Acc. Chem. Res. **2002,** *35,* 455−464

The Role and Perspective of Ab Initio Molecular Dynamics in the Study of Biological Systems

PAOLO CARLONI,*,†
URSULA ROTHLISBERGER,‡ AND
MICHELE PARRINELLO‡,§
International School for Advanced Studies (SISSA/ISAS) and Istituto Nazionale per la Fisica della Materia (INFM), via Beirut 4, I-34014, Trieste, Italy, Department of Chemistry, ETH-Hönggerberg HCI, CH-8093 Zurich, Switzerland, and CSCS-Centro Svizzero di Calcolo Scientifico, Via Cantonale, CH-6928 Manno, Switzerland

Received October 10, 2001

ABSTRACT
Ab initio molecular dynamics (MD) allows realistic simulations to be performed without adjustable parameters. In recent years, the technique has been used on an increasing number of applications to biochemical systems. Here we describe the principles on which ab initio MD is based. We focus on the most popular implementation, based on density functional theory and plane wave basis set. By a survey of recent applications, we show that despite the current limitations of size and time scale, ab initio MD (and hybrid ab initio MD/MM approaches) can play an important role for the modeling of biological systems. Finally, we provide a perspective for the advancement of methodological approaches which may further expand the scope of ab initio MD in biomolecular modeling.

1. Introduction

Molecular dynamics (MD) simulations based on the description of interatomic interactions via empirical force fields have a long and successful record in the study of biomolecular systems.[1] Over the years, ever larger systems have been studied for simulation times that now reach the 100-ns range. In fact, the development of realistic force fields is considered as one of the major achievements of present-day theoretical chemistry.

Still, despite the successes, there are many areas in which the use of effective potentials may be not appropriate, and more sophisticated and accurate approaches are needed. An alternative is offered by ab initio molecular dynamics,[2] which was developed in the early eighties and has been intensively applied in materials science and in chemical problems. At variance with the standard approach, in ab initio MD, the interatomic forces are not empirically derived, but are evaluated from electronic structure calculations as the simulation proceeds. Typically, use is made of density functional theory,[3] but other electronic structure methods have also been suggested. In contrast to the standard approach, in which the electronic degrees of freedom are integrated out, ab initio MD treats the electronic states explicitly and follows their evolution as the nuclei move.

This finer level of description demands a much larger computational cost, which severely restricts the length and time scale of the systems that can be described with the ab initio approach. This goes against the well-justified trend of biomolecular simulations to study larger systems for longer times. Nonetheless, the experience accumulated in recent years has shown that ab initio MD can play a significant and unique role in biomolecular simulations. This is due to the fact that there is a large class of phenomena that depend on the electronic states in such an intricate way that they cannot be modeled via effective potentials.

Despite the usefulness of effective potentials, they are sometimes based on rather drastic approximations, such as the average way in which polarization effects are treated. This leads to difficulties in the presence of strongly polarizing fields, such as those encountered in the sugar−phosphate backbone of nucleic acids or in the presence of highly charged counterions. In empirical models, the intramolecular conformation is determined by potentials that are a combination of pairwise angular and dihedral forces. Such parametrization does not lead to stable aromatic ring structures. These have to be enforced with the use of, for example, improper potentials that are meant to mimic the stabilization effect of aromatic π systems. This type of constraint might at times introduce excessive rigidity. This, together with the problem of correctly capturing polarization effects, might play a role for the well-known difficulty of effective potentials in describing systems, such as RNA and DNA. Similar problems are expected in the simulation of ion channels where large electric fields are present. The continuous increase in computational power makes ab initio MD an increasingly attractive approach to help the theoretical investigations of these systems.

Another area in which traditional modeling has difficulty is in the description of metallic centers, such as those that are encountered in many enzymes and metal/DNA adducts. Here subtle chemical phenomena come into play, such as the ligand field, and the fact that the metal ion ligand bond has a partially covalent nature. In a dynamical environment, the number of ligands and the metal oxidation state can change. This complexity cannot easily be modeled, and no general consensus has yet been reached on the functional form to be used in order to describe metal centers.[4] In the case of most transition metal ions, several spin states are possible, and the ion size and its chemistry can correspondingly vary considerably. Subtle chemical effects can lead to configurations that have very similar structural properties but rather different energies. Since the interaction potentials are

---

Michele Parrinello is Professor of Computational Sciences at ETH, Zurich, and Director of the Swiss Center for Scientific Computing in Manno, Switzerland. Both Paolo Carloni and Ursula Roethlisberger are former postdoctoral associates. Ursula Rothlisberger is now Assistant Professor of Inorganic Chemistry, also at ETH. Paolo Carloni is now Associate Professor of Chemistry at the International School for Advanced Studies in Trieste, Italy.

[†] International School for Advanced Studies and Istituto Nazionale per la Fisica della Materia.

[‡] ETH-Hönggerberg.

[§] CSCS-Centro Svizzero di Calcolo Scientifico.

10.1021/ar010018u CCC: $22.00

functions of the nuclear coordinates only, it would be extremely difficult to discriminate among these effects.

These are certainly important motivations for using ab initio MD, but an even more cogent reason for its use comes from the need to describe properly chemical bond breaking−bond forming processes. Examples include enzymatic reactions, for which, indeed, several investigations have already been reported in the literature,[5] and proton-transfer processes.[6] We recall also that crucial to modeling is the determination of the protonation state, information that can only rarely be obtained from the experiments. Ab initio MD has the accuracy needed to discriminate among different protonation states and naturally allows it to vary.

An important bonus of ab initio MD is that it allows close contact to be made with the experimental data, such as IR, Raman,[7] and NMR[8] spectroscopy data. These quantities can be calculated from the available information on the electronic structure without the need for making additional assumptions. Furthermore, all nonharmonic effects can be taken into account. Recent developments in the theory, such as the introduction of DFT-based methods for excited states such as ROKS[9] or TDDFT,[3] are beginning to permit the study of photochemical processes.

In view of these potential advantages, as the computational resources have increased, so have the number of applications of ab initio MD to biological problems. Since exhaustive reviews of applications of ab initio MD work to biochemical systems have already been published,[5] we will discuss only a few paradigmatic examples that illustrate the usefulness of the ab initio MD approach in biology.

