









# NAMD goes quantum: an integrative suite for hybrid simulations

Marcelo C R Melo<sup>1,2,13</sup> , Rafael C Bernardi<sup>1,13</sup> , Till Rudack<sup>1,3</sup> , Maximilian Scheurer<sup>4,5</sup> , Christoph Riplinger<sup>6</sup>, James C Phillips<sup>1</sup> , Julio D C Maia<sup>7</sup> , Gerd B Rocha<sup>8</sup> , João V Ribeiro<sup>1</sup>, John E Stone<sup>1</sup>, Frank Neese<sup>9</sup>, Klaus Schulten<sup>1,10,12</sup> & Zaida Luthey-Schulten<sup>1,2,10,11</sup> 

**Hybrid methods that combine quantum mechanics (QM) and molecular mechanics (MM) can be applied to studies of reaction mechanisms in locations ranging from active sites of small enzymes to multiple sites in large bioenergetic complexes. By combining the widely used molecular dynamics and visualization programs NAMD and VMD with the quantum chemistry packages ORCA and MOPAC, we created an integrated, comprehensive, customizable, and easy-to-use suite (<http://www.ks.uiuc.edu/Research/qmmm>). Through the QwikMD interface, setup, execution, visualization, and analysis are streamlined for all levels of expertise.**

Though MM force fields are based on quantum mechanical calculations and experimental observations, only QM allows a complete and accurate understanding of many biochemical processes, particularly those that involve chemical reactions or charge redistribution<sup>1</sup>. However, even with the advanced hardware technology available today, the computational cost of studying nanosecond-long dynamics of entire systems solely via QM methodologies is usually prohibitive. A common way to circumvent this cost barrier is to confine the QM formalism to a subregion of a system and include the effects of the surrounding system through MM simulations, an approach referred to as a hybrid QM–MM simulation<sup>2</sup>. QM–MM calculations are used broadly in enzymology<sup>3</sup>, drug discovery<sup>4</sup>, and bioenergetic systems<sup>5</sup>, as well as in combination with serial femtosecond crystallography<sup>6</sup> and other material sciences and structural biology techniques<sup>7</sup>. To deliver accurate results, QM–MM studies require a carefully selected quantum region<sup>5,8</sup>. Unfortunately, most available QM–MM implementations lack a

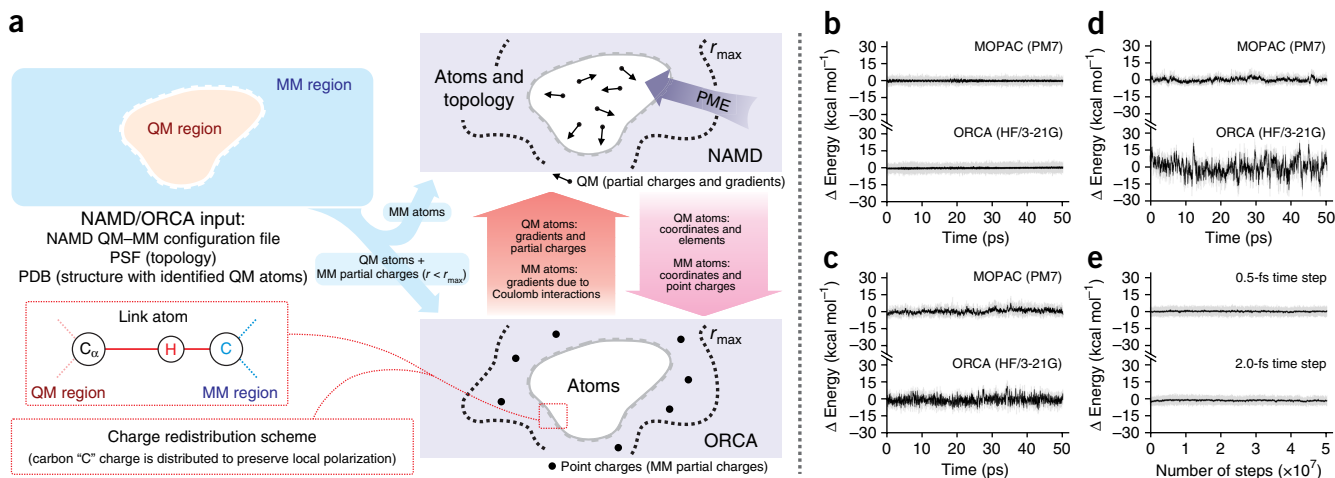
comprehensive, customizable, easy-to-use set of features to make such tools broadly attractive to chemists, structural biologists, and material engineers.

NAMD<sup>9</sup> is a widely used software package for molecular dynamics (MD) simulations, particularly for large biomolecular systems, using supercomputers. NAMD's scalability and large array of enhanced sampling and free energy methods<sup>10</sup>, as well as its seamless integration with VMD<sup>11</sup>, which provides extended setup, visualization, and analysis capabilities, make it an invaluable tool for exploring complex biological systems. Here we present a comprehensive QM–MM suite implemented in NAMD, to provide a broad range of QM–MM methods, and in VMD, for easy setup, visualization, and analysis through the graphical user interface QwikMD<sup>12</sup>. In NAMD, the QM–MM interface supports the simulation of many independent QM regions and smooth integration with a collection of enhanced sampling and alchemical methods, including the combination of QM–MM approaches with state-of-the-art free energy methods, such as extended adaptive biasing force (eABF)<sup>13</sup>. NAMD performs similarly to other QM–MM programs in terms of energy conservation (**Supplementary Table 1**) and allows scalable free energy calculations (**Supplementary Fig. 1**).

In hybrid QM–MM simulations, NAMD offloads part of its standard force and energy calculations to a QM program, either through native interfaces to MOPAC<sup>14,15</sup> or ORCA<sup>16</sup>, or through a flexible generic interface. Other programs, such as Amber<sup>17</sup>, additionally include their own code for QM calculations. We chose not to adopt this strategy here to ensure that users would have the flexibility to easily fine-tune their simulations. Importantly, when file input/output is done in RAM, very little time is lost during communication between NAMD and the QM program. In NAMD, multiple QM–MM coupling schemes have been implemented, which allows both mechanically and electrostatically embedded QM regions to be used (Online Methods). Input files are the same as those used for classical MD, with additional options in the configuration file. Typically, QM and MM atoms that are covalently bound are treated by redistribution of the MM atom's charge over its nearest MM neighbors and by capping of the QM atom with a hydrogen atom, known as the link atom method, as shown in **Figure 1a** and **Supplementary Figure 2** for a calculation using the NAMD–ORCA interface. Link atom variations are described in the Online Methods and in **Supplementary Figures 3a,b, 4, and 5**.

