

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/11321662>

# Role of Conformational Fluctuations in the Enzymatic Reaction of HIV-1 Protease

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · JUNE 2002

Impact Factor: 4.33 · DOI: 10.1016/S0022-2836(02)00301-7 · Source: PubMed

---

CITATIONS

104

---

READS

16

3 AUTHORS, INCLUDING:



Stefano Piana

77 PUBLICATIONS 4,606 CITATIONS

SEE PROFILE



Paolo Carloni

Forschungszentrum Jülich

320 PUBLICATIONS 6,082 CITATIONS

SEE PROFILE

# Role of Conformational Fluctuations in the Enzymatic Reaction of HIV-1 Protease

Stefano Piana<sup>1</sup>, Paolo Carloni<sup>1\*</sup> and Michele Parrinello<sup>2,3</sup>

<sup>1</sup>*Scuola Internazionale  
Superiore di Studi Avanzati  
and Istituto Nazionale di Fisica  
per la Materia, Via Beirut 2-4  
34014 Trieste, Italy*

<sup>2</sup>*CSCS—Centro Svizzero di  
Calcolo Scientifico, Galleria 2  
Via Cantonale, CH-6928  
Manno, Switzerland*

<sup>3</sup>*Physical Chemistry, ETH  
Zurich, Hönggerberg HCI  
CH-8093, Zurich, Switzerland*

The emergence of compensatory drug-resistant mutations in HIV-1 protease challenges the common view of the reaction mechanism of this enzyme. Here, we address this issue by performing classical and *ab initio* molecular dynamics simulations (MD) on a complex between the enzyme and a peptide substrate. The classical MD calculation reveals large-scale protein motions involving the flaps and the cantilever. These motions modulate the conformational properties of the substrate at the cleavage site. The *ab initio* calculations show in turn that substrate motion modulates the activation free energy barrier of the enzymatic reaction dramatically. Thus, the catalytic power of the enzyme does not arise from the presence of a pre-organized active site but from the protein mechanical fluctuations. The implications of this finding for the emergence of drug-resistance are discussed.

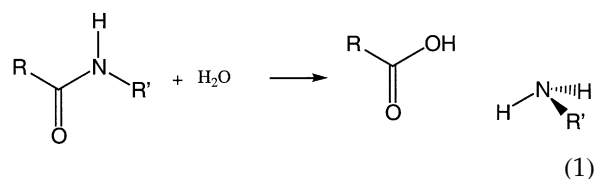
© 2002 Elsevier Science Ltd. All rights reserved

\*Corresponding author

**Keywords:** density functional; HIV-1 protease; molecular dynamics; protein flexibility; reaction mechanism

## Introduction

The human immunodeficiency virus type-1 aspartyl protease (HIV-1 PR) is one of the major targets for anti-AIDS therapy.<sup>1–3</sup> The enzyme recognizes and cleaves specific segments of six to eight amino acid residues of a polyprotein:<sup>4–7</sup>



The resulting peptide segments subsequently fold and form new proteins, which in turn are salvaged into the metabolism of the virus.<sup>5,8,9</sup>

As is usually the case, the details of the enzymatic cycle are hypothesized but not established firmly (Figure 1). It has been suggested that the enzymatic activity may result from the presence at

the active site of a properly engineered electric field, which acts as a supersolvent.<sup>10,11</sup>

In HIV-1 PR, this view appears to be challenged by the emergence of “compensatory” drug-resistant mutations, which may be located 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 function.<sup>12–14</sup> In many cases, these mutations involve chemically similar residues (such as M46I, L47V, L63P, A71T),<sup>15–18</sup> which cannot change the electrostatic field significantly.

Attempts at explaining the catalytic activity of HIV-1 PR have been based on a static picture of the protein derived from the X-ray structures,<sup>19–22</sup> and have been unable to account for the mutagenic behavior described above.

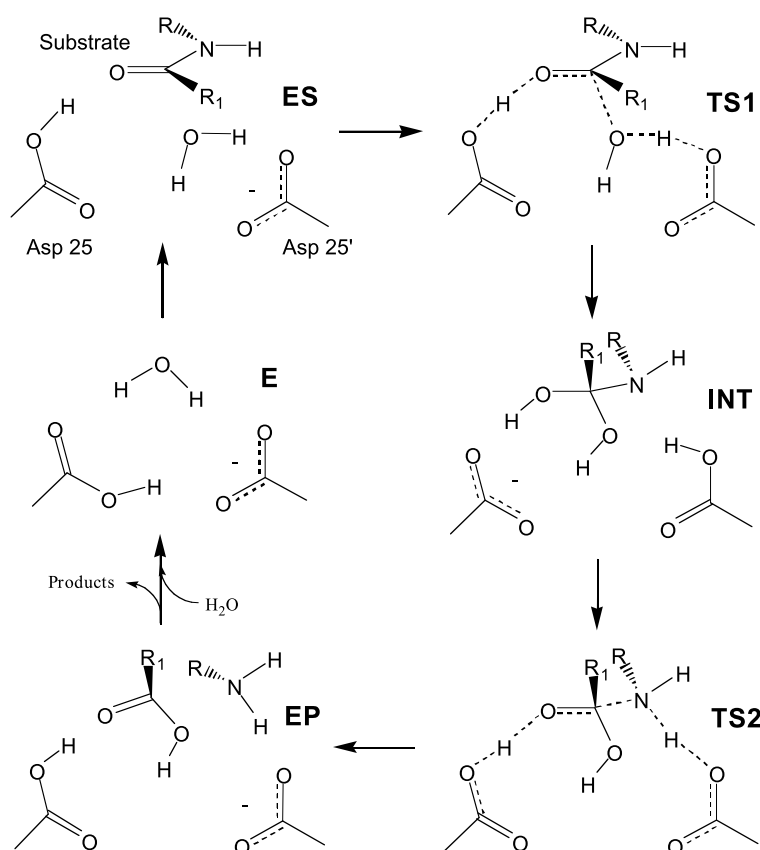
Here, we depart from this conventional approach and show that the large-scale motions of the protein that modulate the geometry of the active site are responsible for the catalytic activity and can offer a clue to understanding the drug-resistant behavior of the protein mutations.

In order to validate this picture, we use a two-step strategy. In the first part of the work, we perform a 2.1 ns simulation of the solvated protein complexed with the *N*-acetyl Thr-Ile-Met-Met-Gln-Arg amide peptide (SUB hereafter) (Figure 2(a)) using a classical molecular dynamics (MD) approach. This simulation reveals the large-scale

Present address: S. Piana, Inorganic Chemistry, ETH Zurich, Hönggerberg HCI, CH-8093, Zurich, Switzerland.

Abbreviations used: DFT, density functional theory; HIV-1, human immunodeficiency virus type 1; MD, molecular dynamics; PR, aspartyl protease; SUB, *N*-acetyl Thr-Ile-Met-Met-Gln-Arg amide peptide.

E-mail address of the corresponding author: carloni@sissa.it



**Figure 1.** Proposed reaction mechanism of HIV-1 PR. In reaction Step 1, substrate **S** binds to the enzyme **E** and forms the enzyme-substrate complex **ES**. A water molecule (WatC), which is polarized by the Asp dyad, attacks the **S** carbonyl carbon atom. In the transition state **TS1**, the Asp dyad transfers a proton to **S** and accepts a proton from Wat. The gem-diol reaction intermediate **INT** is formed subsequently. Reaction Step 2 involves the breaking of the **INT** C–N bond to form the final products (amide and carboxyl acid). Various pathways have been proposed for this step. In the most accredited proposal, an Asp group donates a proton to the **INT** amide group and simultaneously<sup>57,60,83</sup> or subsequently<sup>20</sup> the other accepts a proton from one of the **INT** hydroxyl groups (**TS2**). Alternatively, the amide group could accept the proton directly from one of the **INT** hydroxyl groups.<sup>96</sup> **TS2** decomposes (maybe passing to a second intermediate<sup>20,60</sup>) to give the reaction products (**EP**). These are released to yield the free enzyme (**E**). Variations on the above scheme have been postulated, according to

which **INT** could be a deprotonated carbanion<sup>37</sup> and Step 1 and Step 2 two concerted events.<sup>97</sup>

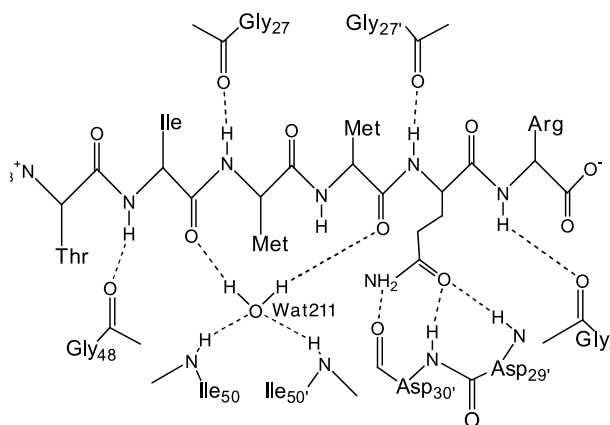
motions of several regions of the protein (such as the flaps) that are crucial for the substrate selectivity.<sup>14,23</sup> The calculation finds that the electric field generated on the active site by the protein residues other than the Asp dyad is small, and independent of time and space. Furthermore, it shows that the substrate motion towards the cleavage site is coupled with the lowest frequency motion of the protein. Since this motion occurs in a long time-scale, it is meaningful to investigate the chemical reactivity of the enzyme for configurations of the protein characterized by different substrate/cleavage site distances ( $d_{SA}$ ).

In the second part, we simulate the first step of the catalytic reaction (from **ES** to **INT**, Figure 1). The quantum-chemical approach adopted here is the *ab initio* Car–Parrinello MD simulation method (*ab initio* MD),<sup>24–26</sup> which has been shown to provide a reliable description of a variety of chemical<sup>27–29</sup> and enzymatic<sup>30–33</sup> reactions.

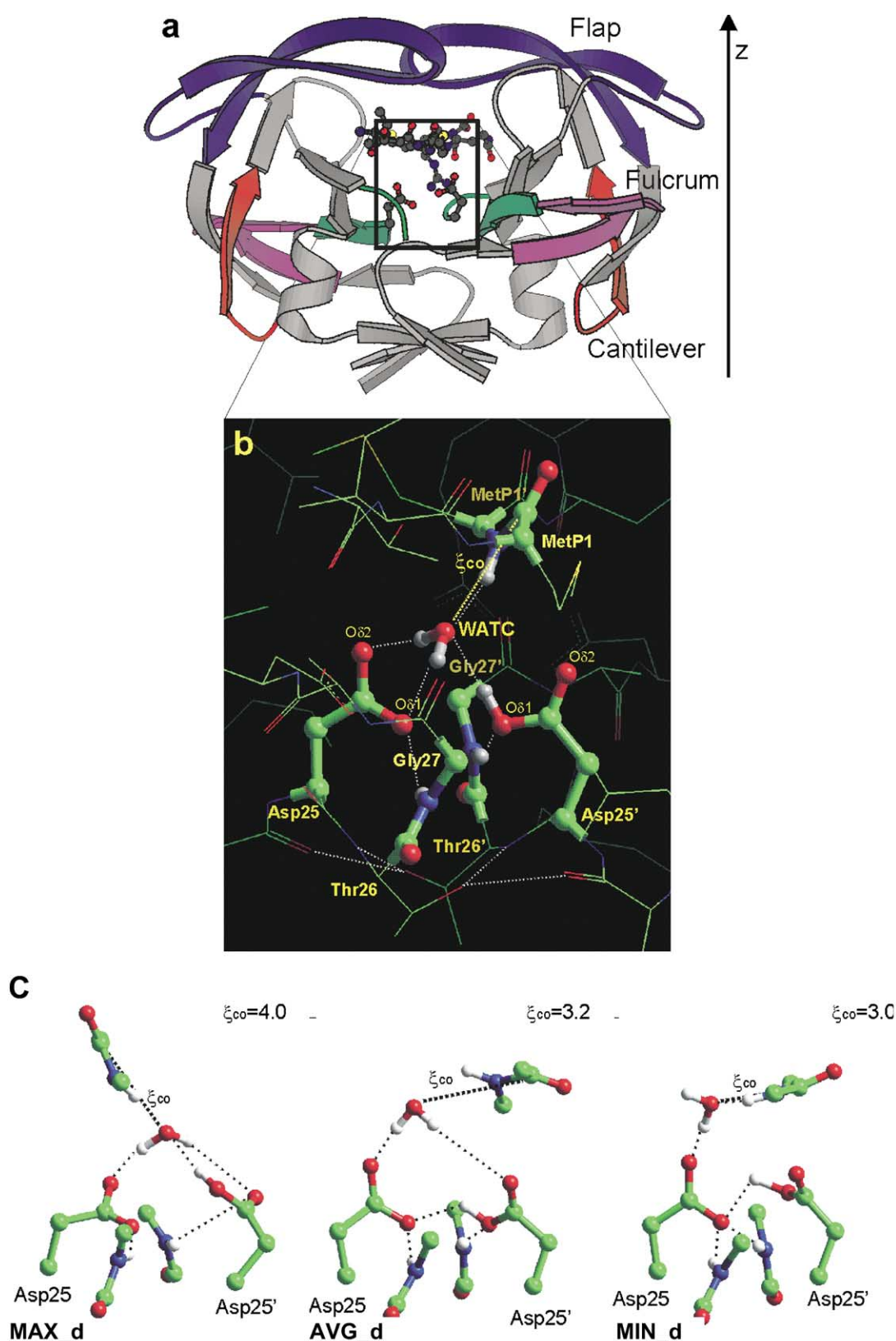
The calculated activation free energy barrier of the reaction decreases strongly with the substrate/cleavage site distance. The protein then acts as an elaborate nutcracker that brings the substrate into a favorable geometrical position for the appropriate fluctuations of the protein frame. This offers a clue to the ease with which drug-resistant mutations far from the active site can emerge.

