

Reaction Mechanism of HIV-1 Protease by Hybrid Car-Parrinello/Classical MD Simulations

Stefano Piana,[†] Denis Bucher,[‡] Paolo Carloni,[§] and Ursula Rothlisberger^{*,‡}

Scuola Internazionale Superiore di Studi Avanzati and INFN - DEMOCRITOS National Simulation Center, Via Beirut 2-4, 34014 Trieste, Italy, Laboratory of Computational Chemistry and Biochemistry, Federal Institute of Technology - EPFL, 1015 Lausanne, Switzerland, and Nanochemistry Research Institute, Curtin University of Technology, P.O. Box U 1987, Perth WA 6845, Australia

Received: November 30, 2003; In Final Form: March 21, 2004

We present a QM/MM *ab initio* molecular dynamics study of the peptide hydrolysis reaction catalyzed by HIV-1 protease. The QM/MM calculations are based on previous extensive classical MD simulations on the protein in complex with a model substrate (Piana, S.; Carloni, P.; Rothlisberger, U. *Protein Sci.* **2002**, *11*, 2393–2402). Gradient-corrected BLYP density functional theory (DFT) describes the reactive part of the active site, and the AMBER force field describes the rest of the protein, the solvent, and the counterions. An unbiased enhanced sampling of the QM/MM free-energy surface is performed to identify a plausible reaction coordinate for the second step of the reaction. The enzymatic reaction is characterized by two reaction free-energy barriers of ~ 18 and ~ 21 kcal mol⁻¹ separated by a metastable gem-diol intermediate. In both steps, a proton transfer that involves the substrate and the two catalytic Asp molecules is observed. The orientation and the flexibility of the reactants, governed by the surrounding protein frame, are the key factors in determining the activation barrier. The calculated value for the barrier of the second step is slightly larger than the value expected from experimental data (~ 16 kcal mol⁻¹). An extensive comparison with calculations on gas-phase model systems at the Hartree–Fock, DFT-BP, DFT-BLYP, DFT-B3LYP, MP2, CCD, and QM/MM DFT-BLYP levels of theory suggests that the DFT-BLYP functional has the tendency to underestimate the energy of the gem-diol intermediate by ~ 5 – 7 kcal mol⁻¹.

The aspartyl protease from human immunodeficiency virus type 1 (HIV-1 PR) targets the AIDS epidemic. The enzyme is essential for viral metabolism¹ because it cleaves the long polypeptide chain that is expressed in infected host cells in specific positions to generate the active proteins that are required for viral maturation.

HIV-1 PR is a homodimer with the active site located at the interface between the two subunits. The cleavage site is an Asp dyad (Asp25 and Asp25', Figure 1), located inside a large active-site pocket that allows the enzyme to recognize and cleave sequences of six amino acids selectively.² Several aspects of the enzymatic reaction mechanism have been the focus of a variety of computational techniques, including molecular mechanics,^{3–5} tight-binding,⁶ semiempirical,^{7,8} and *ab initio*^{9–15} methods. This theoretical work has been complemented by kinetic, thermodynamic, and structural data.^{8,16–29} The picture emerging from these studies can be summarized as follows. The free form of the enzyme (**E**) is stabilized by a low-barrier H bond (LBHB)³⁰ locking the Asp dyad in an almost coplanar conformation. In a first physical step, **E** binds to substrate **SUB** to form the enzyme–substrate complex **ESUB**, which might (**ESUB(a)**, Figure 1) or might not (**ESUB(b)**, Figure 1) maintain the LBHB. ²H and ¹⁵N kinetic isotope effect measurements^{17,18,28} have established that in HIV-1 PR (i) a hydrated intermediate is reversibly formed and (ii) protonation of the peptide bond nitrogen occurs before an irreversible step. To form the hydrated

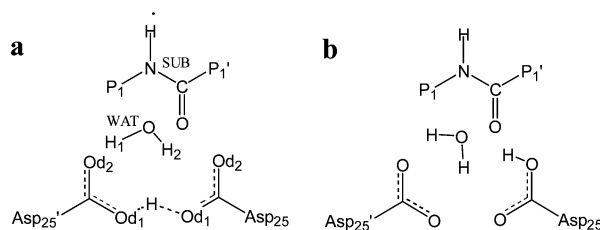


Figure 1. ESUB protonation patterns investigated. In ESUB(a), the low-barrier hydrogen bond between the two aspartic acids is still present after substrate binding.

intermediate, a water molecule (**WAT**) must attack the carbonyl carbon of **SUB**^{15,17,18,21,31} (chemical step 1). This process can be assisted by the Asp dyad, which acts as a proton donor–acceptor group.^{7,8,14,24,32} The possibility of formation of an oxanion intermediate stabilized by tunneling^{6,17,18,21} (as opposed to a gem-diol intermediate) has been proposed.²⁹ Quantum chemical calculations^{7,8,14,15} suggest that the oxanion might be unstable in the active site of HIV-1 PR. This is in contrast to other hydrolases, such as serine proteases, where an oxanion cavity assists in the formation of the anion. Direct nucleophilic attack of the Asp dyad with the formation of a covalently bound intermediate has also been proposed^{12,33} but is in contrast to isotope exchange experiments.^{17,18} Also, the protonation of the nitrogen atom before the nucleophilic attack has been suggested.⁹ In one or more subsequent steps, the C–N bond of **INT** breaks heterolytically (chemical step 2). Also, this process is assisted by the Asp dyad: Asp 25' donates a proton to the **INT** amide group, and simultaneously^{7,8,17} or subsequently^{14,28} Asp 25 accepts a proton from one of the **INT** hydroxyl groups (**TS2**). This double proton transfer can be facilitated by an anti-gauche

* To whom correspondence should be addressed. E-mail: ursula.rothlisberger@epfl.ch. Phone: +41-21-6930321. Fax: +41-21-6930320.

[†] Curtin University of Technology.

[‡] Federal Institute of Technology - EPFL.

[§] Scuola Internazionale Superiore di Studi Avanzati and INFN - DEMOCRITOS National Simulation Center.

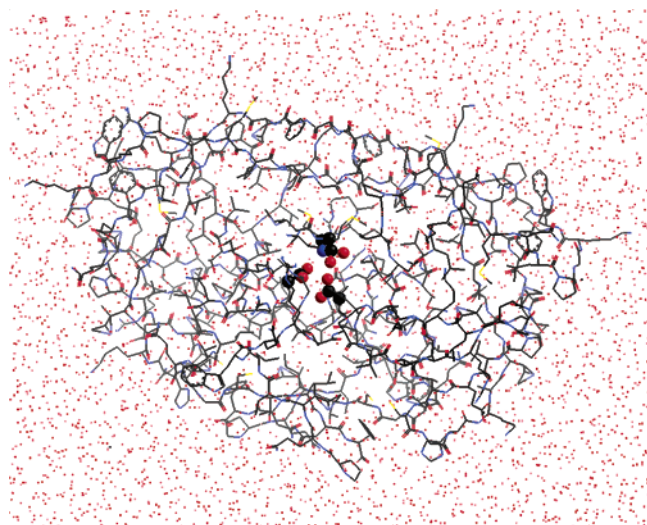


Figure 2. Structure of the HIV-1 PR/INT complex used in the QM/MM simulations. The protein is immersed in a $66.8 \times 55.2 \times 43.0$ Å³ box containing 4170 water molecules. The QM atoms are represented in a ball-and-stick model. Hydrogen atoms are not shown for the sake of clarity.

transition in the gem-diol.⁸ **TS2** decomposes (perhaps passing through a second intermediate^{14,28}) to give the reaction products (**EPROD**), namely, an amine and a carboxylic acid. Finally, the products are released,^{29,34,35} and free form **E** is restored. The rate-determining step depends on the type of substrate, and it may be any of the chemical or physical steps mentioned above.^{17,26,27,29,36}

The present work will be aimed at characterizing the chemical steps that lead from the enzyme substrate complex (**ESUB**) to the hydrated intermediate (**EINT**) to the reaction products (**EPROD**). Previous studies on the conformation and reactivity of the active site of HIV-1 PR have suggested that (i) the flexibility of the protein frame is a fundamental ingredient for the reaction;^{15,32} (ii) the groups interacting with the Asp dyad (in particular, the dipole moment of the Thr26(26')-Gly27(27') peptide bonds) are important for the conformation and stability of the active site^{8,10,11,30} and therefore should be included in the calculations in some form; and (iii) the use of an accurate first principles method and a relatively large basis set is recommended to obtain reliable structural and energetic properties.^{11–13} Here we take explicit account of the ingredients (i.e., i–iii). Toward this aim, we perform hybrid Car–Parrinello/molecular mechanics (QM/MM) simulations³⁷ with the BLYP exchange–correlation functional,³⁸ which has already been used to describe the HIV-1 PR cleavage site,^{15,30,39} and a plane wave basis set (PW).⁴⁰ In this approach, the flexibility of the protein and the interaction of the active site with the rest of the protein are fully taken into account. The calculations are based on extensive classical MD simulations on the **ESUB** and **EINT** protein complexes (Figure 2).³²

The activation free energies of the reaction steps are calculated by the thermodynamic integration of the force acting along a constrained reaction coordinate.⁴¹ For the formation of the hydrated intermediate, this coordinate has been repeatedly recognized as the $C_{\text{(SUB)}}\text{--}O_{\text{(WAT)}}$ distance.^{7,15} An appropriate choice of constraint coordinate for chemical step 2 is less obvious. During this step, one bond ($N_{\text{(INT)}}\text{--}H_{\text{(Asp25')}}$) is formed, and two bonds ($C_{\text{(INT)}}\text{--}N_{\text{(INT)}}$ and $O_{\text{(INT)}}\text{--}H_{\text{(INT)}}$) are broken, thus any of these distances or a combination of the three could be taken as a reaction coordinate. We therefore perform an initial unbiased enhanced sampling of the free-energy surface using

the CAFES formalism.⁴² This procedure suggests that the $N_{\text{(INT)}}\text{--}H_{\text{(Asp25')}}$ distance is a suitable constraint coordinate for chemical step 2.

