



HHS Public Access

Author manuscript

Nat Methods. Author manuscript; available in PMC 2018 September 26.

Published in final edited form as:

Nat Methods. 2018 May ; 15(5): 351–354. doi:10.1038/nmeth.4638.

NAMD goes quantum: An integrative suite for QM/MM simulations

Marcelo C. R. Melo^{1,2,*}, Rafael C. Bernardi^{1,*}, Till Rudack^{1,3}, Maximilian Scheurer^{4,5}, Christoph Riplinger⁶, James C. Phillips¹, Julio D. C. Maia⁷, Gerd B. Rocha⁸, João V. Ribeiro¹, John E. Stone¹, Frank Neese⁹, Klaus Schulten^{1,10,+}, and Zaida Luthey-Schulten^{1,2,9,11,**}

¹NIH Center for Macromolecular Modeling and Bioinformatics, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, 61801, USA

²Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, 61801, USA

³Department of Biophysics, Ruhr-University Bochum, Germany

⁴Biochemistry Center, Heidelberg University, Heidelberg, Germany

⁵Interdisciplinary Center for Scientific Computing, Heidelberg, Germany

⁶FAccTs GmbH, Köln, Germany

⁷Center for Informatics, Federal University of Paraíba, João Pessoa, Brazil

⁸Department of Chemistry, Federal University of Paraíba, João Pessoa, Brazil

⁹Max Planck Institut für Kohlenforschung, Mülheim an der Ruhr, Germany

¹⁰Department of Physics, University of Illinois at Urbana-Champaign, 61801, USA

¹¹Department of Chemistry, University of Illinois at Urbana-Champaign, 61801, USA

Abstract

Applications of hybrid QM/MM methods range from reactions in active sites of small enzymes to multiple sites in large bioenergetic complexes. By combining the widely used molecular dynamics

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

**Corresponding author zan@illinois.edu.

*These authors contributed equally.

+In memoriam.

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 QM/MM interface development. 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 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 FINANCIAL INTERESTS

The authors declare no competing financial interests.

and visualization programs NAMD/VMD with the quantum chemistry packages ORCA, and MOPAC, we provide an integrated, comprehensive, customizable, and easy-to-use suite. Through the QwikMD interface, setup, execution, visualization, and analysis are streamlined for all levels of expertise.

Though molecular mechanics (MM) force-fields are based on quantum mechanical calculations and experimental observations, only quantum mechanics (QM) enables a complete and accurate understanding of many biochemical processes, particularly those involving chemical reactions or charge redistribution [1]. Nevertheless, even with the advanced hardware technology available today, the computational cost of studying nanosecond-long dynamics of entire systems relying solely on QM methodologies is usually prohibitive. A common route to circumvent this cost barrier is to confine the QM formalism to a subregion of a system and to include the effects of the surrounding system through MM simulations, leading to hybrid “QM/MM” simulations [2]. QM/MM calculations are now used broadly in enzymology [3], drug discovery [4], bioenergetic systems [5], as well as in combination with serial femtosecond crystallography [6] and other material sciences and structural biology techniques [7]. To deliver accurate results, these QM/MM studies require a carefully selected quantum region [5, 8]. Unfortunately, most the available QM/MM implementations lack a comprehensive set of features that could make these calculations even more attractive to chemists, structural biologists, and material engineers.

NAMD [9] is one of the most widely used software packages for molecular dynamics (MD) simulations, particularly for large biomolecular systems on supercomputers. NAMD’s scalability and large array of enhanced sampling and free energy methods [10], as well as a seamless integration with VMD [11], which provides extended setup, visualization and analysis capabilities, makes it an invaluable tool for exploring complex biological systems. Here we present a comprehensive QM/MM suite, developed to provide easy setup, visualization and analysis of QM/MM simulations through the graphical user interface QwikMD [12] as well as a broad range of QM methods. Our graphical interface makes our tool unique among widely employed MD codes. The QM/MM interface in NAMD 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) [13]. In addition, NAMD performs similarly to other QM/MM programs regarding 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 [14, 15] or ORCA [16], or through a flexible generic interface. Others, e.g. Amber [17], additionally include their own code for QM calculations. This strategy was not adopted here to ensure flexibility, allowing users to easily fine tune their simulations. Importantly, by performing file I/O 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, allowing for both mechanically and electrostatically embedded QM regions to be used (see Online Methods).

Input files are the same used for classical MD, with additional options in the configuration file. Typically, QM and MM atoms that are covalently bound are treated by redistributing the MM atom's charge over its nearest MM neighbors and by capping the QM atom with a hydrogen atom, known as the Link Atom method, as shown in Fig. 1a and Supplementary Fig. 2 for a calculation using the NAMD/ORCA interface. For all Link Atom variations, see Online Methods and Supplementary Fig. 3a–b, 4, 5a and 5b.

