

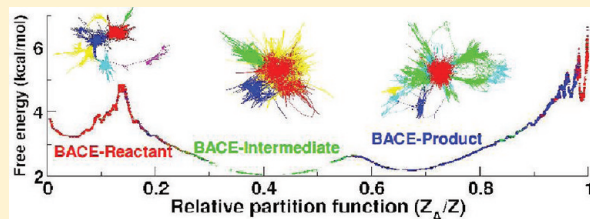
# Dynamics in the Active Site of $\beta$ -Secretase: A Network Analysis of Atomistic Simulations

Sabyashachi Mishra\* and Amedeo Caflisch\*

Department of Biochemistry, University of Zurich, Zurich, Switzerland

**S** Supporting Information

**ABSTRACT:** The aspartic protease  $\beta$ -secretase (BACE) catalyzes the hydrolysis of the amyloid precursor protein (APP) which leads to amyloid- $\beta$  aggregation and, ultimately, the perilous Alzheimer's disease. The conformational dynamics and free energy surfaces of BACE at three steps of the catalytic cycle are studied here by explicit solvent molecular dynamics simulations (multiple runs for a total of 2.2  $\mu$ s). The overall plasticity of BACE is essentially identical for the three states of the substrate: the octapeptide reactant, gem-diol intermediate, and cleavage products. In contrast, the network of hydrogen bonds in the active site is more stable in the complex of BACE with the gem-diol intermediate than the other two states of the substrate. The spontaneous release of the C-terminal (P1'–P4') fragment of the product follows a single-exponential time dependence with a time constant of 50 ns and does not require the opening of the flap. The fast dissociation of the C-terminal fragment is consistent with the transmembrane location and orientation of APP and its further processing by  $\gamma$ -secretase. On the other hand, the N-terminal (P4–P1) fragment of the product does not exit the BACE active site within the simulation time scale of 80 ns. A unified network analysis of the complexes of BACE with the three states of the substrate provides an estimation of the activation free energy associated with the structural rearrangements that involve only noncovalent interactions. The estimated rearrangement barriers are not negligible (up to 3 kcal/mol) but are significantly smaller than the barrier of the peptide bond hydrolysis reaction.



Alzheimer's disease is a fatal neurological disorder whose chances of appearance increase very rapidly with age beyond 65.<sup>1</sup> The pathogenesis of Alzheimer's disease is characterized by accumulation of the 40- or 42-residue  $\beta$ -amyloid ( $A\beta$ ) peptide which is produced by sequential cleavage of membrane bound amyloid precursor protein (APP) by  $\beta$ - and  $\gamma$ -secretase.<sup>2</sup> While  $\gamma$ -secretase cleaves APP in the transmembrane region, the action of  $\beta$ -secretase (termed BACE hereafter), which is the rate-limiting process in  $A\beta$  peptide formation,<sup>3</sup> takes place in the region of APP accessible from the luminal side of the cell membrane.<sup>4,5</sup>

The understanding of the catalytic action of BACE is essential for the search of new inhibitors and, therefore, has inspired several studies in this direction.<sup>6–9</sup> It is generally accepted that the reaction follows a general acid–base reaction mechanism where the two catalytic Asp residues of BACE provide and accept protons to facilitate the hydrolysis of the peptide bond.<sup>10,11</sup> The reaction progresses via a gem-diol tetrahedral intermediate (Figure 1), the direct evidence of which has recently been obtained for a retroviral aspartyl protease employing X-ray crystallography at very low pH conditions which strongly reduces its catalytic activity.<sup>12</sup> The information obtained from such a structural analysis method, however, provides a static picture of the reaction concealing the conformational dynamics of the enzyme during action.

