Comput.*, 2020, **16**, 7123–7134.
- 40 N. D. Mermin, *Phys. Rev.*, 1965, **137**, A1441–A1443.
- 41 R. Evans, *Adv. Phys.*, 1979, **28**, 143–200.
- 42 W. S. B. Dwandaru and M. Schmidt, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2011, **83**, 061133.
- 43 T. A. Wesolowski, S. Shedge and X. W. Zhou, *Chem. Rev.*, 2015, **115**, 5891–5928.
- 44 T. Wesolowski, in *Multiscale Dynamics Simulations: Nano- and Nano-bio Systems in Complex Environments*, ed. D. Salahub and D. Q. Wei, Royal Society of Chemistry, London, 2021, ch. Hohenberg-Kohn theorems as a basis for multi-scale simulations: Frozen-Density Embedding Theory.
- 45 T. A. Wesolowski and A. Warshel, *J. Phys. Chem.*, 1993, **97**, 8050–8053.
- 46 T. Wesolowski and A. Warshel, *J. Phys. Chem.*, 1994, **98**, 5183–5187.
- 47 T. A. Wesolowski and J. Weber, *Chem. Phys. Lett.*, 1996, **248**, 71–76.

- 48 S. V. Shedge and T. A. Wesolowski, *Chem. Phys. Chem.*, 2014, **15**, 3291–3300.
- 49 L. H. Thomas, *Proc. Cambridge Philos. Soc.*, 1927, **23**, 542–548.
- 50 E. Fermi, *Rend. Accad. Naz. Lincei*, 1927, **6**, 602–607.
- 51 J. W. Kaminski, S. Gusarov, T. A. Wesolowski and A. Kovalenko, *J. Phys. Chem. A*, 2010, **114**, 6082–6096.
- 52 A. Warshel and M. Karplus, *J. Am. Chem. Soc.*, 1972, **94**, 5612–5625.
- 53 A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227–249.
- 54 H. M. Senn and W. Thiel, *Top. Curr. Chem.*, 2007, **268**, 173–290.
- 55 H. Hu and W. T. Yang, *Annu. Rev. Phys. Chem.*, 2008, **59**, 573–601.
- 56 H. M. Senn and W. Thiel, *Angew. Chem., Int. Ed.*, 2009, **48**, 1198–1229.
- 57 R. Zhang, B. Lev, J. E. Cuervo, S. Y. Noskov and D. R. Salahub, *Adv. Quantum Chem.*, 2010, **59**, 353–400.
- 58 D. R. Salahub, A. de la Lande, A. Goursot, R. Zhang and Y. Zhang, *Struct. Bonding*, 2013, **150**, 1–64.
- 59 U. Ryde, *Methods Enzymol.*, 2016, **577**, 119–158.
- 60 L. W. Chung, H. Hirao, X. Li and K. Morokuma, *WIREs Comput. Mol. Sci.*, 2012, **2**, 327–350.
- 61 Y. Zhou, S. Wang, Y. Li and Y. Zhang, *Methods Enzymol.*, 2016, **577**, 105–118.
- 62 S. F. Sousa, A. J. M. Ribeiro, R. P. P. Neves, N. F. Bras, N. M. F. S. A. Cerqueira, P. A. Fernandes and M. J. Ramos, *WIREs Comput. Mol. Sci.*, 2017, **7**, e1281.
- 63 D. R. Salahub, S. Y. Noskov, B. Lev, R. Zhang, V. Ngo, A. Goursot, P. Calaminici, A. M. Koster, A. Alvarez-Ibarra, D. Mejia-Rodriguez, J. Rezac, F. Cailliez and A. de la Lande, *Molecules*, 2015, **20**, 4780–4812.
- 64 A. de la Lande, A. Alvarez-Ibarra, K. Hasnaoui, F. Cailliez, X. J. Wu, T. Mineva, J. Cuny, P. Calaminici, L. Lopez-Sosa, G. Geudtner, I. Navizet, C. G. Irieipa, D. R. Salahub and A. M. Koster, *Molecules*, 2019, **24**, 1653.
- 65 S. Ahmadi, L. B. Herrera, M. Chehelamirani, J. Hostas, S. Jalife and D. R. Salahub, *Int. J. Quantum Chem.*, 2018, **118**, e25558.