## Results

Our study is carried out as follows. First, we estimate the electric field at the active site generated by the solvated protein. Then we look at the long-scale protein conformational fluctuations with classical MD simulations. Finally, we investigate the chemistry of the cleavage site by *ab initio* MD.



**Chart 1.**



**Figure 2.** Structural model of the HIV-1 PR/SUB complex. This model was built from the HIV-1 PR/*N*-acetyl-Thr-Ile-Nle- $\Psi$ [CH<sub>2</sub>NH]-Nle-Gln-Arg-amide inhibitor complex<sup>19</sup> (see Methods). (a) The structure of the complex; the protein is a homodimer. The two subunits are related by C<sub>2</sub> symmetry. The z-axis indicated in the Figure is parallel with the C<sub>2</sub> axis of the enzyme. The residue sequence is numbered from 1 (the N terminus) to 99 (the C terminus) in subunit 1 and from 1' to 99' in subunit 2; the z-axis is indicated. The fulcrum (residues 11–21 (11'–21')), fireman's grip (residues 22–28 (22'–28')), flaps (residues 34–59 (34'–59')) and cantilever (residues 64–74 (64'–74'))<sup>40</sup> are colored in purple, green, blue and red, respectively. The flaps and fireman's grip regions are directly in contact with the substrate. (b) The cleavage site region and (c) the model complexes (**MAX\_d**, **AVG\_d** and **MIN\_d**) used for the *ab initio* calculations. The residues in bold were included in the *ab initio* calculations. For the sake of clarity, non-polar hydrogen atoms are not shown.

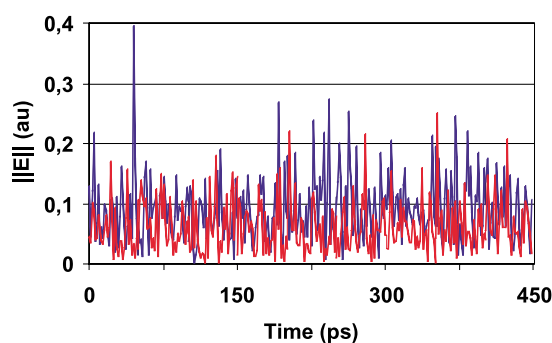
### Polarization of the reactants by the protein

The catalytic reaction is believed to occur *via* an addition of a water molecule (WatC) to the carbonyl group of the substrate (Figure 1).

In this section, we use classical MD and *ab initio* methods to analyze the electrostatics properties of the reactants. These are accommodated in the active site of the protein, which is a large cavity delimited by two large flaps (Figure 2(a)).<sup>34</sup> The substrate is stabilized by a series of H-bond and hydrophobic interactions with several domains of the protein, such as the flaps, the 75–85 (75'–85') loop and the fireman's grip.<sup>5,19,35–37</sup> The cleavage site is a portion of the fireman's grip (Figure 2(b)). It consists of the highly conserved D(25,25')–T(26,26')–G(27,27') catalytic triads located at the subunit–subunit interface.<sup>5</sup> The aspartyl residues, which play a direct role in the enzymatic catalysis, are very close to each other and are almost coplanar (Figure 2(b)).

The electric field  $E$  of the solvated protein on WatC, averaged over the classical MD simulations, is similar in modulus and direction to that on a water molecule in bulk water (Table 1). Furthermore the protein field on the reactants appears to be similar to that in water solution. Indeed,  $E$  calculated on the atoms forming the peptide bond of the substrate is similar to that on a peptide bond on the protein surface (Table 1). Figure 3 shows the modulus of the field on the water oxygen atom as an example.

Thus, within the limitations of our simple model, the reactants in the active site experience an electric field similar to that experienced by the same groups in water solution. A consistent picture emerges from the electronic structure calculation on a model complex of the cleavage site (AVG\_d, Figure 2(c)): the dipole moment of WatC, averaged over several different conformations of the complex, is the same as that reported for water in



**Figure 3.** Electric field in the active site and in water solution. Fluctuations of the total electric field  $\|E\|$  (atomic units). Protein, solvent and catalytic triad included during the dynamics (only the first 0.25 ns is shown).  $\|E\|$  on the oxygen atom belonging to WatC (blue) and on a water molecule in the liquid phase (red).

the liquid phase, calculated with the same computational setup (3.0 (0.2) D and 3.0 D<sup>38</sup>).

Our molecular mechanics model indicates that the polarization on the cleavage site is created essentially by the cleavage site alone (i.e. the catalytic triads): indeed, the protein frame contributes less than 20% to the total field. Consistently, the average water dipole moment in the presence of the field of the catalytic triads only is the same as in the presence of the entire protein field, that is 3.0(0.2) D.

We conclude that the polarization on the reactants is created essentially by the cleavage site alone (i.e. the Asp dyad) and is similar to the polarization that the reactants would experience in bulk water.

### Protein conformational fluctuations

In this section, we investigate the long-scale motion of the substrate/protein complex based on a 2.1 ns classical MD simulation. The structure is well equilibrated after  $\approx 0.7$  ns, as shown by a plot of the rmsd of the backbone atoms as a function of time (Figure 4(a)) (rmsd relative to the X-ray structure  $\approx 1.5$  Å). The overall fold is well maintained (Figure 4(b)). However, the mobility of the protein domains is markedly different. The fireman's grip is rather rigid, as shown by its low rmsd per residue (from 0.5 to 0.7 Å). This relative stiffness may result from the extensive H-bond network formed among the residues in this region, which is highly conserved in all aspartic proteases.<sup>39–41</sup> In contrast, the fulcrum, the cantilever and, even more, the flaps are rather flexible (rmsd from 0.9 to 2.5 Å). Similar findings have been reported by previous MD calculations on the sub-nanosecond<sup>40–45</sup> and multiananosecond<sup>46</sup> time-scales, and are in agreement with the thermal factor calculated from X-ray data.<sup>35</sup>

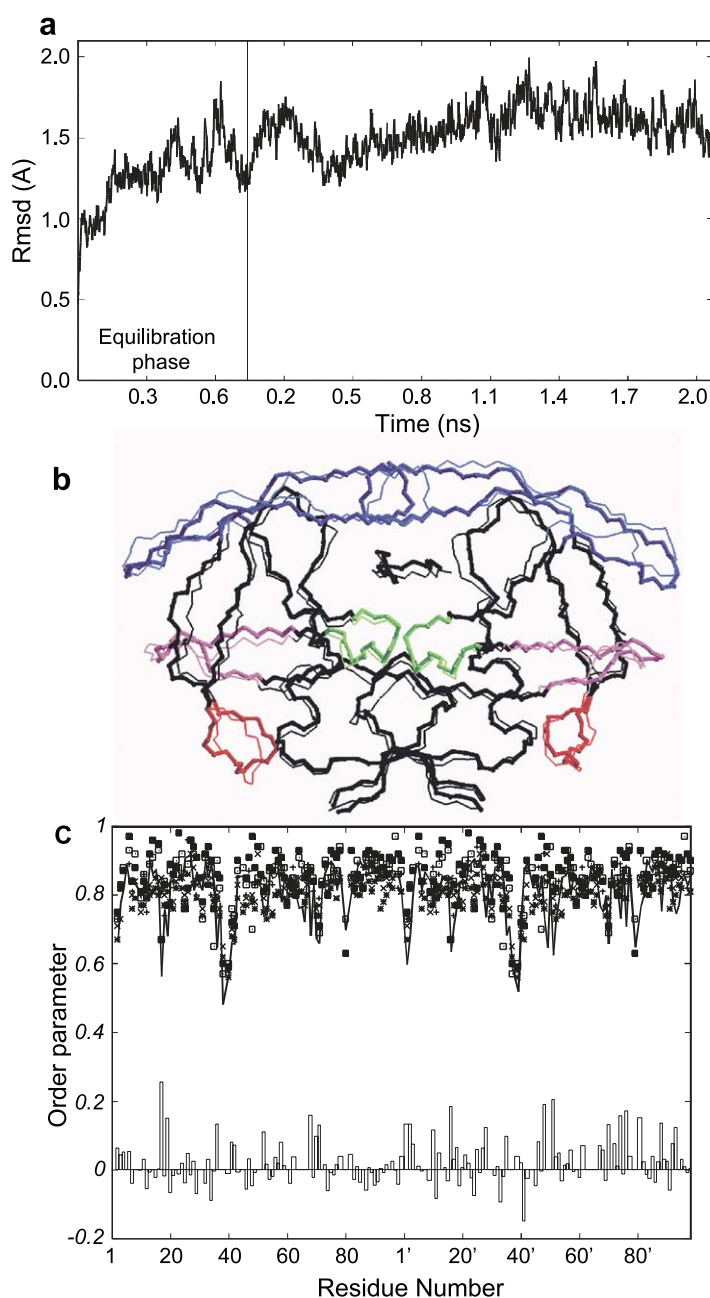
The calculated NMR order parameters  $S^2$ <sup>47,48</sup> are in excellent agreement with the NMR data available for HIV-1 PR complexed with KNI-272<sup>49</sup>,

**Table 1.** Cleavage site electrostatic field

Atom	HIV-1 PR cleavage site		Water solution	
	$\ E\ $ (a.u.)	Direction (deg.)	$\ E\ $ (a.u.)	Direction (deg.)
O <sub>(Wat)</sub>	0.0043(0.0027)	123(24)	0.0039(0.0025)	125(26)
H1 <sub>(Wat)</sub>	0.0053(0.0036)	146(22)	0.0047(0.0031)	148(20)
H2 <sub>(Wat)</sub>	0.0045(0.0038)	138(31)	0.0045(0.0031)	146(19)
N <sub>(PEP)</sub>	0.0029(0.0022)	119(30)	0.0029(0.0024)	110(34)
H <sub>(PEP)</sub>	0.0030(0.0021)	129(29)	0.0031(0.0024)	130(29)
C <sub>(PEP)</sub>	0.0028(0.0021)	122(29)	0.0030(0.0022)	123(28)
O <sub>(PEP)</sub>	0.0026(0.0019)	125(29)	0.0024(0.0021)	120(33)

Modulus ( $\|E\|$ , a.u.) and direction (deg.) of the electrostatic field calculated on the P1–P1' peptide bond and on the catalytic water molecule (WatC) atoms. The field calculated for a bulk water molecule and for the flexible, solvent-exposed Gly16–Gly17 peptide bond is reported for comparison. The direction of the field for the water atoms is computed with respect to the water  $C_2$  axis. That for the peptide bond is calculated with respect to the C=O bond and the N–H bond for the peptide bond C,O and N,H atoms, respectively.



