UCLA

UCLA Previously Published Works

Title

Use of QM/DMD as a Multiscale Approach to Modeling Metalloenzymes.

Permalink

<https://escholarship.org/uc/item/4772b5pr>

Journal

Methods in enzymology, 577

ISSN

0076-6879

Authors

Gallup, NM
Alexandrova, AN

Publication Date

2016

DOI

10.1016/bs.mie.2016.05.018

Peer reviewed

Use of QM/DMD as a Multiscale Approach to Modeling Metalloenzymes

N.M. Gallup*, A.N. Alexandrova*,†,‡

*University of California, Los Angeles, Los Angeles, CA, United States

†California NanoSystems Institute, Los Angeles, CA, United States

‡Corresponding author: e-mail address: ana@chem.ucla.edu

Contents

1. Introduction	320
2. Overview of QM/DMD	321
2.1 Advantages of QM/DMD over Current QM/MM Schemes	321
2.2 QM/DMD Method	323
3. Setting Up QM/DMD	324
3.1 Acquiring the Protein Crystal Structure	325
3.2 Setting Up the DMD Region	326
3.3 Setting Up the QM–DMD Region	330
4. Running QM/DMD and Details of the Procedure	331
4.1 The Blueprint of the Algorithm	331
4.2 Initial Iteration (Iteration 0)	333
4.3 Full Iterations (Iterations 1 Through <i>N</i>)	333
4.4 A Simple Analysis of Convergence	335
4.5 Tasks for Which QM/DMD Is Best Suited	337
4.6 Permissions, Copyrights, and Utilities	337
5. Conclusions	338
References	338

Abstract

Enzymes are complex biomolecules capable of performing unique catalysis under physiological conditions at neutral temperature and pH. However, the architecture of enzymatic catalysis is often a combination of the quantum influence of the immediate active site, as well as the electrostatic and configurational influences of amino acids surrounding the active site. As a result of this cooperation between baseline chemical reactivity and electrostatic assistance, it has become important to model enzymes using multiscale methods that take advantage of treating the active site with quantum mechanical methods, while approximately treating the surrounding protein using cheaper, classically driven force-field molecular mechanics methods. Here we describe

the use of a multiscale engine which utilizes a combination of density functional theory with discrete molecular dynamics (dubbed QM/DMD) to aid in the characterization of metalloenzymes.

1. INTRODUCTION

Enzymes have become particular targets of chemical interest in recent decades due to their marked importance in biological function and the potential for use in industrial applications. However, describing enzyme catalysis, particularly metalloenzymes, possesses unique challenges over typical small-molecule systems (Chung, Li, & Morokuma, 2010; Warshel, 1997). Since the introduction of Warshel and Levitt's simplified protein models (1976), hybrid quantum mechanical and force-field-driven molecular mechanical methods (collectively known as hybrid QM/MM methods) have become an accepted means in describing the catalytic action that takes place at enzyme active sites and has expanded to become a versatile tool in examining solvated small-molecular complexes (Senn & Thiel, 2009). These methods typically rely on partitioning a system into QM and MM subcomponents, and iterating between a QM integration scheme and an MM integration scheme. The QM region usually includes species participating directly in chemical reactions of interest and may expand to include atoms, species, or residues that are thought to exert significant influence on the chemical reactivity that would unlikely be captured via electrostatic or polarizable embedding schemes. The boundary between QM and MM components often bisects covalent bonds. When this occurs, it is common practice in start-to-finish, fixed region schemes to cap these dangling bonds with frozen hydrogen atoms. Dynamic region definition schemes also exist and allow for covering an overall larger QM region, though exhibit more convergence problems. During the QM integration step, the region that surrounds the QM region is usually represented as fixed, polarizable, or non-polarizable point charges. The MM region is normally the region around the active site that one would wish to be described by classical force fields, ie, harmonic bonding potentials, van der Waals, and Coulombic nonbonded interactions. Occasionally, the scheme may contain an intermediate boundary region, which is influenced by both QM and MM. In some methods, such as ONIOM, the MM calculation is done on both MM and QM regions.