To test the QM/MM interface for accuracy, stability, and performance, we carried out standard validation simulations with both MOPAC and ORCA. NAMD achieved energy conservation for a pure QM alanine molecule (Fig. 1b), a hybrid QM/MM tri-alanine molecule in vacuum (Fig. 1c) and in water (Fig. 1d), and a pure QM N-methyl acetamide (NMA) molecule, the latter for up to 100 ns with both 0.5 fs and 2.0 fs time steps (Fig. 1e). A particle mesh Ewald (PME) treatment of long-range electrostatics was also observed to conserve energy (Supplementary Table 2). Compared, NAMD and Amber16 presented equally good QM/MM results (Supplementary Fig. 6). NAMD/MOPAC was shown to perform up to 10 ns/day of QM/MM simulation in a single desktop computer (Supplementary Table 3). In all tests NAMD was shown to perform efficiently while conserving energy.

To provide a quick and easy way of setting up QM/MM simulations, we incorporated the most widely employed features of a hybrid QM/MM molecular dynamics code to VMD's QwikMD [12], setting up 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 lacking 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 have been extended in this work to natively support MOPAC and ORCA outputs, which benefits not only the users of our QM/MM suite but any users of these two packages. New representation schemes were introduced to VMD allowing the selection and visualization of orbital trajectories throughout a simulation, and to allow for the chemical bonds representation to be dynamically updated. The new capabilities in VMD make it a powerful tool to visualize and analyze output from QM/MM calculations as well as outputs directly from ORCA and MOPAC (see Fig. 2b–d, Supplementary Tables 4, 5 and 6, Supplementary Note and Supplementary Video 1).

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^{Glu}), 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, 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 [18]. However, because the anticodon binding domain and catalytic site of GluRS are over 50 angstroms apart, the treatment of the entire pathway, with approximately 5,000 atoms, using QM accuracy would be unfeasible. Using QwikMD to prepare a system with two independent QM regions, we were able to investigate the allosteric signaling pathway of the GluRS:tRNA^{Glu} complex with an accurate QM treatment of the two most critical regions of the system (see Fig. 3a).

The communication pathways that lead to coordinated motion between functionally important QM regions were analyzed using cross-correlation based network analysis [18] (see Online Methods and Supplementary Fig. 8), and found to be significantly degenerate, as previously observed using classical MD [18] (Fig. 3a). The QM/MM treatment of critical regions lead to increased correlations between residues in the active site and anticodon binding site (Supplementary Fig. 9). Suboptimal communication pathways are based on the correlation of atom motion during the simulation and have been employed 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. In this work, 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).

The aminoacylation reaction mechanism was for the first time investigated to distinguish between four possible mechanisms (see Supplementary Fig. 10), all based on a previously proposed concerted exchange where the 3' or the 2'-OH oxygen attacks the carbonyl carbon of Glu-AMP while the proton shifts to the phosphate's oxygen [19]. In addition, the reaction was investigated while the amine group of the adenylate was protonated or deprotonated. QM/MM steered molecular dynamics (SMD) simulations were performed to induce the possible reaction mechanisms, and intermediate states were selected from the SMD trajectories to initiate a String Method optimization [20]. The method utilizes groups of biased MD simulations to find the path with the smallest barrier that describes the chemical reaction. Once an optimized string was achieved, the reaction path was used to perform a parallel extended-Adaptive Biasing Force [13] (eABF) calculation to determine the free energy transformation of each possible mechanism (see Supplementary Fig. 11, 12). The reaction charging the 3-hydroxyl group, with the amine group of the adenylate being deprotonated, was observed to be the most favorable route with a barrier of ≈ 20 kcal/mol and final state at ≈ 8 kcal/mol (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 opens space for hydration of its phosphate group (Supplementary Fig. 14), and was investigated using ABF and classical MD simulations. As the products move apart and water enters the active site, hydrating the phosphate group of AMP-H, we observed a ≈ 22 kcal/mol drop in free energy, making the entire reaction-solvation process exergonic, with a ≈ -15 kcal/mol free energy variation between reactants and products (Fig.

3e–g). It is worth noting that appending QM/MM eABF and MM ABF results leads to a small imprecision in the connection between free energy values (red dashed line in Fig. 3e–f).

Hybrid QM/MM protocols have been applied to a wide range of investigations, however, a comprehensive, customizable, easy-to-use, and freely available suite was not yet available to the broader computational biology and material sciences communities. 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. Test applications presented here highlight the accuracy of this implementation, while the study of the GluRS:tRNA^{Glu}:Glu-AMP complex revealed sub-atomic details of its reaction mechanism. Uniting Network Analysis results for the full complex with the first combination of the string method, parallel eABF, and QM/MM simulations, we provide a unique view for essential steps in establishing the genetic code (Supplementary Fig. 15). In summary, the ease of access through QwikMD and modularity in NAMD will permit this suite to be used in a variety of applications and 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 as a sandbox for the development of new QM tools and QM/MM interactions schemes.