The quality of the calculations is established by an extensive comparison between the structural and energetic features obtained in our first principles calculations with those obtained with other ab initio methods.

Results and Discussion

Protonation State of **ESUB.** The starting structure for **ESUB** was taken from a classical MD simulation of the HIV-1PR/substrate complex.³² The structure obtained after 3.27 ns of MD simulation was selected because, on the basis of previous investigations,¹⁵ it is expected to be a “reactive” conformation. The active-site form of HIV-1 PR that is catalytically competent is monoprotonated.¹⁷ The proton can be located either on Oδ1, with the two Asp groups connected by a low-barrier hydrogen bond (**ESUB(a)**, Figure 1a) or on Oδ2 (in this case, no low-barrier hydrogen bond is present (**ESUB(b)**, Figure 1b)). The free-energy difference between the two protonation patterns has been calculated with the method of constraints. (See Methods.) It turns out that **ESUB(a)** is $1.5(1)$ kcal mol^{−1} more stable than **ESUB(b)**. This small free-energy difference is in line with previous ab initio studies on models of the active site of HIV-1 PR,^{15,30} pepsin and endothiapepsin,¹¹ and suggests that both protonation patterns are significantly populated in **ESUB**. Interconversion between the two patterns occurs through a proton transfer from Asp25' Oδ1 to WAT followed by a proton transfer from WAT to Asp25 Oδ2. The activation energy for this process is $2.0(1)$ kcal mol^{−1}.

QM/MM Simulation of the Formation of the Hydrated Reaction Intermediate. QM/MM simulations were performed starting from **ESUB(a)** and **ESUB(b)**. In our simulation of the formation of the hydrated intermediate, the $C_{\text{(SUB)}}\text{--}O_{\text{(WAT)}}$ distance was constrained, and the constraint was shortened from 3.7 to 1.5 Å in ~6 ps of QM/MM MD simulation. The reaction free energy along the $C_{\text{(SUB)}}\text{--}O_{\text{(WAT)}}$ reaction coordinate was calculated as the integral of the average constraint force.⁴¹

In the simulation starting from **ESUB(a)** (Figure 3), the force on the constraint increases as the water oxygen approaches the carbonyl carbon. When the $C_{\text{(SUB)}}\text{--}O_{\text{(WAT)}}$ bond is shorter than 2.1 Å, the peptide bond loses its planarity, and at a $C_{\text{(SUB)}}\text{--}O_{\text{(WAT)}}$ distance of 1.8 Å, a proton is shared between the water molecule and Asp25', forming an LBHB. At the same time, the formation of an oxyanion species (**EINT(a)**, Figure 3) is observed with a negative accumulated charge on $O_{\text{(SUB)}}$. The free-energy difference between **ESUB(a)** and **EINT(a)** is $36(1.5)$ kcal mol^{−1} (Figure 4). In this simulation, the oxyanion is not stable toward the backward formation of **ESUB(a)**, as indicated by the free-energy profile (Figure 4). Notice that in our simulations the proton dynamics is treated at the classical level, thus it is not possible to establish the proposed relevance of tunneling for stabilizing an oxyanion reaction intermediate.²⁹ A typical $O\cdots H$ tunneling distance is ~1.6 Å. During the QM/MM MD simulation, this distance is never approached by any proton located on the Asp dyad; however, this event might occur on longer time scales than these presently investigated.

In the simulation starting from **ESUB(b)** (Figure 3), after a few picoseconds a hydrogen bond is formed between Oδ2_(Asp25) and $O_{\text{(SUB)}}$. At a $C_{\text{(SUB)}}\text{--}O_{\text{(WAT)}}$ distance of 1.88 Å, a concerted double proton transfer from Oδ2_(Asp25) to $O_{\text{(SUB)}}$ and from $O_{\text{(WAT)}}$ to Oδ2_(Asp25') is observed (**TS1** Figure 3). At the same time, the force on the constraint becomes negative, indicating a transition-state crossing (Figure 4). Subsequently, the gem-diol interme-

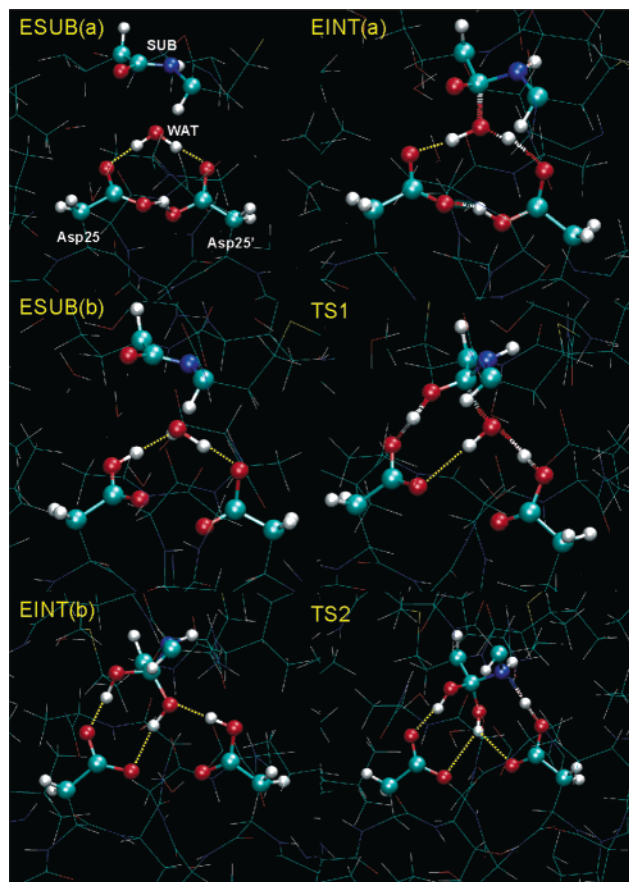


Figure 3. QM/MM simulation of HIV-1 PR: selected snapshots of the reaction representing **ESUB(a)**, **EINT(a)**, **ESUB(b)**, **TS1**, **EINT(b)**, and **TS2**. For labeling, see Figure 1. The QM atoms are represented as balls and sticks; the other atoms are represented as lines. Hydrogen bonds are indicated as yellow dashed lines. Only the active-site residues are shown for clarity.

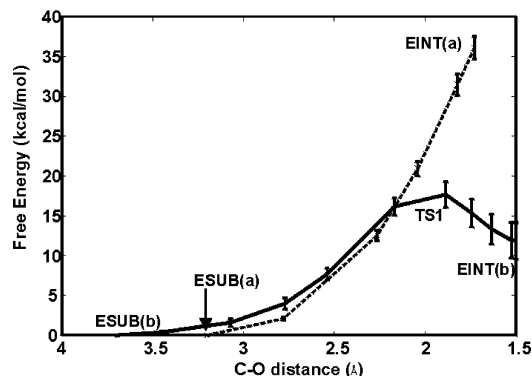


Figure 4. Free-energy profile for the first reaction step. Dashed line: QM/MM MD simulation from **ESUB(a)** to **EINT(a)**. Solid line: QM/MM MD simulation from **ESUB(b)** to **EINT(b)**. Error bars are also indicated.

diolate **EINT(b)** is formed (Figure 3). The calculated activation free energy for the formation of the gem-diol intermediate (ΔF_1) is 18(1.5) kcal mol⁻¹ (Figure 4). The reverse simulation from **EINT(b)** to **ESUB(b)** was also performed, and essentially the same activation free energy ($\Delta F_1 = 18(1)$ kcal mol⁻¹) was obtained. On the basis of ¹⁸O incorporation experiments, it has been suggested that two distinct pathways are possible for the backward reaction from the gem-diol to the substrate.¹⁸ The two pathways are not expected to be equivalent²⁸ because the enzyme's 2-fold symmetry is disrupted by substrate binding. In the present work, only one of the two possible reaction paths

was investigated. In particular, the C_(INTb)–O_{2(INTb)} distance was chosen as a reaction coordinate because O_{2(INTb)} interacts with both of the Asp residues and thus is the most reactive hydroxyl group of the gem-diol. A rotation around the C_(INTb)–N_(INTb) bond would bring O_{1(INTb)} in the same position, thus making this group reactive too. Such a rotation has never been observed during a 7.5-ns classical MD simulation of **EINT**.³²

The calculated ΔF_1 value (18(1.5) kcal mol⁻¹) is much smaller than that obtained for an uncatalyzed peptide hydrolysis in liquid water obtained with a similar computational setup (about 44 kcal mol⁻¹).⁴³ In HIV-1 PR/substrate complex **ESUB(b)**, water molecule **WAT** is located between the two aspartic acids, forming hydrogen bonds with both groups (Figure 2). The calculated dipole moment for **WAT** is 3.2(0.2) D. Although it is considerably larger than the dipole moment calculated for a water molecule in vacuo (1.9 D), it is only slightly more polarized than that of liquid water, as obtained with a very similar computational setup (3.0 D⁴⁴). The polarization of the reactants is similar in the two systems, thus the enzyme appears to lower the free energy by providing a proton donor and a proton acceptor group in the proper positions (**TS1**, Figure 3).