To test the accuracy, stability, and performance of the QM–MM interface, we carried out standard validation simulations with both

<sup>1</sup>NIH Center for Macromolecular Modeling and Bioinformatics, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana–Champaign, Urbana, Illinois, USA. <sup>2</sup>Center for Biophysics and Computational Biology, University of Illinois at Urbana–Champaign, Urbana, Illinois, USA. <sup>3</sup>Department of Biophysics, Ruhr–University Bochum, Bochum, Germany. <sup>4</sup>Biochemistry Center, Heidelberg University, Heidelberg, Germany. <sup>5</sup>Interdisciplinary Center for Scientific Computing, Heidelberg, Germany. <sup>6</sup>FAccTs GmbH, Köln, Germany. <sup>7</sup>Center for Informatics, Federal University of Paraíba, João Pessoa, Brazil. <sup>8</sup>Department of Chemistry, Federal University of Paraíba, João Pessoa, Brazil. <sup>9</sup>Max-Planck-Institut für Kohlenforschung, Mülheim an der Ruhr, Germany. <sup>10</sup>Department of Physics, University of Illinois at Urbana–Champaign, Urbana, Illinois, USA. <sup>11</sup>Department of Chemistry, University of Illinois at Urbana–Champaign, Urbana, Illinois, USA. <sup>12</sup>Deceased. <sup>13</sup>These authors contributed equally to this work. Correspondence should be addressed to Z.L.-S. ([zan@illinois.edu](mailto:zan@illinois.edu)).



**Figure 1** | Hybrid QM-MM simulations in NAMD. **(a)** Schematic of NAMD-ORCA interconnection. The electrostatic embedding, point charge modifications, and charge redistribution methods allow for precise and smooth QM-MM coupling. The contributions of MM charges beyond  $r_{\text{max}}$  are calculated by NAMD (via the particle mesh Ewald (PME) method), and ORCA calculates direct electrostatics. PSF, protein structure file; PDB, Protein Data Bank. **(b-d)** Energy conservation tests were performed during MD simulations of **(b)** a pure QM alanine molecule, **(c)** a QM-MM trialanine molecule, and **(d)** a QM-MM trialanine molecule in water. The basis set used in ORCA tests (3-21G) served only as a technical test. **(e)** Long-scale energy conservation tests, using different time steps, for a pure-QM *N*-methyl acetamide molecule calculated with NAMD-MOPAC at the PM7 level. All energy conservation tests were done with the NVE ensemble. The plot shows the deviation from the mean system energy. Black lines indicate the running average, and gray shading indicates maximal fluctuation. Each line in **b-d** represents  $n = 1$  QM-MM MD simulation.

MOPAC and ORCA. With the NVE ensemble, NAMD achieved energy conservation for a pure QM alanine molecule (Fig. 1b), a hybrid QM-MM trialanine molecule in vacuum (Fig. 1c) and in water (Fig. 1d), and a pure QM *N*-methyl acetamide molecule, the last of which showed energy conservation for up to 100 ns with both 0.5-fs and 2.0-fs time steps (Fig. 1e). We observed that a particle mesh Ewald treatment of long-range electrostatics also conserved energy (Supplementary Table 2). NAMD and Amber16 presented equally good QM-MM results (Supplementary Fig. 6). NAMD-MOPAC performed up to 10 ns/d of QM-MM simulation on a single desktop computer (Supplementary Table 3).

To provide a simple way of setting up QM-MM simulations, we incorporated the most widely used features of a hybrid QM-MM MD code into VMD's QwikMD<sup>12</sup>, which enabled us to create a comprehensive QM-MM workflow (Fig. 2a). QwikMD automates the creation of input and configuration files while checking for common mistakes, and ensures reproducibility of the result. Moreover, QwikMD allows drugs, metabolites, and other molecules that lack MM parameters to be easily added to a QM region. Once prepared, QM-MM simulations can be directly performed either in real time using 'live-view' mode or by calling NAMD using local computer resources or a supercomputing center (Supplementary Fig. 7).

Trajectories from classical or hybrid QM-MM simulations are easily read and analyzed by VMD, whose capabilities were extended in this work to natively support MOPAC and ORCA outputs. We also introduced new representation schemes to VMD to allow for the selection and visualization of orbital trajectories throughout a simulation, and for dynamic updating of chemical bond representations. These capabilities make VMD a powerful tool for visualization and analysis of output from QM-MM calculations, as well as of outputs directly from ORCA and MOPAC. Molecular orbitals can be rendered for both pure QM and QM-MM simulations, in systems ranging from small

molecules to macromolecular complexes in their environment (Fig. 2b-d, Supplementary Tables 4-6, Supplementary Note 1, and Supplementary Video 1).

By combining the new QM-MM capabilities of the NAMD-MOPAC interface with existing tools for enhanced sampling and free energy calculations, we investigated the aminoacylation reaction mechanism of *Thermus thermophilus* glutamyl-tRNA synthetase (GluRS) and its interactions with its cognate tRNA (tRNA<sup>Glu</sup>), using PM7 together with the CHARMM36 force field. To establish the genetic code, GluRS reads the tRNA's anticodon region and uses this information to rearrange its catalytic site, thus facilitating the transfer of an AMP-bound glutamyl to the 3' end of the tRNA. A study of the allosteric information processing and transduction was previously conducted using classical MD<sup>18</sup>. However, because the anticodon binding domain and catalytic site of GluRS are >50 Å apart, treatment of the entire pathway, with approximately 5,000 atoms, by QM would not be feasible. By using QwikMD to prepare a system with two independent QM regions, we were able to investigate the allosteric signaling pathway of the GluRS-tRNA<sup>Glu</sup> complex with accurate QM treatment of the two most critical regions of the system (Fig. 3a).

We analyzed the communication pathways that lead to coordinated motion between functionally important QM regions using cross-correlation-based network analysis<sup>18</sup> (Online Methods and Supplementary Fig. 8) and found them to be notably degenerate, as previously observed with classical MD<sup>18</sup> (Fig. 3a). The QM-MM treatment of critical regions led to increased correlation between residues in the active site and in the anticodon binding site (Supplementary Fig. 9). Suboptimal communication pathways are based on the correlation of atom motion during the simulation and have been used to calculate allosteric signaling and force-propagation pathways. Tightly correlated groups of atoms are clustered into communities, indicating functional domains of biomolecules and important interfaces between multi-molecule