METHODS

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

ONLINE METHODS

In order to study the chemistry of complex biological systems, in particular using MD approaches, one needs to combine QM and MM methods to secure sub-atomic resolution within relevant time and length scales. Yet, recent advances in implementations of QM calculations and hardware improvements are continuously improving our capacity to expand the complexity and comprehensiveness of QM results [21]. Hybrid QM/MM simulations in NAMD divide the system into “MM” and “QM” regions, using a classical force field to treat the classical atoms (or “MM atoms”), and passing the information that describes the quantum atoms in the system (or “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 non-bonded 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 non-bonded 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 in order to compensate for over-polarization that these atoms may exhibit in hybrid QM/MM simulations, though the importance of such modifications has been disputed [22]. 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 is long known that electrostatics, particularly polarization, plays a key role in many biochemical processes [23]. Electrostatic interactions between QM and MM atoms deserve a more detailed discussion due 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, on the other hand, 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 the “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 [1].

Classical point charges handed over to the QM program to be utilized in electrostatic embedding can be altered by NAMD in a variety of ways. First by applying a smoothing function that avoids 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 so that it amounts to a complementary charge to that of the QM region, in which case the sum of charges from QM atoms and classical partial charges add to zero. In the latter, the user is also able to select, atom by atom, which classical partial charges should be used when building the electrostatic embedding, providing further flexibility to the application.

Irrespective 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) [24] 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 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 between 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 re-calculated within the QM module and subtracted from the direct electrostatic calculations, to avoid double counting these contributions. In order to test the impact of such implementation on the simulations, an energy conservation analysis was carried out comparing the use of PME in NAMD when the QM calculations are done using PM3 [25, 26], HF-3C [27] and DFT methods (Supplementary Table 2).

Treatment of covalent bonds involving QM and MM atoms

Hybrid QM/MM simulations of biomolecular systems often present situations where only a part of a molecule should be treated quantum mechanically, usually to save computational resources since the cost of simulating QM regions rises rapidly with the number of atoms. In order to deal with chemical bonds that have one atom in the quantum mechanical (QM) region and another in the classical (MM) region (from here on called “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 vs MM), and minimize errors involved in the QM/MM division of the system. Irrespective of the selected methods to treat the QM-MM bond, the bonded term (between MM1 and QM1 atoms, see 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 the QM regions containing QM-MM bonds is the “link atom” approach [28], where an atom (usually a hydrogen atom) is placed along the bond between the QM atom and the MM atom (see Supplementary Fig. 3c), and does not exist in the classical simulation. The user can fine-tune this process by choosing the element and which method of link atom placement will be used: either fixed (the default) or variable. The former depends on a user-defined distance, which will be used throughout the simulation as d_{L-QM} (see Supplementary Fig. 3c). For the latter, a user-defined fraction is used to define C_L (see 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 due to bond vibration.

In order 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 [29], 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}_{MM-QM} is the unit vector in the QM-MM bond direction, and x_{MM} and x_{QM} are the x components of the positions of the MM1 and QM1 atoms, respectively.

$$F'_{QM1_x} = \vec{F}_L \cdot \left[(1 - C_L) \cdot \hat{i}_x + C_L \cdot \frac{x_{MM} - x_{QM}}{d_{MM-QM}} \cdot \hat{r}_{MM-QM} \right] \quad (1)$$

where

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

Similarly,

$$F'_{MM1_x} = \vec{F}_L \cdot \left[C_L \cdot \hat{i}_x - C_L \cdot \frac{x_{MM} - x_{QM}}{d_{MM-QM}} \cdot \hat{r}_{MM-QM} \right]. \quad (3)$$

The link atom is not the only proposed method to handle QM-MM bonds, but it is the only one that relies entirely on the classical side of a QM/MM simulation, over which we have complete control. Therefore, irrespective of which QM package is 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 its 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 would create very strong electrostatic repulsion forces (or attractions) in case the MM1 partial charge was sent to the QM package for its calculations. Under the “mechanical embedding” scheme, the QM package only receives the atoms in the QM region and the link atoms created to approximate QM-MM bonds, so no manipulation of partial charges is required. On the other hand, 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 (see 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, since 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[30]. 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” (see Supplementary Fig. 4b–d).

The Charge Shifting (CS) method[31] (see Supplementary Fig. 4e) is more elaborate, as it rearranges the partial charge of the MM1 atom (indicated here as q_{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 by creating “virtual” point charges that are passed to the QM package as if they represented partial charges of classical atoms. In this case, the MM1 partial charge is equally distributed across the MM2 atoms, as indicated in the figure with the placement of the charge q_1 at the position of MM2 atoms (where $q_1 = q_{M1}/2$), and the coefficient used in Equations 4 and 5 to redistribute the force on this virtual charge is given by $C_1 = \frac{d_{M1-Q1}}{d_{M1-M2}} = 1$. 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, $C_+ = \frac{d_{M1-Q_+}}{d_{M1-M2}} = 0.94$, while $C_- = \frac{d_{M1-Q_-}}{d_{M1-M2}} = 1.06$ (see

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

In all cases where a virtual charge is created, a force is calculated on it as to balance its electrostatic interactions with QM atoms. Similarly to the Link atom, the total force acting on, e.g., 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

$$\vec{F}_{MM_1} = (1 - C_+) \cdot \vec{F}_{q_+} \quad (4)$$

and

$$\vec{F}_{MM_2} = C_+ \cdot \vec{F}_{q_+} \quad (5)$$

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 [16] or MOPAC [14, 15], or through a standardized interface that utilizes 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 I/O so that external wrapper scripts can be called, providing a translation layer for input and output files from any other QM package. We provide Python scripts that wrap Gaussian [32], TeraChem [33], and Q-CHEM [34], which also serve as templates for the development of new wrapper scripts.