- 66 Y. I. Yang, Q. Shao, J. Zhang, L. J. Yang and Y. Q. Gao, *J. Chem. Phys.*, 2019, **151**, 070902.
- 67 G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187–199.
- 68 A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12562–12566.
- 69 A. Barducci, G. Bussi and M. Parrinello, *Phys. Rev. Lett.*, 2008, **100**, 020603.
- 70 L. Bonati, V. Rizzi and M. Parrinello, *J. Phys. Chem. Lett.*, 2020, **11**, 2998–3004.
- 71 M. Schoberl, N. Zabaras and P. S. Koutsourelakis, *J. Chem. Phys.*, 2019, **150**, 024109.
- 72 W. Chen, A. R. Tan and A. L. Ferguson, *J. Chem. Phys.*, 2018, **149**, 072312.
- 73 C. Wehmeyer and F. Noe, *J. Chem. Phys.*, 2018, **148**, 241703.
- 74 F. Noe, S. Olsson, J. Kohler and H. Wu, *Science*, 2019, **365**, eaaw1147.
- 75 A. Rodriguez, M. d'Errico, E. Facco and A. Laio, *J. Chem. Theory Comput.*, 2018, **14**, 1206–1215.
- 76 Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.*, 1999, **314**, 141–151.
- 77 R. C. Bernardi, M. C. R. Melo and K. Schulten, *Biochim. Biophys. Acta, Gen. Subj.*, 2015, **1850**, 872–877.
- 78 L. Shen and W. T. Yang, *J. Chem. Theory Comput.*, 2018, **14**, 1442–1455.
- 79 L. Boselt, M. Thurleemann and S. Riniker, *J. Chem. Theory Comput.*, 2021, **17**, 2641–2658.
- 80 X. C. Liu and D. R. Salahub, *J. Am. Chem. Soc.*, 2015, **137**, 4249–4259.
- 81 P. Drude, *Ann. Phys.*, 1900, **1**, 566–613.
- 82 Z. Y. Lu and Y. K. Zhang, *J. Chem. Theory Comput.*, 2008, **4**, 1237–1248.
- 83 H. B. Yu, T. W. Whitfield, E. Harder, G. Lamoureux, I. Vorobyov, V. M. Anisimov, A. D. MacKerell and B. Roux, *J. Chem. Theory Comput.*, 2010, **6**, 774–786.
- 84 V. Ngo, M. C. da Silva, M. Kubillus, H. Li, B. Roux, M. Elstner, Q. Cui, D. R. Salahub and S. Y. Noskov, *J. Chem. Theory Comput.*, 2015, **11**, 4992–5001.
- 85 H. Li, V. Ngo, M. C. Da Siva, D. R. Salahub, K. Callahan, B. Roux and S. Y. Noskov, *J. Phys. Chem. B*, 2015, **119**, 9401–9416.
- 86 J. A. Lemkul, J. Huang, B. Roux and A. D. MacKerell, *Chem. Rev.*, 2016, **116**, 4983–5013.
- 87 A. Ganguly, E. Boulanger and W. Thiel, *J. Chem. Theory Comput.*, 2017, **13**, 2954–2961.
- 88 G. Konig, F. C. Pickard, J. Huang, W. Thiel, A. D. MacKerell, B. R. Brooks and D. M. York, *Molecules*, 2018, **23**, 2695.
- 89 E. Flood, C. Boiteux, B. Lev, I. Vorobyov and T. W. Allen, *Chem. Rev.*, 2019, **119**, 7737–7832.
- 90 K. S. Amin, X. Hu, D. R. Salahub, C. Baldauf, C. Lim and S. Noskov, *J. Chem. Phys.*, 2020, **153**, 144102.
- 91 F. Y. Lin, J. Huang, P. Pandey, C. Rupakheti, J. Li, B. T. Roux and A. D. MacKerell, Jr., *J. Chem. Theory Comput.*, 2020, **16**, 3221–3239.
- 92 V. S. Inakollu, D. P. Geerke, C. N. Rowley and H. Yu, *Curr. Opin. Struct. Biol.*, 2020, **61**, 182–190.
- 93 B. T. Thole, *Chem. Phys.*, 1981, **59**, 341–350.
- 94 G. Lamoureux and B. Roux, *J. Chem. Phys.*, 2003, **119**, 3025–3039.