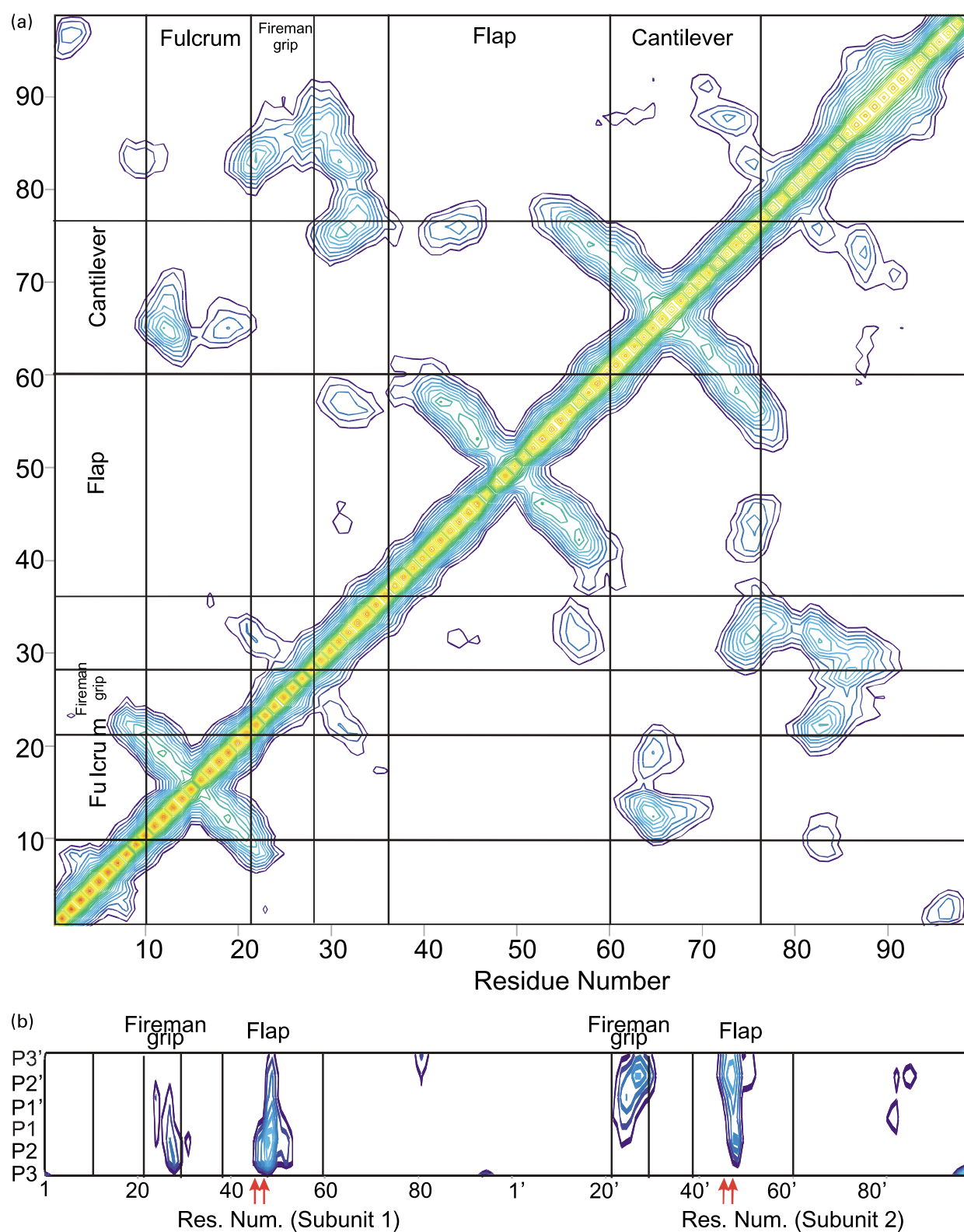


**Figure 4.** Classical MD simulation. (a) The rmsd of non-hydrogen atoms plotted as a function of simulated time. (b) X-ray (thin lines) and MD-averaged backbone structures (thick lines). (c) NMR order parameters as a function of residue number. Calculated order parameters are represented as a continuous line. Those experimentally measured for HIV-1 PR complexed with KNI272,<sup>49</sup> P9941,<sup>50</sup> DMP323 (at 300 MHz),<sup>50</sup> DMP323 (2) (at 600 MHz)<sup>50,51</sup> inhibitors are represented as +,  $\alpha$ ,  $\times$ , \*,  $\bullet$ , respectively. The difference between HIV-1 PR/SUB calculated parameters and PR/KNI272 experimentally measured parameters are represented as bars.

DMP323<sup>50,51</sup> and P9941<sup>50</sup> (Figure 4(c)): small  $S^2$  values are observed in the highly mobile region, which includes residues 38–42. Relatively small  $S^2$  values are observed for residues 17–19, 69–71, 80 and for residues 2 and 3, while the largest values are observed for the residues of the active site (23–27) and for the terminal sequence (87–99). The average difference between the calculated order parameters and the  $S^2$  values determined experimentally ranges from 0.050 to 0.066. These values can be compared with the average difference between different experimental measurements in HIV-1 PR/inhibitor complexes (0.03–0.07),<sup>49–51</sup> with the average experimental error for  $S^2$  measurements (0.02–0.03)<sup>49,50</sup> and with previous  $S^2$  calculations on HIV-1 PR/KNI272 complex (0.042).<sup>41</sup> It is worth noting that the best agreement

with the experimental data is observed for HIV-1 PR complexed to KNI-272,<sup>49</sup> which is the inhibitor resembling a protease substrate most closely.

Correlations of the motion among these regions can be obtained by examining the dynamic cross-correlation map (DCCM) of the  $C^\alpha$  atoms. Figure 5(a) shows that the motion of the flaps is correlated with that of the cantilever, which in turn is correlated with that of the fulcrum. These findings agree with an MD simulation on the free enzyme on the sub-nanosecond time-scale.<sup>40</sup> However, we should note that some regions of the protein are characterized by motions on the micro- to milli-second time-scale.<sup>49–51</sup> These motions are too slow to be observed in our simulation, therefore we cannot rule out the possibility that novel correlated motions could appear in simulations several orders



**Figure 5.** Dynamic cross-correlation matrices. These matrices, calculated for the  $C^\alpha$  atoms, provide information on large-scale correlated motions of the protein.<sup>40</sup> The matrix element  $C_{ij}$  of atoms  $i$  and  $j$  reads:<sup>40</sup>

$$C_{ij} = \frac{\langle \Delta \vec{r}_i(t) \Delta \vec{r}_j(t) \rangle}{(\langle \Delta \vec{r}_i(t)^2 \rangle \langle \Delta \vec{r}_j(t)^2 \rangle)^{1/2}}$$

where  $\langle \rangle$  denotes an MD-averaged quantity and  $\Delta \vec{r}_i(t) = \langle \vec{r}_i(t) \rangle - \vec{r}_i(t)$  the displacement from the average MD position  $\langle \vec{r}_i(t) \rangle$  of atom  $i$  during a generic MD step.  $C_{ij}$  varies from  $-1.0$  for completely anticorrelated motions to  $+1.0$  for completely correlated motions. A value close to  $+1$  reflects a high correlation between the motions of a pair of  $C^\alpha$  atoms. The largest values are obviously found for  $C^\alpha$  atoms belonging to residues  $i$  and  $i \pm a$  with  $a = 0, 1, 2$ . These form a

**Table 2.** HIV-1 PR/peptide interactions

Distance	MD averaged structure (Å)	Crystal structure (Å)
C <sub>P1</sub> –O <sub>Wat</sub>	3.97(0.25)	–
O <sub>Gly27</sub> –N <sub>MetP1</sub>	4.74(0.60)	2.81
O <sub>Gly27</sub> –N <sub>GlnP2'</sub>	2.47(0.45)	2.89
O <sub>Thr26</sub> –N <sub>Thr26'</sub>	3.23(0.23)	3.26
O <sub>Thr26</sub> –O <sub>Leu24'</sub>	2.76(0.14)	2.83
O <sub>Thr26</sub> –N <sub>Thr26</sub>	3.13(0.17)	4.84
O <sub>Thr26</sub> –O <sub>Leu24</sub>	2.83(0.16)	5.17
O <sub>WatA</sub> –N <sub>Ile50</sub>	3.39(0.38)	3.30
O <sub>WatA</sub> –N <sub>Ile50'</sub>	3.16(0.26)	2.67
O <sub>WatA</sub> –O <sub>IleP2</sub>	2.85(0.23)	3.04
O <sub>WatA</sub> –O <sub>MetP1'</sub>	2.79(0.15)	2.48
O <sub>Gly48</sub> –N <sub>IleP2</sub>	2.89(0.19)	2.59
O <sub>Gly48</sub> –N <sub>ArgP3'</sub>	2.98(0.18)	3.00
O <sub>Asp29</sub> –N <sub>ThrP3</sub>	4.82(1.42)	3.09
N <sub>GlnP2'</sub> –O <sub>Asp30'</sub>	2.94(0.16)	5.82
O <sub>GlnP2'</sub> –N <sub>Asp29'</sub>	3.27(0.36)	4.10
O <sub>GlnP2'</sub> –N <sub>Asp3</sub>	3.08(0.20)	3.50

Selected distances between the substrate and the protein for the averaged MD structure of the HIV-1/SUB adduct as obtained with the classical MD simulation. Comparison is made with the corresponding distances of HIV-1 PR/N-acetyl-Thr-Ile-Nle-Ψ[CH<sub>2</sub>NH]-Nle-Gln-Arg-amide inhibitor,<sup>19</sup> which has been used to model the HIV/SUB complex. Standard deviations are given in parentheses.

of magnitude longer than those presented here. Furthermore Figure 5(b) shows that the substrate motion is correlated with motion of residues 24(24')–30(30') (the cleavage site) and 45(45')–55(55') (the flaps).

Substrate/enzyme H-bond and hydrophobic contacts are well maintained during the classical MD simulation (Table 2). However, the fluctuations are markedly different also in the case of the substrate: the largest number of H-bonds is observed at the P2 and P2' sites (Chart 1); accordingly these are the substrate residues characterized by the lowest flexibility (residue rmsd 1.6 and 1.7 Å, respectively). On the other hand, the solvent-exposed terminal residues P3 and P3' (residue rmsd 1.9 and 2.4 Å, respectively) and the cleavage site residues P1 and P1' (residue rmsd 1.8 and 1.9 Å, respectively) are more flexible. To characterize the substrate motions with respect to the active site a distance ( $d_{SA}$ ) was defined as the distance between the C<sup>α</sup> of the Asp dyad and those of the

P1 and P1' methionine residues Figure 6(a)). As the cleavage site is rather rigid (Figure 4(b) and (c)), the fluctuations of  $d_{SA}$  are basically determined solely by the substrate motions.

Figure 6(a), which plots the distance  $d_{SA}$  between the aspartyl pair, the substrate P1 and P1' C<sup>α</sup> atoms, confirms that the substrate experiences rather large fluctuations at the cleavage site:  $d_{SA}$  turns out to oscillate by more than 1 Å around its average value (7.7 Å) in the time-scale investigated. The motion is characterized by two different regimes: one during the first part of the production phase of the simulation ( $\approx 0$ –0.7 ns) and one in a later phase ( $\approx 0.7$ –2.1 ns). The transition between the two regimes is characterized by the breaking of the HN(Asp29)···O(ThrP3) H-bond. However, this hydrogen bond is reformed after 1.7 ns of MD simulation.