An increasing number of studies have provided evidence on the importance of flexibility and dynamics of proteins in enzyme catalysis.<sup>13–17</sup> The time scales between the reactive

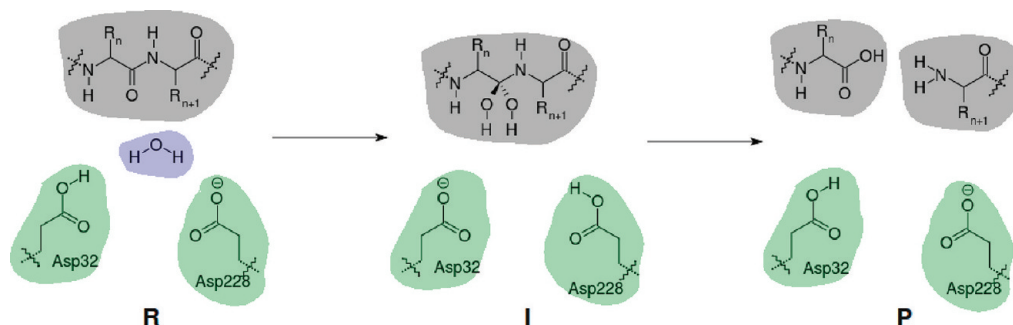
motion along reaction coordinates and conformational fluctuation are not separable for enzymatic reactions.<sup>18</sup> Therefore, the evolving conformational fluctuations in the active site during enzyme action contribute significantly to the chemical kinetics of catalysis.

Molecular dynamics (MD) simulations explore multiple conformational states of proteins. However, classical MD methods are constrained with fixed partial charges and fixed number of bonded and nonbonded interaction terms during time evolution, which makes them inappropriate to study breaking and formation of covalent bonds. These limitations can be surmounted by employing quantum mechanics and treating electron correlation ab initio, except for the fact that quantum mechanical calculations for systems as large as typical enzyme–substrate complexes are prohibitively expensive. Although, in principle, the combination of QM and MD can be used to extract dynamics of enzyme–substrate complexes, in practice the investigation of dynamics beyond a few hundreds of picoseconds of QM/MD becomes very challenging due to the expensive QM calculations, notwithstanding the fact that many interesting phenomena take place in the longer time scale.

**Received:** August 2, 2011

**Revised:** September 21, 2011

**Published:** September 23, 2011



**Figure 1.** Putative mechanism of BACE hydrolysis reaction. The peptide bond of the substrate, catalytic Asp residues, and the catalytic water molecule are highlighted by gray, green, and blue colors, respectively. The substrate in the reactant state (R) has the intact amide bond and in the intermediate state (I) shows a tetrahedral gem-diol structure, while the peptide bond is cleaved in the product state (P). Asp32 is protonated in the reactant and product states whereas Asp228 is protonated in the intermediate state.

Peptide hydrolysis catalyzed by BACE, like many other enzymatic processes, is a complex reaction due to the many degrees of freedom of the system. Analysis of the free energy surface that governs the conformational dynamics in such reactions becomes nontrivial due to the multidimensionality of the system where finding an appropriate reaction coordinate or even order parameters becomes increasingly difficult.<sup>19–24</sup> Commonly used projection based methods are well-known to provide a simplified picture of the free energy surfaces with incorrect barriers. Exploiting the analogy between system kinetics and equilibrium flow-through network, a barrier preserving projection of the free energy surface has been developed by Krivov and Karplus where the minimum cut<sup>25</sup> in a network is used for finding the free energy barriers and the relative partition function arising from the minimum cut is used as the progress coordinate to project the free energy onto.<sup>26,27</sup> This method, the so-called cut-based free energy profile (cFEP), preserves the barriers and minima in the order they appear during the sequence of events. The cFEP approach has been applied to extract protein-folding pathways and mechanisms from MD simulations,<sup>27–29</sup> to analyze kinetics of peptide aggregation,<sup>30</sup> to derive thermodynamics and kinetics of small molecule unbinding from proteins,<sup>31</sup> and has been extended recently to MD sampling by distributed computing.<sup>32</sup> The success of cFEP method relies on the validity of the underlying coarse graining where care should be taken to avoid clustering of kinetically distant snapshots into the same node. Furthermore, the free energy barriers in the cFEPs are typically the lower bound of the real barriers due to possible presence of shortcuts<sup>33</sup> between different parts of conformational space network. Therefore, validity of coarse graining has to be examined by varying clustering cutoff as well as by changing selection of atoms for clustering and verifying the diffusive nature of underlying dynamics.<sup>33</sup>