In general, positions and elements of QM atoms are passed on 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 take an insignificant amount of time (milliseconds) when compared to 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 an SCF calculation on a given set of atoms, a “state-files” containing the result of the SCF procedure is saved. A much faster convergence is achieved in the following 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 ttGluRS were prepared for classical and QM/MM MD simulations employing VMD's [11] QwikMD [12] 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 [35]. The structure of the pre-transfer complex ttGluRS:tRNA^{Glu}: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) [36], with the replacement of the inert analog for the active Glu-AMP [19]. The simulations in the present study were performed employing the NAMD molecular dynamics package [9], MOPAC 2016 [14, 15], using PM7 [37], and ORCA 4.0 [16]. The CHARMM36 force field [38] along with the TIP3P water model, for solvated tests, was used to describe all systems. All calculations involving the ttGluRS:tRNA^{Glu}:Glu-AMP system, were done in the NpT ensemble. In energy conservation tests, the simulations used either the PM3 [25, 26] method (using NAMD's ORCA interface) or RM1 [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 performed in many steps to ensure a reasonable starting conformation for QM-based simulations: (1) First, for systems explicitly solvated, we employed 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, QM-based NpT simulations (either pure-QM or hybrid QM/MM) were carried out. (3) For energy conservation tests, QM-based simulations were performed in the NVE ensemble.

A distance cut-off of 12.0 Å or 17.0 Å (eABF calculations) was applied to short-range, non-bonded interactions, whereas long-range electrostatic interactions were treated using the particle-mesh Ewald (PME) [24] method. For equilibration and classical MM simulations, the equations of motion were integrated using the r-RESPA multiple time step scheme [9]. For all production QM/MM simulations, the van der Waals interactions as well as electrostatic interactions were updated every time step. The time step of integration was chosen to be either 0.5 fs or 2.0 fs (see details below).

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

The PolyAla system is composed of 3 Alanine residues. The same system was used in different ways, either in vacuum or in solution (TIP3 water), with either all 3 residues treated quantum mechanically, or just the central residue as a QM region, leading to the use of 2 Link atoms (Supplementary Fig. 3a and b). In Fig. 1c, the system is used in vacuum, with the middle residue being treated quantum mechanically. In Fig. 1e, the system is used in solution, with the all three residues treated quantum mechanically, and the water molecules treated classically, as also shown in Supplementary Fig. 2a for the NAMD/ORCA interface. Simulations were performed with 0.5 fs time step and the Charge Shift treatment was used for Link Atoms. To reproduce previously reported energy conservation results in Amber, the SPC/FW water model was used in all Amber QM/MM simulations.

Analogously, the N-methyl acetamide (NMA) system was simulated either in vacuum or in solution (TIP3 water), but always entirely treated with QM formalism. In Fig. 1d, the system is 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 step, as shown in Fig. 1d.

Network Analysis

Suboptimal communication pathways and communities were calculated based on Pearson cross-correlation of atom motion, using a similar protocol to the one described in [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 employed, one in the anticodon domain with approximately 500 QM atoms, one in the catalytic region with approximately 200 QM atoms. Briefly, the network analysis protocol uses central atoms (alpha-carbons in amino acids, the N1/N9 in nucleic bases, and the sugar phosphate P) called “nodes” as 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 were less than 4.5Å apart for more than 75% of the total simulated time are considered connected [40]. Nodes that do not fit this cutoff have their cross-correlation multiplied by zero through a mask applied on 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 shortest communication paths between nodes (called “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^{Glu}:Glu-AMP system, first a 200 ps QM/MM simulation was performed with 2 fs time step, followed by a 50 ps QM/MM simulation with 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. Four mechanisms were tested by inducing the reaction steps to occur (Supplementary Fig. 10). The four biased simulations were performed with 0.5 fs time step for 20 to 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

In order 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, between centers of mass of groups of atoms, or even angles between subdomains of a molecular structure. The collective variables (colvars for short) 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 different stages of the transformation. At each iteration, multiple independent MD simulations are initiated from each image, allowing 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 colvars space, and to smooth the reaction coordinate. The iterations proceed until consecutive calculations do not produce significant changes in the mean colvar values, indicating 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 (extended-Adaptive Biasing Force) is used to calculate free energy changes. Using the images that were optimized through the string method, new path collective variables S and Z are created, which constrain the dynamics of the system so it follows the chosen reaction coordinate. The S variable indicates progression along the path, while the Z variable 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 initiates multiple walkers from the different images, we could conduct extensive sampling over the defined path.

Code availability

The reported QM/MM features are publicly available in NAMD 2.12 or later, and VMD 1.9.4 or later. See www.ks.uiuc.edu/Research/qmmm for more details. A Tutorial is available in the same website.

