J|A|C|S
ARTICLES
Published on Web 02/26/2005

Evolutionarily Conserved Functional Mechanics across Pepsin-like and Retroviral Aspartic Proteases

Michele Cascella,[†,‡] Cristian Micheletti,[†] Ursula Rothlisberger,[‡] and Paolo Carloni*[,†]

Contribution from the International School for Advanced Studies (SISSA/ISAS) and INFM—DEMOCRITOS Modeling Center for Research in Atomistic Simulation, Via Beirut 2−4, 34014 Trieste, Italy, and Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

Received September 6, 2004; E-mail: carloni@sissa.it

Abstract: The biological function of the aspartic protease from HIV-1 has recently been related to the conformational flexibility of its structural scaffold. Here, we use a multistep strategy to investigate whether the same mechanism affects the functionality in the pepsin-like fold. (i) We identify the set of conserved residues by using sequence-alignment techniques. These residues cluster in three distinct regions: near the cleavage-site cavity, in the four β-sheets cross-linking the two lobes, and in a solvent-exposed region below the long β-hairpin in the N-terminal lobe. (ii) We elucidate the role played by the conserved residues for the enzymatic functionality of one representative member of the fold family, the human β-secretase, by means of classical molecular dynamics (MD). The conserved regions exhibit little overall mobility and yet are involved into the most important modes of structural fluctuations. These modes influence the substrate−catalytic aspartates distance through a relative rotation of the N- and C-terminal lobes. (iii) We investigate the effects of this modulation by estimating the reaction free energy at different representative substrate/enzyme conformations. The activation free energy is strongly affected by large-scale protein motions, similarly to what has been observed in the HIV-1 enzyme. (iv) We extend our findings to all other members of the two eukaryotic and retroviral fold families by recurring to a simple, topology-based, energy functional. This analysis reveals a sophisticated mechanism of enzymatic activity modulation common to all aspartic proteases. We suggest that aspartic proteases have been evolutionarily selected to possess similar functional motions despite the observed fold variations.

1. Introduction

Aspartic proteases (APs) constitute a ubiquitous class of enzymes which use an Asp dyad to hydrolyze peptide bonds.[1−3] The known structures of APs divide in two distinct fold families.[4,5] APs found in eukaryotic cells are α/β monomers composed of two asymmetric lobes (Figure 1, top), with the catalytic Asp dyad located at the lobe interface and a flap made up of a β-hairpin covering the peptide substrates[6] ("pepsin-like" proteases). Retroviral enzymes, on the other hand, are β homodimers (Figure 1, bottom), where the aspartates are located on two loops at the monomer interface and where two β-hairpins cover the active site. Despite the structural differences and the low sequence identity,[4] it is believed that the pepsin-like and retroviral protease families are evolutionarily related since, in both folds, the cleavage site loops are homologous, the Asp dyad is located at an interface region, and the viral subunits are structurally similar to the N-terminal lobes of the pepsin-like family enzymes (also called "pepsins", hereafter)[7,8] (Figure 1). In the far most characterized viral enzyme, HIV-1 AP,[9] we have recently suggested that conformational fluctuations play a major role in the enzymatic catalysis. In fact, the enzyme acts as sophisticated machinery capable of steering the substrate toward the catalytic Asp dyad to favor a reactive conformation.[10−12] Similar flexibility assisted mechanisms have been recently suggested for other enzymes,[13] including flavin reductase,[14] dihydrofolate reductase,[15] acyl CoA dehydrogenase,[16] orotidine 5′-monophosphate decarboxylase,[17] tyrosyl-tRNA synthetase,[18] and xylose isomerase,[19] as well as in mitogenic signaling processes.[20]

† SISSA/ISAS.
‡ EPFL.

(1) Northrop, D. B. *Acc. Chem. Res.* **2001**, *34*, 790−797.
(2) Davies, D. R. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 189−215.
(3) Simoes, I.; Faro, C. *Eur. J. Biochem.* **2004**, *271*, 2067−2075.
(4) Rawlings, N. D.; Barrett, A. J. *Methods Enzymol.* **1995**, *248*, 105−120.
(5) Dunn, B. M. *Chem. Rev.* **2002**, *102*, 4431−4458.
(6) Sielecki, A. R.; Fujinaga, M.; Read, R. J.; James, M. N. G. *J. Mol. Biol.* **1991**, *219*, 671.
(7) Tang, J.; James, M. N. G.; Hsu, I. N.; Jenkins, J. A.; Blundell, T. L. *Nature* **1978**, *271*, 618.
(8) Blundell, T. L.; Srinivasan, N. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 14243−14248.
(9) Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L. M.; Kent, S. B. H.; Wlodawer, A. *Science* **1989**, *246*, 1149−1152.
(10) Piana, S.; Carloni, P.; Parrinello, M. *J. Mol. Biol.* **2002**, *319*, 567−583.
(11) Piana, S.; Carloni, P.; Rothlisberger, U. *Protein Sci.* **2002**, *11*, 2393−2402.
(12) Perryman, A. L.; Lin, J. H.; McCammon, J. A. *Protein Sci.* **2004**, *13*, 1108−1123.
(13) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186−195.
(14) Yang, H.; Luo, G.; Karnchanaphanurach, P.; Louie, T. M.; Rech, I.; Cova, S.; Xun, L.; Xie, X. S. *Science* **2003**, *302*, 262−266.
(15) Rod, T. H.; Radkiewicz, J. L.; Brooks, C. L., III. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6980−6985.

10.1021/ja044608+ CCC: $30.25

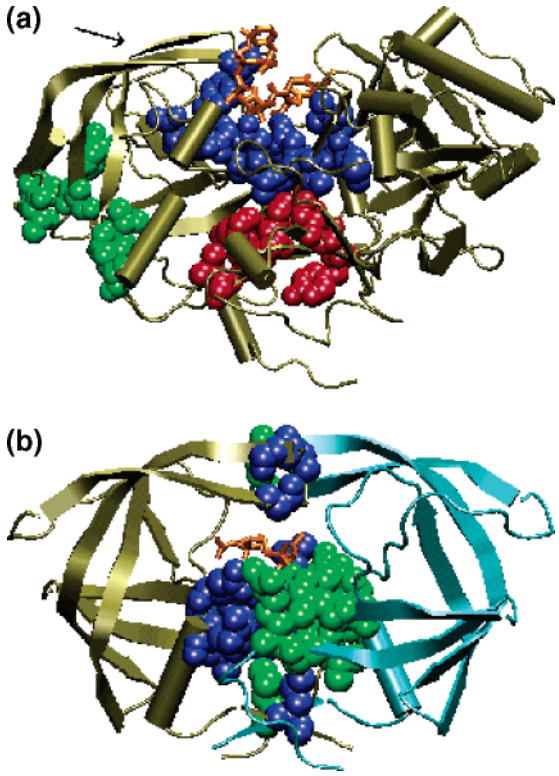

Figure 1. Pepsin-like and retroprotease folds. The structures of BACE and HIV-1 AP are shown in complex with model substrates (orange sticks). The conserved regions are drawn as spheres: in (a), regions **a**−**c** (defined in the text) are colored in blue, red, and green, respectively; the black arrow denotes the flap position. In (b), conserved residues are colored in blue and green.

The key question we pose here is whether mechanical fluctuations are also crucial for the function of all the other pepsin family members (and retroviral proteases other than HIV-1 AP). Structural data point to the presence of large-scale motion of the protein, although their role for the catalytic action is not yet clear: Marcinkeviciene et al.[21] suggested that in porcine pepsin/inhibitor complexes, different conformational isomers could be at equilibrium. Moreover, in plasmepsin 2 from *Plasmodium falciparum*,[22] different images in the crystallographic structure have revealed a rotation of the C-terminal domain with respect to the N-terminal one of 5.2°.[22]

To address this issue, we adopt here a multistep computational strategy. First, we identify the conserved residues within the pepsin-like and retroviral protease families through multiple-sequence alignment; then, we focus on one member of the pepsin-like family, β-secretase (BACE), which is a pharmacologically relevant target for Alzheimer's disease.[23,24] To ascertain

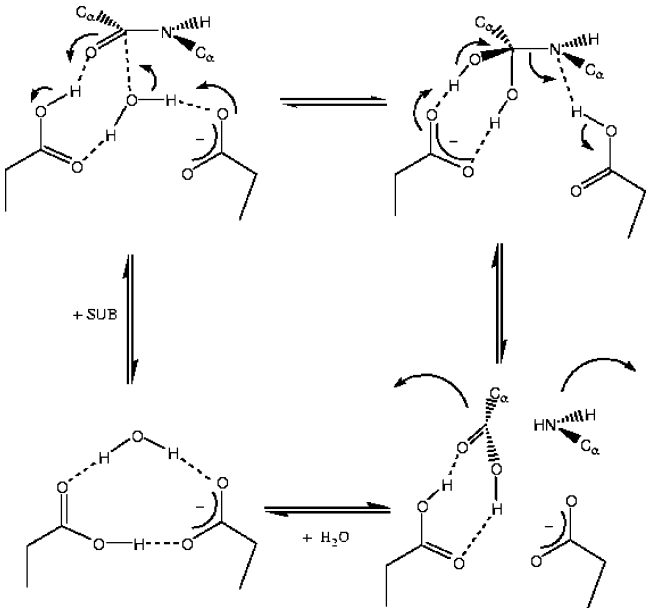

Figure 2. General reaction mechanism of APs. A general reaction mechanism for the chemical catalytic steps of APs. Starting from the lower-left angle and following the reaction clockwise: aspartic dyad in the free form, binding of the substrate, and nucleophilic attack of water; formation of the tetrahedral *gem*-diol intermediate; protonation of the nitrogen atom and formation of the products; release of the products and regeneration of the catalyst.

whether these residues have an important functional role, we perform 20 ns MD simulations on the enzyme in complex with a model substrate, identifying conformational changes in the complex which affect the structure of the active site. In particular, the substrate−cleavage site distance shows two characteristic regimes, which interchange into one another in a multinanosecond time scale. This bimodal distribution is related to two different active-site conformations which are distinguished in terms of hydrogen-bond pattern among the substrate, the aspartic dyad, and the catalytic water. Subsequently, we perform hybrid Car-Parrinello[25]/molecular mechanics (QM/MM)[26,27] simulations to estimate the enzymatic reaction barrier in both conformations.

At present, all the mechanical details of the catalytic action are not clear yet. Indirect observations on inhibitor complexes[28,29] or kinetic measurements[1] agree on the fact that the hydrolysis reaction should proceed in two chemical steps of similar free energy barriers[1,5,30] (exp. $k_{cat} \approx 18$ kcal/mol[24]). All proposals put forward so far indicate that as the first chemical step of the reaction,[1,5,10,30] an Asp group at the cleavage site should act as a base by deprotonating the catalytic water (Figure 2). In the second step, the Asp dyad yields a proton to the N-atom of the substrate, inducing the breaking of the C−N bond. The hydrolysis reaction should be preceded and followed by a

(16) Alhambra, C.; Corchado, J.; Sanchez, M. L.; Garcia-Viloca, M.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. B* **2001**, *105*, 11326−11340.
(17) Radzicka, A.; Wolfenden, R. *Science* **1995**, *267*, 90−93.
(18) Winter, G.; Fersht, A.; Wilkinson, A. J.; Zoller, M.; Smith, M. *Nature* **1982**, *299*, 756−758.
(19) Garcia-Viloca, M.; Alhambra, C.; Truhlar, D. G.; Gao, J. *J. Am. Chem. Soc.* **2002**, *124*, 7268−7269.
(20) Suenaga, A.; Kiyatkin, A. B.; Hatakeyama, M.; Futatsugi, N.; Okimoto, N.; Hirano, Y. et al. *J. Biol. Chem.* **2004**, *279*, 4657−4662.
(21) Marcinkeviciene, J.; Kopcho, L. M.; Yang, T.; Copeland, R. A.; Glass, B. M.; Combs, A. P.; Falahatpisheh, N.; Thompson, L. *J. Biol. Chem.* **2002**, *277*, 28677−28682.
(22) Silva, A. M.; Gulnik, S. V.; Majer, P.; Collins, J.; Baht, T. N.; Collins, P. J.; Cachau, R. E.; Luker, K. E.; Gluzman, I. Y.; Francis, S. E.; Oksman, A.; Goldberg, D. E.; Erickson, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10034−10039.
(23) Selkoe, D. J. *Nature* **1999**, *399A*, 23−31.
(24) Lin, X.; Koelsch, G.; Wu, S.; Downs, D.; Dashti, A.; Tang, J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 1456−1460.
(25) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471−2474.
(26) Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941−6947.
(27) Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Phys. Chem. B* **2002**, *106*, 7300−7307.
(28) James, M. N. G.; Sielecki, A. R.; Hayakawa, K.; Gelb, M. H. *Biochemistry* **1992**, *31*, 3872−3886.
(29) Veerapandian, B.; Cooper, J. B.; Sali, A.; Blundell, T. L.; Rosati, R. L.; Dominy, B. W.; Damon, D. B.; Hoover, D. J. *Protein Sci.* **1992**, *1*, 322−328.
(30) Piana, S.; Bucher, D.; Carloni, P.; Rothlisberger, U. *J. Phys. Chem. B* **2004**, *108*, 11139−11149.

series of physical steps that should be able to re-form the catalyst, but over which a full consensus has not been found yet.[1,5,31−33]

In this work, we focus on the first chemical step, namely, the nucleophilic attack of water to the carbonyl carbon, over which there is a general agreement.[5] Our calculations suggest that, during this step, the enzyme is able to catalyze the reaction only in one of the two geometries highlighted before, similar to what was found in the viral isoenzyme.

The protocols used in our classical and QM/MM simulations have already proven to be reliable to address these issues, as they were able to rationalize spectroscopic and kinetic experimental data in HIV-1 protease,[1,34,35] as well as in metalloproteases.[36]

Finally, we attempt to extend our conclusions to all pepsin-like enzymes by elucidating the functional role played by these conserved residues. To accomplish this latter step, we have characterized the near-native protein fluctuations by using a computationally efficient coarse-grained approach.[37] It is found that regions of BACE involved in conformational changes relevant for biological activity are conserved both at the level of sequence and at the structural organization, across other members of the pepsin family, also showing interesting analogies with retroviral proteases.

2. Methods

2.1. Multiple-Sequence Alignment. A Hidden Markov Model search in the sequence database,[38] performed at the UCSC server (www.cse.ucsc.edu/research/compbio/HMM-apps), allowed alignment of the sequence of BACE to other members of the pepsin family, up to a final number of 868 sequences (alignment available at www.sissa.it/sbp/bc/). The multiple alignment procedure was repeated for members of the retroviral protease family, starting from the HIV-1 AP (PDB code: 1A30)[39] sequence.

2.2. BACE. Structural Model. Our model of the BACE/EVNLAAEF substrate complex was obtained by replacing the hydroxyl−vinyl group of the OM99-2 inhibitor with the amide group in the X-ray structure of BACE (PDB code: 1FKN; X-ray structure, 1.9 Å resolution[40]). One of the two catalytic aspartates (Asp32) was protonated, following refs 41 and 42, while other ionizable groups were assigned their charged forms. Histidines were protonated according to their putative H-bond pattern in the X-ray structure.

Classical MD Simulations. The protein was immersed in a water box with a size of 75 × 87 × 90 Å^3. Overall charge neutrality was achieved by adding nine sodium counterions. The whole system, composed of about 48 000 atoms, underwent 20 ns of MD at 300 K, and 1 atm pressure simulations were carried out with the GROMACS program.[43,44] The Amber force field (Parm98[45]) was used to describe the protein and the counterions, while the TIP3P model was used for water.[46] Particle mesh Ewald routines were used to treat long-range electrostatic interactions.[47,48] A cutoff of 12 Å was used for the van der Waals interactions and the real part of the electrostatic interactions. NPT ensemble was simulated by a Nose−Hoover chain of thermostats[49−51] and the Parrinello−Rahman[52,53] scheme. The LINCS constraint algorithm[54] applied on bond distances allowed a time step of 1.5 fs. Large-scale motions were calculated by principal modes analysis, through diagonalization of the C_α covariance matrix.[55] The statistical significance of the MD conformational sampling was checked through the measure of the cosine content in the principal modes, as recently proposed by Hess.[56,57] Hinge regions for modes were identified by inspecting the spacial location of the 40 residues with the smallest values in the corresponding entry of the eigenvector square module.

QM/MM Calculations. Two selected MD snapshots, in which the distance between substrate and cleavage site (ξ in Figure 6) was 7.6 and 8.0 Å, respectively, were used as starting points for QM/MM simulations. The protocol used was similar to that in refs 58 and 59. The system was partitioned into (i) a QM region, described at the DFT/Car−Parrinello level as in refs 60 and 61, comprising Asp32 and Asp228 side chains, the Leu-Ala substrate peptide, the catalytic water, Ser35 and Thr231 side chains, and Thr33-Gly34 and Ser229-Gly230 peptide backbones; (ii) a MM region, comprising the rest, treated as above.

The free energy of the reaction was computed by thermodynamic integration,[62] using as the approximate unidimensional reaction coordinate (ζ_{CO}) the distance between the oxygen atom of the catalytic water and the carbon atom of the peptide bond of the substrate. For both systems, ζ_{CO} was fixed at increasingly shorter distances, from the equilibrium distance up to the formation of the C−O bond. For each computed distance, about 2 ps of QM/MM simulations was carried out to achieve an acceptable convergence of the force on the constraint. This methodology allows us to follow dynamics in the active site without choosing any reaction mechanism "a priori". As atoms are allowed to rearrange freely in the space, the most favorable reaction mechanism can be "a posteriori" verified.

2.3. Coarse-Grained Mechanical Model. Because the MD calculations were computationally too demanding, we have used the recently developed Beta-Gaussian model[37] to calculate large-scale motions of pepsins other than BACE, namely, pepsins from *Candida tropicalis* (PDB code: 1J71),[63] *Candida albicans* (PDB code: 1EAG),[64] Plasmodium (PDB code: 1LEE)[65] and human cathepsin D (PDB code: 1LYA),[66] and uropepsin (PDB code: 1FLH),[67] which share a low

(31) Meek, T. D.; Rodriguez, E. J.; Angeles, T. S. *Methods Enzymol.* **1994**, *241*, 127−156.
(32) Polgar, L.; Szeltner, Z.; Boros, I. *Biochemistry* **1994**, *33*, 9351−9357.
(33) Porter, D. J. T.; Hanlon, M. H.; Furfine, E. S. *Biochemistry* **2002**, *41*, 1302−1307.
(34) Piana, S.; Carloni, P. *Proteins* **2000**, *39*, 26−36.
(35) Piana, S.; Sebastiani, D.; Carloni, P.; Parrinello, M. *J. Am. Chem. Soc.* **2001**, *123*, 8730−8737.
(36) Dal Peraro, M.; Llarrull, L. I.; Rothlisberger, U.; Vila, A. J.; Carloni, P. *J. Am. Chem. Soc.* **2004**, *126*, 12661−12668.
(37) Micheletti, C.; Carloni, P.; Maritan, A. *Proteins* **2004**, *55*, 635−645.
(38) Karplus, K.; Barrett, C.; Hughey, R. *Bioinformatics* **1998**, *14*, 846−856.
(39) Louis, J. M.; Dyda, F.; Nashed, N. T.; Kimmel, A. R.; Davies, D. R. *Biochemistry* **1998**, *37*, 2105.
(40) Hong, L.; Koelsch, G.; Lin, X.; Wu, S.; Terzyan, S.; Ghosh, A. K.; Zhang, X. C.; Tang, J. *Science* **2000**, *290*, 150−153.
(41) Beveridge, A. *Proteins* **1996**, *24*, 322−334.
(42) Park, H.; Lee, S. *J. Am. Chem. Soc.* **2003**, *125*, 16416−16422.
(43) Berendsen, H. J. C.; van der Spoel, D.; Vandrunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43−56.
(44) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306−317.
(45) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(46) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(47) Darden, T.; York, D.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(48) Essman, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(49) Nosé, S. J. *J. Chem. Phys.* **1984**, *81*, 511−519.
(50) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695−1697.
(51) Martyna, G.; Tuckerman, M.; Klein, M. L. *J. Chem. Phys.* **1992**, *97*, 2635−2643.
(52) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(53) Nosé, S. J.; Klein, M. L. *Mol. Phys.* **1983**, *50*, 1055−1076.
(54) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(55) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins* **1993**, *17*, 412−425.
(56) Hess, B. *Phys. Rev. E* **2000**, *62*, 8438−8448.
(57) Hess, B. *Phys. Rev. E* **2002**, *65*, 031910.
(58) Sulpizi, M.; Laio, A.; VandeVondele, J.; Cattaneo, A.; Rothlisberger, U.; Carloni, P. *Proteins* **2003**, *52*, 212−224.
(59) Mordasini, T.; Curioni, A.; Andreoni, W. *J. Biol. Chem.* **2003**, *278*, 4381−4384.
(60) Martinez, G. R.; Mascio, P. D.; Bonini, G. M.; Augusto, O.; Briviba, K.; Sies, H.; Maurer, P.; Rothlisberger, U.; Herold, S.; Koppenol, W. H. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10307−10312.
(61) Sulpizi, M.; Schelling, P.; Folkers, G.; Carloni, P.; Scapozza, L. *J. Biol. Chem.* **2001**, *276*, 21692−21697.
(62) Carter, E. A.; Ciccotti, G.; Heynes, J. T.; Kapral, R. *Chem. Phys. Lett.* **1989**, *156*, 472−477.

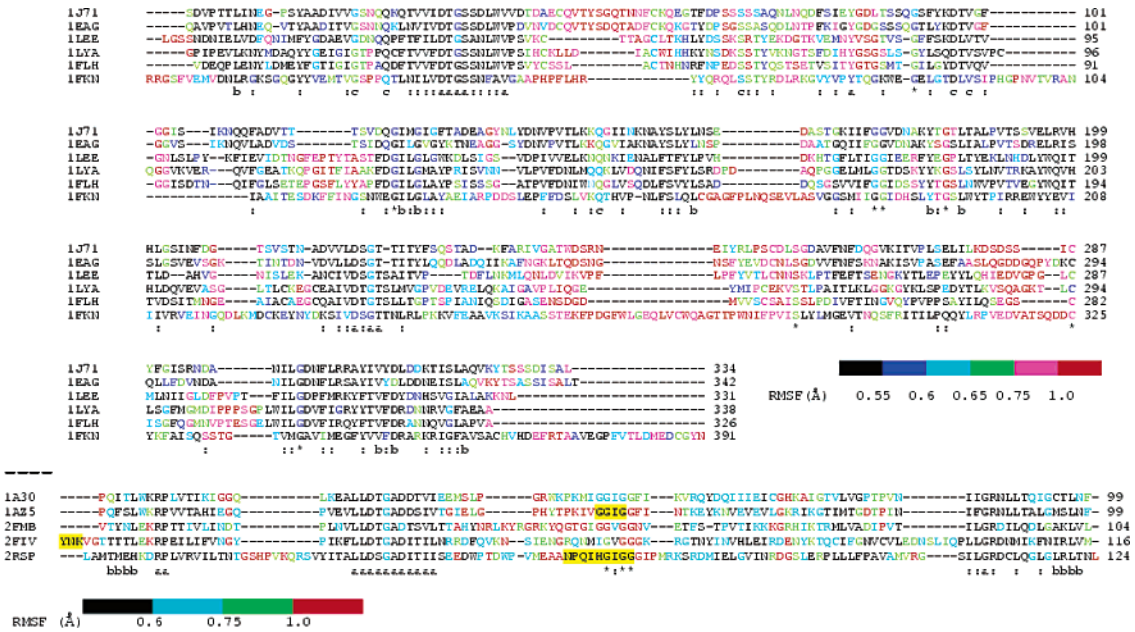

Figure 3. Sequence alignment and residue mobility of APs. Top: sequences of pepsin-like aspartic proteases from *Candida tropicalis* (PDB code: 1J71),[63] *Candida albicans* (1EAG),[64] Plasmodium (1LEE)[65] and human cathepsin D (1LYA),[66] uropepsin (1FLH),[67] and BACE (1FKN).[40] Bottom: viral proteases from human immunodeficiency virus 1 (1A30),[9] simian immunodeficiency virus (1AZ5),[76] equine anemia virus (2FMB),[77] feline immunodeficiency virus (2FIV),[78] and Rous sarcoma virus (2RSP).[79] Conserved residues in regions structurally homologous to regions **a** and **b** in pepsins are labeled by a and b, respectively, conserved residues not belonging to these regions by asterisks, and conservative mutations by columns. Conserved residues belonging to regions **a**−**c** are marked by a, b, and c, respectively, conserved residues not belonging to these regions by asterisks, and conservative mutations by columns. Residues are colored according to their root mean square fluctuations (rmsf), based on the MD simulations (BACE, this work HIV-1 AP, ref 11) or coarse-grained computations (all the others, ref 37).

sequence identity among themselves (about 10%). This scheme adopts a simplified structural description where each amino acid is represented by a C_α and (except for glycines) a C_β centroid. As for other Gaussian network models (see refs 68 and 69 and references therein), the energy function consists of harmonic couplings between pairs of spatially close centroids which allow decomposition of the protein vibrational mechanics in terms of independent modes.[70] This model allows the use of different harmonic couplings for covalent bonds (e.g., peptide and/or disulfide bonds), $C_\alpha-C_\alpha$, $C_\beta-C_\alpha$, and $C_\beta-C_\beta$ interactions. Previous studies on HIV-1 AP and NGF-trkA have shown that the resulting covariance matrix agrees well with the one obtained from all-atom MD simulation.[37] The package implementing the Beta-Gaussian model is made freely available upon request.

3. Results

Our study is carried out as follows. First, we find the most conserved regions of pepsins and viral proteases by means of sequence-alignment techniques. To reveal the possible involvement of these residues in the functional mechanics, we recourse to classical MD. Within this approach, we identify the large-scale motions of BACE and then investigate the enzymatic reaction by hybrid Car−Parrinello ab initio MD/classical MD simulation. Finally, by adopting a simple topology-based energy functional, we establish that the mechanical response of BACE not only is shared by arbitrary members of the pepsin family but also shares important traits with the functional mechanics of retroviral proteases.

3.1. Multiple Alignment of Families of APs. Pepsins. The alignment of sequences of the pepsin family shows that overall ∼20% of the residues either are conserved or exhibit highly conservative mutations (Figure 3, top). Most of these residues (81%) belong to three regions (Figure 1, top).

As expected, one region (**a** in Figures 1, 3) is the cleavage site region, comprising the two aspartate-carrying loops and the β-sheets located at their edges, localized at the lobes' interface. Region **b** is made up of the four antiparallel β-sheets located opposite the substrate binding pocket, with respect to the cleavage site. This is the only structured region that cross-links the two lobes of the enzyme. Finally, region **c** is localized on the surface of the N-terminal lobe, approximately perpendicular to the flap ends. It comprises mostly polar residues involved in specific contacts on the protein surface. In particular, the conserved Ser57, Thr59, and Asp83 residues (Figure 3), which are bridged by an H-bond, are located in the loop preceding the beginning of the first β-sheet of the flap and at the other β-sheet end.

Retroviral Proteases. Until now, only six aspartic proteases have been characterized from retroviruses.[71] On the basis of this relatively small database, one can establish that only the residues located at the interface, which stabilize the dimeric structure, are conserved[72] (Figures 1 and 3). This conserved region shares structurally conserved regions with pepsins, namely, the cleavage site region, with the two Asp carrying

(63) Symersky, J.; Monod, M.; Foundling, S. I. *Biochemistry* **1997**, *36*, 12700.
(64) Cutfield, S. M.; Dodson, E. J.; Anderson, B. F.; Moody, P. C.; Marshall, C. J.; Sullivan, P. A.; Cutfield, J. M. *Structure* **1995**, *3*, 1261.
(65) Asojo, O. A.; Afonina, E.; Gulnik, S. V.; Yu, B.; Erickson, J. W.; Randad, R.; Mehadjed, D.; Silva, A. M. *Acta Crystallogr. D* **2002**, *58*, 2001.
(66) Baldwin, E. T.; Bhat, T. N.; Gulnik, S.; Hosur, M. V.; Sowder, R. C., II; Cachau, R. E.; Collins, J.; Silva, A. M.; Erickson, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6796−6800.
(67) Canduri, F.; Teodoro, L. G. V. L.; Fadel, V.; Lorenzi, C. C. B.; Hial, V.; Gomes, R. A. S.; Neto, J. R.; de Azevedo, W. F. *Acta Crystallogr. D* **2001**, *57*, 1560−1570.
(68) Noguti, T.; Go, N. *Nature* **1982**, *296*, 776−778.
(69) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. *Biophys. J.* **2001**, *80*, 505−515.
(70) Levitt, M.; Sander, C.; Stern, P. S. *J. Mol. Biol.* **1985**, *181*, 423.
(71) Wlodawer, A.; Gustchina, A. *Biochim. Biophys. Acta* **2000**, *1477*, 16−34.
(72) Weber, I. T. *Gene* **1989**, *85*, 565−566.

loops (**a**) and the four antiparallel β-sheets that cross-link the two subunits (**b**). This last region is slightly modified from the one present in pepsins, as the strand alternates sheets belonging to the two subunits, while in pepsins, this does not occur. The flap region, on the contrary, is structurally different from that of pepsins since due to the homodimeric structure, there are two flaps present rather than the one in pepsins.

3.2. BACE Conformational Fluctuations. The structure of BACE complexed with the model substrate equilibrates after 0.8 ns of dynamics. After this time interval, the root mean square deviation (rmsd) of the backbone atoms fluctuates around an average value of 1.5 Å with respect to the X-ray structure (see the Supporting Information).

The overall fold of the protein is well maintained. Root mean square fluctuation (rmsf) calculations are in qualitative agreement with experimental B-factors, and they show that the protein core is rather rigid whereas the solvent-exposed loops are mobile (rmsf > 1.0 Å). The main difference with experimental B-factors is in the Thr254-Pro258 loop, which in our simulations, is found to be more mobile than in the X-ray structure. This difference may be due to crystal packing effects, as in the X-ray structure, this loop is in direct contact (interatomic distances <3.0 Å) with the pro-sequence in the other protein image. The rmsf plot shows that the conserved regions at the interface of the lobes (**a** and **b**) are rigid (rmsf ≈ 0.5 Å).

The flap on the N-terminal lobe is also rather rigid (rmsf (always) < 1.0 Å), especially in the turn over the binding pocket (rmsf ≈ 0.5 Å). This fact marks a difference with MD simulations of HIV-1 AP, where the flap had shown a larger mobility. The conserved region at the basis of the flap (**c**) is also characterized by a relatively low mobility (rmsf ≈ 0.6 Å).

Almost all of the residues of the substrate conserve, on average, the contacts with the protein described in the crystal structure,[40] including Gly11−N-terminal substrate (d(O−N) ≈ 2.8 Å), typical of BACE, or H-bonds between the backbone of the substrate and the protein: N@Val_{P_3}−O_γ@Thr232 ≈ 2.9 Å, O@Val_{P_3}−N@Thr232 ≈ 2.9 Å, O@Asn_{P_2}−N@Gln73 ≈ 3.0 Å, N@Leu_{P_1}−O@Gly230 ≈ 2.9 Å, O@$Ala_{P'_1}$−N@Thr72 ≈ 2.9 Å, N@$Ala_{P'_2}$−O@Gly34 ≈ 3.1 Å.

Instead, the C-terminal Glu-Phe residues do not conserve the contacts and are rather mobile (rmsf ≈ 2.5 Å), consistent with the fact that these residues have a different conformation in the two images of the crystal.[40] Notice that, instead, in the structure of BACE in complex with the OM00−3 inhibitor (at 2.1 Å resolution, PDB code: 1M4H[73]), the location of the side chain of these residues is unambiguously identified.

At the active site, Asp32 has been protonated, and Asp228 is ionized, following refs 10, 28, 29, and 41. The hydroxyl−vinyl group, chemically mutated into a peptide bond, is H-bonded with its carbonyl either to Asp32 or to the catalytic water, while the N-atom is hydrated by a water molecule in the binding cleft. The deviation from coplanarity of the Asp dyad is more pronounced than in other ligand-bound pepsin structures (∼50°[40]). Both Thr33-Gly34 and Ser229-Gly230 backbone amide groups interact preferably with Asp228, while Asp32 is only H-bonded to the Ser35 side chain. This deviation from the coplanarity of the two aspartates is kept during the MD run.

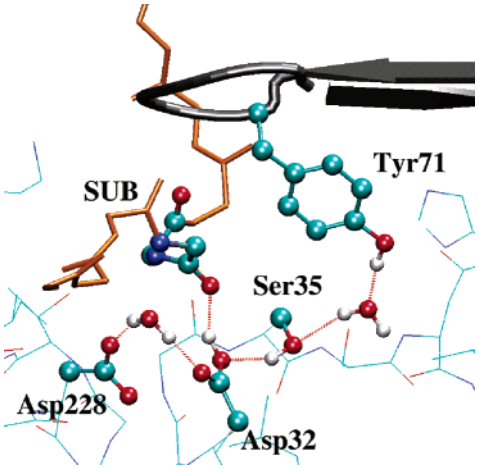

Figure 4. Tyr71 H-bond pattern in the active site of BACE. Tyr71, Ser35, Asp32, and Asp228 side chains, an ordered (buried, crystallographic) water molecule, and the catalytic water molecule are shown in balls-and-sticks. Only hydrogens bound to polar groups are shown for clarity. The substrate backbone is represented in orange cylinders, except for Leu_{P_1} and $Ala_{P'_1}$ atoms, which are drawn as spheres; the flap is represented by a cartoon. The rest of the protein is drawn in lines. This H-bond pattern is conserved in all pepsins investigated here.

The largest displacements of the system can be rationalized by analysis of the covariance matrix[55,74] and through the principal modes obtained by its diagonalization.[55,74] Two main features emerge from this analysis. First, the motion of the fully conserved Tyr71[5] is correlated to that of the cleavage site, possibly because it is located opposite the cleavage site with respect to the substrate, and it is directly linked to it through a buried water detected in the X-ray structure[40] and the side chain of Ser35 (Figure 4). The H-bond pattern featured by Tyr71 might provide an efficient way for clamping the substrate (Figure 4). The importance of this structural motif for the enzymatic catalysis has been already suggested by Andreeva and Rumsh,[75] although these authors proposed a different H-bond pattern. Second, and most importantly for the present discussion, the substrate motion at the cleavage site turns out to be highly correlated to its neighbors and particularly with region a (despite its low mobility) and the turn of the substrate-binding flap (see Figure 5).

The three largest eigenvectors of the covariance matrix are dominant with respect to the others. The first one is as large as 46 Å^2, while the second and the third are ∼30 Å^2. The other modes are negligible in amplitude (fourth eigenvalue is less than 10 Å^2, other eigenvalues are less than 5 Å^2). However, the second eigenvector involves oscillations of the hydrated loops on the protein surface; the fact that the projection of the MD trajectory onto this vector is characterized by a large cosine content (0.58) suggests that this is a mode with a high random diffusion character.[57] Instead, the first and the third principal modes are characterized by a low cosine content (0.17 and 0.04, respectively), suggesting that these modes should describe genuine vibrations of the protein.[57] These two modes involve regions **a**, **b**, and the substrate, with a contraction/expansion of the relative distances in the active site and steering of the substrate alternatively toward one of the two aspartates of the catalytic dyad (Table 1, see also animation at

(73) Hong, L.; Turner, R. T.; Koelsch, G.; Shin, D.; Ghosh, A. K.; Tang, J. *Biochemistry* **2002**, *41*, 10963.
(74) Garcia, A. E. *Phys. Rev. Lett.* **1992**, *68*, 2696−2699.
(75) Andreeva, N. S.; Rumsh, L. D. *Protein Sci.* **2001**, *10*, 2439.

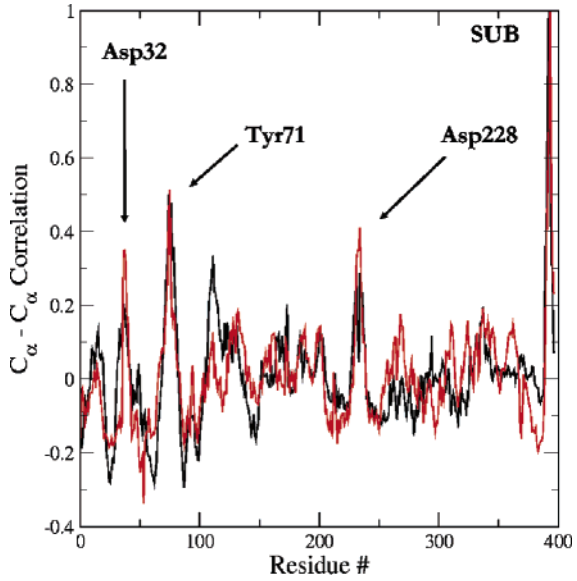

Figure 5. Correlation of substrate to BACE. Correlation values of Leu_{P_1} and $Ala_{P'_1}$ residues with the protein, obtained from normalization of the covariance matrix,[55] are drawn in black and red lines, respectively. Positions of the two catalytic aspartates and Tyr71 in the protein sequence are shown by black arrows.

Table 1. Variation Δ (in Å) of the Relative Distances in the Active Site, along the First and Third Eigenvector of the Covariance Matrix

mode	Asp32-Leu	Asp32-Ala	Asp228-Leu	Asp228-Ala
1	0.51	1.05	−0.36	0.08
3	0.77	1.31	−0.46	−0.09

www.sissa.it/sbp/bc). In both of the modes mentioned above, region **b** acts as a hinge between the two subunits, which have rigid displacements toward each other. As a result, the distance ξ between the cleavage site and the substrate defined, following ref 10, as the distance between the C_α of the Asp dyad and those of the P1 and P1′ $Leu_{P_1}-Ala_{P'_1}$ residues (Figure 6), shows two regimes around $\xi_1 = 7.7$ Å and $\xi_2 = 8.1$ Å. As for HIV-1 AP,[10] the fluctuations of the substrate allow oscillations of more than 1 Å around these typical distances; these conformations are characterized by two distinct H-bond patterns.

At $\xi = \xi_1$, Asp32 H-bonds to the Leu_{P_1} backbone, with the catalytic water (WAT, hereafter) bridging the two aspartates (Figure 7a). At $\xi = \xi_2$, the Asp32 carboxyl group is rotated by 180° and points its hydrogen toward WAT, which acts as an H-bond donor toward Leu_{P_1}'s backbone and Asp228 (Figure 7b).

QM/MM Simulations. We now investigate the enzymatic activity during the two different regimes observed in the MD run. The reaction has two steps (see Figure 2). Here, we shall concentrate on the first step of the reaction, which is the one that has been found in HIV-1 AP to be highly modulated by conformational fluctuations.

Within our picture, the free energy of first step of the enzymatic reaction depends, at least, on both ξ and the distance between water oxygen and the carbonyl carbon (ζ_{CO} in Figure 7a,b).

Here, we assume, as in refs 10 and 30, that the modes associated with ξ can be decoupled from those associated with ζ_{CO}, which are much slower. Thus, we consider two different

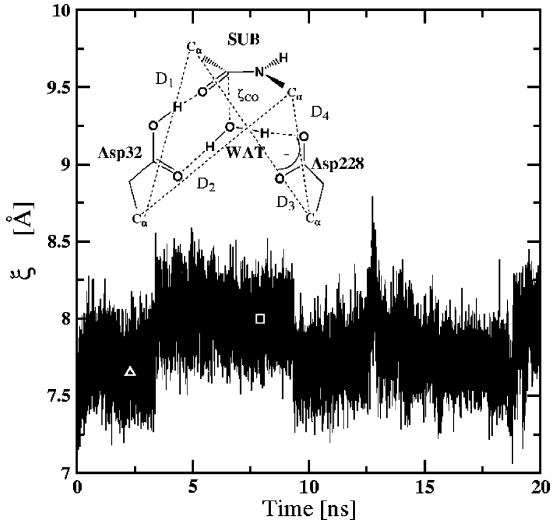

Figure 6. Fluctuation of the substrate in BACE active site. ξ distance ($\xi = (D_1 + D_2 + D_3 + D_4)/4$)[10] is plotted as a function of the simulated time. The white triangle and square represent the two snapshots used as a starting point for the QM/MM simulations at $\xi = \xi_1 = 7.7$ Å and $\xi_2 = 8.1$ Å, respectively.

randomly chosen conformations, one for each ξ regime, and perform thermodynamic integration only on the ζ_{CO} coordinate.

At $\xi = \xi_1$, the free energy barrier for the first step is around 20 kcal/mol (see the Supporting Information), similar to the one computed in the HIV-1 AP with the same procedure (∼18−20 kcal/mol[30]) and in fair agreement with the experimental barrier of the reaction (∼18 kcal/mol),[24] although such comparison should be taken into account, at this stage, only at a qualitative level, as in our computations, we are lacking the contributions to the k_{cat} coming from the second chemical step and other physical steps that might occur in the process. In our simulations, Asp32 transfers its proton to Leu_{P_1}'s carbonyl at $\zeta_{CO} \approx 1.9-2.0$ Å, well before reaching the transition state (TS) ($\zeta_{CO} = 1.73$ Å). After deprotonation, the side chain of Asp32 rotates and recovers a better coplanarity with the Asp228 side chain (dihedral angle from ∼50 to 20°).

At the TS, WAT protonates Asp228. As a result, the two hydroxyl groups of the gem-diol intermediate can H-bond to Asp32 (see Figure 7c,d). This reaction mechanism is similar to that of HIV-1 AP[10,30] except that in that case, the proton transfer from Asp32 to the substrate's carbonyl is concerted, as it occurs at the TS.

The conformation associated with the $\xi = \xi_2$ value is found to be much less reactive ($\Delta G >$ 40 kcal/mol; see the Supporting Information) and the geometry very similar to that found for large ξ values in HIV-1 AP;[10] as ζ_{CO} becomes as short as 2.5 Å, Asp32 H-bonds to Asp228 and does not interact with WAT (see Figure 7e). As a result, Asp228 is not able to deprotonate WAT anymore, and reaction does not occur even at very short WAT−substrate distances. During simulations at $\xi = \xi_2$, the two aspartates are never found to recover a better coplanarity.

3.3. Mobility of Other Pepsins. Performing classical MD simulations on other members of the pepsin family to investigate their mechanical properties would be computationally highly demanding. Here, we adopt a computer-efficient coarse-grained model where suitable quadratic energy penalties are introduced for structural distortions around the native state[37] (Figure 3). The reliability of the model has been checked by comparing

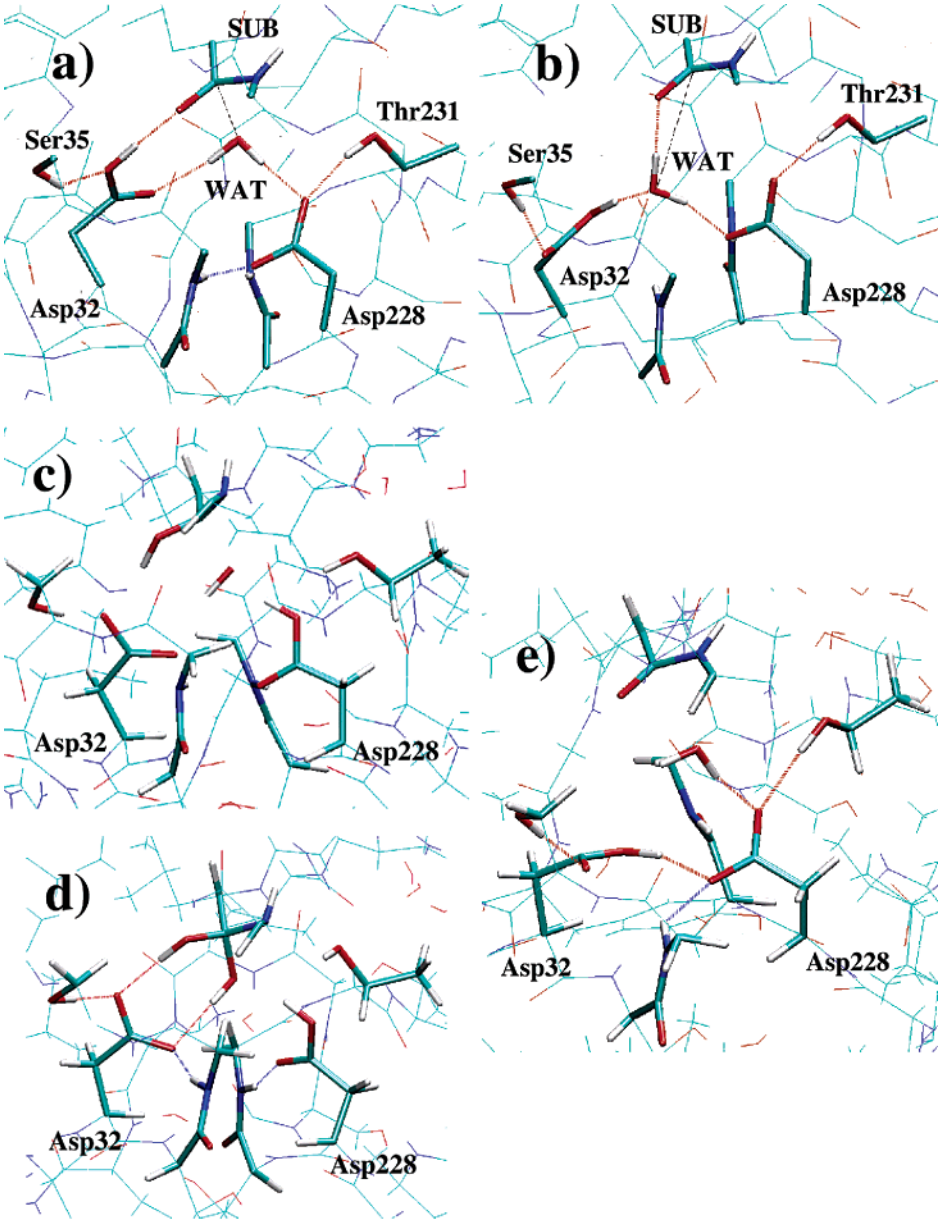

Figure 7. BACE active site geometries. Configurations at ξ_1 (a) and ξ_2 (b), TS geometry (c), gem-diol intermediate (d), and H-bond rearrangement at ξ_2, when ζ_{CO} is as short as 2.5 Å (e). Atoms drawn in cylinders are treated as quantum atoms during the QM/MM calculations. Hydrogen bonds are drawn in red and blue lines. In a and b, only hydrogens bonded to polar groups are shown for the sake of clarity. The dashed black line in a and b represents the reaction coordinate, ζ_{CO}.

the C_α covariance matrix computed for BACE with that previously obtained by MD simulations. The linear correlation coefficient of the covariance matrixes turns out to be as high as 0.8 (measured for over 78 000 distinct entries). In particular, the diagonal elements of the model and MD covariance matrix, which reflect the root mean square fluctuations (rmsf) of the C_αs, were also in very good agreement (Figure 8), thus confirming the viability of the coarse-grained approach to characterize the vibrational properties of this type of protein structures.

We focus on five randomly chosen pepsins from different organisms (Figure 3). Remarkably, aligned residues among these pepsins exhibit similar mobility properties (Figure 3). In particular, highly conserved regions are rather rigid, while nonconserved regions show a larger mobility.

The alignment underlines that the total number and position of S−S bridges may vary from sequence to sequence.[2,5] However, our coarse-grained calculations suggest that the presence of S−S bonds does not affect the low-frequency large-scale dynamics around the folded state (see Figure 8). Thus, their presence might impact more strongly on the folding dynamics or native-state stability rather than the near-native vibrational properties.

3.4. Mobility of Retroviral Proteases. Mobility of aspartic proteases from simian immunodeficiency virus (1AZ5),[76] equine anemia virus (2FMB),[77] feline immunodeficiency virus (2FIV),[78] and Rous sarcoma virus (2RSP)[79] have been computed and matched to data from MD of HIV AP from refs 10 and 11. The

(76) Rose, R. B.; Craik, C. S.; Stroud, R. M. *Biochemistry* **1998**, *37*, 2607−2621.

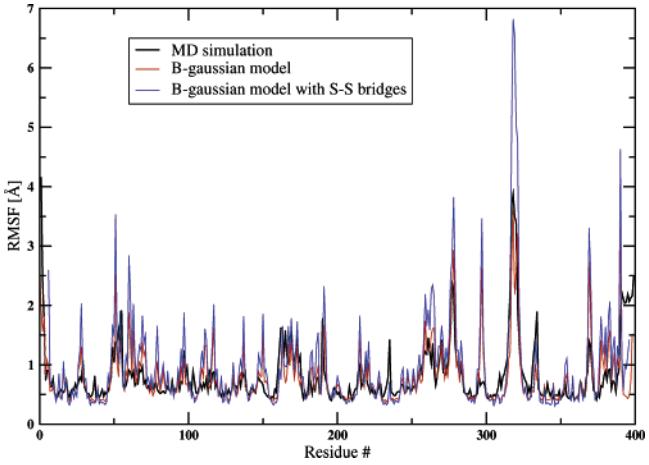

Figure 8. Root mean square fluctuations (rmsf) of BACE. Rmsf of the enzyme/substrate complex residues computed from MD simulation (black line), from B-Gaussian model (without explicit S−S bridges interactions) (red line) and from B-Gaussian model where explicit S−S bond interactions have been included (blue line). The plot shows that the B-Gaussian model is able to reproduce the mobility of the system, which is almost independent of the presence of the S−S bridges.

structurally conserved regions have been found to show a relatively low mobility, in MD simulations[10−12] and by our coarse-grained model (Figure 3).

The structural[9,40] and dynamical analyses[10−12] allow for the identification of a structural and functional correspondence between the conserved regions **a** and **b**.

The flaps of the retroproteases (residue numbers 38−60 in the HIV-1 AP sequence) seem to be highly mobile, especially when the enzyme is not bound to any ligand. In fact, it has not always been possible to determine their crystallographic structure (Figure 3). MD runs of HIV-1 AP from refs 10 and 11 have shown flaps less mobile than those of coarse-grained computations. Indeed, in that case, MD simulations of closed conformation of the ligand-bound enzyme[10] were performed. This contrasts the findings in pepsins, where similar mobility has been shown for the flaps in both ligand-bound and ligand-free enzymes.

4. Discussion and Concluding Remarks

Our calculations suggest that BACE explores different geometric conformations of the active site that are able to enhance or suppress its catalytic power. This efficient way of modulating BACE enzymatic activity is achieved by large-scale conformational changes of the protein. The relative rotations of the two lobes found during our dynamics are compatible with those described in the structure of plasmepsin 2 by Silva et al.[22]

In addition, the fluctuations of the substrate in the binding cavity can reduce the Asp32-Leu$_{P_1}$ distance enough so that an H-bond can be established among the carboxyl group of the former and the carbonyl one of the latter. The resulting rearrangement of the H-bond pattern leads to WAT reorientation into a favorable geometry for nucleophilic attack to Leu$_{P_1}$ carbonyl, with a resulting free energy of ∼20 kcal/mol, in fair agreement with experimental data (∼18 kcal/mol).[24]

The MD trajectory confirms the low mobility of regions **a**−**c** of Figure 1, which are found to undergo uncorrelated motion. The relevance of these regions for the enzymatic activity of pepsins is highlighted by the fact that 81% of conserved residues in pepsins are localized in them, while only the remaining 19% (mainly involving glycines in loops) do not group into specific regions. These regions exhibit low mobility also in the entire set of pepsins considered here (Figure 3), for which the 3D structure is available. The inspection of the principal components of the covariance matrix allows one to gain insight into the putative role played by these different regions, which is summarized below.

Functional Role of Regions a and b. These regions turn out to be involved in the large-scale motions of all considered pepsins, and most likely of all pepsins in general, owing to the strong fold similarities.

The displacement of the substrate with respect to the Asp dyad, located in the active site of region **a**, depends almost exclusively on the two principal modes, where the substrate is found to be correlated only to its nearest neighbors. These motions involve relative rotations of the N- and C-terminal lobes, and in both cases, region **b** acts as a hinge.

Interestingly, both regions **a** and **b** have similar relatively rigid structural elements (the cleavage site loops and the terminal β-sheets, respectively[11,12]) in HIV-1 AP, for which a similar mechanical coupling has been proven to modulate its enzymatic activity.[10] Because these regions are present in all retroproteases of known structure, it is reasonable to assume that similar mechanisms may occur in the entire family. We conclude that these conserved regions play a key role in ensuring the correct mechanics between the two lobes/subunits of the enzymes.

Functional Role of Region c. Region c does not appear to play a significant role in the enzymatic catalysis itself. In fact, region **c** is *not* involved in the relevant large-scale motions of the enzymes.

Inspection of *all* available 3D structures of pepsins in the free form and in complex with inhibitors (BACE,[40,80] renin,[81] endothiapepsin,[82,83] human pepsin,[84] and cathepsin D[66]) suggests that the flap residues are the ones that undergo the major displacements upon ligand binding (see Table 2 and ref 80), whereas region **c** does not rearrange significantly upon ligand binding.

Thus, the kink of the flap where the fully conserved Gly78 is located (Figure 1) could act as the fulcrum of the open-to-close conformation of the flaps upon substrate binding. Within this proposal, we anticipate that **c** could act as a rigid cantilever for this conformational change.

The flap opening functional mode could not be described by unbiased MD, as the relatively high free energy barrier for the close-to-open conformational change of the flap (estimated as ∼20 kcal/mol for the HIV-1 isoenzyme[85]) makes this mode

(77) Kervinen, J.; Lubrowski, J.; Zdanov, A.; Bhatt, D.; Dunn, B. M.; Hui, K. Y.; Powell, D. J.; Kay, J.; Wlodawer, A.; Gustchina, A. *Protein Sci.* **1998**, *7*, 2314−2323.
(78) Laco, G. S.; Schalk-Hihi, C.; Lubkowski, J.; Morris, G.; Zdanov, A.; Olson, A.; Elder, J. H.; Wlodawer, A.; Gustchina, A. *Biochemistry* **1997**, *36*, 10696−10708.
(79) Miller, M.; Jaskolski, M.; Rao, J. K. M.; Leis, J.; Wlodawer, A. *Nature* **1989**, *337*, 576−579.
(80) Hong, L.; Tang, J. *Biochemistry* **2004**, *43*, 4689−4695.
(81) Rahuel, J.; Priestle, J. P.; Gruetter, M. G. *J. Struct. Biol.* **1991**, *107*, 227−236.
(82) Blundell, T. L.; Jenkins, J. A.; Sewell, B. T.; Pearl, L. H.; Cooper, J. B.; Tickle, I. J.; Veerpandian, B.; Wood, S. P. *J. Mol. Biol.* **1990**, *211*, 919−941.
(83) Veerpandian, B.; Cooper, J. B.; Sali, A.; Blundell, T. L. *J. Mol. Biol.* **1990**, *216*, 1017−1029.
(84) Sielecki, A. R.; Fedorov, A. A.; Boodhoo, A.; Andreeva, N. S.; James, M. N. G. *J. Mol. Biol.* **1990**, *214*, 143−170.

Table 2. Comparison among Root Mean Square Deviations between Bound and Free Form of Different Pepsins in the Turn Region of the Flap, in Region c, and of the Whole Structure

pepsin	flap (Å)	region c (Å)	total (Å)
renin[a]	1.7	0.2	0.3
endothiapepsin[b]	0.6	0.1	0.3
human pepsin[c]	2.8	0.4	0.7
cathepsin D[d]	1.9	0.2	0.4
BACE[e]	3.0	0.2	0.6

[a] PDB codes: 2REN, 1HRN (ref 81). [b] PDB codes: 1OEW, 1ENT (refs 82 and 83). [c] PDB codes: 4PEP, 1PSO (ref 84). [d] PDB codes: 1LYA, 1LYB (ref 66). [e] PDB codes: 1SGZ (ref 80), 1FKN (ref 40).

unlikely to occur during the typical time span covered by MD simulations.

Region **c** has no rigid counterpart in HIV-1 AP, as in the other retroprotease structures, where flap terminal regions are much less structured than those in pepsins (Figure 1). It has to be stressed, as well, that structural modifications of the C-lobe, with respect to the N- one, have produced an asymmetry in the binding pocket and the presence of only one flap. On the contrary, the homodimeric structure of retroproteases leads to a symmetric binding pocket and to the presence of two flaps.[2] Therefore, it is reasonable to expect that the mechanisms of molecular recognition and binding between the enzyme and the substrate could be different in the two families, and that functional regions implied in the flap mechanics might have been modified in the evolutionary path of the two families.

In conclusion, our calculations suggest that BACE modulates the motion of the substrate in the active site which, in turn, impacts on its enzymatic activity. An analogous mechanism may occur in retroproteases, despite the differences between the folds of the eukaryotic and viral enzymes. The correct functional mechanics is achieved by ensuring rigidity in two specific regions (**a** and **b** in Figure 1). On the basis of coarse-grained models and sequence alignments of pepsins, we propose here that pepsins have evolved conserving not only the cleavage site (region **a**) but also the residues in the β-sheets linking the two lobes of the protein (region **b**) and around the flap ends (region **c**), and that all these regions should be important for the dynamical properties of this family of enzymes. Region **c**, which is putatively relevant for substrate binding, is typical of only the pepsin family. Instead, regions **a** and **b**, which are directly involved in the enzymatic activity, have their counterparts in the fold of the functionally related viral isoenzymes. Thus, the folds of retroproteases and pepsins, although very different, feature two conserved similar regions related to specific mechanical functions. In both families, it is found that the free energy barrier associated with the functional modulations of the enzyme/substrate complex is sufficiently low that thermal fluctuations can effectively trigger the conformational changes in time scales much shorter than the duration of the complete reaction.

By synergically exploiting structural and dynamical information, it has been possible to characterize three structural regions with different functional roles in pepsins and two related ones in retroproteases. This methodology could be applied in the future as a useful tool for broadening the actual capabilities of identifying the function for relevant targets in structural genomics research.

Acknowledgment. The authors thank INFM for providing grants for CPU time at the CINECA consortium.

Supporting Information Available: Root mean square deviation plot of the protein/substrate complex during MD simulations, and free energy plots from QM/MM simulations for both conformations. This material is available free of charge via the Internet at http://pubs.acs.org.

JA044608+

(85) Zhu, Z. W.; Schuster, D. I.; Tuckerman, M. E. *Biochemistry* **2003**, *42*, 1326−1333.