SISSA ISAS

Ab initio molecular dynamics studies on HIV-1 protease.

Thesis submitted for the Degree of
Doctor Philosophiae

Candidate

Stefano Piana Agostinetti

Supervisor

Paolo Carloni

October 2000

1. __INTRODUCTION__

1.1. The Human Immunodeficiency Virus.

In the 80's the Human Immunodeficiency Virus of type 1 (HIV-1) was identified as the causative agent of AIDS.
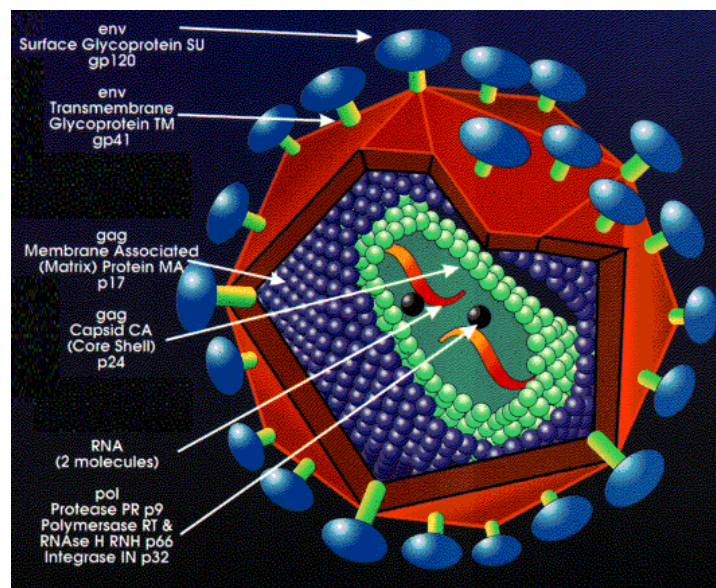
Figure 1. The Human Immunodeficiency Virus Type-1 (HIV-1).
The HIV-1 protease is enclosed in a capside (green) along with RNA and the Reverse Trascriptase enzyme (RT).

HIV-1 is a retrovirus, i.e. the genetic information of the virus is stored on a double filament of RNA. In contrast to the other organisms which use DNA. The RNA filament is protected by a shell of proteins called capsid that is immersed in the cytoplasm[1] (Figure 1).

The viral genome codes for a total of 15 proteins of which only few are actually essential for the viral life cycle and can be exploited as targets for anti-AIDS therapy[2;3].

One of these major targets is the HIV-1 protease enzyme[*] (PR)[2;4-7]. This enzyme is the subject of the present study.
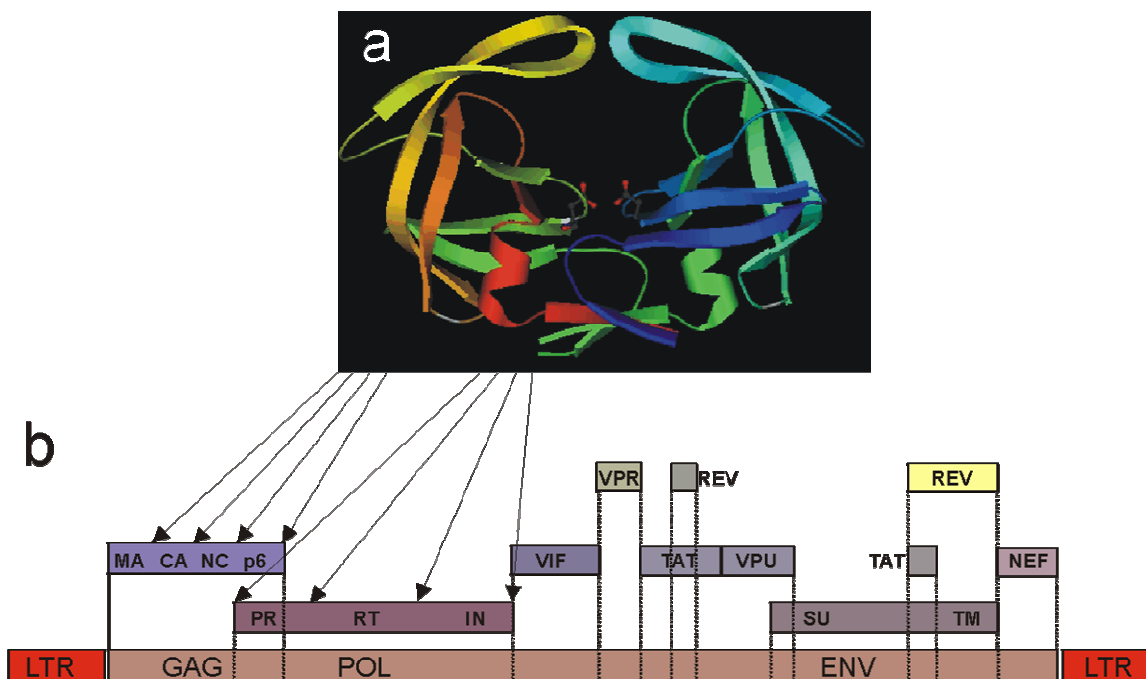
Figure 2. HIV-1 PR structure and sequence cleaved.

(a) structure of the free enzyme; (b) the GAG and POL polypeptide chains are cleaved in the positions indicated by the arrows to yield the active enzymes. MA, CA, NC, p6 are the structural proteins that form the viral shell and capsid; RT is the reverse transcriptase enzyme and IN is the integrase enzyme. Note that also HIV-1 PR itself is part of the POL polypeptide chain[8].

1.2. The HIV-1 Protease enzyme.

HIV-1 PR is required for the correct maturation of the viral particles[4;9]. It cleaves long chains of amino acid, expressed by the virus, in specific positions[†] (Figure 2)[5].

---

[*] Enzymes catalyze the chemical reactions necessary for the metabolism. To this aim they are engineered by the evolution to provide stabilizing interactions for the transition states of the reaction they catalyze. In this way they are able to speed up chemical reactions by several order of magnitude.

1.2.1. Biochemistry of the enzyme.

HIV-1 PR catalyzes the reaction of hydrolysis of the peptide bond in the Gag-Pol polypeptide chain of aminoacids. This enzyme was subject of theoretical[10-30] and experimental[9;31-56] studies aimed at elucidating structural and functional aspects of its biochemistry (For a review see[57;58]). Here we summarize the main findings.

In the early stage of the viral life cycle HIV-1 PR is not active. It is capable to perform its activity only during the maturation phase. Premature activation leads to the production of uninfective immature virion particles[59], thus the regulation of the enzyme activity appears to be important for the correct development of new viral particles.

The enzyme is highly selective and is able to recognize sequences of 6 to 8 aminoacids[6]. 14 of such sequences have been identified in the Gag-Pol polypeptide chain[38](Figure 2). HIV-1 PR itself contains 3 splicing sites, thus it is subject to autolysis[60].

HIV-1 PR is relatively slower than all the other isoenzymes. Typical kinetic constants for the enzymatic activity, measured in vitro, are $K_m \approx 100$ µM and $k_{cat} \approx 20$ s^{-1} [61]; but these values vary considerably depending on the nature of the substrate and ionic strength[42-44].

1.2.2. Enzyme structure.

Due to its crucial importance as target for anti-AIDS therapies, the free enzyme and the complexes with inhibitors have been the subject of a large number of X-ray crystallography[33;35;36;40;62-64] and nuclear magnetic resonance (NMR)[47-52] studies aimed at elucidating the 3D structure.

HIV-1 PR is a homodimeric aspartic protease [6;9;53]composed of two identical subunits of 99 amino acids each. The sequence of residues is commonly

---

[†] In the early stages of the viral life-cycle the viral genome is translated into very long chains of amino acids. These chains assemble into immature particles that are able to escape from the host cell membrane, but lack the enzymatic functions required to infect a new cell[179;180].

numbered from 1 (N-terminus) to 99 (C-terminus) in subunit 1 and from 1' to 99' in subunit 2.
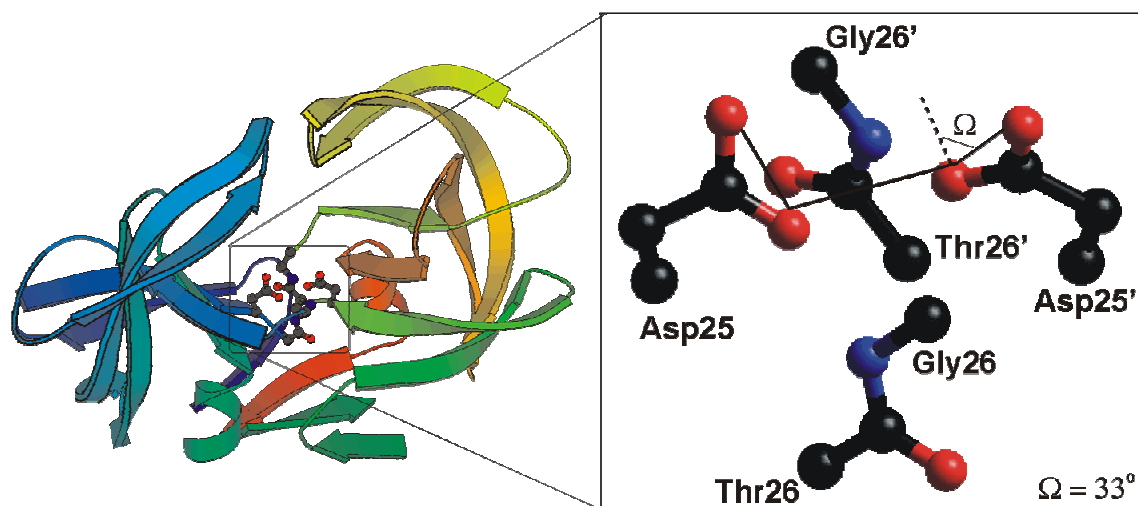
Figure 3. The active site of HIV-1 PR.

The active site of HIV-1 PR is located at the interface between the subunits. According to the numbering in the text the cleavage site is composed by Asp25(25')-Thr26(26')-Gly27(27').

The two subunits are related by a C_2 symmetry axis.

As in all the other aspartic proteases known so far, the active site is characterized by a pair of Aspartate(Asp)-Threoneine(Thr)-Glycine(Gly) sequences[65]. The cleavage site is located at the interface between the two subunits; each subunit contributing with one triplet of residues. Only the two Asps (Asp dyad) are directly involved in the chemical catalysis. Figure 3 shows that these groups are located very close to each other and almost coplanar. This peculiar conformation of the active site Asp dyad is a distinctive feature of aspartic proteases and is crucial for enzymatic activity and drug binding[6].
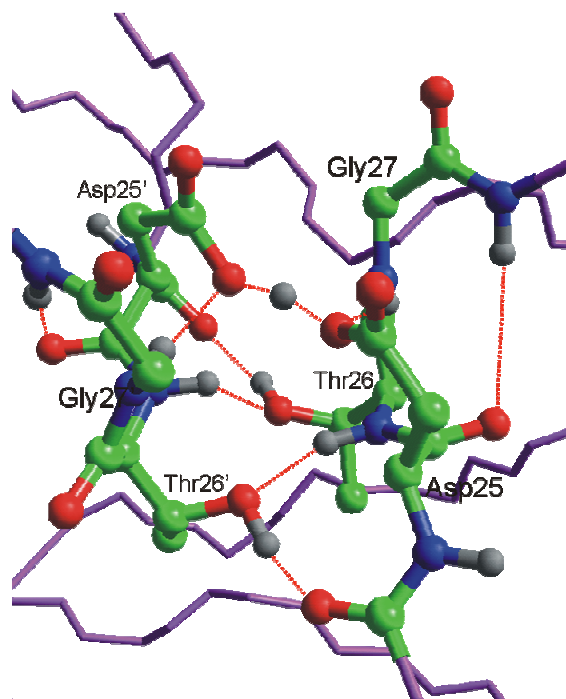
Figure 4. The "fireman's grip"[66].

H-bonds are represented as dashed lines.

The highly conserved Thr26(26') and Gly27(27') residues[65] do not participate directly in the catalysis. The crystal structures of all aspartic proteases reveal that they are involved in a network of hydrogen bonds surrounding the active site (forming the so called *fireman's grip*)[66] thus rendering this region rather rigid (Figure 4). This hydrogen bond network is characteristic of aspartic proteases and essential for enzymatic activity[41]. In the free enzyme the active site is located at the bottom of a large water-filled cavity. When the enzyme is bound to a ligand (namely substrates or inhibitors molecules) the cavity is almost completely occupied by the ligand atoms (Figure 5). The Gibbs free energy of inhibitor binding is dominated by entropy changes[56], a distinctive feature of a binding process driven by hydrophobic interactions. Indeed most of the residues surrounding the active site are hydrophobic.

On the opposite side of the Asp dyad, two large flaps delimit the active site. The flaps are flexible[10;21;49;67;68] and experience large rearrangements after inhibitor binding[21;35;49;55;57]; these rearrangements have large effects on the energetics

and on the protease stability[69] as the flaps contribute to most of the binding energy for the inhibitors[70].
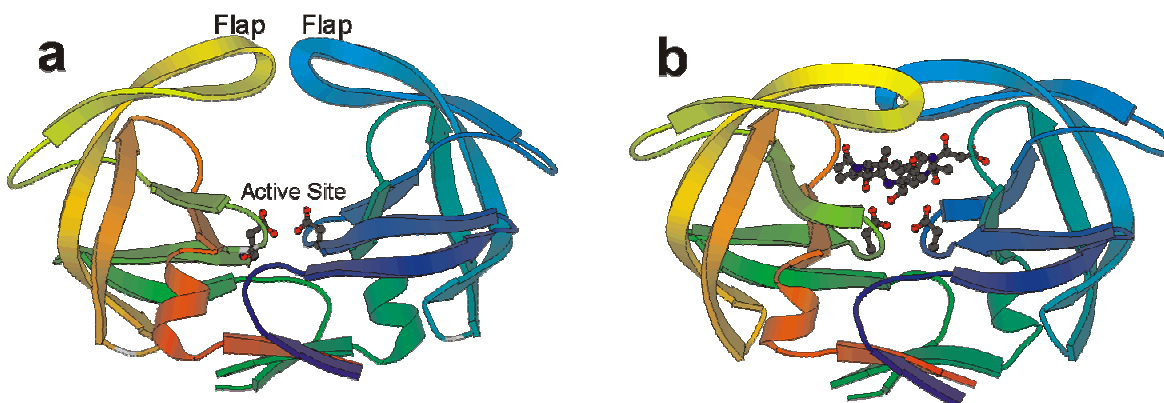
Figure 5. Binding of substrates and inhibitors to HIV-1 PR.

a) Free enzyme[66]. **b**) Enzyme/pepstatin complex[71]. The ligand molecules are bound in the large cavity above the active site. Beside the large rearrangements observed in the flaps that close up on the ligand molecule, the overall structure of the rest of the protease and of the active site remains unaltered upon ligand binding.

1.2.3. Reaction mechanism.

The reaction mechanism has not yet been completely elucidated.

Several reaction pathways have been proposed for aspartic proteases[65]. In the most accepted mechanism (Figure 6)[27;28;43;44;72] the free enzyme (**E**) binds the substrate to form an enzyme-substrate complex (**ES**). Subsequently a water molecule, polarized by the Asp groups, attacks the carbonyl carbon. One proton is donated from the Asp dyad to the substrate while another proton is transferred from the water molecule to the Asp dyad (**TS1**) to form the reaction intermediate **INT**. Subsequently, the substrate nitrogen atom is protonated while a proton is released from the intermediate to the Asp dyad (**TS2**). As a consequence the reaction intermediate decomposes to give the reaction products (**EP**) that are released to yield the free enzyme (**E**) again.

The rate-limiting step of the reaction depends on the chemical structure of the substrate[42] and can be either the formation (**TS1**, Figure 6)[27] or the decay (**TS2**, Figure 6)[44] of the reaction intermediate **INT**.
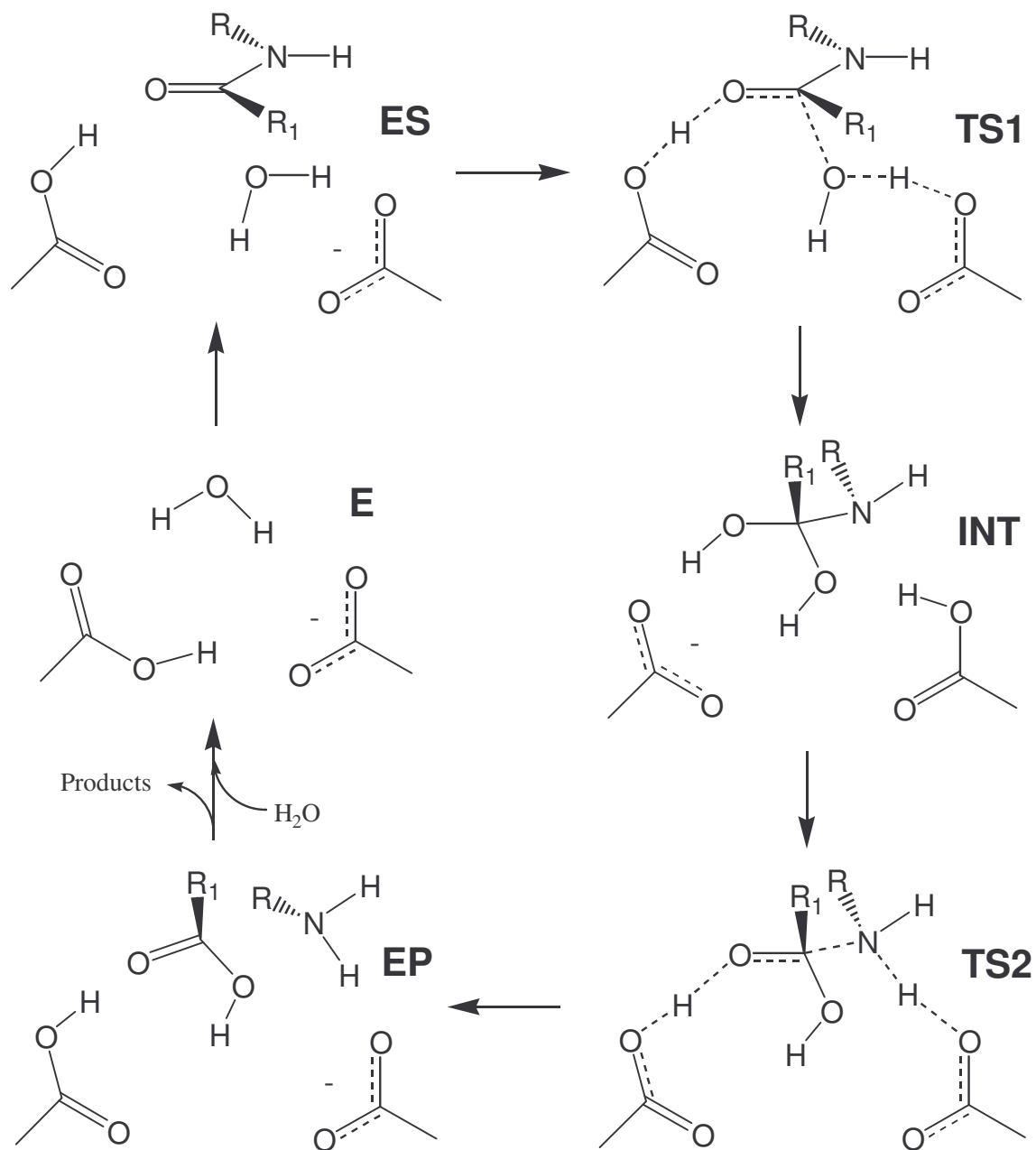
Figure 6. Most accepted mechanism for HIV-1 PR[43].

In TS1 and TS2 a negative charge is spread across the substrate and the Asp groups.

1.2.4. Drug design.

Since the discovery that inhibition of the protease activity with pepstatin leads to the production of immature, non-infectious viral particles[5;7;73;74] HIV-1 PR has become a fundamental target for the computer aided design[‡] of anti-AIDS drugs[57;75-80]. Four drugs are now routinely used in anti AIDS treatments[75].

As the enzyme is expected to strongly interact with transition-state mimics[81], most of the inhibitors are designed such as to resemble closely either the transition states or the reaction intermediate (**TS1**, **TS2**, **INT** Figure 6)[75].

The interaction between inhibitors in the active site cavity (Figure 3) and the cleavage site depends significantly on the Asp dyad protonation state[75]. The latter may vary depending on the type of inhibitor. Unfortunately protons are invisible to the X-ray experiments and their location is not usually determined in the X-ray studies. The exact charge state of carboxylic groups in proteins is then determined by ^{13}C NMR spectroscopy. Chemical shifts on ^{13}C-enriched HIV-1 PR have been measured in the free enzyme[31] and in inhibitor/enzyme complexes[31;32;45;47-52]. However, the interpretation of the signals has proved to be challenging [47;50] and in some cases the experimental data could not unequivocally determine the number and location of the protons in the active site of the enzyme.

1.2.4.1. Drug resistance.

Despite the relative successes obtained, the development of drugs against HIV-1 PR is continuously frustrated by the fast development of mutations that confer to the virus resistance against a particular pharmacological agent. Several of these mutations are known and have been studied (Figure 7).

[‡] Drug design is based on simulated docking experiments. In these simulations the interaction between HIV-1 PR and the functional groups of a putative drug is modeled with an effective potential. Within this potential, a search is performed for the compounds that minimize the enzyme/drug interaction energy.
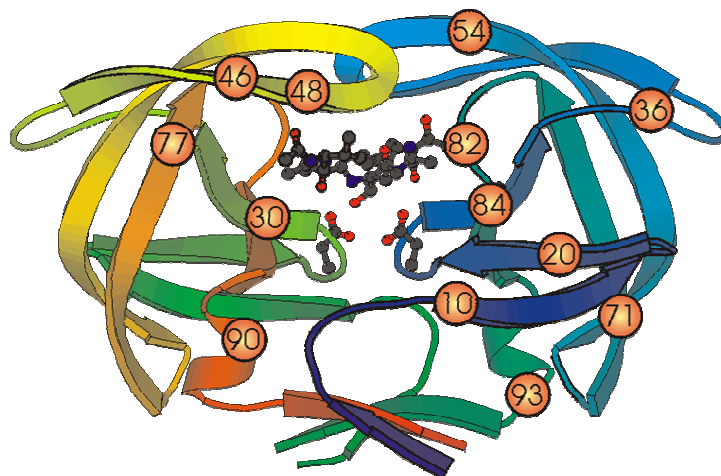
Figure 7. HIV-1 PR Mutants.
Positions of the most important mutations observed in vivo that result in viral particles resistant to the four clinically approved HIV-1 PR inhibitors[82]. For clarity the mutations are indicated on only one of the two subunits.

Most of the characterized mutations belongs to residues close to the active site and directly disrupt favorable enzyme-drug interactions. Nevertheless there are mutations that affect residues far from the active site (Figure 7). The mechanisms that lead to drug-resistance in these cases are often not known.

1.3. Motivation and outline of present work.

Due to its importance in modern pharmacology, HIV-1 PR is one of the most studied enzymes. Force field-based calculations have provided great advance in understanding structure/function relationships in the protein. However, it was soon realized that the force field approach encounters problems towards correctly describing the molecular interactions at the cleavage site[25] (the Asp dyad[24])[26]. As a result, geometric constraints[26] or *ad hoc* charge distributions[25] were used to reproduce the coplanarity and close proximity of the Asp dyad. These *a posteriori* models could reproduce the correct structural properties, yet they do not provide the physical origin of the stability of the Asp dyad.

Car Parrinello ab initio molecular dynamics simulations[83] appears to be a promising technique to investigate the physical origin of the stability of the active site. In its implementation with BLYP exchange-correlation functional, DFT allows a reliable description of relatively large systems. Furthermore, temperature effects, of outmost importance for biological function[84], are fully included.

As only a portion of the protein can be treated by ab initio methods, an estimation of the effect of the protein environment on the cleavage site has been the first step of our investigation. It turns out that the electric field is small and almost constant in the active site (see Appendix 1). This is consistent with the fact that the region around the active site is mostly hydrophobic. Therefore we expect that ab initio calculations on models of the active site capture most of the chemistry of the Asp dyad.

We have therefore carried out an ab initio Car-Parrinello molecular dynamics study on models of increasing complexity of the cleavage site. Calculations focused on the Asp-Thr-Gly triad and water in the active site. Our simulations turn out to provide the physical origin of the Asp dyad stability and allow identifying a low-barrier hydrogen bond and a dipole-charge interaction as a key factors in determining the cleavage site conformation observed experimentally. Analysis of the electronic structure of the cleavage site provides a rationale for the observed discrepancy in the experimental measurements of the active site protonation state[31;44].

Having become confident that our model complex describe accurately the conformational flexibility of the aspartyl dyad, we have subsequently addressed two issues of importance in the design of new inhibitors:

1. <u>Asp dyad protonation state</u>. Determination of the Asp dyad protonation state, which governs drug/HIV-1 PR interactions, is crucial for the design of new drugs. The method of choice to determine the protonation state is ^{13}C NMR; however the interpretation of the ^{13}C NMR data has proven to be challenging[31]. We have investigated the protonation and H-bond pattern of the HIV-1 PR/pepstatin complex. For this system, a recent ^{13}C NMR experiment[31] on the Asp25-Asp25' dyad measured two separate resonance

lines. Based on this result the authors suggested that the active site is singly protonated. We address this issue by performing *ab-initio* molecular dynamics calculations on a model for this site accompanied by calculations of the NMR chemical shifts. We find that already on the picosecond time-scale the model proposed[31] is not stable and evolves towards a structure whose NMR pattern differs from the experimental one. We suggest instead a different and novel protonation state in which both aspartic groups are protonated. In spite of the symmetric protonation state, the calculated ^{13}C NMR shift exhibits two lines with a split in good agreement with experiments. We rationalize this result with a simple valence bond model, which explains the chemical inequality of the two C sites.

2. <u>The transition state</u>. Inhibitors are designed such as to resemble the transition state[81]. We have investigated by classical and ab initio molecular dynamics aspects of the reaction mechanism to obtain structural information on the transition state and the relevance of the substrate fluctuations for the enzymatic function. In turns out that during the MD simulations the substrate experiences significant thermal fluctuations correlated to global protein motions on the ns timescale. The ab initio calculations show that these fluctuations are very important for the enzymatic reaction.

2. <u>METHODS</u>

2.1. The Schröedinger equation.

In the non relativistic case, the energy of a stationary state of given number of M atoms and N electrons is given by the time independent Schröedinger equation

$$H\Psi = E\Psi$$

where the Hamiltonian operator H is

$$H = v_{ext} + T + V_{ee}$$

Atomic units are assumed in the entire chapter. The external potential v_{ext} is determined by the nuclear configuration

$$v_{ext} = -\sum_{i=1}^{N}\sum_{I=1}^{M}\frac{Z_A}{\left|\vec{R}_I - \vec{r}_i\right|}$$

T is the kinetic energy operator

$$T = -\tfrac{1}{2}\sum_{i=1}^{N}\nabla_i^2$$

V_{ee} is the electron-electron repulsion operator

$$V_{ee} = \sum_{i=1}^{N-1}\sum_{i<j}^{N}\frac{1}{r_{ij}}$$

and Ψ is the electronic wavefunction

$$\Psi = \Psi(\vec{r}_1, \vec{r}_2, ... \vec{r}_N)$$

Ψ is in principle also function of the spin coordinates (s_1, s_2..., s_N). As in this work we will always deal with closed-shell systems, we will assume double occupancy of any electronic energy level and will neglect the electron spin dependence in the Hamiltonian.

2.2. Density functional theory.

The Density functional theory (DFT) approach to the solution of the quantum problem is based on the Hohemberg-Kohn theorems.

2.2.1. The Hohemberg-Kohn Theorems.

In their first theorem[85], Hohemberg and Kohn demonstrated that, in the ground state, the external potential v_{ext} is determined within an additive constant, by the electron density

$$\rho(\vec{r}_1) = N \int d\vec{r}_2 d\vec{r}_3 \dots d\vec{r}_N \left| \Psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N) \right|^2$$

Where $\vec{r}_i$ is any of the N spatial variables of Ψ.

As v_{ext} determines univocally H it follows that the ground state energy and all the observables are a functional of the electron density $\rho(\vec{r})$.

Their second theorem[85] states that, in the ground state, for any trial density $\tilde{\rho}(\vec{r})$ such that $\tilde{\rho}(\vec{r}) \geq 0$ and $\int \tilde{\rho}(\vec{r}) d\vec{r} = N$

$$E_0 \leq E_v[\tilde{\rho}] \text{ where } E_v[\rho] = T[\rho] + V_{ne}[\rho] + V_{ee}[\rho]$$

$T[\rho]$ is the electron kinetic energy $T = \sum_{i=1}^{N} n_i \langle \psi_i | -\tfrac{1}{2}\nabla^2 | \psi_i \rangle$ where n_i is the occupation number of each orbital; V_{ee} is the electron-electron interaction that includes the classical electrostatic electron-electron repulsion $J[\rho] = \int\int d\vec{r}d\vec{r}' \dfrac{\rho(\vec{r})\rho(\vec{r}')}{|\vec{r}-\vec{r}'|}$ and a non-classical part.

Thus, the energy can be expressed as a functional of the electron density and the variational principle can be used to find the ground state density for any given nuclear configuration.

The total energy of the electronic ground state of a system with external potential *v(r)*, written in terms of electron density $\rho(\vec{r})$ reads

$$E_v[\rho] = \int \rho(\vec{r})v(\vec{r})d\vec{r} + F_{HK}[\rho]$$

$$F_{HK}[\rho] = T[\rho] + V_{ee}[\rho]$$

According to the variational principle, with the constraint $\int \tilde{\rho}(\vec{r})d\vec{r} = N$, the ground state density satisfies the stationary principle

$$\delta\{E_v[\rho] - \mu[\int \rho(\vec{r})d\vec{r} - N]\} = 0$$

giving the Euler-Lagrange equation

$$\mu = v(\vec{r}) + \frac{\partial F_{HK}[\rho]}{\partial \rho(\vec{r})}$$

where μ is the Lagrange multiplier associated with the constraint

$$\int \rho(\vec{r})d\vec{r} = N$$

2.2.2. The Kohn-Sham equations.

A direct approach to the determination of the ground state energy of a system is not possible as the functional form of $F_{HK}[\rho]$ is not known. Kohn and Sham introduced an indirect approach[86] to the problem mapping the system of interest on a system of non interacting electrons.