Here, we analyze the free energy surface associated with the hydrolysis of a BACE–substrate complex using explicit solvent MD simulations and the cFEP approach. The BACE–substrate complex is modeled at three stages of the hydrolysis reaction corresponding to reactant, intermediate, and product (see Figure 1). Multiple MD runs are carried out for each of the three states of the hydrolysis reaction, and the trajectories are analyzed through network analysis methods.<sup>27,34</sup> The simulation results indicate that despite similar overall flexibility of BACE for the three forms of the substrate the active site network of hydrogen bonds is most stable for the gem-diol intermediate. Moreover, spontaneous release of the C-terminal

fragment (CTF) of the cleaved product (i.e., P1'–P4' residues) is observed on a 50 ns time scale, whereas the N-terminal fragment (NTF) (i.e., P4–P1) does not leave the active site of BACE during this time scale.

## COMPUTATIONAL METHODS

**Simulation Setup.** The initial structure of the complex was obtained from the crystal structure of BACE bound to the OM99-2 inhibitor (Glu-Val-Asn-Leu- $\phi$ [CHOH-CH<sub>2</sub>]-Ala-Ala-Glu-Phe), where  $\phi$ [CHOH-CH<sub>2</sub>] represents a hydroxyethylene isostere of the peptide bond (PDB ID 1FKN).<sup>35</sup> Chain A together with the crystal water molecules was considered for MD simulation. All Glu, Asp (other than catalytic Asps, see below), Lys, and Arg side chains were modeled as charged, and all His were kept neutral. Three disulfide bonds between the Cys residue pairs 155–359, 217–382, and 269–319 were taken into account for MD simulations. The Swedish variant of the wild-type APP substrate (double mutant Lys(P2)Asn and Met(P1)Leu with the additional mutation Asp(P1')Ala as in OM99-2) was modeled in three forms representing the reactant, intermediate, and product states of hydrolysis reaction. In the reactant form of the substrate, the hydroxyethylene isostere was replaced by a carbonyl group which results in the peptide bond. On the other hand, the intermediate form of the substrate was modeled by replacing the hydroxyethylene isostere by a tetrahedral gem-diol (C-(OH)<sub>2</sub>) group. The product state of the substrate was modeled by converting the peptide bond between the Leu(P1) and Ala(P1') into –COOH and H<sub>2</sub>N– groups, respectively. In all three forms, the P4 terminal amine and P4' carboxyl groups of the substrate were modeled as NH<sub>3</sub><sup>+</sup> and COO<sup>–</sup>, respectively. Of the two catalytic aspartates, the Asp32 was protonated for the reactant and product, while the Asp228 was protonated for the intermediate, which is in accordance to the general mechanism of acid–base hydrolysis reaction (see Figure 1). The catalytic water molecule is not present in the BACE–OM99-2 complex (PDB ID 1FKN). In the reactant state this water molecule was placed in the catalytic active site according to the position of the water molecule present in the active site of BACE apo crystal structure (PDB ID 1W50). The missing hydrogen atoms were added, and the protein was immersed in a cubic water box of side 93 Å, which ensured a 12 Å distance between the boundary of water box and any solute atom. The overall charge (+12) was neutralized by adding 12 chloride ions, and 150 mM salt concentration was achieved by adding 60 sodium and chloride

**Table 1. Details of MD Simulations**

state of the substrate	no. of trajectories	length of individual trajectories (ns)	total simulation time (ns)
reactant	9	61 (×2), 65, 78, 81, 84 (×4)	682
intermediate	9	62, 69, 75, 80, 81, 84 (×4)	703
product	10	61 + (61, 64, 67) <sup>a</sup> , 81, 83, 84 (×5)	837

<sup>a</sup>The trajectories in the parentheses were started from the last snapshot of the 61 ns trajectory with different initial velocities.

ions. The simulation system contained about 23 000 water molecules and a total of about 76 000 atoms.