2. The Method

In the course of an MD run, the interatomic forces need to be evaluated a rather large number of times. It is, therefore, imperative in any type of MD to make this part of the calculation as efficient as possible. This is even more important in ab initio MD, where the interatomic potential is calculated as

$$\Phi(\vec{R}_I) = E_0(\vec{R}_I) + V_{\mathrm{NN}}(\vec{R}_I)$$

where $V_{\mathrm{NN}}(\vec{R}_I)$ is the direct internuclear interaction and $E_0(\vec{R}_I)$ is the electronic ground-state energy evaluated at fixed nuclear positions $\vec{R}_I$. Here, we are assuming the validity of the Born−Oppenheimer approximation. Clearly computing $E_0(\vec{R}_I)$ is the most demanding part of the calculation. Since an exact evaluation of $E_0(\vec{R}_I)$ is not possible except for the smallest systems, it is necessary to resort to some form of approximation. If one wishes to apply ab initio MD to large systems, one needs to strike a balance between accuracy and computational efficiency, since high accuracy such as that provided by high-level quantum chemical approaches, such as MP2 or coupled cluster theory, is too expensive for use in ab initio MD. One scheme that provides useful accuracy at affordable cost is density functional theory (DFT),[3] and in fact, the overwhelming majority of ab initio MD are based on DFT. In the Kohn and Sham (KS) formulation of DFT,[3]

$$E_0(\vec{R}_I) = \min_{\psi_i} E_{\mathrm{KS}}[\psi_i;\vec{R}_I] \tag{1}$$

where $E_{\mathrm{KS}}[\psi_i;\vec{R}_I]$ is the KS energy density functional which depends on both the electronic $[\psi_i]$ and the nuclear $[\vec{R}_I]$-degrees of freedom.

$$E_{\mathrm{KS}}[\psi_i;\vec{R}_I] = -\frac{1}{2}\sum_i \int \mathrm{d}\vec{r}\ \psi^*_i(\vec{r})\nabla\psi_i(\vec{r}) + \int \mathrm{d}\vec{r}\ \mathrm{V_N}(\vec{r})\rho(\vec{r}) + \frac{1}{2}\int \mathrm{d}\vec{r}\ \mathrm{d}\vec{r}\,'\ \rho(\vec{r})\frac{1}{|\vec{r}-\vec{r}\,'|}\rho(\vec{r}\,') + E_{\mathrm{xc}}[\rho(\vec{r})] \tag{2}$$

In eq 2, $E_{\mathrm{KS}}[\psi_i;\vec{R}_I]$ is decomposed into a kinetic energy term, a term that takes into account the interaction with the nuclei; the Hartree term; and E_{xc}, the exchange correlation energy functional, which contains all of the intricacies of the many-body problem.[3]

The electronic density $\rho(\vec{r})$ is written in terms of the occupied KS orbitals $\psi_i(\vec{r})$ as

$$\rho(\vec{r}) = 2\sum_i \psi^*_i(\vec{r})\psi_i(\vec{r}) \tag{3}$$

The exact form of $E_{\mathrm{xc}}[\rho(\vec{r})]$ is not known, but excellent approximations are available. The minimum has to be sought subject to the orthogonality condition $\int \mathrm{d}\vec{r}\ \psi^*_i(\vec{r})\psi_j(\vec{r}) = \delta_{i,j}$.

To solve numerically the minimum problem of eq 1, we need to introduce a basis set in which to expand the orbitals. Schemes for ab initio MD based on the traditional quantum chemistry (QC) approach of using Gaussian basis sets have been implemented, but generally, the most popular choice until now has been to expand the $\psi_i(\vec{r})$ in plane waves (PW).[10] Such a basis set is unusual in QC, but it is very popular in condensed matter theory. There are many advantages to using a PW basis set. PW are totally unbiased, they automatically enforce the periodic boundary conditions that are usually employed in ab initio MD, they do not introduce basis set superposition errors, and FFT routines allow very efficient numerical implementation. Their major drawback is the very large number of PWs that is needed in order to achieve a satisfactory representation of the $\psi_i(\vec{r})$. Very great help in reducing the number of PW comes from the realization that there are two length scales in the problem, a short one associated with the core electrons, which vary very rapidly, and a larger one associated with the valence electrons. The latter is the one determining the chemical behavior, while the core electrons remain essentially unchanged during chemical processes and their effect can be expressed in terms of relatively smooth pseudopotentials acting only on the valence electrons.[10] Thus, in this approach, the core is treated by these pseudopotentials, and only the valence orbitals are treated with plane waves. The number of the latter depends on the type of the pseudopotential and the size of the simulation box that is used.[10]

During the simulation, the KS minimum has to be determined several times. To this effect, efficient optimization methods have been put into place.[10] Alternatively, a very stable dynamics can be obtained by propagating ions and electrons simultaneously according to the Lagrangian,[3]

$$L = \frac{1}{2}\mu\sum_i \int \mathrm{d}\vec{r}\ \dot{\psi}^*_i(\vec{r})\ \dot{\psi}_j(\vec{r}) + \frac{1}{2}\sum_I M_I\dot{\vec{R}}_I^2 - E_{KS}[\psi_I;\vec{R}_I] - \sum_{i,j}\Lambda_{i,j}\left(\int \mathrm{d}\vec{r}\ \psi^*_i(\vec{r})\psi_j(\vec{r}) - \delta_{i,j}\right) \quad (5)$$

where μ is a fictitious classical kinetic energy term, M_I the nuclear masses, and the Lagrange parameters $\Lambda_{i,j}$ enforce orthogonality. The ensuing second-order dynamics is very stable, and provided that the frequency $\sqrt{E_G/\mu}$, where E_G is the HOMO−LUMO gap, is much higher than the nuclear frequency, the nuclear dynamics is very close to the Born−Oppenheimer evolution. These conditions guarantee an adiabatic decoupling of the nuclear and electronic degrees of freedom, and the electrons follow the nuclear motion while remaining all the time close to the ground state.

The overall scaling of this algorithm is $O(N^3)$, where N is the number of electrons. This scaling is less than that of more elaborate QC schemes, but it still presents an enormous barrier toward application to systems as large as those of interest to biology. To circumvent the $O(N^3)$ barrier, a host of $O(N)$ methods have been introduced that exploit the locality of the chemical bond.[11] The usefulness of such methods for large-scale accurate calculations, however, remains to be demonstrated.

An alternative approach that is increasingly being applied is that of partitioning the whole system into a chemically active region and the rest, which is treated with empirical potentials.[12] This is the so-called quantum mechanics/molecular mechanics (QM/MM) approach based on a mixed Hamiltonian of the form

$$H = H_{\mathrm{QM}} + H_{\mathrm{MM}} + H_{\mathrm{QM-MM}} \quad (6)$$

H_{QM} represents the ab initio Hamiltonian, and H_{MM} is the purely classical Hamiltonian, which is described by a standard biomolecular force field,

$$H_{\mathrm{MM}} = H^{\mathrm{b}}_{\mathrm{MM}} + H^{\mathrm{nb}}_{\mathrm{MM}} \quad (7)$$

$H^{\mathrm{b}}_{\mathrm{MM}}$ takes into account intramolecular interactions (stretching, bending and torsional terms), whereas $H^{\mathrm{nb}}_{\mathrm{MM}}$ includes the van der Waals and electrostatic interactions.