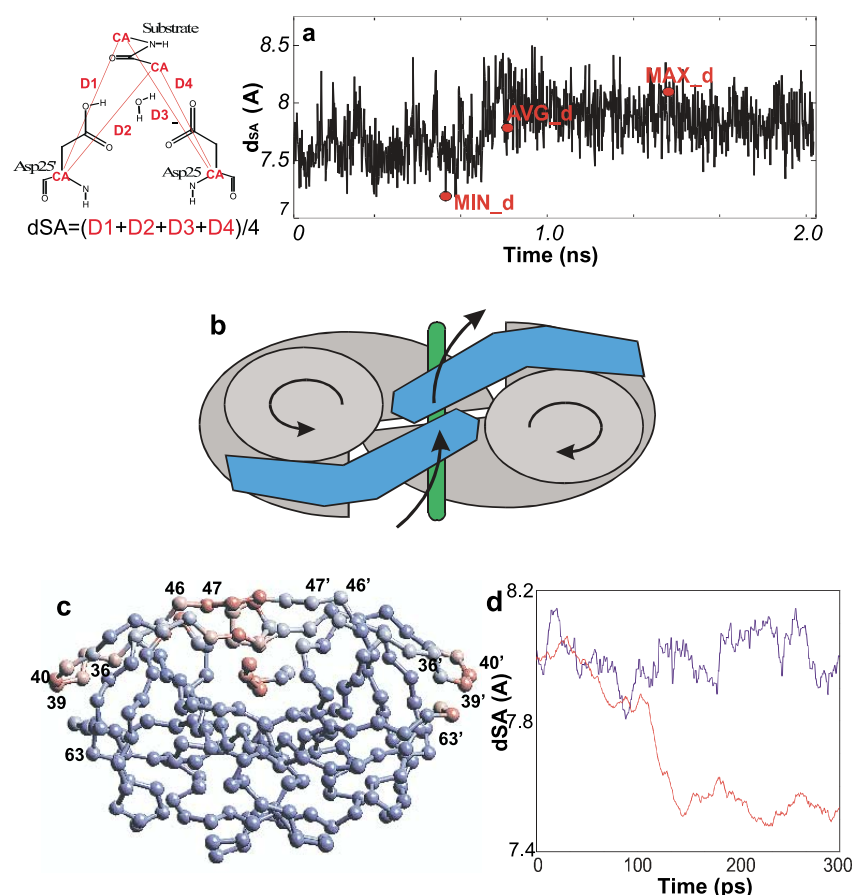
The largest displacements of the biomolecular complex can be calculated from a diagonalization of the covariance matrix. The modes associated with the largest eigenvalues of the matrix describe a long-time, significant global protein motion.<sup>52,53</sup> Here, the calculated largest eigenvector, as large as 43 Å<sup>2</sup>, is clearly dominant: the second largest eigenvector is only 25% of the first one and the others do not contribute significantly. This mode involves a rotation of solvent-exposed regions in the two subunits: the largest components of the eigenvector are observed for the flaps, cantilever and fulcrum; large displacements of the substrate along the protein C<sub>2</sub> symmetry axis are found (Figure 6(b)) (see also the animation available on the web†).

Based on the calculation of the covariance matrix eigenvectors, it is possible to identify the regions of the protein that are mostly relevant for substrate motion. This can be achieved by calculating the substrate displacement upon application of a force on any C<sup>α</sup> atom (see Methods). Figure 6(c) shows that, as expected, some of these regions include the substrate itself and the flap tips, directly interacting with it. However, it is found that the flap elbows, relatively far from the substrate, do play a role for substrate motion. It is interesting to notice that several compensatory mutations (M36I, M46I-L, L47V) are located in these regions (Figure 6).

diagonal in the map. High correlations are found for residues  $i$  and  $i \pm 4$  belonging to  $\alpha$ -helices (visible as an enlargement of the diagonal) and for antiparallel  $\beta$ -sheets, which are characterized by patterns orthogonal to the diagonal that originate from the coupling of residues whose numbering flows in opposite directions. Off-diagonal terms represent correlated motions between non-bonded residues not involved in secondary structure elements. The maps have been calculated for the C<sup>α</sup> atoms from the 2.1 ns classical MD simulation. Very similar patterns are obtained if the map is calculated on 1 ns of trajectory only. (a) A map of HIV-1 PR subunit 1. The map of the other subunit is practically identical with that of subunit 1, and is not reported. (b) Map of SUB (y-axis from P3 to P3')/HIV-1 PR subunit 1 (residues numbered from 1 to 99) and of SUB/HIV-1 PR subunit 2 (residues numbered from 1'–99'). The coupling between the motion of Met46 and Leu47 and that of the substrate is indicated with red arrows.

† [http://www.sissa.it/sbp/bc/publications/publications\\_5.html](http://www.sissa.it/sbp/bc/publications/publications_5.html)





**Figure 6.** Substrate motion. (a) Left: Definition of the SUB/Asp dyad distance  $d_{SA}$ . Right:  $d_{SA}$  plotted as a function of simulated time. For the sake of clarity, the first 0.7 ns (equilibration phase) is not shown. (b) A representation of the large-scale motion associated with the largest eigenvector. This motion involves a complex rotation of the two subunits (gray and blue in the Figure), which leads to a significant rearrangement of the flaps (blue) and substrate (green). The view is orthogonal to the protein  $C_2$  axis. (c) Substrate displacements upon application of a unitary force on each  $C^\alpha$  atom. Left: the  $C^\alpha$  atoms associated with an increasingly large displacement are colored from blue to red. The positions of the residues undergoing compensatory mutations are indicated (36, 46, 47, 63, 36', 46', 47', 63'). Right:  $d_{SA}$  plotted as function of time in the unconstrained (blue) MD simulations and in an MD simulation in which a force is exerted on the 39(39') and 40(40')  $C^\alpha$  atoms (red).

A consistent picture is obtained by an MD simulation in which constraints along the elbow are applied. Indeed, after few tens of nanoseconds the substrate is displaced towards the cleavage site as a consequence of the constraint (Figure 6(d)). Such a displacement is not present in the unconstrained simulation.

### The enzymatic reaction

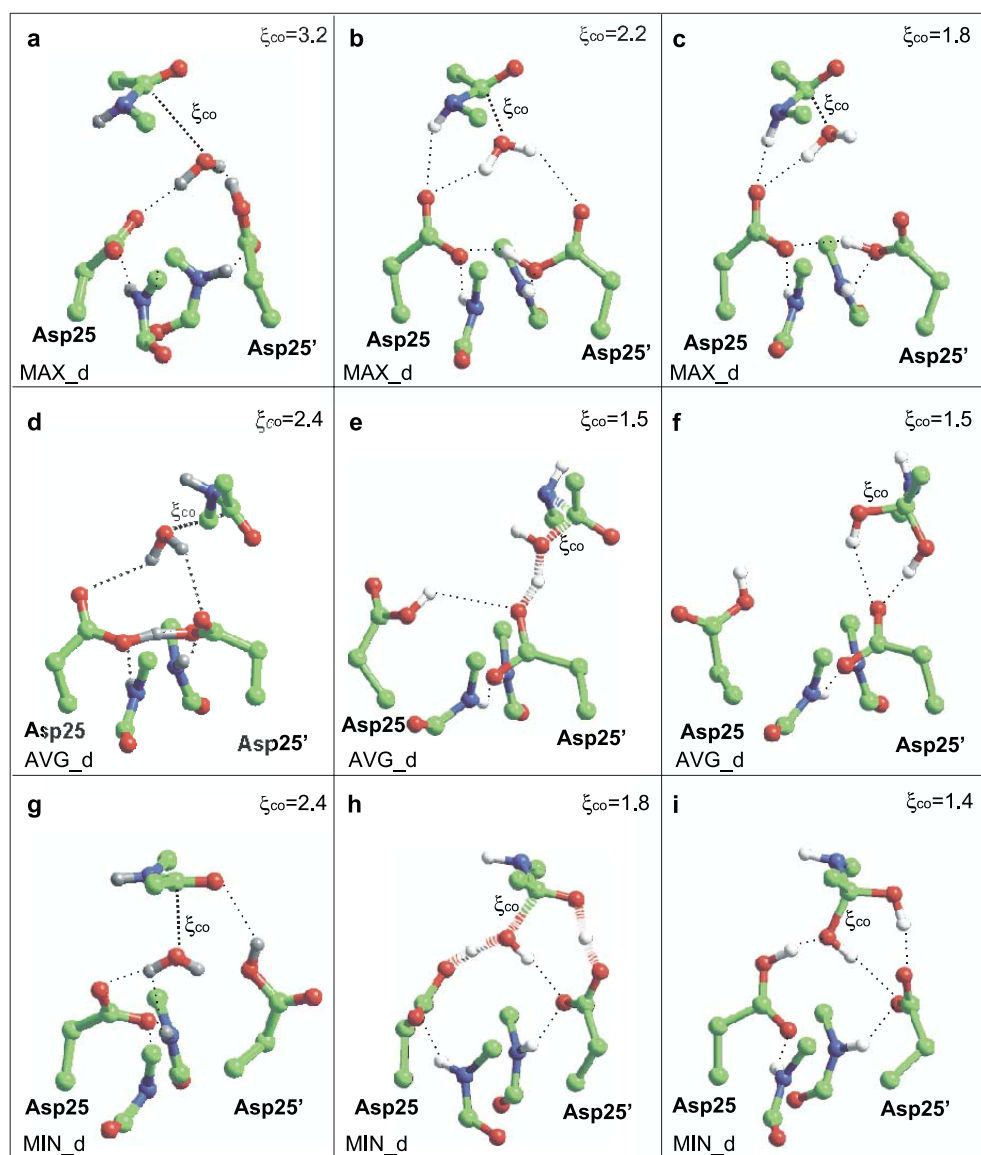
In the previous section we have seen that a large displacement of the substrate towards the Asp dyad is coupled with a complex motion of the entire protein. We now attempt to estimate the influence of the substrate thermal fluctuations on reaction Step 1 of the enzymatic reaction; that is, the nucleophilic attack of the catalytic water (WatC) on the carbonyl carbon atom (Figure 1).

The free energy of the process depends on at least two reaction coordinates, the carbonyl carbon distance ( $\xi_{CO}$ ) and  $d_{SA}$ . Thus, use of thermodynamic integration involves a calculation along the two reaction coordinates. Here, we make the simplifying and plausible assumption that the modes associated with  $d_{SA}$  are so slow relative to those associated with  $\xi_{CO}$  that they can be separated (see also Methods for an evaluation of this assumption). Then, we consider three conformations of the cleavage site/substrate complex extracted by the classical MD simulation, characterized by different

distances  $d_{SA}$  (Figures 2(c) and 6(a)): (i) conformation **MAX\_d**, where  $d_{SA} = 8.1$  Å, which is close to the maximum Asp dyad/substrate during the classical MD simulation (Figure 6(a)); (ii) conformation **AVG\_d**, in which  $d_{SA} = 7.7$  Å, which is the MD-averaged distance; (iii) conformation **MIN\_d**, in which  $d_{SA}$  reaches its minimum value during the 2.1 ns MD (7.17 Å). Finally, we calculate the activation free energy for each complex by integrating the constraint force along the  $\xi_{CO}$  distance (Figure 2(b)) (see Methods). These calculations are carried out on model complexes that include a water molecule, a model of the substrate, the Asp dyad and the Thr–Gly peptide bond unit (Figure 2(c)). The latter residues provide almost all the polarization experienced by the reactants (see the previous sections).

### Complex MAX\_d

For  $\xi_{CO}$  ranging from 4.00 to 2.33 Å, WatC forms an H-bond with both Asp25' and Asp25, and the large backbone unit dipole<sup>54</sup> of the substrate points towards the Asp pair negative charge (Figure 7(a) and (b)). The approach of WatC to the carbonyl carbon atom ( $\xi_{CO}$  ranging from 2.33 to 1.75 Å) is accompanied by the disruption of the H-bond with Asp25' (Figure 7(c)) and by the loss of the planarity of the substrate peptide bond. The  $C^\alpha$ –N–C–O torsion angle exhibits large deviations



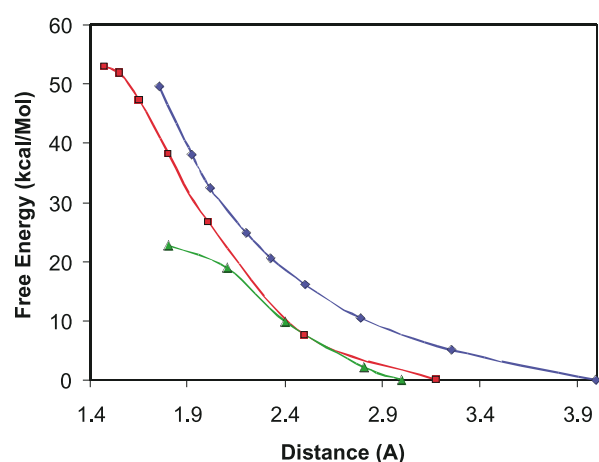
**Figure 7.** Constrained *ab initio* MD calculations. Selected intermediate (a, b, d, e, g, h) and final (c, f, i) MD snapshots of complexes MAX\_d, AVG\_d, and MIN\_d are shown. In (b), (e), (i)  $\xi_{co} = 2.0, 1.47$  and  $1.81$  Å, respectively.

from  $0^\circ$ , which increase with the decrement of  $\xi_{co}$ . However, the stereochemistry of the carbonyl carbon atom is still very close to that of an  $sp^2$  carbon atom: the ( $C^\alpha$ -C-O) and (N-C-O) bond angles are close to  $120^\circ$ ; consistently, the C-O bond length does not change significantly, indicating that the latter retains its double bond character along the simulation. On the other hand, the peptide C-N bond length increases by as much as  $0.1$  Å during the simulation. In conclusion, close proximity of the WatC oxygen atom is followed by loss of peptide unit planarity and C-N bond elongation without affecting the  $sp^2$  geometry at the carbonyl carbon atom. At  $\xi_{co}$  as small as  $1.75$  Å (Figure 7(c)), the reaction has not occurred and the calculated activation energy (Figure 8) is much larger than the experimental value (namely, 50 and 16–18 kcal/mol,<sup>55</sup> respectively:  $1 \text{ cal} = 4.184 \text{ J}$ ). We decided therefore to interrupt the simulation at this stage and proceed to the next complex.