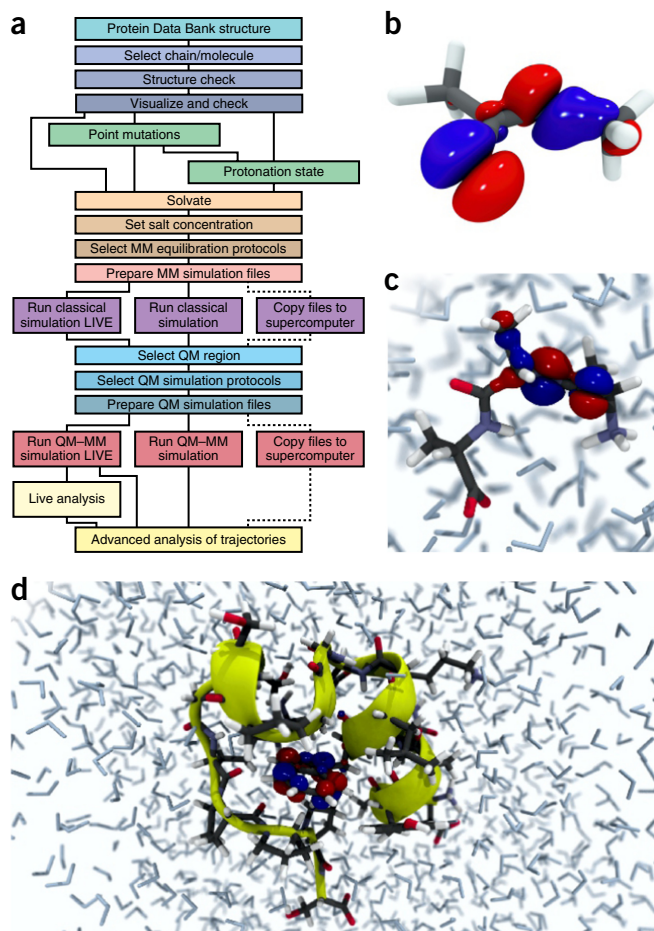
complexes. The presence of multiple communities within a QM region and the occurrence of communities that cover both classical and quantum atoms highlight the seamless integration of QM and MM regions (Fig. 3b).

We were able to distinguish between four possible aminoacylation reaction mechanisms (Supplementary Fig. 10), all based on a previously proposed concerted exchange in which the 3'- or 2'-OH oxygen attacks the carbonyl carbon of Glu-AMP while the proton shifts to the phosphate's oxygen<sup>19</sup>. We also investigated the reaction while the amine group of the adenylate was protonated or deprotonated. We carried out QM-MM steered MD simulations to induce the possible reaction mechanisms, and selected intermediate states from the steered MD trajectories to initiate a string-method optimization<sup>20</sup>. This method uses groups of biased MD simulations to find the path with the smallest barrier that describes the chemical reaction. Once an optimized string had been achieved, we used the reaction path to perform a parallel eABF<sup>13</sup> calculation to determine the free energy transformation of each possible mechanism (Supplementary Figs. 11 and 12). We observed that the reaction charging the 3-hydroxyl group, with the amine group of the adenylate being deprotonated, was the most favorable route, with a barrier of  $\sim 20$  kcal mol<sup>-1</sup> and final state at  $\sim 8$  kcal mol<sup>-1</sup> (Fig. 3c-e).

Initially, all possible reaction mechanisms tested were observed to be endergonic (Supplementary Fig. 13). However, closer examination of the final states revealed that the AMP's phosphate was only 3 Å away from the charged tRNA. Dissociation of the AMP-H from the charged tRNA opened space for hydration of its phosphate group (Supplementary Fig. 14), and was investigated via adaptive biasing force (ABF) and classical MD simulations. As the products moved apart and water entered the active site, hydrating the phosphate group of AMP-H, we observed an  $\sim 22$  kcal mol<sup>-1</sup> decrease in free energy that made the entire reaction-solvation process exergonic, with free energy variation of approximately  $-15$  kcal mol<sup>-1</sup> between reactants and products (Fig. 3e,f). In the final state of the eABF calculation, the distance between Glu-AMP:Carb-C and Glu-AMP:P-O was 3 Å, and we used this distance as the connecting point (red dashed line in Fig. 3e) between the free energy profiles calculated with eABF and ABF. It is worth noting that appending QM-MM eABF and MM ABF results led to a small amount of imprecision in the connection between free energy values.

Building upon the synergy between NAMD and VMD, we developed a robust and user-friendly QM-MM suite to prepare, perform, and analyze QM-MM simulations. Our test applications highlight the accuracy of this implementation, and our study of the GluRS:tRNA<sup>Glu</sup>:Glu-AMP complex revealed subatomic details of its reaction mechanism. By uniting network analysis results for the full complex with the first combination of the string method, parallel eABF, and QM-MM simulations, we were able to provide a unique view that models the essential steps in establishing the genetic code (Supplementary Fig. 15).

The modularity in NAMD will enable this suite to be used in a variety of contexts, including as a teaching tool, a research interface, a platform for mixing and matching QM-MM with free energy and enhanced sampling methods, and even a sandbox for the development of new QM tools and QM-MM interaction schemes. Applications for this suite include the intelligent design of enzymatic inhibitors, as well as efforts to improve the precision of gene editing technologies.



**Figure 2 | Hybrid QM-MM VMD features.** (a) QwikMD provides a graphical user interface in VMD for performing QM-MM simulations. The workflow illustrates how to prepare, run, analyze, and visualize a hybrid QM-MM MD simulation. (b) The highest occupied molecular orbital of an alanine molecule in vacuum. (c) Alanine's highest occupied molecular orbital in a solvated QM-MM trialanine. (d) Trp-cage protein's highest occupied molecular orbital in water solution. Color-coding in b–d is as follows: gray, carbon; red, oxygen; blue, nitrogen; white, hydrogen. Orbitals are in glossy red and blue for clarity. Yellow ribbon in d represents secondary structure.