Such a hierarchical approach has the advantage that the computational effort can be concentrated on the part of the system where it is most needed, whereas the effects of the surroundings are taken into account with a more expedient model. In this way, large biomolecular systems can be treated in a computationally efficient manner and for longer time scales. The pitfalls of QM/MM methods lie in the challenge of finding an appropriate and rigorous treatment of the coupling between QM and MM regions, as described by the interaction Hamiltonian $H_{\mathrm{QM-MM}}$. Special care has to be taken that the QM/MM interface is treated in an accurate and consistent way, in particular in combination with a plane-wave-based Car−Parrinello scheme. Several mixed QM/MM Car−Parrinello methods have recently been implemented. In the fully Hamiltonian coupling scheme of ref 13, bonds between the QM and MM parts of the system are treated with specifically designed monovalent pseudopotentials, whereas the remaining bonding interactions of the interface region, that is, the angle bending and dihedral distortions, are described on the level of the classical force field. The same holds for the van der Waals interactions between the QM and MM parts of the system. The electrostatic effects of the classical environment, on the other hand, are taken into account in the quantum mechanical description as an additional contribution to the external field of the quantum system,

$$H^{\mathrm{ele}}_{\mathrm{QM-MM}} = \sum_{i\in\mathrm{MM}} q_i \int \mathrm{d}r\ \rho(r)v_i(|r - r_i|) \quad (8)$$

where q_i is the classical point charge located at r_i and $v_i(|r - r_i|)$ is a Coulombic interaction potential modified at short range in such a way as to avoid spill-out of the electron density to nearby positively charged classical point charges.[13] In the context of a plane-wave-based Car−Parrinello scheme, a direct evaluation of eq 10 is prohibitive, because it involves of the order of $N_{\mathrm{r}}N_{\mathrm{MM}}$ operations, where N_{r} is the number of real space grid points (typically, $\sim 100^3$), and N_{MM} is the number of classical atoms (usually of the order of 10 000 or more in systems of biochemical relevance). Therefore, the interaction between the QM system and the more distant MM atoms is included with a Hamiltonian term explicitly coupling the multipole moments of the quantum charge distribution with the classical point charges. In this way, efficient and consistent QM/MM Car−Parrinello simulations of complex extended systems can be performed in which the steric and electrostatic effects of the surroundings are taken explicitly into account.

3. Applications

In the following Sections, we present a few selected examples of Car−Parrinello applications to biological systems of special current interest. These are meant to whet your appetite and illustrate some of the particular advantages and possibilities of such an ab initio approach to biology. The first is an investigation of the binding of copper ions to the murine prion protein, which provides a convincing illustration of the intricate nature of the transition metal−protein interactions that would be difficult to capture in a simplified empirical interaction potential. The second illustrates the precious benefit due to the explicit treatment of the electronic degrees of freedom, which allows for a direct simulation of the bond-breaking and -forming processes that occur in chemical reactions. This special feature can be used to study, for example, the occurrence of low-barrier hydrogen bond (LBHB) or enzymatic mechanisms of the anti-AIDS target HIV-1 protease. The last two parts give a short glimpse

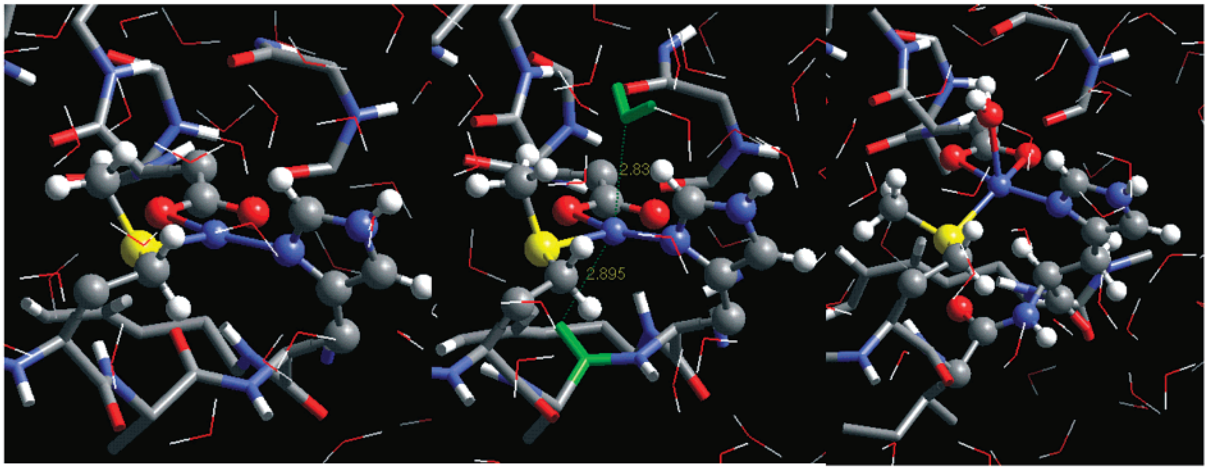


FIGURE 1. Sequence of snapshots from a mixed QM/MM Car–Parrinello simulation of a potential Cu^{2+} binding site in the mouse prion protein.[17] The QM part is shown in cylinders, and the classical environment is given in a stick representation.[18] The copper ion is shown in yellow. The total length of the simulation is ~5 ps. Note the significant dynamical changes that occur in the copper coordination sphere.

into recent extensions of the Car–Parrinello approach for the in situ calculation of NMR chemical shifts and the investigation of excited states.

1. Copper-Binding Sites in Prions. Prions are infectious agents that play a central role for a group of invariably fatal, neurodegenerative diseases affecting animals such as sheep (scrapie), cattle (BSE), and humans (e.g. variants of the Creutzfeldt–Jacob disease). It is now widely established that these diseases are caused by an abnormal isoform PrP^{Sc} of the normal cellular prion protein PrP^{C}.[14] Recent experiments, both in vivo and in vitro, indicate that prions are able to bind metal ions, in particular Cu^{2+}.[15] It has generally been assumed that Cu^{2+} ions bind to the histidine-rich unstructured part.[15] However, recent pulse EPR and ENDOR experiments[15] suggest that copper binds with higher affinity in the structured part (amino acid residues 125–228). Three pH-dependent copper binding sites have been identified together with some information about the nature of the immediate copper coordination sphere. The exact position of the metal binding sites within the protein, however, could not be localized experimentally.