- 95 D. Loco, L. Lagardere, O. Adjoua and J. P. Piquemal, *Acc. Chem. Res.*, 2021, **54**, 2812–2822.
- 96 P. Y. Ren and J. W. Ponder, *J. Phys. Chem. B*, 2003, **107**, 5933–5947.
- 97 A. Grossfield, P. Y. Ren and J. W. Ponder, *J. Am. Chem. Soc.*, 2003, **125**, 15671–15682.
- 98 J. W. Ponder, C. J. Wu, P. Y. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson and T. Head-Gordon, *J. Phys. Chem. B*, 2010, **114**, 2549–2564.
- 99 D. Loco, E. Polack, S. Caprasecca, L. Lagardere, F. Lipparini, J. P. Piquemal and B. Mennucci, *J. Chem. Theory Comput.*, 2016, **12**, 3654–3661.

- 100 X. J. Wu, J. M. Teuler, F. Cailliez, C. Clavaguera, D. R. Salahub and A. de la Lande, *J. Chem. Theory Comput.*, 2017, **13**, 3985–4002.
- 101 G. Donati, A. Wildman, S. Caprasecca, D. B. Lingerfelt, F. Lipparini, B. Mennucci and X. S. Li, *J. Phys. Chem. Lett.*, 2017, **8**, 5283–5289.
- 102 A. Alvarez-Ibarra, A. M. Koster, R. Zhang and D. R. Salahub, *J. Chem. Theory Comput.*, 2012, **8**, 4232–4238.
- 103 N. T. Maitra, *J. Phys.: Condens. Matter*, 2017, **29**, 423001.
- 104 W. Magnus, *Commun. Pur. Appl. Math.*, 1954, **7**, 649–673.
- 105 X. J. Wu, A. Alvarez-Ibarra, D. R. Salahub and A. de la Lande, *Eur. Phys. J. D*, 2018, **72**, 206.
- 106 A. de la Lande, S. Denisov and M. Mostafavi, *Phys. Chem. Chem. Phys.*, 2021, **23**, 21148–21162.
- 107 K. A. Omar, K. Hasnaoui and A. de la Lande, *Annu. Rev. Phys. Chem.*, 2021, **72**, 445–465.
- 108 S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. de Vries, *J. Phys. Chem. B*, 2007, **111**, 7812–7824.
- 109 P. C. T. Souza, R. Alessandri, J. Barnoud, S. Thallmair, I. Faustino, F. Grunewald, I. Patmanidis, H. Abdizadeh, B. M. H. Bruininks, T. A. Wassenaar, P. C. Kroon, J. Melcr, V. Nieto, V. Corradi, H. M. Khan, J. Domanski, M. Javanainen, H. Martinez-Seara, N. Reuter, R. B. Best, I. Vattulainen, L. Monticelli, X. Periole, D. P. Tieleman, A. H. de Vries and S. J. Marrink, *Nat. Methods*, 2021, **18**, 382–388.
- 110 N. Liguori, R. Croce, S. J. Marrink and S. Thallmair, *Photosynth. Res.*, 2020, **144**, 273–295.
- 111 G. S. Ayton, E. Lyman and G. A. Voth, *Faraday Discuss.*, 2010, **144**, 347–357.
- 112 L. L. Olenick, J. M. Troiano, A. Vartanian, E. S. Melby, A. C. Mensch, L. L. Zhang, J. W. Hong, O. Mesele, T. Qiu, J. Bozich, S. Lohse, X. Zhang, T. R. Kuech, A. Millevolte, I. Gunsolus, A. C. McGeachy, M. Dogangun, T. Z. Li, D. H. Hu, S. R. Walter, A. Mohaimani, A. Schmoldt, M. D. Torelli, K. R. Hurley, J. Dalluge, G. Chong, Z. V. Feng, C. L. Haynes, R. J. Hamers, J. A. Pedersen, Q. Cui, R. Hernandez, R. Klaper, G. Orr, C. J. Murphy and F. M. Geiger, *Chem.*, 2018, **4**, 2709–2723.
- 113 C. Lamprakis, I. Andreadelis, J. Manchester, C. Velez-Vega, J. S. Duca and Z. Cournia, *J. Chem. Theory Comput.*, 2021, **17**, 3088–3102.