ORCA is available at <https://orcaforum.cec.mpg.de> and <https://www.faccts.de>.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank M. F. Herbst, C. Chipot and G. Fiorin for helpful discussions. This work was supported by National Science Foundation (NSF) grants MCB-1616590 and MCB-1244570, and National Institutes of Health (NIH) grant P41-GM104601. Z.L.S. acknowledges support from the NSF PHY1430124. M.C.R.M. and Z.L.S. acknowledge support from the Keck Foundation (grant no. 206231). T.R. acknowledges support as a Feodor Lynen von Humboldt Postdoctoral Fellow. J.D.C.M acknowledges a fellowship from the Brazilian Coordination for Improvement of Higher Educational Personnel (CAPES). G.B.R. acknowledges support from the Brazilian National Council for Scientific and Technological Development (CNPq 305271/2013-0), and CAPES AUXPE1375/2014. 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 the Blue Waters sustained-petascale computing, which is supported by the state of Illinois and the NSF OCI-0725070 and ACI-1238993. This work is part of the Petascale Computational Resource (PRAC) grant, which is supported by the NSF ACI-1713784.

References

1. Senn HM, Thiel W. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed. Engl.* 2009; 48:1198–229. [PubMed: 19173328]
2. Field MJ, Bash PA, Karplus M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* 1990; 11:700–733.
3. van der Kamp MW, Mulholland AJ. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry.* 2013; 52:2708–2728. [PubMed: 23557014]
4. Bernardi RC, Pascutti PG. Hybrid QM/MM Molecular Dynamics Study of Benzocaine in a Membrane Environment: How Does a Quantum Mechanical Treatment of Both Anesthetic and Lipids Affect Their Interaction. *J. Chem. Theor. Comp.* 2012; 8:2197–2203.
5. Retegan M, Neese F, Pantazis DA. Convergence of QM/MM and Cluster Models for the Spectroscopic Properties of the Oxygen-Evolving Complex in Photosystem II. *J. Chem. Theory Comput.* 2013; 9:3832–3842. [PubMed: 26584129]
6. Coquelle N, et al. Chromophore twisting in the excited state of a photoswitchable fluorescent protein captured by time-resolved serial femtosecond crystallography. *Nature Chemistry.* 2017
7. Lin H, Truhlar DG. QM/MM: what have we learned, where are we, where do we go from here? *Theor. Chim. Acta.* 2007; 117:185–199.
8. Jindal G, Warshel A. Exploring the dependence of QM/MM calculations of enzyme catalysis on the size of the QM region. *J. Phys. Chem. B.* 2016; 120:9913–9921. [PubMed: 27552257]
9. Phillips JC, et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 2005; 26:1781–1802. [PubMed: 16222654]
10. Jiang W, et al. Generalized scalable multiple copy algorithms for molecular dynamics simulations in NAMD. *Comput. Phys. Commun.* 2014; 185:908–916. [PubMed: 24944348]
11. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J. Mol. Graphics.* 1996; 14:33–38.
12. Ribeiro JV, et al. QwikMD Integrative Molecular Dynamics Toolkit for Novices and Experts. *Sci. Rep.* 2016; 6:1–14. [PubMed: 28442746]
13. Fu H, Shao X, Chipot C, Cai W. Extended Adaptive Biasing Force Algorithm. An On-the-Fly Implementation for Accurate Free-Energy Calculations. *J. Chem. Theory Comput.* 2016; 12:3506–3513. [PubMed: 27398726]