### Complex AVG\_d

The distance between the substrate and the Asp dyad distance is  $\approx 0.4$  Å shorter than that in MAX\_d. For  $\xi_{co}$  ranging from  $3.17$  to  $2.50$  Å, WatC prevalently H-bonds to the negatively charged Asp25 (Figure 7(d) and (e)). As  $\xi_{co}$  is shortened (from  $2.50$  to  $1.80$  Å), WatC forms strong H-bonds with both Asp groups, and proton hopping between the aspartyl  $O^\delta$  oxygen atoms occurs in the picosecond time-scale, as with the free enzyme.<sup>54</sup> The approach of WatC is accompanied in this case by a deviation from planarity of the peptide unit. The peptide C-N bond length increases significantly, while the peptide C-O bond is essentially unaffected.

At  $\xi_{co} \approx 1.55$  Å, WatC transfers a proton to Asp25' (Figure 7(e)): a negatively charged, reactive intermediate is formed. The C-N distance in the intermediate fluctuates from  $1.6$  up to  $1.9$  Å. The



**Figure 8.** Activation free energies. The free energies for reaction Step 1 as obtained from the constrained *ab initio* MD simulations are plotted as a function of the constrained coordinate  $\xi_{\text{CO}}$ . Green, red, and blue refer to models MAX\_d, AVG\_d and MIN\_d, respectively.

C–O bond is still short (1.28 (0.02) Å) and the C $^{\alpha}$ –C–N and N–C–O bond angles are close to 120°: thus, at this stage, the carbon atom is still hybridized sp<sup>2</sup> and forms a double bond with the carbonyl oxygen atom. Within a few hundred femtoseconds, the intermediate evolves to a stable gem-diol by a proton transfer from Asp25 to the carboxylic oxygen atom of the substrate (Figure 7(f)).

The pathway observed is not consistent with the commonly accepted mechanism (Figure 1), as it involves a single and not a double proton transfer event. Furthermore, the calculated free energy (50 kcal/mol, see Figure 8) is still much larger than the experimental value. We conclude that this conformer does not exhibit the correct geometry for the reaction.

### Complex MIN\_d

The substrate-aspartyl pair distance here assumes the shortest value during the MD simulation ( $d_{\text{SA}} = 7.1$  Å). For  $\xi_{\text{CO}}$  ranging from 3.0 to 2.1 Å, WatC binds to Asp25, which is deprotonated. Also here, the WatC dipole points to the aspartyl charge (Figure 7(g)). The corresponding hydrogen bond pattern is similar to that of AVG\_d. Asp25' does not form stable hydrogen bonds with the other active site residues and is therefore highly mobile. For  $\xi_{\text{CO}}$  from 3.0 to 2.1 Å, the peptide bond loses its planarity (Figure 7(g)); the peptide C–O bond length is not affected greatly, as in the other complexes. In contrast, the peptide C–N bond length increases significantly from 1.37 to 1.41 Å. At  $\xi_{\text{CO}} = 1.8$  Å WatC transfers a proton to Asp25 and, simultaneously, Asp25' transfers its proton to the substrate carbonyl oxygen atom (Figure 7(h)). As a consequence, the average force acting on the  $\xi_{\text{CO}}$  distance constraints

changes its sign, an indication that the transition state has been reached.<sup>56</sup> This concerted transfer stabilizes the very reactive OH<sup>−</sup> species, which readily attacks the carbonyl carbon atom to form a stable gem-diol.

The transition state is characterized by a substrate carbonyl carbon atom still largely sp<sup>2</sup>, as indicated by its bond angles (see Figure 7(h)). Full sp<sup>3</sup> geometry of the carbon atom is achieved only once the C–O<sub>WatC</sub> bond is fully formed (Figure 7(i)).

The transition state is stabilized by a double proton transfer, in agreement with the commonly accepted mechanism (Figure 1).<sup>57</sup>

The calculated reaction free energy is 20 kcal/mol. This estimation has to be corrected for the probability that the system is in a favorable conformation. To better estimate this probability, we have carried out a new, much longer MD simulation (overall 8 ns, see Methods). Over the 8 ns run, we find that the probability takes place ~4% of the time. Thus, a correction of ~1.5 kcal/mol should be added to the calculated 20 kcal/mol value, which still leaves our estimation in fair agreement with the experimental value (16–18 kcal/mol).<sup>55,57,58</sup>

In conclusion, the catalytic reaction appears to take place only when  $d_{\text{SA}}$  is short.

The structural and dynamical properties of the intermediate are explored by carrying out 1.4 ps of unconstrained *ab initio* MD starting from the last snapshot of the free energy calculations. The intermediate, which H-bonds both Asp25 and Asp25' (Figure 7(i)), is relatively rigid (rmsd 0.8 Å). The MD-averaged structure compares very well with available X-ray structures of aspartic proteases complexed with an important class of transition state analog inhibitors, the difluorostatone derivatives,<sup>59–61</sup> which contain the gem-diol form of the type –CF<sub>2</sub>–C(OH)<sub>2</sub> (Table 3).

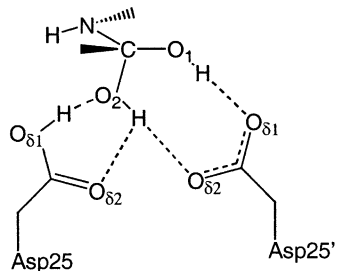
### Discussion

We have investigated the conformational flexibility of the enzyme-substrate complex by computer simulations. The combination of electrostatic modeling, classical MD and *ab initio* calculations has allowed us to explore the flexibility of the enzyme-substrate complex on a relatively long time-scale and to elucidate the influence of the thermal fluctuation on the enzymatic activity.

Our electrostatic modeling suggests that the electric field generated by the protein residues beside the Asp dyad during the dynamics is small, consistent with the fact that the residues around the cleavage site are mostly hydrophobic.<sup>36,62</sup> the field at the cleavage site is essentially that created by the highly conserved D(25,25')–T(26,26')–G(27,27') residues.<sup>5</sup>

Our classical MD calculations provide information on the conformational flexibility of the biomolecular complex on the nanosecond

**Table 3.** Gem-diol intermediate

	MD-averages (this work)	HIV-1 PR	Endothiapep.	Penicillopep.
				
C–O1	1.42(0.03)	1.37	1.38	1.40
C–O2 ( $\xi_{CO}$ )	1.50(0.06)	1.40	1.47	1.47
O1...O <sup><math>\delta</math>1</sup> (Asp25')	2.69(0.12)	2.65	2.58	2.62
O2...O <sup><math>\delta</math>2</sup> (Asp25')	2.92(0.34)	2.83	2.90	2.67
O2...O <sup><math>\delta</math>2</sup> (Asp25)	2.80(0.14)	3.15	2.58	2.63

Comparison of average bond distances from the *ab initio* MD with those of HIV-1 PR<sup>60</sup>, penicillopepsin<sup>59</sup> and endothiapepsin–difluorostatone<sup>61</sup> complexes.

time-scale. The fireman's grip region around the active site is rather rigid and does not exhibit significant motions on the nanosecond time-scale. In contrast, the flaps, cantilever and fulcrum exhibit larger flexibility during the dynamics. These findings are in qualitative agreement with NMR,<sup>49–51</sup> crystallographic data<sup>35</sup> and previous MD investigations on the sub-nanosecond time-scale.<sup>40–43,46</sup> The calculated NMR order parameters are in excellent agreement with the available experimental data.<sup>49–51</sup>

Novel information is obtained on the substrate motion. The substrate fluctuates along the *z*-direction (which is displayed in Figure 2) backwards and towards the Asp dyad on the nanosecond time-scale (Figure 6(a)). This peculiar motion of the substrate towards the cleavage region is coupled with that of several protein regions, in particular the flap tips and the flap elbows (Figure 6(b)). Indeed, these regions are the most relevant for substrate fluctuations (Figure 6(c)); consistently, a new MD simulation, in which a constraint is imposed on the flap elbows (Figure 6(d)), causes a motion of the substrate towards the cleavage site.

Thus, our calculations provide support to the suggestion that residues far from the active site can be very important for modulating the substrate motions. Most interestingly, several compensatory mutations (M36I, M46I-L, L47V, L63P) are located in these regions or in their close proximity (Figure 6(c)).

Is this motion relevant for the enzymatic catalysis? To address this issue we have investigated by *ab initio* methods Step 1 of the hydrolysis; that is, the nucleophilic attack of the catalytic water molecule on the substrate carbonyl carbon atom (Figure 1). Constrained *ab initio* MD simulations have been carried out on model complexes extracted by selected conformations of the classical MD simulations, characterized by different substrate/cleavage site distances ( $d_{SA}$  in Figure 6). At large  $d_{SA}$ , the activation free energy of the reaction

is unphysically large. At  $d_{SA}$  corresponding to the average value of the classical MD simulation, the activation free energy is still excessively high and the reaction takes place through a single proton transfer event from water to the Asp dyad, in disagreement with the most widely accepted mechanism (Figure 1).<sup>57</sup> In contrast, at the shortest  $d_{SA}$  observed during the classical MD simulation (MIN\_d in Figure 6(a)), a concerted double proton transfer between the Asp dyad, WatC and substrate takes place (Figure 7(h)): a proton is transferred from Asp25' to the substrate, while simultaneously a proton is transferred from WatC to Asp25. This event, which is in complete agreement with the currently accepted picture of the enzymatic mechanism (Figure 1), leads to the formation of the highly nucleophilic agent OH<sup>–</sup>, which reacts readily to form the intermediate INT (Figure 1). Thus, the crucial role of the Asp dyad cleavage site for this step of the reaction is to stabilize the OH-reactive species by exploiting the optimal acid–base properties of the protonated and deprotonated aspartyl groups. We expect the activation barrier of the same reaction in solution to be larger than that in the enzyme, even in the case where the reactants are located in the same precatalytic conformation as in the protein. This is due to the different acid–base properties of water with respect to those of the Asp dyad.

The structural and dynamical properties of INT (whose existence has been postulated on the basis of experimental<sup>58,63</sup> and theoretical<sup>20,60</sup> studies) are in very close agreement with those of potent intermediate-mimic inhibitors of the enzyme.<sup>59,61</sup> The activation energy ( $\sim 21.5$  kcal/mol) is much lower than that obtained with the previous simulations, and it is fairly close to the value of 16–18 kcal/mol that can be estimated from the experimental kinetic constants.<sup>55,57</sup> The discrepancy between theory and experiment can possibly be ascribed to the use of the BLYP approximation<sup>28</sup> and, more importantly, of a model complex in the gas phase.



For example, the Asp groups are modeled as propanoic acids in our quantum chemical calculations. Use of QM/MM approaches may improve the description of the chemical properties such as the proton affinity. However, we stress here that our models (i) include the most important electrostatic interactions; (ii) capture the key aspects of the enzyme-catalyzed nucleophilic reaction, in complete agreement with the most widely accepted mechanism; (iii) provide a structure for the Step 1 transition state (TS1; see Figure 1) that is in excellent agreement with the available X-ray data.

In conclusion, our findings provide a consistent picture in which TS1 can be formed only at short  $d_{SA}$ , where the Asp dyad can perform a concerted double proton transfer. The short  $d_{SA}$  result in turn from the mechanical fluctuations induced by the protein motions.

Our results support the suggestion that protein dynamical fluctuation plays a key role in enzymatic catalysis<sup>64–68</sup>, and, more specifically, that correlated motions can be crucial for specific steps of enzymatic reactions.<sup>69</sup>

Furthermore, they validate the “rack mechanism” proposed by R. Lumry for aspartic proteases.<sup>70–72</sup> According to this mechanism, proteases can be subdivided into regions called “knots” and “matrices”. The knots, which include the relatively rigid protein core, are essential for the protein structure and very often contain the catalytic residues; the matrices are more flexible protein domains characterized by large-amplitude/low-frequency motions and are essential for substrate recognition.<sup>70</sup> In particular, in HIV-1 PR, the fireman’s grip holding the two monomers is a knot<sup>71</sup> and it is expected to be relatively rigid, as found in our calculations. The flexible flap, tips and elbows, which are involved in the large-scale motion, are the matrices.