## METHODS

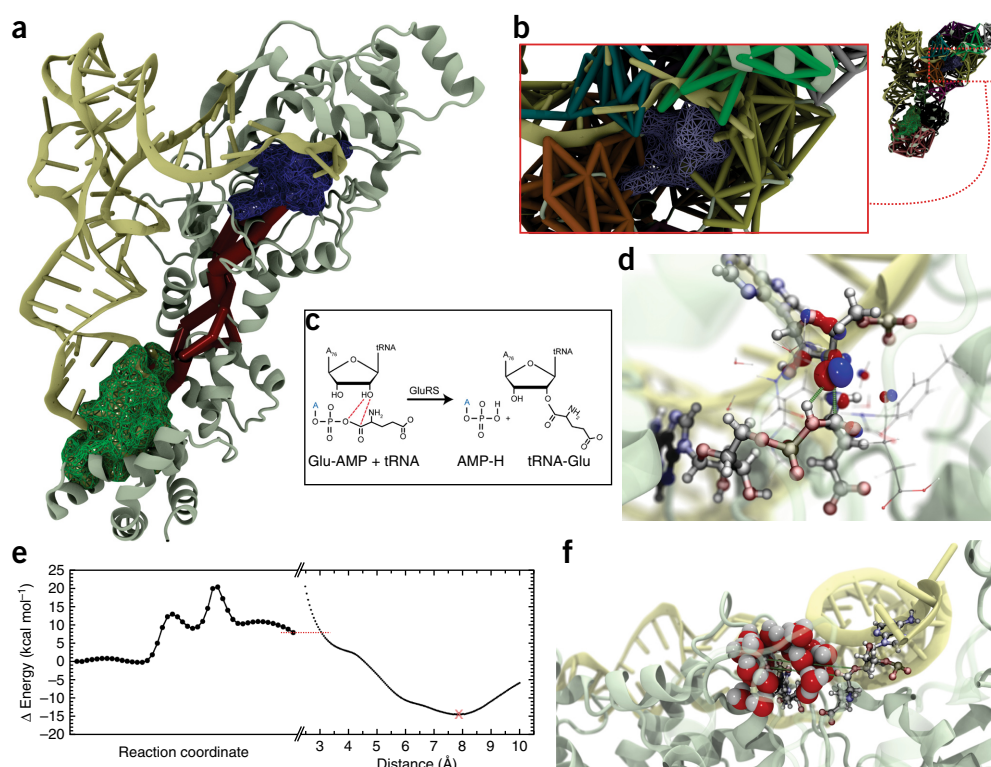
Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

## ACKNOWLEDGMENTS

The authors thank M.F. Herbst, C. Chipot, and G. Fiorin for helpful discussions. This work was supported by the National Science Foundation (NSF) (grants MCB-1616590, MCB-1244570, and PHY1430124 to Z.L.-S.), the US National Institutes of Health (NIH) (grant P41-GM104601 to Z.L.-S.), the Keck Foundation (grant 206231 to M.C.R.M. and Z.L.-S.), the Alexander von Humboldt Foundation (Feodor Lynen Postdoctoral Fellowship to T.R.), the Brazilian Coordination for Improvement of Higher Educational Personnel (CAPES; fellowship to J.D.C.M.; grant AUXPE1375/2014 to G.B.R.), and the Brazilian National Council for Scientific and Technological Development (CNPq 305271/2013-0 to G.B.R.). F.N. and C.R. acknowledge support for the development of ORCA by the Max Planck society (MPG) and the Germans Science Foundation (DFG). This research made use of Blue Waters sustained-petascale computing, which is supported by the state of Illinois and the NSF (OCI-0725070 and ACI-1238993). This work is





**Figure 3** | Mechanism of glutamyl-tRNA synthetase. **(a)** The glutamyl-tRNA synthetase allosteric pathway (red). Two independent QM regions are highlighted, indicating the active site (blue) and the anticodon binding region (green). Yellow and white ribbons represent tRNA and synthetase, respectively. **(b)** Cross-correlation-based community analysis of the active site calculated from QM-MM trajectories. Color-coded as in **a**. **(c)** Reaction mechanism of glutamyl-tRNA synthetase. **(d)** Intermediate state of the glutamyl-tRNA synthetase reaction, showing the highest occupied molecular orbital. **(e)** Left, free energy profile of the most favorable glutamyl-tRNA synthetase reaction mechanism, calculated by eABF after a string-method path optimization. Both eABF and the string method were carried out with QM-MM MD simulations with NAMD-MOPAC and PM7. Right, free energy profile of the distancing and solvation of AMP calculated via ABF and classical MD simulations. “X” indicates the minimum energy state. The red dashed line indicates the connecting point between the eABF and ABF free energy profiles.  $n = 1$  experiment each for eABF and ABF. **(f)** Snapshot of the minimum energy state (red X in **e**) during the release of the AMP, showing the solvation of the phosphate group. Color-coding in **d,f** defined as in **Figure 2**, except for gold element structures (phosphatase) and white and yellow ribbons (as in **a**).

part of the Petascale Computational Resource (PRAC) grant, which is supported by the NSF (ACI-1713784).

#### AUTHOR CONTRIBUTIONS

M.C.R.M., R.C.B., T.R., K.S., and Z.L.-S. conceived the project. M.C.R.M. implemented the QM-MM interface. R.C.B., T.R., M.C.R.M., G.B.R., and K.S. discussed QM-MM features. J.D.C.M., G.B.R., C.R., and F.N. provided guidance on the development of the QM-MM interface. J.C.P. assisted in adapting NAMD. M.C.R.M. and M.S. prepared Python scripts for the interfaces of selected QM software packages. R.C.B. and M.C.R.M. performed all NAMD tests and simulations. J.D.C.M. and G.B.R. performed Amber calculations. R.C.B. and M.C.R.M. performed all free-energy calculations and analysis. M.S. and J.E.S. implemented the orbital visualization in VMD. J.V.R. and J.E.S. implemented the QM-MM graphical interface in QwikMD. R.C.B., M.C.R.M., T.R., M.S., G.B.R., F.N., and Z.L.-S. wrote and edited the manuscript. K.S. and Z.L.-S. supervised the project.

#### COMPETING INTERESTS

The authors declare no competing interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Senn, H.M. & Thiel, W. *Angew. Chem. Int. Edn Engl.* **48**, 1198–1229 (2009).

2. Field, M.J., Bash, P.A. & Karplus, M. *J. Comput. Chem.* **11**, 700–733 (1990).
3. van der Kamp, M.W. & Mulholland, A.J. *Biochemistry* **52**, 2708–2728 (2013).
4. Bernardi, R.C. & Pascutti, P.G. *J. Chem. Theory Comput.* **8**, 2197–2203 (2012).
5. Retegan, M., Neese, F. & Pantazis, D.A. *J. Chem. Theory Comput.* **9**, 3832–3842 (2013).
6. Coquelle, N. *et al. Nat. Chem.* **10**, 31–37 (2018).
7. Lin, H. & Truhlar, D.G. *Theor. Chem. Acc.* **117**, 185–199 (2007).
8. Jindal, G. & Warshel, A. *J. Phys. Chem. B* **120**, 9913–9921 (2016).
9. Phillips, J.C. *et al. J. Comput. Chem.* **26**, 1781–1802 (2005).
10. Jiang, W. *et al. Comput. Phys. Commun.* **185**, 908–916 (2014).
11. Humphrey, W., Dalke, A. & Schulten, K. *J. Mol. Graph.* **14**, 33–38 (1996).
12. Ribeiro, J.V. *et al. Sci. Rep.* **6**, 26536 (2016).
13. Fu, H., Shao, X., Chipot, C. & Cai, W. *J. Chem. Theory Comput.* **12**, 3506–3513 (2016).
14. Stewart, J.J. *J. Comput. Aided Mol. Des.* **4**, 1–105 (1990).
15. Maia, J.D.C. *et al. J. Chem. Theory Comput.* **8**, 3072–3081 (2012).
16. Neese, F. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **8**, e1327 (2018).
17. Götz, A.W., Clark, M.A. & Walker, R.C. *J. Comput. Chem.* **35**, 95–108 (2014).
18. Sethi, A., Eargle, J., Black, A.A. & Luthey-Schulten, Z. *Proc. Natl. Acad. Sci. USA* **106**, 6620–6625 (2009).
19. Black Pyrkosz, A., Eargle, J., Sethi, A. & Luthey-Schulten, Z. *J. Mol. Biol.* **397**, 1350–1371 (2010).
20. Pan, A.C., Sezer, D. & Roux, B. *J. Phys. Chem. B* **112**, 3432–3440 (2008).