Our calculated value of ΔF_1 is also smaller (by about 2 kcal mol⁻¹) than that obtained with full ab initio CPMD calculations on gas-phase models of the adduct.¹⁵ This small difference can be ascribed in principle to the presence of the protein electrostatic field and/or to the lower flexibility of the complex, which are included only in the QM/MM approach. In fact, in the gas-phase models, the protein electric field, besides that generated by the Thr26(26')–Gly27(27') peptide bond, was not included and the terminal C α atoms were rigidly constrained to their starting positions,¹⁵ whereas they are linked to the flexible protein frame in the present QM/MM approach. Indeed, the mean displacements of the Asp dyad and the substrate C α atoms relative to the initial structure are small but significant (~ 0.3 Å for each atom). This could cause a decrease in ΔF_1 by few kcal mol⁻¹ because ΔF_1 has been shown to be very sensitive to the substrate–Asp distance.¹⁵

We now attempt to dissect the relative importance of these two effects (the protein electric field and protein flexibility).

To investigate the effect of the protein electric field, we performed QM/MM MD simulations of the first reaction step in the absence of such a field. The calculated structural properties of **ESUB(b)** complexes and the dipole moment of **WAT** (3.2(0.2) D) turn out to be very similar to those of our QM/MM calculations (3.2(0.2) D). However, ΔF_1 is significantly smaller (15.0(0.5) kcal mol⁻¹). This result, surprising at first, can be rationalized by realizing that a fundamental ingredient of HIV-1 PR catalytic power is the low polarity of the cleavage site environment¹⁵ that destabilizes the negatively charged Asp dyad. The removal of the electrostatic field of the protein (and, in particular, of the Thr26(26')–Gly27(27') peptide units, which directly interact with the Asp dyad negative charge^{11,30}) decreases the polarity of the dielectric medium. As a consequence, the negative charge of the cleavage site is further destabilized, thus enhancing the catalytic power of the Asp dyad and causing a decrease in ΔF_1 . We conclude that the protein electric field outside the cleavage site does not decrease the calculated ΔF_1 .

To estimate the effect of the protein flexibility, an analogous test QM/MM MD simulation of the first reaction step was performed in which the C α atoms were kept fixed in their initial positions. The calculated ΔF_1 turns out to be 25(0.2) kcal mol⁻¹, which is significantly larger than that of the calculated values without position constraints. We conclude that the flexibility

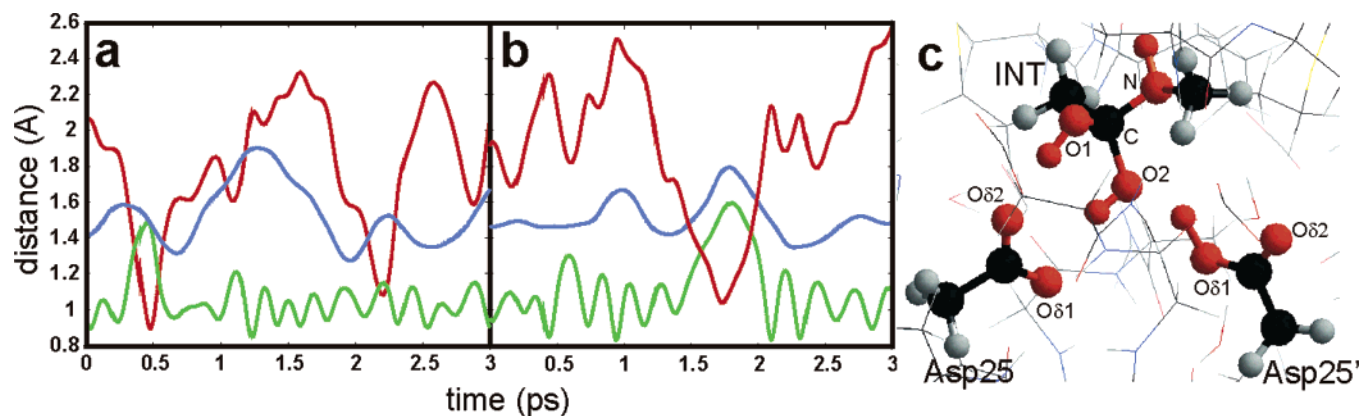


Figure 5. CAFES QM/MM simulation of the HIV-1 PR/INT complex. (a, b) Selected distances (Å) are plotted as a function of simulated time (ps). (a) $O2_{(INT)}-H_{(INT)}\cdots O\delta1_{(Asp25)}$ hydrogen bond distance (red), $O2_{(INT)}-H_{(INT)}$ bond distance (green), and $C_{(INT)}-O1_{(INT)}$ bond distance (blue). (b) $O\delta1-H_{(Asp25')}\cdots N_{(INT)}$ hydrogen bond distance (red), $O\delta1-H_{(Asp25')}$ bond distance (green), and $C_{(INT)}-N_{(INT)}$ bond distance (blue). (c) Structure of HIV-1 PR/INTb active site. The atoms held at 3000 K in the CAFES simulation are red.

of the protein frame is very important for determining the reaction free energy.

The calculated free energy for gem-diol intermediate **EINT(b)** is 12(2) kcal mol⁻¹ (Figure 3); a slightly larger value (13-(2) kcal mol⁻¹) was obtained for the reverse simulation.

Influence of the Choice of QM and MM Regions. In a QM/MM calculation, it is important to select carefully the region of the system to be treated at the ab initio level. The residues that were treated at the QM level in the present calculation were the two carboxylic acids (up to the C β) and the gem-diol (up to the C α) (Figures 2 and 3). Previous ab initio calculations³⁰ have shown that the interaction between the Asp dyad and the Thr26(26')-Gly27(27') peptide bond dipole moment is important in order to maintain Asp dyad coplanarity. Other studies have also suggested that this interaction is important for the active-site energetics.^{8,10,11} To investigate the influence of the level of theory (QM or MM) in treating these moieties, we performed QM/MM calculations of the energy difference between **ESUB(b)** and **EINT(b)** in two identical systems characterized by different quantum regions. In the first system (large QM), the Thr26(26')-Gly27(27') peptide bond was included in the QM region; in the second (small QM), it was not. A free-energy calculation with the method of constraints on the large QM system is computationally too demanding to be performed at the present stage. For this reason, total energy differences between geometry-optimized structures were calculated. Single-point calculations were also performed with the BP^{45,46} and the PBE⁴⁷ exchange correlation functionals using the geometries optimized at the BLYP level of theory. It turns out that the differences in relative energies between the large QM and small QM systems are only 0.9 (QM/MM BLYP) and 0.1 (QM/MM BP and QM/MM PBE) kcal mol⁻¹. Moreover, the rmsd, calculated for the QM atoms, between the two systems is less than 0.1 Å. These results indicate that the MM representation of the Thr26(26')-Gly27(27') peptide bonds is sufficient and that their interaction with the Asp dyad is correctly described by the QM/MM approach; for this reason, the Thr26(26')-Gly27(27') peptide bonds were not treated at the QM level in the present work.

Reaction Coordinate Identification for the Decomposition of the Hydrated Intermediate. An appropriate approximate reaction coordinate for the decomposition of the hydrated intermediate has not been established yet. Therefore, a total of 3 ps of QM/MM canonical adiabatic free-energy sampling (CAFES)⁴² was performed to identify a suitable approximate reaction coordinate. In this approach, a small group of "reactive"

atoms are heated to high temperature (i.e., 3000 K). These atoms cross free-energy barriers more easily than the others, sampling a larger part of the free-energy surface in an unbiased way, whereas the rest of the system is kept at room temperature. To avoid heat transfer between the two subsystems, they are dynamically decoupled by assigning to the reactive atoms masses 100 times larger than those of the rest.

The QM/MM-CAFES calculations were performed on the complex between the protein and **INT(b)** (Figure 3), as obtained from previous classical MD calculations.³² The reactive atoms were the oxygen and the polar hydrogen atoms of the Asp dyad and of the gem-diol as well as the gem-diol nitrogen atom (Figure 5c). In the QM/MM-CAFES simulation, two different events are observed: (i) Proton transfer from the gem-diol oxygen belonging to **INT(b)** ($O2_{(INT)}$) to Asp25 (Figure 5a). This event occurs twice in the simulation and leads to the formation of a short-lived oxyanion species. However, it is not followed by reactive events leading to the formation of the reactants (i.e., the weakening of the $C_{(INT)}-O1_{(INT)}$ bond and the double proton-transfer event stabilizing **TS1**, Figure 3) or products, namely, the weakening of the $C_{(INT)}-N_{(INT)}$ bond of the gem-diol. Thus, our QM/MM-CAFES calculation supports the proposal¹⁸ that the gem-diol is a stable reaction intermediate rather than a short-lived transient state. Furthermore, it indicates that, in the active site of HIV-1 PR, the gem-diol is more stable than the oxyanion. (ii) Proton transfer from Asp25' to the gem-diol intermediate nitrogen atom, followed by a reactive event (i.e., the reaction toward the products, namely, the weakening and consequently the elongation of the $C_{(INT)}-N_{(INT)}$ bond, Figure 5b). This proton transfer is not correlated with the weakening of other bonds involving **INT(b)** such as the $C_{(INT)}-O1_{(INT)}$ or $C_{(INT)}-O2_{(INT)}$ bond (Figure 5a and b). Although in the CAFES simulation the transition state for the decomposition of the hydrated intermediate is not reached, because no decomposition is observed, the $N_{(INT)}-H_{(Asp25')}$ distance appears to be a suitable independent reaction coordinate to approach **TS2**. This finding is also consistent with a previous ab initio study.¹⁴

Decomposition of Hydrated Intermediate **EINT(b).** The second step investigated here is the decomposition of the gem-diol and the formation of the reaction products (from **EINT(b)** to **EPROD**). Here, the $N_{(INT)}-H_{(Asp25')}$ distance was assumed to be the reaction coordinate based on the CAFES calculations reported above. This distance was constrained and progressively shortened from 3.17 to 1.05 Å during 10 ps of QM/MM simulation. It has been proposed that the gem-diol intermediate

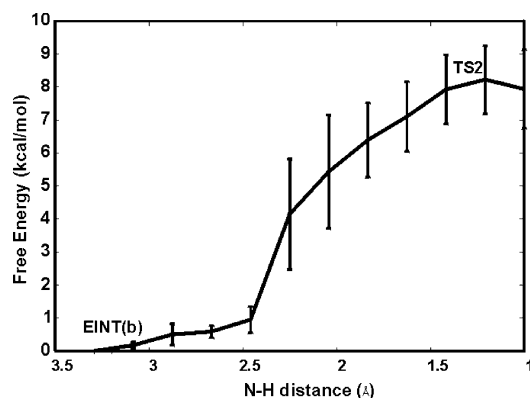


Figure 6. Free-energy profile for the second reaction step. QM/MM MD simulation from **EINT(b)** to **TS2**. Error bars are indicated.

is characterized by an anti and a gauche conformation of similar energy and that the gauche conformation is the conformation relevant to catalysis.⁸ Our results are consistent with this proposal. In the classical MD simulation of the gem-diol intermediate,³² both conformations are populated by the anti conformation about four times more frequently than by the gauche form. In the early steps of the QM/MM MD simulation ($d_{\text{N-H}} \approx 3\text{--}2.5$ Å), the gem-diol is in an anti conformation, and $\text{H}_{(\text{Asp25}')} \cdots \text{H}_{(\text{O1}_{\text{INT}})}$ forms a strong, short hydrogen bond with $\text{O2}_{(\text{INT})}$ (Figure 3). As the reaction proceeds, this hydrogen bond is broken to form a weaker hydrogen bond between Asp25' and the gem-diol nitrogen $\text{N}_{(\text{INT})}$. At the same time, an anti-gauche transition is observed ($d_{\text{N-H}} \approx 2.5\text{--}2$ Å). The QM/MM calculation indicates that the overall process requires 4 to 6 kcal mol⁻¹ to occur. This value is consistent with the spontaneous formation of a short-lived $\text{N}_{(\text{INT})} \cdots \text{H}_{(\text{Asp25}')} \cdots \text{H}_{(\text{O1}_{\text{INT}})}$ hydrogen bond in an 11-ns classical MD simulation of **EINT(b)**.³²

At a $\text{N}_{(\text{INT})}\text{--}\text{H}_{(\text{Asp25}')} \cdots \text{H}_{(\text{O1}_{\text{INT}})}$ distance of 1.35 Å, the constraint force changes its sign, indicating transition-state crossing and the formation of the N-H covalent bond (Figure 6). After this point, the constraint was released, and 2 ps of QM/MM simulation without constraints was performed. After a few tenths of a picosecond, a proton is transferred from the gem-diol $\text{O1}_{(\text{INT})}$ to Asp25 and the $\text{C}_{(\text{INT})}\text{--}\text{N}_{(\text{INT})}$ peptide bond spontaneously breaks to form the reaction products, namely, a carboxylic acid and an amine. The calculated activation free-energy barrier (ΔF_2) is 8(1) kcal mol⁻¹ with respect to gem-diol intermediate complex **EINT(b)** (Figure 6) or 21(3) kcal mol⁻¹ with respect to reactants **ESUB(a)**. Also, the transition from **EINT(b)** to **EPROD** state is characterized by a double proton transfer: one from Asp25' to the gem-diol nitrogen and one from the gem-diol O1 to Asp25. However, in this case, the second proton transfer is not concerted with the first but follows the formation of a metastable N-protonated hydrated intermediate.

The experimental k_{cat} for the TIMMQR substrate used in this study is on the order of 10 s⁻¹.^{25,48} This value puts an upper bound of ~ 16 kcal mol⁻¹ on the activation free energy of any step of the catalysis. Clearly, the ΔF_2 calculated here for the second reaction step (21(3) kcal mol⁻¹ with respect to the reactants) is somewhat too large with respect to the value expected from the experimental measurements. A comparison of the performance of several ab initio methods for gas phase models of the catalyzed peptide bond hydrolysis reaction (see Methods) strongly suggest that most of the discrepancy is likely to arise from the tendency of the DFT-BLYP method to overestimate the energy of the gem-diol intermediate and all of the subsequent steps. We expect that the value of the transition-state energies could change significantly in substrates

that are characterized by a markedly different flexibility, as in the case of the cleavage of the Phe-Pro bond. In this respect, it is interesting that in a previous ab initio investigation at the HF level of the Phe-Pro peptide bond cleavage by HIV-1 PR the gem-diol intermediate is found to be 5 kcal mol⁻¹ more stable than the reactants.¹⁴ This surprising result might be ascribed to the lower resonance energy present in the Phe-Pro double bond due to the geometric constraints imposed by the proline ring that destabilizes the reactants with respect to the gem-diol intermediate. However, the reported activation energies for **TS1** (16.26 kcal mol⁻¹) and **TS2** (23.95 kcal mol⁻¹)¹⁴ are similar to those calculated in the present work.

The ¹⁵N kinetic isotope effect in the second step of the reaction was estimated from the zero-point energy difference between the hydrated intermediate and the transition state (see Methods). To this aim, vibrational modes for the QM atoms were calculated from the QM/MM MD simulations by the diagonalization of the mass-weighted covariance matrix. An inverse ¹⁵N kinetic isotope effect is found, in agreement with the experiment.²⁸ However, the calculated effect of 0.97(0.02) is larger than the effect observed experimentally, possibly because of the particular properties of the substrate investigated in the present work²⁵ and/or the numerical noise in the vibrational mode calculation arising from the limited length of the QM/MM MD trajectories.

Our final QM/MM structure can be compared with the available structural data, namely, with the X-ray structure of HIV-1 PR in complex with reaction products Ac-S-L-N-F↓ and ↓P-I-V at 2.2–2.3-Å resolution.²⁰ From the examination of the crystal structure of the HIV-1 PR/product complex, the authors suggested that no strain is induced on the reactants by the protein frame.²⁰ Our results support this hypothesis; in the 2 ps of QM/MM simulation, after the $\text{C}_{(\text{INT})}\text{--}\text{N}_{(\text{INT})}$ bond is broken, the two tails of the substrate do not drift apart but remain in van der Waals contact.¹ Furthermore, the conformation of the reaction products at the cleavage site turns out to be similar in the two structures; in particular, the characteristic bifurcated binding mode of the substrate –COOH tail that was observed in the X-ray structure is reproduced by the QM/MM simulation, as indicated by the low rmsd (calculated for the reaction products and the Asp dyad) of 0.9 Å.

In an attempt to calculate the relative energy of reaction products **EPROD**, additional QM/MM calculations with constraints on the N-H or C-N bond distances were performed. In both cases, the constraint force exhibits large oscillations, and it was not possible to obtain a converged average within the picosecond time scale of QM/MM simulations. This indicates that the N-H and C-N distances alone are not appropriate reaction coordinates for the simulation of the decomposition of the N-protonated hydrated intermediate.

Concluding Remarks

The overall reaction of peptide bond hydrolysis in HIV-1 PR has been described using QM/MM simulations. The cleavage site has been studied at the DFT-BLYP level of theory with a PW basis set. The two reaction coordinates that were constrained in order to simulate the hydrolysis reaction were (i) the $\text{C}_{(\text{SUB})}\text{--}\text{O}_{(\text{WAT})}$ distance, which has already been suggested in previous studies,^{6,7,15} and (ii) the $\text{N}_{(\text{INT})}\text{--}\text{H}_{(\text{Asp25})}$ distance, which has been identified in this work by performing an unbiased scan of the free-energy surface with the CAFES method.⁴² The reaction is characterized by two distinct energy barriers, one for the nucleophilic attack of the water molecule (**TS1**) and the second for the protonation of the peptide bond nitrogen (**TS2**) that leads

TABLE 1: Energetics of Gas-Phase Model 1^a

	6-31G	6-31G (d)	6-31G (d,p)	6-31+G (df,2p)	6-31++G (3df,3p)
HF	16.3	18.6	18.7	20.1	19.3
BP86	20.3	21.6	19.9	21.2	20.7
BLYP	22.6	23.1	22.9	24.6	24.8
BLYP (PW)	23.2 ^{§50}	26.4 ^{§60}	26.4 ^{§70}	25.9 ^{§110}	25.8 ^{§150}
B3LYP	19.2	19.9	19.8	21.3	21.5
MP2		17.7	16.8	17.1	17.1 ^b
MP4			14.5		
CCD			14.4		

^a Energy differences (kcal mol⁻¹) between 1b and 1a for model 1 (ΔE in the text) calculated with a variety of ab initio methods and with different basis sets (see Methods). All structures were fully geometry optimized. ^b This number has been obtained with the 6-31+G(2df,2p) basis set. For BLYP(PW) calculations, DFT energies were calculated with the program CPMD,⁵⁴ the BLYP exchange-correlation functional, and a plane wave basis set. The size of the basis set is determined by the energy cutoff: ^{§50} cutoff 50 Ry; ^{§60} cutoff 60 Ry; ^{§70} cutoff 70 Ry; ^{§110} cutoff 110 Ry; ^{§150} cutoff 150 Ry. BLYP and B3LYP DFT energies were calculated with Gaussian 98⁶⁴ and the BLYP and B3LYP exchange correlation functionals; MP2 post-Hartree–Fock energies were calculated with Gaussian 98⁶⁴ and Møller–Plesset second-order perturbation theory. MP4 post-Hartree–Fock energies were calculated with Gaussian 98⁶⁴ and Møller–Plesset fourth-order perturbation theory. CCD post-Hartree–Fock energies were calculated with Gaussian 98⁶⁴ and the coupled-cluster doubles method.

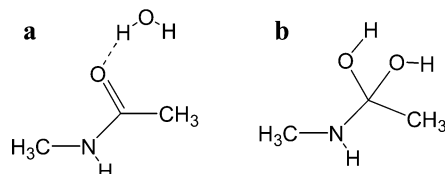


Figure 7. Chemical structure of model 1. (a) Peptide bond with a hydrogen-bonded water molecule. (b) gem-diol peptide bond hydrate.

to the decomposition of a hydrated amide intermediate. The calculated reaction barriers are 18(1) and 21(3) kcal mol⁻¹, respectively.

A comparison of our QM/MM results with previous studies and with the analogous QM/MM simulations in which the protein electric field is switched off indicates that, at least for the first reaction step, the electrostatic field generated by the residues surrounding the cleavage site plays a fundamental role in stabilizing the Asp dyad. However, the net effect of this field on the reaction free energy is an increase of 3 kcal mol⁻¹. Moreover, QM/MM calculations, in which constraints are applied to the Asp dyad and the substrate C α atoms, establish that the local flexibility of the cleavage site is a key factor for the energetics of the reaction barrier, consistent with our previous proposals.¹⁵ This finding highlights the importance of a computational method that explicitly takes into account the constraints and flexibility imposed by the protein frame in the studies of the reaction mechanism of HIV-1 PR.

The accuracy of the method used here has been investigated by comparison to a variety of ab initio (HF, DFT, MP2, and CCD, Tables 1 and 3) methods and the QM/MM approach on gas-phase models 1 and 2 (Figures 8 and 9). The conclusions are that (i) the convergence with respect to the basis set size within 1 kcal mol⁻¹ can be obtained using a PW basis set with a cutoff of 70 Ry or more and (ii) the DFT methods tested (BLYP, B3LYP, BP86) and the QM/MM method predict reasonable transition-state energies (within 2 kcal mol⁻¹), whereas they severely underestimate the stability of the gem-diol intermediate relative to MP2 and CCD calculations.

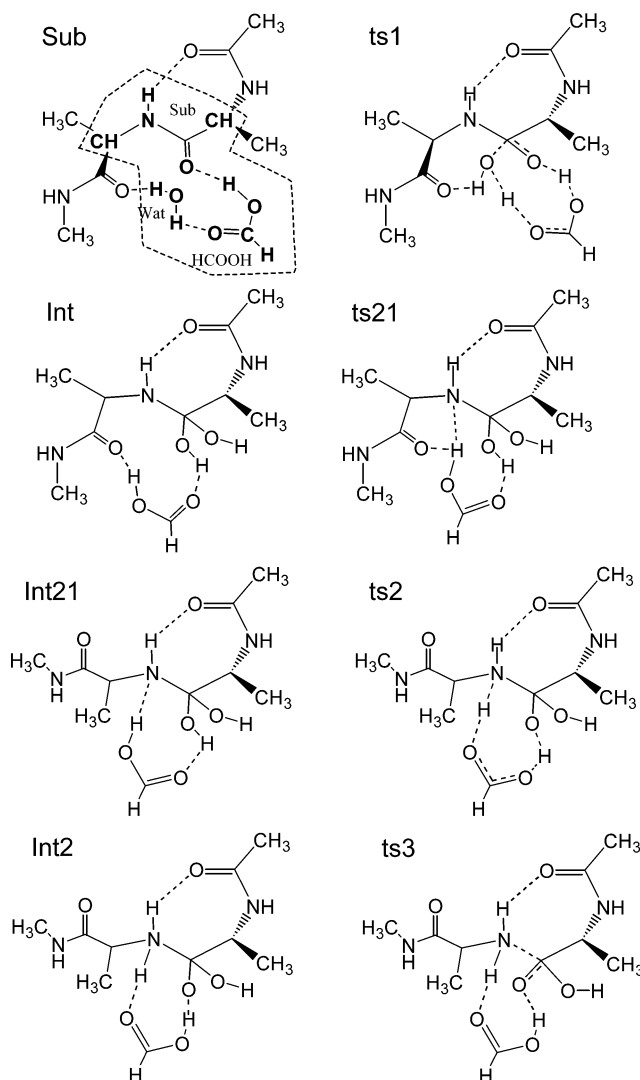


Figure 8. Chemical structure of model 2. In **Sub**, the QM/MM partitioning of the system is also indicated: atoms enclosed within the dashed line (in bold) were included in the QM part of the system.

Methods

Structural Models. The structural model of HIV-1 PR complexed with the substrate and the gem-diol reaction intermediate (**ESUB** and **EINT**, SUB = Thr-Ile-Met-Met-Gln-Arg) were built from the X-ray structure of HIV-1 PR complexed with MVT101^{18,25} (4HVP⁴⁹). Protein residues belonging to subunit 1 were numbered from 1 to 99, and those belonging to subunit 2, from 1' to 99'. Substrate residues binding to subunit 1 (2) were numbered from P1 (P1') to P3 (P3'): ⁵⁰ Thr (P3)-Ile (P2)-Met (P1)-Met (P1')-Gln (P2')-Arg (P3'). The peptide bond to be cleaved by the enzyme belongs to residues P1 and P1' (Met-Met). In the simulation of the reaction intermediate, the carbonyl group of Met P1' was substituted by a gem-diol. The total system was composed of 15 749 atoms (Figure 2).

Starting structures were taken from 7.5-ns classical molecular dynamics (MD) of **ESUB** and **EINT**.³² For **ESUB**, the structure obtained after 3270 ps of classical MD simulation was chosen because it is expected to be the most reactive conformation sampled.¹⁵ For **EINT**, the structure obtained after 478 ps of classical MD simulation was chosen as a starting structure because it closely resembled the average MD structure (C α atoms' rmsd = 1.0 Å).

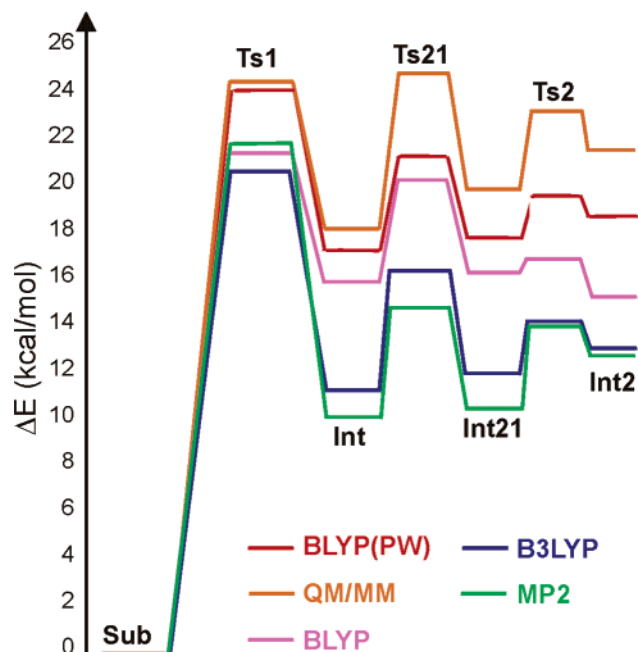


Figure 9. Energetics of gas-phase model 2. Labeling is the same as in Figure 8 and Tables 1 and 3.

Complexes **1** and **2** were built starting from the structure of **EINT**. In complex **1**, only the atoms belonging to the P1–P1' gem-diol moiety were retained (Figure 7). In complex **2**, P1 and P1' were included up to the C β , and the dangling peptide bonds were terminated with acetyl and *N*-methyl on the C and N termini, respectively. The carboxyl group of Asp25 was also included. All of the remaining dangling bonds were capped with hydrogen atoms. The final system was composed of 40 atoms (Figure 8).

QM/MM Calculations. These calculations were carried out with the QM/MM scheme developed in ref 37.

The quantum part included the side chains of Asp 25 and Asp 25', the hydrated peptide bond, and a water molecule (Figure 3). Dangling bonds in the quantum part were terminated with hydrogen atoms. The small unbalanced charge (0.1 e) resulting from the choice of the QM/MM partitioning was redistributed over the 500 MM atoms located within 7.5 Å of the Asp dyad.

The quantum problem was solved using density functional theory (DFT). Exchange and correlation functionals were those of Becke³⁸ and Lee, Yang, and Parr⁴⁰ (BLYP), respectively. The Kohn–Sham orbitals were expanded in plane waves up to 70 Ry. Martins–Troullier⁵¹ pseudopotentials were used to describe the interactions between the ionic cores and the valence electrons. A $14.0 \times 14.0 \times 10.5$ Å³ quantum cell was used. The systems were treated as if they were isolated, as in Barnett and Landman.⁵² DFT-based MD simulations were performed according to the Car–Parrinello approach⁵³ using the CPMD program.⁵⁴ A time step of 5 au and a fictitious electron mass of 600 au were used. Constant temperature was achieved by coupling the systems to a Nosé–Hoover thermostat⁵⁵ of 500-cm⁻¹ frequency. The GROMOS96⁵⁶ program combined with the Amber94 force field⁵⁷ was used to treat the classical system. A cutoff of 12.0 Å was used for nonbonded interactions, and the P3M method⁵⁸ was used to describe long-range electrostatics.