An identification of the copper binding sites in the structured part of prion proteins and a detailed characterization of the metal protein interactions are of great interest, because it has been suggested that metal ions could influence the structural stability and, potentially, also the transition to the infectious scrapie form. Moreover, the interplay with metal ions is a common denominator for several other neurodegenerative disorders such as Alzheimer's, Parkinson's, and amylotrophic sclerosis, and there is increasing evidence that the interaction with metal ions plays a crucial role in all of them.[16]

Starting from the available NMR structure of the metal-free mouse PrP^{C},[17] we have performed mixed QM/MM Car–Parrinello simulations with the aim of identifying likely locations for Cu^{2+} binding.[18]

Locating (transition) metal ions in a protein structure is a particularly challenging task, since the intricate nature of the bonding properties of the metal necessitates the use of a quantum mechanical electronic structure approach. At the same time, the protein environment and solvent have to be taken into account as well as dynamical finite temperature effects. A QM/MM approach appears, therefore, to be the method of choice. However, since it is a priori not known which ligands are involved in binding, the necessary partitioning of the system into QM and MM regions is far from trivial. To circumvent this problem, we first performed a statistical analysis of all available high-resolution (≤2 Å) crystal structures of copper proteins to determine the relative propensity of different amino acids for copper ligation. Using this probability map, we scanned the structure of the mouse prion protein for the most likely Cu^{2+} binding regions. The resulting candidate structures were ranked by increasing probability, and their stability was subsequently tested in QM/MM Car–Parrinello simulations of several picoseconds around room temperature (300 K).

The starting point of the simulations was the experimentally determined NMR structure, which was immersed in a box of water and equilibrated by classical MD simulations based on the GROMOS96 force field.[1c] Subsequently, a Cu^{2+} ion was placed in the vicinity of selected residues of the QM region, and the exact location of the Cu^{2+} ion and its coordination were determined in mixed QM/MM Car–Parrinello simulations. Characteristic snapshots of one of these dynamics runs are shown in Figure 1. This possible binding site involves a Met, a His, and an Asp residue. As is apparent from Figure 1, the system undergoes a number of changes in the coordination sphere and the assumed ligand geometry. In fact, in several simulations, the QM region of the system had to be adapted dynamically in such a way as to allow for coordination with alternative ligands whose potential participation emerged only during the course of the simulation. This illustrates the highly complex and dynamical nature of the metal binding properties, which

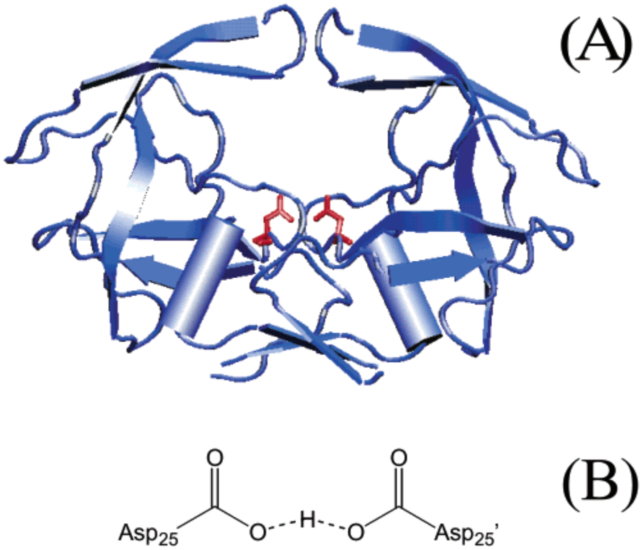


FIGURE 2. (a) Three-dimensional structure of HIV-1 protease.[19] This enzyme, which cleaves peptide segments at specific locations, is essential for the virus' metabolism.[19] The enzyme is composed of two identical subunits. A C_2 symmetry axis passes through the cleavage site, which consists of the Asp catalytic dyad (in red). (b) LBHB between the Asp pair in the free enzyme, as obtained by ab initio MD simulations on model systems.[21]

are not easily captured in a simplified empirical force field description.

2. HIV-1 Protease Cleavage Site. The protease from the human immunodeficiency virus type 1 (HIV-1 PR) is one of the major targets against the AIDS epidemic.[19] The enzyme is a dimer composed of two identical subunits (Figure 2a). The cleavage site consists of two aspartyl residues (Asp25 and Asp25′) located at the subunit/subunit interface. The two groups are almost coplanar and rather close to each other.

Force-field-based MD methods have encountered unexpected difficulties in describing the conformational properties of the cleavage site,[20] which are believed to be crucial for drug binding and enzymatic function. As a result, ad-hoc models imposing geometric constraints have been introduced.[20] These models succeed in reproducing the structural properties of the cleavage site, although they do not provide the physicochemical origin of the stability in the active site.

Ab initio MD simulations on models of the active site have been used to provide an understanding of the molecular interactions between the Asp pair.[21]

The calculations suggest that the *close proximity* of the cleavage site is achieved by the formation of a LBHB[6] connecting the two Asp groups (Figure 2b). This also helps explain the chemical equivalence of the two Asp carboxyl carbons in the free enzyme, as observed in a recent ^{13}C NMR study.[22] Furthermore, it accounts for the inverse solvent isotope effect in the free enzyme, because LBHB have a very low fractionation factor.[6] Finally, it permits answering many mechanistic questions about aspartic proteases.[6]

The calculations further suggest that the electrostatic interaction between the aspartyl negative charge and the rather rigid Thr 26(26′)−Gly 27(27′) dipoles provide a very important contribution to the stabilization of the *coplanar orientation* of the two carboxy groups. This finding appears to be consistent with site-directed mutagenesis experiments on this position.[21]

From the ab initio MD simulations, we have an indication that several ingredients, such as the polarization forces, the treatment of bond-forming/breaking processes, and temperature effects, play a crucial role in the HIV-1 PR active site. On the basis of these findings, specific force fields could now be developed for this system, which in turn might allow for a more accurate modeling of HIV-1 PR-drug interactions with empirical force-fields.

3. Reaction Mechanism of HIV-1 PR. The catalytic power of enzymes is believed to result from the presence of a properly engineered electric field[23] at the active site.

In HIV-1 PR, this view appears to be challenged by the emergence of "compensatory" drug-resistant mutations, which may be located either close to or far from the active site. These mutations, which are usually associated with mutations in the region of the active site, are believed to enhance the enzymatic reaction.[19] In many cases, these mutations involve chemically similar residues (such as M46I, L47V, L63P, A71T), which cannot significantly change the electrostatic field.

Attempts at explaining the catalytic activity of HIV-1 PR based on a static picture of the protein derived from the X-ray structures have been unable to account for the mutagenic behavior described above.

In a recent investigation,[24] we departed from this conventional approach and investigated the role of large-scale protein motions for enzymatic activity. To this aim, we have adopted a two-step strategy.

First, we have carried out classical MD calculations on a protein/substrate complex in aqueous solution. The simulation reveals large-scale protein motions on the nanosecond time-scale involving the flaps of the protein (Figure 3). The substrate motion turns out to be coupled with the lowest frequency motion of the protein.