This chapter will focus on the use of a hybrid QM and discrete molecular dynamics (QM/DMD) engine developed with particular focus on metalloenzymes (Sparta, Shirvanyants, Ding, Dokholyan, & Alexandrova, 2012). The QM/DMD engine was designed to adequately handle the additional challenges that embody metalloenzyme catalysis, and the substantial Coulombic influence and configurational preference that active site metals exert on their surroundings. Additionally it was also designed with scalability and speed in mind. QM/DMD takes advantage of the established motif of partitioning a system into QM and MM subcomponents. The QM region is treated with density functional theory (DFT), although other *ab initio* methods could easily be used. The MM region is treated with DMD which was developed by the Dokholyan group (Dokholyan, Buldyrev, Stanley, & Shakhnovich, 1998). Later, we describe in detail the qualities of QM/DMD that we believe are advantageous over other QM/MM methods. We then provide a thorough description of the method, setup, analysis of the results, and best known uses.

2. OVERVIEW OF QM/DMD

2.1 Advantages of QM/DMD over Current QM/MM Schemes

The hybrid QM/DMD engine attempts to overcome some of the issues relating to the use of traditional QM/MM schemes with enzymes and metalloenzymes. One of the most significant shortcomings of general QM/MM schemes is the difficulty in providing adequate sampling of the active site for a sufficient number of QM configurations. This becomes particularly expensive for QM/MM schemes that utilize Born–Oppenheimer molecular dynamics (BO–MD) for their QM methods. BO–MD seeks to integrate atomic motions of the QM region on the QM potential energy surface while damping the velocity of those motions with a thermostat set to a given temperature, eg, 300 K. While BO–MD provides a highly accurate description of the dynamics of the active site, the large number of calculations required often renders it prohibitively computationally expensive. More traditional QM/MM engines make use of *ab initio* geometry optimizations instead of QM dynamics. This is a more economical approach, but lacks sampling. In QM/DMD, we take advantage of DMD’s ability to provide extensive sampling, but with a frequent reference to QM for the active site geometry. The only region that is never sampled dynamically is the immediate vicinity of the metal.

QM/MM schemes have employed a large variety of QM methods. Hartree–Fock (HF) and post-HF methods are often used, and post-HF description is essentially the only reliable tool when excited state QM/MM simulations are performed. For ground state simulations, DFT became much more popular over HF and post-HF methods, due the increased accuracy and improved scaling with respect to the number of electrons in the system. Semiempirical methods are also used, and are cheap and allow for extensive sampling, but accuracy is questionable. In addition, parameterization, particularly for metals, is extremely limited. The original aim of QM/DMD is to accurately and efficiently sample the ground state ensemble of metalloproteins, and therefore, DFT is the most natural choice for the QM method in QM/DMD.

DMD is particularly exciting part of QM/DMD, as it allows for very rapid sampling, and overcome some of the traditional problems with MM schemes. For example, it is generally acknowledged that solvation can play a critical role not only in providing the electrostatics necessary to facilitate a given reaction but also in the conformation of biomolecules. Generally, QM/MM schemes elect to account for this effect either via the inclusion of explicit water molecules or by freezing the initial backbone configuration of the protein for the duration of the QM/MM simulation.

The inclusion of explicit water molecules is an excellent way to model solvation. For large enzymes and biomolecules, however, a sufficiently solvated solute may require hundreds or thousands of water molecules. Under periodic boundary conditions that try to replicate the effects of a solvent continuum, this number is particularly large, since the solvent box must be large enough to prevent interactions of solutes from neighboring cells. Sometimes, this model is replaced with a finite solvent cap. In either implementation, this massive number of water molecules can drastically increase the computational expense of both the QM and MM portions, and sampling may become intractable, especially if a quality QM method is in use. In QM/DMD, solvation is implicit, apart from a few water molecules that might be directly bound to the active site and/or play a role in the catalyzed reaction. This is more granular, but our extensive tests show that the performance is adequate for a variety of tasks of interest, which will be discussed later. However, one can solvate the DMD-sampled enzyme in another, non-DMD simulation, if explicit placement of water molecules is essential for the project at hand.

Simulations start with a high-quality initial configuration of the protein typically obtained at the Protein Data Bank (PDB) (Berman et al., 2000).

The structures at the PDB typically are X-ray structures, crystallized at non-physiological, and exceptionally low temperatures to allow for accurate imaging, and often in the presence of unusual solutes or ions that help induce crystallization. It is possible that the protein configuration in the PDB is not the same as at equilibrium at room temperature and normal conditions. Often, however, the protein backbone is kept frozen in QM/MM simulations, because backbone sampling, especially in explicit solvent, adds to computational expense, and equilibration becomes hard to reach. When this is done, it is implicitly asserted that the free-energy surface of the crystallized protein at its low temperature and in its current minima is approximate to neutral, physiological conditions. This assumption may not be valid, since entropy is as large as enthalpy for flexible systems in solution at finite temperatures. Thus, efficient sampling of the protein backbone is essential. DMD is a highly efficient and rapid sampling technique, and with it, we seek to overcome the potential intractability of explicit water, as well as the issues with frozen backbone approximation.

Another pitfall of QM/MM MD is the discontinuities of the potential at the QM–MM boundary. When forces acting on atoms need to be calculated in MD, these discontinuities create cusps. Various force-matching algorithms exist to overcome the problem. In QM/DMD, this problem does not occur, because of DMD's formulation which relies on ballistic equations of motion over Newtonian.

2.2 QM/DMD Method

QM/DMD relies on iterating between a QM integration scheme and an MM integration scheme, like most QM/MM methods. The QM region is defined at the beginning and is fixed start-to-finish. Where the QM region bisects covalent bonds, hydrogen atoms are used to cap the QM portion of the protein. The QM integration scheme consists of a geometry optimization via gradients with DFT to allow for accessible scaling, but the implementation of other *ab initio* methods is, of course, possible. The protein environment surrounding the QM region during the QM step is modeled using the conductor-like screening continuum solvation model, COSMO (Klamt & Schüürmann, 1993), to increase speed with mild penalty to accuracy. The QM region (in QM/DMD we call it QM–DMD region, because QM and DMD regions overlap in the method) should consist of all meaningful elements of chemical interest and expand outward to up to two or

Fig. 1 The purple and red regions represent the QM–DMD boundary, QM–DMD region, and QM region. The rest of the protein constitutes the DMD-only region. The *inset* on the right is the cutout of the QM–DMD region.

even three coordination spheres if necessary. A representation of the cooperative quantum and classical regions can be found in Fig. 1.

The DMD region should consist of the entire protein, except for the active site metals and atoms directly coordinating to them. DMD provides for extensive sampling of protein conformations. It relies on an atomistic, event-driven integration of nuclei along spatial, not temporal, coordinates. The DMD method makes use of a slightly coarse-grained force field containing implicit solvation of amino acids in proteins, enabling exclusion of explicit water solvation and subsequently drastically reducing computational costs, while still including the effect of solvation on protein configuration.

The QM and DMD regions have a large overlap, allowing for structural information to transmit across boundaries between each integration step. We find that this approach saves many troubles having to do with the accuracy of treating of the QM–MM boundary. However, the energy of the shared QM–DMD region is double counted in QM/DMD due to this boundary overlap.

3. SETTING UP QM/DMD

Let us walk through a typical setup of QM/DMD. We will discuss the intuition one must utilize when creating the necessary constraints on the system, and then examine some of the potential results and insight that QM/DMD could provide. While much of the information contained

herein pertains primarily to QM/DMD, a fair amount of the intuition that goes into selection of QM and DMD regions, respectively, is widely applicable to other QM/MM engines. We will use the metalloenzyme histone deacetylase 8, HDAC8, as an example. The expression of this protein is associated with a variety of diseases, and whose targeted inhibition is an FDA-approved therapy for some cancers (Marks, 2007).

3.1 Acquiring the Protein Crystal Structure

All protein crystal structures at the PDB are assigned unique accession codes. There is often a number of crystal structures for any given enzyme, whose structure could differ based on crystallization environment, degree of success in crystallizing with substrates or important metals, or even mutants that facilitated crystallization of important states. Here, we are going to select the PDB structure [2V5W](#), which was published in 2007 (Vannini et al., 2007) and has a resolution of 2.0 Å. When selecting a PDB structure, it is important to weight the resolution heavily in your consideration because this represents the experimental confidence in the location of all atoms. If the resolution is extremely low, the orientations of entire residues, or even their identities, may be obfuscated. Most PDB structures have too low a resolution to resolve the position of hydrogen atoms, thus we will need to manually add them later. Besides the resolution, [2V5W](#) contains many other essential components that will allow it to be studied, including the bound substrate and as additional metal ion in the vicinity of the active site. More often than not, it is not possible to find a crystal structure with bound substrate, in which case the substrate can be manually docked to the protein during the QM/DMD setup.

The [2V5W](#) structure is unique in that it is a noncatalytic variant of wild-type HDAC8. In order to facilitate the crystallization of HDAC8 with bound substrate, the catalysis had to be inhibited, which was done by mutating Y306 into F306, whose original identity had been implicated in being essential for catalysis. Since we are interested in describing this catalysis, we need to reverse this mutation. To start, we need to clean up the PDB file of several extraneous components. As seen in [Fig. 2](#), our PDB contains several HDAC8 images. This represents the repeat unit cell of the crystal structure, and the orientation of each enzyme upon crystallization. Differences in these enzymes can sometimes reveal interesting properties about the enzymes themselves or their substrates, but in this case they are redundant and only

Fig. 2 The two crystallized enzymes are colored in *blue* (*gray* in the print version) and *orange* (*light gray* in the print version). The interface contains a substrate analog and substantial water oxygens.

one is necessary. We kept only one chain (chain A) for QM/DMD simulations.

However, our structure still has a substantial amount of unnecessary components: solvent oxygens and excess substrate; other solutes that were part of the crystallization process but not critical to the functioning of the enzyme may also be present. These excess components are more visible in Fig. 3 and should be manually removed by editing the PDB text file or using a program such as *PyMol* or *UCSF Chimera*. Now our structure has a single enzyme unit, no extraneous solvent, and a viable substrate is already bound (Fig. 4). It will be receiving a few final modifications to make it a high-quality structure for use in QM/DMD.

3.2 Setting Up the DMD Region

The DMD force field includes only polar hydrogens, eg, the ϵH of a His residue whose coordinates need to be provided. However, particular attention must be given to the active site and substrate, as they may exhibit non-typical protonation states. Since the substrate is nonnative to the DMD force field, it will need to be manually protonated. Since DMD makes a distinction between polar (labeled H in the PDB) and nonpolar hydrogens (Eh in the PDB), one must exercise his/her chemical intuition as to what each hydrogen of the substrate is. It is also important to choose correct protonation states for metal-bound amino acids. His residues often switch their protonation from the ϵN to the δN , if the former is bound to a metal.

Fig. 3 The cocrystallized solvent is more apparent. Each water oxygen is represented by a red (gray in the print version) sphere. The substrate analogs bound to the remaining enzyme, and the former enzyme can also be observed (tan).

Asp, Glu, and Cys are deprotonated. QM/DMD has built-in ways to override the protonation states of amino acids for use in the active site. Additionally, DMD only has built-in parameters for the Zn ion, thus it is necessary to change the identity of our catalytic metal at the active site to that of Zn, but only for the DMD integration step. The metal's identity will be reversed for the QM step. This is likely to be a reasonable general approximation as, outside of the local bonding environment, any divalent metal ion is likely to exert similar influences to its second coordination sphere, and furthermore, the immediate metal coordination is not sampled by DMD in QM/DMD. The geometry of the local bonding environment can be changed by QM, and QM will see the metal with its true identity.

While DMD offers significant advantages in terms of speed and subsequently sampling, it does come with caveats whose effects should be minimized. In particular, because the substrate is unlikely to be natively included in DMD's force field, DMD will generate parameters for it on-the-fly. These parameters, however, are unlikely to be as high quality as native parameters, and if left unchecked, may facilitate unusual geometries with its coordinating neighbors. Therefore, it might be necessary to adjust the force-field parameters, or apply constraints to the system to prevent it from

Fig. 4 The final PDB structure. Few modifications remain to make this functional for QM/DMD.

taking on unphysical configurations. Constraints are also always applied to the metal–ligand interactions. [Fig. 5](#) highlights some example considerations at the active site of **2V5W**.

Generally speaking, the entire substrate is frozen for the DMD integration step to prevent DMD from contorting it into unnatural configurations. Nonbonded interactions of the substrate with other protein parts are included in DMD, however. Likewise, coordinating ligand bonds to the substrate are also good targets for constraints. As an example, one potential candidate for a constraint is a hydrogen bond highlighted in red in [Fig. 5](#). A small variety of constraints are also available within DMD, such as frozen nuclei positions in Cartesian space or frozen atom–atom distances. DMD’s atom–atom constraints are analogous to the common harmonic constraint that can be applied in MD simulations to a pair of atoms; DMD’s atom–atom constraint is modeled as a square well, in line with DMD’s use of step-wise and discontinuous potentials. It is generally good practice to constrain important substrate–ligand bonds, such as hydrogen bonds, using atom–atom constraints to allow for some sampling of these bonds. Bond length

Fig. 5 Red marks indicate potential targets for DMD constraints. The red line connects a hydrogen bond between the substrate and the backbone of the protein, while the red circle highlights the coordinating ligands around the active site metal.

constraints on the order of ± 0.1 to $1.0 \text{ \AA}$ is generally adequate in providing a compromise between sampling of reasonable bond lengths and disallowing unusual configurations. Note again that all constraints introduced for DMD are released during QM, during every iteration. Therefore, sampling in these areas does take place, but only at the QM level.

The active site metal, circled red in Fig. 5, also requires constraints, to prevent DMD from controlling its coordination, and instead follow only the physically relevant QM gradients. It is best practice to apply freezes to all the Cartesian coordinates of the metal and all its coordinating atoms in DMD. Then, DMD would sample the backbone and side chain atoms of the metal ligands, excluding only the metal-coordinating atoms.

Sometimes, it is also important to constrain parts of the system in the second or third coordination spheres, if they exhibit unphysical and systematic fluctuations when other parts of the protein are absent or present. This is often the case for His–His dyads, whose conservation is often important in catalysis, while configuration can vary dramatically if allowed. It is also generally accepted that in the case of a His–His dyad, these two residues will exhibit significant parallel planarity relative to each other. His–His dyads are unique in that it is probably best practice to apply two atom–atom constraints between the two residues to both reduce ligand–ligand distance sampling, but also reduce the degree to which the plane of these two residues can

become orthonormal. These constraints can generally be loose (± 1.0 Å), but may need to become more strict depending on the extent of the charge relay network such as in the case of a triad and so on, if the orientation of these residues could have significant impact on the QM region energy. In general, with all these considerations, the number of constraints should be as small as possible, to avoid biases in sampling.

3.3 Setting Up the QM–DMD Region

The QM–DMD region is shared by QM and DMD in QM/DMD. In other QM/MM methods, it would be the analog of a pure-QM region. QM/DMD has been designed to work with *Turbomole* using DFT with the resolution of identity approximation for acceleration, and COSMO implicit solvation scheme. However, QM/DMD has been coded in such a way that the use of any QM software package or method could be feasible. It is also possible to utilize explicit point charge embedding, if desired. Explicit solvation, apart from a few QM water molecules present at the active site and participating in the chemistry, should, again, not be used. This is because DMD is parameterized to include solvation implicitly, and additional solvent would create double counting.

The choice of the shared QM–DMD region should be heavily guided by chemical intuition. It will only consist of a small fraction of the overall protein. In the case of **2V5W**, the QM–DMD region should encompass the active site metal, its coordinating ligands, the substrate, the coordinated water molecule, and any other ligands that may play an important role in catalysis. There are no hard rules for what should be included in the QM–DMD region. In **2V5W**, the substrate is quite large, and it may be reasonable to truncate it at some point in the chain so that computational resources are not wasted on extraneous parts of the substrate that do not participate in catalysis. In general, QM/DMD allows for significantly expanded QM–DMD regions, but chemical accuracy must be appropriately balanced with the available computational resources. QM–DMD regions on the order of 100–200 atoms are common and tractable.

It is very likely that the choice of QM–DMD region will bisect some covalent bonds. If left unmodified, covalent bond cleavage could lead to an excessive buildup of charge around the dangling bonds and lead to artifacts. There are a variety of ways to ameliorate this issue; however, QM/DMD caps dangling bonds with hydrogen atoms. Both atoms along the dangling bond are frozen in Cartesian space to reflect the attachment to the rest of the protein structure, which remains untreated with QM.

Before the QM step, the DMD ensemble of structures is clustered, as described in detail later, and a single structure is chosen to undergo a QM relaxation. During the QM step, the selected structure undergoes a geometric relaxation, either to convergence or to a predefined maximum number of optimization steps. The relaxed QM geometry is then reinstalled back into the protein. It is important to note that the QM–DMD region may consist of a significant number of atoms that are not frozen in the DMD scheme. This flexibility is important as it allows QM/DMD to relay structural information in this boundary region between the QM and DMD integration steps. QM bond lengths are also used to reparameterize the DMD force field for the QM–DMD region on the fly, by recentering the potential wells.

4. RUNNING QM/DMD AND DETAILS OF THE PROCEDURE

This section will cover, in more technical detail, how QM/DMD actually interfaces the QM and DMD integrations steps. It will describe the manner in which QM/DMD selects appropriate configurations with which to continue additional QM/DMD simulations, as well as some of the metrics that can be used to probe convergence of geometries. It will also highlight the actual processes that take place during the QM and DMD steps. Almost all of the operations described here are automated and take place without user oversight. We will discuss how to use the produced sets of configurations for further studies.

4.1 The Blueprint of the Algorithm

Once QM/DMD has been properly setup, a user needs to define only a few additional parameters, such as the starting annealing temperature, duration of the annealing step in DMD, and duration of the equilibration stage. After that, QM/DMD software is ready to run. The QM/DMD engine will perform the necessary steps in line with its original design philosophy (Fig. 6). First, a QM optimization is performed to relax any poor contacts at the active site. After, this new active site configuration is installed back into the PDB, and a DMD integration step is performed. The enzyme structure begins sampling at an elevated temperature and cooled until it reaches the desired temperature specified by the user at the beginning of the QM/DMD simulation, similar to simulated annealing. Once the target temperature is reached, equilibrium DMD is carried out for a length of time also specified by the user and data collected. Once DMD has concluded, the ensemble is

Fig. 6 Schematic overview of the steps QM/DMD goes through in generating publishable configurations for mechanistic analysis. Final configurations are highlighted in *purple* (dark gray in the print version). QM optimizations are represented in *light blue* (gray in the print version), and steps present only in Iterations >0 are represented in *orange* (light gray in the print version) (DMD sampling and handling steps).

clustered, several configurations are selected, their QM–DMD regions are extracted and capped with hydrogens, and single-point QM energies are calculated. The most suitable configuration is then chosen based on both QM and DMD energies, and its geometry is optimized using QM. The optimized QM–DMD region is installed back in the protein and the cycle begins anew. Each step will be described in more detail as we proceed through this section.

QM/DMD subdivides its progress into what have been dubbed Iterations. Each Iteration consists of a DMD integration step, followed by a QM integration step, yields the energy, and coordinates of a DMD-sampled, QM- protein configuration. As QM/DMD progresses, these iterations will, on average, decrease in total energy and eventually converge toward a minimum geometry. Looser constraints during the setup will produce greater fluctuations in energy and will likely require a greater number of Iterations

to reach an appropriate minimum. Overly rigid constraints may produce less energetic fluctuations at the expense of potentially preventing the protein from reaching a lower energy structure where catalysis may be more favorable. A balance must be struck between constraint rigidity and desired sampling. It is also generally recommended to increase the temperature at the annealing, and the duration of the simulation, if the starting configuration is suspected to be far from equilibrium. However, convergence can generally be reached within 20–100 Iterations.

4.2 Initial Iteration (Iteration 0)

The QM/DMD engine's initial Iteration is not a complete Iteration, in the sense that it does not consist of both a DMD and QM steps. At the beginning of the simulation, bonding contacts have not been relaxed in either the QM–DMD or DMD regions, and so it is important to relax this configuration with priority given to the active site geometry. For this reason, the initial Iteration (Iteration 0) consists only of a QM gradient-following relaxation on the QM–DMD region. Once an optimized QM geometry is acquired, it is reinstalled back into the enzyme structure, and subsequent Iterations can take place using this as an initial reference geometry.

4.3 Full Iterations (Iterations 1 Through *N*)

All subsequent Iterations after Iteration 0 utilize both a DMD and QM integration step, in that order. For a given set of Iterations, the lowest energy structure is used as the reference structure. This means that, if the simulation takes the system toward a lower energy conformation, this new conformation will be used to start a new iteration; if, however, a particular Iteration produces a higher energy species, this species is retained for the statistics of the ensemble, but the next Iteration will not start from it, and will use the lowest energy structure instead. This is implemented for equilibrium sampling near a minimum. This option can be turned off, if desired, with a single keyword in the input file. For example, if one might wish to cross low free-energy barriers and access new minima, the option must be turned off.

DMD begins with slowly cooling the structure from an elevated to the target temperature, and then sampling takes place for a specified period of time. The DMD step within each Iteration produces approximately 10,000 configurations, of which every 10th configuration is saved by default, producing 1000 saved snap-shots. To produce the most rigorous results, QM single points would be calculated for each of the configurations along

the DMD trajectory. However, due to the amount of sampling DMD performs, this quickly becomes intractable. Instead, a special selection scheme is utilized to sample likely candidates and produce reliable convergence among Iterations. For each configuration saved from the DMD trajectory, the Kabsch RMSD (Kabsch, 1976) is calculated. Configurations with significant similarities are then clustered together into bins of configurations. The maximum number of clusters can be defined prior to executing QM/DMD, allowing the user to specify the degree of granularity implemented for the QM/DMD engine. A typical recommendation would be to use three to five such bins. For each bin, either the structure closest to the centroid, or the lowest energy structure in the bin, or both, can be chosen as the representative geometries for QM single-point calculations to follow. A representative image of this process can be found in Fig. 7.

Representative geometries' QM–DMD regions are then extracted, capped with hydrogens, and evaluated with single-point QM calculations. The DMD energy and QM energy for each structure are combined, to choose the best candidate structure. In traditional QM/MM schemes, complex coupling terms are often employed to provide balance between these

Fig. 7 Circled regions represent geometrically similar clusters based on Kabsch RMSD. The configurations marked by a "x" represent the chosen representative configurations based on centroid proximity or lowest energy structure.

two terms, but in QM/DMD these two energies are weighted equally. Thus the total QM/DMD energy becomes

$$E_{\text{total}} = \frac{1}{2}(E_{\text{DMD}} + E_{\text{QM}})$$

where E_{total} is the total energy, E_{DMD} is the energetic contributions from DMD, and E_{QM} is the energetic contributions from the QM routine. The total energy is the means by which QM/DMD evaluates which configuration to proceed with for subsequent treatment. Note that it is not the true total energy of the system, due to double counting in the shared QM–DMD region.

The lowest energy structure chosen from cluster single points undergoes a geometric optimization at the QM level, with the constraints representing its attachment to the rest of the protein being observed. After this, the QM–DMD region is reinstalled into the protein, the QM–DMD boundary shrinks back to the QM-only region, and the wells in the DMD potential are recentered in accord with QM-optimized bond length. The system is ready for the DMD step of the next Iteration, when QM information can propagate to the rest of the protein.

4.4 A Simple Analysis of Convergence

An essential goal for QM/DMD is to quickly converge to a reasonable set of geometries that would exist at room temperature and neutral conditions. The simplest means to evaluate convergence over a set of Iterations is to calculate the RMSD between them and compare fluctuations in the total energy. All-atom RMSD of the active site (QM–DMD region), and backbone-only RMSD of the entire protein are monitored. Additionally, the convergence in terms of DMD and QM energies is evaluated. Note, however, that the DMD energy fluctuations can be very large (ca. 20 kcal/mol) and should not be considered on par with the QM energies, which are much more narrowly distributed. This arises because of the discontinuous form of the DMD potential: upon a small change in the protein structure, some interactions might be turned on or completely off, in accord with the square-well potentials. This produces large fluctuations in energy. This behavior of DMD is normal. [Fig. 8](#) shows an example of potential RMSD and energy curves that could be expected from a QM/DMD run.

Once a converged set of geometries has been reached, one can then use a set of these geometries, or the lowest energy conformer for further QM analysis (ie, a mechanistic study). The spread in computed values of interest,

Fig. 8 Comparison of QM and DMD energies (*top*) and RMSDs (*bottom*) plotted against Iteration number. The combination of a flat RMSD and total energy curve seen here is indicative of geometric convergence.

such as atomic charges or reaction barriers, will represent the ensemble. Further averaging could be done using Boltzmann statistics.

4.5 Tasks for Which QM/DMD Is Best Suited

QM/DMD has been tested and shown to perform well for such tasks as recapitulation of natural protein structures (Sparta et al., 2012), recovery of these structures back to the equilibrium after moderate distortions such as temporary removal and reinstallation of the metal (Sparta et al., 2012), prediction of structural changes upon mutagenesis (Sparta et al., 2012), flexible docking of substrates to metalloproteins that involve motion of large protein parts such as loops (Valdez, Sparta, & Alexandrova, 2013), predictions of protein structure after replacement of the metal or change of its oxidation state and accompanying changes in the coordination sphere geometry and number of ligands (Nedd, Redler, Proctor, & Dokholyan, 2014; Sparta et al., 2012; Sparta, Valdez, & Alexandrova, 2013; Valdez, Gallup, & Alexandrova, 2014), and predictions of protein rearrangements after removal of the metal in cases where the protein does not completely unfold, or the use of a short sequence within 20–30 amino acids (Nedd et al., 2014). Structures produced by QM/DMD are in good agreement with quality X-ray structures, if available (Sparta et al., 2012). For the sets of structures produced by QM/DMD, calculated parameters, such as reaction barriers, are found in good agreement with the experiment (Sparta et al., 2012, 2013; Valdez et al., 2013).

There are also tasks for which QM/DMD is not adequate. Predictions of the absolute values of the reduction potential of the metal cannot be satisfactorily computed with QM/DMD. For these types of calculations, one can use QM/DMD structures and improve the level of QM theory by increasing the basis set, and including the electrostatic embedding. Likewise, small (within 50 mV) changes in the reduction potential, for example, upon mutations around the active site, are not well captured (Sparta et al., 2012). Additionally, the use of QM/DMD for the full metalloprotein folding is not recommended. For small metalloproteins, folding possibly could be done, but this has not been fully tested yet.

4.6 Permissions, Copyrights, and Utilities

QM/DMD is available free of charge. However, in the current implementation it uses Turbomole for the QM calculations, and also DMD, both of which are proprietary software. The user is responsible for obtaining

individual licenses for these programs. If one desires to use QM/DMD in research, a request with a short justification must be submitted to the developers (Anastassia Alexandrova, ana@chem.ucla.edu, and Nikolay Dokholyan, nikolay_dokholyan@med.unc.edu). Every publication using QM/DMD must cite *Biophysical Journal*, 2012, 103, 767–776. In addition, in order to make QM/DMD easy to use, we provide a script for setting up a simulation for any metalloprotein starting from a PDB file. It takes care of troubleshooting, or points at possible problems with the provided PDB file, such as presence of unresolvable geometric clashes or the absence of a metal site. We also accompany the QM/DMD distribution with the user manual and are willing to provide reasonable amount of advice for setting up the runs and interpreting the results.

5. CONCLUSIONS

QM/DMD is a hybrid engine designed specifically with biomolecules in mind. It combines DMD with DFT, although other ab initio methods can feasibly be employed, to provide for fast sampling of a protein. The use of DMD over more traditional MD methods overcomes many of the computationally costly obstacles presented by the need for explicit solvation, and algorithmic challenges notorious for QM/MM, such as boundary force matching. QM/DMD provides a means to acquire converged ensembles of geometries for biomolecules at a finite temperature, in a short period of time, for a subsequent more focused study by the user.

REFERENCES

- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank (www.rcsb.org). *Nucleic Acids Research*, 28, 235–242.
- Chung, L. W., Li, X., & Morokuma, K. (2010). Modeling enzymatic reactions in metalloenzymes and photobiology by quantum mechanics (QM) and quantum mechanics/molecular mechanics (QM/MM) calculations. *Quantum Biochemistry*, 1, 85–130.
- Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E., & Shakhnovich, E. I. (1998). Discrete molecular dynamics studies of the folding of a protein-like model. *Folding & Design*, 3, 577–587.
- Kabsch, W. (1976). A solution of the best rotation to relate two sets of vectors. *Acta Crystallographica. Section A*, 32, 922–923.
- Klamt, A., & Schüürmann, G. (1993). COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society, Perkin Transactions 2*, 2, 799–805.
- Marks, P. A. (2007). Discovery and development of SAHA as an anticancer agent. *Oncogene*, 26, 1351–1356.

- Nedd, S., Redler, R. L., Proctor, E. A., & Dokholyan, N. V. (2014). Cu, Zn-superoxide dismutase without Zn is folded but catalytically inactive. *Journal of Molecular Biology*, 426, 4112–4124.
- Senn, H. M., & Thiel, W. (2009). QM/MM methods for biomolecular systems. *Angewandte Chemie International Edition*, 48, 1198–1229.
- Sparta, M., & Alexandrova, A. N. (2012). How metal substitution affects the enzymatic activity of COMT. *PloS One*, 7, e47172.
- Sparta, M., Shirvanyants, D., Ding, F., Dokholyan, N. V., & Alexandrova, A. N. (2012). Hybrid dynamics simulation engine for metalloproteins. *Biophysical Journal*, 103, 767–776.
- Sparta, M., Valdez, C. E., & Alexandrova, A. N. (2013). Metal-dependent activity of Fe and Ni acireductone dioxygenases: How two electrons reroute the catalytic pathway. *Journal of Molecular Biology*, 245, 3007–3018.
- Valdez, C. E., Gallup, N. M., & Alexandrova, A. N. (2014). Co^{2+} acireductone dioxygenase: Fe^{2+} mechanism, Ni^{2+} mechanism, or something else? *Chemical Physics Letters*, 604, 77–82.
- Valdez, C. E., Sparta, M., & Alexandrova, A. N. (2013). The role of the flexible L43-S54 protein loop of the CcrA metallo-beta-lactamase in binding structurally dissimilar beta-lactam antibiotics. *Journal of Chemical Theory and Computation*, 9, 730–737.
- Vannini, A., Volpari, C., Gallinari, P., Jones, P., Mattu, M., Carfi, A., et al. (2007). Substrate binding to histone deacetylases as revealed by crystal structure of Hdac8-substrate complex. *EMBO Reports*, 8, 879.
- Warshel, A. (1997). Computer modelling of chemical reactions in enzymes and solutions. *Trends in Biotechnology*, 15, 439.
- Warshel, A., & Levitt, M. (1976). Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, 103, 227–249.