- 114 H. I. Ingolfsson, C. A. Lopez, J. J. Uusitalo, D. H. de Jong, S. M. Gopal, X. Periole and S. J. Marrink, *WIREs Comput. Mol. Sci.*, 2014, **4**, 225–248.
- 115 A. Yu, A. J. Pak, P. He, V. Monje-Galvan, L. Casalino, Z. Gaieb, A. C. Dommer, R. E. Amaro and G. A. Voth, *Biophys. J.*, 2021, **120**, 1097–1104.
- 116 Z. Jarin, J. Newhouse and G. A. Voth, *J. Chem. Theory Comput.*, 2021, **17**, 1170–1180.
- 117 A. V. Mironenko and G. A. Voth, *J. Chem. Theory Comput.*, 2020, **16**, 6329–6342.
- 118 J. Hutter, M. Iannuzzi, F. Schiffmann and J. VandeVondele, *WIREs Comput. Mol. Sci.*, 2014, **4**, 15–25.
- 119 E. van der Giessen, P. A. Schultz, N. Bertin, V. V. Bulatov, W. Cai, G. Csanyi, S. M. Foiles, M. G. D. Geers, C. Gonzalez, M. Hutter, W. K. Kim, D. M. Kochmann, J. LLorca, A. E. Mattsson, J. Rottler, A. Shluger, R. B. Sills, I. Steinbach, A. Strachan and E. B. Tadmor, *Model. Numer. Simul. Mater. Sci.*, 2020, **28**, 043001.
- 120 W. K. Kim and E. B. Tadmor, *Philos. Mag.*, 2017, **97**, 2284–2316.
- 121 D. E. Shaw, *Biophys. J.*, 2013, **104**, 45a–45a.
- 122 R. E. Amaro and A. J. Mulholland, *Nat. Rev. Chem.*, 2018, **2**, 0148.
- 123 M. Feig and Y. Sugita, *Annu. Rev. Cell Dev. Biol.*, 2019, **35**, 191–211.
- 124 A. Singhary, C. Maffeo, K. H. Delgado-Magnero, D. J. K. Swainsbury, M. Sener, U. Kleinekathofer, J. W. Vant, J. Nguyen, A. Hitchcock, B. Isralewitz, I. Teo, D. E. Chandler, J. E. Stone, J. C. Phillips, T. V. Pogorelov, M. I. Mallus, C. Chipot, Z. Luthey-Schulten, D. P. Tieleman, C. N. Hunter, E. Tajkhorshid, A. Aksimentiev and K. Schulten, *Cell*, 2019, **179**, 1098–1111.
- 125 M. Feig, R. Harada, T. Mori, I. Yu, K. Takahashi and Y. Sugita, *J Mol Graph Model*, 2015, **58**, 1–9.
- 126 Y. Chu, X. Wang, Q. Dai, Y. Wang, Q. Wang, S. Peng, X. Wei, J. Qiu, D. R. Salahub, Y. Xiong and D. Q. Wei, *Briefings Bioinf.*, 2021, **22**, bbab165.
- 127 Y. Chu, X. Shan, T. Chen, M. Jiang, Y. Wang, Q. Wang, D. R. Salahub, Y. Xiong and D. Q. Wei, *Briefings Bioinf.*, 2021, **22**, bbaa205.
- 128 Y. Chu, A. C. Kaushik, X. Wang, W. Wang, Y. Zhang, X. Shan, D. R. Salahub, Y. Xiong and D. Q. Wei, *Briefings Bioinf.*, 2021, **22**, 451–462.
- 129 K. P. Murphy, *Machine Learning - a probabilistic perspective*, The MIT Press, Cambridge, Massachusetts, 2012.
- 130 A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* O'Reilly, Sebastopol, CA, USA, 2021.
- 131 M. Rupp, O. A. von Lilienfeld and K. Burke, *J. Chem. Phys.*, 2018, **148**, 241401.
- 132 O. A. von Lilienfeld and K. Burke, *Nat. Commun.*, 2020, **11**, 4895.
- 133 A. Tkatchenko, *Nat. Commun.*, 2020, **11**, 4125.
- 134 O. V. Prezhdo, *J. Phys. Chem. Lett.*, 2020, **11**, 9656–9658.