The MD simulations were carried out at constant temperature (300 K) and constant pressure (1 atm) using the Nosé-Hoover Langevin thermostat and piston.<sup>36,37</sup> The long-range electrostatic interactions were treated by the particle mesh Ewald method with a 12 Å cutoff. The van der Waals interactions were truncated at a cutoff of 12 Å, and a switch function was activated starting at 10 Å.<sup>38</sup> The SHAKE algorithm<sup>39</sup> was used to fix the length of the covalent bonds involving hydrogen atoms, which allowed an integration time step of 2 fs. Structures were saved every 5 ps for analysis. All simulations were performed employing the NAMD program<sup>40</sup> with the CHARMM22 force field<sup>41</sup> and TIP3 potential for water molecules.<sup>42</sup> Nine MD trajectories were calculated for both BACE–reactant and BACE–intermediate complexes, while 10 trajectories were run for the BACE–product state (see Table 1). A total of 2.2 μs of MD simulation was analyzed by the programs CHARMM,<sup>43,44</sup> VMD,<sup>45</sup> and WORDOM.<sup>46</sup>

**Network Analysis.** The MD trajectories of the complexes were first clustered according to structural criteria and then analyzed by network-based methods. The distance root-mean-square (DRMS) was chosen for clustering. The DRMS between two MD snapshots (*a* and *b*) is defined as the square root of the averaged squared distance deviations ( $d_{ij}$ ) between pairs of atoms *i* and *j*, i.e.,  $DRMS = [(1/N) \sum_{ij} (d_{ij}^a - d_{ij}^b)^2]^{1/2}$ . The DRMS for *n* atoms, thus, takes into account the fluctuations of  $N = n(n - 1)/2$  distance pairs. All pairs of non-hydrogen atoms in the BACE active site (residues 32–35, 71–73, 76, 198–199, 227–228, and 231–232) and substrate (except Phe(P4′), which undergoes large fluctuations<sup>35</sup> and does not interact directly with BACE) were considered for DRMS clustering. This ensures the clustering to be sensitive toward all interactions involving intra-BACE, intrasubstrate, and between BACE and substrate. The leader algorithm, as implemented in WORDOM,<sup>46</sup> was employed for the clustering procedure where a DRMS cutoff of 0.5 Å was used (Table 2).

**Table 2. Details of DRMS Clustering**

state of the substrate	no. of snapshots	no. of clusters	no. of links	population of the largest cluster
reactant	136710	1240	8804	5023
intermediate	140705	2256	25255	6251
product	167562	3590	27122	10935
all	444977	14658	104857	19203

The robustness of clustering was examined with changing atom selection and varying DRMS cutoff (Figures S8 and S9 in Supporting Information).