14. Stewart JJ. MOPAC: a semiempirical molecular orbital program. *J. Comput.-Aided Mol. Des.* 1990; 4:1–103. [PubMed: 2197373]
15. Maia JDC, et al. GPU linear algebra libraries and GPGPU programming for accelerating MOPAC semiempirical quantum chemistry calculations. *J. Chem. Theor. Comp.* 2012; 8:3072–3081.
16. Neese F. Software update: the ORCA program system, version 4.0. *Wiley Interdisciplinary Reviews: Computational Molecular Science.* 2018; 8:e1327.
17. Götz AW, Clark MA, Walker RC. An extensible interface for QM/MM molecular dynamics simulations with AMBER. *J. Comput. Chem.* 2014; 35:95–108. [PubMed: 24122798]
18. Sethi A, Eargle J, Black AA, Luthey-Schulten Z. Dynamical networks in tRNA:protein complexes. *Proc. Natl. Acad. Sci.* 2009; 106:6620–6625. [PubMed: 19351898]
19. Black Pyrkosz A, Eargle J, Sethi A, Luthey-Schulten Z. Exit Strategies for Charged tRNA from GluRS. *J. Mol. Biol.* 2010; 397:1350–1371. [PubMed: 20156451]
20. Pan AC, Sezer D, Roux B. Finding Transition Pathways Using the String Method with Swarms of Trajectories. *J. Phys. Chem. B.* 2008; 112:3432–3440. [PubMed: 18290641]
21. Riplinger C, Pinski P, Becker U, Valeev EF, Neese F. Sparse maps—A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory. *J. Chem. Phys.* 2016; 144:024109. [PubMed: 26772556]
22. Riccardi D, Li G, Cui Q. Importance of van der Waals Interactions in QM/MM Simulations. *J. Phys. Chem. B.* 2004; 108:6467–6478. [PubMed: 18950136]
23. Warshel A, Levitt M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* 1976; 103:227–249. [PubMed: 985660]
24. Darden T, York D, Pedersen L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* 1993; 98:10089–10092.
25. Stewart JJP. Optimization of parameters for semiempirical methods I. *Method. J. Comput. Chem.* 1989; 10:209–220.
26. Stewart JJP. Optimization of parameters for semiempirical methods II. Applications. *J. Comput. Chem.* 1989; 10:221–264.
27. Sure R, Grimme S. Corrected small basis set Hartree-Fock method for large systems. *J. Comput. Chem.* 2013; 34:1672–1685. [PubMed: 23670872]
28. Singh UC, Kollman PA. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the CH₃Cl + Cl exchange reaction and gas phase protonation of polyethers. *J. Comput. Chem.* 1986; 7:718–730.
29. Walker RC, Crowley MF, Case DA. The implementation of a fast and accurate QM/MM potential method in Amber. *J. Comput. Chem.* 2008; 29:1019–1031. [PubMed: 18072177]
30. Lin H, Truhlar DG. Redistributed charge and dipole schemes for combined quantum mechanical and molecular mechanical calculations. *J. Phys. Chem. A.* 2005; 109:3991–4004. [PubMed: 16833721]
31. Sherwood P, et al. Computer simulation of zeolite structure and reactivity using embedded cluster methods. *Faraday Discuss.* 1997; 106:79–92.
32. Frisch MJ, et al. Gaussian 09, Revision A.02. Gaussian Inc. Wallingford CT. 2016
33. Titov AV, Ufimtsev IS, Luehr N, Martinez TJ. Generating Efficient Quantum Chemistry Codes for Novel Architectures. *J. Chem. Theory Comput.* 2013; 9:213–221. [PubMed: 26589024]
34. Shao Y, et al. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* 2015; 113:184–215.
35. Kulik HJ, Zhang J, Klinman JP, Martínez TJ. How Large Should the QM Region Be in QM/MM Calculations? The Case of Catechol O-Methyltransferase. *J. Phys. Chem. B.* 2016; 120:11381–11394. [PubMed: 27704827]
36. Sekine S-i, et al. ATP binding by glutamyl-tRNA synthetase is switched to the productive mode by tRNA binding. *EMBO J.* 2003; 22:676–688. [PubMed: 12554668]
37. Stewart JJP. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* 2013; 19:1–32. [PubMed: 23187683]

38. Best RB, et al. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain χ 1 and χ 2 Dihedral Angles. *J. Chem. Theor. Comp.* 2012; 8:3257–3273.
39. Dutra JDL, Filho MAM, Rocha GB, Simas AM, Freire RO. RM1 Semiempirical Quantum Chemistry: Parameters for Trivalent Lanthanum, Cerium and Praseodymium. *PLoS One.* 2015; 10:e0124372. [PubMed: 26132289]
40. Schoeler C, et al. Mapping mechanical force propagation through biomolecular complexes. *Nano Lett.* 2015; 15:7370–7376. [PubMed: 26259544]

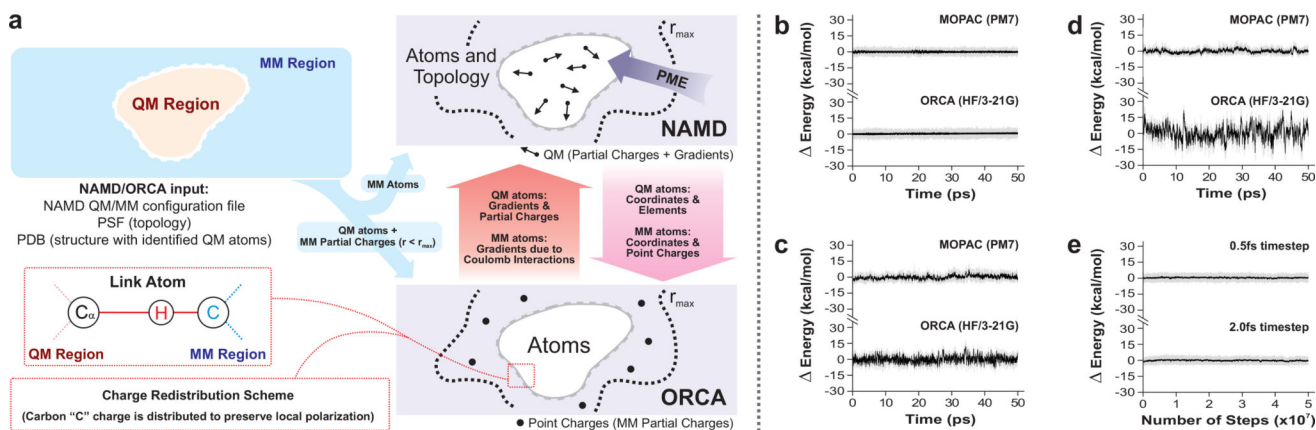


Figure 1. Hybrid QM/MM NAMD

a) Schematic of NAMD-ORCA interconnection. The contribution of MM charges beyond r_{max} are calculated by NAMD (via PME), while ORCA calculates direct electrostatics. **b)** Energy conservation test for a pure-QM alanine molecule. All energy conservation plots display the deviation from the mean system energy. The black line indicates a running average and the gray line indicates maximal fluctuation. **c)** Energy conservation test for a QM/MM tri-alanine molecule. **d)** Energy conservation test for a QM/MM tri-alanine molecule in water. The basis set employed in ORCA tests (3-21G) serves only as a technical test. **e)** Energy conservation test for a pure-QM NMA molecule using NAMD/MOPAC with PM7.