- 135 G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto and L. Zdeborova, *Rev. Mod. Phys.*, 2019, **91**, 045002.
- 136 M. Ceriotti, C. Clementi and O. A. von Lilienfeld, *J. Chem. Phys.*, 2021, **154**, 160401.
- 137 P. O. Dral, *J. Phys. Chem. Lett.*, 2020, **11**, 2336–2347.
- 138 J. Behler, *Chem. Rev.*, 2021, **121**, 10037–10072.
- 139 O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltaysky, K. T. Schutt, A. Tkatchenko and K. R. Muller, *Chem. Rev.*, 2021, **121**, 10142–10186.
- 140 F. Noe, A. Tkatchenko, K. R. Muller and C. Clementi, *Annu. Rev. Phys. Chem.*, 2020, **71**, 361–390.
- 141 A. Glielmo, B. E. Husic, A. Rodriguez, C. Clementi, F. Noe and A. Laio, *Chem. Rev.*, 2021, **121**, 9722–9758.
- 142 A. Aspuru-Guzik, R. Lindh and M. Reiher, *ACS Cent. Sci.*, 2018, **4**, 144–152.
- 143 M. Ceriotti, *J. Chem. Phys.*, 2019, **150**, 150901.
- 144 V. L. Deringer, M. A. Caro and G. Csanyi, *Adv. Mater.*, 2019, **31**, 1902765.

- 145 V. L. Deringer, A. P. Bartok, N. Bernstein, D. M. Wilkins, M. Ceriotti and G. Csanyi, *Chem. Rev.*, 2021, **121**, 10073–10141.
- 146 F. Musil, A. Grisafi, A. P. Bartok, C. Ortner, G. Csanyi and M. Ceriotti, *Chem. Rev.*, 2021, **121**, 9759–9815.
- 147 T. Lookman, P. V. Balachandran, D. Z. Xue and R. H. Yuan, *npj Comput. Mater.*, 2019, **5**, 21.
- 148 L. Himanen, A. Geurts, A. S. Foster and P. Rinke, *Adv. Sci.*, 2019, **6**, 1902765.
- 149 J. Schmidt, M. R. G. Marques, S. Botti and M. A. L. Marques, *npj Comput. Mater.*, 2019, **5**, 83.
- 150 S. M. Moosavi, K. M. Jablonka and B. Smit, *J. Am. Chem. Soc.*, 2020, **142**, 20273–20287.
- 151 P. S. Lamoureux, K. T. Winther, J. A. G. Torres, V. Streibel, M. Zhao, M. Bajdich, F. Abild-Pedersen and T. Bligaard, *ChemCatChem*, 2019, **11**, 3579–3599.
- 152 C. R. Duan, F. Liu, A. Nandy and H. J. Kulik, *J. Phys. Chem. Lett.*, 2021, **12**, 4628–4637.
- 153 B. R. L. Galvao, L. P. Viegas, D. R. Salahub and M. P. Lourenco, *J. Mol. Model.*, 2020, **26**, 303.
- 154 J. Hostas, A. Tchagang, M. P. Lourenco, A. M. Koster and D. R. Salahub, *Theor. Chem. Acc.*, 2021, **140**, 44.
- 155 M. P. Lourenco, B. R. L. Galvao, L. B. Herrera, J. Hostas, A. Tchagang, M. X. Silva and D. R. Salahub, *Theor. Chem. Acc.*, 2021, **140**, 62.
- 156 M. P. Lourenco, L. B. Herrera, J. Hostas, P. Calaminici, A. M. Koster, A. Tchagang and D. R. Salahub, *Theor. Chem. Acc.*, 2021, **140**, 116.
- 157 V. L. Deringer, N. Bernstein, G. Csanyi, C. Ben Mahmoud, M. Ceriotti, M. Wilson, D. A. Drabold and S. R. Elliott, *Nature*, 2021, **589**, 59–64.
- 158 F. Hase, I. F. Galvan, A. Aspuru-Guzik, R. Lindh and M. Vacher, *Chem. Sci.*, 2019, **10**, 2298–2307.
- 159 J. Rogal, E. Schneider and M. E. Tuckerman, *Phys. Rev. Lett.*, 2019, **123**, 245701.
- 160 F. Noe and E. Rosta, *J. Chem. Phys.*, 2019, **151**, 190401.