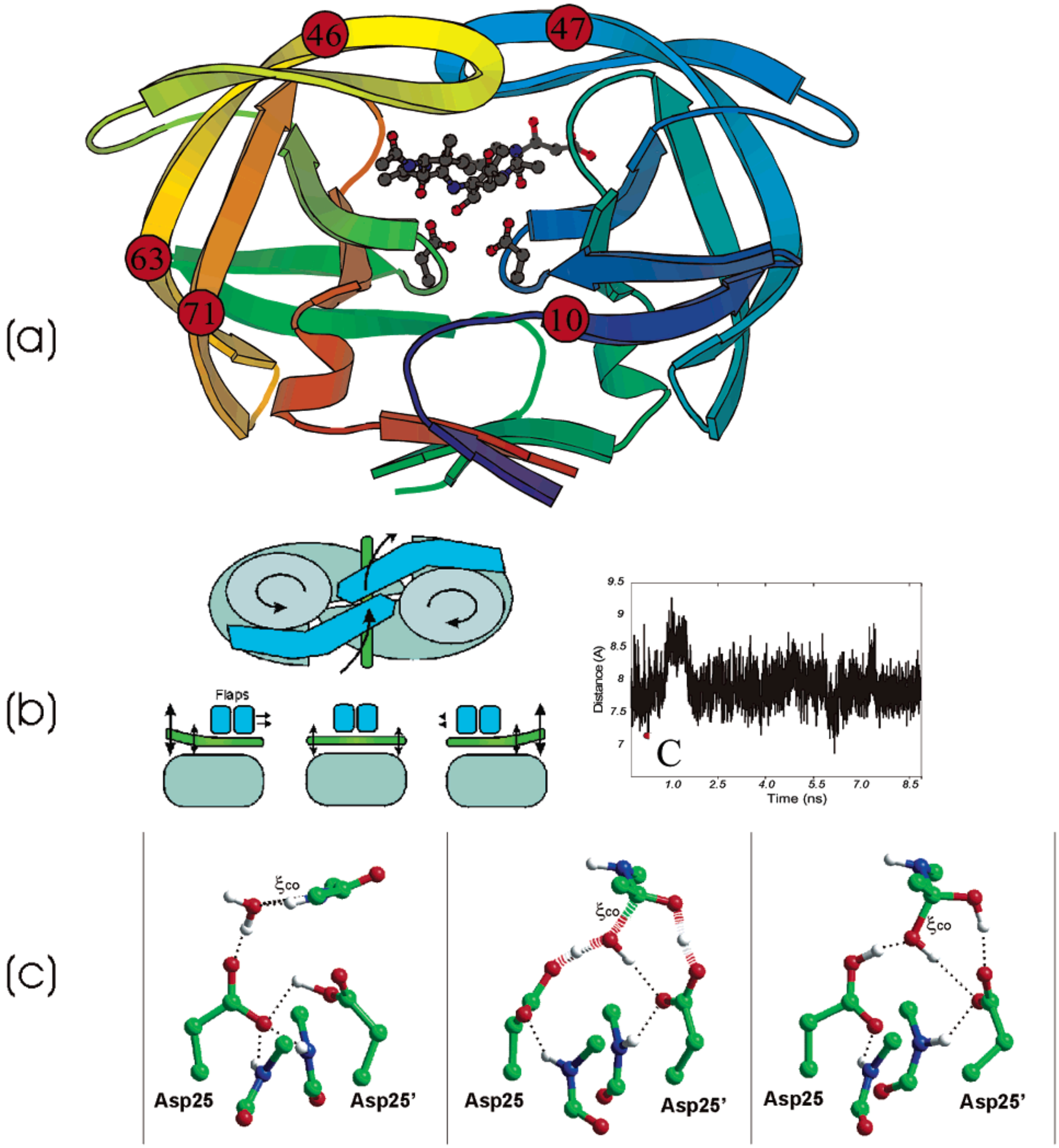


FIGURE 3. (a) Reaction mechanism and drug resistance in HIV-1 PR. The complex between the enzyme and a model substrate used in ab initio and classical MD calculations is shown.[24] Selected positions undergoing compensatory mutations are also shown.[19] (b) Classical MD simulation: (left) Schematic representation of the lowest-frequency motion of the protein, on the nanosecond time scale. The motion is a rotation of the flaps (blue) and other regions of the protein (gray), coupled to a movement of the substrate (green) along the protein C_2 axis. View from the top and from a plane containing the C_2 symmetry axis. (Right) Distance (measured on the four Cα's, see ref 24) between the substrate and Asp dyad Cα carbons plotted as a function of the simulated time. (c) Ab initio MD simulation: selected snapshots at the shortest substrate/Asp dyad distance investigated (labeled as C in the plot of part B of this figure).

As the second step, we have investigated the chemical reactivity of the enzyme at different configurations of the protein. We have focused on the first step of the catalytic reaction (from ES to INT, Chart 1) and estimated the activation free energy with ab initio MD.[25] The quantum-chemical calculations show that at conformations in which the substrate/enzyme distance corresponds to the average value of the classical simulation, the activation free energy is by far larger (several tens of kcal/mol) than that obtained experimentally (16−18 kcal[26]). In contrast, at very short substrate/Asp dyad distances (indicated with C in Figure 5b), a concerted double proton transfer among the Asp dyad, the catalytic water and the substrate can take place (Figure 3c). This stabilizes the highly nucleo-

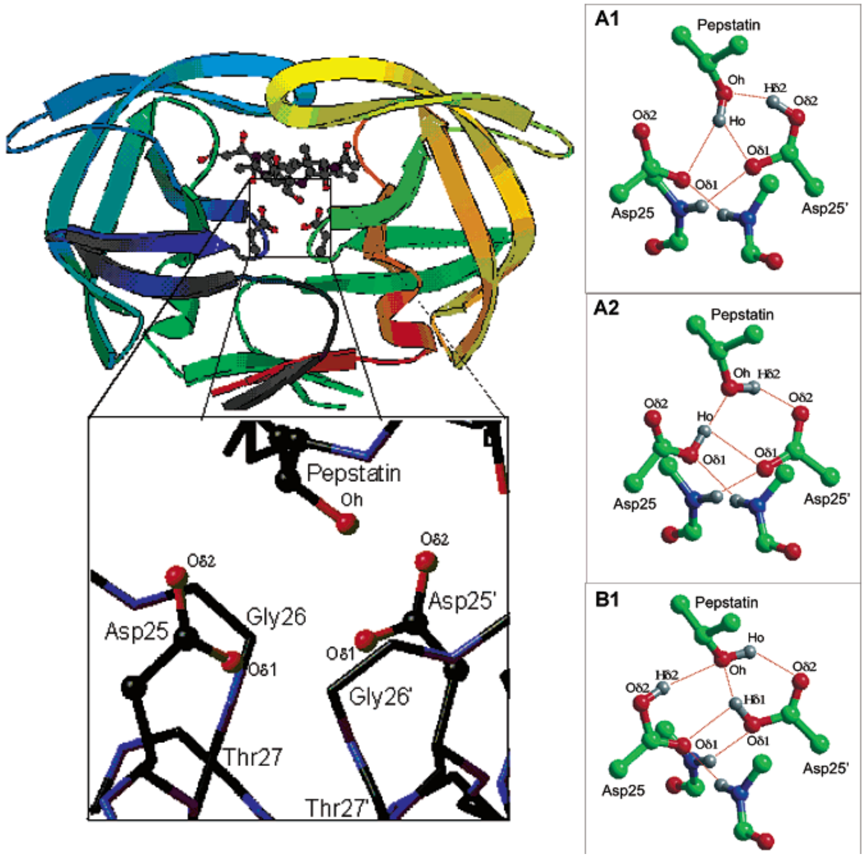
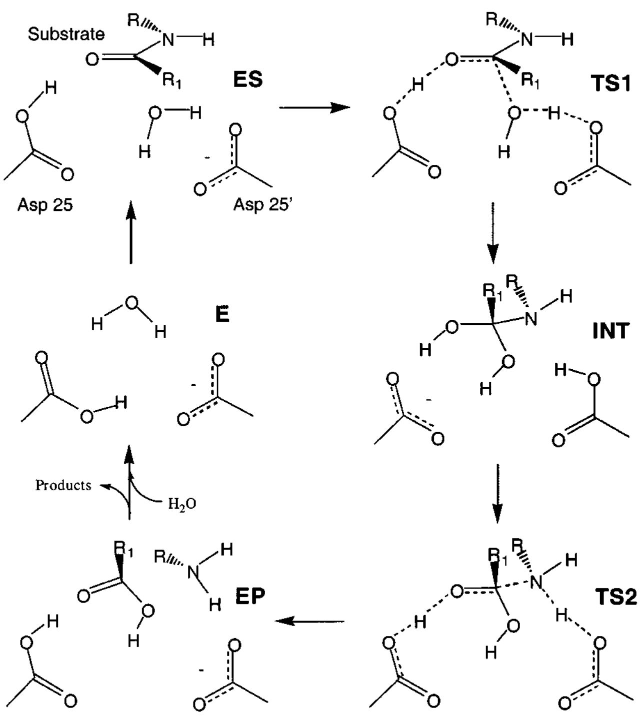


FIGURE 4. HIV-1 PR/pepstatin complex.[28] The protomer A1, postulated on the basis of ^{13}C NMR data,[22] evolves to A2 during the ab initio MD. The diprotonated form B1, proposed in ref 28, is stable during the dynamics and exhibits two clearly distinct chemical shifts.

Chart 1



philic agent OH$^-$, which readily reacts to form a gem-diol intermediate INT (Chart 1) with a calculated activation energy comparable to the experimental value.

Thus, the protein appears to act as an "elaborate nutcracker" that brings the substrate into a favorable geometrical position for the appropriate mechanical fluctuations of the protein frame. This mechanism validates the proposal of Lumry for HIV-1 PR in particular and mechanical catalysis in general.[27] Furthermore, it offers a clue to how the influence of drug-resistant mutations far from the active site can emerge. Indeed, these mutations are expected to enhance the catalytic rate of HIV-1 PR mutants by affecting the flexibility of the protein, which in turn may affect the enzymatic activity.

4. NMR Properties of HIV-1 PR Cleavage Site. In the HIV-1PR/pepstatin adduct (Figure 4), a recent ^{13}C NMR experiment[22] measured two separate resonance lines. On the basis of these results and the X-ray structure of the complex, the Asp dyad was suggested to be monoprotonated (pattern A1 in Figure 4). The assignment appeared to be surprising, because the chemical shifts are reversed with respect to the adduct with the chemically similar inhibitor KNI-272.

Ab initio MD on models of the pepstatin/enzyme adduct[28] (A1, Figure 4) have provided evidence that the proposed pattern is not stable. It evolves to protomer A2 (Figure 4), whose calculated ^{13}C chemical shifts are almost equivalent, in disagreement with the experimental results, whereas the other possible patterns of the monoprotonated form exhibited NMR properties that disagree with experiment or are highly unstable.

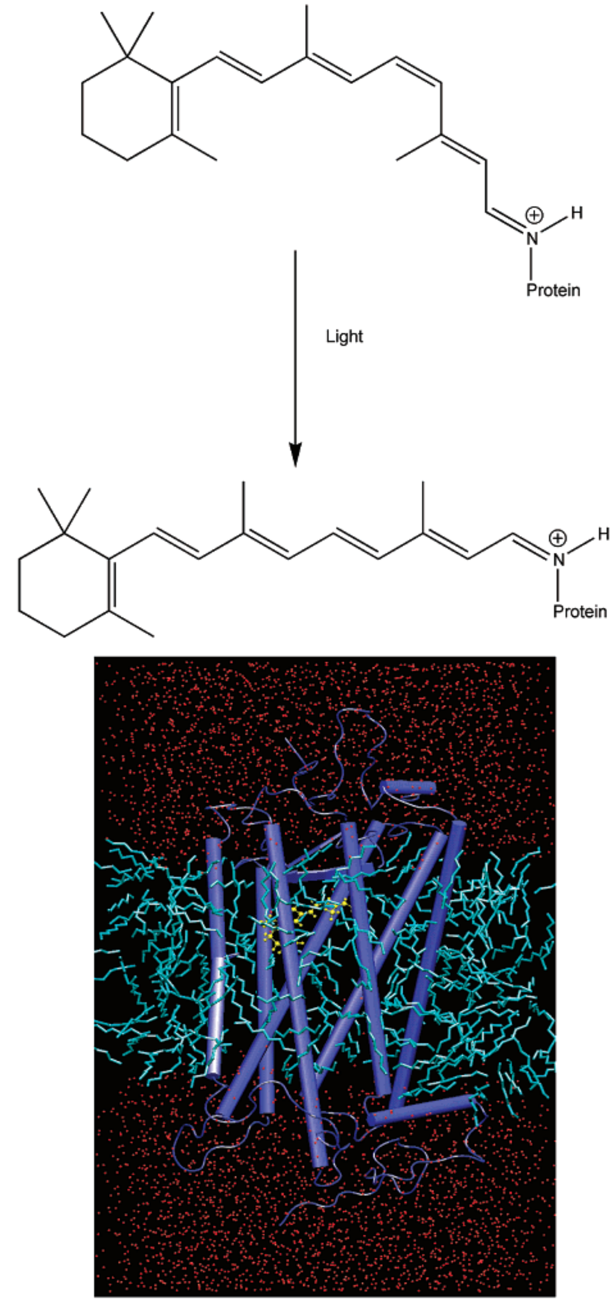


FIGURE 5. Simulation of rhodopsin: (above) photoinduced cis–trans isomerization of the protontated Schiff base of 11-*cis* retinal; (below) simulation system for rhodopsin. The protein (blue) contains seven transmembrane helices that accommodate the chromophore (yellow) in the middle.[30] The membrane is mimicked by an octane solution (turquoise) surrounded by water (red).

We suggest, therefore, that the Asp dyad is doubly protonated. The diprotonated form B1 (Figure 4) turns out to be stable in the time scale investigated (few picoseconds) and exhibits different ^{13}C chemical shifts, in qualitative agreement with experimental data[22] (Figure 4).

We expect that the powerful combination of ab initio MD with calculations of NMR chemical shifts will be widely used in the future to interpret spectral data in biomolecular systems.

5. Cis/Trans Photoisomerization in Rhodopsin. Human vision is a highly efficient process in which light is transformed into a neuronal signal that is detected in the brain. In the first step of this biochemical reaction chain, the transmembrane protein rhodopsin, which is located in the retina, plays the predominant role.[29] Rhodopsin consists of seven transmembrane helices accommodating in their middle the covalently bound chromophore, the protonated Schiff base of 11-*cis* retinal. Upon exposure to light, retinal isomerizes from 11-*cis* to all-trans in an extremely fast and efficient process, which is completed within 200 fs with a quantum yield of 0.67. This induces a structural change of the intracellular part of the protein and triggers the signal transduction to another receptor protein of the signal cascade. A mechanistic understanding of this primary reaction is not only essential with regard to the cure of some eye diseases but also of great interest for the development of optical switches.

The recent X-ray structure of rhodopsin[30] has paved the way to a molecular investigation of the activation mechanism.

Detailed mechanistic understanding has remained an open question. The role of the protein environment that promotes the isomerization reaction is not well-understood. In addition, the conformational changes that accompany the accommodation of all-*trans* retinal within the tight binding pocket are unknown. From a theoretical point of view, the investigation of these questions requires a mixed quantum/classical approach, because the excited-state dynamics of retinal (50 atoms) is a quantum chemical process, but the description of the whole protein (about 5600 atoms) is only feasible within a classical framework.

We have recently combined a mixed QM/MM Car–Parrinello approach[18] with the restricted open-shell Kohn–Sham (ROKS)[9] method for describing the dynamics in the first excited-singlet state. This combination allows the investigation of photochemical processes in complex biological systems. As a first test case study, we are currently investigating the cis–trans photoisomerization in rhodopsin. Our model system (24 000 atoms, see Figure 5) contains the protein, a hydrophobic membrane–mimetic environment consisting of octane and a saline solution. This setup is based on the most recent crystal structure of rhodopsin and results in a stable classical MD of several nanoseconds without imposing any type of structural constraint. The key configurations of these classical preequilibration runs are currently used for mixed QM/MM simulations of the initial reactant state and the ROKS-QM/MM Car–Parrinello simulations of the isomerization in the excited state.

4. Concluding Remarks

In this Account, we have shown that despite the current limitations of size and time scale, ab initio MD can play an important role in the modeling of biological systems. Particularly useful is the QM/MM approach that allows local chemical accuracy to be combined with a proper description of the environmental effects. As is clearly demonstrated by the examples discussed, it is now possible through the use of the ab initio approach to tackle problems of great current interest that cannot be solved by a different approach. Nevertheless these successes cannot hide the fact that much needs to be done in order to further expand the scope of ab initio MD.

The accuracy of the potential energy surface in ab initio MD is determined by the exchange and correlation functionals used. These are quite accurate but have known limitations, such as approximate transition state energies. Some of these limitations can be overcome by locally using higher-order quantum chemical methods. Less localized correlation effects such as van der Waals interactions are more difficult to include in an ab initio way[31] and have to be added empirically whenever they are expected to play a significant role.

As discussed earlier, size is an issue. Whenever a small region where quantum chemical effects are important can be identified and this region is clearly separated from the rest, the QM/MM approach can be of great help. However, a large class of problems requires the explicit quantum chemical description of rather large systems. For instance, electron transfer can occur over quite large distances in electron-transfer proteins[32] and proton pumps induce long-range proton motions.[33] Given the difficulty of modeling nucleic acids with classical potentials, we expect this to be an area in which ab initio MD will have a great impact. The sheer size of these systems has until now been an impediment to wider applications. Although technological progress can, indeed, push forward the limit of what is presently possible, a decisive impulse in this direction will come when $O(N)$ methods reach maturity.

The limited time scale over which a system can be simulated is a general problem of all MD simulations. This is all the more cogent in ab initio MD, in which the computational cost of a single time step is so much higher. QM/MM methods improve matters, because they reduce the time step cost, but they are not a solution for modeling processes that take place over time scales of nanoseconds or more. Extending the time scale of MD simulations is an area of great current interest and important progress is being made wherever the relevant free energy surface is composed of local minima separated by large barriers. In this case, the system spends most of its time in these minima, but occasionally it jumps very quickly to another minimum. These jumps are infrequent but crucial, since they lead to important changes in the system. For instance, they can correspond to a chemical reaction or to a change in conformation. The locality in time of these rare but important events can be exploited, and a variety of powerful methods have been suggested.[34] There are, however, problems in which the different time scales cannot be separated, one example out of the many being actin dynamics. For this class of difficult problems, novel solutions need to be found, since it is not to be expected that technology can deliver the necessary computational power in the foreseeable future.

We briefly mention the fact that a proper description of the proton motion and other light particles require that their dynamics be treated quantum mechanically. This is a formidable problem which, despite the progress achieved, has not been seen solved except for relatively simple models.[35] Finally, we wish to emphasize that many of the applications so far presented have used the Born–Oppenheimer approximation. Going beyond this approximation requires a method not only for calculating accurately and efficiently the potential energy surfaces of excited states, but also for evolving the dynamics of the coupled nonadiabatic electron ion motions.

From this rather cursory analysis, it is clear that the challenges before us are manifold and difficult. We nonetheless wish to close on a positive note by pointing out that we have already reached the stage where ab initio MD has a sufficient level of maturity to make it the method of choice for a very large class of problems.

References

(1) (a) Karplus, M. Molecular dynamics simulations in biology. *Nature* **1990**, *18*, *347*, 631–639. (b) van Gunsteren, W. F.; Luque, F. J.; Timms, D.; Torda, A. E. Molecular mechanics in biology: from structure to function, taking account of solvation. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 847–863.

(2) Car, R.; Parrinello, M. Unified Approach for Density Functional Theory and Molecular Dynamics. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.

(3) Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Phys. Rev. B* **1964**, *136*, 864–871. Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev. A* **1965**, *140*, 1133–1138. Parr, R. G.; Yang, W. *Density Functional Theory of Atoms and Molecules*, Oxford University Press: New York: 1989.