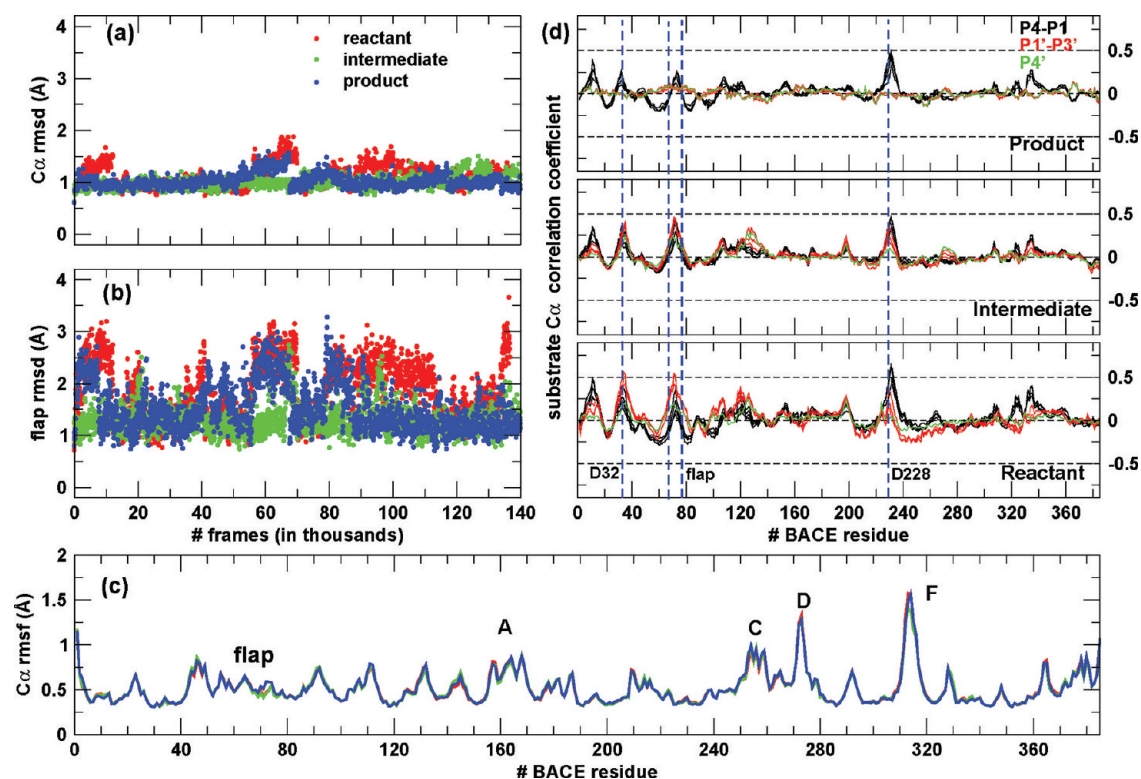
The above clustering provides a conformational space network<sup>34</sup> where the DRMS clusters are the nodes which are connected by edges that account for the transitions sampled during MD simulations with a saving frequency of 5 ps. For such a transition network the partition function of a node (*i*) is

given by  $Z_i = \sum_j c_{ij}$ , i.e., the number of times the node *i* is visited, where  $c_{ij}$  the edge capacity from node *j* to node *i*, is proportional to the number of direct transitions from *j* to *i* observed along a trajectory. Detailed balance is not imposed, and a comparison of networks with and without detailed balance shows essentially identical cFEPs (Figure S10 in Supporting Information). In the cFEP procedure, the transition barrier between two nodes of the network is defined as the surface with the minimal partition function that divides the network into two groups (*A* and *B*) each containing one of the two nodes of interest.<sup>26,27</sup> In other words, the partition function of the barrier is equal to the partition function of the cutting surface. For each minimum-cut calculation, the nodes are partitioned into two groups, *A* and *B*, with  $Z_A = \sum_{i \in A} Z_i$  and  $Z_B = \sum_{i \in B} Z_i$ , where  $Z_A$  and  $Z_B$  are the partition functions of the regions *A* and *B*. The partition function of the cutting surface is then  $Z_{AB} = \sum_{i \in A, j \in B} c_{ij}$ , where  $c_{ij}$  is the number of transitions between the nodes in regions, *A* and *B*. Thereby, the free energy of the barrier can be written as  $G = -k_B T \ln(Z_{AB}/Z)$ , where  $Z$  is the partition function of the entire network. This minimum-cut procedure, in principle, can be carried out for each pair of nodes, and the corresponding free energy barrier can be estimated. In practice, however, the nodes are sorted according to their values of mean first passage time (MFPT) with respect to a target node (often the most populated node), followed by the minimum-cut calculation to obtain  $x = Z_A/Z$  and  $y = -k_B T \ln(Z_{AB}/Z)$  for each value of MFPT, which then results in a one-dimensional free energy profile that features free energy basins separated by energy barriers for escaping the basin. The relative partition function is a progress variable that takes into account all possible pathways from the reference node<sup>27</sup> (see also Figure 1 in ref 29). All nodes to the left of the first barrier make up the first basin. Basins to the right of the first barrier are potentially overlapping. Thus, each basin of the cFEP profile is individually isolated by recalculating the cFEP profile from the most populated node of the concerned basin. To ensure that the clustering procedure preserves the diffusive behavior of the dynamics, the cFEPs were evaluated using the same DRMS clustering but considering every other MD snapshot, i.e., by doubling the time interval (*dt*) between consecutive structures (Figures S3–S5 in Supporting Information). The clustering preserves the intrinsic diffusivity of the simulated system (i.e., the diffusivity test is successful) when the free energy profiles constructed from the MD snapshots saved every 2*dt* steps is equidistant from the profile constructed from the data at saving intervals *dt* with a vertical spacing of  $0.35k_B T$ .<sup>33</sup>