$F_{HK}[\rho]$ can be written as

$$F_{HK}[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho]$$

$T_s[\rho]$ is thereafter the kinetic energy of a system of N non-interacting electrons that occupy N/2 electronic states and thus reads

$$T_s[\rho] = 2\sum_{i=1}^{\frac{N}{2}} \langle \psi_i | -\tfrac{1}{2}\nabla^2 | \psi_i \rangle$$

$E_{xc}[\rho]$ contains the difference between the kinetic energy of the real system and the kinetic energy of the non interacting electrons, and the non-classical part of the electron-electron interaction $V_{ee}[\rho]$.

The Euler equation then reads

$$\mu = v(\vec{r}) + \frac{\partial F_{HK}[\rho]}{\partial \rho(\vec{r})} = v(\vec{r}) + \frac{\partial T_s[\rho]}{\partial \rho(\vec{r})} + \frac{\partial J[\rho]}{\partial \rho(\vec{r})} + \frac{\partial E_{xc}[\rho]}{\partial \rho(\vec{r})}$$

Let us introduce the effective potential v_{eff} defined as

$$v_{eff}(\vec{r}) = v(\vec{r}) + \frac{\partial J[\rho]}{\partial \rho(\vec{r})} + \frac{\partial E_{xc}[\rho]}{\partial \rho(\vec{r})} = v(\vec{r}) + \int \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} + v_{xc}(\vec{r})$$

with $v_{xc}(\vec{r}) = \dfrac{\partial E_{xc}[\rho]}{\partial \rho(\vec{r})}$

Then the density of the new system, composed of non interacting electrons moving in the field v_{eff}, corresponds to the density of the original system and is calculated as the sum of the one electron contributions

$$\rho(\vec{r}) = \sum_{i=1}^{N} |\psi_i(\vec{r})|^2$$

The corresponding total energy is obtained as

$$E = \sum_{i=1}^{N} \varepsilon_i - \frac{1}{2} \int \frac{\rho(\vec{r})\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} dr dr' + E_{xc}[\rho] - \int v_{xc} \rho(\vec{r}) d\vec{r}$$

where ε_i are "orbital energies" of the non interacting reference system. According to the variational principle, imposing $\dfrac{\delta E}{\delta \psi_i} = 0$ the ground state can be obtained solving the Kohn-Sham equations

$$\left[-\tfrac{1}{2}\nabla^2 + v_{eff}(\vec{r}) \right]\psi_i = \varepsilon_i \psi_i$$

These equations are non-linear and are to be solved self-consistently, as v_{eff} depends upon the total electron density of the system.

The Kohn-Sham equations are in principle exact. Unfortunately the functional form of $E_{xc}[\rho]$ is not known and one has to resort to approximated forms.

2.2.3. Exchange-correlation functionals.

A successful approximation for the exchange-correlation energy is based on the results obtained for the uniform electron gas applied to small portions of atomic

system[86]. Within this approximation (Local Density Approximation) the exchange correlation energy reads

$$E_{xc}^{LDA}[\rho] = -\frac{3}{2}\left(\frac{3}{4\pi}\right)^{1/3}\int \rho^{4/3}(\vec{r})d\vec{r}$$

Local density approximation (LDA) calculations, although successful in the description of many systems, exhibit some deficiencies when studying hydrogen bonded systems like those of interest in biological systems. To improve the results of the calculations on those systems it is necessary to go beyond LDA and explicitly include a dependence upon the gradient of the electron density in the exchange-correlation functional. This correction is called general gradient approximation[87;88;88-92].

The gradient-corrected Becke exchange[87] and Lee, Yang, Parr correlation[89] (BLYP) functional has shown to describe accurately structural and dynamical properties of liquid water and biological molecules[93;94]. Thus it was used in this work.[95-99]

2.2.3.1.   Becke exchange functional.

The Becke exchange functional[100] was introduced to reproduce the exact asintotic behavior of the exchange energy density U_x

$$\lim_{r\to\infty} U_x = -\frac{1}{2}$$

and reads

$$E_x[\rho] = E_{xc}^{LDA}[\rho] - \beta\int \rho^{4/3}(\vec{r})\frac{x^2}{\left(1+6\beta\sinh^{-1}x\right)}d\vec{r}$$

where

$$E_{xc}^{LDA}[\rho] = -\frac{3}{2}\left(\frac{3}{4\pi}\right)^{1/3}\int \rho^{4/3}(\vec{r})d\vec{r}$$

$$x = \frac{|\nabla\rho(\vec{r})|}{\rho^{4/3}(\vec{r})}$$

The parameter β is determined by a fit on the exact HF data and was suggested by the author to be 0.0042 au.

2.2.3.2.   Lee Yang Parr correlation functional.

The LYP correlation energy functional[89] was derived from the Colle-Salvetti formula[101] to calculate the correlation energies from the HF second order density matrix. It reads

$$E_c[\rho] = -a \int \frac{1}{1+d\rho^{-1/3}(\vec{r})} \left\{ \rho(\vec{r}) + b\rho^{-2/3}(\vec{r}) \left[C_F \rho^{5/3}(\vec{r}) - 2t_\omega + \left(\tfrac{1}{9} t_\omega + \tfrac{1}{12} \nabla^2 \rho(\vec{r}) \right) e^{-c\rho^{-1/3}(\vec{r})} \right] \right\} d\vec{r}$$

and for the correlation potential

$$\frac{\delta E_c[\rho]}{\delta\rho(\vec{r})} = -a\left(\dot{F}_1\rho(\vec{r}) + F_1\right) - abC_F\rho^{5/3}(\vec{r})\left(\dot{G}_1\rho(\vec{r}) + \tfrac{8}{3}G_1\right) +$$

$$-\frac{ab}{4}\left[\ddot{G}_1\rho(\vec{r})|\nabla\rho(\vec{r})|^2 + \dot{G}_1\left(3|\nabla\rho(\vec{r})|^2 + 2\rho(\vec{r})\nabla^2\rho(\vec{r})\right) + 4G_1\nabla^2\rho(\vec{r})\right] + C_F = \tfrac{3}{10}(3\pi^2)^{2/3}$$

$$-\frac{ab}{72}\left[3\ddot{G}_1\rho(\vec{r})|\nabla\rho(\vec{r})|^2 + \dot{G}_1\left(5|\nabla\rho(\vec{r})|^2 + 6\rho(\vec{r})\nabla^2\rho(\vec{r})\right) + 4G_1\nabla^2\rho(\vec{r})\right]$$

$$F_1 = \frac{1}{1+d\rho^{-1/3}(\vec{r})}$$

$$G_1 = F_1\rho^{-5/3}(\vec{r})e^{-c\rho^{-1/3}}$$

$\dot{F}_1, \dot{G}_1, \ddot{G}_1$ are the derivatives of F_1 and G_1 with respect to the density; *a=0.04918, b=0.132, c=0.2533* and *d=0.349* are parameters derived by fitting the HF results for the helium atom.

2.2.4. Plane waves basis set.

Our basis set was expanded in plane waves

$$\psi_i(\vec{r}) = \frac{1}{\sqrt{V}} \sum_{\vec{g}} C_i(\vec{g}) e^{i\vec{g}\cdot\vec{r}}$$

where V is the volume of the system, the sum runs over all the reciprocal lattice vectors $\vec{g}$ and $C_i(\vec{g})$ are the coefficients of the plane wave expansion.

in an orthorhombic cell the wavevectors $\vec{g}$ are defined by

$$\vec{g} = 2\pi\left(\frac{K_x}{L_x}, \frac{K_y}{L_y}, \frac{K_z}{L_z}\right)$$

where K_{xyz} can be any integer number.

The size of the basis set can be truncated up to a spherical cutoff on the momentum $\vec{g}$

$C_i(\vec{g}) = 0$ when $|\vec{g}| > g_{max}$

where g_{max} satisfies the following equations

$$\frac{N_x}{L_x} = \frac{N_y}{L_y} = \frac{N_z}{L_z} = \frac{g_{max}}{\pi}$$

with $-\dfrac{N_{xyz}}{2} \leq K_{xyz} \leq \dfrac{N_{xyz}}{2}$

Large cutoff values correspond to a large basis set and accurate (and expensive) calculations. Thus the optimal cutoff value is a tradeoff between accuracy and computational cost.

The solution of the KS equations within this basis set is an eigenvalue equation of dimension N$_{pw}$ equal to the size of the basis set employed

$$\sum_{\vec{g}'}^{N_{pw}} H_{KS}^{pw}(\vec{g}', \vec{g}) C_i(\vec{g}') = \varepsilon_i C_i(\vec{g})$$

where $N_{pw} = N_x N_y N_z$ and $H_{KS}^{pw}(\vec{g}', \vec{g})$ read

$$H_{KS}^{pw}(\vec{g}', \vec{g}) = \frac{1}{2}\vec{g}^2 + \frac{4\pi}{V}\frac{|\tilde{\rho}(\vec{g}-\vec{g}')|}{|\vec{g}-\vec{g}'|^2} + \tilde{v}_{xc}(\vec{g}-\vec{g}') + \sum_M e^{-i(\vec{g}-\vec{g}')R_M}\tilde{V}_{Ion}^{PP}(\vec{g})$$

$$\tilde{\rho}(\vec{g}) = \frac{1}{V}\int \rho(\vec{r})e^{-i\vec{g}\vec{r}}d\vec{r}$$

$$\tilde{v}_{xc}(\vec{g}) = \frac{1}{V}\int v_{xc}(\vec{r})e^{-i\vec{g}\vec{r}}d\vec{r}$$

$$\tilde{V}_{Ion}^{PP}(\vec{g}) = \frac{1}{V}\int V_{Ion}^{PP}(\vec{r})e^{-i\vec{g}\vec{r}}d\vec{r}$$

v_{xc} is the exchange-correlation potential; V_{Ion}^{PP} is the potential corresponding to the ion-electron interaction, described as pseudopotentials.

2.2.4.1.   Isolated system calculations.

The molecules were treated as isolated following the procedure of Barnett and Landmann[102;103]. To enforce non-periodic boundary conditions the one electron wavefunctions were expanded in plane waves within the cell

$$\psi_i(\vec{r}) = \frac{1}{\sqrt{V}} \sum_{\vec{g}} C_i(\vec{g}) e^{i\vec{g}\vec{r}} \quad \text{for r inside the cell}$$

$$\psi_i(\vec{r}) = 0 \quad \text{for r outside the cell}$$

with the boundary condition that both $\psi_i(\vec{r})$ and $\nabla \psi_i(\vec{r})$ are negligibly small at the cell borders.

2.2.5. Pseudopotentials for the electron ion interactions.

The plane-waves expansions is inefficient to describe core electron wave functions. Indeed the latter exhibit sharp spatial oscillations around the nuclei and a very large basis set is required for their proper description. On the other hand, the core region of the atoms does not participate in the chemical bonding and remains essentially unaffected by changes in the surrounding chemical environment. Thus the core electrons can be assumed to be frozen and one can treat explicitly only the valence electrons. The interactions of the valence electrons with the nuclei and the core electrons are described by pseudopotentials[104;105].

Several pseudopotential types have been proposed (for a review see [106]). Here we used the ab initio pseudopotentials of Martins-Troulliers (MT)[107] type.

Martins Troulliers pseudopotentials are of the *smooth norm-conserving* type: i) they contain no nodes; ii) beyond a chosen cutoff radius r_{cl} the normalized atomic radial pseudo-wave function R_l^{PP} is equal to the normalized radial all electron wave function R_l^{AE}

$$R_l^{PP}(r) = R_l^{AE}(r) \quad \text{for } r > r_{cl}$$

iii) the charge enclosed within r_{cl} is equal to the all electron charge;

$$\int_0^{r_{cl}} \left| R_l^{PP}(r) \right|^2 r^2 dr = \int_0^{r_{cl}} \left| R_l^{AE}(r) \right|^2 r^2 dr$$

iv) the valence pseudopotential eigenvalues are equal to the valence all electron eigenvalues;

$$\varepsilon_l^{PP} = \varepsilon_l^{AE}$$

v) at $r = r_{cl}$ the pseudo wavefunction and its first four derivatives should be continuous;

vi) the pseudo potential should have zero curvature at the origin.

The radial part has the general form

$$R_l^{PP}(r) = \begin{cases} R_l^{AE}(r); r > r_{cl} \\ r^l e^{p(r)}; r \leq r_{cl} \end{cases}$$

$$p(r) = c_0 + c_2 r^2 + c_4 r^4 + c_6 r^6 + c_8 r^8 + c_{10} r^{10} + c_{12} r^{12}$$

where the seven coefficient of the polynomial *p(r)* are obtained by imposing iii), v) and vi). The local and non local parts of the pseudopotential can be fully separated using the Kleinman-Bylander approach[108].

2.3. Car-Parrinello ab initio molecular dynamics.

Ab initio molecular dynamics (AIMD) simulations were performed following the Car-Parrinello approach[83].

The energy of an electronic system $\langle \Psi_0 | H_e | \Psi_0 \rangle$ is a function of the nuclear positions $\vec{R}_{Ion}$; but, at the same time, it can be considered as a functional of the electronic wavefunction Ψ_0, that is, of the one particle Kohn-Sham orbitals ψ_i.

In the Car-Parrinello approach the electronic degrees of freedom are treated as classical particles of mass μ. As the forces on the nuclei are obtained by a derivative of the Lagrangian with respect to the nuclear positions; a functional derivative with respect to the one particle orbitals gives the "forces" on the orbitals.

The Lagrangian proposed by Car and Parrinello to describe this motion is

$$L = \sum_i \tfrac{1}{2}\mu \int \left| \dot{\psi}_i(\vec{r}) \right|^2 d\vec{r} + \sum_{Ion} \tfrac{1}{2} M \dot{\vec{R}}_{Ion} - E\left[\{\psi_i\}, \{\vec{R}_{Ion}\} \right] + \sum_{ij} \lambda_{ij} \left(\left\langle \psi_i(\vec{r}) \middle| \psi_j(\vec{r}) \right\rangle - \delta_{ij} \right)$$

With λ_{ij} are the Lagrange multipliers associated with the constraint necessary to ensure orthonormality of the one electron wavefunctions during the dynamics

$$\int \psi_i(\vec{r}) \psi_j^*(\vec{r}) d\vec{r} = \delta_{ij}$$

The corresponding equations of motion for both the electronic and the ionic degrees of freedom can be obtained from the Euler-Lagrange equations

$$\frac{\delta}{\delta t} \frac{\delta L}{\delta \dot{\vec{R}}_{Ion}} = \frac{\delta L}{\delta \vec{R}_{Ion}}$$

$$\frac{\delta}{\delta t} \frac{\delta L}{\delta \dot{\psi}_i^*} = \frac{\delta L}{\delta \psi_i^*}$$

$$\begin{cases} \mu \ddot{\psi}_i(\vec{r}) = -H_{KS} \psi_i(\vec{r}) + \sum_j \lambda_{ij} \psi_j(\vec{r}) \\ M_{Ion} \ddot{\vec{R}}_{Ion} = -\dfrac{dE\left[\{\psi_i\}, \{\vec{r}_i\} \right]}{d\vec{R}_{Ion}} \end{cases}$$

and the orthonormality constraint is enforced at each step by a RATTLE algorithm[109].

In insulators (such as the system object of the present study), the value of μ is chosen such that $\mu \ll M_{Ion}$.

In this case the frequency spectra of ions ω_{Ion} and electrons ω_i are well separated as

$$\omega_i \propto \sqrt{\frac{1}{\mu}} \text{ and } \omega_{Ion} \propto \sqrt{\frac{K_{Ion}}{M_{Ion}}}$$

where K_{Ion} is the kinetic energy of the nuclei.

Under these conditions there is no energy transfer between the electronic and the ionic degrees of freedom and the dynamic is adiabatic.

2.3.1. Integration of the equations of motion.

A Velocity Verlet[110] algorithm was used to integrate the equations of motion over discrete time intervals δt. In this algorithm the position at time $t+\delta t$ is first obtained

$$\vec{r}(t + \delta t) = \vec{r}(t) + \delta t \dot{\vec{r}}(t) + \tfrac{1}{2} \delta t^2 \ddot{\vec{r}}(t)$$

and the velocities are calculated as

$$\dot{\vec{r}}(t + \delta t) = \dot{\vec{r}}(t) + \tfrac{1}{2} \delta t \left[\ddot{\vec{r}}(t) + \ddot{\vec{r}}(t + \delta t) \right]$$

2.3.2. The Nosé thermostat.

Constant temperature MD can be obtained by coupling the ionic degrees of freedom to a Nosé thermal bath[111] with kinetic energy

$$K_{Nose} = \tfrac{1}{2} Q \dot{s}^2$$

where Q is a parameter and s is the additional degree of freedom coming from the thermal bath. s is associated with an extra potential energy term that reads

$$V_{Nose} = (F + 1) k_B T \ln s$$

where F is the total number of the degrees of freedom of the system. The Lagrangian that describes the ionic motions coupled to the thermal bath is

$$L = K + K_{Nose} - V - V_{Nose}$$

for the equations of motion

$$\ddot{\vec{r}}_i = \frac{1}{m_i s^2} \vec{f}_i - 2 \frac{\dot{s}\dot{\vec{r}}_i}{s}$$

$$Q\ddot{s} = \sum_i m \dot{\vec{r}}_i^2 s - (F + 1) k_B \frac{T}{s}$$

Q can be chosen arbitrarily and its value determines the strength of the coupling; high values of Q result in a very low coupling and vice versa.

2.4. Calculated quantum mechanical properties.

2.4.1. Maximally localized Wannier functions as tool to visualize atomic orbitals.

To visualize chemical concepts as bonds or lone pairs in molecules and solid state it is convenient to make an unitary transformation from the one particle Kohn-Sham eigenstates to a set of localized molecular orbitals (maximally localized Wannier functions)[112;113]. The positions of the centers of the Wannier functions allow an accurate description of polarization effects and to calculate dipole moments for complex systems.

Such a transformation can be performed according to Marzari and Vanderbilt[114;115]. The Maximally localized Wannier functions in cell R associated with the eigenstate *i* are defined as the Bloch functions

$$w(\vec{r} - \vec{R}) = \frac{V}{(2\pi)^3} \int e^{-i\vec{k}\vec{R}} \left| \psi_{i\vec{k}} \right\rangle d\vec{k}$$

Those minimize the spreading functional

$$\Omega = \sum_{i=1}^{N/2} \left[\left\langle \vec{r}_i^2 \right\rangle - \bar{r}_i^2 \right]$$

$$\bar{r}_i = \left\langle \vec{r} \right\rangle_i$$

Note that, for periodic systems, the position expectation value $\left\langle \vec{r} \right\rangle$ can be defined as in ref.[116]

$$\left\langle \vec{r} \right\rangle = \frac{L}{2\pi} \mathrm{Im} \left\langle \Psi^{(0)} \left| e^{i\frac{2\pi}{L}X} \right| \Psi^{(0)} \right\rangle$$

where X is the position operator and L is the cell dimension.

2.4.1.1. Dipole moments.

The resulting Wannier centers $\bar{r}_i$ provide a straightforward way to calculate the dipole moment of a molecule in a complex system[117] as

$$\vec{\mu} = 2e \sum_{i=1}^{N/2} \vec{r}_i + e \sum_I Z_I \vec{R}_I$$

Also the position of the Wannier center from the atomic nuclei along a chemical bond can be taken as an estimation of the bond polarity[99] and used to compare the polarization of the atoms in different chemical environments[118].

2.4.2. ^{13}C NMR chemical shifts.

Nuclear Magnetic Resonance experiments measure the resonance frequencies of the nuclear spins of a sample in a uniform magnetic field $\vec{B}_{ext}$. The resonance frequency of each nucleus is expressed in terms of *chemical shifts* σ with respect to the resonance frequency of a reference compound. Chemical shifts are defined as the difference between the nuclear spin resonance frequencies of atoms of the same type in different chemical environments.

The resonance frequency depends upon the strength of the local magnetic field $\vec{B}_{loc}(\vec{r})$ experienced by each nucleus. Thus the chemical shifts are measures of differences in the local field $\vec{B}_{loc}$ at the nuclear positions $\vec{R}$.

$$\vec{B}_{loc}(\vec{R}) = \vec{B}_{ext} + \vec{B}_{ind}(\vec{R}) = B_{ext} \left[1 - \overline{\sigma}(\vec{R})\right]$$

where $\vec{B}_{ind}(\vec{R}) = \vec{B}_{ext} \overline{\sigma}(\vec{R})$ is the field induced by the applied uniform magnetic field $\vec{B}_{ext}$ and $\vec{\sigma}(\vec{R})$ is the shielding tensor. In the isotropic case

$$\sigma = \frac{1}{3} Tr\left[\vec{\sigma}(\vec{R})\right]$$

σ is dimensionless, usually very small $\approx 10^{-6}$-10^{-4} and measured in part per million (ppm).

$\vec{B}_{ind}(\vec{r})$ is generated by $\vec{J}_{ind}(\vec{r})$; that is the electronic current density generated by the electrons as linear response to $\vec{B}_{ext}$ ($\vec{J}_{ind}(\vec{r}) \propto \vec{B}_{ext}$).

Thus $\vec{B}_{ind}$ can be calculated from $\vec{J}_{ind}(\vec{r})$ as

$$\vec{B}_{ind}(\vec{r}) = \frac{1}{c}\int d\vec{r}\vec{J}_{ind}(\vec{r}') \wedge \frac{(\vec{r}-\vec{r}')}{|\vec{r}-\vec{r}'|^3}$$

The interaction between the electrons and a magnetic field in a non-magnetic insulator is described by the non-relativistic time-independent Hamiltonian

$$H = \frac{1}{2m}\left[\vec{p}+\frac{e}{c}\vec{A}(\vec{r})\right]^2 + V(\vec{r})$$

where $\vec{p}=-i\hbar\nabla$ and $\vec{A}(\vec{r})$ is the vector potential associated to the magnetic field $\vec{B}(\vec{r})$

$$\vec{B}(\vec{r}) = \nabla \wedge \vec{A}(\vec{r})$$

The induced current density $\vec{J}^{ind}(\vec{r})$ is defined as

$$\vec{J}^{ind}(\vec{r}) = \frac{e\hbar}{2mi}\left(\Psi^*(\vec{r})\nabla\Psi(\vec{r})-\Psi(\vec{r})\nabla\Psi^*(\vec{r})\right)-\frac{e^2}{m}\vec{A}(\vec{r})\Psi^*(\vec{r})\Psi(\vec{r})$$

For low magnetic field the first order linear correction to the ground state wavefunction is

$$|\Psi\rangle = |\Psi^{(0)}\rangle + B|\Psi^{(1)}\rangle$$

$\Psi^{(0)}$ is real and the perturbation $H^{(1)}=\frac{e}{2}\vec{p}\cdot\vec{A}$ is purely imaginary; thus $\Psi^{(1)}$ is purely imaginary and we can then rewrite the induced electronic current density as

$$\vec{J}^{ind}(\vec{r}) = -\frac{e}{m}\left\{\left\langle\Psi^{(0)}\left|\left[|\vec{r}'\rangle\langle\vec{r}'|\vec{p}+\vec{p}|\vec{r}'\rangle\langle\vec{r}'|\right]\right|\Psi^{(1)}\right\rangle\right\}-\frac{e^2}{m}\left\langle\Psi^{(0)}\left|\vec{A}(\vec{r}')\right|\Psi^{(0)}\right\rangle+O(B^2)$$

The induced current density $\vec{J}^{ind}(\vec{r})$ is composed by two terms of opposite sign

$$\vec{J}^{ind}(\vec{r}) = \vec{J}^{para}(\vec{r}) + \vec{J}^{dia}(\vec{r})$$

$$\vec{J}^{para}(\vec{r}) = -\frac{e}{m}\left\{\left\langle\Psi^{(0)}\left|\left[|\vec{r}'\rangle\langle\vec{r}'|\vec{p}+\vec{p}|\vec{r}'\rangle\langle\vec{r}'|\right]\right|\Psi^{(1)}\right\rangle\right\}$$

$$\vec{J}^{dia}(\vec{r}) = -\frac{e^2}{m}\left\langle\Psi^{(0)}\left|\vec{A}(\vec{r}')\right|\Psi^{(0)}\right\rangle$$

In a uniform magnetic field it is particularly convenient to choose for the vector potential $\vec{A}(\vec{r})$ the form

$$\vec{A}(\vec{r}) = -\frac{1}{2}(\vec{r}-\vec{d})\wedge B$$

where $\vec{d}$ is the gauge origin.

The corresponding Hamiltonian reads

$$H = \tfrac{1}{2m}\vec{p}^2 + \tfrac{e}{2mc}B(\vec{r}-\vec{d})\wedge\vec{p} + \tfrac{e}{8mc^2}\left\|B\wedge(\vec{r}-\vec{d})\right\|^2 + V(\vec{r})$$

and the induced current density, in terms of orbital contributions, can be computed as

$$\vec{J}^{ind}(\vec{r}) = \sum_i \vec{J}_i^{ind}(\vec{r}) = \sum_i \vec{J}_i^{para}(\vec{r}) + \vec{J}_i^{dia}(\vec{r})$$

$$\vec{J}_i^{para}(\vec{r}) = -\tfrac{e}{m}\sum_l \left\langle \psi_i^{(0)} \left| \left[\vec{p}|\vec{r}\rangle\langle\vec{r}| + |\vec{r}\rangle\langle\vec{r}|\vec{p} \right] \frac{\left|\psi_l^{(0)}\right\rangle\left\langle\psi_l^{(0)}\right|}{\varepsilon_i - \varepsilon_l} \left[(\vec{r}-\vec{d})\times\vec{p}\right]B \right| \psi_i^{(0)} \right\rangle$$

$$J_i^{dia}(\vec{r}) = -\tfrac{e^2}{m}\rho_i(\vec{r})B\times(\vec{r}-\vec{d})$$

Where the sum over l runs over the unoccupied orbitals.

The choice of the gauge origin $\vec{d}$ is arbitrary. The total current is independent of the gauge origin chosen, but the two terms are not and they individually diverge for $\left|\vec{r}-\vec{d}\right|\to\infty$. In particular $\vec{J}^{para}(\vec{r})$ converges slowly with the size of the basis set in points far from the gauge origin $\vec{d}$. This would not be a concern if an infinite basis set was used, but for finite basis sets the numerical calculations are not exactly gauge invariant and require very large basis sets for evaluating the current far from the gauge origin.

As $\vec{J}^{ind}(\vec{r})$ is the result of the difference between two large numbers, its calculation is numerically slow to converge. To overcome this problem several methods, that employ different gauge origins, were proposed (For a review see[119;120]).

In our calculations the "d(r)=r" version of the Continuous Set of Gauge Transformations (CSGT) method[121] was used. In this method a gauge transformation with $\vec{d}=\vec{r}$ is performed for each point where $\vec{J}^{ind}(\vec{r})$ is evaluated. In this way the diamagnetic contribution to the current density vanishes analytically everywhere and only the paramagnetic part should be evaluated, thus $\vec{J}^{ind}(\vec{r})$ is not calculated as the difference of two large terms. Within this gauge transformation $\vec{J}^{ind}(\vec{r})$ reads

$$\vec{J}_i^{ind}(\vec{r}) = -\frac{e}{m}\sum_l \left\langle \psi_i^{(0)} \left| [\vec{p}|\vec{r}\rangle\langle\vec{r}| + |\vec{r}\rangle\langle\vec{r}|\vec{p}] \frac{\left|\psi_l^{(0)}\right\rangle\left\langle\psi_l^{(0)}\right|}{\varepsilon_i - \varepsilon_l} \left\{ [\vec{r}\times\vec{p}]B \left|\psi_i^{(0)}\right\rangle - \vec{p}\left|\psi_i^{(0)}\right\rangle[\vec{B}\times\vec{r}] \right\} \right.$$

To compute the electron density current it is necessary to calculate the first order correction to the ground state wavefunction that can be determined through a minimization of the second-order energy functional[122;123] which is variational in $\left|\Psi^{(1)}\right\rangle$

$$E^{(2)} = \sum_i \left\langle \psi_i^{(1)} \left| H^{(0)} - \varepsilon_i^{(0)} \right| \psi_i^{(1)} \right\rangle + \sum_i \left[\left\langle \psi_i^{(1)} \left| H^{(1)} \right| \psi_i^{(0)} \right\rangle + \left\langle \psi_i^{(0)} \left| H^{(1)} \right| \psi_i^{(1)} \right\rangle \right] +$$

$$+ \frac{1}{2}\int d\vec{r}\,d\vec{r}' \frac{\delta^2 E_{Hxc}\left[\rho^{(0)}\right]}{\delta\rho(\vec{r})\delta\rho(\vec{r}')} \rho^{(1)}(\vec{r})\rho^{(1)}(\vec{r}')$$

Where $E_{Hxc}\left[\rho^{(0)}\right]$ is the Hartree, exchange and correlation functional and $\varepsilon_i^{(0)}$ are the eigenvalues of the unperturbed Hamiltonian.

As the perturbation is a magnetic field, the first order correction to the ground state wavefunction is purely imaginary. Thus $\rho^{(1)}(\vec{r}) = 0$ and the last term of the second order energy functional vanishes analytically leaving

$$E^{(2)} = \sum_i \left\langle \psi_i^{(1)} \left| H^{(0)} - \varepsilon_i^{(0)} \right| \psi_i^{(1)} \right\rangle + \sum_i \left[\left\langle \psi_i^{(1)} \left| H^{(1)} \right| \psi_i^{(0)} \right\rangle + \left\langle \psi_i^{(0)} \left| H^{(1)} \right| \psi_i^{(1)} \right\rangle \right]$$

In our calculations the electronic problem was solved within DFT with a plane waves basis set and pseudopotentials to represent the core contributions to the all electron wave function. Thus only the contributions of the valence electrons to the induced first order quantum electric current were taken into account.

In this way three contributions were neglected: I) the diamagnetic core contributions that comes from the core electron density and are expected to be independent of the chemical environment; ii) the contributions to the paramagnetic term of the transitions from valence to core and from core to valence states; iii) the contribution due to the difference between the pseudopotential wave function and the all electron wave function. Nevertheless it has been shown for several atom types that these contributions determine an

additive constant to the calculated chemical shift tensors[119]. This constant turns out to be independent of the chemical environment.

In principle, the dependence of the exchange-correlation functional upon the electronic current should be explicitly taken into account as in Current Density Functional Theory (CDFT)[124]. In practice, the corrections to the chemical shift due to the use of CDFT functionals are negligible[125]; therefore this dependence is omitted in the chemical shift calculations.

The experimentally measured chemical shifts are bulk averages over all the possible phase space conformations explored by the system. To compare with the experimental data, in our calculations some picoseconds of Car-Parrinello MD were performed and the chemical shifts were calculated as the time average over a number of conformations taken at regular intervals from the MD trajectory.

2.4.3. Isotopic shift.

The isotopic substitution effect on the chemical shift, *isotopic shift* hereafter, is a small displacement (usually of the order of less than 1 ppm) of the nuclear magnetic resonance frequency of some nuclei that occurs when an atom of a chemical compound is substituted with one of its isotopes.

Such a substitution leaves the ground state electronic wavefunction unchanged. As it affects the distribution of masses of the molecule, it will induce a perturbation in the dynamical properties of the system. In particular, the vibrational levels of the molecule change. This leads to differences in the average bond lengths of bonds involving atoms whose masses have been modified by the isotopic substitution.

In general these effects are very small. When they can be measured, they provide useful information on the molecular properties of the systems. An important case is the substitution of hydrogen with deuterium[126;127]. In a chemical bond, the ground state vibrational frequency $\omega_0 \propto \mu_r^{-\frac{1}{2}}$ where μ_r is the reduced mass. When hydrogen is substituted with deuterium the reduced mass is almost doubled. This leads to a considerable difference in ω_0 for the two isotopes[128]. To

calculate ω_0 the potential energy along the A-H bond can be fitted to a Morse function[128]

$$E = D_e \left(1 - e^{-a(R-R_{eq})}\right)^2$$

for the Morse oscillator

$$E_v = \hbar\omega_e(v+\tfrac{1}{2}) - \chi_e\hbar\omega_e(v+\tfrac{1}{2})^2 + O(v^3)$$

$$\omega_e = a\left(\frac{2D_e}{\mu_r}\right)^{\frac{1}{2}}$$

$$\chi_e = \frac{\hbar\omega_e}{4D_e}$$

Assuming that only the first vibrational level is populated, the isotopic shift δ_σ is equal to

$$\delta_\sigma = \frac{d\sigma}{dr_{A-H}}\left(r_0^H - r_0^D\right)$$

where $\dfrac{d\sigma}{dr_{A-H}}$ was obtained from a linear fit of the values of σ around the equilibrium distance r_0^H.

The isotopic shifts calculated with this methodology are in good agreement with the experimental measurements[128]. It should be pointed out that two factors contributes in determining the magnitude of the isotopic shift; namely the difference between r_0^H and r_0^D, related to the anharmonicity of the stretching potential; and the susceptibility of the chemical shift to variation of the A-H bond. As $\omega_0 \propto \mu_r^{-\frac{1}{2}}$; $r_0^H - r_0^D$ is always negative, whereas $\dfrac{d\sigma}{dr_{A-H}}$ can have different signs and either direct or inverse isotopic shifts can be observed (see Figure 33, chapter 4.2.4).

If the substituted hydrogen atom is involved in an hydrogen bond, the shape of the Morse potential and, as a consequence, $r_0^H - r_0^D$, would depend strongly upon the hydrogen bonding strength[126;127;129;130]. Thus isotopic shift

measurements can provide a method to evaluate the relative strength of the hydrogen bonds of a system.

2.4.4. Vibrational spectrum.

The vibrational frequencies were calculated as the Fourier transform of the velocity-velocity autocorrelation function $C_{vv}(t)$ of the system

$$C_{vv}(t) = \frac{1}{t_{max}} \sum_{t=t_0}^{t_{max}} \vec{v}(t_0)\vec{v}(t_0 + t)$$

where t_{max} corresponds to the simulation time and $\vec{v}(t)$ are the velocity vectors. As we calculate bond stretching frequency we define $\vec{v}(t)$ by the projections on the atomic distance vectors.

2.4.5. Reaction free energies.

The reaction free energies were calculated by rigidly constraining a reaction coordinate[131] and calculating the integral of the average force f_s acting on the constraint along the reaction coordinate Q.

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

We define $W(Q') = -KT \ln\left[\dfrac{P(Q')}{P_u}\right]$ as the reversible work associated with the reaction coordinate Q' where P_u represents the uniform probability density of Q'. The mean force is defined as the derivative of the reversible work with respect to the constrained reaction coordinate

$$f_s(Q') = -\frac{dW(Q')}{dQ'}$$

where the prime indicates that the average force is calculated along the constrained MD simulation.

The relative free energy that connects two states can be calculated from the integral of the mean force along the reaction coordinate. In the general case the mean force is be obtained by proper reweigh of the average constraint force[131;132]. However, if the constrained coordinate is a simple interatomic distance $r_1 - r_0$, the mean force reduces to the average value of the Lagrange multiplier λ associated with the distance constraint[133] and the reaction free energy reads

$$\Delta F(r_1 - r_0) = \int_{r_0}^{r_1} \langle f_s(r) \rangle dr = \int_{r_0}^{r_1} \langle \lambda(r) \rangle dr$$

$\langle \lambda(r) \rangle$ is computed for a set of values of r and the integral is numerically evaluated along the constraint distance.

2.4.6. Atomic charges and bond orders.

The atomic integrated electron density was calculated according to the Hirshfeld method[134]. Mayer bond orders[135] were calculated by projecting the wavefunction into an atomic basis set.

2.5. Environment effects.

To calculate the effect of the protein electrostatic field on the active site, we calculated the protein electrostatic potential $V(\vec{r})$ as

$$V(\vec{r}) = \frac{1}{\varepsilon_0} \sum_{i=1}^{N} \frac{q_i}{|\vec{r} - \vec{r}_i|}$$

where q_I is the atomic charge of atom I and $\vec{r} - r_i$ is the distance from the grid points r. The values of the atomic charges are taken from the AMBER[136;137] force field (see next session). In our calculations, we assume a dielectric constant $\varepsilon_0 = 1$.

2.6. Force field-based molecular dynamics simulations.

2.6.1. The force field.

Force field-based calculations were performed using the AMBER 5.0 suite of program[137] and the AMBER 5.0 force field[136]. The corresponding potential energy function is

$$E = \frac{1}{2} \sum_{bonds} K_b (b - b_0)^2 +$$

$$+ \frac{1}{2} \sum_{angles} K_\theta (\theta - \theta_0)^2 +$$

$$+ \frac{1}{2} \sum_{dihedrals} \sum_n K_\phi^{(n)} \left[1 + \cos(n\phi - \delta) \right] +$$

$$+ \sum_{ij} \left(\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^6} + \frac{q_i q_j}{r_{ij}} \right)$$

The first and second term are harmonic approximations to the bond stretching and angle bending motions. The force constants K_b, K_θ and the equilibrium values b_0 and θ_0 are obtained by fittings over spectroscopy data and ab initio calculations. The third term describes the rotations around bonds as a periodic function with n minima each characterized by a force constant $K_\phi^{(n)}$.

The non bonded interactions are represented as an electrostatic and a Lennard-Johnes pair potentials where the *ij* sum runs over all the non bonded pairs of atoms. The Lennard-Johnes interactions are calculated only for $r_{ij} < 9\,\text{Å}$.

2.6.2. The Ewald sum method.

The systems were treated as periodic and the electrostatic interactions were calculated with the Ewald sum method[138-142].
Wihitin this method, to each charge in the system are added two charges distributions of equal magnitude. One (A) of the same sign and another (B) of opposite sign.These distributions are written as gaussian functions

$$\rho_i(r) = q_i \beta^3 e^{-\frac{\beta^2 r^2}{\pi^{3/2}}}$$

where β is an arbitrary parameter that determines the shape of the Gaussian distribution. The interaction between the original charges (O) and the charge distribution B (including the O-O, the O-B and the B-B terms) reads

$$V_{ij}^{OB} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} q_i q_j \frac{erfc(\beta r_{ij})}{r_{ij}}$$

it is short ranged and can be accurately calculated using a relatively short cutoff (~9Å).

The O-A interaction V_{ij}^{OA} is then calculated in the reciprocal space and the result transformed back in real space. The resulting energy term reads

$$V_{ij}^{OA} = \frac{1}{\pi V} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{\vec{n} \neq 0} \frac{q_i q_j}{\vec{n}^2} e^{-\frac{\pi^2 \vec{n}^2}{\beta^2}} e^{2\pi i \vec{n} \vec{r}_{ij}}$$

where $\vec{n}$ are the reciprocal space lattice vectors and V is the unit cell volume. A third term to remove the self-interaction of each original charge with its image in A has to be added. The complete expression for the potential energy is then

$$V_{ij}^{ELE} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} q_i q_j \frac{erfc(\beta r_{ij})}{r_{ij}} + \frac{1}{\pi V} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{\vec{n} \neq 0} \frac{q_i q_j}{\vec{n}^2} e^{-\frac{\pi^2 \vec{n}^2}{\beta^2}} e^{2\pi i \vec{n} \vec{r}_{ij}} - \frac{\beta}{\sqrt{\pi}} \sum_{i} q_i^2 + J(\vec{D})$$

where $J(\vec{D})$ is a further correction term that depends upon the macroscopic dipole moment $\vec{D}$ and the shape of the unit cell. The accuracy of the calculation can be improved by increasing the number of $\vec{n}$-vectors used in the reciprocal-space sum.

2.6.3. The Berendsen thermostat.

Constant temperature MD simulations were obtained by coupling to a Berendsen thermal bath[143]. At each step the velocities are scaled by a factor

$$\chi = \left(1 + \frac{\delta t}{\tau_T} \left(\frac{T}{T_0} - 1\right)\right)$$

where T_0 is the reference temperature and τ_T is a time constant that determines the strength of the coupling between the system and the thermal bath. A similar algorithm[143] can be used to obtain constant pressure by periodic scaling of the simulation cell size and atomic positions.

2.6.4. The SHAKE algorithm.

In the MD simulations the size of time step should be smaller than the motions characterized by the highest frequencies. In classical MD simulations these are typically the bond stretching motions involving hydrogen atoms. In order to use a larger time step all the bonds involving hydrogen atoms were constrained to their equilibrium position with the SHAKE algorithm[144]. In this algorithm at each step a correction $\vec{g}_a^{(r)}$, directed along the bond, is introduced in the forces such as to guarantee that the constraint is satisfied.

$$\vec{r}(t+\delta t) = \vec{r}'(t+\delta t) + \frac{\delta t^2}{m_a}\vec{g}_a^{(r)}$$

where $\vec{r}'(t+\delta t)$ is the position that the system would have reached in the absence of the constraint. The procedure is then applied on the next constraints. In a chain of atoms the correction applied on constraint i+1 would partially disrupt constraint i. For this reason the correction is applied cyclically on each constraint in the system, until the desired convergence is reached for all the constraints.

2.6.5. Calculated properties.

i) The root mean square deviation (rmsd) between two fitted structures of coordinates r_{i1} and r_{i2} is defined as

$$rmsd = \left(\sum_{i=1}^{N}\frac{(r_{i1}-r_{i2})^2}{n}\right)^{\frac{1}{2}}$$

The rmsds per residue are calculated as the MD average of the rmsd, calculated on each residue backbone atoms, with respect to the average MD structure.

ii) The dynamical cross correlation matrix (DCCM) of the $C\alpha$ atoms provides information on large-scale motion of the protein[67]. The matrix element C_{ij} of atoms i and j reads[67]

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

Where $\langle \rangle$ denotes a MD-averaged quantity and $\Delta \vec{r}_i(t) = \left\langle \vec{r}_i(t) \right\rangle - \vec{r}_i(t)$ the displacement from the MD average position $\left\langle \vec{r}_i(t) \right\rangle$ 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.

iii) Normal modes were calculated diagonalizing the covariance matrix[145]. The modes having the largest eigenvectors are associated to the largest atomic displacements[145].

3. <u>SYSTEMS AND COMPUTATIONAL DETAILS</u>

3.1. Model systems representing the active site of HIV-1 PR.

At present it is not possible to treat the entire protein in solution by ab initio methods On the other hand, the most relevant contributions to describe the chemistry at the cleavage site may arise from a few residues surrounding the catalytic Asp dyad itself (see 1.3 and Appendix 1[*]). Thus we built small model systems representing only the relevant functional groups and neglecting the rest of the protein. Calculations performed on these models can give a reliable picture of the active site of the enzyme. Nevertheless the results obtained depend upon how the model systems are built and should be carefully tested against the experimental data and calculations performed on models at different levels of sophistication.

In the "real" protein, the fragments included in the model systems are linked to the protein frame and not free to move. To reproduce this feature the terminal atoms of each fragment were constrained to their starting position. This corresponds to approximate the protein as a rigid frame. The approximation is valid until the protein is in fact rigid; i.e. it does not rearrange significantly on the ad initio MD timescale. In the active site of HIV-1 PR this assumption appears to be reasonable as in all the available crystal structures of HIV-1 PR complexed with different inhibitors the cleavage site region is almost identical (rmsd on the backbone atoms of Asp25(25')-Thr26(26')-Gly27(27') < 0.5). This is an indirect indication that its conformation should be rather rigid.

---

[*] Appendix 1 shows that, for HIV-1 PR, this is actually the case as the electrostatic field of the residues external to the active site is small and does not exhibit sharp spatial oscillations.

To summarize, there are two major approximations in our model system representation of the active site of HIV-1 PR:

i)   only a few of the thousands atoms that constitute HIV-1 PR are considered;

ii)   the protein is regarded as a rigid frame that surrounds the simulation cell.

The influence of these approximations on the results will be discussed in detail in the subsequent sections.

3.1.1. Free HIV-1 PR (PR).

Three models (**PR A**, **PR B** and **PR C** Figure 8) were built to account for the protonation states proposed for the free enzyme[31;44].
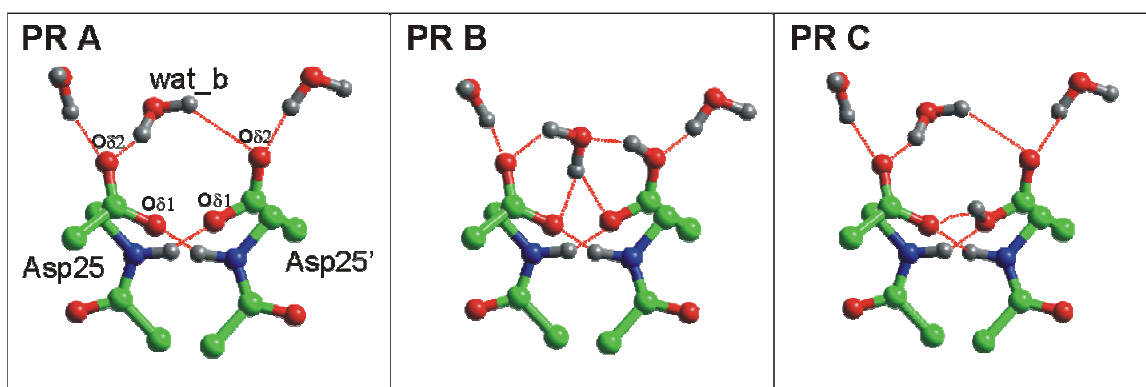

Figure 8. Model systems representing the free HIV-1 PR.
PR A: doubly deprotonated Asp dyad proposed on the basis of NMR measurements. **PR B**, **PR C**: possible patterns of the monoprotonated form of the enzyme. Non polar hydrogen atoms are not shown for clarity.

Each model included: Asp25(25') modeled as acetic acid/acetate molecules; the Thr26(26')-Gly27(27') peptide bonds, modeled as N-methyl acetamide and three water molecules that are present in the active site channel. The complexes were built starting from the structure of the free enzyme (3PHV entry[146] in the PDB database[147], 2.7 Å resolution). In model **PR A** both Asp groups were treated as deprotonated and the overall charge of the model was –2; in model **PR B** and **PR**

C, one aspartate was protonated on Oδ2 and Oδ1, respectively. The overall charge of these models was –1.

The Asp dyad is partially exposed to the solvent, thus several water molecules can in principle interact with it. One is known to bridge the two Asp groups (wat_b hereafter) and was positioned according to the protonation patterns already proposed for the eukariotic isoenzyme[148]. The other water molecules were not detected by X-ray crystallography[62;66]. The number and positions of the other water molecules were determined performing a force field-based molecular dynamics simulation on a shell of water molecules surrounding the free enzyme. The average number of water molecules hydrogen bonded to the Asp dyad was then calculated according to the following scheme: for each step of the MD simulation a water-Asp hydrogen bond was considered to be formed if $d\left(O\delta2_{Asp} - O_{Wat}\right) \leq 3.5\,\text{Å}$ and $\angle(O\delta2_{Asp} - H_{Wat} - O_{Wat}) \geq 60^0$. It turned out that, on average, 2.8 water molecules interact with the Asp dyad. Thus, three water molecules were added in the model system **PR A**, **PR B** and **PR C**.
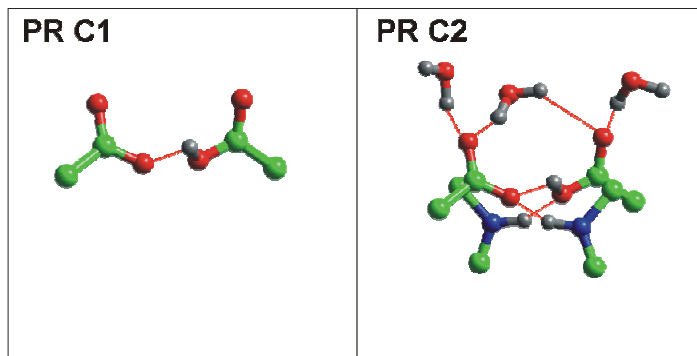

Figure 9. Additional model systems for protonation pattern C.

In **PR C1** only the two Asp groups are included; in **PR C2** the Thr-Gly peptide bond is modeled as dimethylamine. Non polar hydrogens are not shown for clarity.

For the protonation state **PR C** four additional models were built, namely **PR C1-4**. Complexes **PR C1** and **PR C2** (Figure 9) differ from **PR C** in the Thr26(26')-Gly27(27') moiety: in **PR C1** it was completely omitted and in **PR C2** it was modeled as dimetylamine. Complexes **PR C3** and **PR C4** (not shown) differ from

PR C in the number of water molecules present: only one in **PR C3** and two in **PR C4**.

3.1.2. HIV-1 PR/pepstatin complex (PEP).

The structure of pepstatin A/HIV-1 PR complex has not been solved yet. Instead the structure of the complex with acetyl pepstatin is known at 2.0 Å resolution[71] (5HVP entry in the PDB database[147]).

Acetyl pepstatin is chemically and structurally extremely similar to pepstatin A (Figure 9) and its portion binding to the active site is identical to that of pepstatin A. The initial configuration of pepstatin A/HIV-1 PR was therefore built from this X-ray structure.
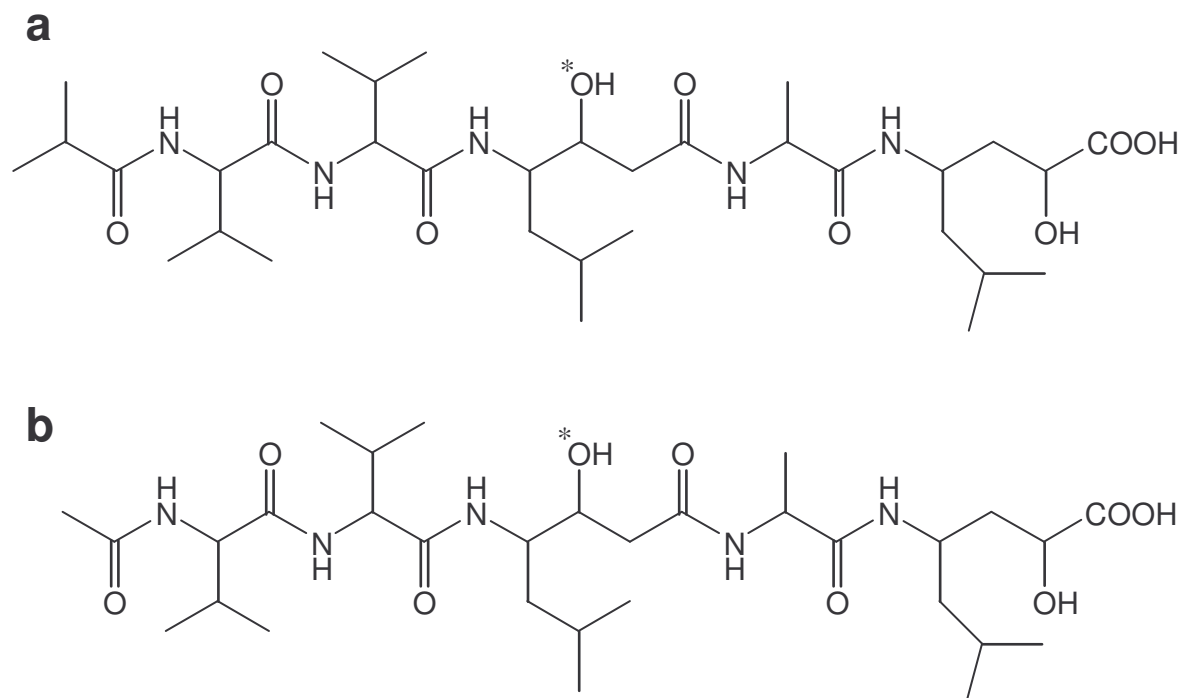

Figure 10. Structure of pepstatin A and Acetyl pepstatin.

Pepstatin A (a) and Acetyl pepstatin (b) differs by the group bound to the N-terminus of the chain: isopropionate for pepstatin and acetate for acetyl pepstatin. This class of inhibitors mimics the transition state of a putative peptide bond hydrolysis reaction and binds to the Asp dyad with the central hydroxyl group (marked with a star).

Our model complexes included the side-chains of Asp 25/Asp 25' (modeled as acetic acid and acetate, respectively) and groups interacting with them: i) Thr 26(26')-Gly 27(27') peptide bonds (modeled as N-methyl formamide); ii) the alcoholic moiety of acetyl pepstatin A (modeled as isopropanole). A model at a similar level of sophistication was used for the free enzyme[149].
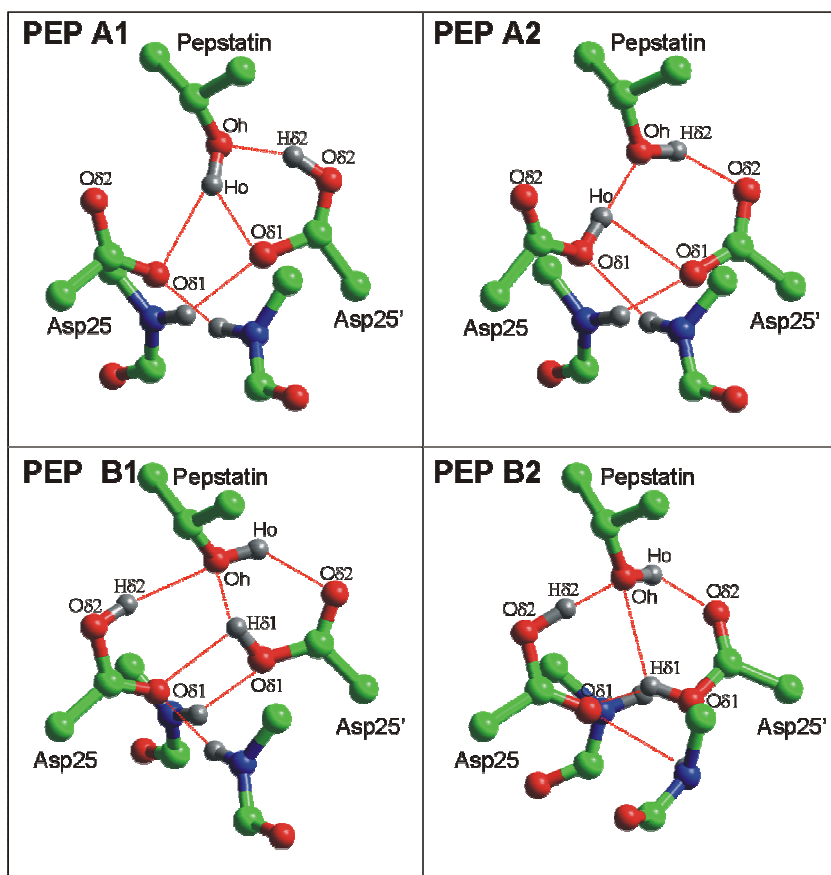

Figure 11. Model systems for the simulation of the HIV-1 PR/pepstatin complex.
PEP A1 and **PEP A2** correspond to the monoprotonated form of the enzyme at the beginning and at the end of the ab initio MD simulation, respectively. **PEP B1** and **PEP B2** represent the diprotonated enzyme at the beginning and at the end of the ab initio MD simulation. Non polar hydrogens not shown for clarity.

Two protonation states were considered (Figure 11): in **PEP A1**/**A2**, one of the two Asp side chains was protonated (overall net charge −1); in **PEP B1**, both Asp groups were protonated. **PEP A1** represents the protonation pattern of HIV-1

PR/pepstatin A proposed on the basis of ^{13}C NMR measurements[31]. The pattern **PEP B1** was proposed by us to account for the observed abnormal isotope shift. This pattern is present in other transition-state analog/protein complexes in which the inhibitor binds as pepstatin A to the catalytic Asp dyad through its hydroxyl group[50][†].

3.1.3. HIV-1 PR/substrate complex (SUB).

The model systems built to represent the active site of the HIV-1 PR/substrate complex included the side-chains of Asp 25(25') (modeled as propionic acid and propionate, respectively), the cleavage region of the substrate (modeled ad N-methyl acetamide), the catalytic water molecule (WatC) and the Thr 26(26')-Gly 27(27') peptide bonds (modeled as N-methyl formamide) (Figure 12).
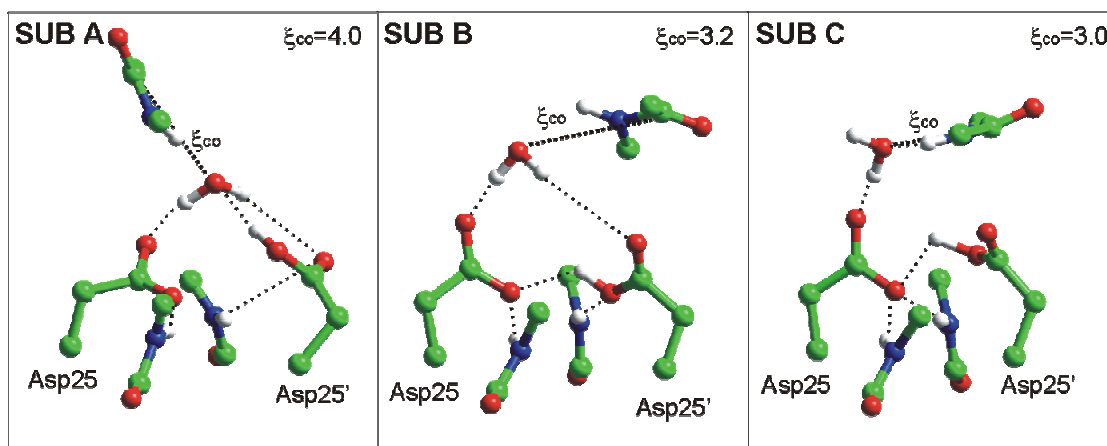

Figure 12. Model complexes for the simulation of the HIV-1 PR/Substrate complex.
For each complex the starting distance between the carbonyl carbon of the substrate and the oxygen of the water molecule (ξ_{CO}) is reported in the upper right corner. Non polar hydrogen atoms are not shown for clarity.

[†]   On the other hand, the fully deprotonated state has not been observed and is not considered in our calculations.

Three complexes (**SUB A**, **SUB B** and **SUB C** Figure 12) were built. Initial structures were taken from the force field-based calculations of the HIV-1 PR/substrate complex (see below). The main difference between the three models is in the substrate position: **SUB A** is the *ab initio* geometry-optimized structural model used also as initial model for the classical MD simulation (Model **SUB 3**, see also force field-based calculations); in **SUB B**, the substrate was located as in the classical MD-averaged structure; in **SUB C,** it was located as in the classical MD snapshot for which d_{SA} is the shortest (see RESULTS, force field-based calculations).

Ab initio MD simulations on the HIV-1 PR/Intermediate adduct (Figure 13) were performed starting from the last configuration of the constrained ab initio MD of model **SUB C**.
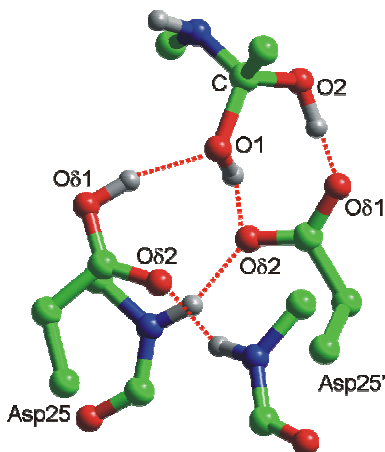

Figure 13. Model system for the simulation of the HIV-1 PR/Intermediate complex.
The structure of the reaction intermediate was built starting from the coordinates of the last step of the constrained ab initio molecular dynamics simulation of model **SUB C** and releasing the CO distance constraint. Non polar hydrogen atoms not shown for clarity.

3.1.4. Formic acid/water complexes.

Formic acid (**F 0**) and its conjugated base (**F 0-**) were constructed assuming standard bond lengths and bond angles (Figure 14).
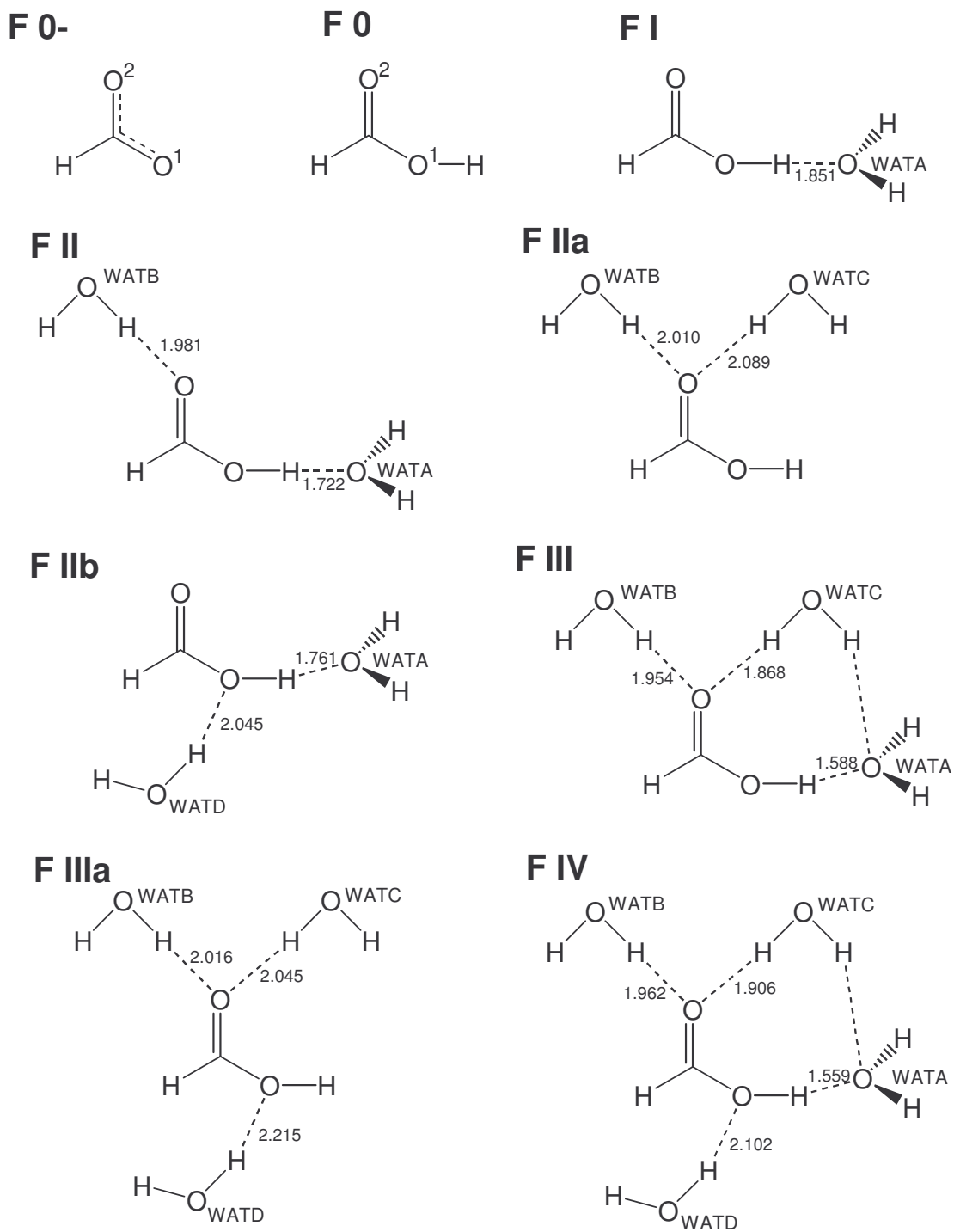
Figure 14. Formic acid/water model complexes.

The complexes are numerated according to the number of water molecules present. Hydrogen bonding distances are indicated in Å.

Water/formic acid complexes **F I-IV** (Figure 14) were built adding an increasing number of water molecules to **F 0[‡]**. **F IIa** (**IIc**) was the same as **F III** except that WATA(WATB) was removed; **F IIb** was the same as **F IV** except that WATB and WATC were removed.

3.1.5. Computational details.

Unless otherwise stated, all the DFT calculations were performed in orthorhombic cells. Martins Troulliers pseudopotentials and BLYP exchange correlation functionals were used. The one-electron wavefunctions were expanded in plane waves up to a cutoff of 70 ry.

Car Parrinello ab initio molecular dynamics simulations were performed using a fictitious electron mass of 400 au and a time step of 0.096 fs; a Nosé thermostat with a characteristic frequency of 500 cm^{-1} and a reference temperature of 300 K was applied to the ionic degrees of freedom.

3.1.5.1.   Free HIV-1 PR.

Geometry optimization were carried out on models **PR A, B, C** using the direct inversion of the iterative subspace method[150] up to a convergence of 5 10^{-4} au in the maximum gradient of the ionic forces.

0.9, 0.9, 0.9, 1.0, 2.5, 4.5 ps of Car Parrinello ab initio molecular dynamics simulations were performed on model **PR A**, **PR C1-4** and **PR C**, respectively.

3.1.5.2.   HIV-1 PR/pepstatin complex.

2.3 ps and 5.2 ps of Car-Parrinello ab initio molecular dynamics simulations were performed for complexes **PEP A1** and **PEP B1**, respectively.

[‡] Complexes are named according to the number of water models included in the calculations.

The chemical shifts[§] were calculated every 77.3 fs of ab initio molecular dynamics simulation. Since the configuration **PEP A1** (Figure 11) is unstable and evolves to **PEP A2**, the ^{13}C chemical shifts were calculated for **PEP A2** only, using the sampling interval of 77.3 fs. This corresponds to averaging over 20 configurations (Figure 30). In the case of **PEP B1**, 55 configurations were considered (Figure 31).

In order to investigate the dependence of our results on the choice of pseudopotentials and basis set, test calculations were also carried out on 10 configurations taken from **PEP B** with Goedecker pseudopotentials[151;152] and 100 Ry of cutoff. It turns out that the relative shifts change by as little as ~0.1 ppm.

3.1.5.3.   *Formic acid/water complexes.*

^{13}C NMR chemical shifts were calculated for all the formic acid/water complexes (**F 0-IV**, **IIa**, **IIb**, **IIc**) after full ab initio geometry optimization[150] up to a convergence of 5 10^{-4} on the largest component of the ionic forces. Deuterium isotope effects on the ^{13}C chemical shift were calculated for complex **F IIA**, **F IIB** and **F IIC.**

3.1.5.4.   *HIV-1 PR/substrate complex.*

Calculations of the free energy required for the nucleophylic attack of a water molecule (WATC) on the substrate peptide bond were carried out on **SUB A**, **SUB B**, **SUB C.** The $C_{(sub)}$-$O_{(WATC)}$ distance ξ_{co} (Figure 6) was chosen as the reaction coordinate for water nucleophilic attack to substrate carbonyl[26;132;133] and the activation free energy of the reaction was calculated as point integration of the average constraint force along ξ_{co}[153] (Table 4). Each constraint point was sampled until the force acting on the constraint was completely equilibrated. This usually took about 0.3 ps. On average each point was sampled for about 1 ps of

---

[§] We calculated the chemical shifts for all the nuclei of the system but focalized our attention on the NMR signal of the Asp γ carbon atoms as they are the most sensitive to differences in the chemical environment of the Asp dyad

ab initio MD simulation. The first 0.3 ps of dynamics of each point were discarded. Overall, ≈ 10 ps were sampled for each model.

The conformational flexibility of the intermediate (Figure 13) was studied by performing 1.4 ps of *ab initio* MD calculations with no constraint starting from the last configuration of model **SUB C**.

The dipole moment of WATC was calculated from the centers of the maximally localized Wannier orbitals in 6 snapshot of the *ab initio* MD simulation of **SUB B** (ξ_{CO}= 3.17 Å).

3.2. Force field-based calculations.

Force field-based calculations were used as a tool to determine the position of the hydrogen atoms in the X-ray structures of the free HIV-1 PR (Figure 5a)[66] and of the HIV-1 PR/pepstatin A complex (Figure 5b)[71]. The missing hydrogen atoms were added to the crystal structures with the PROTONATE module of AMBER 5.0 and their position refined with 1000 steps of conjugate gradient energy minimization keeping the heavy atoms fixed to their starting position.

3.2.1. HIV-1 PR/substrate complex.

The enzyme/substrate complex was constructed starting from the X-ray structure of a HIV-1 PR/peptide inhibitor complex (4HVP entry in the PDB database[147]). The inhibitor sequence reads: N-acetyl Thr-Ile-Nle-Ψ[CH_2NH]-Nle-Gln-Arg-amide[35] (Nle = Norleucine). Nle was replaced with Met and the CH_2NH group was changed in a peptide bond so as to transform the inhibitor in the N-acetyl

Thr-Ile-Met-Met-Gln-Arg.amide peptide substrate (SUB_PEP hereafter)[**]. (Figure 15).
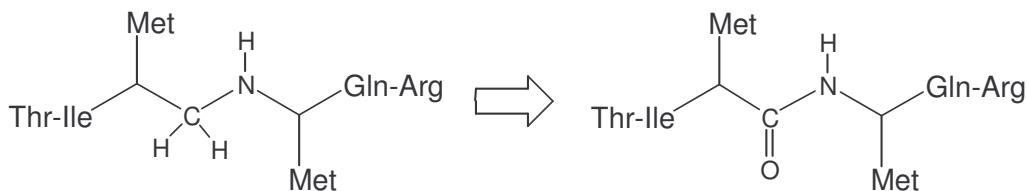

Figure 15. Chemical substitutions to obtain a substrate from an inhibitor of HIV-1 PR.

Substrate residues binding to subunit 1(2) were numbered from P1(P1') to P3(P3')[154]: Thr(P3)-Ile(P2)-Met(P1)-Met(P1')-Gln(P2')-Arg(P3'). Thus, the peptide bond cleaved by HIV-1 PR lies between residues P1 and P1' (Met-Met).

[**] Thr26' OH group is rotated towards a hydrophobic region and does not form the typical hydrogen bonding pattern observed in all X-ray structures of HIV-1 PR/inhibitors so far determined[35]. As stated by the crystallographers (A. Wlodawer, personal communication), the reason for the OH flipping is unknown[35]. In our structural models we rotated the side chain of Thr26' around the $C\alpha$-$C\beta$ bond such as to resort to the typical hydrogen bond network of HIV-1 PR and of Thr26.
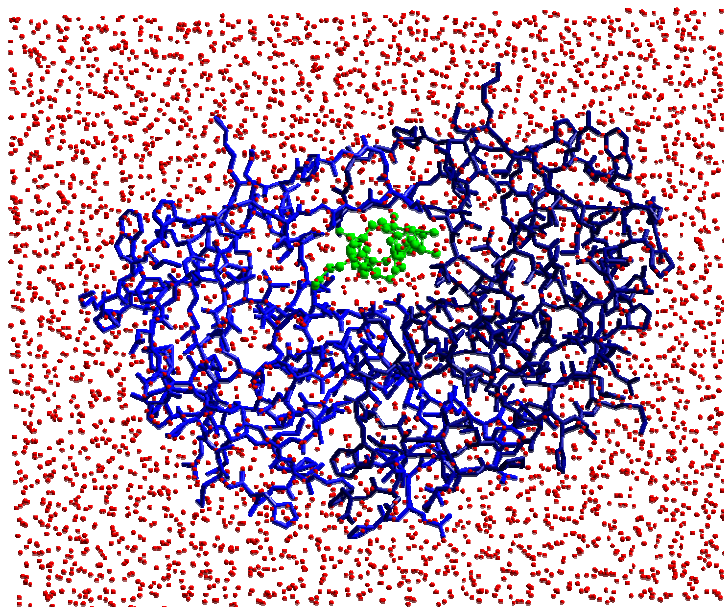
Figure 16. HIV-1 PR/substrate complex.

Simulation box used for the force field-based molecular dynamics simulations of the HIV-1 protease/substrate complex. The water molecules are represented as red spheres. The substrate is colored in green and rendered as ball and sticks. The protease subunits are colored in light and dark blue. Hydrogen atoms are not shown for clarity.

The overall positive charge of the HIV-1 PR-peptide complex (+5) was neutralized adding five chlorine ions close to five positively charged groups not involved in salt bridges in the crystal structure of the enzyme. The complex was immersed in a box containing 4170 water molecules. The starting size of the periodic simulation cell was 66.8 x 55.2 x 43.0 Å. The final system was composed of 15749 atoms (Figure 16).

Upon binding of PEP_SUB, the aspartyl dyad is monprotonated[44]. As proton location is unknown, four different protomers (**SUB 1**-**4** Figure 17) have to be considered: **SUB 1**, Asp25' Oδ2 protonated; **SUB 2**, Asp25 Oδ2 protonated; **SUB 3** Asp 25' Oδ1 protonated and **SUB 4**, Asp25 Oδ1 protonated.

In the four complexes, the catalytic water molecule bridges the two aspartic groups. Three hydrogen bonds are formed when Oδ1 is protonated; only two when Oδ2 is protonated (Figure 17).
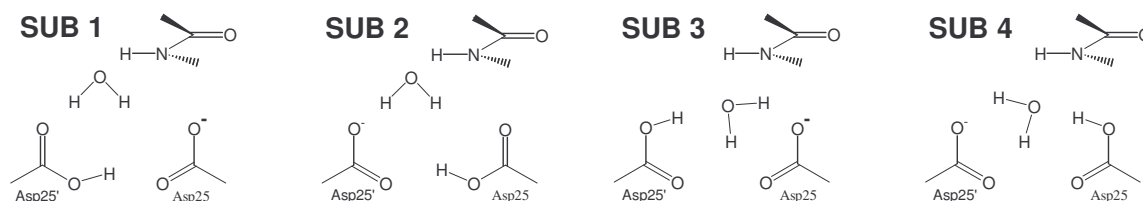
Figure 17. Protonation patterns investigated for the HIV-1 PR/substrate complex.

The most stable protomer was determined according to the following procedure. Force-field based geometry optimizations were performed on the four protomers of the HIV-1 PR/PEP_SUB complex in water solution underwent. Then, the geometry of the cleavage site was further geometry-optimized *ab initio* using the direct inversion in the iterative subspace method[155] with a convergence criteria of $5 \cdot 10^{-4}$ for the largest component of the atomic forces.

The energetically most stable protomer turns out to be **SUB 3** (in which Asp 25' $O\delta 1$ is protonated) which is 2 Kcal/mol lower in energy than **SUB 2**.

The HIV-1 PR/PEP_SUB complex structure corresponding to this protomer was heated to 150K for 15 ps and finally equilibrated with 720 ps of constant pressure MD at 300K. During the equilibration time the catalytic Aspartates were restrained to their starting positions with a harmonic restraint (1.0 Kcal Mol^{-1} Å^{-1}). Also the hydrogen bonds between $HN_{(Gly27)}$ and $O_{(MetP1)}$ that are conserved in all the peptido-mimetic inhibitors of HIV-1 PR were restrained with a harmonic constraint of 0.5 Kcal Mol^{-1} Å^{-1} centered at 2.3 Å. After the equilibration, the constraints were released and 2.1 ns of constant pressure MD were performed.

4. <u>RESULTS AND DISCUSSION</u>

4.1. Conformational flexibility of the Asp dyad in the free enzyme.

4.1.1. Protonation state.

The first step towards a characterization of the molecular interactions that stabilize the active site conformation in HIV-1 PR is the determination of the protonation state of the Asp dyad. At a pH optimal for the enzymatic activity (~5-6), the Asp dyad could in principle have three different protonation states: completely deprotonated (total charge –2), monoprotonated (total charge –1) or completely protonated (total charge 0). As two oxygen atoms on each aspartate could be in turn protonated, the three protonation state give rise to 9 possible protonation patterns (Figure 18)
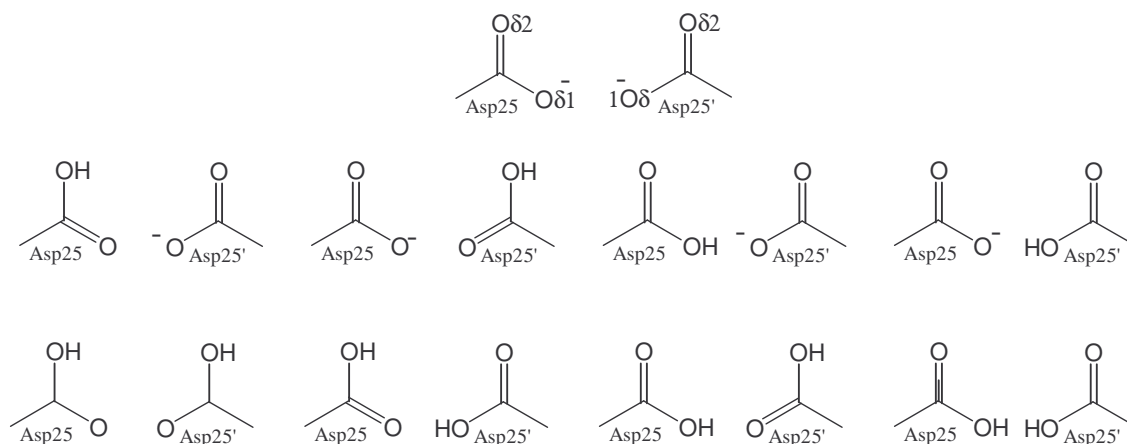

Figure 18. Possible protonation states and protonation patterns in free HIV-1 PR.

There are both experimental[44] and theoretical[16] indications that the Asp dyad is monoprotonated at pH 5-6. On the other hand, on the basis of ^{13}C NMR studies[31] it has been suggested that the enzyme could be completely deprotonated at pH

5-6; however this study has been subjected to severe criticism and is not conclusive.

The doubly protonated form has never been proposed so far and was not considered in the present study.

Because of the active site symmetry, the two aspartates are equivalent; thus three protonation patterns were studied: **PR A**, representing the completely deprotonated form; **PR B** and **PR C** corresponding to the monoprotonated form of the enzyme (Figure 19)
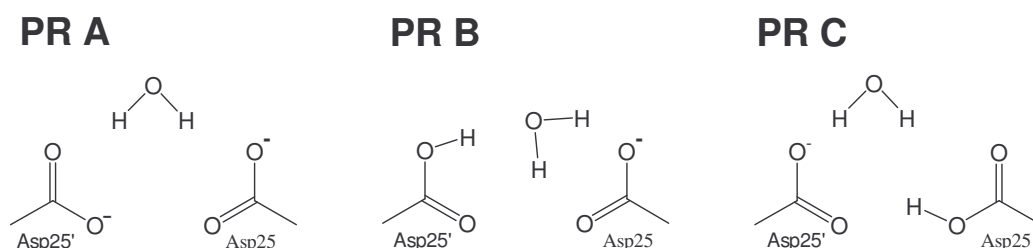

Figure 19 Protonation patterns considered in the studies on the free HIV-1 PR

4.1.1.1.   *Structure and dynamics of the deprotonated specie*

Geometry optimization and 0.9 ps of ab initio molecular dynamics simulation were performed on model **PR A**. We note that, already during the geometry optimization, the two negatively charged aspartates rearranges to an highly distorted conformation (rmsd with respect to the crystallographic structure 0.5 Å). Also the position of the bridging water molecule in the geometry-optimized structure is different from the electron density observed in the X-ray crystallography experiments (Figure 20)[62].

During the ab initio molecular dynamics even a larger rearrangement is observed as the two Asp groups move away from each other (rmsd after the simulation 1.16 Å). These are clear indications that the local interactions are not able to counterbalance the strong repulsion of the two negatively charged groups (~30 kcal/mol from point charge estimation of the electrostatic repulsion energy).

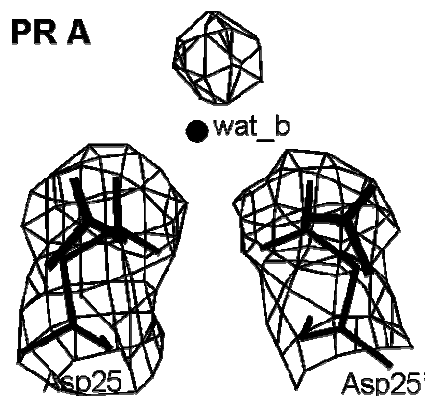We conclude that the doubly deprotonated form is not stable.

Figure 20. Geometry optimization of the doubly deprotonated free HIV-1 PR.

Starting (thin line) and final (bold line) configuration of the Asp dyad in the ab initio geometry optimization of model PR A. The experimentally measured electron density isosurface[62] is shown for comparison.

Inspection of the X-ray structures show that there are no polar or charged groups around the active site that could provide additional stabilization to the Asp pair in the doubly deprotonated state. Thus we expect that these results would not change by using larger models of the active site.

4.1.1.2.   Protonation pattern in the monoprotonated form.

High level ab initio calculations[148] have established in the Eukariotic Aspartic Proteases (E-PR) the proton is located on Oδ2 and wat_b forms three hydrogen bonds with the Asp dyad as in model **PR B** (Figure 8).

However there are many structural differences between the active site of HIV-1 PR and those E-PR as: i) in E-PR the Asp dyad is tightly hydrogen bonded to a Serine hydroxyl group, this group is missing in HIV-1 PR; ii) in HIV-1 the Asp dyad is closer (d$_{(O\delta2-O\delta2)}$ ~4.8Å instead of ~5.8Å) and less coplanar (Ω~33^0 instead of 8^0); iii) the electron density maps of the active site of HIV-1 PR and E-PR are suggestive of a different location of wat_b[66].

It is then possible that the differences in the active site characteristics of E-PR and HIV-1 PR would in result in a different protonation pattern for the two enzymes.
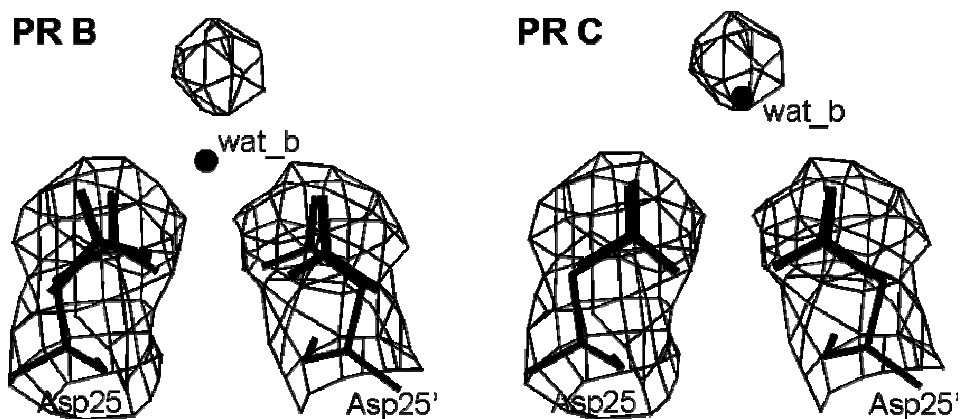
53

Figure 21. Ab initio geometry optimization of the monoprotonated free HIV-1 PR.
Starting (thin line) and final (bold line) configurations of the Asp dyad in the ab initio geometry optimization of model **PR B** and **PR C**. The experimentally measured electron density isosurface[62] is shown for comparison. Note that in **PR C** both the water molecule position and the Asp dyad conformation are compatible with the experimentally determined electron density; while in **PR B** they are not.

Ab initio geometry optimizations performed on model **PR B** and **PR C** indeed confirms that this is the case as: i) **PR C** is more stable than **PR B** by 1.1 kcal/mol[*]; ii) the conformation of **PR C** is closer to the X-ray structure than those of **PR B** (Table 1 and Figure 21); iii) the position of wat_b in **PR B** is compatible with the observed electron density while those of **PR B** is not (Figure 21). Calculations performed with larger models containing up to 5 water molecules did not changed these findings.

In conclusion, our calculations do not support the doubly deprotonated state and provide strong evidence for the presence of **PR C** in the active site of free HIV-1

---

[*] This is a relatively small difference in view of the fact that, to obtain the correct relative energies for the different protomers in E-PR, it was necessary to resort to very involved CI calculations. Nevertheless, DFT-based test calculations performed on the four different protomers of endothiapepsin gave the same order of stability found in previous CI calculations. Thus it appears that our level of theory and out modeling is adequate for describing the energetic at the active site of aspartic proteases.

PR. For these reasons all subsequent calculations on the free enzyme were performed according to the protonation pattern of **PR C**.

Measure	*Exp*	*PR A*	*PR B*	*PR C*
$O\delta2$-$O\delta2$	4.87	7.59	5.70	4.64
$O\delta1$-$O\delta1$	2.95	5.34	3.98	2.51
$C\gamma$-$C\gamma$	4.70	6.81	5.52	4.60
$O\delta1_{Asp25}$-N_{Gly27}	2.55	1.92	2.90	2.83
$O\delta1_{Asp25'}$-$N_{Gly27'}$	2.55	2.28	2.83	2.87
ω	33^0	21^0	48^0	29^0
rmsd	-	1.16	0.59	0.26

Table 1

Selected structural parameters resulting from geometry optimizations of complex **PR A**, **PR B** and **PR C** compared with the experimental structure.

4.1.2. Ab initio MD simulations of the monoprotonated form.

Ab initio MD simulations were performed on models of increasing complexity (**PR C1**, **PR C2**, **PR C** Figure 8) to study the effect of the surrounding groups on the conformational flexibility of the Asp dyad.

PR C1 contains only the Asp dyad modeled as acetic acid/acetate. Ab initio molecular dynamics simulations performed on this model indicates that, after a few femtoseconds, a low barrier hydrogen bond (LBHB) is formed between the two $O\delta1$ atoms. As a consequence the distance between the two Asp groups is considerably shortened (from 2.8 to 2.5 Å) and a fast proton hopping on the fs timescale between the two groups is observed (Figure 22)
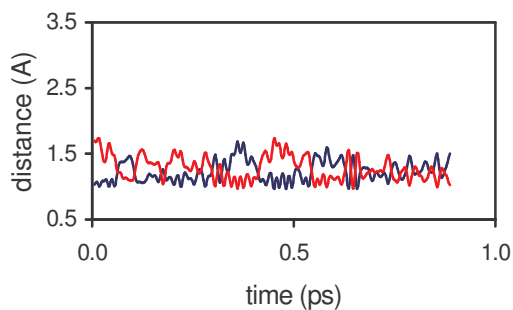
Figure 22. MD simulation of model PR C1.

Plot of the Oδ1…H…Oδ1 distances during the ab initio MD simulation of model PR C1. During the simulation the proton is transferred several times and almost delocalized between the aspartates.

While the LBHB is maintained during the entire simulation, the rest of the complex turns out to be intrinsically unstable. After 0.2 ps the two aspartates start to rotate around the Cβ-Cγ bond and, at the end of the simulation, the coplanarity is completely disrupted (Figure 23).

Figure 23. MD simulation of model PR C1.

Superposition of the final (bold line) and starting (thin line) configurations of the Asp dyad in the MD simulation of model PR C1. The low barrier hydrogen bond is indicated as a dotted line.

We conclude that the LBHB is able to keep the two Asp groups close to each other, but the repulsion of the outer oxygen atoms renders the system unstable. This is an indication that the interaction with other groups is important in maintaining the observed Asp coplanarity.

Inclusion of hydration and of the hydrogen bond interaction with the glycine residues modeled as dimethyl amine (model **PR C2** Figure 8) is not sufficient to obtain a stable system.

Figure 24. MD simulation of model PR C2.

Plot of the $O\delta1...H...O\delta1$ distances during the ab initio MD simulation of model PR C2. After a few hundreds of femtoseconds the hydrogen bonding between the two Asp groups is lost.

Indeed, after a few hundreds of fs of ab initio MD simulations on model **PR C2** the Asp groups rearrange significantly and the coplanarity is lost. Moreover the LBHB that was keeping the two groups close to each other in model **PR C1** is not maintained (Figure 24).

As a result the distance between the two Asp groups increases from 3.1 to 7.0 Å. In the last MD configuration each Asp group interact with the solvent molecules and with the $HN_{Gly27(27')}$ but any Asp-Asp interaction is lost (Figure 25).
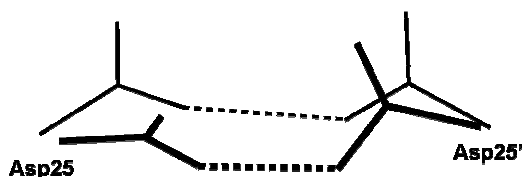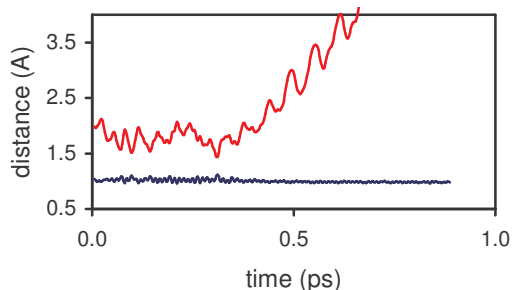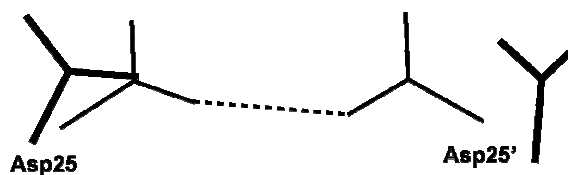

Figure 25. MD simulation of model PR C1.

Superposition of the final (bold line) and starting (thin line) configurations of the Asp dyad in the MD simulation of model PR C1. The low barrier hydrogen bond is indicated as a dotted line.

Inspection of the available X-ray structures of free[62;66;156] and complexed[27;35;36;72;157] HIV-1 PR reveals that, in all cases, the Asp dyad does *not* form an optimal hydrogen bond with $HN_{Gly27(27')}$ ($\angle$(N-H...$O\delta1$) ranging from 125 to 153^0). In model **PR C2**, after 0.4 ps the Asp groups rearrange such as to maximize the hydrogen bond interaction with $HN_{Gly27(27')}$ (maximum $\angle$(N-H...$O\delta1$)

$=179^0$); as a consequence the Asp-Asp hydrogen bond is disrupted and the peculiar conformation of the active site is lost.

We conclude that the $HN_{Gly27(27')}$ hydrogen bond is not a key factor for the stability of the Asp pair.

An detailed inspection of the active site residues close to the Asp dyad suggest that a role could be played by the strong Thr26(26')-Gly27(27') peptide bond dipole moment. Indeed, calculations of the electrostatic field generated by the Asp dyad, reveal that this dipole is perfectly aligned along the field (Figure 26).

The resulting electrostatic interaction turns out to be rather large (~7.8 kcal/mol from a point charge estimation).
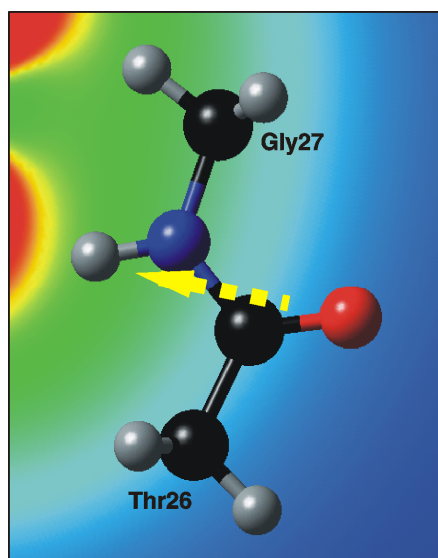
Figure 26. Interaction between the Asp dyad and the Thr-Gly peptide unit dipole.
Plot of the Thr26-Gly27 peptide unit dipole moment (yellow arrow) in the electrostatic potential generated by the Asp dyad. Coloring varies continuously from red (negative potential) to blue (positive potential).

In the simulation where this peptide bond is explicitly considered (model **PR C**) the systems is remarkably stable during the relatively long time scale investigated (4.5 ps).

During the ab initio MD simulation the water molecules experience large mobility and wat_b exchanges with the other water molecules. Also the Asp dyad is very

flexible, as shown by the relatively large standard deviation on the angle that describes the Asp coplanarity $(28(16)^0)$. In this complex a proton hopping between the carboxylates is observed after 4 ps of ab initio MD simulation (Figure 27).
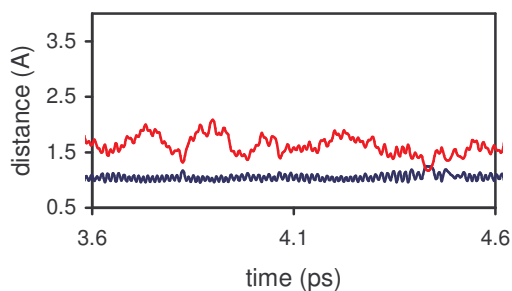
Figure 27. MD simulation of model PR C2.

Plot of the $O\delta1\ldots H\ldots O\delta1$ distances during the ab initio MD simulation of model PR C. The system is stable within the timescale investigated and a proton transfer is observed after about 4 ps. Only the last ps where the proton transfer occurs is shown.

The timescale of the proton hopping in this complex is much longer than those observed for **PR C1**. This difference can be explained in terms of the larger stabilization of the deprotonated Asp group in **PR C** coming from the larger number of hydrogen bond donor groups present. Indeed, in simulations with only one or two water molecules (model **PR C3** and **PR C4**) the overall active site conformation is well maintained, but the proton transfer is observed on shorter time scales (~0.5 and ~2 ps, respectively). In all the three cases the proton is transferred only when wat_b is tightly bound to both the Asp groups. Thus it appears that the role of this water in the proton transfer process is to achieve the correct geometry so as to render the carboxylate oxygens close enough for the transfer to occur.

The calculated stretching frequency for the $O\delta2$-H bond (2470 cm^{-1}) is that characteristic for a weakly bound proton[158], as a proton involved in a LBHB; the typical O-H stretching frequencies for the carboxylic acids fall in the range 2500-3000 cm^{-1}.

4.1.3. Characterization of the chemical bonding.

To characterize the nature of the O-H chemical bonding we analyzed the electronic structure during the proton transfer process in model **PR C**. To this aim the centers of the maximally localized Wannier orbitals[114] were calculated for an isolated carboxylic acid and for 10 snapshots of the MD simulation. In the isolated, protonated Asp side chain the center of the Wannier orbital relative to the O-H bond is located roughly midway between the oxygen and the hydrogen atom (Figure 28**b**). In HIV-1 PR the center is shifted towards the oxygen (Figure 28**c**), indicating a strong polarization of the O-H bond.
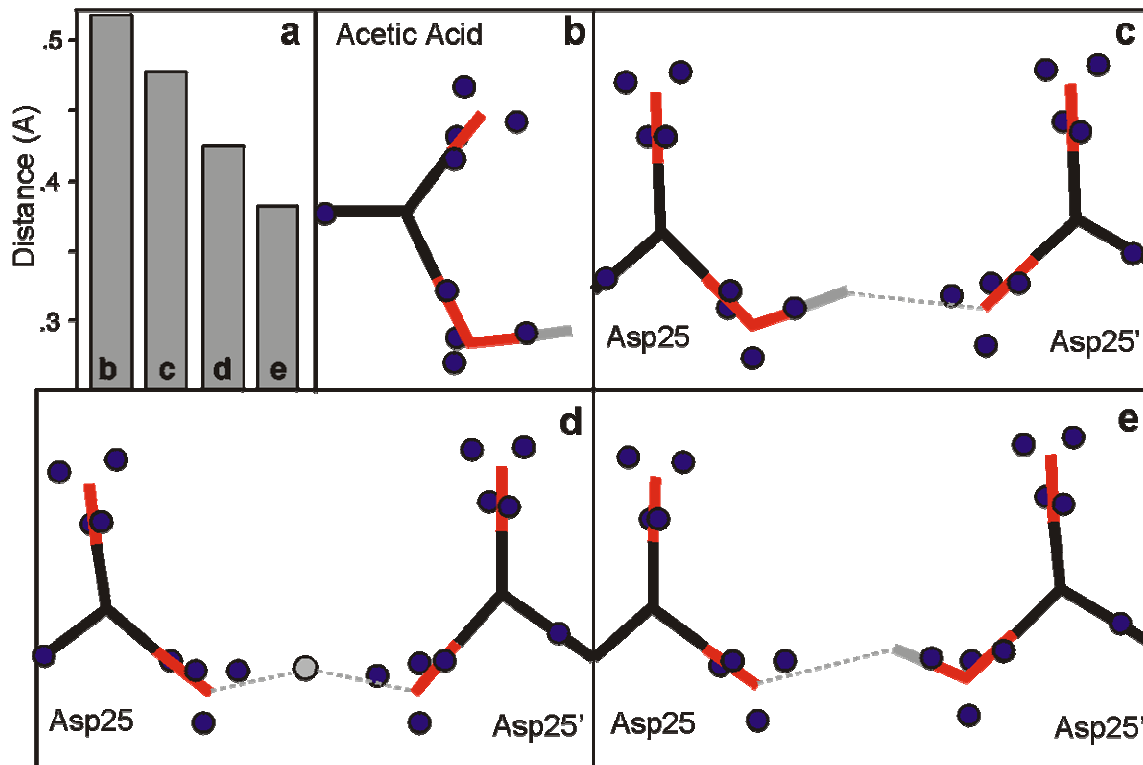
Figure 28. Wannier centers analysis of model PR C.

Isolated acetic acid molecule and proton transfer in model **PR C**. The position of the Wannier centers are represented as blue circles. **a**) Distance between the Wannier center relative to the O-H bond and Oδ1(Asp25). Shorter values indicate larger polarization of the O-H bond. **b**) isolated acetic acid molecule and before (**c**), during (**d**) and after (**e**) the proton transfer.

During the proton transfer a further shift of the center occurs (Figure 28**d**); now the O-H bond is almost ionic in nature. At the end of the proton transfer the proton is bound to the second Asp residue and the nature of this bond is similar to those of **c** due to the system symmetry.

We conclude that the aspartyl O-H bond is strongly polarized and, during the proton transfer, essentially ionic in character. Thus we suggest that the protomer can be described in terms of an almost ionic $O^{\delta-}...H^{+}...O^{\delta-}$ complex. This would explain the relative success obtained in reproducing the structural properties of the active site of the protease with a force-field model in which the two Asp are considered deprotonated and a proton is located between the two groups[24].

4.1.4. Environment effects.

A large hydrophobic region surrounds the active site of the protease. However, on the protein surface, there are a large number of charged residues that, in principle, could induce significant polarization effects in the active site.
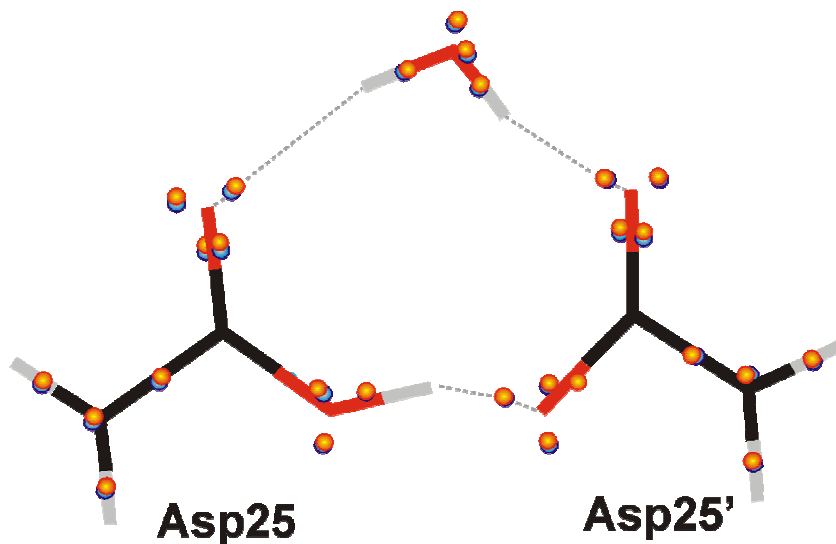

Figure 29. Effect of the protein field

Wannier centers displacement induced by the external protein field. The location of the centers in the isolated system calculation is represented as blue circles; the location of the centers in presence of the protein field is indicated as red circles. The Thr26(26')-Gly27(27') peptide bond moieties are omitted for clarity.

To estimate the amount of polarization due to the rest of the protein, we performed calculations of the Wannier centers in gas-phase and in presence of the external field generated by the protein residues. Only the contributions from atoms more than three chemical bonds away from any atom explicitly included in the ab initio models were considered as it has been found that inclusion of the charges of atoms too close to the model would lead to unphysical overpolarization effects[159].

As an example we report the position of the centers with and without the external potential during the proton transfer (Figure 29)

It turns out that the relative displacement of the Wannier centers is very small, of the order of 1% with respect to the displacement induced by the residues directly included in the ab initio model. These results indicate that the residues far from the active site do not play a key role in the proton transfer process and in determining the conformational stability of the cleavage site, as expected from the preliminary molecular mechanics studies (see Appendix 1). Note that, as the protein dielectric constant is not known, the dielectric constant of vacuum was used and the resulting electrostatic potential is an upper bound to the "real" electrostatic potential.

4.1.5. Discussion.

The protonation state of free HIV-1 was studied by ab initio DFT-based energy calculations.

^{13}C NMR measurements indicate the presence of two equivalent, deprotonated Asp groups in the active site of HIV-1 PR at pH 5-6[31]. Our ab initio calculations on the free enzyme indicate that the doubly deprotonated specie, suggested on the basis of NMR data, is not stable due to the strong repulsion of the two negatively charged Asp groups.

This result can be an artifact of our limited-size model calculations, nevertheless there is no other group in close proximity of the active site that could provide

extra stabilization to the Asp pair. Thus, even using larger models we do not expect our qualitative findings to change.

On the other hand, our calculations indicate that the monoprotonated form is stable and reproduces the structural properties of the Asp dyad observed experimentally. The electronic structure of the monoprotonated form suggests that the latter can be described in terms of an *almost ionic* $O^{\delta-}...H^{+}...O^{\delta-}$ complex. This observation would reconcile the NMR data, suggestive of the presence of two identical ionized Asp groups, with the pH measurements[44] that indicate the presence of a proton in the active site of the enzyme at pH 5-6.

The lowest energy protonation pattern found in HIV-1 PR differs from the pattern proposed for the eukariotic isoenzymes[148]. This result is not unexpected as the active sites of the two enzymes exhibit substantial differences.

Car-Parrinello molecular dynamics simulations allowed identifying a low barrier hydrogen bond and a dipole-charge interaction as key factors for maintaining the active site conformation observed experimentally.

Our calculations suggest that the proton hopping frequency is on the ps timescale. However this result should be taken at the qualitative level because of the relatively short dynamics investigated, the BLYP approximation[132] and the neglect of tunneling effects[160].[133]

While the LBHB keeps the two Asp groups close to each other, the Thr26-Gly27 backbone dipoles are responsible for the almost coplanar orientation that is typical of the active site of aspartic proteases. These findings are completely consistent with the observation that substitution of Gly27 with Valine leads to a completely inactive enzyme[39]. Indeed the bulk group of Valine induces a reorientation in the backbone dipoles that, in turn, would lead to a non-active conformation of the active site of the enzyme, thus confirming the fundamental role for this dipole.

4.2. Protonation state of the Asp dyad in the HIV-1 PR/pepstatin complex.

The protonation state of the active site of HIV-1 protease is of crucial importance in determining the energetic of binding of the inhibitors[11;12;16;17]. One of the most widely used techniques to study the protonation states in proteins is ^{13}C NMR. This technique has been applied to HIV-1 protease[31;47;50], yet the interpretation of the experimental data sometimes can be difficult.

An example is given by the ^{13}C NMR spectrum of the adduct with the transition-state analog pepstatin A[31].

The complex exhibits two distinct signals at 172.4 and 178.8 ppm in the pH range of 2.5 to 6.5[31]. Of the two peaks, the low field one undergoes an isotopic shift, while the other remains unchanged upon deuteration. Based on these results and on the X-ray structure (Figure 5**b**) of the complex with the analog acetyl pepstatin A[71] (Figure 9), the authors proposed the H-bond pattern **PEP A1** shown in Figure 11. In this structure, both aspartyl groups interact with the inhibitor's hydroxyl hydrogen and Asp25' forms an additional H-bond to the hydroxyl oxygen. Thus, the two carboxylates are chemically non-equivalent, which might explain the NMR signals. Furthermore, as one of the aspartyl groups is ionized, the pattern can also explain the different isotropic behavior.

However, also in the eyes of the authors, this assignment is surprising. In fact the chemical shifts reverse relative to those of the carboxylic acids in aqueous solution[161] and in the adduct with the chemically similar inhibitor KNI-272[47].

Clearly, an *ab-initio* calculation of the NMR chemical shift[119;128;162-169] could be of great value in explaining these results. In this part of the work we calculated NMR chemical shifts within the frame of Car-Parrinello molecular dynamics[122]. This combination is particularly powerful, since it allows us to investigate the effects of temperature on the NMR signals, which are expected to be of crucial relevance in biological systems[84].

4.2.1. Monoprotonated form.

The H-bond pattern proposed based on ^{13}C NMR data[31] (Figure 11 PEP A1) is unstable already on the sub-ps time-scale: after $\cong$ 0.5 ps, Hδ2(Asp25') is transferred to Oh(pepstatin A) and, simultaneously, Ho(pepstatin A) is transferred to Oδ2(Asp25). In the new structure (PEP A2 in Figure 11), which is stable on the timescale of the simulation, both aspartyl groups interact strongly with the pepstatin A hydroxyl group, one as H-bond acceptor (Asp25') and one as H-bond donor (Asp25).

The calculated ^{13}C chemical shifts of the carboxyl carbon atoms fluctuate considerably (Figure 30) but they average out to very similar values. These are 128.4(1.1) and 128.7(1.0) ppm for Cγ(Asp25) and Cγ(Asp25'), respectively. Thus, the structure PEP A2 is not compatible with experiment[31].
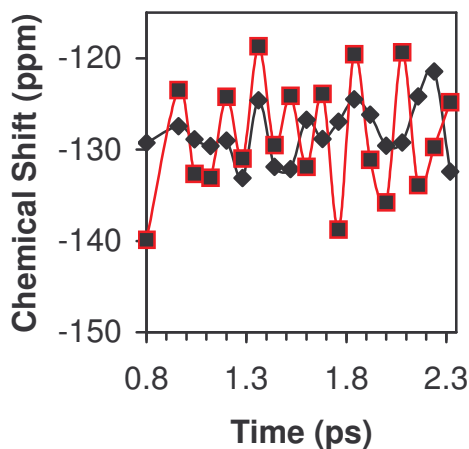
Figure 30. Calculated ^{13}C NMR chemical shifts for model A2.

In red squares are reported the values relative to Asp25'; in black those relative to Asp25

The discrepancy with experiment could arise in principle from the relatively small size of our model and the short time-scale. However, calculated chemical shifts are well converged in the time-scale considered, as shown by their small standard deviation. Furthermore, our complex appears to capture the relevant

chemical interactions at the active site as i) it includes all the groups interacting with the Asp dyad; ii) the cleavage site is buried inside a hydrophobic region which is unlikely to generate a strong electric field[71]; iii) the protein electric field in the cleavage region is small compared to that of other enzymes. We are therefore forced to abandon the monoprotonated state hypothesis and seek another explanation for the NMR signal.

We suggest that both aspartyl side chains are protonated (model **PEP B1** in Figure 11), but, due to the asymmetric nature of the inhibitor, Asp25' has a weaker interaction with its surroundings, leading to a different ^{13}C chemical shift signal.

The observed isotope effect can be rationalized by recalling the theoretical[128] and experimental[126;127;129] evidence that in ^{13}C links the isotopic shift to the strength of the H-bond. This will lead to a large isotope effect in the strongly interacting Asp25 and to a negligible one in the weakly bonded Asp25'. The hypothesis that one of the two aspartyl groups is in a low polar medium was also made by the authors of ref.[31] on the basis of the rather low resonance frequency (172.4 ppm compared to ≈ 176 ppm for protonated and ≈ 180 ppm for deprotonated aspartic acid in water[77].

Our model for the diprotonated form (**PEP B1**) is not dissimilar from what is known on other transition state analog/HIV-1 PR complexes[16;17;50].

4.2.2. Diprotonated form.

The H-bond pattern is maintained and the system is stable during the entire simulation time (5.2 ps). The asymmetric inhibitor hydroxyl group forms a very strong H-bond with Asp25 (dOH=1.65(0.2) Å, Θ=158(11)º) and weaker H-bond interactions with Asp25' (dOH=2.07(0.2) Å, Θ=146(15)º) (Figure 11). Furthermore, during the simulation, each Asp is stabilized by two additional hydrogen bonds: one with the other Asp group (Oδ1-H···Oδ1) and another with the Gly27, Gly 27' amide groups (N-H···Oδ1 in Figure 11).

The calculated chemical shifts differ by 4 ppm. In fact, Cγ(Asp25') and Cγ(Asp25) resonate at 129.2(0.4) and 132.9(0.6) ppm respectively (Figure 31). Thus, this pattern exhibits ^{13}C NMR peaks in much better agreement than those of **PEP A**, evidence of the inequivalence of the two groups in the doubly protonated form.

In principle, the calculated chemical shift difference could be dependent upon the choice of the pseudopotential and the size of the basis set. For this reason the calculated chemical shifts were tested against calculations with different pseudopotentials (Goedecker pseudopotentials) and a larger basis set (100 Ry cutoff instead of 70 Ry cutoff). It turns out that the pseudopotential contributions to the ^{13}C chemical shift are constant within less than 0.1 ppm.

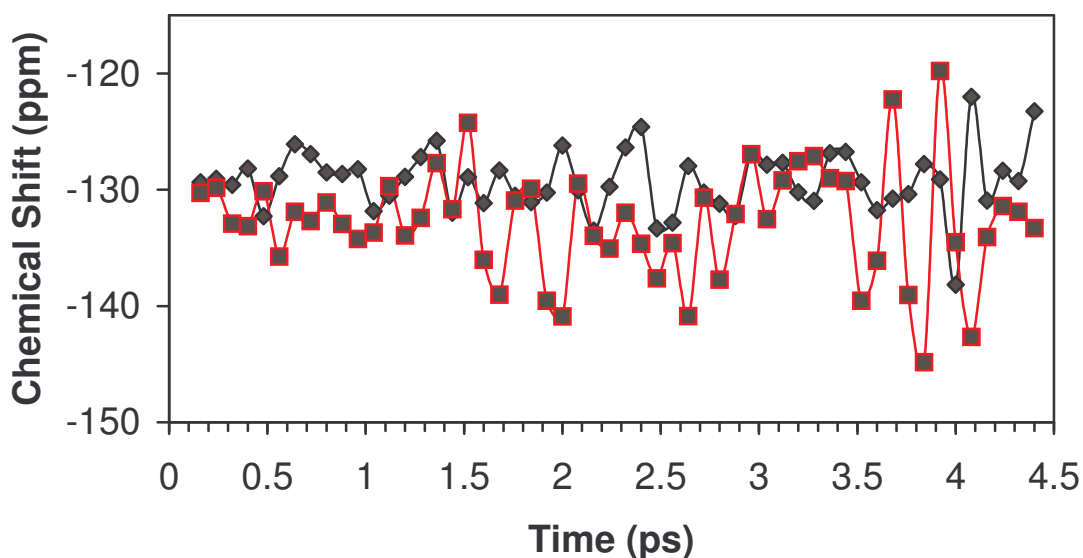**Figure 31. Calculated ^{13}C NMR chemical shifts for model B.**

In red squares are reported the values relative to Asp25'; in black those relative to Asp25

4.2.3. A simple model for interpreting ^{13}C NMR chemical shifts of H-bonded carboxylic acids.

It may seem somewhat surprising that the two Asp groups in the same protonation state (Model **PEP B**) resonate at markedly different frequencies. In

an attempt to rationalize this fact, we now investigate more systematically the effects of H-bonding on protonated carboxylic acids in a low-dielectric medium. To this end, we calculate the ^{13}C NMR properties of formic acid/water H-bond complexes in vacuum (Figure 14) and use the result to interpret the ^{13}C NMR signals of the Asp dyad in complex **PEP B**.

We first perform calculations on formiate and formic acid (**F 0** and **F 0-** Figure 14). These show a carbon deshielding of $\approx$ 2 ppm for formiate (Table 2). A low-field shift can only be due to a change in magnetic anisotropy (MA) and it has to be larger than the two effects producing high-field shifts, namely the increase of diamagnetic contributions due to an increase of electronic density and a decrease in the polarity of the C-O bond.

Complex	Water molecules included	N_C (electrons)	MBO ratio	σ (ppm)
0-	-	4.141	1.00	113.8
0	-	3.937	1.68	111.5
I	WATA	3.950	1.58	113.6
II	WATA, WATB	3.933	1.41	115.7
III	WATA, WATB, WATC	3.917	1.34	116.6
IV	WATA, WATB, WATC, WATD	3.909	1.38	116.0
IIa	WATB, WATC	3.903	1.55	113.6
IIb	WATA, WATD	3.938	1.62	112.8
IIc	WATA, WATC	3.928	1.48	116.3

Table 2.

Selected calculated properties of formic acid/water complexes. N_C is the integrated number of electrons around C. The MBO ratio is the ratio between the C-O^2 and C-O^1 Mayer Bond Orders. σ are the carboxyl carbon ^{13}C NMR chemical shifts.

The origin of MA lies in the carboxylate π-electron system[170]. In the deprotonated form, the π-electrons are more delocalized and the valence bond structures more important in $HCOO^-$ (Figure 32). This is consistent with the calculation of Mayer Bond Orders[135] (MBO's), which give a lower $C\text{-}O^2$ over $C\text{-}O^1$ ratio in the deprotonated form (Table 2).
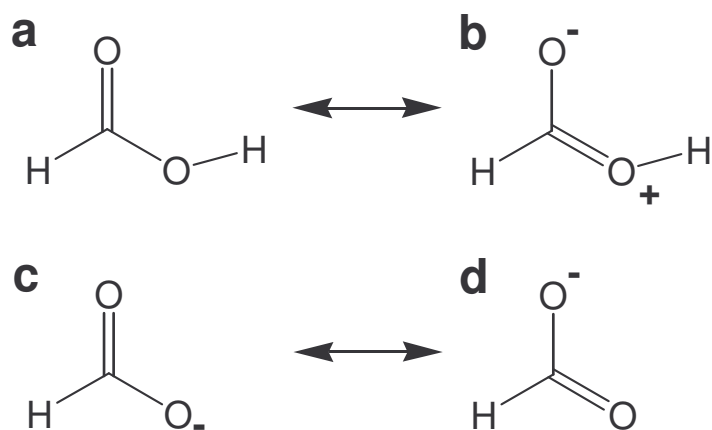
Figure 32. Resonance structures of formic acid.

Valence bond representation of protonated (**a**-**b**) and deprotonated (**c**-**d**) formic acid. The electronic properties of the molecule can be described in terms of linear combination of resonance structures. For each protomer the two structures that provide the largest contribution to the electronic structure are drawn.

These results for the isolated molecules give the key for interpreting a series of calculations performed on formic acid/water complexes (Figure 14). The results are summarized in Table 2. Clearly the complexes can be classified into those that stabilize the π-resonance (complexes **F I**, **F II**, **F IIc, F III**) and those that do not (complex **F IIa** and **F IIb**). In **F IV**, both stabilizing and destabilizing interactions are present, although the latter predominates (Table 2). The MBO analysis confirms this picture. We also find that an H-bond gives a roughly additive contribution to the shielding of about 1-2 ppm, leading to an average value of 1.5(1) ppm.

We can use the results of this model calculation in order to interpret the HIV-1 PR results. In fact, the Asp side chains are in a low dielectric medium environment[75] and their behavior is expected to be similar to that in the gas phase.

In order to make a semi-quantitative estimate let us enumerate the shielding and deshielding contributions: i) the Asp-Asp H-bond, which deshields both aspartyl groups; ii) the H-bonds of Asp25 with Gly27 and with the inhibitor[†] , which deshields Asp25; iii) the Asp25'-Gly27 H-bond, which shields Asp25'; iv) The Asp25'-inhibitor H-bond[‡], which deshields Asp25'.

Assuming that the H-bonding contributions are additive and similar to that of water, we arrive at an estimate of 3.0 ppm, in fairly good agreement with the calculated 3.7 ppm value. This increases our confidence in our results and our simplified model can be used to obtain rough estimates in situations where a full quantum-chemical calculation is not possible.

4.2.4. The isotopic shift.

We now turn our attention to the ^{13}C isotopic shifts of protonated carboxylic acids where hydrogen is replaced with deuterium; i.e.
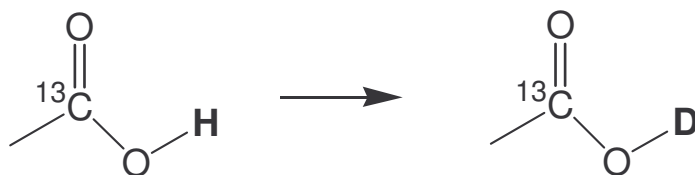
This type of isotopic shift is called *secondary* as the carbon atom is separated by two bonds from the substituted hydrogen atom. A larger *primary* effect is observed on the oxygen atom; directly bound to the substituted hydrogen.

[†] $d(Oh\text{-}O\delta2_{Asp25})= 3.5(0.2)$ Å

[‡] $d(Oh\text{-}O\delta2_{Asp25'})= 3.8(0.2)$Å

In the active site of HIV-1 PR secondary isotopic shifts with an absolute value of
~ 0.1 ppm were measured and used to determine the protonation state of the
catalytic aspartic acid side chains. Experimentally it has been found that only the
low field carboxylic carbon is subject to isotopic shift[31], apparently suggesting
that only one aspartyl is protonated.[126;127] Here we quantified this effect by
performing a series of NMR isotopic shift calculations of formiate/water
complexes (Figure 33). These confirm the dependence of the isotope shift on the
H-bond strength[128].
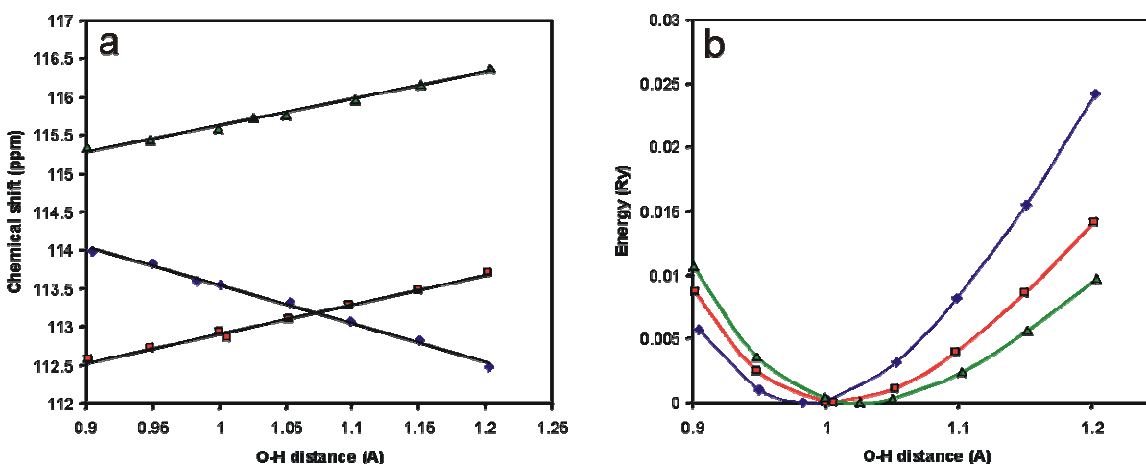
Figure 33 Chemical shift susceptibility (a) and Morse potentials (b).
a) Chemical shift as a function of the carboxylic O-H bond stretching for model system **F IIa**
(blue), **F IIb** (red) and **F IIc** (green); b) Total energy as a function of the carboxylic O-H bond
stretching for model system **F IIa** (blue), **F IIb** (red) and **F IIc** (green).

The isotopic shift can be as small as 0.030 ppm for a weak H-bond interaction.
Since Asp25' is weakly bound, its isotope shift could easily fall within
experimental error. Thus, the origin of isotopic shifts is not necessarily evidence
for a deprotonated carboxylic acid.

4.2.5. Discussion.

Calculations performed models of the enzyme bound to pepstatin indicate that
the protonation pattern suggested on the basis of NMR data[31] is not stable.

Indeed, after a few tenths of picosecond a double proton transfer is observed and a different pattern is formed, stable for the rest of the simulation.

In this complex the two Asp groups exhibit very similar resonances. This result is not in agreement with the experimental data that indicate the presence of two markedly distinct signals in the active site of the protease[31].

To account for the observed chemical inequivalence of the two groups a new protonation pattern was proposed where both the Asp groups were protonated. The calculated ^{13}C NMR chemical shifts for the two Asp groups are different by as much as 4 ppm. The fact that two identical groups in a symmetric enzyme could resonate at different frequency can be explained in terms of the local asymmetry induced by the inhibitor bound to the protease. In particular, despite of the fact that both Asp groups have the same protonation state, the hydrogen-bonding pattern around each group is different. Calculations performed on small formic acid/water complexes indicate that the different hydrogen-bonding pattern could easily explain the observed difference in the chemical shift.

These results are in complete agreement with calorimetric measurements that indicate the uptake of 0.4 protons upon pepstatin binding to HIV-1 PR[82].

It should be remarked that the calculated chemical shift difference (3.7 ppm) is significantly smaller than those experimentally measured (6.2 ppm). This discrepancy can arise from the small size of the model employed in the calculations. The field generated by the residues external to the protease active site, even if apparently not important in describing the dynamical properties of the active site of the free enzyme, could have non-negligible effects on the density currents that determine the chemical shift. Also the short time scale and the use of small fragments to represent the active site residues leads to incomplete sampling and poor thermal averaging of the calculated ^{13}C NMR values. Finally another source of error could come from the use of MT pseudopotentials within the BLYP approximation and the limited size of the basis set employed.

All these factors do not allow, at present, a more accurate calculation of the chemical shifts in large systems. Nevertheless our calculations were able to capture the difference in chemical shift of two identical groups in different

chemical environments and provided a rationale for the observed NMR data. We expect that the use of larger models and basis sets and possibly MM/QM calculations could improve the present results without qualitatively change our findings.

4.3. The reaction mechanism.

In this part of our study study we investigated the influence of the large scale enzyme dynamics on the catalytic activity and structural aspects of the transition state[§].

To this aim we performed a force field-based MD simulation of the HIV-1 PR/substrate complex in water solution to a much longer trajectory than previous works and characterized the protein motions on the ns timescale. Subsequently we calculated ab initio the reaction free energy for the first reaction step in three representative configurations sampled during the classical MD simulation.

Comparison was finally made between the structure of the reaction intermediate obtained from the ab initio MD simulations and the X-ray structures of several aspartic proteases complexed with difluorostatone inhibitors.

4.3.1. Classical MD simulation.

4.3.1.1.   Overall fold dynamics.

The protein equilibrates in the first 700 ps, as indicated by a plot of the rmsd of the backbone atoms as a function of the simulated time (Figure 34).

[§] The proteolysis reaction catalyzed by HIV-1 PR has been subject of several theoretical studies at the classical[23], semiempirical[26] and *ab initio*[27-30] level.

QM/MM calculations pointed to the nucleophilic attack of the water molecule, concerted with a proton transfer from water to Asp 25[26], as the slow step for the reaction[26;27]; on the other hand, ab initio calculations at the RHF/6-31G* level proposed protonation of the amide nitrogen atom either as first reaction step[29] or as rate-limiting step[28]. Nevertheless, these results cannot be taken as conclusive as, to achieve an accurate description of the energetic at the active site of the enzyme, it is necessary to resort to an higher level of theory[30;148] and to include the effect of the protein residues interacting with the Asp dyad[148].
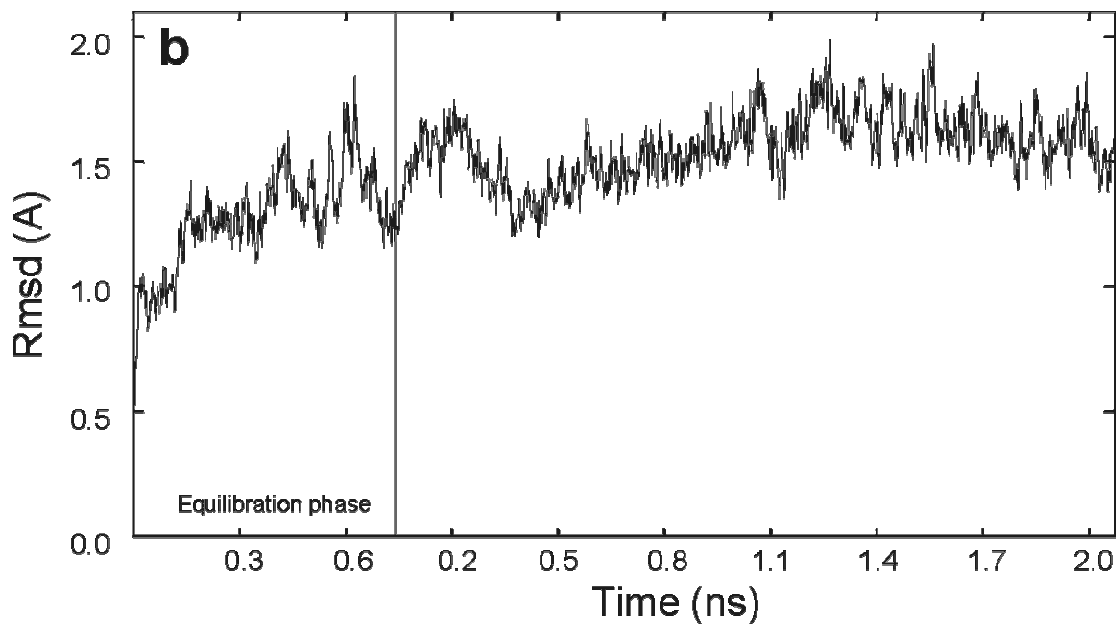
Figure 34. Backbone rmsd with respect to the starting structure.

Comparison of the final structure with the X-ray one (rmsd 1.4 Å) indicates that overall the overall fold is well-maintained (Figure 35).
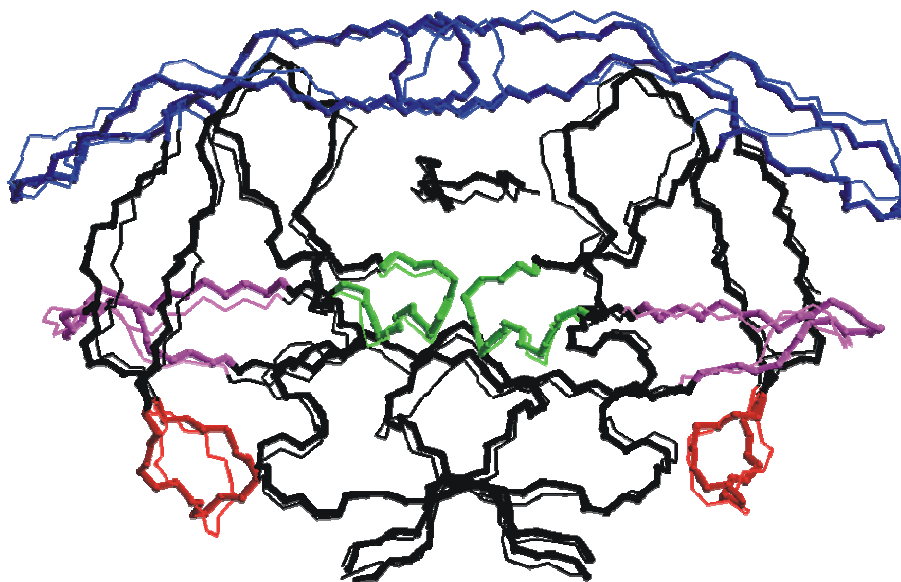
Figure 35. MD simulation of the HIV-1 PR/Substrate complex.

Starting (thin line) and final (bold line) structures of the HIV-1 PR/Substrate complex. The flaps region is indicated in blue; the *fireman's grip* in green; the fulcrum region in red and the cantilever in purple.

In particular, the calculation reproduces fairly well the conformational properties of fingerprint regions' backbone: i) the *fireman's grip* (residues 22-28(22'-28'), rmsd 0.6 Å), which encloses the fully conserved catalytic Asp-Thr-Gly triad; ii) the *flaps* (residues 36-60(36'-60'), rmsd 1.2 Å), which have extended contacts with the substrate and provide most of the protease-substrate binding energy[69]; iii) the *fulcrum* (residues 10-21(10'-21'), rmsd 0.9 A) and the *cantilever* (residues 61-76(61'-76'), rmsd 0.9 A), which modulate the flaps rearrangements and fluctuations[67].

The MD-averaged rmsd fluctuations per residue (for the last 2.1 ns) provide a measure of the local flexibility of the enzyme (Figure 36).

The fireman grip is rather rigid, as a result of the extensive H-bond network conserved in all aspartic proteases; indeed it is the region of the final structure with the lowest rmsd with respect to the starting one. In contrast, the flaps and cantilevers experience large fluctuations, in agreement with previous MD calculations[10;21;24;67;68] and NMR experimental data[49]. These regions usually provide the largest contributions to the binding energy of the protease inhibitors[56;69].
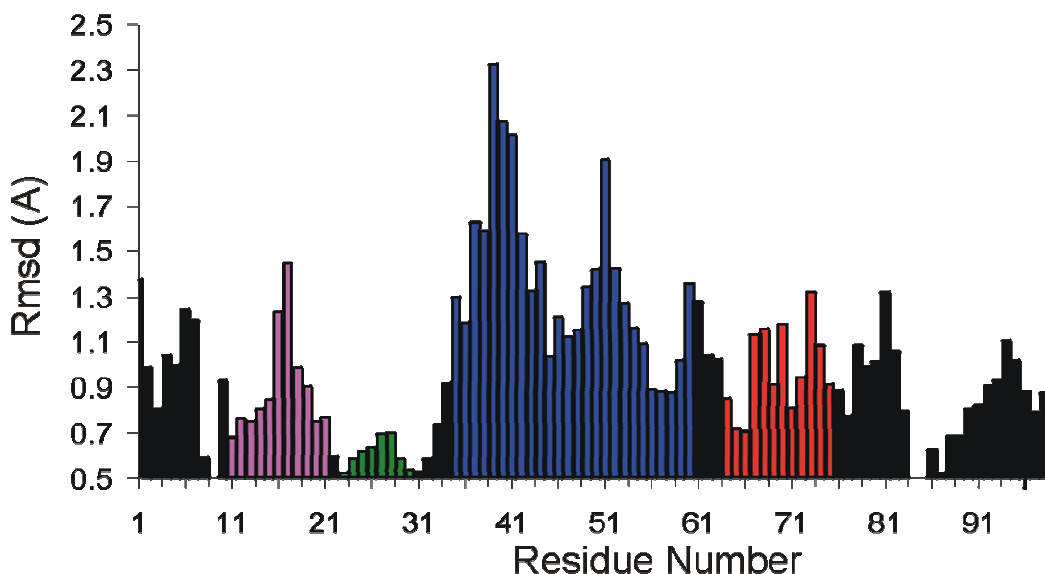
Figure 36. Per-residue average rmsd with respect to the average MD structure.

The flaps region is indicated in blue; the *fireman's grip* in green; the fulcrum region in red and the cantilever in purple.

The dynamical cross correlation map of the Cα atoms provides further insights in the dynamics of these highly mobile regions[67]. The map of the cross correlations of subunit 1 is represented in Figure 37[67]. The map corresponding to subunit 2 exhibits very similar patterns.
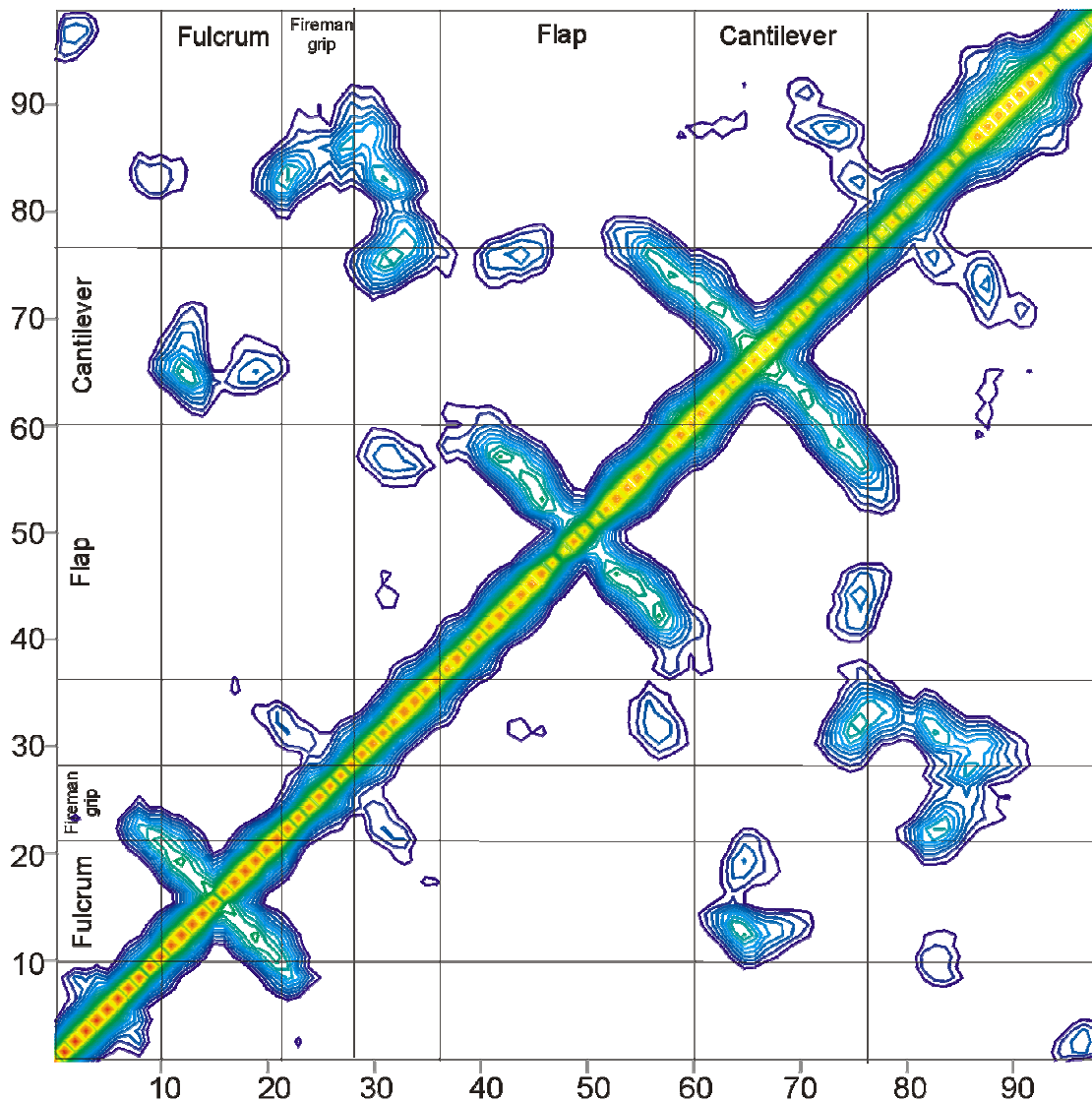
Figure 37. DCCM map of subunit 1 of the HIV-1 PR/Substrate complex.

Both axes go from 1 to 99 according to the residue numbering in subunit 1. The different functional regions of the protease are indicated. Coloring goes from red (high correlation) to blue (lower correlation).

A large value in the map results from a high-correlation between the motions of two Cα atoms. The largest values are obviously found for Cα atoms belonging to residues *i* and *i* $\pm a$ with *a*=0,1,2, which form a diagonal in the map. High correlations are also found for residues *i* and *i* $\pm$4 belonging to α-helices (visible as enlargement of the diagonal) and for *antiparallel* β-sheets, which are characterized by patterns *orthogonal* to the diagonal that originate from the coupling of residues whose numbering flows in opposite directions. The most interesting features are given by the off-diagonal terms, which represent correlations between non-bonded residues not involved in secondary structure elements.

The map shows that the motion of the flaps (35-60, 35'-60') is correlated with those of cantilever (60-76,60'-76') whose motions are strongly correlated with those of the fulcrum (10-21,10'-21'). Similar findings have been reported for the free enzyme[67]. More importantly, Figure 38, which represents the substrate correlations, show that *the motions of the tip of the flaps (48-52, 48'-52') and substrate motions are highly correlated*[171]. Note that, as the protease binds the substrate asymmetrically, the first subunit (1-99) interacts mostly with residues P3-P1, while the second with residues P1'-P3' (Figure 38).
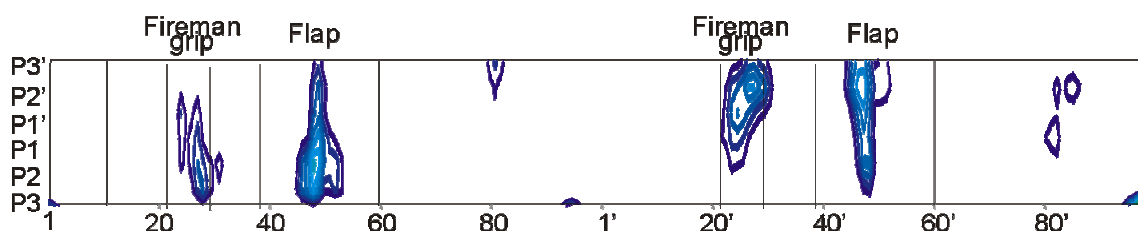
Figure 38. DCCM map of the substrate in the HIV-1 PR/Substrate complex.
On X and Y axis are indicated the HIV-1 PR and the substrate residues, respectively. Numbering goes from P3 to P3' for the substrate and from 1-99,1'-99' for the residues of the two subunits of HIV-1 PR. The functional regions of the protease that mostly interacts with the substrate are indicated. Coloring goes from light to deep blue indicating higher and lower correlation, respectively.

78

The normal modes can be calculated by diagonalizing of the correlation map[**]. The mode with the largest eigenvector describes the most significant global protein motion[145].

Figure 39a superimposes the average MD structure with a vector representation of the mode with the largest eigenvector. This mode involves a rotation of solvent-exposed regions in the two subunits (flaps, cantilever and fulcrum, where the largest components of the eigenvector are observed) and a displacement of the substrate along the substrate cleavage-site axis. Consequently the average distance d_{SA}[††] between Asp 25 (25') and P1(P1') C_α oscillates by more than 1 Å around an average of 7.7 Å (Figure 39b).
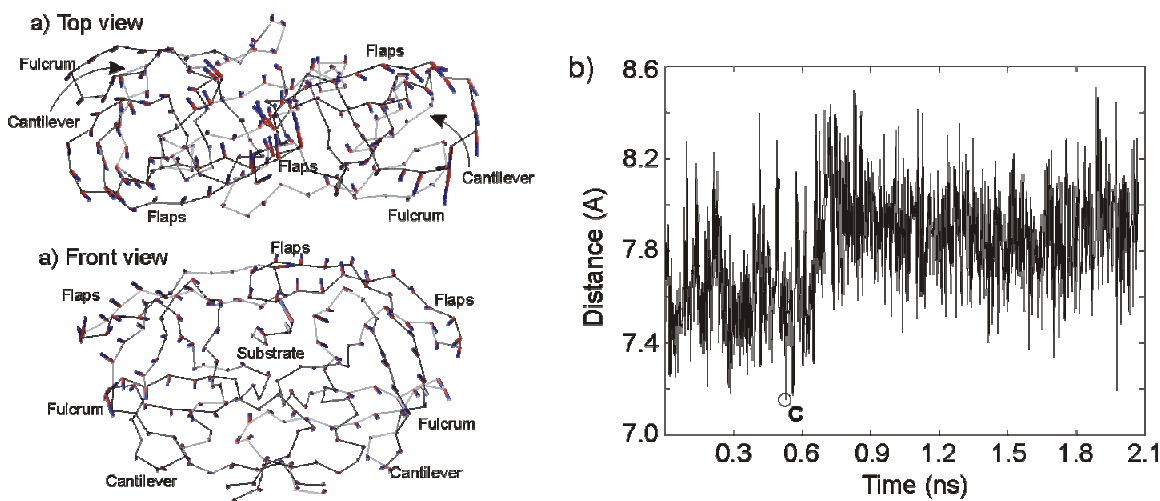
Figure 39. Global motions of the protease and the active site.

a) Front and top view of the projection of the largest eigenvector normal modes (red-blue bars) on the C_α atoms of the protein and protein backbone (thin black line). b) Plot of the substrate-active site distance during the 2.1 ns of classical MD simulation.

[**]   The time-scale investigated (2.1 ns) is expected to be adequate to capture the concerted large scale protein motions of different functional domains relevant for the catalysis[145;171].

[††]   The substrate-active site distance (d_{SA}) was defined as the average distance (measured on the four $C\alpha$) between the Asp dyad and the P1 and P1' Methionine residues:

$d_{SA} = (d(C\alpha_{25}\text{-}C\alpha_{P1}) + d(C\alpha_{25}\text{-}C\alpha_{P1'}) + d(C\alpha_{25'}\text{-}C\alpha_{P1}) + d(C\alpha_{25'}\text{-}C\alpha_{P1'}))/4.0$

We conclude therefore that significant motions of the substrate occur towards the Asp dyad and that these motions are coupled to global motions of several regions of the enzyme, both close and far from the active site.

4.3.1.2. H-bonding at the active site.

The network of hydrogen bonds at the active site is described in Figure 40; the corresponding distances are summarized in Table 3.

Distance	Crystal structure (Å)	MD average (Å)	σ
C_{P1}-O_{WATC}	-	3.97	0.25
O_{Gly27}-N_{MetP1}	2.81	4.74	0.60
$O_{Gly27'}$-$N_{GlnP2'}$	2.89	2.47	0.45
$O\gamma_{Thr26}$-$N_{Thr26'}$	3.26	3.23	0.23
$O\gamma_{Thr26}$-$O_{Leu24'}$	2.83	2.76	0.14
$O\gamma_{Thr26'}$-N_{Thr26}	4.84	3.13	0.17
$O\gamma_{Thr26'}$-O_{Leu24}	5.17	2.83	0.16
O_{WatA}-N_{Ile50}	3.30	3.39	0.38
O_{WatA}-$N_{Ile50'}$	2.67	3.16	0.26
O_{WatA}-O_{IleP2}	3.04	2.85	0.23
O_{WatA}-$O_{MetP1'}$	2.48	2.79	0.15
O_{Gly48}-N_{IleP2}	2.59	2.89	0.19
$O_{Gly48'}$-$N_{ArgP3'}$	3.00	2.98	0.18
O_{Asp29}-N_{ThrP3}	3.09	4.82	1.42
$N\epsilon_{GlnP2'}$-$O_{Asp30'}$	5.82	2.94	0.16
$O\epsilon_{GlnP2'}$-$N_{Asp29'}$	4.10	3.27	0.36
$O\epsilon_{GlnP2'}$-$N_{Asp30'}$	3.50	3.08	0.20

Table 3.

H-bond distances in HIV-1 PR/substrate complex. Comparison between the crystal structure of HIV-1 PR/inhibitor complex and the average MD structure of HIV-1 PR/substrate complex. The standard deviation of the MD averaged values is reported in the last column.

The substrate is directly linked to the flaps by two hydrogen bonds (O_{Gly48}-HN_{IleP2}, $O_{Gly48'}$-$HN_{ArgP3'}$ Figure 40); two additional hydrogen bonds, mediated by a water molecule (Wat 211), link the substrate to Ile50(50') (Figure 40). This water molecule has been observed in all the crystal structure of the enzyme bound with peptido-mimetic inhibitors and is essential for the substrate binding[53]. The side chain of the substrate $GlnP_2$' is hydrogen bonded to the backbone amide hydrogen atoms of Asp29 and Asp30 and to the backbone carboxyl oxygen of Asp30. All these interactions are well kept during the dynamics as demonstrated by the MD-averaged values and standard deviations (Table 3).
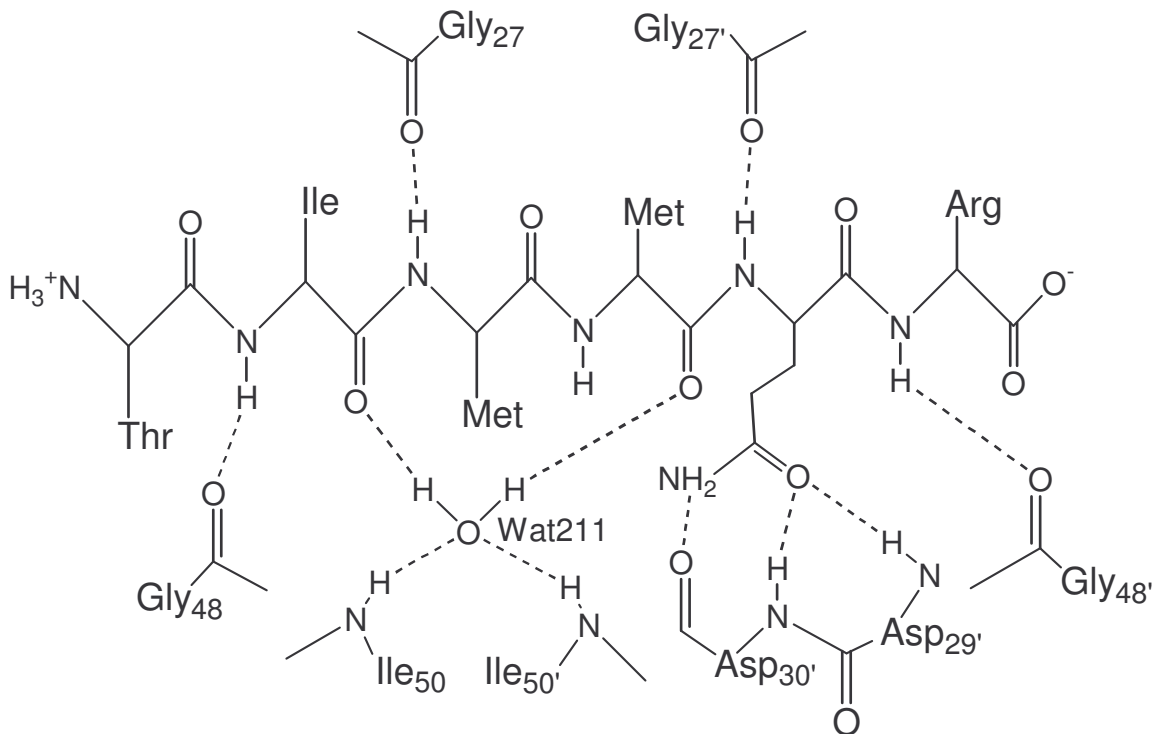

Figure 40. Binding interactions of the substrate with HIV-1 PR.

In the protease-inhibitor complex the terminal amide group of residue P_2' points towards the solvent and does not interact with the carboxyl oxygen of Asp30. In

our MD simulation this hydrogen bond is formed after few hundreds of picoseconds and then is conserved along the entire trajectory investigated. The binding mode of GlnP2' is of particular interest as, in the vast majority of the peptide sequences cleaved by the protease, only Gln or Asp are observed in the P_2' position[38] therefore this residue has been predicted to be important for the catalysis[42].

Finally two hydrogen bonds link the substrate to the carbonyl groups of Gly27(27') (Figure 40). These hydrogen bonds are ubiquitous in the structure of the aspartic proteases complexed with peptido-mimetic inhibitors; yet previous MD calculations reported some difficulties in maintaining them[26]. Thus we restrained them with a weak harmonic constraint (force constant 0.5 kcal mol^{-1} Å^2, R_0= 2.3 Å) during the equilibration phase. The constraint was then released during the 2.1 ns of MD simulation. The O(Gly27')··HN(GlnP2') hydrogen bond is maintained during the entire MD simulation (Table 3), on the other hand the O(Gly27)··HN(MetP1) hydrogen bond is much weaker and forms and breaks several times. The reason for the different dynamics of the two hydrogen bonds can be ascribed to the presence of WATC H-bonded to the deprotonated Asp 25. This water molecule is placed asymmetrically inside the active site of the enzyme and is prevalently located between the substrate and the active site, close to Gly27. Therefore its presence does not allow the substrate to come too close to Gly 27. WATC is not present in the enzyme-inhibitor complexes that mimic the reaction intermediate, as, in the reaction intermediate, it is covalently bound to the substrate to form a gem-diol. Thus our calculations indicate that this hydrogen bond is very weak in the enzyme-substrate complex and becomes much stronger in the enzyme-intermediate structure. We suggest that its formation could be important in lowering the reaction free energy for the first reaction step.

4.3.2. Constrained ab initio molecular dynamics.

Force field-based MD simulations indicate that the HIV-1 PR/substrate complex is flexible. In particular the substrate experiences considerable motions toward the Asp dyad. These fluctuations are coupled to large-scale protein motions.
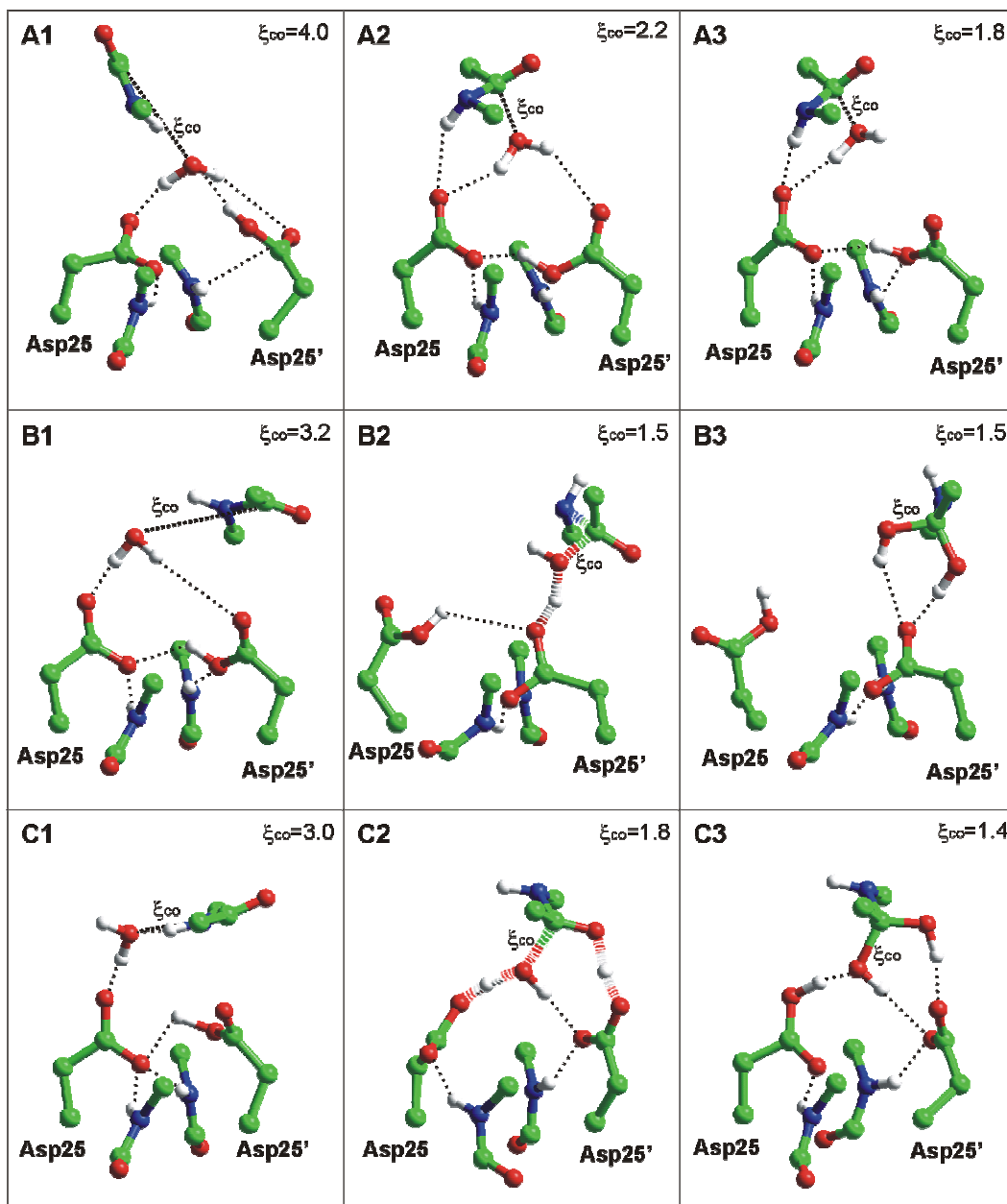

Figure 41. Constrained ab initio MD simulations of model SUB A-C.

Selected snapshots from the constrained ab initio MD simulation of model **SUB A-C**. Starting (1), intermediate (2) and final (3) structures. The value of the constraint distance (ξ_{CO}) is reported in each frame. Partially formed-broken bonds are drawn as discontinuous.

To elucidate the influence of these thermal fluctuations on the catalytic activity of HIV-1 PR we investigated ab initio the dynamics and energetic of the first step of the reaction, that is the nucleophylic attack of the catalytic water (WATC) to the carbonyl carbon (Figure 6 from ES to INT).

To this aim we performed constrained *ab initio* MD simulations starting from different substrate/enzyme conformations. Complex **SUB A** (Figure 12) is obtained by geometry optimization of the initial structural model. In this complex, the average distance between substrate/Asp dyad is maximum d_{sa} = 8.1 Å, that is close to the maximum distance during the classical MD (Figure 39b). **SUB B** (Figure 12) is the snapshot of the classical MD simulation after 201 ps. In this configuration, d_{sa} =7.7 Å, that is the MD-averaged distance (Figure 39b). The last (**SUB C,** Figure 12) is the snapshot the classical MD simulation after 545 ps, for which d_{sa} is minimum (d_{sa}=7.17 Å Figure 39b).

In the ab initio MD simulations of the three system, the substrate carbonyl carbon-water oxygen distance (ξ_{co}) was progressively reduced from ~4.0 to ~1.5 Å. The activation free energy was obtained by point integration along the path (Figure 42)

4.3.2.1.   Complex **SUB A**.

In the starting structure (Figure 41, **SUB A1**) WATC is bridging the two Asp groups, forming two H-bonds with Asp25' and one with Asp25. These interactions are maintained for ξ_{oc} ranging 4.0 to 2.33 Å. As ξ_{oc} further decreases the water molecule moves away from Asp25' and the H_{WAT}-$O\delta1_{Asp25}$ hydrogen bond is disrupted (Figure 41, **SUB A3**).

During the simulation, the dipole moment of the substrate is oriented along the electrostatic field generated by the negatively charged Asp residue (Figure 41, **SUB A2**). As WATC approach substrate carbonyl carbon, the peptide bond partially loses its planarity, as described by the Cα-N-C-O torsion angle that progressively experiences larger deviations from 0° (Table 4). Despite the relative loss of planarity of the peptide bond, the arrangement around the

carbonyl carbon is still planar, as indicated by the Cα-C-O and N-C-O angles close to 120° (Table 4).

Model	point	ξ_{CO}	C-N(Å)	C-O(A)	Cα-C-N(°)	N-C-O(°)	Cα-N-C-Cα(°)
A	1	4.00	1.36(0.03)	1.24(0.02)	120(2)	123(2)	-5(9)
	2	3.25	1.37(0.04)	1.25(0.03	120(3)	123(3)	-8(11)
	3	2.79	1.38(0.03)	1.24(0.02)	121(3)	123(3)	-11(14)
	4	2.51	1.38(0.04)	1.24(0.02)	121(5)	123(4)	-13(10)
	5	2.33	1.39(0.03)	1.24(0.01)	121(4)	122(3)	-17(11)
	6	2.20	1.40(0.03)	1.24(0.02)	121(3)	121(3)	-20(9)
	7	2.02	1.42(0.03)	1.25(0.02)	121(3)	120(3)	-26(8)
	8	1.92	1.44(0.05)	1.25(0.02)	120(5)	121(6)	-25(24)
	9	1.75	1.46(0.07)	1.27(0.04)	119(4)	119(4)	-31(18)
B	1	3.17	1.38(0.04)	1.25(0.03)	121(4)	122(5)	3(20)
	2	2.50	1.40(0.04)	1.24(0.04)	120(6)	121(5)	17(16)
	3	2.00	1.44(0.06)	1.24(0.02)	120(5)	120(4)	21(10)
	4	1.80	1.45(0.05)	1.26(0.03)	119(4)	118(4)	24(9)
	5	1.65	1.52(0.07)	1.27(0.02)	118(3)	117(4)	24(19)
	6	1.55	1.56(0.05)	1.28(0.02)	116(3)	116(3)	27(21)
	7	1.47	1.53(0.09)	1.39(0.08)	110(7)	109(5)	49(28)
C	1	3.00	1.37(0.02)	1.26(0.03)	119(5)	121(3)	-3(14)
	2	2.80	1.38(0.03)	1.24(0.02)	120(5)	121(4)	13(9)
	3	2.40	1.40(0.03)	1.24(0.02)	121(3)	121(3)	16(16)
	4	2.10	1.41(0.03)	1.25(0.02)	121(3)	120(4)	28(12)
	5	1.80	1.43(0.04)	1.30(0.04)	117(5)	116(4)	34(14)
	6	1.50[a]	1.48(0.05)	1.42(0.03)	110(4)	109(4)	55(18)

Table 4.

Selected structural parameters of the substrate molecule from the constrained MD simulations of model **SUB A-C**: constraint distance; C-N and C-O bond lengths; Cα-C-N and N-C-O angles that define the transition of the carbonyl carbon coordination from triangular (angles of ~120°) to tetrahedral (angles of ~109°); Cα-N-C-O torsion angle that defines the peptide bond planarity (planar ~0°, tetrahedral ~60°).
In the last row (model **SUB C**, point n.**6**) the results of the *unconstrained ab initio* MD simulation on the reaction intermediate are reported. [a]In **SUB C6** the distance was not constrained; the reported value is the MD average value.

Furthermore, while decreasing ξ_{OC} from 4.0 to 1.9 Å, the C-N bond elongates of as much as 0.1 Å while the C-O bond length changes by only 0.01 Å (Table 4). Thus approaching of the water oxygen to the carbonyl carbon leads to loss of peptide bond planarity and C-N bond elongation but little effect is observed on the C-O bond length and carbonyl bond angles.

These findings support the idea that the C-N bond (de)stabilization is independent from the (de)stabilization of the C-O double bond and are consistent with previous ab initio calculations on formamide molecules[172;173].

At ξ_{oc} = 1.75 Å (Figure 41 **SUB A3**) C-N distance is 1.46(0.07) Å and the Cα-N-C-O torsion angle is 31(18)º (Table 4 and Figure 41, **SUB A3**). WATC interacts only with Asp25 but the average O_{WATC}-$O\delta1_{Asp25}$ distance is 2.7 Å, too long for any proton transfer to occur. The distances of the reactants (WATC and substrate) from the other active site residues are even longer. Thus, in this model, the active site residues are too far to provide proper stabilization for a transition state and the simulation was interrupted.

The corresponding calculated free energy that is required for the first reaction step in model **SUB A** is larger than 50 kcal/mol (Figure 42).
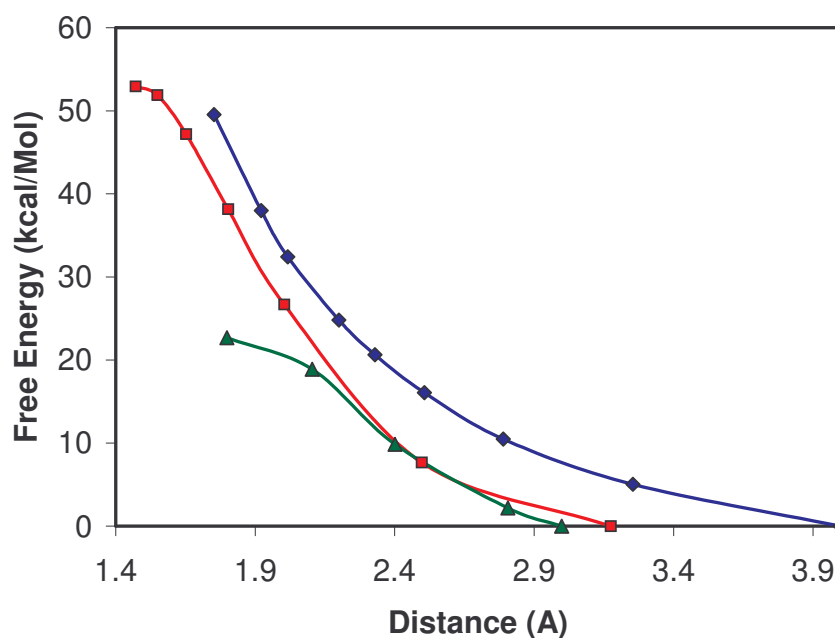

Figure 42. Calculated free energy for the first reaction step as a function of ξ_{co}.
Calculated reaction free energies for model **SUB A** (blue), **SUB B** (red) and **SUB C** (green)

4.3.2.2.   Complex **SUB B.**

In complex **SUB B** the substrate is $\approx$ 0.4 Å closer to the Asp dyad than it is in **SUB A**. In the first part of the ab initio MD simulation (ξ_{oc} ranging from 3.17 to 2.50 Å) Asp 25' is protonated and WATC is prevalently located close to Asp 25 (Figure 41 **SUB B1**). When ξ_{oc} is further reduced (from 2.50 to 1.65 Å), the water molecule is strongly hydrogen bonded to both the Asp groups to form an hydrogen bonding pattern very similar to those found in the free enzyme. As in the free enzyme, in this part of the simulation a low barrier hydrogen bond is formed between the two Asp groups that exchange a proton several times in the subps timescale.

As ξ_{oc} is decreased from 3.17 to 1.80 Å the average C-N bond length increases from 1.38 to 1.45 Å and the Cα-N-C-Cα torsion angle increases from 3 to 24°. In contrast little or no effect is observed on the C-O bond length and angles until ξ_{oc} becomes smaller that 1.8 Å (Table 4).

In the last part of the ab initio MD simulation (ξ_{oc} ranging from 1.57 to 1.45 Å), Asp 25 is protonated; thus the proton location is different from that of the starting protomer. At $\xi_{oc} \approx$ 1.55 Å, a proton in transferred from WATC to the deprotonated carboxyl group of Asp 25' (Figure 41 **SUB B2**) and a negatively charged substrate intermediate is formed.

This intermediate is characterized by an exceptionally long and weak C-N bond (average C-N distance 1.63(0.1) Å with fluctuations up to 1.9 Å). The C-O bond length (1.28 Å compared to a typical C-OH bond length of ~ 1.40 Å) and bond angles (Cα-C-N and N-C-O ~116° Table 4) indicate that the C-O bond yet retains a large double bond character. This intermediate is very reactive: within a few hundreds of femtoseconds, a proton is transferred from Asp 25 to the carboxylic oxygen of the substrate to form a stable gem-diol (Figure 41, **SUB B3**).

The calculated free energy for this two-step process is 50 Kcal/mol (Figure 42).

4.3.2.3.   Complex **SUB C.**

Model **SUB C** was built starting from the force field-based MD snapshot with the shortest **d_{sa}** (7.1 Å). In the first part of the simulation (ξ_{oc} ranging from 3.0 to 2.1

Å) the water molecule is located close to the deprotonated Asp 25 and the substrate dipole is oriented along the electrostatic field generated by the negatively charged Asp 25 (Figure 41 **SUB C1**). The corresponding hydrogen bond pattern is very similar to those observed for **SUB B**. Asp 25' does not form stable hydrogen bonds with the other active site residues and is highly mobile in this part of the simulation. As ξ_{oc} decreases from 3.0 to 2.1 Å the peptide bond loses its planarity and the C-N bond stretches from 1.37 to 1.41 Å. Consistently with model **SUB A** and **SUB B** no significant effect is observed at the C-O bond (Table 4). At ξ_{oc} = 1.8 Å, two protons are simultaneously transferred: one from WATC to Asp25 and the second from Asp25' to the substrate carbonyl oxygen (Figure 41 **SUB C2**). Consequently, a stable gem-diol is formed (Figure 41 **SUB C3**).

This result is in complete agreement with the experimental findings that the rate limiting step is associated with the exchange of two protons[42;44]. Moreover both the Asp groups are essential for the catalysis as mutation of any one of the two Asp groups results in a completely inactive enzyme[41;174]. Indeed, in the double proton transfer, the presence of both the Asps is required. The observed reaction pathway and intermediate are consistent with the suggestions based on previous calculations[26;27].

The transition state is characterized by a substrate carbonyl carbon atom still largely Sp2, as indicated by its bond angles (~115º see Figure 41 **SUB C2** and Table 4). Full Sp3 geometry around the carbon atom is achieved only after the transition state, when the C-O$_{WATC}$ bond is fully formed (Figure 41 **SUB C3** and Table 4).

The calculated reaction free energy for the nucleophylic attack concerted with a double proton transfer is 20 kcal/Mol. The calculated free energy barrier is therefore much lower than those calculated for the two-step reaction pathway occurring in model **SUB B** (Figure 42).

4.3.3. The gem-diol Intermediate.

Figure 43 represent the structure of the gem-diol intermediate bound to the active site of the protease from our 1.4 ps of ab initio MD simulations starting from the last configuration of model **SUB C**.
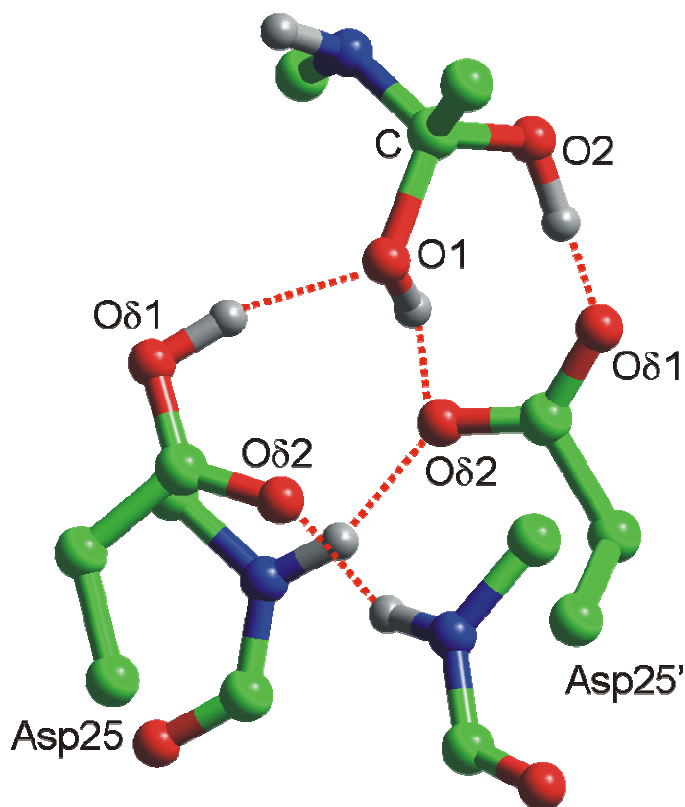
Figure 43. Model of the reaction intermediate (INT).

During the ab initio MD simulation the gem-diol forms two hydrogen bonds with the deprotonated aspartate and one with the protonated. Additional stabilization to the Asp dyad is provided by interaction with the Thr26(26')-Gly27(27') peptide bond unit.

The calculated structure of the reaction intermediate **INT** was validated by comparison of the structural parameters obtained from the ab initio MD simulation of the gem-diol with those of difluorostatone inhibitors bound to HIV-1 PR, penicillopepsin and endothiapepsin[27] (Table 5)[175;176]. Difluorostatone inhibitors contain the –CF_2-CO- moiety that is easily hydrolyzed to the gem-diol –

CF_2-$C(OH)_2$-. Several compounds containing this group are powerful inhibitors of the aspartic proteases as the gem-diol form mimic the reaction intermediate **INT**.

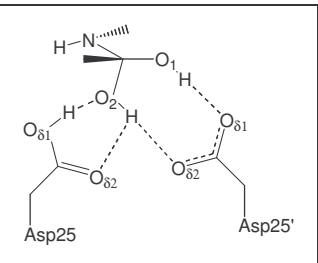|	**MD-averages (This work)**	**HIV-1 PR[27]**	**Endothiap.[176]**	**Penicillop.[175]**
C - O1	1.42(0.03)	1.37	1.38	1.40
C - O2	1.50(0.06)	1.40	1.47	1.47
O1···Oδ1(Asp25')	2.69(0.12)	2.65	2.58	2.62
O2···Oδ2(Asp25')	2.92(0.34)	2.83	2.90	2.67
O2···Oδ2(Asp25)	2.80(0.14)	3.15	2.58	2.63

Table 5.

Selected bond lengths and hydrogen bond distances (Å) of the gem-diol intermediate. Average values of the *ab initio* MD of the gem-diol intermediate (Figure 13) compared with the values found in the X-ray structures of aspartic proteases complexed with difluorostatone inhibitors.

The crystal structures of the difluorostatone/aspartic protease adducts suggest that the difluorostatone binding mode is remarkably similar to the binding mode calculated for the gem-diol intermediate (Figure 43). Table 5 shows that both the calculated bond lengths and hydrogen bond distances are in excellent agreement with the experimental data.

In particular, the calculated C-O_1 bond length is very close to the X-ray values (calculated 1.41(0.03) Å, X-ray 1.38-1.40 Å); whence a slightly larger deviation is observed for the C-O_2 bond (calculated 1.51(0.06) Å, X-ray 1.41-1.47 Å).

The C-O_2 bond is significantly longer than C-O_1. We suggest that this difference reflects the larger reactivity of O_2 towards the dissociation from the carbon atom to form a water molecule following a reaction path reversed with respect to those

investigated by us. To support this hypothesis we note that the shortest $C-O_2$ distance (1.407, HIV-1 PR[27]) is observed in the crystal structure where the $O_2\cdots O_{\delta 1}$ hydrogen bond is the weakest (3.15 Å)[27].

4.3.4. Reactants polarization.

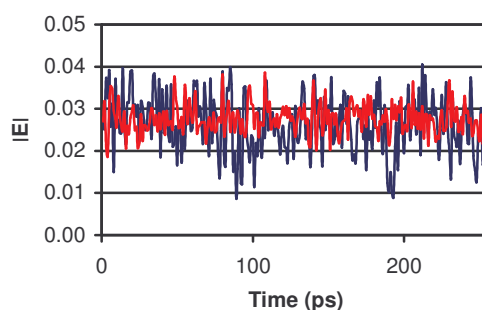We now turn our attention towards the polarization of the reactants within the active site of HIV-1 PR.

Figure 44. Electric field acting on WATC and on a solvent water molecule.
Modulus of the electrostatic field in eVÅ[-1] acting on WATC (red) and on a solvent water molecule (blue). Note that the fluctuations around the average value are much larger for the solvent water molecule as compared to WATC.

The peptide bond hydrolysis occurs 10^4 times faster in the active site of HIV-1 PR than in bulk water. Electrostatic polarization of the reactants may be a key factor in the enzymatic catalysis[177]; therefore we investigated the electrostatics at the cleavage site both classically and ab initio and compared the results with those obtained for bulk water.

The electrostatic field experienced by WATC during the classical MD simulation is very similar to those experienced by a water molecule in bulk water (average modulus of the electric field 0.0266(2) eV/Å compared to 0.0265(1) eV/Å Figure 44).

The carbonyl carbon of the substrate experiences an electrostatic field similar to those experienced by a carbonyl carbon atom of a solvent-exposed residue at

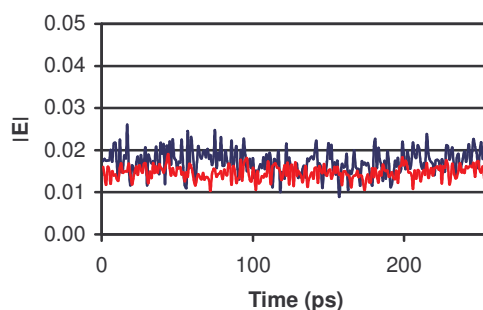the protein surface (average modulus of the electric field 0.015(1) eV/Å compared to 0.018(1) eV/Å Figure 45).

Figure 45. Electric field on a carbonyl carbon in the active site and at the protein surface.
Modulus of the electrostatic field in eVÅ^{-1} acting on the carbonyl carbon of the scissile bond (blue) and on the carbonyl carbon of a solvent exposed peptide bond (Gly 17, red)

Our force fields based MD calculations therefore suggest that the reactants in the active site do not experience electrostatic fields larger than those experienced by the same groups in water.

A consistent picture is obtained by the ab initio calculations. The calculated dipole moment of WATC is the same as those of a water molecule in water (3.0(0.2) D and 3.0 D[117], respectively).

Two independent lines of evidence therefore indicate that the reactants are polarized similarly in the active site of HIV-1 PR and in bulk water. We conclude that polarization is *not* the source of the catalytic power of HIV-1 PR.

4.3.4.1.   Environment effects.

The values of the dipole moment of WATC, calculated in the presence of the protein field, were compared with those obtained for the gas phase model. It turns out that the calculated water molecule dipole moment has the same value in both cases. These findings suggest that, for this system, long range electrostatic interactions do not provide a large contribution in determining the electronic structure of the active site. This confirms that use of a relatively small

gas-phase model to reproduce the reaction pathway appears to be justified for this particular system.

4.3.5. Discussion

In our study we investigated the conformational flexibility of the enzyme-substrate complex by force field-based MD simulations. Constrained ab initio molecular dynamics simulations were then performed on models of the enzyme-substrate complex obtained from the classical MD simulation.

The combination of classical and ab initio molecular dynamics simulation allows 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 calculations indicate that the fireman's grip region encompassing the active site is rather rigid and does not exhibit significant motions on the ns timescale. This may be taken as a validation of our approximation to keep restrained the terminal atoms in the ab initio calculations (see 3.1). In contrast, the substrate-cleavage site distance is modulated by global protein motions and fluctuates considerably on the ns timescale.

Several classical MD studies have been reported in the literature[23]. However, the timescale investigated is much smaller (~0.1 ns). As we focus on the dynamic motions on the ns timescale, comparison between our findings and the previous work is difficult.

The free energy associated to the nucleophylic attack of water to the substrate carbonyl carbon has been subsequently investigated by constrained ab initio molecular dynamics. Simulations were carried out on models of MD-averaged, large and small substrate-Asp dyad distances d_{CO}. At MD-averaged or larger d_{CO} distances the transition state is achieved through a proton transfer from water to Asp 25'. The process is highly activated and requires more than 50 kcal/mol.

At short d_{CO}, the Asp dyad can perform a *concerted* double proton transfer with water and substrate. This event stabilizes the nucleophylic agent OH$^-$ by the

formation of a complex H-bond network at the cleavage site (Figure 41 **SUB C2**). As a result, the activation energy is much lower (20 kcal/mol), close to the experimental value of 16-18 kcal/mol.

The structural properties of the reaction intermediate turn out to be in very close agreement with those of intermediate-mimic inhibitors of the enzyme[27].

We conclude that the key event, represented by the concerted double proton transfer, can occur only at short d_{CO} values, which, in turn, result from the fluctuations induced by the protein motions.

A comment is in order on the fact that the calculated reaction free energy for the first reaction step is a few kcal/mol larger than the experimental value. The discrepancy between experiment and theory can be ascribed to i) the use of a small model and ii) to use of constraints for the terminal carbons; i.e. on the assumption that the protein region around the active site is rigid. Use of an MM/QM approach is expected to improve the agreement with the experimental data.

However, our modeling appears to capture the key aspects of the enzyme catalyzed nucleophylic reaction as indicated also by the striking agreement of the calculated structural parameters for the reaction intermediate with the available crystallographyc data.

5. <u>CONCLUSIONS</u>

Ab initio MD calculations have been carried out on models of HIV-1 protease cleavage site.

The first study, which has focused on the free enzyme, has shed light on key factors for the stability of the apartyl dyad. *Close proximity* of the two Asp groups is achieved by formation of a low barrier hydrogen between the two groups*. The coplanarity* of the two carboxy groups is instead governed by the interaction between the aspartyl dyad charge and the electric dipole moments of the Gly27 Thr26 peptide bond units. These results provide a rationale for ^{13}C NMR data[31] as well as site directed mutagenesis experiments[39] on the Gly27 position[39]. Novel force fields, based on our findings, could now be developed which accurately describe the cleavage site structural and dynamical properties. This, in turn, could lead to a more accurate computer-aided modeling of drug/enzyme interactions.

The second step of our study has focused on the protonation state and H-bond pattern on the HIV-1 PR/pepstatin complex. Smith *et al*[31] proposed that the cleavage site is negatively charged on the basis of the ^{13}C NMR chemical shifts. According to the NMR chemical shifts usually observed in aspartic acids, the measured high and low field shifts should correspond to a deprotonated and to a protonated Asp, respectively. Surprisingly, the assignment, made on the basis of isotopic substitution, suggested the reversed protonation state.
Our ab initio NMR shift calculations indicate that the pattern postulated by Smith et al is not stable. It evolves in the subps timescale to a new pattern in which the ^{13}C NMR signals are almost equal. We propose instead that the Asp dyad is *diprotonated.* Indeed, the diprotonated form turns out to be stable in the timescale investigated (5 ps) and the calculated ^{13}C NMR shifts are in much

better agreement with the experimental data[*]. For this form, the calculated isotopic effect on the chemical shift is not observed for Asp25' as it is so low to fall within the experimental error. Therefore, isotopic shifts in ^{13}C spectra are not diagnostic for assigning a specific protonation state to an Asp group.

The final step of our investigation has been the simulation of the first step of the enzymatic mechanism, that is the nucleophylic attack of water on the carbonyl carbon of the substrate. We first performed a classical MD simulation to explore the mobility of the substrate. We then carried out an ab initio MD simulation on different substrate/cleavage site distances. It turns out that only when the substrate is very close to the Asp dyad can the reactive specie (an OH$^-$ group) be stabilized through a double proton transfer. Thus, the protein frame appears to modulate the motion of the substrate in the active site whereas the Asp dyad exchanges protons with the substrate and water so as to stabilize the OH$^-$ highly nucleophylic agent.

In conclusion, our ab initio MD calculations are able to account for the observed conformational flexibility and NMR properties of the Asp dyad in the free enzyme and to describe accurately aspects of the reaction mechanism and the protonation state of an inhibitor/enzyme complex.

The main limitations of the present calculations are represented by the relatively small size of the structural models and the relatively short time-scale investigated. These limitations may offer an explanation for the significant discrepancies between calculated and experimental activation free energy and ^{13}C chemical shifts (20 kcal/mol and 4 ppm instead of ~16 kcal/mol and 6 ppm).

---

[*] The chemical inequivalence of the carboxylate carbons is a consequence of the asymmetric nature of the pepstatin inhibitor.

We may expect that for larger models, the use of QM/MM approaches[178] and a longer time-scale would improve the agreement without qualitatively changing our findings.

6. APPENDIX 1

In order to be suitable for an ab initio study, a structural model should be large enough so that environment effects do not play a major role in the properties of interest. As a preliminary step, before performing computationally costly ab initio calculations, we characterized the electrostatic properties of the cleavage site using molecular mechanics.

To address this issue we performed an analysis of the electrostatic field experienced by the water molecule inside the active site and comparison was made with the electrostatic field generated by the residues included in the simulation cell using a simple point-charge model.
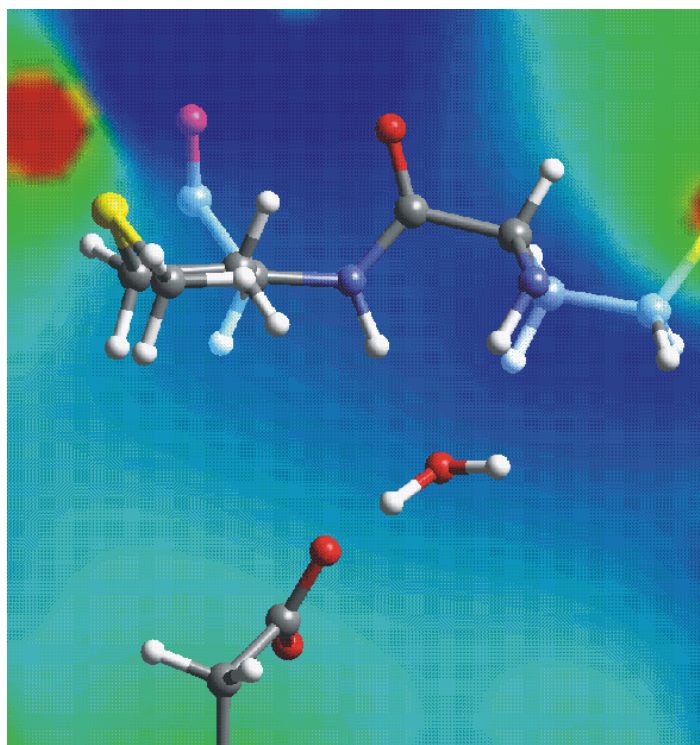

Figure 46. HIV-1 PR electric field.

Electrostatic potential generated on the active site by the charge distribution of the atoms external to the active site. Coloring goes from red (negative potential) to blue (positive potential).

It turns out that the global protein field is roughly directed along the active site-substrate axis. This field assumes similar values within the active site (Figure 46) and is mostly generated by the asymmetric distribution of the charged residues on the protein surface, thus relatively far from the active site.

Within this simple electrostatic model, it turns out that the residues not included in the ab initio calculation contribute to only ~20% to the total field experienced by the cleavage site. Thus it appears that the protein contributions to the electrostatic field inside the active site of HIV-1 PR is small. To crosscheck the results we performed calculations in the presence of the electrostatic field induced by the external protein residues on the simulation cell: it is found, consistently, that the Wannier functions displacements induced by the external field are small (see 4.1.4 and 4.3.4.1)

7. REFERENCE LIST

(1) Pantaleo, G.; Fauci, A. S. *Annu Rev Microbiol* **1996**, *50*, 825.

(2) Frankel, A. D.; Young, J. A. T. *Annu Rev Biochem* **1998**, *67*, 1.

(3) Turner, B. G.; Summers, M. F. *J Mol Biol* **1999**, *285*, 1.

(4) Debouck, C.; Gorniack, J. G.; Strickler, J. E.; Meek, T. D.; Metcalf, B. W.; Rosenberg, M. *Biochemistry* **1999**, *84*, 8903.

(5) Seelmeier, S.; Schmidt, H.; Turk, V.; von der Helm, K. *Proc Natl Acad Sci U S A* **1988** , *85*, 6612.

(6) Fitzgerald, P. M. D.; Springer, J. P. *Annu Rev Biophys Biophys Chem* **1991**, *20,* 299.

(7) Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R. A. F.; Scolnick, E. M.; Sigal, I. S. *Proc Natl Acad Sci U S A* **1988**, *85*, 4686.

(8) Debouck, C.; Gorniack, J. G.; Strickler, J. E.; Meek, T. D.; Metcalf, B. W.; Rosenberg, M. *Biochemistry* **1999**, *84*, 8903.

(9) Hong, L.; Hartsuck, J. A.; Foundling, S.; Ermolieff, J.; Tang, J. *Protein Sci* **1998**, *7*, 300.

(10) Collins, J. R.; Burt, S. K.; Erickson, J. W. *Nature Struct Biol* **1995**, *2*, 334.

(11) Chen, X.; Tropsha, A. *J Med Chem* **1995**, *38*, 42.

(12) Tropsha, A.; Hermans, J. *Protein Eng* **1992**, *5*, 29.

(13) Shaffer, L.; Verkhivker, G. M. *PROTEINS: Str Funct Gen* **1998**, *33*, 295.

(14) Weber, I. T.; Miller, M.; Jaskolski, M.; Leis, J.; Skalka, A. M.; Wlodawer, A. *Science* **1989**, *243*, 928.

(15) Hodge, C. N.; Lam, P. Y. S.; Eyermann, C. J.; Jadhav, P. K.; Ru, Y.; Fernandez, C. H.; De Lucca, G. V.; Chang, C. H.; Kaltenbach III, R. F.; Holler, E. R.; Woerner, F.; Daneker, W. F.; Emmet, G.; Calabrese, J. C.; Aldrich, P. E. *J Am Chem Soc* **1998**, *120*, 4570.

(16) Trylska, J.; Antosiewicz, J.; Geller, M.; Hodge, C. N.; Klabe, R. M.; Head, M. S.; Gilson, M. K. *Protein Sci* **1999**, *8*, 180.

(17) Harte, W. E.; Beveridge, D. L. *J Am Chem Soc* **1993**, *115*, 3883.

(18) Geller, M.; Miller, M.; Swansom, S. M.; Maizel, J. *PROTEINS: Str Funct Gen* **1997**, *27*, 195.

(19) Holloway, M. K.; Wai, J. M.; Halgren, T. A.; Fitzgerald, P. M. D.; Vacca, J. P.; Dorsey, B. D.; Levin, R. B.; Thompson, W. J.; Chen, L. J.; deSolms, S. J.; Gaffin, N.; Ghosh, A. K.; Giuliani, E. A.; Graham, S. L.; Guare, J. P.; Hungate, R. W.; Lyle, T. A.; Sanders, W. M.; Tucker, T. J.; Wiggins, M.; Wiscount, C. M.; Woltersdorf, O. W.; Young, S. D.; Darke, P. L.; Zugay, J. A. *J Med Chem* **1995**, *38*, 305.

(20) Rao, B. G.; Tilton, R. F.; Singh, U. C. *J Am Chem Soc* **1992**, *114*, 4447.

(21) Luo, X.; Kato, R.; Collins, J. R. *J Am Chem Soc* **1998**, *120* , 12410.

(22) Tawa, G. J.; Topol, I. A.; Burt, S. K.; Erickson, J. W. *J Am Chem Soc* **1998**, *120*, 8856.

(23) Okimoto, N.; Tsukui, T.; Kitayama, K.; Hata, M.; Hoshimo, T.; Tsuda, M. *J Am Chem Soc* **2000**.

(24) Harrison, R. W.; Weber, I. T. *Protein Eng* **1994**, *7*, 1353.

(25) Chatfield, D. C.; Brooks, B. R. *J Am Chem Soc* **1995**, *117*, 5561.

(26) Liu, H.; Müller-Plathe, F.; Van Gusteren, W. F. *J Mol Biol* **1996**, *261*, 454.

(27) Silva, A. M.; Cachau, R. E.; Sham, H. L.; Erickson, J. W. *J Mol Biol* **1996**, *255*, 321.

(28) Okimoto, N.; Tsukui, T.; Hata, M.; Hoshino, T.; Tsuda, M. *J Am Chem Soc* **1999**, *121*, 7349.

(29) Lee, H.; Darden, T. A.; Pedersen, L. G. *J Am Chem Soc* **1996**, *118*, 3946.

(30) Venturini, A.; López-Ortiz, F.; Alvarez, J. M.; Gonzalez, J. *J Am Chem Soc* **1998**, *120*, 1110.

(31) Smith, R.; Brereton, I. M.; Chai, R. Y.; Kent, S. B. H. *Nature Struct Biol* **1996**, *3*, 946.

(32) Katoh, E.; Yamazaki, T.; Kiso, Y.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J. D.; Torchia, D. A. *J Am Chem Soc* **1999**, *121*, 2607.

(33) Baldwin, E.; Bhat, T. N.; Gulnik, S.; Liu, B.; Topol, I. A.; Kiso, Y.; Mimoto, T.; Mitsuya, H.; Erickson, J. W. *Structure* **1995**, *3*, 581.

(34) Ridky, T. W.; Kikonyogo, A.; Leis, J.; Gulnik, S.; Copeland, T.; Erickson, J. W.; Wlodawer, A.; Kurinov, I.; Harrison, R. W.; Weber, I. T. *Biochemistry* **1998**, *37*, 13835.

(35) 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.

(36) Erickson, J. W.; Neidhart, D. J.; VanDrie, J.; Kempf, D. J.; Wang, X. C.; Norbeck, D. W.; Plattner, J. J.; Rittenhouse, J. W.; Turon, M.; Wideburg,

N.; Kohlbrenner, W. E.; Simmer, R.; Helfrich, R.; Paul, D. A.; Knigge, M. *Science* **1990**, *249*, 527.

(37) Lee, T.; Le, V.-D.; Lim, D.; Lin, Y.-C.; Morris, G. M.; Wong, A. L.; Olson, A. J.; Elder, J. H.; Wong, C.-H. *J Am Chem Soc* **1999**, *121*, 1145.

(38) Poorman, R. A.; Tomasselli, A. G.; Heinrikson, R. L.; Kezdy, F. L. *J Biol Chem* **1991** , *266*, 14554.

(39) Bagossi, P.; Cheng, Y.-S. E.; Oroszlan, S.; Tözsér, J. *Protein Eng* **1996**, *9*, 997.

(40) Rose, R.; Craik, C. S.; Douglas, N. L.; Stroud, R. M. *Biochemistry* **1996**, *35*, 12933.

(41) Loeb, D. D.; Swanstrom, R.; Everitt, L.; Manchester, M.; Stamper, S. E.; Hutchinson, C. A. *Nature* **1989**, *340*, 397.

(42) Polgár, L.; Szeltner, Z.; Boros, I. *Biochemistry* **1994**, *33*, 9351.

(43) 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.

(44) Hyland, L. J.; Tomaszek, T. A.; Meek, T. D. *Biochemistry* **1991**, *30*, 8454.

(45) Ishima, R.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J. D.; Torchia, D. A. *J Am Chem Soc* **1998**, *120*, 10534.

(46) Keinan, S.; Avnir, D. *J Am Chem Soc* **2000**, *122*, 4378.

(47) Wang, Y. X.; Freedberg, D. I.; Yamazaki, T.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J. D.; Kiso, Y.; Torchia, D. A. *Biochemistry* **1996**, *35*, 9945.

(48)   Wang, Y. X.; Freedberg, D. I.; Grzesiek, S.; Torchia, D. A.; Wingfield, P. T.; Kaufman, J. D.; Stahl, S. J.; Chang, C. H.; Hodge, C. N. *Biochemistry* **1996**, *35,* 12694.

(49)   Freedberg, D. I.; Wang, Y. X.; Stahl, S. J.; Kaufman, J. D.; Wingfield, P. T.; Kiso, Y.; Torchia, D. A. *J Am Chem Soc* **1998**, *120*, 7916.

(50)   Yamazaki, T.; Nicholson, L. K.; Torchia, D. A.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J. D.; Eyermann, C. J.; Hodge, C. N.; Lam, P. Y. S.; Ru, Y.; Jadhav, P. K.; Chang, C. H.; Weber, P. C. *J Am Chem Soc* **1994**, *116*, 10791.

(51)   Yamazaki, T.; Nicholson, L. K.; Torchia, D. A.; Stahl, S. J.; Kaufman, J. D.; Wingfield, P. T.; Domaille, P. J.; Campbell-Burk, S. *Eur J Biochem* **1994**, *219*, 707.

(52)   Grzesiek, S.; Bax, A.; Nicholson, L. K.; Yamazaki, T.; Wingfield, P. T.; Stahl, S. J.; Eyermann, C. J.; Torchia, D. A.; Hodge, C. N.; Lam, P. Y. S.; Jadhav, P. K.; Chang, C. H. *J Am Chem Soc* **1994**, *116*, 1581.

(53)   Baca, M.; Kent, S. B. H. *Proc Natl Acad Sci U S A* **1993**, *90*, 11638.

(54)   Todd, J. M.; Semo, N.; Freire, E. *J Mol Biol* **1998**, *283*, 475.

(55)   Todd, J. M.; Freire, E. *PROTEINS: Str Funct Gen* **1999**, *36*, 147.

(56)   Velasquez-Campoy, A.; Todd, J. M.; Freire, E. *Biochemistry* **2000**, *39*, 2201.

(57)   Wlodawer, A.; Erickson, J. W. *Annu Rev Biochem* **1993**, *62*, 543.

(58)   Wlodawer, A.; Vondrasek, J. *Annu Rev Biophys Biomol Struct* **1998**, *27*, 249.

(59)   Kräusslich, H. G. *Proc Natl Acad Sci U S A* **1991**, *88,* 3213.

(60) Tomasselli, A. G.; Mildner, A.; Rothrock, D. J.; Sarcich, J. L.; Lull, J.; Leone, J.; Heinrikson, R. L. *Adv Exp Med Biol* **1995**, *362*, 387.

(61) Richards, A. D.; Phylip, L. H.; Farmeire, W. G.; Scarborough, P. E.; Alvarez, A.; Dunn, B.; Hirel, P.-H.; Konvalinka, J.; Strop, P.; Pavlickova, L.; Kotska, V.; Kay, J. *J Biol Chem* **1990**, *265*, 7733.

(62) Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Clawson, L.; Schneider, J.; Kent, S. B. H. *Science* **1989**, 616.

(63) Miller, M.; Jaskolski, M.; Rao; Mohana, J. K.; Leis, J.; Wlodawer, A. *Nature* **1989**, *337*, 576.

(64) Navia, M. A.; Fitzgerald, P. M. D.; McKeever, B. M.; Leu, C.-T.; Heimbach, J. C.; Herber, W. K.; Sigal, I. S.; Darke, P. L.; Springer, J. P. *Nature* **1989**, *337*, 615.

(65) Davies, D. R. *Annu Rev Biophys Biophys Chem* **1990**, *19*, 189.

(66) Navia, M. A.; Fitzgerald, P. M. D.; McKeever, B. M.; Leu, C.-T.; Heimbach, J. C.; Herber, W. K.; Sigal, I. S.; Darke, P. L.; Springer, J. P. *Nature* **1989**, *337*, 615.

(67) Harte, W. E.; Swaminathan, S.; Mansuri, M. M.; Martin, J. C.; Rosenberg, I. E.; Beveridge, D. L. *Proc Natl Acad Sci U S A* **1990**, *87*, 8864.

(68) York, D.; Darden, T. A.; Pedersen, L. G.; Anderson, M. W. *Biochemistry* **1993**, *32*, 1443.

(69) Todd, J. M.; Freire, E. *PROTEINS: Str Funct Gen* **1999**, *36*, 147.

(70) Luque, I.; Todd, J. M.; Gomez, J.; Semo, N.; Freire, E. *Biochemistry* **1998**, *37*, 5791.

(71) Fitzgerald, P. M. D.; McKeever, B. M.; VanMiddlesworth, J. F.; Springer, J. P.; Heimbach, J. C.; Leu, C.-T.; Herber, W. K.; Dixon, R. A. F.; Darke, P. L. *J Biol Chem* **1990**, *265*, 14209.

(72) Suguna, K.; Padlan, E. A.; Smith, C. W.; Carlson, W. D.; Davies, D. R. *Proc Natl Acad Sci U S A* **1987**, *84*, 7009.

(73) McQuade, T. J.; Tomasselli, A. G.; Liu, L.; Karacostas, V.; Moss, B.; Sawyer, T. K.; Heinrikson, R. L.; Tarpley, W. G. *Science* **1990**, *247*, 454.

(74) Kaplan, A. H.; Zack, J. A.; Knigge, M.; Paul, D. A.; Kempf, D. J.; Norbeck, D. W.; Swanstrom, M. *J Virol* **1993**, *67*, 4050.

(75) Wlodawer, A.; Vondrasek, J. *Annu Rev Biophys Biomol Struct* **1998**, *27*, 249.

(76) Gulnik, S.; Erickson, J. W.; Xie, D. *Vitamins and Hormones* **2000**, *58*, 213.

(77) Vacca, J. P. Design of tight-binding human immunodeficiency virus type 1 protease inhibitors.; In *Methods in enzymology*; Kuo, L. C., Shafer, J. A., eds. Academic Press: San Diego, CA, 1994; pp 311-334.

(78) Vacca, J. P.; Condra, J. H. *Drug Des Discovery* **1997**, *2*, 261.

(79) Flexner, C. *N.Engl.J.Med.* **1998**, *338*, 1281.

(80) Erickson, J. W.; Eissenstat, M. E. HIV protease as a target for the design of antiviral agents for AIDS; In *Proteases of infectious agents*; Dunn, B., ed. Academic Press: San Diego, CA, 1999; pp 1-60.

(81) Pauling, L. *Nature* **1974**, *248*, 769.

(82) Xie, D.; Gulnik, S.; Collins, L.; Gustchina, E.; Bhat, T. N.; Erickson, J. W. *Adv Exp Med Biol* **1998**, *436*, 381.

(83) Car, R.; Parrinello, M. *Phys Rev Lett* **1985**, *55*, 2471.

(84) Brooks, C. L.; Karplus, M.; Pettitt, B. M. *Proteins: a Perspective of Dynamics, Structure, Thermodynamics.*; J Wiley and Sons: New York, 1988.

(85) Hohemberg, P.; Kohn, W. *Phys Rev B* **1964**, *136*, 864.

(86) Kohn, W.; Sham, L. J. *Phys Rev A* **1965**, *140*, 1133.

(87) Becke, A. *Phys Rev A* **1988**, *38*, 3098.

(88) Perdew, J. P. *Phys Rev B* **1986**, *33*, 8822.

(89) Lee, C.; Yang, W.; Parr, R. G. *Phys Rev B* **1988**, *37*, 785.

(90) Perdew, J. P.; Wang, Y. *Phys Rev B* **1992**, *45*, 13244.

(91) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys Rev B* **1992**, *46*, 6671.

(92) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys Rev Lett* **1996**, *77*, 3865.

(93) Hutter, J.; Carloni, P.; Parrinello, M. *J Am Chem Soc* **1996**, *118*, 8710.

(94) Sprik, M.; Hutter, J.; Parrinello, M. *J Chem Phys* **1996**, *105*, 1142.

(95) Laasonen, K.; Sprik, M.; Parrinello, M.; Car, R. *J Chem Phys* **1993**, *99*, 9080.

(96) Tuckerman, M.; Laasonen, K.; Sprik, M.; Parrinello, M. *J Chem Phys* **1995**, *103*, 150.

(97) Tuckerman, M.; Laasonen, K.; Sprik, M.; Parrinello, M. *J Chem Phys* **1995**, *105*, 1142.

(98) Laasonen, K.; Klein, M. L. *J Phys Chem A* **1997**, *101*, 98.

(99) Alber, F.; Folkers, G.; Carloni, P. *J Phys Chem* **1999**, *103* , 6121.

(100) Becke, A. *Phys Rev A* **1988**, *38*, 3098.

(101) Colle, R.; Salvetti, D. *Theor Chim Acta* **1975**, *37*, 329.

(102) Barnett, R. N.; Landman, U. *Phys Rev B* **1993**, *48*, 2081.

(103) Hockney, R. W. *Methods Comput Phys* **1970**, *9*, 136.

(104) Hellmann, H. *J Chem Phys* **1935**, *3*, 61.

(105) Hellmann, H.; Kasatotschkin, W. *J Chem Phys* **1936**, *4*, 324.

(106) Fuchs, M.; Scheffler, M. *Comput Phys Comm* **1999**, *119*, 67.

(107) Troullier, N.; Martins, J. L. *Phys Rev B* **1991**, *43*, 1943.

(108) Kleinman, L.; Bylander, D. M. *Phys Rev Lett* **1982**, *48*, 1425.

(109) Andersen, H. C. *J Comp Chem* **1983**, *52*, 24.

(110) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *J Chem Phys* **1982**, *76*, 637.

(111) Nose, S. *J Chem Phys* **1984**, *81*, 511.

(112) Wannier, G. H. *Phys Rev* **1937**, *52*, 191.

(113) Kohn, W. *Phys Rev* **1959**, *115*, 809.

(114) Marzari, N.; Vanderbilt, D. *Phys Rev B* **1997**, *56*, 12847.

(115) Silvestrelli, P. L.; Marzari, N.; Vanderbilt, D.; Parrinello, M. *Solid State Commun* **1998**, *107*, 7.

(116) Resta, R. *Phys Rev Lett* **1998**, *80*, 1800.

(117) Silvestrelli, P. L.; Parrinello, M. *Phys Rev Lett* **1999**, *82*, 3308.

(118)   Rovira, C.; Carloni, P.; Parrinello, M. *J Phys Chem B* **1999**, *103*, 7031.

(119)   Gregor, T.; Mauri, F.; Car, R. *J Chem Phys* **1999**, *111*, 1815.

(120)   Cheeseman, J. R.; Trucks, G. W.; Keith, T. A.; Frisch, M. J. *J Chem Phys* **1996**, *104*, 5497.

(121)   Keith, T. A.; Bader, R. F. W. *Chem Phys Lett* **1993**, *210*, 223.

(122)   Putrino, A, Sebastiani, D, Hutter, J, and Parrinello, M. J Chem Phys . 2000. *In Press*

(123)   Gonze, X. *Phys Rev B* **1997**, *55*, 10337.

(124)   Vignale, G.; Rasolt, M. *Phys Rev Lett* **1987**, *59*, 2360.

(125)   Lee, A. M.; Handy, N. C.; Colwell, S. M. *J Chem Phys* **1995**, *103*, 10095.

(126)   Reuben, J. *J Am Chem Soc* **1985**, *108*, 1735.

(127)   Reuben, J. *J Am Chem Soc* **1986**, *109*, 316.

(128)   Abildgaard, J.; Bolvig, S.; Hansen, P. E. *J Am Chem Soc* **1998**, *120*, 9063.

(129)   Bordner, J.; Hammen, P. D.; Whipple, E. B. *J Am Chem Soc* **1989**, *111*, 6572.

(130)   Ottiger, M.; Bax, A. *J Am Chem Soc* **1997**, *119*, 8070.

(131)   Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chem Phys Lett* **1989**, *156*, 472.

(132)   Meijer, E. J.; Sprik, M. *J Am Chem Soc* **1998**, *120*, 6345.

(133)   Curioni, A.; Sprik, M.; Andreoni, W.; Schiffer, H.; Hutter, J.; Parrinello, M. *J Am Chem Soc* **1997**, *119*, 7218.

(134)   Hirshfeld, F. *Israel J of Chem* **1977**, *16*, 198.

(135)   Mayer, I. *Chem Phys Lett* **1983**, *97*, 270.

(136)   Cornell , W. D.; Cieplack , P.; Bayly , C. I.; Gould , I. R.; Merz , K. M.;
        Ferguson , D. M.; Spellmeyer , D. C.; Fox , T.; Ca *J Am Chem Soc* **1995**,
        117, 5179.

(137)   Case, D A, Pearlman, D A, Caldwell, J W, Cheatham III, T E, Ross, W S,
        Simmerling, C L, Darden, T A, Merz, K M, Stanton, R V, Cheng, A L,
        Vincent, J J, Crowley, M F, Ferguson, D M, Radmer, R J, Singh, U C,
        Weiner, P K, and Kollman, P A. AMBER. (5.0). 1997. San Francisco
        California, University of California.

(138)   Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*; Clarendon
        Press: Oxford, 1987.

(139)   Darden, T. A.; York, D. *J Chem Phys* **1993**, *98*, 10089.

(140)   Essman, U.; Perera, L.; Berkowitz, M. L.; Darden, T. A.; Lee, H.;
        Pedersen, L. G. *J Chem Phys* **1995**, *103*, 8577.

(141)   Weerasinghe, S.; Smith, P. E.; Mohan, V.; Cheng, Y.-K.; Pettitt, B. M. *J
        Am Chem Soc* **1995**, *117*, 2147.

(142)   Duan, Y.; Kollman, P. A. *Science* **1999**, *282*, 740.

(143)   Berendsen, H. J. C.; Postma, J. P. M.; Van Gusteren, W. F.; DiNola, A.;
        Haak, J. R. *J Chem Phys* **1984**, *81*, 3684.

(144)   Armstrong, R. N. *Curr Opin Chem Biol* **1998**, *5*, 618.

(145)   Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *PROTEINS: Str
        Funct Gen* **1993**, *17*, 412.

(146) McKeever, B. M.; Navia, M. A.; Fitzgerald, P. M. D.; Springer, J. P.; Leu, C.-T.; Heimbach, J. C.; Herber, W. K.; Sigal, I. S.; Darke, P. L. *J Biol Chem* **1989**, *264*, 1919.

(147) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G. L.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Research* **2000**, *28*, 235.

(148) Beveridge, A. J.; Heywood, G. C. *Biochemistry* **1993**, *32*, 3325.

(149) Piana, S.; Carloni, P. *PROTEINS: Str Funct Gen* **2000**, *39*, 26.

(150) Csazar, P.; Pulay, P. *J Mol Stuct (Theochem)* **1984**, *114*, 31.

(151) Hartwigsen, C.; Goedecker, S.; Hutter, J. *Phys Rev B* **1998**, *58*, 3641.

(152) Goedecker, S.; Teter, M.; Hutter, J. *Phys Rev B* **1996**, *54* , 1703.

(153) Ciccotti, G.; Ferrario, M.; Hynes, J. T.; Kapral, R. *Chem Phys* **1989**, *129*, 241.

(154) Henderson, L. E.; Benveniste, R. E.; Sowder, R. C.; Copeland, T.; Schutz, A. M.; Oroszlan, S. *J Virol* **1988**, *62*, 2587.

(155) Hutter, J.; Luthi, H. P.; Parrinello, M. *Comput Mat Sci* **1994**, *2*, 244.

(156) Lapatto, R.; Blundell, T.; Hemmings, A.; Overington, J.; Wilderspin, A.; Wood, S.; Merson, J. R.; Whittle, P. J.; Danley, D. E.; Geoghegan, K. F.; Hawrylik, S. J.; Lee, S. E.; Scheld, K. G.; Hobart, P. M. *Nature* **1999**, *342*, 299.

(157) Kempf, D. J.; Marsh, K. C.; Denissen, J. F.; McDonald, E.; Vasavanonda, S.; Flentge, C. A.; Green, B. E.; Fino, L.; Park, C. H.; Kong, X. P.; Wideburg, N.; Saldivar, A.; Ruiz, L.; Kati, W. M.; Sham, H. L.; Robins, T.; Stewart, K. D.; Hsu, A.; Plattner, J. J.; Leonard, J. M.; Norbeck, D. W. *Proc Natl Acad Sci U S A* **1995**, *92*, 2484.

(158) Bertolasi, V.; Gilli, P.; Ferretti, V.; Gilli, G. *J Am Chem Soc* **1991**, *113*, 4917.

(159) Carloni, P.; Roethlisberger, U. Simulations of enzymatic systems: perspectives from Car-Parrinello molecular dynamics simulations; In *Theoretical biochemistry - processes and properties of biological systems.*; Ericksson, L., ed. Elsevier Science: New York, 2000.

(160) Meijer, E. J.; Sprik, M. *J Phys Chem* **1998**, *102*, 2893.

(161) Hagen, R.; Roberts, J. D. *J Am Chem Soc* **1969**, *91*, 4504.

(162) Farcasiu, D.; Hancu, D. *J Phys Chem A* **1999**, *103*, 754.

(163) Kubicki, J. D.; Sykes, D.; Apitz, S. E. *J Phys Chem A* **1999**, *103*, 903.

(164) Sykes, D.; Kubicki, J. D.; Farrar, T. C. *J Phys Chem A* **1997**, *101*, 2715.

(165) Rodriguez-Fortea, A.; Alemany, P.; Ziegler, T. *J Phys Chem* **1999**, *103*, 8288.

(166) Valerio, G.; Goursot, A. *J Phys Chem B* **1999**, *103*, 51.

(167) Mauri, F.; Pfrommer, B. G.; Louie, S. G. *Phys Rev Lett* **1996**, *77*, 5300.

(168) Mauri, F.; Pfrommer, B. G.; Louie, S. G. *Phys Rev Lett* **1997**, *79*, 2340.

(169) Pfrommer, B. G.; Mauri, F.; Louie, S. G. *J Am Chem Soc* **2000**, *122*, 123.

(170) Streitwieser, A.; Heatcock, C. H.; Kosower, E. M. *Introduction to organic chemistry*; MacMillian Publishing Company: New York, 1992.

(171) Amadei, A.; Ceruso, M. A.; Di Nola, A. D. *PROTEINS: Str Funct Gen* **1999**, *36*, 419.

(172) Wiberg, K. B.; Laidig, K. E. *J Am Chem Soc* **1987**, *109*, 5935.

(173) Wiberg, K. B.; Breneman, C. M. *J Am Chem Soc* **1992**, *114*, 831.

(174) Prabu-Jeyabalan, M.; Nalivaika, M.; Schiffer, C. *J Mol Biol* **2000**, *301*, 1207.

(175) James, M. N. G.; Sielecki, A. R.; Hayakawa, K.; Gelb, M. H. *Biochemistry* **1992**, *31*, 3872.

(176) Veerapandian, B.; Cooper, J. B.; Sali, A.; Blundell, T.; Rosati, R. L.; Dominy, B. W.; Damon, D. B.; Hoover, D. J. *Protein Sci* **1992**, *1*, 322.

(177) Warshel, A. *J Biol Chem* **1998**, *273*, 27035.

(178) Rothlisberger, U.; Carloni, P.; Docklo, K.; Parrinello, M. *J Biol Inorg Chem* **2000**, *5*, 236.

(179) Fisher, A. G.; Ensoli, B.; Ivanoff, L.; Chamberlain, M.; Petteway, S.; Ratner, L.; Gallo, R. C.; Wong-Staal, F. *Science* **1999**, *237*, 888.

(180) Strebel, K.; Daugherty, D.; Clouse, K.; Cohen, D.; Folks, T.; Martin, T. A. *Nature* **1986**, *28*, 728.