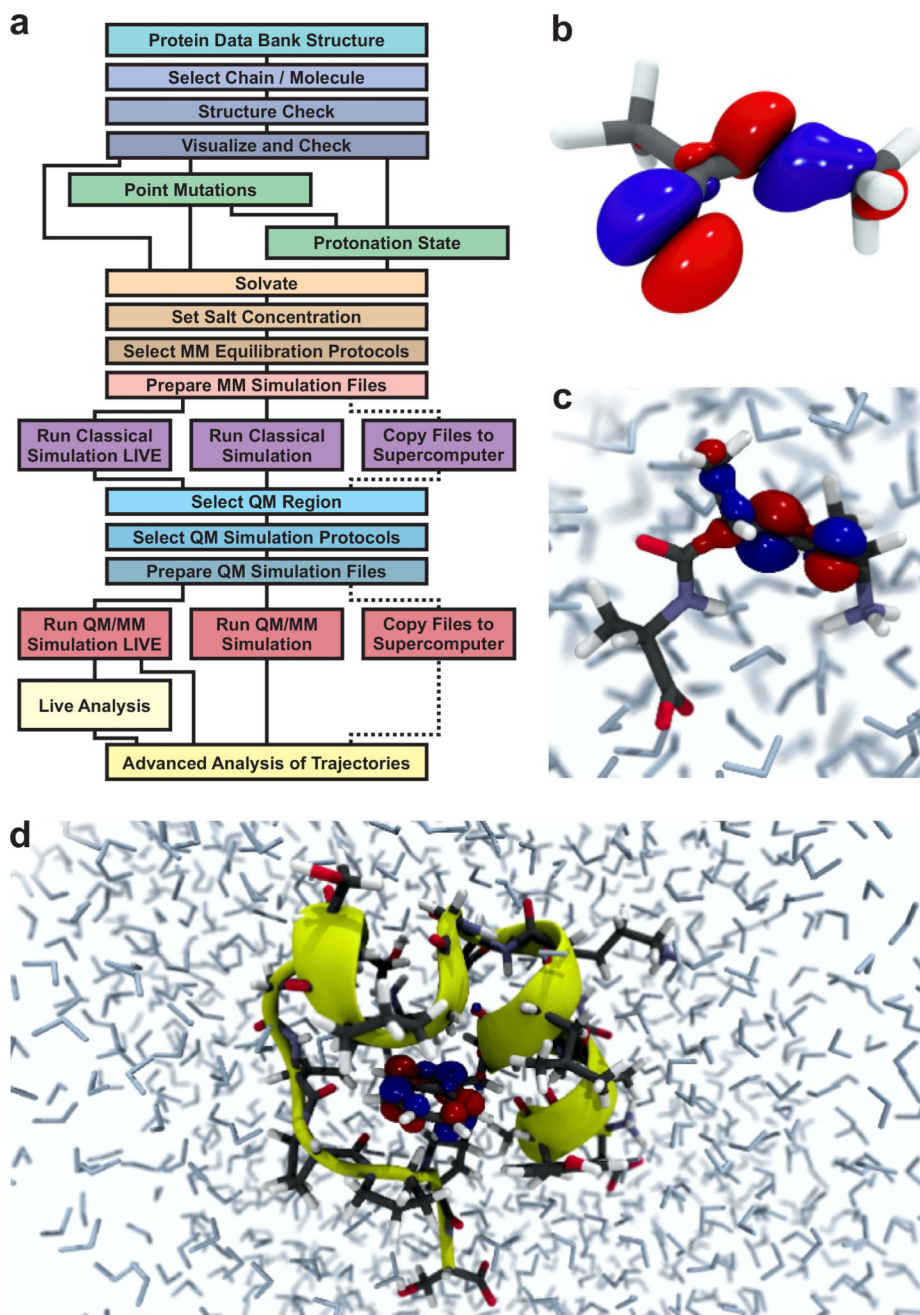


Figure 2. Hybrid QM/MM VMD features

a) QwikMD provides a graphical user interface (GUI) in VMD for performing QM/MM simulations. The image shows the workflow to prepare, run, analyze, and visualize a hybrid QM/MM molecular dynamics simulation. **b**) Highest occupied molecular orbital of an alanine molecule in vacuum. **c**) Alanine's highest occupied molecular orbital in a solvated QM/MM tri-alanine. **d**) Trp-Cage protein highest occupied molecular orbital in water solution.