## RESULTS AND DISCUSSION

**Plasticity of BACE during Hydrolysis of Peptide Substrate.** BACE has a bilobar structure where the N-terminus lobe (residue 1–150) and the C-terminus lobe (residue 151–385) form a substrate binding cleft which is partially covered by a hairpin loop also known as the flap (residue 67–77). There is a rigid framework with extended β-





**Figure 2.** Fluctuations and correlated motion analysis. RMSD of (a) BACE  $C_{\alpha}$  atoms (excluding the insertion regions) and (b) non-hydrogen atoms of the flap (BACE residues 67–77) along BACE–reactant (red), BACE–intermediate (green), and BACE–product (blue) MD trajectories. The RMSD is calculated with respect to the X-ray structure (PDB ID 1FKN), and the frames (separated by 5 ps) are from the concatenated MD trajectories (Table 1). Note that some abrupt changes are spurious since they represent the beginning of a new trajectory. (c) Root-mean-square fluctuation of BACE  $C_{\alpha}$  atoms averaged over intervals of 2 ns after excluding the first 20 ns of each MD trajectory. Same colors as in (a) and (b). The flap and insertion regions (A, C, D, and F) are labeled. (d) The correlation of fluctuations of the eight substrate  $C_{\alpha}$  atoms with the BACE  $C_{\alpha}$  atoms is obtained from the normalized covariance matrices of the corresponding  $C_{\alpha}$  atom displacements and averaging over intervals of 2 ns after excluding the first 20 ns of each MD trajectory. The correlation of the N-terminus (P4–P1) substrate residues are shown in black, while that of the C-terminus residues (P1'–P3') are in red. The green curve represents the correlation of the terminal residue (P4'). The flap region and the two catalytic Asp residues (Asp32 and Asp228) are indicated by vertical dashed lines.

sheets in the N-terminus lobe. The C-terminus lobe, on the other hand, features several flexible insertions as loops or  $\alpha$ -helices.<sup>35</sup> The  $C_{\alpha}$  RMSD time series for BACE bound to the substrate in reactant form (as octapeptide), intermediate form (backbone carbonyl group of Leu(P1) is replaced by a tetrahedral gem-diol group), and product form (with peptide bond between Leu(P1)-Ala(P1') cleaved) show a stable BACE structure during all MD runs (Figure 2a). The fluctuations in the flap region also remain limited irrespective of the form of the substrate bound to BACE, as seen from the RMSD of the heavy atoms of the flap (Figure 2b). This observation is in agreement with earlier simulations and experimental data where the flap was found in closed state when an inhibitor is bound to the BACE active site,<sup>35,47–49</sup> whereas the apo form of BACE exhibited an open conformation of the flap.<sup>49–51</sup> The extent of flap opening varies significantly among structures and among different protease enzymes.<sup>11,52</sup> The flap RMSD is found similar for all three states of the substrate (i.e., reactant, intermediate, and product) and is mainly in the closed state. The root-mean-square fluctuations (RMSF) of BACE  $C_{\alpha}$  atoms (Figure 2c) reveal a stable N-terminus lobe while the C-terminus lobe features large fluctuations associated with the insertion regions (A, C, D, and F). The extent of correlation between the substrate and BACE  $C_{\alpha}$  atoms is obtained by calculating the normalized covariance

$$C_{ij} = \frac{\langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle}{\sqrt{\langle (x_i - \langle x_i \rangle)^2 \rangle \langle (x_j - \langle x_j \rangle)^2 \rangle}}$$

between their fluctuations, where  $x_i$  is Cartesian coordinate of  $C_{\alpha}$  atom and the brackets represent the average over all configurations obtained from 2 ns segments of the MD trajectories. The values of  $C_{ij}$  are further averaged over all 2 ns segments after excluding the first 20 ns of each trajectory. The dynamical cross-correlation between substrate and BACE (Figure 2d) shows that the displacement of most BACE residues are not correlated with the substrate, reflected by the correlation coefficients ranging mainly between  $-0.2$  and  $0.2$ . Significant BACE–substrate correlation is observed only for the BACE segments around the two catalytic Asp residues (Asp32 and Asp228) and the flap (residues 67–77) (see Figure 2d). The terminal Phe(P4') residue of the substrate, which is disordered in the X-ray structure,<sup>35</sup> shows the smallest dynamical correlation with BACE.

From the above analysis, it emerges that the overall plasticity of BACE is essentially identical for the three states of the peptide substrate. This observation is further supported by the BACE  $C_{\alpha}$  correlation maps which are similar irrespective of the form of the substrate (Figure S1 in Supporting Information) and similar distribution of number of hydrogen bonds within BACE (Figure S2a in Supporting Information).

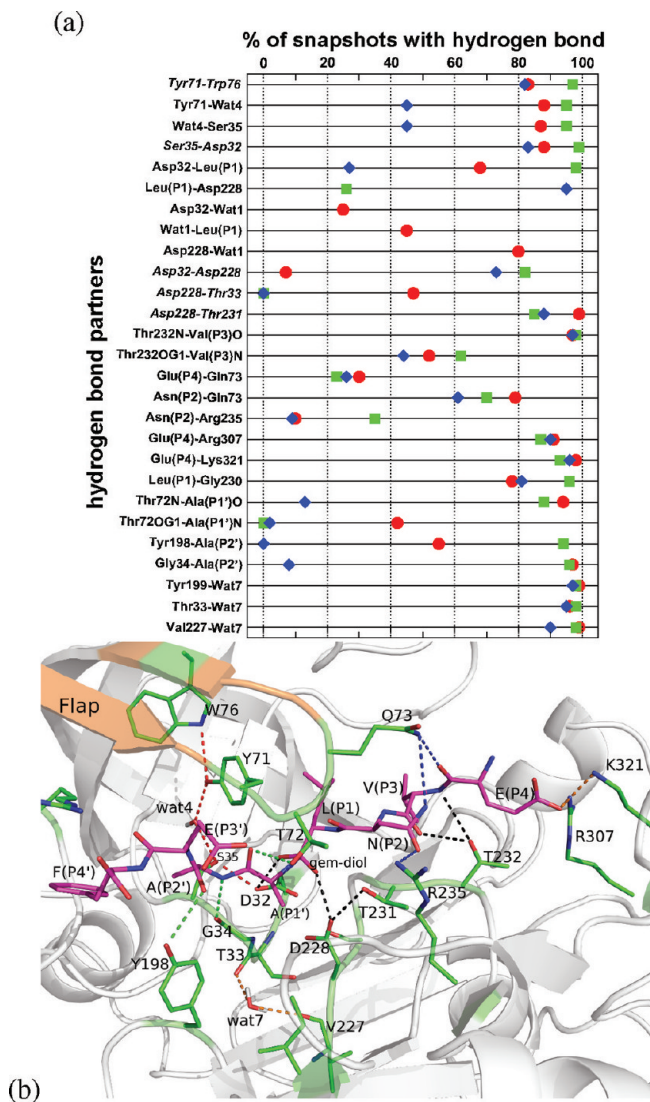
**Hydrogen Bond Network in the BACE Active Site.** The active site of BACE features an extended network of hydrogen bonds connecting the substrate with several residues from both N- and C-terminal lobes and, in particular, with the flap.<sup>47,49,53</sup> Notably, this hydrogen bond network is most stable (i.e., most populated) in the BACE–intermediate state (Figure 3a). Thus,