Finally, our data may provide a rationale for some of the observed drug-resistant mutations,<sup>12</sup> i.e. the capability of drastically reducing drug affinity without affecting the enzymatic reaction significantly through the mutations of residues in the protein. In drug-resistant mutants, the active-site mutations are very often associated with mutations that partially restore the enzymatic function.<sup>12–14,82</sup> Several of these “compensatory” mutations have been observed in the flap or cantilever region that are important for modulating the substrate motion (M36I, M46I, L47V, L63P; Figure 6(c)) and they do not involve any dramatic change of the polarity of the residues.<sup>13,15–18</sup> The motion of Met46 and Leu47 is coupled directly to that of the substrate (Figure 5(b)); furthermore, all the above mutations are expected to change the structural properties of the enzyme. Thus, our results strongly support the idea<sup>13,43</sup> that the mutations in these positions may enhance the catalytic rate of HIV-1 PR mutants by affecting the flexibility of the protein, which in turn may affect the conformational properties of the enzyme.

## Methods

### Classical molecular dynamics calculations

The structural model for the enzyme/*N*-acetyl Thr-Ile-Met-Met-Gln-Arg amide-substrate (SUB) complex was constructed starting from an HIV-1 PR/peptide inhibitor X-ray structure<sup>19</sup> (4HVP entry in the PDB database<sup>73</sup>). The inhibitor sequence is: *N*-acetyl Thr-Ile-Nle-Ψ[CH<sub>2</sub>NH]-Nle-Gln-Arg-amide<sup>19</sup> (Nle = norleucine). To transform the inhibitor in the substrate *N*-acetyl Thr-Ile-Met-Met-Gln-Arg amide, Nle was replaced with Met and the CH<sub>2</sub>NH group was changed to a peptide bond. In the crystal structure, the Thr26′ OH group is rotated towards a hydrophobic region and does not form the typical hydrogen bonding pattern observed in all the X-ray structures of HIV-1 PR/inhibitors so far determined. As stated by the crystallographers (A. Wlodawer, personal communication), the reason for the OH flipping is unknown. In our structural models, we rotated the side-chain of Thr26′ around the C–C bond so as to reform the typical hydrogen bond network observed in HIV-1 PR.

Protein residues belonging to subunit 1 were numbered from 1 to 99. Those belonging to subunit 2 were numbered from 1′ to 99′. Substrate residues binding to subunit 1 (2) were numbered from P1 (P1′) to P3 (P3′)<sup>74</sup>: Thr(P3)-Ile(P2)-Met(P1)-Met(P1′)-Gln(P2′)-Arg(P3′). Thus, the peptide bond to be cleaved by the enzyme belongs to residues P1 and P1′ (Met–Met).

Upon binding of SUB, the aspartyl dyad is believed to be monoprotonated.<sup>57</sup> As the proton’s location is unknown, the four different protomers resulting from protonation of the four carboxyl oxygen atoms of the Asp dyad were initially considered. The most stable protomer was determined by *ab initio* DFT-based geometry optimizations using the direct inversion in the iterative subspace method<sup>75</sup> with a convergence criteria of  $5 \times 10^{-3}$  a.u. for the largest component of the atomic forces (for the details of the DFT calculations, see below). The energetically most stable protomers at the *ab initio* level are those in which Asp25′ O<sup>6</sup> is protonated, which is  $\approx 2$  kcal/mol lower in energy than the second most stable conformer.

The HIV-1 PR/SUB complex was immersed in a  $66.8 \text{ \AA} \times 55.2 \text{ \AA} \times 43.0 \text{ \AA}$  box containing 4170 water molecules. The positive charge of the HIV-1 PR-peptide complex (+5) was neutralized by adding five chlorine ions close to five positively charged groups not involved in salt-bridges in the crystal structure of the enzyme. The total system was composed of 15,749 atoms.

The classical MD simulation was carried out using the AMBER suite of programs†. The force-field parameters of the protein, SUB and Cl<sup>−</sup> were those of the AMBER force-field<sup>76</sup>. For water, we used the TIP3P model.

Electrostatic interactions were calculated with the Ewald particle mesh method, using a  $64 \times 54 \times 45$  grid.<sup>77–80</sup> A cubic interpolation between the points was used. The average estimated RMS error on the forces during the simulation was  $\sim 0.0003 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ . The mass center motions were removed every 30 ps. Constant temperature and pressure conditions were achieved by coupling the system to Berendsen’s thermostat and barostat<sup>81</sup> with a relaxation time of 0.2 ps. Bonds involving hydrogen atoms were constrained to

† <http://www.amber.ucsf.edu>

their equilibrium position with the SHAKE algorithm. The time-step was 1.5 fs.

The system was heated at 150 K for 15 ps, then at 300 K for 15 ps. Subsequently, 705 ps of equilibration at 300 K was carried out. Following Liu *et al.*,<sup>83</sup> we imposed position restraints in the cleavage site during this equilibration phase. Specifically: (i) we restrained the position of the Asp dyad with an harmonic restraint of 1.0 kcal mol<sup>-1</sup> Å<sup>2</sup>; (ii) the HN<sub>(Gly27,27')</sub>-O<sub>(MetP1,P2')</sub> H-bonds (see Chart 1) were restrained with an harmonic constraint of 0.5 kcal mol<sup>-1</sup> Å<sup>2</sup> centered at 2.3 Å. These H-bonds are conserved in the complexes of HIV-1 PR with peptido-mimetic inhibitors for which the X-ray structure has been determined. On the other hand, previous studies reported difficulties in maintaining them.<sup>83</sup> After the equilibration phase, these constraints were released and three MD simulations at 300 K and 1 atm pressure (101,325 Pa) were carried out. The first, which was used to extract the conformations of the quantum-chemical calculations (see below), was obtained by prolonging the simulation by a further 2.1 ns, which was collected for analysis. The second, which was carried out to check the convergence of the first calculation, was carried out for 8.0 ns, of which 7.8 ns was collected for data analysis. The third MD simulation, which was used to validate our force analysis (see below), was performed for 0.35 ns by imposing a soft harmonic restraint ( $k_{\text{HARM}} = 0.1$  kcal mol<sup>-1</sup> Å<sup>2</sup>,  $r_{\text{EQ}} = 28$  Å) on the C<sup>α</sup> atoms having the largest effect on the substrate motion according to the large-scale motion analysis reported in the next section (C<sub>(68)</sub><sup>α</sup>-C<sub>(39)</sub><sup>α</sup>, C<sub>(68')</sub><sup>α</sup>-C<sub>(39')</sub><sup>α</sup>, C<sub>(69)</sub><sup>α</sup>-C<sub>(40)</sub><sup>α</sup> and C<sub>(69')</sub><sup>α</sup>-C<sub>(40')</sub><sup>α</sup> atoms belonging to the flap elbow). The last 0.3 ns was collected for analysis. In the last two simulations, random starting velocities to the atoms were assigned.

### Calculated properties

NMR order parameters ( $S^2$ ) were calculated using the model-free formulation<sup>47,48</sup> and following the procedure already described for HIV-1 PR.<sup>41</sup> Order parameters were calculated separately for the first (2.1 ns) and second (8.0 ns) MD simulations. No significant differences were observed in the sets of calculated values.

Global protein modes eigenvectors  $\vec{A}_i(\vec{a}_{i1}, \vec{a}_{i2}, \dots, \vec{a}_{iN})$  and their eigenvalues  $\lambda_i$  ( $i = 1, 3N$  where  $N$  is the total number of C<sup>α</sup> atoms and  $\vec{a}_{ij}$  is the component on the  $j$ th atom of the  $3N$  dimensional eigenvector), were calculated by diagonalizing the coordinates covariance matrix,<sup>52,53</sup> whose elements read:

$$c_{ij} = \langle \Delta \vec{r}_i(t) \Delta \vec{r}_j(t) \rangle$$

where  $\Delta \vec{r}_i = \langle \vec{r}_i(t) \rangle - \vec{r}_i(t)$ . The modes having the largest eigenvalues are associated with the largest global atomic displacements<sup>52</sup> and define the so-called “essential subspace”. The calculated eigenvectors of the 2.1 and 8.0 ns MD simulations turned out to be very similar. Indeed, the root-mean-square of the inner products of the ten largest eigenvectors was as large as 0.75.<sup>84</sup>

The calculation of the covariance matrix eigenvectors allows identification of the C<sup>α</sup> atoms most relevant for the substrate motion. To this end, we calculate the largest displacement along  $d_{\text{SA}}$  caused by a unitary force  $\vec{f}_j$  applied to the  $j$ th C<sup>α</sup> atom. The orientations that cause the largest displacement  $\delta_j$  are those for

which:

$$\vec{f}_j = \frac{\sum_i^{3N-6} \vec{a}_{ij} \frac{(\vec{A}_i \times \vec{D}_{\text{SA}})}{K_i}}{\left| \sum_i^{3N-6} \vec{a}_{ij} \frac{(\vec{A}_i \times \vec{D}_{\text{SA}})}{K_i} \right|}$$

The vector  $\vec{D}_{\text{SA}}(\vec{d}_1, \vec{d}_2, \dots, \vec{d}_{\text{P1}}, \vec{d}_{\text{P2}}, \dots, \vec{d}_N)$ , which describes the motion of the substrate along  $d_{\text{SA}}$ , has non-zero components only for  $\vec{d}_{\text{P1}} = (\vec{D}_1 + \vec{D}_3)/2$  and  $\vec{d}_{\text{P1}'} = (\vec{D}_2 + \vec{D}_4)/2$  (Figure 6(a)).  $K_i$  is the force constant associated with the  $i$ th eigenvector ( $\vec{A}_i$ ). In the approximation that  $\delta_j$  is small, the forces can be assumed to be harmonic and  $K_i$  approximated as:  $K_i \approx k_{\text{B}}T/\lambda_i$ .<sup>85</sup>

Thus:

$$\delta_j = \sum_i^{3N-6} \vec{f}_j \times \vec{a}_{ij} \frac{(\vec{A}_i \times \vec{D}_{\text{SA}})}{K_i}$$

that is:

$$\delta_j \approx \left| \sum_i^{3N-6} \lambda_i \times \vec{a}_{ij} \frac{(\vec{A}_i \times \vec{D}_{\text{SA}})}{k_{\text{B}}T} \right|$$

$\delta_j$  is then evaluated for each C<sup>α</sup> atom. Finally, we select the C<sup>α</sup> atoms that are associated with the largest displacement (Figure 6(c)).

The maximum displacements computed based on the 2.1 and 8.0 ns MD simulations do not differ significantly, as can be expected from the similarity of the eigenvector spectrum.

### Ab initio quantum chemical calculations

In this work, we use *ab initio* MD simulations to investigate a key step of HIV-1 PR enzymatic reaction (Figure 1). All calculations were carried out with the CPMD program†.

We focus on Step 1 as: (i) it is firmly established<sup>58,60</sup> as opposed to Step 2, which is still the subject of controversy; (ii) it is easy to construct a model of the substrate/enzyme reaction; (iii) it is possible to compare structural properties of the gem-diol transition state (see Figure 1) with those of the transition state analog inhibitors of the enzyme;<sup>59–61</sup> (iv) the two steps have similar energy barriers.<sup>20,21,60</sup>

The structural models included the Asp25 (25') side-chains (modeled as propionic acid and propionate, respectively), the cleavage region of the substrate (modeled as *N*-methyl acetamide), the catalytic water molecule (Wat) and the Thr26(26')–Gly27(27') peptide bonds (modeled as *N*-methyl formamide) (Figure 2(c)).

The quantum problem was solved using density functional theory (DFT). Exchange and correlation functionals were those of Becke<sup>86</sup> and Lee *et al.*,<sup>87</sup> respectively. The Kohn-Sham orbitals were expanded in plane waves up to 70 Ry. Troullier & Martin<sup>88</sup> pseudopotentials were used to describe the interactions between the ionic cores and the valence electrons. A 14 Å × 14 Å × 16 Å cell was used. The systems were treated as isolated, as was done by Barnett & Landmann.<sup>89</sup> Test calculations with very large cells (up to 20 Å × 20 Å × 22 Å) indicated that the relative energies of different protomers were well converged with respect

† <http://www.cpmd.org>

to the cell size (the differences being of the order of  $\sim 0.1$  kcal mol $^{-1}$ ).