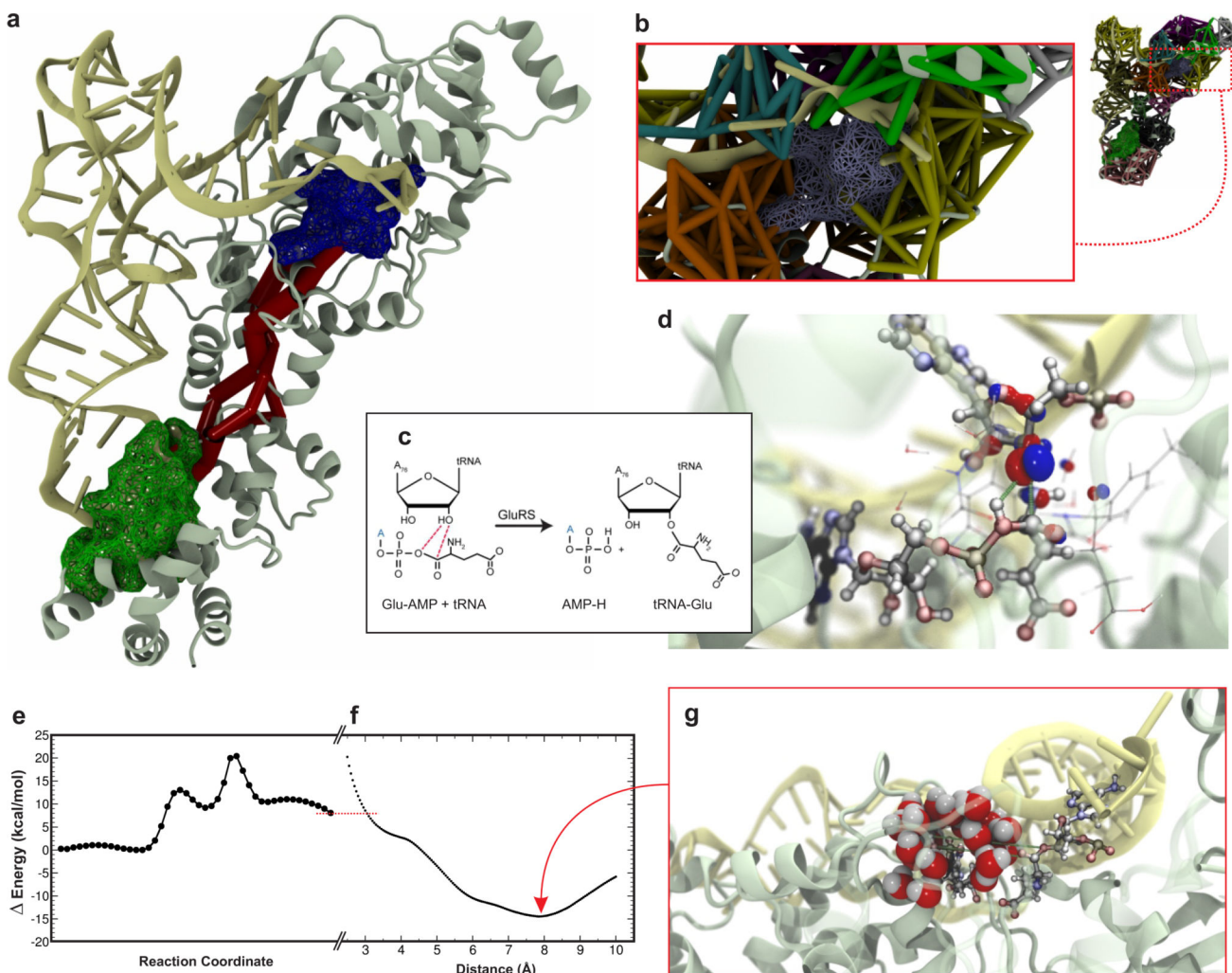


Figure 3. Mechanism of glutamyl-tRNA synthetase

a) Representation of the glutamyl-tRNA synthetase allosteric pathway (red). Two independent QM regions are highlighted, indicating the active site (blue) and the anticodon binding region (green). **b)** Community analysis showing multiple communities involved in the active site. **c)** Reaction mechanism of glutamyl-tRNA synthetase. **d)** Intermediate state of the glutamyl-tRNA synthetase reaction showing the highest occupied molecular orbital. **e)** Free energy profile of the glutamyl-tRNA synthetase reaction mechanism, calculated using eABF after a string method path optimization. Both eABF and the string method were performed using QM/MM MD simulations with NAMD/MOPAC and PM7. **f)** Free energy profile of the distancing and solvation of AMP calculated using ABF and classical MD simulations. **g)** Snapshot of the minimum energy state during the release of the AMP, showing the solvation of the phosphate group. The eABF free energy profile in **e)** corresponds to the most favorable mechanism (**d)**). The final state of the reaction revealed that the AMP-H phosphate was very close to the glutamyl-tRNA, keeping water from entering the active site. As the AMP-H moves away from the charged tRNA, water hydrates the AMP-H, as shown in **f)** and **g)**. Since in the final state of the eABF calculation the distance between Glu-AMP:Carb-C and Glu-AMP:P-O was of 3 Å, we used this distance as the

connecting point (see red dashed line between **e** and **f**) between the free energy profiles calculated with eABF and ABF.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript