CHAPTER TWENTY-FOUR

Nitric Oxide Reactivity with Globins as Investigated Through Computer Simulation

Marcelo A. Marti,* Luciana Capece,* Axel Bidon-Chanal,† Alejandro Crespo,* Victor Guallar,‡ F. Javier Luque,† *and* Dario A. Estrin*

Contents

* Departamento de Química Inorgánica, Analítica y Química Física/INQUIMAE-CONICET, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Buenos Aires, Argentina
† Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Barcelona, Spain
‡ Catalan Institute for Research and Advanced Studies (ICREA), Computational Biology Program, Barcelona Supercomputing Center, Barcelona, Spain

Methods in Enzymology, Volume 437
ISSN 0076-6879, DOI: 10.1016/S0076-6879(07)37024-9

Abstract

This chapter reviews the application of classical and quantum-mechanical atomistic simulation tools used in the investigation of several relevant issues in nitric oxide reactivity with globins and presents different simulation strategies based on classical force fields: standard molecular dynamics, essential dynamics, umbrella sampling, multiple steering molecular dynamics, and a novel technique for exploring the protein energy landscape. It also presents hybrid quantum-classical schemes as a tool to obtain relevant information regarding binding energies and chemical reactivity of globins. As illustrative examples, investigations of the structural flexibility, ligand migration profiles, oxygen affinity, and reactivity toward nitric oxide of truncated hemoglobin N of *Mycobacterium tuberculosis* are presented.

1. Introduction

Computational techniques for modeling large biological molecules have emerged during the last decades as important tools to complement experimental information. The *in silico*-generated models and the information obtained by these studies are essential to complement the structural, energetic, and kinetic data of biological systems obtained through experimental methods and to shed light onto the relationships between structure and function. Computer simulations also afford a systematic and economical tool to analyze the dependence of the property of interest not only on the static structure (e.g., amino acid sequence), but also on dynamic behavior. Moreover, because of the increase in computer power and the accuracy of the models, *in silico* experiments are valuable to propose new hypothesis and to draw biologically relevant conclusions about the molecular mechanisms that operate in biological systems (Karplus and Petsko, 1990; Leach, 2001).

Modeling of biological processes that do not involve formation and/or breaking of chemical bonds can be achieved by employing classical force fields. Among the most widely used biomolecular force fields are AMBER (Perlman *et al.*, 1995) and CHARMM (MacKerell *et al.*, 1998). Although the timescales accessible in atomistic simulations with present hardware technology are still limited to the nanosecond/microsecond range, the predictive power of simulation schemes applied to biomolecules has increased by the development of enhanced sampling techniques.

However, reactive processes are a key ingredient in understanding nitric oxide (NO) physiology. In these cases, one has to resort to quantum

mechanical (QM)-based schemes. There are two main strategies for the investigation of reactive chemical processes in biomolecules. The first one consists of performing QM electronic structure calculations on adequate model systems, which are chosen to represent the main features of the active site and eventually the most relevant region of the surrounding environment. The second strategy is to employ hybrid quantum mechanical/molecular mechanical (QM-MM) schemes (Capece *et al.*, 2006; Crespo *et al.*, 2003; Elola *et al.*, 1999; Friesner and Guallar, 2005; Guallar and Friesner, 2004; Warshel and Levitt, 1976), which represent the chemically relevant part of the system at the QM level using different strategies (valence bond theory, semiempirical, or Hartree–Fock methods and density functional theory), whereas the rest of the system is treated at the less expensive MM level.

This work describes the implementation of several computational methods partly developed by our groups, based on the schemes mentioned earlier, to investigate the molecular basis of NO reactivity with heme proteins. Selected examples are given to illustrate the capabilities of these methods. Finally, a critical analysis of the described computational schemes is presented.

2. Molecular Dynamics (MD) Methods

Molecular dynamics simulations have become a powerful tool in computational biology and are widely used to obtain information about processes involving conformational changes. In MD simulations, atoms are treated as charged spheres connected by springs and the whole system is described by a potential energy function, that is, the so-called force field, composed of various terms, each of which represents a portion of the interactions present in the system. The basic terms describe bond distances, angles, dihedrals, electrostatic (Coulombic), and other nonbonded interactions between atoms, and there might be also additional terms used to describe specific interactions such as hydrogen bonds or nuclear magnetic resonance restraints.

The AMBER package (Perlman *et al.*, 1995), for example, uses a simple but very efficient potential energy function, given by

$$\begin{aligned} U(R) = & \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 \\ & + \sum_{dihedrals} \frac{V_n}{2} (1 + \cos[n\phi - \gamma]) \\ & + \sum_{i<j}^{atoms} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} \right) + \sum_{i<j}^{atoms} \frac{q_i q_j}{\varepsilon R_{ij}} \end{aligned} \tag{24.1}$$

In Eq. (24.1), the first three terms describe the so-called bonded interactions, namely bond stretching, angle bending, and dihedral torsions, respectively. The last two terms correspond to van der Waals and electrostatic potentials, the so-called nonbonded interactions. The parameters K_r, K_θ, and V_n represent the force constants for stretching and bending of bonds, and the barrier for torsional angles, respectively. The subindex *eq* specifies the values of bond length and angle at the equilibrium geometry, whereas n and γ denote the periodicity and phase angle of torsions. Finally, A and B represent van der Waals parameters, q denotes the atomic partial charges, ϵ is the permittivity, and R_{ij} stands for interatomic distances. Particles in a given system must be defined with fixed parameters to account for the complete set of bonded and nonbonded interactions.

Given a set of coordinates for the particles of the system, the dynamic behavior can be examined from the trajectory obtained by solving Newton's equations of motion numerically with finite difference methods (Leach, 2001).

2.1. Setup of the system

When starting the study of a given protein through MD simulations, one has to choose the starting structure and the force field. The starting structure is typically taken from the crystal structure of the protein, when available. When several structures are available, that solved at the highest resolution is usually the best choice. However, care must be taken in selecting the structure that most likely represents the desired simulation conditions. For example, myoglobin (Mb) is found crystallized in several oxidation and coordination states (Fe^{III}, Fe^{II}, Fe^{II}-O_2, Fe^{II}-CO, Fe^{II}-NO, among others) (Brunori *et al.*, 2004, Chu *et al.*, 2000, Copeland *et al.*, 2003, 2006), and the best choice for simulation of the nitrosyl protein would be the crystallized Mb-NO structure (Copeland *et al.*, 2003). If the precise desired state is unavailable, the most similar structure should be taken as the starting point, and the desired state is built through *in silico* modification. If the crystal structure of the system is not available, homology modeling is an option (Blundell, 1987). For a biochemical study, however, the level of identity/similarity in the active site needs to be of the order of 70–80%.

Once the initial structure and the force field have been chosen, the next step consists in setting up the system. As X-ray techniques usually do not provide information about the hydrogen atoms, positions, they have to be added to the experimental structure. Most MD packages have an automated H-adding function, although some user specifications are needed. Of particular relevance in globins is the histidine case, as this residue has three representations in most force fields. Two of them are neutral, one protonated in the Nϵ and the other protonated in the Nδ, whereas the third form is a positively charged doubly protonated histidine. Many globins such as

Mb or hemoglobin (Hb) have two conserved histidines: the proximal histidine F8 is nonprotonated, as Nϵ is coordinated to the iron of the heme group, and the distal histidine E7 is a protonated residue with the NϵH group capable of forming a hydrogen bond with ligands. Clearly, this suffices to remark that a close inspection of the possible interactions is essential to describe the system properly.

The last step in the setup of the system is the choice of the solvent, typically water, and the addition of ions to mimic a given ionic atmosphere. MD packages are able to perform MD simulations in different solvation conditions, such as implicit water with the generalized Born approximation (Onufriev *et al.*, 2004), a cluster of explicit water models, and the insertion of the protein in a box containing water molecules, employing periodic boundary conditions to represent the bulk solvation.

2.2. Equilibration

The equilibration protocol is critical to allow for initial relaxation and thermalization yielding an equilibrated structure at the desired simulation temperature. Most equilibration protocols perform an initial optimization of the system, followed by a slow heating up to the desired temperature using, for instance, the Berendsen thermostat (Berendsen *et al.*, 1984). Similarly, the pressure can be kept fixed at a given value by coupling the system to a barostat (Berendsen *et al.*, 1984). In our simulations, heating is generally performed in about 100 to 200 ps at constant volume. Once the system is equilibrated, the different MD runs are performed at constant temperature and pressure (NPT ensemble), as these conditions resemble the typical experimental situation.

2.3. Essential dynamics

This technique determines the essential motions of the simulated system that explain most of the structural variance detected during the trajectory (Amadei *et al.*, 1993). Technically, this implies diagonalization of the covariance matrix of the positional deviations given in Eq. (24.2), which affords a set of $3N$ (N = number of atoms in the system) eigenvectors and their associated eigenvalues:

$$C = \mathrm{cov}(\mathbf{x}) = \langle (\mathbf{x} - \langle \mathbf{x} \rangle)(\mathbf{x} - \langle \mathbf{x} \rangle)^T \rangle, \qquad (24.2)$$

where $\mathbf{x}$ denotes a $3N$-dimensional vector of all atomic coordinates, and <>stands for an average over time.

The eigenvectors can be considered to be $3N$-dimensional vectors representing the nature of the essential motions. Each eigenvalue represents

the percentage of structural variance explained by each motion. By using harmonic approximations, eigenvalues can be translated into frequencies, which indicate the softness of a given essential motion. The eigenvectors of one trajectory can be compared with those of another trajectory, deriving quantitative measures of the similarity between the essential motions of two independent trajectories [Eq. (24.3)]:

$$\gamma_{AB} = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}(v_i^A \cdot v_j^B)^2, \tag{24.3}$$

where n stands for the minimum number of eigenvectors, which explain more than a given threshold of the variance.

Essential dynamics calculations are very useful in identifying structural fluctuations in MD simulations. However, caution is needed in the analysis, as the technique is very sensitive to the extension of the trajectory and to numerical errors in the calculations and can neglect important local distortions in favor of general but perhaps less relevant movements. Additional sources of error exist in the definition of a common reference system for the trajectories and on the elimination of translational and rotational degrees of freedom of the molecule. It is very important to keep in mind that essential motions are detected only if they occur in the MD trajectory, but slow motions might be difficult to detect in current "state-of-the-art" simulations (5–10 ns). Caution and common sense are then necessary for a reasonable use of the very powerful essential dynamics tool.

2.4. Free energy profile calculations

In the study of proteins through MD, relevant information about a given process is obtained from the mean force potential (PMF) associated with the process (Leach, 2001). Examples of the potential applications of this tool are ligand migration through the protein matrix (Bidon-Chanal *et al.*, 2006) or the conformational change of the side chain movement of a residue between two different stable conformations (Marti *et al.*, 2006a). The PMF can be compared directly related to the experimental results, as it takes into account thermal and entropic effects.

If the free energy barriers are of the same magnitude as the thermal fluctuations, it is feasible to obtain the free energy profiles associated with a given process directly from classical MD simulations. In these cases, an adequate sampling of the relevant configurations may be achieved in accessible simulation times, and the free energy profile can be obtained by computing the probability distribution along the selected reaction coordinate, $\boldsymbol{P}(\xi)$:

$$-\beta A(\xi) \approx \ln[p(\xi)], \tag{24.4}$$

where $\beta^{-1} = k_B T$ is the Boltzmann constant times the temperature and $\boldsymbol{A}$ is the free energy.

To have an appropriate sampling in 5- to 10-ns simulation times, the barriers should be ≈2 kcal/mol. In cases where barriers are suspected to be higher, biased sampling is required to obtain the PMF. Two different biased sampling methods are now presented: umbrella sampling and steered molecular dynamics.

2.4.1. Umbrella sampling

This method (Torrie and Valleau, 1977) attempts to overcome the sampling problem by modifying the potential function so that the unfavorable states are sampled sufficiently. The potential function is modified by adding a weighting function that usually takes a harmonic form according to

$$E'(r) = E(r) + k(\xi - \xi o)^2, \tag{24.5}$$

where $E(r)$ is the potential energy of the protein for a given configuration, denoted by r, and ξ denotes a specific reaction coordinate.

For configurations that are far from the equilibrium state, the weighting function assumes large values and so the simulation using the modified energy function $E'(r)$ is biased along the reaction coordinate. An umbrella sampling calculation involves a series of stages (called simulation windows), each characterized by a particular value of the reaction coordinate. The PMF is then obtained by superposing the results obtained for all the series of windows. To obtain the PMF in the unbiased simulation, the PMF obtained from the simulation is corrected by subtracting the contribution due to the weighting function. Although the method itself seems to be implemented easily in multiprocessor parallel computers, several technical details must be taken into account: (i) two consecutive windows must overlap in order to correctly superimpose them, (ii) the force constant for each stage has to be chosen carefully in order to make an efficient sampling of the potential energy surface, and (iii) the initial thermalization simulation must be performed for all the windows.

2.4.2. Multiple Steering Molecular Dynamics (MSMD)

The multiple steering molecular dynamics approach, originally proposed by Jarzynski (1997) is based on the following relation between nonequilibrium dynamics and equilibrium properties,

$$\exp[-\Delta A(\xi)/k_B T] = \langle \exp[-W(\xi)/k_B T] \rangle, \tag{24.6}$$

where $W(\xi)$ is the external work performed on the system as it evolves from the initial to the final state along the reaction coordinate ξ.

In MSMD, the original potential is modified by adding a time-dependent external potential, usually harmonic, to the potential energy that moves the system along the reaction coordinate by varying the potential well according to

$$E'(r) = E(r) + k[\xi - (\xi_o + v\Delta t)]^2, \tag{24.7}$$

where v is the pulling speed that moves the system along the reaction coordinate.

The PMF is obtained by performing several MSMD runs, collecting the work done at each time step, and then averaging it properly, according to Jarzynski's equation [Eq. (6)]. Usually, the pulling speed is chosen so that the system moves smoothly, but faster than in a true reversible simulation (Crespo *et al.*, 2005a; Hummer and Szabo, 2001; Park and Schulten, 2004; Xiong *et al.*, 2006).

2.5. Protein energy landscape exploring

Although MD simulations have been widely used to study biomolecular systems, the modeling of long time dynamics still remains a challenge. In the last decade, there has been a significant effort toward the development of theoretical methods for protein structure prediction based on the use of rotamer libraries (Dwyer *et al.*, 2004). Jacobson and colleagues (2002) have developed a program for protein modeling using specialized sampling algorithms for side chain prediction. The sampling algorithms include the use of highly detailed rotamer state libraries for side chain conformational searching, hierarchical screening methods based on steric overlap and approximate electrostatics to rapidly eliminate obviously incorrect conformations, and a multiscale minimization algorithm, one to two orders of magnitude more efficient than conventional approaches. Using these technological advances in protein structure prediction methods, a new approach to study the protein energy landscape exploration (PELE) associated with long time dynamics events has been developed (Borrelli *et al.*, 2005). The heuristic algorithm for the PELE method involves consecutive iteration of three main moves: localized perturbation, side chain sampling, and minimization.

2.5.1. Local perturbation

The procedure begins with the generation of a local perturbation. When studying ligand diffusion in a protein, this first step involves a ligand perturbation. Initially, the ligand is treated as a rigid body. Hundreds of perturbations are generated within seconds and the one with the best energy is selected.

2.5.2. Side chain sampling

The algorithm proceeds by placing all side chains local to the atoms perturbed in step 1 using the algorithms described earlier (Jacobson *et al.*, 2002).

2.5.3. Minimization

The last step in every move involves minimization of a region including, at least, all residues local to the atoms involved in steps 1 and 2.

These three steps compose a move that is accepted (defining a new minima) or rejected based on a Metropolis criterion for a given temperature. The collection of accepted steps forms a stochastic trajectory. The changes are propagated to the nonlocal environment by means of diffusion of the local perturbed area and by intercalated longer range (larger local area) steps. Many trajectories are typically run in parallel with a collective task. Processors ahead in the collective task will spawn the coordinates to those left behind. The task protocol allows for the simulation of reaction coordinates or to focus any search in a given trajectory.

2.6. Heme group parameters

A critical point in MD simulations of heme proteins is the parameters associated with the heme group. Typically, all bonded and Lennard–Jones parameters can be obtained from the force field. However, it is crucial to select the atomic charges very carefully because they are essential to obtain accurate results, taking into account the different states of the heme group. Typically, partial charges for the heme moiety in the different coordination/oxidation states are obtained according to the standard protocol, which consists in performing Hartree–Fock or density functional theory (DFT) calculations for the isolated heme group and subsequently deriving electrostatic potential-fitted charges (Wang *et al.*, 2000). Other electronic structure methods can also be used to determine the electrostatic potential.

3. Quantum Mechanical-Molecular Mechanical Methods

Several electronic structure computations can be used to study reactive processes in proteins. This chapter shows results obtained with a QM-MM scheme at the DFT level with the SIESTA code. DFT is, to our knowledge,

the method with the best cost-benefit ratio, as it provides quite accurate results at a reasonable computational cost (Soler *et al.*, 2002). The SIESTA code has shown excellent performance for medium and large systems, including biomolecules, and heme models in particular (Capece *et al.*, 2006; Crespo *et al.*, 2003, 2005b; Fernandez *et al.*, 2005; Marti *et al.*, 2004, 2005, 2006a,b). Basis functions consist of localized (numerical) pseudo-atomic orbitals, projected on a real space grid to compute the matrix elements. In the examples presented in this chapter, basis sets of double zeta plus polarization quality were employed for all atoms. Calculations were performed using the generalized gradient approximation to the exchange-correlation energy proposed by Perdew and colleagues (1996). Such a combination of exchange-correlation functional, basis sets, and grid parameters has been widely validated for heme models (Capece *et al.*, 2006; Crespo *et al.*, 2005b; Fernandez *et al.*, 2005; Marti *et al.*, 2004, 2005, 2006a,b).

3.1. Selection of the QM subsystem

Selection of the QM subsystem in simulations of NO reactivity with globins is quite intuitive. In order to take into account the chemical reactivity of the heme group, the iron porphyrinate and the bound ligands, e.g., the proximal histidine and the distal O_2, NO, etc., must be selected as the quantum subsystem. In cases where other residues play a relevant role in the chemical reaction, they should also be included in the QM subsystem. The rest of the system is treated at the MM level, and the interface between QM and MM portions is treated using several formalisms, such as the scaled position link atom method (Eichinger *et al.*, 1999), adapted to the SIESTA QM-MM code. This method completes the valence of the QM system with hydrogen atoms placed along the QM carbon atom–MM carbon atom bond.

3.2. Quantum mechanical-molecular mechanical optimizations

The computational cost of QM-MM calculations limits the feasibility to explore the dynamics of the system. Instead, geometry optimizations of the whole system are less expensive and afford the energy of the potential energy minimum. To obtain an initial structure representative of the protein conformation in physiological conditions for QM-MM optimization, a conformational trapping protocol is performed. This is done by starting from an equilibrated snapshot of the protein obtained in a classical MD run and cooling the system down slowly to 0 K. As shown later, this method must be used starting from different snapshots to obtain representative structures of the protein. The cooling process is followed by QM-MM geometry optimizations using a conjugate gradient algorithm. Because we

are only interested in the changes that take part in the active site, only residues located less than 10 Å apart from the heme reactive center are allowed to move freely in QM-MM runs. This protocol yields structures of the protein–ligand complex that can be compared with experimental data.

3.3. Binding energy calculations

A relevant feature of a globin is its ligand-binding properties, which can be examined by the ligand-binding energy (ΔE_L) determined from Eq. (24.8). Using Eq. (24.8), no entropy contribution is included in calculations. However, for comparative analysis, it can be omitted in a first approximation, and the QM-MM interaction energies are then expected to be useful.

$$\Delta E_L = E_{prot-L} - (E_{prot} + E_L), \tag{24.8}$$

where $E_{\text{prot-L}}$ is the energy of the ligand–protein complex obtained from a QM-MM optimization, E_{prot} is the QM-MM energy of the free protein, and E_{L} is the QM energy of the isolated ligand.

It is worth mentioning that the DFT at the generalized gradient approximation level exhibits a bias in the description of the spin-state energetics of iron-porphyrins, in general favoring low-spin configurations. In particular, DFT energies for the free heme, whose electronic structure corresponds to a high-spin state, are somehow overestimated with respect to the ligand-bound form, which presents a low-spin ground state (Deeth and Fey, 2004, Franzen, 2002). However, even though the estimates for ΔE may be sometimes above the experimental values, the predicted trends are in agreement with the experimental results.

3.4. Reaction pathway search

As noted earlier, QM-MM calculations permit identifying the minimum energy point on the potential energy surface through geometry optimizations. However, much chemical interest lies in the free energy barrier between stable minima or for a given chemical reaction between reactants and products. The energy barrier is defined as the maximum energy along the minimum energy path (i.e., the reaction path) that connects the two minima. The conversion of one minimum energy (reactants) into another (products) may sometimes occur primarily along just one or two coordinates, denoted reaction coordinates. In such cases, an approximation to the reaction pathway, called restrained minimization, can be obtained by changing these coordinates gradually, allowing the system to relax at each stage while keeping the chosen coordinates fixed. The point of higher

energy on the path is an approximation to the transition state, and structures generated during the course of the calculation can be considered to represent a sequence of points on the interconversion pathway (Leach, 2001).

To carry out restrained minimizations, an additional term is added to the potential energy, as noted in Eq. (24.9):

$$V_R = \frac{1}{2}k(\xi - \xi_0)^2, \tag{24.9}$$

in which k is an adjustable force constant, ξ is the value of the reaction coordinate, and ξ_0 is the value of the reaction coordinate for a particular configuration.

The path is constructed by (i) unrestrained minimization of reactant or product to generate an initial configuration for the reaction path, (ii) adding V_R to the potential energy, varying ξ_0, and finally (iii) performing a series of energy minimizations at each step. The actual energy of each configuration along the reaction path is obtained by subtracting the restraint term from the total energy.

4. Illustrative Examples

4.1. Structural flexibility of globins as studied by MD simulations: *Mycobacterium tuberculosis* truncated hemoglobin N

Flexibility of proteins can play an important role in modulating ligand diffusion and binding to the active site. This is the case of the truncated hemoglobin N (trHbN) from *M. tuberculosis*, which is a small globin that presents NO-dioxygenase activity in its oxygenated form (Ouellet *et al.*, 2002). Thus, the protein contributes to the detoxification of NO through the conversion to nitrate anion [Eq. (24.10)]:

$$\mathrm{Fe(II)O_2 + NO \rightarrow Fe(III) + NO_3^-}. \tag{24.10}$$

To carry out the reaction, the protein must ensure the presence of oxygen in the active site before entrance of NO. This is achieved by means of a dual-path ligand-induced regulation mechanism underlying diffusion of both O_2 and NO through two distinct branches of the ligand migration tunnel. It has been proposed that this mechanism might control diatomic ligand migration to/from the heme as the rate-limiting step in NO conversion to nitrate (Bidon-Chanal *et al.*, 2006). The short branch of the tunnel (around 8 Å) is mainly defined by residues pertaining to helices

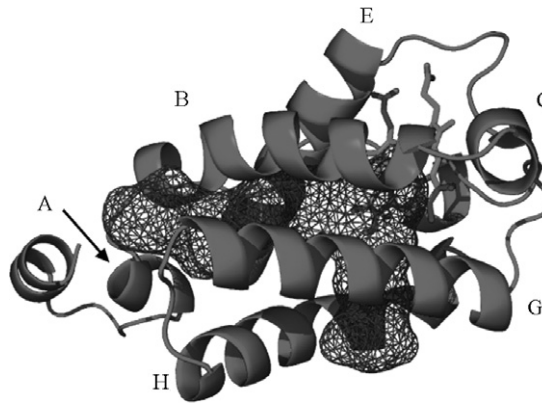

Figure 24.1 Long and short tunnel branches in *M. tuberculosis* trHbN. Helix names are indicated in capital letters.

G and H, and the large branch (around 20 Å) is mainly formed by residues in helices B and E (Milani *et al.*, 2001)(Fig. 24.1).