DFT-based *ab initio* MD simulations were performed according to the Car–Parrinello approach.<sup>24</sup> A time-step of 0.096 fs and a fictitious electron mass of 400 a.u. were used. Constant temperature was achieved by coupling the systems to a Nosé thermostat of 500 cm $^{-1}$  frequency. The positions of the terminal atoms were kept fixed during the *ab initio* calculation so as to mimic the rigid protein frame.<sup>35,41,49</sup> This appears as a plausible assumption as: (i) during our classical MD, the fireman's grip backbone, which includes the cleavage site, turns out to be rigid (Figure 4(c)); (ii) this procedure has been shown to describe accurately structural and dynamical properties of the cleavage site in the free enzyme and in that complexed with pepstatin.<sup>54,90</sup>

The reactivity of three conformations (MIN\_d, AVG\_d, and MAX\_d, corresponding to MD snapshots after 585, 980, and 1560 ps of the classical MD simulation, see Figure 6(a)) was investigated. These conformations, characterized by different substrate/Asp dyad distances  $d_{SA}$ , were investigated. While AVG\_d and MAX\_d conformations are encountered several times during the dynamics, conformations with  $d_{SA}$  shorter than 7.4 Å (roughly corresponding to MIN\_d) represent 2% of the simulation. This is, however, caused by the finite length of the MD simulation that samples infrequent events inadequately. Indeed, in the 8 ns MD simulations, the number of events leading to conformations similar to MIN\_d on this time-scale is larger (about 4% of the simulation).

Reaction free energies were calculated with the method of the constraints.<sup>26–28,91</sup> The reaction coordinate was  $\xi_{CO} = d(C_{(sub)} - O_{(WatC)})$ . This choice has been reported in the literature for water nucleophilic attack to substrate carbonyl group in HIV-1 PR<sup>83</sup> (Figure 2(b)). The activation free energy was calculated as an integral of the average force  $f_s$  acting on the constraint along the reaction coordinate  $Q$ :<sup>92</sup>

$$\Delta F(Q) = \int_{Q_0}^Q dQ f_s(Q)$$

Each point along  $\xi_{CO}$  was sampled for  $\approx 1.0$  ps of *ab initio* MD. The initial 0.3 ps was discarded. In the relatively short dynamics explored (1.0 ps), the average constrained force  $f_s$  appears well equilibrated, the averages taken from the first and second half of the simulation differ by less than 5%. The only exceptions are those points where the transition state is crossed. In these cases, only the part of the trajectory following the reactive event was considered. Overall,  $\approx 10$  ps was sampled for each model.

This calculation implicitly assumes that the time-scale of protein fluctuations is much larger than the other modes. To verify this hypothesis, three constrained classical simulations were performed starting from MD snapshots after 585, 980, and 1560 ps of the classical MD simulation, that is the starting points of the constrained *ab initio* MD simulations (see above). The  $\xi_{CO}$  distance was constrained and reduced progressively from 3.8 to 1.8 Å during the simulations. The simulation time was the same as that used for the *ab initio* simulation (see below). No major rearrangement of the substrate position and orientation was observed during these simulations, confirming that long time-scale, slow protein fluctuations are well separated from the other near constraints subspace modes.<sup>84</sup>

The gem-diol intermediate obtained at the end of the *ab initio* MD from model MIN\_d underwent 1.4 ps of *ab initio* MD calculations with no distance constraints. Some calculations were performed in the presence of the electrostatic field of the protein frame as was done by Piana & Carloni.<sup>54</sup> Centers of Boys orbitals WOC<sup>93,94</sup> were calculated as reported.<sup>95</sup>

The electric dipole moment of WatC was calculated from the Boys orbital centers of model AVG\_d. The dipole moment calculation was averaged over six different configurations taken from the *ab initio* MD simulation of model AVG\_d = 3.17 Å.

## Acknowledgments

We thank COFIN-MURST for financial support. INFN is gratefully acknowledged for giving us the computer resources. S.P. thanks Regione Friuli Venezia Giulia for the "Giovani Ricercatori" grant. We thank Dexter Northrop and Rufus Lumry for pointing out, after the paper was submitted, the relevance of our calculations for the rack mechanism proposed by R. Lumry. We thank Leonardo Guidoni for many useful discussions.

## References

- Wlodawer, A. & Vondrasek, J. (1998). Inhibitors of HIV-1 protease: a major success of structure-based drug design. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 249–284.
- Wlodawer, A. & Erickson, J. W. (1993). Structure-based inhibitors of HIV-1 protease. *Annu. Rev. Biochem.* **62**, 543–585.
- Tomasselli, A. G., Howe, W. J., Sawyer, T. K., Wlodawer, A. & Heinrikson, R. L. (1991). The complexities of AIDS: an assessment of the HIV protease as a therapeutic target. *Chim. Oggi*, **5**, 6–27.
- Debouck, C., Gornick, J. G., Strickler, J. E., Meek, T. D., Metcalf, B. W. & Rosenberg, M. (1987). Human immunodeficiency virus protease expressed in *Escherichia coli* exhibits autoprocessing and specific maturation of the gag precursor. *Proc. Natl Acad. Sci. USA*, **84**, 8903–8906.
- Fitzgerald, P. M. D. & Springer, J. P. (1991). Structure and function of retroviral proteases. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 299–320.
- Poorman, R. A., Tomasselli, A. G., Heinrikson, R. L. & Kezdy, F. L. (1991). A cumulative specificity model for proteases from human immunodeficiency virus types 1 and 2, inferred from statistical analysis of an extended substrate data base. *J. Biol. Chem.* **266**, 14554–14561.
- Tomasselli, A. G., Hui, J. O., Sawyer, T. K., Staples, D. G., Fitzgerald, D. J., Chaudary, V. K. *et al.* (1990). Interdomain hydrolysis of a truncated pseudomonas exotoxin by the human immunodeficiency virus-1 protease. *J. Biol. Chem.* **265**, 408–413.
- Turner, B. G. & Summers, M. F. (1999). Structural biology of HIV. *J. Mol. Biol.* **285**, 1–32.
- McQuade, T. J., Tomasselli, A. G., Liu, L., Karacostas, V., Moss, B., Sawyer, T. K. *et al.* (1990). A synthetic HIV-1 protease inhibitor with antiviral activity arrests HIV-like particle maturation. *Science*, **247**, 454–456.

10. Fersht, A. R. (1999). *Structure and Mechanism in Protein Science: a Guide to Enzyme Catalysis and Protein Folding*, W.H. Freeman, New York pp. 74.
11. Warshel, A. (1998). Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J. Biol. Chem.* **273**, 27035–27038.
12. Gulnik, S., Erickson, J. W. & Xie, D. (2000). HIV protease: enzyme function and drug resistance. *Vitamins Hormon.* **58**, 213–256.
13. Schock, H. B., Garsky, V. M. & Kuo, L. C. (1996). Mutational anatomy of an HIV-1 protease variant conferring cross-resistance to protease inhibitors in clinical trials. Compensatory modulations of binding and activity. *J. Biol. Chem.* **271**, 31957–31963.
14. Ridky, T. W., Kikonyogo, A., Leis, J., Gulnik, S., Copeland, T., Erickson, J. W. *et al.* (1998). Drug-resistant HIV-1 protease identify enzyme residues important for substrate selection and catalytic rate. *Biochemistry*, **37**, 13835–13845.
15. Pazhanisamy, S., Stuver, C. M., Cullinan, A. B., Margolin, N., Rao, B. G. & Livingston, D. J. (1996). Kinetic characterization of human immunodeficiency virus type-1 protease-resistant variants. *J. Biol. Chem.* **271**, 17979–17985.
16. Martinez-Picado, J., Savara, A. V., Sutton, L. & D'Aquila, R. T. (1999). Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1. *J. Virol.* **73**, 3744–3752.
17. Markowitz, M., Mo, H., Kempf, D. J., Norbeck, D. W., Bhat, T. N., Erickson, J. W. & Ho, D. D. (1995). Selection and analysis of human immunodeficiency virus type 1 variants with increased resistance to ABT-538, a novel protease inhibitor. *J. Virol.* **69**, 701–706.
18. Nijhuis, M., Schuurman, R., de Jong, D., Erickson, J., Gustchina, E., Albert, J. *et al.* (1999). Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *AIDS*, **13**, 2349–2359.
19. Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L. *et al.* (1989). Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. *Science*, **246**, 1149–1152.
20. Okimoto, N., Tsukui, T., Hata, M., Hoshino, T. & Tsuda, M. (1999). Hydrolysis mechanism of the phenylalanine-proline peptide bond specific to HIV-1 protease: investigation by the *ab initio* molecular orbital method. *J. Am. Chem. Soc.* **121**, 7349–7354.
21. Lee, H., Darden, T. A. & Pedersen, L. G. (1996). An *ab initio* quantum mechanical model for the catalytic mechanism of HIV-1 protease. *J. Am. Chem. Soc.* **118**, 3946–3950.
22. Venturini, A., López-Ortiz, F., Alvarez, J. M. & Gonzalez, J. (1998). theoretical proposal of a catalytic mechanism for the HIV-1 protease involving an enzyme-bound tetrahedral intermediate. *J. Am. Chem. Soc.* **120**, 1110–1111.
23. Stebbins, J., Towler, E. M., Tennant, M. G., Deckman, I. C. & Debouck, C. (1997). The 80's loop (residues 78 to 85) is important for the differential activity of retroviral proteases. *J. Mol. Biol.* **267**, 467–475.
24. Car, R. & Parrinello, M. (1985). Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Letters*, **55**, 2471–2474.
25. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical integration of the cartesian equation of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341.
26. Carter, E. A., Ciccotti, G., Hynes, J. T. & Kapral, R. (1989). Constrained reaction coordinate dynamics for the simulation of rare events. *Chem. Phys. Letters*, **156**, 472–477.
27. Curioni, A., Sprik, M., Andreoni, W., Schiffer, H., Hutter, J. & Parrinello, M. (1997). Density functional theory-based molecular dynamics simulation of acid-catalyzed chemical reactions in liquid trioxane. *J. Am. Chem. Soc.* **119**, 7218–7229.
28. Meijer, E. J. & Sprik, M. (1998). *Ab initio* molecular dynamics study of the reaction of water with formaldehyde in sulfuric acid solution. *J. Am. Chem. Soc.* **120**, 6345–6355.
29. Alber, F., Folkers, G. & Carloni, P. (1999). Conformational analysis of dimethyl phosphate in aqueous solution: a density functional theory based-molecular dynamics study. *J. Mol. Struct. (Theochem.)*, **489**, 237–245.
30. Pantano, S., Alber, F. & Carloni, P. (2000). Proton dynamics in an enzyme substrate: an *ab initio* molecular dynamics study. *J. Mol. Struct. (Theochem.)*, **530**, 177–181.
31. Röthlisberger, U., Carloni, P., Doclo, K. & Parrinello, M. (2000). A comparative study of galactose oxidase and active site analogs based on QM/MM Car-Parrinello simulations. *J. Biol. Inorg. Chem.* **5**, 236–250.
32. De Santis, L. & Carloni, P. (1999). Serine proteases: an *ab initio* molecular dynamics study. *Proteins: Struct. Funct. Genet.* **37**, 611–618.
33. Röthlisberger, U. (1998). *Ab initio* and hybrid molecular dynamics simulations of the active site of human carbonic anhydrase II: a test case study. In *Combined Quantum Mechanical and Molecular Mechanical Methods* (Gao, J. & Thompson, M. A., eds), pp. 264–274, American Chemical Society, Washington, DC.
34. Dreyer, G. B., Metcalf, B. W., Tomaszek, T. A., Jr, Carr, T. J., Chandler, A. C., III, Hyland, L. *et al.* (1989). Inhibition of human immunodeficiency virus 1 protease in vitro: rational design of substrate analogue inhibitors. *Proc. Natl Acad. Sci. USA*, **86**, 9752–9756.
35. Prabu-Jeyabalan, M., Nalivaika, M. & Schiffer, C. (2000). How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. *J. Mol. Biol.* **301**, 1207–1220.
36. Ishima, R., Louis, J. M. & Torchia, D. A. (2001). Characterization of two hydrophobic methyl clusters in HIV-1 protease by NMR spin relaxation in solution. *J. Mol. Biol.* **305**, 515–521.
37. Suguna, K., Padlan, E. A., Smith, C. W., Carlson, W. D. & Davies, D. R. (1987). Binding of a reduced peptide inhibitor to the aspartic proteinase from *Rhizopus chinensis*: implications for the mechanism of action. *Proc. Natl Acad. Sci. USA*, **84**, 7009–7013.
38. Silvestrelli, P. L. & Parrinello, M. (1999). Water molecule dipole in the gas phase and in the liquid phase. *Phys. Rev. Letters*, **82**, 3308–3311.
39. Davies, D. R. (1990). The structure and function of the aspartic proteinases. *Annu. Rev. Biophys. Biophys. Chem.* **19**, 189–215.
40. Harte, W. E., Swaminathan, S., Mansuri, M. M., Martin, J. C., Rosenberg, I. E. & Beveridge, D. L. (1990). Domain communication in the dynamical structure of human immunodeficiency virus 1 protease. *Proc. Natl Acad. Sci. USA*, **87**, 8864–8868.
41. Luo, X., Kato, R. & Collins, J. R. (1998). Dynamic flexibility of protein-inhibitor complexes: a study of