QM/MM geometry optimizations were performed on large QM and small QM systems with a conjugate gradient algorithm up to a convergence of 5×10^{-3} hartree Å⁻¹ on the forces and 2×10^{-6} hartree step⁻¹ on the energies. In the small QM

systems, the Asp dyad and the hydrated peptide bond were treated at the QM level. In the large QM systems, the Thr26-(26')-Gly27(27') peptide bond was also treated at the QM level. During the geometry optimizations, only the Asp dyad, the hydrated peptide bond, and the Thr26(26')-Gly27(27') peptide bonds were allowed to move.

Free energies were calculated with the method of constraints.^{41,59–61} The activation free energies were calculated as an integral of the average force f_s acting on the constraint along the reaction coordinate Q .⁶²

The selected constraint does not necessarily have to correspond to the entire reaction coordinate. However, it turns out that the time required to obtain a converged MD average of the force for each constraint point depends critically on how close the chosen constraint matches the slowest part of the reaction coordinate. For this reason, it is important to evaluate which geometric parameter most closely matches the reaction coordinate for each step of the reaction. To calculate the structure and relative free energy of **ESUB(a)** and **ESUB(b)**, the **ESUB(a)** complex was first equilibrated with 0.5 ps of QM/MM MD. Subsequently, the O δ 2_(Asp25)–H1_(WAT) distance was constrained and reduced from 3.7 to 1.07 Å in 4 ps of constrained QM/MM MD simulation. The obtained **ESUB(a)** and **ESUB(b)** structures were used for the simulation of the first reaction step, namely, the formation of the hydrated intermediate.

The first step involves the nucleophilic attack of a water molecule on the carbonyl carbon of the peptide bond of the substrate. Previous studies have established that $\xi_{CO} = d(C_{(SUB)}-O_{(WAT)})$ is an appropriate reaction coordinate for describing this part of the reaction.^{6,7,15}

For chemical step 2, several reaction coordinates are possible. For this reason, 3.0 ps of QM/MM–CAFES⁴² simulation were performed to obtain qualitative information about the possible reaction coordinates. In the CAFES simulation, the temperature of the QM oxygen and nitrogen atoms and polar protons (Figure 3c) was set to 3000 K via coupling to a Nosé–Hoover thermostat⁵⁵ with $T = 3000$ K and a coupling frequency of 500 cm⁻¹. The masses of these atoms were increased by a factor of 100 to obtain decoupling between the “hot” and “cold” parts of the system. As a result of these calculations, the N_(INT)–H_(Asp25) distance (ξ_{NH}) turned out to be the most plausible reaction coordinate.

Each point along ξ_{CO} and ξ_{NH} was sampled until the MD-averaged forces calculated from the first and second parts of the trajectory differed by less than 10%. About 1.2 ps of ab initio MD was required for most of the points. About 2.0 ps was required for the points closer to the transition states. The initial 0.3 ps was always discarded. Overall, ~15 and 10 ps were sampled for chemical step 1 and chemical step 2, respectively.

A reverse simulation of chemical step 1 was also performed starting from **EINT(b)**; this control simulation gave essentially the same results in terms of transition-state structure and energy as the direct reaction. Two additional QM/MM MD simulations of chemical step 1 were also performed with the same computational scheme outlined above but without electrostatic interactions between the classical and the quantum system or with position constraints on the Asp dyad and gem-diol C α atoms. The error in the calculated free energy is reported in parentheses in the text and was estimated from the difference between the values calculated for the first and second halves of the simulation.

The dipole moment of the catalytic water molecule in the active site of HIV-1 PR was calculated as an average over 40

structures selected at regular intervals from the simulation of the **ESUB(b)** complex (Figure 3).

Kinetic isotope effects (KIE) were calculated as in ref 63

$$\text{KIE} = \text{MMI} \cdot \text{ZPE} \cdot \text{EXC}$$

where MMI is the contribution due to the change in the moments of inertia

$$\text{MMI} = \frac{\left(\frac{{}^{14}M}{{}^{15}M} \right)^{3/2} \frac{{}^{14}I_x {}^{14}I_y {}^{14}I_z}{{}^{15}I_x {}^{15}I_y {}^{15}I_z}}{\left(\frac{{}^{14}M}{{}^{15}M} \right)^{3/2} \frac{{}^{14}I_x {}^{14}I_y {}^{14}I_z}{{}^{15}I_x {}^{15}I_y {}^{15}I_z}}_{\text{reactant}}$$

ZPE is the zero-point energy contribution⁶³

$$\text{ZPE} = \frac{\left[\prod_i^{3N-1} e^{-(14u_i - 15u_i)/2} \right]_{\text{TS}}}{\left[\prod_i^{3N-6} e^{-(14u_i - 15u_i)/2} \right]_{\text{reactant}}}$$

and EXC is the excited-state contribution

$$\text{EXC} = \frac{\left[\prod_i^{3N-1} \frac{1 - e^{-(15u_i)}}{1 - e^{-(14u_i)}} \right]_{\text{TS}}}{\left[\prod_i^{3N-6} \frac{1 - e^{-(15u_i)}}{1 - e^{-(14u_i)}} \right]_{\text{reactant}}}$$

where N is the number of atoms in the system. ${}^{14}M$ is the mass of the ${}^{14}\text{N}$ system, and ${}^{14}I_x$ is the moment of inertia of the ${}^{14}\text{N}$ system with respect to the x axis. The number of vibrational modes is $3N - 6$ for the reactants and $3N - 7$ for the TS as, in this approach, one degree of freedom is constrained. ${}^{14}u_i$ is the i th vibrational mode for the ${}^{14}\text{N}$ system, and ${}^{15}u_i$ is the i th vibrational mode for the ${}^{15}\text{N}$ system:

$$u_i = \frac{h\nu_i}{k_{\text{B}}T}$$

ν_i is the frequency of the i th vibrational mode. Vibrational modes were calculated for the QM atoms only from diagonalization of the mass-weighted covariance matrix of the QM atoms:

$$C_{ij} = \langle M_{ii}^{1/2}(x_i - \langle x_i \rangle) M_{jj}^{1/2}(x_j - \langle x_j \rangle) \rangle$$

M is the diagonal matrix of atomic masses. A comparison between formamide and protonated formamide ZPEs calculated with this method and with Gaussian 98⁶⁴ indicates that about 0.5 ps of ab initio MD trajectory is required to obtain ZPEs converged to within 10%.

Gas-Phase Calculations. All of the calculations with localized basis sets were carried out with the Gaussian 98⁶⁴ program on two models, **1** and **2** (Figures 7 and 8).

1 was energy minimized until the largest component of the force was smaller than 2.5×10^{-3} a.u. and the largest displacement was smaller than 1.0×10^{-2} Å. Basis sets of increasing size, ranging from 6-31G to 6-31++G(3df,3p), were used. DFT calculations were performed with the BLYP, BP86, and B3LYP exchange-correlation functionals. Post-Hartree–Fock calcula-

tions were performed according to Møller–Plesset second and fourth order perturbation theory (MP2 and MP4) and with the coupled-cluster doubles method (CCD).

For model **2**, the reaction profiles of chemical steps 1 and 2 were investigated by a scan along the $\text{C}_{(\text{Sub})}-\text{O}_{(\text{Wat})}$ and $\text{N}_{(\text{Int})}-\text{H}_{(\text{HCOOH})}$ distances, respectively. The $\text{C}_{(\text{Sub})}-\text{O}_{(\text{Wat})}$ and $\text{N}_{(\text{Int})}-\text{H}_{(\text{HCOOH})}$ distances were scanned at the B3LYP-6-31G(d) level of theory. The 6-31G(d) basis set was used because it strikes a reasonable compromise between accuracy and computational cost (Table 1). A control calculation was also performed with B3LYP and the 6-31G(d,p) basis set and gave essentially the same results in terms of structure and energetics.

The $\text{C}_{(\text{Sub})}-\text{O}_{(\text{Wat})}$ distance was constrained, and the constraint value decreased from 3.61 to 1.21 Å in 22 steps; the $\text{N}_{(\text{Int})}-\text{H}_{(\text{HCOOH})}$ distance was constrained, and the constraint decreased from 4.77 to 0.87 Å in 33 steps. In each step, the structure was energy minimized until the largest component of the force was smaller than 2.5×10^{-3} au and the largest displacement was smaller than 1.0×10^{-2} Å.

The intermediates or transition-state structures were also energy minimized at the BLYP-6-31G(d) and BLYP-PW levels of theory with a 70-Ry cutoff, as well as with the QM/MM scheme outlined in the previous section. In the QM/MM calculations, the QM part of the system included the water molecule, the formic acid, and the Ala-Ala peptide bond represented as *N*-methyl amide (Figure 6). Dangling bonds were terminated with hydrogen atoms. Classical atoms were described with the Amber94⁵⁷ force field. Finally, single-point calculations were also performed with Møller–Plesset second-order perturbation theory (MP2) using the B3LYP geometries.

Accuracy of the Methodology. The accuracy of the method used here (QM/MM DFT-BLYP using a plane wave (PW) basis set expanded up to 70 Ry) is established by performing first principles (HF, MP2, DFT-BP86, DFT-B3LYP, DFT-BLYP) and QM/MM calculations using both PW and Gaussian (G) basis sets. The calculations are carried out on two model systems representing some of the groups involved in the enzymatic reaction (Figures 7 and 8).

Complex **1** consists of a water molecule and either *N*-methyl acetamide (**1a**) or the corresponding gem diol hydrate (**1b**, Figure 7). The energy difference between the two systems (ΔE hereafter) is expected to be very sensitive to the type of approach used to calculate the exchange and correlation contributions because the two systems differ in the number of hydrogen bonds (Figure 7) and in the type of chemical bonds, namely, the *N*-methyl acetamide peptide bond is turned into two σ bonds in the intermediate. We exhaustively address this issue by energy minimizing **1** with different methods (BLYP, BP86, B3LYP, HF, MP2, MP4, and CCD) and basis sets (6-31G, 6-31G(d), 6-31G(d,p), 6-31G+(df,2p), and 6-31G++(3df,2p); see Table 1).

All of the methods employed turn out to predict similar geometries (Table 2). B3LYP is the DFT method that provides the most similar structural properties to those obtained with the CCD method, which is taken here as a reference (Table 2); in particular, this method accurately reproduces the hydrogen bond distance (Table 2). In contrast, ΔE depends largely on the level of theory. It is underestimated by all DFT methods by 5–7 kcal mol⁻¹ both using G and PW basis sets. This large deviation indicates that DFT significantly overestimates the stability of a resonant double bond (the peptide bond) plus a hydrogen bond with respect to two single bonds. Indeed, the inclusion of diffuse functions (basis sets 6-31G+(df,2p) and 6-31G++(3df,3p)), which is expected to improve the description of the peptide and

TABLE 2: Structural Properties of Gas-Phase Model 1^a

	HF	BLYP	BP86	B3LYP	MP2	CCD
C=O	1.245	1.248	1.232	1.232	1.232	1.222
N—C	1.375	1.375	1.363	1.363	1.362	1.362
C—C α	1.532	1.531	1.517	1.517	1.517	1.521
N—C α	1.464	1.465	1.450	1.450	1.450	1.452
C—N—C α	123.3	123.4	121.6	121.6	121.5	121.5
N—C—O	121.8	121.9	121.8	121.8	121.8	121.9
N—C—C α	115.5	115.6	115.0	115.0	115.1	115.1
O _{wat} —H ₁	0.987	0.992	0.968	0.968	0.968	0.964
O _{wat} —H ₂	0.974	0.980	0.974	0.958	0.958	0.957
H ₁ —O _{wat} —H ₂	101.9	102.3	102.1	102.1	102.2	102.6
H ₁ ...O	1.907	1.905	1.932	1.932	1.924	1.960
SD bonds	0.026	0.028	0.011	0.008	0.010	
SD angles	0.75	0.68	0.20	0.20	0.13	

^a The geometric parameters are calculated with different ab initio methods. The standard deviation (sd) for bonds and angles with respect to the CCD calculation is also reported. Abbreviations are the same as those in Table 1. All of the reported structures refer to the calculations with the 6-31G(d) basis set.

hydrogen bonds, causes an increase of ΔE by 1–2 kcal mol^{−1}. In contrast, the MP2 calculation of ΔE is not largely affected by the type of basis set used here (Table 1). ΔE calculated with BLYP-PW turns out not to be very sensitive to the size of the basis set used as long as a cutoff larger than 60 Ry is used (Table 1) and is slightly larger than those calculated with G basis sets, consistent with the observation that a diffuse basis set stabilizes **1a** with respect to **1b**.

We conclude that the DFT methods used here (BLYP, B3LYP, and BP86) provide accurate structural information but significantly underestimate the stability of gem-diol **1b** relative to that obtained from post-Hartree–Fock methods (MP2, MP4, CCD). Although these results do not appear to depend significantly on the quality of the PW basis set used (as far as the cutoff is larger than 70 Ry), in the case of a G basis set, the use of diffuse functions improves the accuracy of ΔE by 1–2 kcal mol^{−1}.

Complex **2** includes all of the elements required for a minimal description of the reaction catalyzed by a proton-donor–proton-acceptor group, namely, (Ace-Ala-Ala-Nme) (Sub hereafter), a water molecule (Wat), and a proton-donor–proton-acceptor group (HCOOH). In this case, all of the species depicted in Figure 6 (describing chemical steps 1 and 2) are energy minimized at constrained C_(Sub)–O_(Wat) or N_(Int)–H_(HCOOH) distances. Thus, this model allows us to describe the peptide bond hydrolysis catalyzed by formic acid as proton-donor–proton-acceptor group. In particular, it also allows us to compare the relative energies of reaction intermediates and transition states calculated with the QM/MM method used in the HIV-1 PR simulation with the energies calculated within a full QM description of the system at the BLYP-PW, BLYP-G, B3LYP-G, and MP2 levels of theory.

In the QM/MM calculations on complex **2**, the formic acid, the water molecule, and the peptide bond are treated at the first principles level, and the rest of the system is described with the Amber94 force field (Figure 8, Sub). The calculated structural parameters of the Sub, ts1, Int, ts21, Int21, ts2, and Int2 species calculated with the different QM methodologies are remarkably similar, the largest rmsd between two structures being as small as 0.1 Å. The rmsd between the structures predicted by the QM/MM calculations and the structures predicted by the full QM BLYP-PW calculations is in all cases smaller than 0.6 Å. The largest discrepancies between the QM/MM and the BLYP-PW structures are observed in the length of the hydrogen bonds between a classical and a quantum atom;

TABLE 3: Energetics of Gas-Phase Model 2^a

	Sub	ts1	Int	ts21	Int21	ts2	Int2
BLYP(PW)	0	23.9	16.9	20.9	17.6	19.1	18.1
QM/MM	0	24.2	18.0	24.8	20.0	23.3	21.7
BLYP	0	21.3	15.9	20.1	15.6	16.3	14.9
B3LYP	0	20.2	11.5	16.2	12.0	14.1	13.0
MP2	0	21.8	9.8	14.4	10.3	13.8	12.9

^a Profiles (in kcal mol^{−1}) for the hydrolysis reaction (Figure 8) obtained by a distance scan along the C–O and N–H distances. In the DFT calculations, all points along the reaction path were fully optimized. The MP2 calculations were carried out on the B3LYP geometries. Abbreviation are the same as those in Table 1. Gaussian 98⁶⁴ calculations were performed with the 6-31G(d) basis set. For BLYP(PW) and QM/MM calculations, the BLYP exchange–correlation functional and a plane wave basis set with an energy cutoff of 70 Ry were used. QM/MM DFT energies were calculated with the QM/MM approach of ref 37. The QM/MM partitioning is reported in Figure 8 and described in the text.

these are in general 0.05–0.2 Å shorter in the QM/MM calculations with respect to those in BLYP-PW. This observation indicates that the strength of the hydrogen bonds between classical and quantum atoms might be slightly overestimated by the QM/MM approach adopted here.

We now turn our attention to the energetics of the transition states and intermediates of the reaction. In chemical step 1, the energies calculated for the transition state (ts1, Figure 8) are in fair agreement and range from 20.2 for B3LYP to 24.2 kcal mol^{−1} for the BLYP/MM calculation (Table 3 and Figure 9). It turns out that the ts1 energy calculated with the QM/MM scheme (24.2 kcal mol^{−1}) is remarkably similar to the QM energy calculated with the BLYP exchange–correlation functional and plane wave basis set (23.9 kcal mol^{−1}) and 2.4 kcal mol^{−1} larger than the value calculated at the MP2 level. Larger discrepancies are observed for the predicted stability of the gem-diol intermediate (Int). As in the small-model calculations, the BLYP functional largely underestimates the stability of the gem-diol (between 15.9 and 18.0 kcal mol^{−1} compared to 9.8 kcal mol^{−1} obtained from the MP2 results) (Table 3 and Figure 9). The B3LYP functional performs much better in this respect (11.5 kcal mol^{−1}) and provides a substantial correction to the BLYP overestimation, leading to results that are intermediate between the BLYP and MP2 calculations.

Chemical step 2 in complex **2** is characterized by three transition states (ts21, ts2, and ts3; Figure 6). The first has the largest barrier (ts21, ~4–6 kcal/mol; Figure 9), and it is determined by the rupture of the O_(HCOOH)–H...O_(Int) hydrogen bond and the formation of the O_(HCOOH)–H...N_(Int) hydrogen bond. This barrier is well reproduced by all ab initio methods investigated and is only slightly overestimated in the QM/MM calculation (6.8 kcal mol^{−1}). The subsequent proton transfer from formic acid to the gem-diol nitrogen is characterized by a lower barrier (ts2, 2–3 kcal/mol; Figure 9) and leads to the formation of a third metastable zwitterionic intermediate (Int2). BLYP underestimates this barrier by about 1–2 kcal mol^{−1} with respect to B3LYP and MP2, whereas the QM/MM method somewhat compensates for the BLYP error and gives results in good agreement with the MP2 value (Table 3). A further scan along the C_(Int)–N_(Int) bond reveals that a very small barrier on the order of $k_B T$ (ts3) has to be crossed in order to break the C_(Int)–N_(Int) bond and form the final products.