Binding of O_2, which accesses the heme cavity through the short branch, modulates the specific pattern of hydrogen-bonded contacts between residues TyrB10 and GlnE11, which in turn alters the global dynamics of the protein, favoring the relative displacement of helices B and E. The combined effect of both local conformational changes in the TyrB10–GlnE11 pair and global structural changes in the protein facilitates the conformational transition. This transition then leads to opening of the tunnel long branch, which consequently allows the diffusion of small ligands (NO) to the O_2-bound heme.

Inspection of the essential dynamics of the oxygenated and deoxygenated forms of trHbN shows clear differences that can be related to the distinct migration pathways of O_2 and NO (Crespo *et al.*, 2005b). As shown in Fig. 24.2, the major motions affect helixes C, G, and H in deoxy-trHbN. In contrast, the major motions in oxy-trHbN involve the relative displacement of helices B and E, which largely contribute to delineate the shape of the tunnel long branch. Finally, the difference in the main structural fluctuations of deoxy-trHbN and oxy-trHbN can be measured by means of a similarity index, which takes into account not only the nature of the essential movements, but also their contribution to the structural variance of the protein. Thus, for the 10 most relevant essential motions of the backbone skeleton in the core region of trHbN, which account for around 70% of the structural variance, the similarity index between deoxy and oxy forms of trHbN only amounts to 0.17. Overall, these findings reveal the different nature of the motions in the oxygenated and deoxygenated forms of the protein (Crespo *et al.*, 2005b).

Figure 24.2 Representation of structural distortions in the peptide backbone of trHbN associated with the first essential motion in (left) oxygenated and (right) deoxygenated forms of the protein.

5. Ligand Migration Profiles from MSMD and PELE Simulations: Exploring Ligand Entry Pathways in *M. tuberculosis* trHbN

From the plain MD simulations mentioned earlier, we observed two states for the long tunnel branch, where PheE15, which acts as a gate, adopts two conformational states corresponding to open and closed forms of the tunnel. Interestingly, the open state was only found in the oxygenated protein. Using MSMD, we computed the free energy profile for ligand migration along the two branches of the tunnel in the oxy and deoxy forms of trHbN (Bidon-Chanal *et al.*, 2006). As an illustrative example, results for the open and closed states of the tunnel long branch in the oxy-trHbN are shown later (see Fig. 24.3).

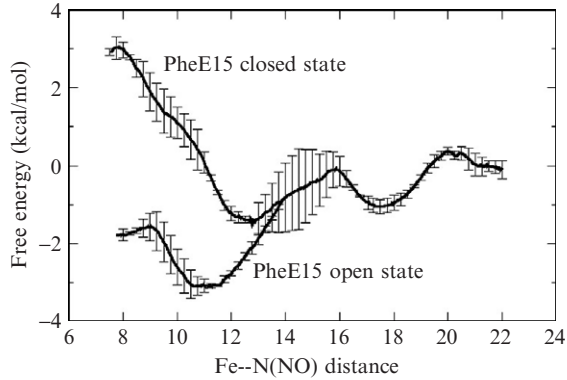

Figure 24.3 Free energy profiles for ligand diffusion along the tunnel long branch in open and closed states defined by the orientation of PheE15 in oxy-trHbN.

In the closed state, access of the ligand to the heme cavity is impeded by PheE15, whose side chain protrudes into the channel, and the ligand remains preferably in the secondary docking sites. However, in the open state, where the benzene ring of PheE15 lies parallel to the axis of the tunnel, the ligand is directed toward the active site, as the highest barrier is around 1.5 kcal/mol. The opening of the tunnel long branch in oxy-trHbN allowed us to explain why the NO detoxification rate, determined by NO migration into the oxy protein, is around 100 times faster than O_2 binding to the deoxy protein, as observed experimentally (Bidon-Chanal *et al.*, 2006).

As already mentioned in the computational methods section, an alternative strategy for the study of migration processes is to use the PELE algorithm. Figure 24.4 displays results for the carbon monoxide migration in deoxy trHbN. The CO migration was studied with 10 different runs. Each run produced a complete escape of the CO from the active site to the solvent and took about 120–192 CPU hours. Results show that the ligand first spends a large amount of cycles in the active site until Phe32 allows it to move into a cavity defined by Phe32 and Phe62. From this point, the CO

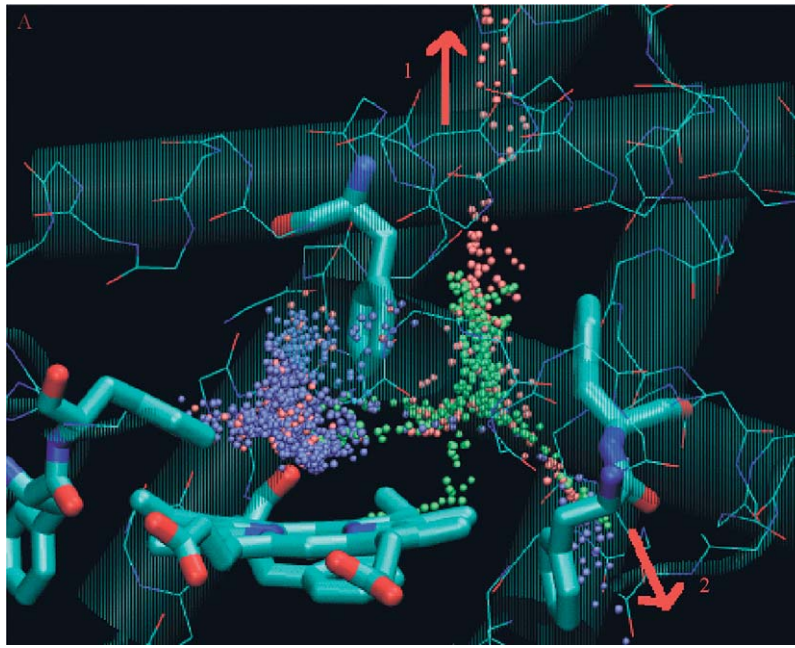

Figure 24.4 CO migration in *M. tuberculosis* trHbN as determined by the PELE algorithm. The heme group and relevant Phe residues are shown as sticks. Each CO position accepted by the algorithm along the run is shown as a dot. The arrows indicate the two migration paths.

ligand can escape through different pathways, shown in Fig. 24.4 with arrows and numbers 1 and 2. These exit pathways are in good agreement with the crystallographic and MSMD results mentioned previously.