## ONLINE METHODS

**Method overview.** In order to study the chemistry of complex biological systems, in particular by MD approaches, one needs to combine QM and MM methods to secure subatomic resolution within relevant time and length scales. However, recent advances in implementations of QM calculations and hardware improvements are continuously increasing the capacity to expand the complexity and comprehensiveness of QM results<sup>21</sup>. Hybrid QM–MM simulations in NAMD divide the system into MM and QM regions, using a classical force field to treat the classical atoms ('MM atoms'), and passing the information that describes the quantum atoms in the system ('QM atoms') to a quantum chemistry package, which is expected to calculate forces for all QM atoms, as well as the total energy of the QM region (and optionally partial charges). All bonded and nonbonded interactions among MM atoms are handled by NAMD's CHARMM force field, whereas all interactions among QM atoms are handled by the quantum chemistry package in its chosen theory level.

The nonbonded interactions between QM and MM atoms can be modified and regulated by the user. Lennard–Jones interactions are always calculated by NAMD, and specific parameters can be provided for QM atoms. QM-specific modifications have been proposed to compensate for the overpolarization that these atoms sometimes exhibit in hybrid QM–MM simulations, although the importance of such modifications has been disputed<sup>22</sup>. In all simulations presented in this work, the standard CHARMM36 Lennard–Jones parameters were used for all MM and QM atoms.

**Mechanical and electrostatic embedding.** It has long been known that electrostatics, particularly polarization, plays a key role in many biochemical processes<sup>23</sup>. Electrostatic interactions between QM and MM atoms deserve a more detailed discussion owing to the abundance and diversity of available alternatives. In the 'mechanical embedding' scheme, no electrostatic influences from MM atoms are accounted for while the QM package computes forces, charge distribution, and energy in the QM region. Only positions and elements of atoms in the QM region are passed on to the QM package, and QM and MM atoms interact only through NAMD-calculated Lennard–Jones and electrostatic potentials (see below for special treatment of QM–MM bonds).

In the 'electrostatic embedding' scheme, in contrast, the partial charges of MM atoms surrounding all QM atoms are used to approximate the electrostatic environment of the QM atoms. The selection of classical point charges can be done automatically by NAMD, in which case a cutoff value is used to effectively create a shell of point charges around the QM region. This type of embedding is the most frequently used in biomolecular simulations<sup>1</sup>.

Classical point charges provided to the QM program to be used in electrostatic embedding can be altered by NAMD in a variety of ways. First, a smoothing function can be applied to avoid abrupt changes in electrostatic forces due to the cutoff used in the selection of surrounding point charges. Second, partial charges can be further modified so that their sum is a whole number or a charge complementary to that of the QM region, in which case the sum of charges from QM atoms and classical partial charges will be zero. In the latter case, the user is also able to select, atom by atom, the classical partial charges used to build the electrostatic embedding, which provides further flexibility to the application.

Regardless of the chosen embedding method, the calculated charge distribution for QM atoms can be used by NAMD to update the partial charges of QM atoms for the calculation of short- and long-range electrostatic interactions. In particular, if particle mesh Ewald (PME)<sup>24</sup> is being used, NAMD can apply the newly determined charges for QM atoms to the calculation of long-range electrostatics in both QM and MM regions of the system. In this case, the necessary corrections are calculated so as to subtract from the PME forces all interactions already calculated by the QM package between QM–QM atom pairs, and by NAMD's QM module between QM–point charge pairs.

PME forces and energy calculations in NAMD are carried out using classical charges for all classical atoms and the user's choice of either classical charges or updated charges for QM atoms. The forces and energy derived from PME interactions between QM atoms and between QM atoms and the surrounding point charges are recalculated within the QM module and subtracted from the direct electrostatic calculations, to avoid double-counting of these contributions. To test the effects of such implementation on the simulations, we carried out an energy-conservation analysis to compare the use of PME in NAMD when the QM calculations were done with the PM3 (refs. 25,26), HF-3C (Hartee-Fock-3C)<sup>27</sup>, and DFT (density functional theory) methods (**Supplementary Table 2**).

### Treatment of covalent bonds involving QM and MM atoms.

Hybrid QM–MM simulations of biomolecular systems often present situations in which only a part of a molecule should be treated quantum-mechanically, usually to save computational resources, as the cost of simulating QM regions increases rapidly with the number of atoms. To deal with chemical bonds that have one atom in the quantum mechanical region and another in the classical (MM) region (hereinafter referred to as QM–MM bonds), NAMD offers several methods that can be combined to alter the molecular system in order to bridge differences in simulation type (QM versus MM) and minimize errors involved in the QM/MM division of the system. Regardless of the methods selected to treat the QM–MM bond, the bonded term (between MM1 and QM1 atoms; **Supplementary Fig. 3c**) will still be calculated by NAMD, along with all proper and improper dihedral terms and angle terms that cross the QM–MM barrier but still involve at least one classical atom.

**Link atoms.** The most widely used method to cap QM regions containing QM–MM bonds is the link atom approach<sup>28</sup>, whereby an atom (usually a hydrogen atom) is placed along the bond between the QM atom and the MM atom (**Supplementary Fig. 3c**) and does not exist in the classical simulation. The user can fine-tune this process by choosing the element and method of link-atom placement to be used: either fixed (the default) or variable. The former depends on a user-defined distance  $d_{L-QM}$ , which is used throughout the simulation (**Supplementary Fig. 3c**). For the latter, a user-defined fraction is used to define  $C_L$  (equation (2)), and this fixed value is used by NAMD to calculate  $d_{L-QM}$  at every step as a function of  $d_{MM-QM}$ , which in turn varies over the course of the simulation because of bond vibration.