In summary, the peptide hydrolysis performed by complex **2** is characterized by a barrier of about 22 kcal mol^{−1} for the first step of the reaction (ts1) and three smaller barriers of ~4, ~3, and <1 kcal mol^{−1} for the decay of the gem-diol intermediate (ts21, ts2, and ts3; Figures 8 and 9) using the first principles

and QM/MM methods indicated in Table 3. The QM/MM calculations provide a reasonable description of the reaction energetics because the relative energy of ts1 is only slightly overestimated by about 2 kcal mol⁻¹ and the proton-transfer barrier (ts2) is underestimated by about 1–2 kcal mol⁻¹. However, because of the use of the BLYP exchange-correlation functional, the QM/MM calculations adopted here largely underestimate the stability of the gem-diol intermediate and all subsequent structures with respect to B3LYP and MP2 calculations by about 5–8 kcal mol⁻¹.

The energy landscape of gas-phase model 2 differs from that obtained for the real protein. Several factors may play a role in the observed differences: (i) the different structure of the proton shuttle system (two aspartic acid residues in HIV-1 PR, only one formic acid in model 2); (ii) in the HIV-1 PR QM/MM calculations, the constraints imposed by the protein frame do not allow the formation of a strong H_(Asp25)...N_(Int) hydrogen bond, which stabilizes intermediate Int21 (Figure 8); and (iii) thermal effects, which are included only in HIV-1 PR QM/MM calculations, are on the order of the energy of ts3.

A direct comparison between MP2 and QM/MM BLYP energies of structures taken from the QM/MM simulation of HIV-1 PR is not possible at the present stage because the MP2 calculations have not been implemented in our QM/MM code. **EINT(b)** single-point energies relative to those of **ESUB(b)** have been calculated with the BP and PBE exchange-correlation functionals for the large and small QM systems; these are 5 kcal mol⁻¹ lower than BLYP energies, in line with model 1 calculations. These results again indicate that most of the discrepancy between the calculated and expected ts2 should be ascribed to the use of the BLYP exchange-correlation functional

Acknowledgment. We thank the INFM for financial support and Dexter Northrop for useful discussions.

References and Notes

- Wlodawer, A.; Vondrasek, J. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 249–284.
- Fitzgerald, P. M. D.; Springer, J. P. *Annu. Rev. Biophys. Biomol. Struct.* **1991**, *20*, 299–320.
- Weber, I. T.; Harrison, R. W. *Protein Eng.* **1996**, *9*, 679–690.
- Harrison, R. W.; Weber, I. T. *Protein Eng.* **1994**, *7*, 1353–1363.
- Okimoto, N.; Tsukui, T.; Kitayama, K.; Hata, M.; Hoshimo, T.; Tsuda, M. *J. Am. Chem. Soc.* **2000**, *122*, 5613–5622.
- Trylska, J.; Bala, P.; Geller, M.; Grochowski, P. *Biophys. J.* **2002**, *83*, 794–807.
- Liu, H.; Müller-Plathe, F.; Van Gasteren, W. F. *J. Mol. Biol.* **1996**, *261*, 454–469.
- Silva, A. M.; Cachau, R. E.; Sham, H. L.; Erickson, J. W. *J. Mol. Biol.* **1996**, *255*, 321–346.
- Lee, H.; Darden, T. A.; Pedersen, L. G. *J. Am. Chem. Soc.* **1996**, *118*, 3946–3950.
- Goldblum, A. *Biochemistry* **1988**, *27*, 1653–1658.
- Beveridge, A. J.; Heywood, G. C. *Biochemistry* **1993**, *32*, 3325–3333.
- Park, H.; Suh, J.; Lee, S. *J. Am. Chem. Soc.* **2000**, *122*, 3901–3908.
- Venturini, A.; López-Ortiz, F.; Alvarez, J. M.; Gonzalez, J. *J. Am. Chem. Soc.* **1998**, *120*, 1110–1111.
- Okimoto, N.; Tsukui, T.; Hata, M.; Hoshino, T.; Tsuda, M. *J. Am. Chem. Soc.* **1999**, *121*, 7349–7354.
- Piana, S.; Parrinello, M.; Carloni, P. *J. Mol. Biol.* **2002**, *319*, 567–583.
- Gulnik, S.; Erickson, J. W.; Xie, D. *Vitam. Horm.* **2000**, *58*, 213–256.
- Hyland, L. J.; Tomaszek, T. A.; Meek, T. D. *Biochemistry* **1991**, *30*, 8454–8463.
- Hyland, L. J.; Tomaszek, T. A.; Roberts, G. D.; Carr, S. A.; Maagard, V. W.; Bryan, H. L.; Fakhoury, S. A.; Moore, M. L.; Minnich, M. D.; Culp, J. S.; DesJarlais, R. L.; Meek, T. D. *Biochemistry* **1991**, *30*, 8441–8453.
- Hong, L.; Hartsuck, J. A.; Foundling, S.; Ermoliev, J.; Tang, J. *Protein Sci.* **1998**, *7*, 300–305.
- Rose, R.; Craik, C. S.; Douglas, N. L.; Stroud, R. M. *Biochemistry* **1996**, *35*, 12933–12944.
- Suguna, K.; Padlan, E. A.; Smith, C. W.; Carlson, W. D.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7009–7013.
- Xie, D.; Gulnik, S.; Collins, L.; Gustchina, E.; Bhat, T. N.; Erickson, J. W. *Adv. Exp. Med. Biol.* **1998**, *436*, 381–386.
- Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. A. *J. Mol. Biol.* **2000**, *301*, 1207–1220.
- Baca, M.; Kent, S. B. H. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11638–11642.
- Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L. M.; Kent, S. B. H.; Wlodawer, A. *Science* **1989**, *246*, 1149–1152.
- Polgár, L.; Szeltner, Z.; Boros, I. *Biochemistry* **1994**, *33*, 9351–9357.
- Szeltner, Z.; Polgár, L. *J. Biol. Chem.* **1996**, *271*, 32180–32184.
- Rodriguez, E. J.; Angeles, T. S.; Meek, T. D. *Biochemistry* **1993**, *32*, 12380–12385.
- Northrop, D. B. *Acc. Chem. Res.* **2001**, *34*, 790–797.
- Piana, S.; Carloni, P. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 26–36.
- Brancolini, C.; Lazarevic, D.; Rodriguez, J.; Schneider, C. *J. Cell Biol.* **1997**, *139*, 759–771.
- Piana, S.; Carloni, P.; Rothlisberger, U. *Protein Sci.* **2002**, *11*, 2393–2402.
- Chatfield, D. C.; Brooks, B. R. *J. Am. Chem. Soc.* **1995**, *117*, 5561–5572.
- Cho, Y. K.; Northrop, D. B. *J. Biol. Chem.* **1998**, *273*, 24305–24308.
- Cho, Y. K.; Rebholz, K. L.; Northrop, D. B. *Biochemistry* **1994**, *33*, 9637–9642.
- Meek, T. D.; Rodriguez, E. J.; Angeles, T. S. *Methods Enzymol.* **1994**, *241*, 127–156.
- Laio, A.; Van de Vondel, J.; Rothlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941–6947.
- Becke, A. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- Piana, S.; Sebastiani, D.; Carloni, P.; Parrinello, M. *J. Am. Chem. Soc.* **2001**, *123*, 8730–8737.
- Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chem. Phys. Lett.* **1989**, *156*, 472–477.
- VandeVondele, J.; Rothlisberger, U. *J. Phys. Chem. B* **2002**, *103*, 206–208.
- Cascella, M.; Raugei, S.; Carloni, P. *J. Phys. Chem. B* **2004**, *108*, 369–375.
- Silvestrelli, P. L.; Parrinello, M. *Phys. Rev. Lett.* **1999**, *82*, 3308–3311.
- Becke, A. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822–8824.
- Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- Schock, H. B.; Garsky, V. M.; Kuo, L. C. *J. Biol. Chem.* **1996**, *271*, 31957–31963.
- Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G. L.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- Berger, A.; Schechter, I. *Philos. Trans. R. Soc. London, Ser. B* **1970**, *257*, 249–264.
- Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1943–2006.
- Barnett, R. N.; Landman, U. *Phys. Rev. B* **1993**, *48*, 2081–2097.
- Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.
- Hutter, J.; Ballone, P.; Bernasconi, M.; Focher, P.; Foiss, E.; Goedecker, S.; Parrinello, M.; Tuckerman, M. *CPMD 3.3. MPI für Festkörperforschung und IBM Zurich Research Laboratory: Zurich*, 1999.
- Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- Van Gasteren, W. F.; Billeter, S. R.; Eising, A. A.; Hunenberger, P. H.; Kruger, P. K. H. C.; Mark, A. E.; Scott, W. R.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; Hochschulverlag AG: Zürich, 1996.
- Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- Hunenberger, P. H. *J. Chem. Phys.* **2000**, *113*, 10464–10476.
- Curioni, A.; Sprik, M.; Andreoni, W.; Schiffer, H.; Hutter, J.; Parrinello, M. *J. Am. Chem. Soc.* **1997**, *119*, 7218–7229.
- Sprik, M.; Ciccotti, G. *J. Chem. Phys.* **1998**, *109*, 7737–7744.
- Meijer, E. J.; Sprik, M. *J. Am. Chem. Soc.* **1998**, *120*, 6345–6355.
- Ciccotti, G.; Ferrario, M.; Hynes, J. T.; Kapral, R. *Chem. Phys.* **1989**, *129*, 241–251.

(63) Berti, P. J. Determining Transition States from Kinetic Isotope Effects. In *Enzyme Kinetics and Mechanism*; Schramm, V. L., Purich, D. L., Eds.; Academic Press: San Diego, CA, 1999; pp 355–397.

(64) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.;

Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.11.4; Gaussian, Inc.: Pittsburgh, PA, 1998.