- the HIV-1 protease/KNI-272 complex. *J. Am. Chem. Soc.* **120**, 12410–12418.
42. Ringhofer, S., Kallen, J., Dutzler, R., Billich, A., Visser, A. J., Scholz, D. *et al.* (1999). X-ray structure and conformational dynamics of the HIV-1 protease in complex with the inhibitor SDZ283-910: agreement of time-resolved spectroscopy and molecular dynamics simulations. *J. Mol. Biol.* **286**, 1147–1159.
  43. Collins, J. R., Burt, S. K. & Erickson, J. W. (1995). Flap opening in HIV-1 protease simulated by activated molecular dynamics. *Nature Struct. Biol.* **2**, 334–338.
  44. York, D., Darden, T. A., Pedersen, L. G. & Anderson, M. W. (1993). Molecular dynamics simulations of HIV-1 protease in a crystalline environment and in solution. *Biochemistry*, **32**, 1443–1453.
  45. Harrison, R. W. & Weber, I. T. (1994). Molecular dynamics simulations of HIV-1 protease with peptide substrate. *Protein Eng.* **7**, 1353–1363.
  46. Scott, W. R. & Schiffer, C. A. (2000). Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. *Struct. Fold Des.* **8**, 1259–1265.
  47. Lipari, G. & Szabo, A. (1982). Model free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.* **104**, 4559–4570.
  48. Lipari, G. & Szabo, A. (1982). Model free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546–4559.
  49. Freedberg, D. I., Wang, Y. X., Stahl, S. J., Kaufman, J. D., Wingfield, P. T., Kiso, Y. & Torchia, D. A. (1998). Flexibility and function in HIV protease: dynamics of the HIV-1 protease bound to the asymmetric inhibitor Kynostatin 272 (KNI-272). *J. Am. Chem. Soc.* **120**, 7916–7923.
  50. Nicholson, L. K., Yamazaki, T., Torchia, D. A., Grzesiek, S., Bax, A., Stahl, S. J. *et al.* (1995). Flexibility and function in HIV-1 protease. *Nature Struct. Biol.* **2**, 274–280.
  51. Tjandra, N., Wingfield, P., Stahl, S. J. & Bax, A. (1996). Anisotropic rotational diffusion of perdeuterated HIV protease from  $^{15}\text{N}$  NMR relaxation measurements at two magnetic fields. *J. Biol. NMR*, **8**, 273–284.
  52. Amadei, A., Linssen, A. B. M. & Berendsen, H. J. C. (1993). Essential dynamics in proteins. *Proteins: Struct. Funct. Genet.* **17**, 412–425.
  53. Garcia, A. E. & Garcia, A. E. (1992). Large-amplitude nonlinear motions in proteins. *Phys. Rev. Letters*, **68**, 2696–2699.
  54. Piana, S. & Carloni, P. (2000). Conformational flexibility of the catalytic Asp dyad in HIV-1 protease: an *ab Initio* study on the free enzyme. *Proteins: Struct. Funct. Genet.* **39**, 26–36.
  55. Polgár, L., Szeltnér, Z. & Boros, I. (1994). Substrate-dependent mechanism in the catalysis of human immunodeficiency virus protease. *Biochemistry*, **33**, 9351–9357.
  56. Carloni, P., Sprik, M. & Andreoni, W. (2000). Key steps of the cisplatin-DNA interaction. Density functional theory-based molecular dynamics simulations. *J. Phys. Chem. sect. B*, **104**, 823–835.
  57. Hyland, L. J., Tomaszek, T. A. & Meek, T. D. (1991). Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. *Biochemistry*, **30**, 8454–8463.
  58. Hyland, L. J., Tomaszek, T. A., Roberts, G. D., Carr, S. A., Maagard, V. W., Bryan, H. L. *et al.* (1991). Human immunodeficiency virus-1 protease. 1. Initial velocity studies and kinetic characterization of reaction intermediates by  $^{18}\text{O}$  isotope exchange. *Biochemistry*, **30**, 8441–8453.
  59. James, M. N. G., Sielecki, A. R., Hayakawa, K. & Gelb, M. H. (1992). Crystallographic analysis of transition state mimics bound to penicillopepsin: difluorostatine- and difluorostatone-containing peptides. *Biochemistry*, **31**, 3872–3886.
  60. Silva, A. M., Cachau, R. E., Sham, H. L. & Erickson, J. W. (1996). Inhibition and catalytic mechanism of HIV-1 aspartic protease. *J. Mol. Biol.* **255**, 321–346.
  61. Veerapandian, B., Cooper, J. B., Sali, A., Blundell, T. L., Rosati, R. L., Dominy, B. W. *et al.* (1992). Direct observation by X-ray analysis of the tetrahedral intermediate of aspartic proteinases. *Protein Sci.* **1**, 322–328.
  62. Erickson, J. W., Neidhart, D. J., Van Drie, J., Kempf, D. J., Wang, X. C., Norbeck, D. W. *et al.* (1990). Design, activity and 2.8 Å crystal structure of a C2 symmetric inhibitor complexed to HIV-1 protease. *Science*, **249**, 527–533.
  63. Fitzgerald, P. M. D., McKeever, B. M., Van Middlesworth, J. F., Springer, J. P., Heimbach, J. C., Leu, C.-T. *et al.* (1990). Crystallographic analysis of a complex between human immunodeficiency virus type 1 protease and acetyl-pepstatin at 2.0 Å resolution. *J. Biol. Chem.* **265**, 14209–14219.
  64. Karplus, M. & McCammon, J. A. (1983). Dynamics of proteins: elements and function. *Annu. Rev. Biochem.* **52**, 263–300.
  65. Lau, E. Y. & Bruice, T. C. (1998). Importance of correlated motions in forming highly reactive near attack conformations in catechol O-methyltransferase. *J. Am. Chem. Soc.* **120**, 12387–12394.
  66. McCammon, J. A., Wolynes, P. G. & Karplus, M. (1979). Picosecond dynamics of tyrosine side chains in proteins. *Biochemistry*, **18**, 927–942.
  67. Lau, E. Y. & Bruice, T. C. (2000). Comparison of the dynamics for ground-state and transition-state structures in the active site of catechol O-methyl transferase. *J. Am. Chem. Soc.* **122**, 7165–7171.
  68. Cannon, W. R., Singleton, S. F. & Benkovic, S. J. (1996). A perspective on biological catalysis. *Nature Struct. Biol.* **3**, 821–833.
  69. Radkiewicz, J. L. & Brooks, C. L. (2000). Protein dynamics in enzymatic catalysis: exploration of dihydrofolate reductase. *J. Am. Chem. Soc.* **122**, 225–231.
  70. Lumry, R. (1991). Mechanical force, hydration and conformational fluctuations in enzymic catalysis. In *A Study of Enzymes* (Kuby, S. A., ed.), pp. 3–81, CRC Press, Boca Raton, FL.
  71. Lumry, R. (1995). On the interpretation of data from isothermal processes. *Methods Enzymol.* **259**, 628–720.
  72. Lumry, R. & Eyring, H. (1954). Conformational changes of proteins. *J. Phys. Chem.* **58**, 110.
  73. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G. L., Bhat, T. N., Weissig, H. *et al.* (2000). The Protein Data Bank. *Nucl. Acids Res.* **28**, 235–242.
  74. Berger, A. & Schechter, I. (1970). Mapping the active site of papain with the aid of peptide substrates and inhibitors. *Phil. Trans. R. Soc. ser. B*, **257**, 249–264.
  75. Hutter, J., Luthi, H. P. & Parrinello, M. (1994). Electronic structure optimization in plane-wave-based density functional calculations by direct inversion in the iterative subspace. *Comput. Mat. Sci.* **2**, 244–248.

76. Cornell, W. D., Cieplack, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C. *et al.* (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197.
77. Darden, T. A. & York, D. (1993). Particle mesh Ewald: an  $N \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10094.
78. Essman, U., Perera, L., Berkowitz, M. L., Darden, T. A., Lee, H. & Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593.
79. Weerasinghe, S., Smith, P. E., Mohan, V., Cheng, Y.-K. & Pettitt, B. M. (1995). Nanosecond dynamics and structure of a model DNA triple helix in saltwater solution. *J. Am. Chem. Soc.* **117**, 2147–2158.
80. Duan, Y. & Kollman, P. A. (1999). Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, **282**, 740–744.
81. Berendsen, H. J. C., Postma, J. P. M., Van Gusteren, W. F., DiNola, A. & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690.
82. Cecconi, F., Micheletti, C., Carloni, P. & Maritan, A. (2001). Molecular dynamics studies on HIV-1 protease: drug resistance and folding pathways. *Proteins: Struct. Funct. Genet.* **43**, 365–372.
83. Liu, H., Müller-Plathe, F. & Van Gusteren, W. F. (1996). A combined quantum/classical molecular dynamics study of the catalytic mechanism of HIV protease. *J. Mol. Biol.* **261**, 454–469.
84. Amadei, A., Ceruso, M. A. & Di Nola, A. D. (1999). On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins: Struct. Funct. Genet.* **36**, 419–424.
85. Horiuchi, T. & Go, N. (1991). Projection of Monte Carlo and molecular dynamics trajectories onto the normal mode axes: human lysozyme. *Proteins: Struct. Funct. Genet.* **10**, 106–116.
86. Becke, A. (1988). Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. sect. A*, **38**, 3098–3100.
87. Lee, C., Yang, W. & Parr, R. G. (1988). Development of the Colle–Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. sect. B*, **37**, 785–789.
88. Troullier, N. & Martins, J. L. (1991). Efficient pseudo-potentials for plane-wave calculation. *Phys. Rev. sect. B*, **43**, 1943–2006.
89. Barnett, R. N. & Landman, U. (1993). Born–Oppenheimer molecular dynamics simulations of finite systems: structure and dynamics of  $(\text{H}_2\text{O})_2$ . *Phys. Rev. sect. B*, **48**, 2081–2097.
90. Piana, S., Sebastiani, D., Carloni, P. & Parrinello, M. (2001). Ionization states of catalytic residues in HIV-1 protease-pepstatin complex: a reexamination of  $^{13}\text{C}$  NMR data based on density functional theory. *J. Am. Chem. Soc.* **123**, 8730–8737.
91. Sprik, M. & Ciccotti, G. (1998). Free energy from constrained molecular dynamics. *J. Chem. Phys.* **109**, 7737–7744.
92. Ciccotti, G., Ferrario, M., Hynes, J. T. & Kapral, R. (1989). Constrained molecular dynamics and the mean potential for an ion pair in a polar solvent. *Chem. Phys.* **129**, 241–251.
93. Marzari, N., Vanderbilt, D., Silvestrelli, P. L. & Parrinello, M. (1998). Maximally-localized Wannier functions for disordered systems: application to amorphous silicon. *Solid State Commun.* **107**, 7–11.
94. Marzari, N. & Vanderbilt, D. (1997). Maximally-localized Wannier functions for composite energy bands. *Phys. Rev. sect. B*, **56**, 12847–12865.
95. Alber, F., Folkers, G. & Carloni, P. (1999). Dimethyl phosphate: stereoelectronic *versus* environments effects. *J. Phys. Chem.* **103**, 6121–6126.
96. Rose, R., Craik, C. S., Douglas, N. L. & Stroud, R. M. (1996). Three-dimensional structures of HIV-1 and SIV protease product complexes. *Biochemistry*, **35**, 12933–12944.
97. Jaskolski, M., Tomasselli, A. G., Sawyer, T. K., Staples, D. G., Heinrikson, R. L., Schneider, J. *et al.* (1991). Structure at 2.5-Å resolution of chemically synthesized human immunodeficiency virus type 1 protease complexed with a hydroxyethylene-based inhibitor. *Biochemistry*, **30**, 1600–1609.

Edited by P. Wright

(Received 14 March 2001; received in revised form 11 February 2002; accepted 12 February 2002)