5.1. Oxygen affinity of wild-type and mutant *M. tuberculosis* trHbN as compared to myoglobin

Quantum mechanical-molecular mechanical calculations provide useful information about structural features. The comparison of structures for different proteins, or between wild-type and single residue mutants, is useful in gaining insight into the relationship between structure and function. A key issue in globin function is ligand affinity, which is modulated by the association (k_{on}) and dissociation (k_{off}) rate constants (Olson and Phillips, 1997; Scott *et al.*, 2001).

The theoretical investigation of the ligand association process is rather complex. However, ligand release is mainly controlled by the Fe-ligand bond breaking and is therefore intimately related to the calculated ΔE of ligand release (Marti *et al.*, 2006b). As an example of how to study the structure and ligand affinity characteristics of several globins, Fig. 24.5 and Table 24.1 show the structural features and the binding energy of the oxy complexes of wild-type and mutant trHbN, as compared with those of Mb.

We selected the TyrB10→Ala mutant in trHbN and the HisE7→Gly mutant in Mb, as wild-type residues are the main residues affecting the complex structure and oxygen affinity. Thus, Fig. 24.5 clearly shows that TyrB10 in trHbN and HisE7 in Mb are the main hydrogen bond donors to the oxygen ligand. Mutation of these two residues reduces the oxygen

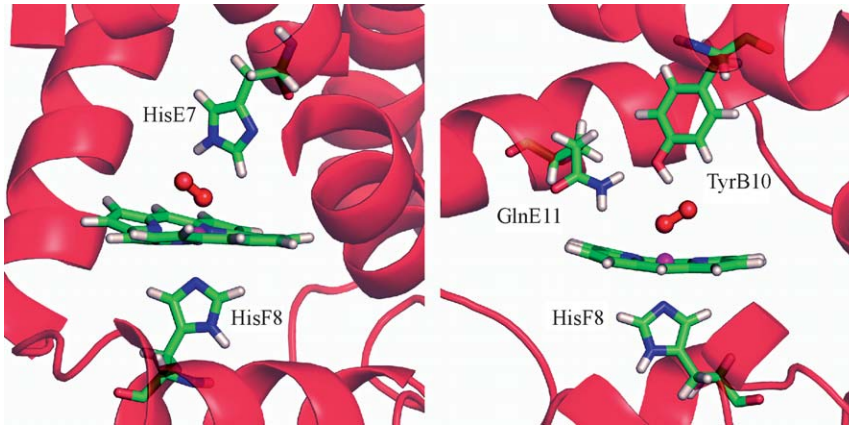

Figure 24.5 Optimized structures of the active site of (right) wild-type trHbN and (left) myoglobin.

Table 24.1 Geometrical parameters (angstroms and degrees), charge received by the O_2 (ΔqO_2) and donated by the proximal HisF8 Δq_{prox} (in *e*), O_2 binding energy ΔE (kcal/mol), and oxygen dissociation rate constant k_{off} (s^{-1})

	trHb N	trHb N TyrB10→Ala	Wild-type Mb	HisE7→Gly Mb
d Fe-O	1.84	1.86	1.84	1.76
d O-O	1.31	1.31	1.30	1.29
dFe-NεHisF8	2.06	2.09	2.18	2.18
dFe-NεHisF8 (free protein)	2.13	2.10	2.194	2.190
ΔqO_2	–0.360	–0.35	–0.214	–0.219
Δq_{prox}	0.160	0.160	0.146	0.142
ΔE	37.2	29.8	27.7	18.2
k_{off}	0.2[a]	45[a]	12[b]	1600[b]

[a] From Ouellet *et al.* (2002)
[b] From Scott *et al.* (2001)

affinity (see Table 24.1), which agrees with experimental data (Scott *et al.*, 2001). However, in trHbN TyrB10→Ala the affinity is still similar to that of wild-type Mb because of the presence of GlnE11, which now interacts strongly with the bound ligand. The presence of two hydrogen bond donors in trHbN (TyrB10 and GlnE11) also explains the higher O_2 affinity as compared to Mb. Another important finding is that hydrogen bond interactions and π-backbonding effects produce a net increase in the oxygen negative charge, ΔqO_2. The π-backbonding effect is also different in both proteins, being more important in trHbN, probably because of a closer proximal histidine (HisF8) (Capece *et al.*, 2006).

5.2. Chemical reactions between NO and globins: NO detoxification in *M. tuberculosis* trHbN

The toxic effects of NO can be reduced or even eliminated by the development of resistance mechanisms, which consist of the oxidation of nitric oxide with heme-bound O_2 to yield the harmless nitrate ion. NO scavenging functions have been observed in red blood cell hemoglobin, muscle Mb (Blomberg *et al.*, 2004; Eich *et al.*, 1996), neuroglobin within neuronal cells, and leghemoglobin, as well as in flavohemoglobins and truncated hemoglobins (Gardner *et al.*, 2000, 2006). The study of the chemical reaction itself in the protein environment requires a hybrid QM-MM treatment. This section shows results obtained through this methodology for the heme-controlled oxidation of NO in *M. tuberculosis* trHbN (Crespo *et al.*, 2005a).

We start by performing QM-MM optimizations of the oxy-trHbN with NO located in the active site. The resulting structure shows that NO is close to O_2 and interacts with Tyr33 and Gln58. Starting from this structure, we performed reaction pathway searches for the three reactions steps involved in the detoxification mechanism. In the first step, a peroxynitrite ion is formed from the attack of NO to the coordinated O_2, as noted in Eq. (24.11). Subsequently, the metal center catalyzes the two-step isomerization of peroxynitrite to nitrate through an iron-oxo intermediate [Eqs. (24.12) and (24.13)]:

$$\mathrm{Fe(II)} - \mathrm{O_2} + \mathrm{NO} \rightarrow \mathrm{Fe(III)[^{-}OONO]} \tag{24.11}$$

$$\mathrm{Fe(III)[^{-}OONO]} \rightarrow \mathrm{Fe(IV)} = \mathrm{O} + \mathrm{NO_2} \tag{24.12}$$

$$\mathrm{Fe(IV)} = \mathrm{O} + \mathrm{NO_2} \rightarrow \mathrm{Fe(III)[NO_3^{-}]} \tag{24.13}$$