To conserve force and energy, the total force acting on the link atom ( $\vec{F}_L$ ), as calculated by the QM code, is redistributed by NAMD over the QM1 and MM1 atoms<sup>29</sup> and added to the total

forces calculated on those atoms by the QM code and NAMD, respectively. Equations (1) and (3) describe the calculation of the  $x$  component of the redistributed force applied on the QM1 and MM1 atoms, respectively. Analogous equations are used for the other two axes. Here,  $\hat{i}_x$  is the unit vector on the  $x$ -axis,  $\hat{r}_{\text{MM-QM}}$  is the unit vector in the QM–MM bond direction, and  $x_{\text{MM}}$  and  $x_{\text{QM}}$  are the  $x$  components of the positions of the MM1 and QM1 atoms, respectively.

$$F'_{\text{QM1}_x} = \vec{F}_L \cdot \left[ (1 - C_L) \cdot \hat{i}_x + C_L \cdot \frac{x_{\text{MM}} - x_{\text{QM}}}{d_{\text{MM-QM}}} \cdot \hat{r}_{\text{MM-QM}} \right] \quad (1)$$

where

$$C_L = \frac{d_{\text{L-QM}}}{d_{\text{MM-QM}}} \quad (2)$$

Similarly,

$$F'_{\text{MM1}_x} = \vec{F}_L \cdot \left[ C_L \cdot \hat{i}_x - C_L \cdot \frac{x_{\text{MM}} - x_{\text{QM}}}{d_{\text{MM-QM}}} \cdot \hat{r}_{\text{MM-QM}} \right] \quad (3)$$

The link atom approach is not the only proposed method for handling QM–MM bonds, but it is the only one that relies entirely on the classical side of a QM–MM simulation, over which the user has complete control. Therefore, regardless of the QM package chosen to carry out the QM calculations of the simulation, NAMD's QM–MM interface always guarantees the creation of proper conditions to simulate QM–MM bonds, which grants it great flexibility.

**Point charge alteration and redistribution.** In any system containing a QM–MM bond, the link atom will invariably be placed very near the MM1 atom, which creates very strong electrostatic repulsion forces (or attractions) in case the MM1 partial charge is sent to the QM package for its calculations. Under the mechanical embedding scheme, the QM package receives only the atoms in the QM region and the link atoms created to approximate QM–MM bonds, so no manipulation of partial charges is required. In contrast, the more usual electrostatic embedding scheme requires special treatment of nearby classical partial charges.

Several methods have been proposed to handle the conditioning of classical partial charges surrounding a QM–MM bond, and the QM–MM interface developed here offers the most widely accepted ones (**Supplementary Fig. 4**). In all methods implemented here, the classical atom participating in the QM–MM bond (MM1 atom) does not have its partial charge passed on to the QM package, as this would create excessive repulsion (or attraction) on the link atom. This is, in fact, the entirety of the 'Z1' method: ignoring the partial charge of the MM1 atom<sup>30</sup>. Analogously, Z2 ignores the MM1 partial charge and all partial charges of atoms bound to MM1, called MM2 atoms, and Z3 extends the approach by ignoring all partial charges of atoms bound to MM2 atoms, called MM3 atoms (**Supplementary Fig. 4b–d**).

The charge-shifting method<sup>31</sup> (**Supplementary Fig. 4e**) is more elaborate, as it rearranges the partial charge of the MM1 atom (denoted here by  $q_{\text{M1}}$ ) so that the total charge of the region is maintained, while approximating the dipole moments of the bonds between MM1 and MM2 atoms. This is done through the creation of 'virtual' point charges that are passed to the QM package as if they represent partial charges of classical

atoms. In this case, the MM1 partial charge is equally distributed across the MM2 atoms, with the placement of the charge  $q_1$  at the position of MM2 atoms (where  $q_1 = q_{\text{M1}}/2$ ), and the coefficient used in equations (4) and (5) to redistribute the force on this virtual charge is given by  $C_1 = d_{\text{M1-Q1}}/d_{\text{M1-M2}} = 1$ . Moreover, two virtual point charges are placed along the direction of the MM1–MM2 bond, one before the MM2 atom ( $q_+ = q_0$ ) and one after ( $q_- = -1 \cdot q_0$ ). In our implementation,  $C_+ = d_{\text{M1-}q_+}/d_{\text{M1-M2}} = 0.94$ , and  $C_- = d_{\text{M1-}q_-}/d_{\text{M1-M2}} = 1.06$  (**Supplementary Fig. 4e**). This method will keep the total charge of the region constant while trying to preserve the local dipoles formed by all MM1–MM2 bonds.

The 'redistributed charge and dipole' method<sup>30</sup> (**Supplementary Fig. 4f**) follows a similar arrangement. A virtual point charge is created in the middle of all MM1–MM2 bonds ( $C_2 = d_{\text{M1-Q2}}/d_{\text{M1-M2}} = 0.5$ ) with a charge  $q_2 = 2 \cdot q_{\text{M1}}/2$ , and a charge  $q_1 = -1 \cdot q_{\text{M1}}/2$  is placed in the positions of all MM2 atoms ( $C_1 = 1$ ). This arrangement maintains a constant total charge of the region while attempting to mimic local dipoles.

In all cases where a virtual charge is created, a force is calculated on it to balance its electrostatic interactions with QM atoms. Similarly as for the link atom, the total force acting on, for example, the virtual point charge  $q_+$  ( $\vec{F}_{q_+}$ ) is redistributed over the MM1 and MM2 atoms that were used to define it. The fraction of the total force that is applied on the MM1 and MM2 atoms is defined, respectively, by

$$\vec{F}_{\text{MM1}} = (1 - C_+) \cdot \vec{F}_{q_+} \quad (4)$$

and

$$\vec{F}_{\text{MM2}} = C_+ \cdot \vec{F}_{q_+} \quad (5)$$

**Implementation and interaction with quantum chemistry packages.** The implementation of the QM–MM module in NAMD was done entirely in C/C++, making use of Charm++ message-driven load-balancing tools so that simulations can be run in parallel over one or several computer nodes. All quantum mechanical calculations are offloaded to a QM package, either through hard-coded interfaces to ORCA<sup>16</sup> or MOPAC<sup>14,15</sup>, or through a standardized interface that uses Python scripts (or any other tool) to wrap and convert input and output formats between NAMD and any other arbitrary QM package. A standardized interface was created for data input/output so that external wrapper scripts can be called, thus providing a translation layer for input and output files from any other QM package. We provide Python scripts that wrap Gaussian<sup>32</sup>, TeraChem<sup>33</sup>, and Q-CHEM<sup>34</sup>, which also serve as templates for the development of new wrapper scripts.

In general, positions and elements of QM atoms are passed on to the QM package along with positions and magnitudes of partial charges representing the local MM environment. In return, NAMD expects forces, total energy, and partial charges for QM atoms, and possibly forces acting on MM partial charges due to electrostatic interactions. The exchange of information between NAMD and quantum chemistry packages is preferably done through files written to RAM, which takes an insignificant amount of time (milliseconds) compared with the time scale of QM calculations (seconds or more). By not embedding the QM code directly into NAMD, we allow users to choose their preferred quantum chemistry package and level of theory.

Additionally, NAMD takes advantage of the current advanced state of quantum chemistry packages, which are prepared to be



sequentially called by an external software, such as in a QM-MM context. After a self-consistent field (SCF) calculation on a given set of atoms, a ‘state-file’ containing the result of the SCF procedure is saved. A much faster convergence is achieved in the subsequent SCF calculation when the QM package is re-initialized using the saved state-files, as it is already initialized with a very close ‘guess’ of what the SCF solution should be. This occurs because atom positions vary only slightly between consecutive calculations.

Multiple QM regions can be simulated through simultaneous and independent executions of the chosen QM package, one per independent QM region defined in the biomolecular system.

**MD simulations of test systems and ttGluRS.** Structures for the test systems and for *T. thermophilus* GluRS (ttGluRS) were prepared for classical and QM-MM MD simulations using VMD’s<sup>11</sup> QwikMD<sup>12</sup> graphical interface. While QwikMD assists users in selecting QM regions residue by residue through a point-and-click interface, other methods have been developed to provide automated QM-region selection<sup>35</sup>. The structure of the pre-transfer complex ttGluRS:tRNA<sup>Glu</sup>:Glu-AMP had been previously solved by means of X-ray crystallography at 2.1-Å resolution and is available at the Protein Data Bank (PDB 1N78)<sup>36</sup>, with the replacement of the inert analog for the active Glu-AMP<sup>19</sup>. The simulations in the present study were done with the NAMD MD package<sup>9</sup>, MOPAC 2016 (refs. 14,15), PM7 (ref. 37), and ORCA 4.0 (ref. 16). The CHARMM36 force field<sup>38</sup> and the TIP3P water model, for solvated tests, were used to describe all systems. All calculations involving the ttGluRS:tRNA<sup>Glu</sup>:Glu-AMP system were done in the NpT ensemble. In energy-conservation tests, the simulations used either the PM3 (refs. 25,26) method (using NAMD’s ORCA interface) or RM1 (ref. 39) (using NAMD’s MOPAC interface). Results from Amber used its SQM implementation for both PM3 and RM1.

For solvated systems, the simulations were carried out under periodic boundary conditions. Simulations were carried out in many steps to ensure a reasonable starting conformation for QM-based simulations. (1) First, for systems explicitly solvated, we used classical MD in the NpT ensemble with temperature maintained at 300 K using Langevin dynamics for both pressure, kept at 1 bar, and temperature coupling. (2) Using the same parameters from classical NpT simulations, we carried out QM-based NpT simulations (either pure QM or hybrid QM-MM). (3) For energy-conservation tests, QM-based simulations were performed in the NVE ensemble.

A distance cutoff of 12.0 Å or 17.0 Å (eABF calculations) was applied to short-range, nonbonded interactions, whereas long-range electrostatic interactions were treated using the PME<sup>24</sup> method. For equilibration and classical MM simulations, the equations of motion were integrated using the r-RESPA multiple-time-step scheme<sup>9</sup>. For all production QM-MM simulations, the van der Waals interactions and electrostatic interactions were updated at every time step. The time step of integration was either 0.5 fs or 2.0 fs (details below).

The MonoAlanine system is composed of a single alanine residue in vacuum (gas phase). In the experiments described in **Figure 1b**, the simulations for verification of energy conservation were carried out as described above with a 0.5-fs time step.

The PolyAla system is composed of three alanine residues. The same system was used in different ways, either in vacuum or in solution (TIP3 water), with either all three residues treated quantum-mechanically, or just the central residue treated as a QM region, leading to the use of two link atoms (**Supplementary Fig. 3a,b**). **Figure 1c** depicts the system used in vacuum, with the middle residue being treated quantum-mechanically. **Figure 1e** shows the system used in solution, with all three residues treated quantum-mechanically and the water molecules treated classically, as also shown in **Supplementary Figure 2a** for the NAMD-ORCA interface. Simulations were performed with 0.5-fs time steps, and the charge-shift treatment was used for link atoms. To reproduce previously reported energy-conservation results in Amber, we used the SPC/Fw water model (a flexible variant of the simple point charge (SPC) model) in all Amber QM-MM simulations.

Analogously, the *N*-methyl acetamide system was simulated either in vacuum or in solution (TIP3 water), but always entirely treated with QM formalism. **Figure 1d** shows the system used in vacuum (gas phase), with the entire molecule being treated quantum-mechanically. Simulations were performed with both 0.5-fs and 2.0-fs time steps, as shown in **Figure 1d**.

**Network analysis.** Suboptimal communication pathways and communities were calculated on the basis of Pearson cross-correlation of atom motion, using a protocol similar to the one described in ref. 18. Our calculations were based on ten QM-MM trajectories that used 2-fs time steps for a total of 100 ps, totaling 1 ns of QM-MM simulations. For this system two independent QM regions were used, one in the anticodon domain with approximately 500 QM atoms, and one in the catalytic region with approximately 200 QM atoms. Briefly, the network analysis protocol uses central atoms ( $\alpha$ -carbons in amino acids, N1/N9 in nucleic bases, and the sugar phosphate P) called nodes as a proxy for the motion of residues, and their positions throughout MD simulations are used to calculate the Pearson cross-correlation of motion. Only nodes from residues whose non-hydrogen atoms are less than 4.5 Å apart for more than 75% of the total simulated time are considered connected<sup>40</sup>. Nodes that do not meet this cutoff have their cross-correlation multiplied by zero through a mask applied to the cross-correlation matrix. The resulting set of correlations between connected nodes forms a weighted matrix that is used by the Floyd-Warshall algorithm to find the shortest communication paths between nodes (‘suboptimal’ paths), and by the Girvan-Newman algorithm to find optimal communities between highly interacting nodes. The suboptimal paths represent the shortest set of connected nodes that could transmit information from two distant sites, such as an allosteric regulator site and an enzyme’s active site.

**QM-MM simulations.** To study the reaction mechanism of the ttGluRS:tRNA<sup>Glu</sup>:Glu-AMP system, we carried out a 200-ps QM-MM simulation with a 2-fs time step, followed by a 50-ps QM-MM simulation with a 0.5-fs time step. These two simulations were used to equilibrate the system in a conformation that was favorable for the reaction mechanism to be investigated. We tested four mechanisms by inducing the reaction steps to occur (**Supplementary Fig. 10**). The four biased simulations were performed with a 0.5-fs time step for 20–100 ps. All simulations were carried out using the charge-shift method to treat link atoms,

and a ‘shift’ function was applied to surrounding classical partial charges in the electrostatic embedding scheme.

**String method and eABF.** To study a transformation that occurs in a biomolecular system, such as a chemical reaction or a conformational change, one defines collective variables such as distances between atoms or between centers of mass of groups of atoms, or even angles between subdomains of a molecular structure. The collective variables (colvars) are used to track the changes in the system as it undergoes the transformation being studied, and can be used to define a reaction coordinate.

The string method is an iterative process that optimizes a reaction coordinate in order to find the path of least resistance from the initial to the final state of the system. The method uses a discretized representation of the reaction coordinate composed of ‘images’, where each image is a copy of the entire simulated system at a different stage of the transformation. At each iteration, multiple independent MD simulations are initiated from each image, which allows the systems to explore their energy surface and drift toward local minima. Then, average values for the colvars are determined, and biases are applied to keep consecutive images approximately equidistant in colvar space and to smooth the reaction coordinate. The iterations proceed until consecutive calculations do not produce significant changes in the mean colvar values, which indicates convergence of the method.

Once the string has been optimized, the images are used to define a continuous path that defines the transformation, and eABF is used to calculate free energy changes. Using the images that were optimized via the string method, new path colvars *S* and *Z* are created, which constrain the dynamics of the system so that it follows the chosen reaction coordinate. *S* indicates progression along the path, and *Z* indicates a perpendicular distance to the path. Taking advantage of NAMD’s outstanding scalability and NCSA’s Blue Waters supercomputer, we used a parallel strategy that initiated multiple walkers from the different images, and were thus able to conduct extensive sampling over the defined path.

**Code availability.** The reported QM–MM features are publicly available in NAMD 2.12 or later and in VMD 1.9.4 or later. More details are available at <http://www.ks.uiuc.edu/Research/qmmm>, as well as source code and a tutorial. ORCA is available at <https://orcaforum.cec.mpg.de> and <https://www.faccts.de>.

**Life Sciences Reporting Summary.** Further information on experimental design is available in the **Life Sciences Reporting Summary**.

**Data availability.** The data that support the findings of this study are available from the corresponding author upon reasonable request. Source data for **Figures 1** and **3** and for **Supplementary Figures 1, 5, 6, 8, 9, and 12–14** are available online.

21. Riplinger, C., Pinski, P., Becker, U., Valeev, E.F. & Neese, F. *J. Chem. Phys.* **144**, 024109 (2016).
22. Riccardi, D., Li, G. & Cui, Q. *J. Phys. Chem. B* **108**, 6467–6478 (2004).
23. Warshel, A. & Levitt, M. *J. Mol. Biol.* **103**, 227–249 (1976).
24. Darden, T., York, D. & Pedersen, L. *J. Chem. Phys.* **98**, 10089–10092 (1993).
25. Stewart, J.J.P. *J. Comput. Chem.* **10**, 209–220 (1989).
26. Stewart, J.J.P. *J. Comput. Chem.* **10**, 221–264 (1989).
27. Sure, R. & Grimme, S. *J. Comput. Chem.* **34**, 1672–1685 (2013).
28. Singh, U.C. & Kollman, P.A. *J. Comput. Chem.* **7**, 718–730 (1986).
29. Walker, R.C., Crowley, M.F. & Case, D.A. *J. Comput. Chem.* **29**, 1019–1031 (2008).
30. Lin, H. & Truhlar, D.G. *J. Phys. Chem. A* **109**, 3991–4004 (2005).
31. Sherwood, P. *et al. Faraday Discuss.* **106**, 79–92 (1997).
32. Frisch, M.J. *et al. Gaussian 09, Revision A.02.* (Gaussian, Inc., 2016).
33. Titov, A.V., Ufimtsev, I.S., Luehr, N. & Martinez, T.J. *J. Chem. Theory Comput.* **9**, 213–221 (2013).
34. Shao, Y. *et al. Mol. Phys.* **113**, 184–215 (2015).
35. Kulik, H.J., Zhang, J., Klinman, J.P. & Martínez, T.J. *J. Phys. Chem. B* **120**, 11381–11394 (2016).
36. Sekine, S. *et al. EMBO J.* **22**, 676–688 (2003).
37. Stewart, J.J.P. *J. Mol. Model.* **19**, 1–32 (2013).
38. Best, R.B. *et al. J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
39. Dutra, J.D.L., Filho, M.A.M., Rocha, G.B., Simas, A.M. & Freire, R.O. *PLoS One* **10**, e0124372 (2015).
40. Schoeler, C. *et al. Nano Lett.* **15**, 7370–7376 (2015).



## Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study. For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

### ► Experimental design

#### 1. Sample size

Describe how sample size was determined.

Not applicable. No significant variability is expected when employing our software with the same input files.

#### 2. Data exclusions

Describe any data exclusions.

No data was excluded.

#### 3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

Not applicable. No significant variability is expected when employing our software with the same input files.

#### 4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Not applicable. All experiments were done in silico.

#### 5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Not applicable. All experiments were done in silico.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

#### 6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Confirmed  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The <u>exact sample size</u> ( <i>n</i> ) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement indicating how many times each experiment was replicated   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The statistical test(s) used and whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>                              |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as an adjustment for multiple comparisons   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Test values indicating whether an effect is present<br><i>Provide confidence intervals or give results of significance tests (e.g. <i>P</i> values) as exact values whenever appropriate and with effect sizes noted.</i> |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A clear description of statistics including <u>central tendency</u> (e.g. median, mean) and <u>variation</u> (e.g. standard deviation, interquartile range)  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clearly defined error bars in <u>all</u> relevant figure captions (with explicit mention of central tendency and variation)   |

See the web collection on [statistics for biologists](#) for further resources and guidance.

## ► Software

Policy information about [availability of computer code](#)

### 7. Software

Describe the software used to analyze the data in this study.

All computational tools employed to analyze data are described in the text. Plots and running averages were made using either the Grace or R software package.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* [guidance for providing algorithms and software for publication](#) provides further information on this topic.

## ► Materials and reagents

Policy information about [availability of materials](#)

### 8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

No unique materials were used.

### 9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

Not applicable. All experiments were done in silico. No antibodies were used in this study.

### 10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

Not applicable. All experiments were done in silico. No eukaryotic cells were used in this study.

b. Describe the method of cell line authentication used.

Not applicable. All experiments were done in silico. No eukaryotic cells were used in this study.

c. Report whether the cell lines were tested for mycoplasma contamination.

Not applicable. All experiments were done in silico. No eukaryotic cells were used in this study.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

Not applicable. All experiments were done in silico. No eukaryotic cells were used in this study.

## ► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

### 11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

Not applicable. All experiments were done in silico. No animals were used in this study.

Policy information about [studies involving human research participants](#)

### 12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Not applicable. All experiments were done in silico. This study did not involve human research participants.