(4) Banci, L., Comba, P., Eds.; *Molecular modeling and dynamics of bioinorganic compounds*; Kluwer Academic Publishers: Dorderecht, Boston, London 1997.

(5) See, for example, Carloni, P.; Rothlisberger, U. Simulations of enzymatic systems: perspectives from Car-Parrinello molecular dynamics simulations. In *Theoretical biochemistry–processes and properties of biological systems*. Eriksson L., Ed. Elsevier Science: New York, 2000.

(6) Northrop, D. Follow the protons: a low-barrier hydrogen bond unifies the mechanism of the aspartic proteases. *Acc. Chem. Res.* **2001**, *34,* 790–797.

(7) Silvestrelli, P. L.; Parrinello, M. Water molecule dipole in the gas phase and in the liquid phase. *Phys. Rev Lett.* **1999**, *82*, 3308–3311.

(8) Sebastiani, D.; Parrinello, M. A new ab initio approach for NMR chemical shifts in periodic systems. *J. Phys. Chem. A* **2001**, *105*, 1951–1958.

(9) Frank, I.; Hutter, J.; Marx, D.; Parrinello, M. Molecular dynamics in low-spin excited states. *J. Chem. Phys.* **1998**, *108*, 4060–4069.

(10) Marx, D.; Hutter, J. Ab initio Molecular Dynamics: Theory and Implementation. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; John von Neumann Institute for Computing: Julich, 2000; 301–409.

(11) Goedecker, S. Linear scaling electronic structure methods. *Rev. Mod. Phys.* **1999**, *71*, 1085–1123.

(12) Aqvist, J.; Warshel, A. Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. *Chem. Rev.* **1993**, *93,* 2523–2544.

(13) Laio, A.; VandeVondele, J.; Rothlisberger, U. A Hamiltonian Electrostatic Coupling Scheme for Hybrid Car–Parrinello Molecular Dynamics Simulations, *J. Chem. Phys.*, **2002**, *116*, 6941–6947.

(14) Horwich, A. L.; Weissmann, J. S. Deadly conformations–Protein misfolding in prion disease. *Cell* **1997**, *89,* 499–510.

(15) Van Doorslaer, S.; Cereghetti, G. M.; Glockshuber, R.; Schweiger, A. Unraveling the Cu^{2+} binding sites in the C-terminal domain of the murine prion protein: A pulse EPR and ENDOR study. *J. Phys. Chem.* **2001**, *105,* 1631–1639.

(16) Price, D. L.; Sisodia, S. S.; Borchelt, D. R. Genetic neurodegenerative diseases: The human illness and transgenic models. *Science* **1998**, *282*, 1079–1083.

(17) Riek, R.; Hornemann, S.; Wider, G.; Billeter, M.; Glockshuber, R.; Wuthrich, K. NMR structure of the mouse prion protein domain PrP(121–231). *Nature* **1996**, *382*, 180–182.

(18) Colombo, M. C.; Guidoni, L.; Laio, A.; Magistrato, A.; Maurer, P.; Piana, S.; Roehrig, U.; Spiegel, K.; VandeVondele, J.; Zumstein, M.; Rothlisberger, U. Hybrid QM/MM Car-Parrinello Simulations of Catalytic and Enzymatic Reactions. *Chimia* **2002**, *56*, 11−17.

(19) Ogden, R. C., Flexner, C. W., Eds.; *Protease Inhibitors in AIDS Therapy,* Marcel Dekker: New York: 2001.

(20) Straatsma, T. P.; Zacharias, M.; McCammon, J. A. Free energy difference calculations in biomolecular systems. In van Gunsteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds. *Computer simulations of biomolecular systems*. ESCOM.: Leiden, The Netherlands, 1993.

(21) Piana, S.; Carloni, P. Conformational Flexibility of the Catalytic Asp Dyad in HIV-1 Protease: an Ab Initio Study on the Free Enzyme. *Proteins: Struct. Funct. Genet.* **2000**, *39*, 26−36.

(22) Smith, R.; Brereton, I. M.; Chai, R. Y.; Kent, S. B. H. Ionization states of the catalytic residues in HIV-1 protease. *Nature Struct. Biol.* **1996**, *3*, 946−950.

(23) Warshel, A. Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J. Biol. Chem.* **1998**, *273*, 27035−27038.

(24) Piana, S.; Carloni, P.; Parrinello, M. The role of conformational fluctuations in the enzymatic mechanism of HIV-1 protease, *J. Mol. Biol.,* in press.

(25) The free energy is calculated using constraint dynamics (Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. Constrained reaction coordinate dynamics for the simulation of rare events. *Chem. Phys. Lett.* **1989,** *156*, 472−477). The reaction constraint is the distance between the catalytic water and the substrate carbonyl carbon.

(26) Polgár, L.; Szeltner, Z.; Boros, I. Substrate-dependent mechanism in the catalysis of human immunodeficiency virus protease. *Biochemistry* **1994**, *33*, 9351−9357.

(27) Lumry, R. On the interpretation of data from isothermal processes. *Methods Enzymol.* **1995**, *259*, 628−720.

(28) Piana, S.; Sebastiani, D.; Carloni, P.; Parrinello, M. An ab initio molecular dynamics-based assignment of the protonation state of pepstatin A/HIV-1 protease cleavage site. *J. Am. Chem. Soc.* **2001**, *123*, 8730−8737.

(29) For a review, see: Gether U. Uncovering molecular mechanisms involved in activation of G protein-coupled receptor. *Endocr. Rev.* **2000**, *21*, 90−113.

(30) Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **2000**, *289*, 739−745.

(31) Meijer, E. J.; Sprik, M. A density-functional study of the intermolecular interactions of benzene. *J. Chem. Phys.* **1996**, *105*, 8684−8689.

(32) Warshel, A. Protein Control of Redox Potentials of Iron−Sulfur Proteins, *Chem. Rev.* **1996**, *96* (7), 2491−2514.

(33) Hofacker, I.; Schulten, K. Oxygen and proton pathways in cytochrome *c* oxidase. *Proteins: Struct. Funct. Genet.* **1998**, *30,* 100−107.

(34) VandeVondele, J.; Rothlisberger, U. Efficient Multidimensional Free Energy Calculations for ab initio Molecular Dynamics using Classical Bias Potentials, *J. Chem. Phys.* **2000**, *113*, 4863−4868. Passerone, A.; Parrinello, M. Action-derived molecular dynamics in the study of rare events. *Phys. Rev. Lett.* **2001**, 8710, 8302.

(35) Webb, S. P.; Agarwal, P. K.; Hammes-Schiffer, S. Combining electronic structure methods with the calculation of hydrogen vibrational wave functions: application to hydride transfer in liver alcohol dehydrogenase. *J. Phys. Chem.* **2000**, *104*, 8884−8894.

AR010018U