The selected reaction coordinates for describing reactions (11), (12), and (13) were the O_2–N_{NO} distance, the O1–O2 distance, and the O1–N_{NO2} distances, respectively (O1 is the Fe-bound O_2 oxygen atom, and O2 is the distal O_2 oxygen atom). For comparative purposes, calculations were performed for the isolated model system (vacuum), for the same model system solvated in aqueous solution (by using a cluster of 1061 water molecules), in the wild-type protein, and in the Tyr33→Phe mutant. Results indicate that the protein catalyzes the chemical reactions, leading to the formation of nitrate with no significant contributions of the protein environment, as the potential energy profiles were almost barrierless in vacuum, in the aqueous solution, and in the protein (Crespo *et al.*, 2005b). This suggests that the rate-limiting process is ligand diffusion to the heme, as can be seen from the experimental bimolecular rate constant (745 $\mu M^{-1}s^{-1}$) (Ouellet *et al.*, 2002) typical of an almost diffusion-controlled process.

6. Conclusions

The present results make apparent the contribution of computational classical, quantum mechanical, and hybrid quantum-classical techniques to the investigation of NO reactivity with globins. In particular, it has been shown how structural and conformational features and ligand migration paths can be investigated with classical MD simulations in combination with advanced sampling tools to yield information about free energy barriers and possible secondary docking sites. However, it was shown that QM-MM schemes are specially suited to investigate ligand binding and other chemical processes. By covering different time and space scales, these two

frameworks, classical MD and QM-MM, complement each other in providing a complete picture of the protein function. The agreement with experimental data, when available, constitutes a stringent test to the reliability of these approaches. Computational simulation is a valuable tool even if it is often used to validate and contrast experimental observations because it offers a microscopic view, sometimes resolved in real time, that is very difficult to attain with any other method. With refinement of the methodology and available computers, classical and QM-MM simulations will be more and more at the center of the stage in this field, likely to become independent tools with the same hierarchy as any experimental technique.

ACKNOWLEDGMENTS

This work was partially supported by the University of Buenos Aires, Agencia Nacional de Promoción Científica y Tecnológica (Project PICT 25667), CONICET (PIP 02508), and the Spanish Ministerio de Educación y Ciencia (Grant CTQ2005-08797-C02-01/BQU). The Barcelona Supercomputer Center is kindly acknowledged for providing access to the Marenostrum supercomputer.

REFERENCES

Amadei, A., Linssen, A. B. M., and Berendsen, H. J. C. (1993). Essential dynamics of proteins. *Proteins* **17,** 412–425.

Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3684–3690.

Bidon-Chanal, A., Marti, M. A., Crespo, A., Milani, M., Orozco, M., Bolognesi, M., Luque, F. J., and Estrin, D. A. (2006). Ligand-induced dynamical regulation of NO conversion in *Mycobacterium tuberculosis* truncated hemoglobin-N. *Proteins* **64,** 457–464.

Blomberg, L. M., Blomberg, M. R., and Siegbahn, P. E. (2004). A theoretical study of myoglobin working as a nitric oxide scavenger. *J. Biol. Inorg. Chem.* **9,** 923–935.

Blundell, T. L., Sibanda, B. L., Sternberg, M. J. E., and Thornton, J. M. (1987). Knowledge-based prediction of protein structures and the design of novel molecules. *Nature* **326,** 347–352.

Borrelli, K. W., Vitalis, A., Alcantara, R., and Guallar, V. (2005). PELE: Protein Energy Landscape Exploration. A novel Monte Carlo based technique. *J. Chem. Theory Comput.* **1,** 1304–1311.

Brunori, M., Bourgeois, D., and Vallone, B. (2004). The structural dynamics of myoglobin. *J. Struct. Biol.* **147,** 223–234.

Capece, L., Marti, M. A., Crespo, A., Doctorovich, F., and Estrin, D. A. (2006). Heme protein oxygen affinity regulation exerted by proximal effects. *J. Am. Chem. Soc.* **128,** 12455–12461.

Chu, K., Vojtchovsky, J., McMahon, B. H., Sweet, R. M., Berendzen, J., and Schlichting, I. (2000). Structure of a ligand-binding intermediate in wild-type carbonmonoxy myoglobin. *Nature* **403,** 921–923.

Copeland, D. M., Soares, A. S., West, A. H., and Richter-Addo, G. B. (2006). Crystal structures of the nitrite and nitric oxide complexes of horse heart myoglobin. *J. Inorg. Biochem.* **100,** 1413–1425.

Copeland, D. M., West, A. H., and Richter-Addo, G. B. (2003). Crystal structures of ferrous horse heart myoglobin complexed with nitric oxide and nitrosoethane. *Proteins* **53,** 182–192.

Crespo, A., Marti, M. A., Estrin, D. A., and Roitberg, A. E. (2005a). Multiple-steering QM-MM calculation of the free energy profile in chorismate mutase. *J. Am. Chem. Soc.* **127,** 6940–6941.

Crespo, A., Marti, M. A., Kalko, S. G., Morreale, A., Orozco, M., Gelpi, J. L., Luque, F. J., and Estrin, D. A. (2005b). Theoretical study of the truncated hemoglobin HbN: Exploring the molecular basis of the NO detoxification mechanism. *J. Am. Chem. Soc.* **127,** 4433–4444.

Crespo, A., Scherlis, D. A., Marti, M. A., Ordejón, P., Roitberg, A. E., and Estrin, D. A. (2003). A DFT-based QM-MM approach designed for the treatment of large molecular systems: Application to chorismate mutase. *J. Phys. Chem. B* **107,** 13728–13736.

Deeth, R. J., and Fey, N. (2004). The performance of nonhybrid density functionals for calculating the structures and spin states of Fe(II) and Fe(III) complexes. *J. Comput. Chem.* **25,** 1840–1848.

Dwyer, M. A., Looger, L. L., and Hellinga, H. W. (2004). Computational design of a biologically active enzyme. *Science* **304,** 1967–1971.

Eich, R. F., Li, T., Lemon, D. D., Doherty, D. H., Curry, S. R., Aitken, J. F., Mathews, A. J., Johnson, K. A., Smith, R. D., Phillips, G. N., Jr., and Olson, J. S. (1996). Mechanism of NO-induced oxidation of myoglobin and hemoglobin. *Biochemistry* **35,** 6976–6983.

Eichinger, M., Tavan, P., Hutter, J., and Parrinello, M. (1999). A hybrid method for solutes in complex solvents: Density functional theory combined with empirical force fields. *J. Chem. Phys.* **110,** 10452–10467.

Elola, M. D., Estrin, D. A., and Laria, D. (1999). Hybrid quantum classical molecular dynamics simulation of the proton-transfer reaction of HO- with HBr in aqueous clusters. *J. Phys. Chem. A* **103,** 5105–5112.

Fernandez, M. L., Marti, M. A., Crespo, A., and Estrin, D. A. (2005). Proximal effects in the modulation of nitric oxide synthase reactivity: A QM-MM study. *J. Biol. Inorg. Chem.* **10,** 595–604.

Franzen, S. (2002). Spin-dependent mechanism for diatomic ligand binding to heme. *Proc. Natl. Acad. Sci. USA* **99,** 16754–16759.

Friesner, R. A., and Guallar, V. (2005). Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. *Annu. Rev. Phys. Chem.* **56,** 389–427.

Gardner, A. M., Martin, L. A., Gardner, P. R., Dou, Y., and Olson, J. S. (2000). Steady-state and transient kinetics of *Escherichia coli* nitric-oxide dioxygenase (flavohemoglobin): The B10 tyrosine hydroxyl is essential for dioxygen binding and catalysis. *J. Biol. Chem.* **275,** 12581–12589.

Gardner, P. R., Gardner, A. M., Brashear, W. T., Suzuki, T., Hvitved, A. N., Setchell, K. D., and Olson, J. S. (2006). Hemoglobins dioxygenate nitric oxide with high fidelity. *J. Inorg. Biochem.* **100,** 542–550.

Guallar, V., and Friesner, R. A. (2004). Cytochrome P450CAM enzymatic catalysis cycle: A quantum mechanics/molecular mechanics study. *J. Am. Chem. Soc.* **126,** 8501–8508.

Hummer, G., and Szabo, A. (2001). Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. USA* **98,** 3658–3661.

Jacobson, M. P., Friesner, R. A., Xiang, Z., and Honig, B. (2002). On the role of the crystal environment in determining protein side-chain conformations. *J. Mol. Biol.* **320,** 597–608.

Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **78,** 2690–2693.

Karplus, M., and Petsko, G. A. (1990). Molecular dynamics simulations in biology. *Nature* **347,** 631–639.

Leach, A. R. (2001). "Molecular Modelling: Principles and Applications." : Pearson Education EMA.

MacKerell, A. D., Jr., Bashford, D., Bellott, M., Dunbrack, R. L., Jr., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., III, Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102,** 3586–3616.

Marti, M. A., Bikiel, D. E., Crespo, A., Nardini, M., Bolognesi, M., and Estrin, D. A. (2006a). Two distinct heme distal site states define *Cerebratulus lacteus* mini-hemoglobin oxygen affinity. *Proteins* **62,** 641–648.

Marti, M. A., Capece, L., Crespo, A., Doctorovich, F., and Estrin, D. A. (2005). Nitric oxide interaction with cytochrome c′ and its relevance to guanylate cyclase: Why does the iron histidine bond break? *J. Am. Chem. Soc.* **127,** 7721–7728.

Marti, M. A., Crespo, A., Bari, S. E., Doctorovich, F. A., and Estrin, D. A. (2004). QM-MM study of nitrite reduction by nitrite reductase of *Pseudomonas aeruginosa*. *J. Phys. Chem. B* **108,** 18073–18080.

Marti, M. A., Crespo, A., Capece, L., Boechi, L., Bikiel, D. E., Scherlis, D. A., and Estrin, D. A. (2006b). Dioxygen affinity in heme proteins investigated by computer simulation. *J. Inorg. Biochem.* **100,** 761–770.

Milani, M., Pesce, A., Ouellet, Y., Ascenzi, P., Guertin, M., and Bolognesi, M. (2001). *Mycobacterium tuberculosis* hemoglobin N displays a protein tunnel suited for O2 diffusion to the heme. *EMBO J.* **20,** 3902–3909.

Olson, J. S., and Phillips, G. N., Jr. (1997). Myoglobin discriminates between O2, NO, and CO by electrostatic interactions with the bound ligand. *J. Biol. Inorg. Chem.* **2,** 544–552.

Onufriev, A., Bashford, D., and Case, D. A. (2004). Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **55,** 383–394.

Ouellet, H., Ouellet, Y., Richard, C., Labarre, M., Wittenberg, B., Wittenberg, J., and Guertin, M. (2002). Truncated hemoglobin HbN protects *Mycobacterium bovis* from nitric oxide. *Proc. Natl. Acad. Sci. USA* **99,** 5902–5907.

Park, S., and Schulten, K. (2004). Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys.* **120,** 5946–5961.

Perlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., III, DeBolt, S., Ferguson, D., Seibel, G., and Kollman, P. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **91,** 1–41.

Perdew, J. P., Burke, K., and Ernzerhof, M. (1996). Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77,** 3865–3868.

Scott, E. E., Gibson, Q. H., and Olson, J. S. (2001). Mapping the pathways for O2 entry into and exit from myoglobin. *J. Biol. Chem.* **276,** 5177–5188.

Soler, J. M., Artacho, E., Gale, J. D., García, A., Junquera, J., Ordejón, P., and Sánchez-Portal, D. (2002). The SIESTA method for ab initio order-N materials simulation. *J. Phys. Cond. Matt.* **14,** 2745–2779.

Torrie, G. M., and Valleau, J. P. (1977). Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **23,** 187–199.

Wang, J., Cieplak, P., and Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21,** 1049–1074.

Warshel, A., and Levitt, M. (1976). Theoretical studies of enzymatic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **103,** 227–249.

Xiong, H., Crespo, A., Marti, M., Estrin, D., and Roitberg, A. E. (2006). Free energy calculations with non-equilibrium methods: Applications of the Jarzynski relationship. *Theor. Chem. Acc.* **116,** 338–346.