Figure 3b illustrates the hydrogen bond network using a representative snapshot of the BACE–intermediate complex. The side chain of Trp76 forms an intraflap hydrogen bond with the Tyr71 side chain<sup>47</sup> which orients itself as a proton donor to a water molecule (observed in the crystal structure), which in turn participates in a hydrogen bond with Ser35. Furthermore, Ser35 donates a hydrogen bond to the carboxylate group of the catalytic Asp32. This chain of hydrogen bonds, i.e., Trp76–Tyr71–Wat–Ser35–Asp32 (shown by red dashed lines in Figure 3b), provides a contact between the flap and the catalytic active site. While the Trp76–Tyr71 hydrogen bond is present in about 97% of simulation time for BACE–intermediate, this hydrogen bond is less stable in the reactant and product forms (83 and 82%, respectively, see Figure 3a). In about half of the BACE–product snapshots, the water molecule connecting Tyr71 and Ser35 escaped the protein matrix whereas it is very stable when BACE is bound to reactant and intermediate (Figure 3a). For BACE–intermediate complex, the chain of hydrogen bonds described above is present in 95% of the simulation, while it is only 71 and 42% when BACE is bound to reactant and product, respectively.

When the substrate is in the reactant form, i.e., when the peptide bond exists between P1 and P1', Asp32 takes part in hydrogen bond either with the carbonyl oxygen of Leu(P1) or the catalytic water molecule in the active site. The catalytic water molecule is engaged in 80% of the snapshots in a hydrogen bond with the catalytic Asp228, which lies in the C-terminus lobe of BACE. In the intermediate and product states, Asp32 and Asp228 are involved in hydrogen bonds with the gem-diol hydroxyl group (shown in Figure 3b) and terminal carboxyl group of Leu(P1), respectively. The carboxyl group of Asp228 makes a hydrogen bond with Thr231, and Thr232 stabilizes the backbone O of Val(P3). This way, the previously described chain of hydrogen bonds is extended by an additional hydrogen-bond chain comprising of Asp32–Leu(P1)–Asp228–Thr231...Thr232–Val(P3) (shown by black dashed lines in Figure 3b), where the dots indicate a connection through protein backbone. The side chain of Gln73 (in the flap) participates in hydrogen bond with the backbone O of Glu(P4) as well as Asn(P2) side chain. The latter also takes part in hydrogen bond with the side chain of Arg235, thus making a chain of hydrogen bonds Glu(P4)–Gln73–Asn(P2)–Arg235 (blue dashed lines in Figure 3b). Finally, the side chain of Glu(P4) is almost always involved in salt bridges with Arg307 and Lys321 (>90% of simulation time in all three states of the substrate, Figure 3a).

While NTF residues of the substrate are involved in multiple hydrogen bonds with BACE residues, fewer hydrogen bonds exist between BACE and the CTF of the substrate (Figure S2b in Supporting Information). The flap Thr72 side chain makes a hydrogen bond with Ala(P1') backbone O, while the backbone O and N of Ala(P2') are involved in hydrogen bond with Gly34 and Tyr198, respectively (dashed green lines in Figure 3b). The terminal residue of the substrate Phe(P4') does not participate in any hydrogen bond with BACE. For the BACE–product state, the stability of the hydrogen bonds between BACE and the CTF of the product reduces drastically (Figure 3a) since the latter undergoes spontaneous unbinding upon hydrolytic cleavage (see below).

**Clustering and Structural Analysis of BACE–Substrate Interaction.** The extensive network of interactions between BACE and substrate is rich with information regarding the role of the enzyme in the three states of the substrate during



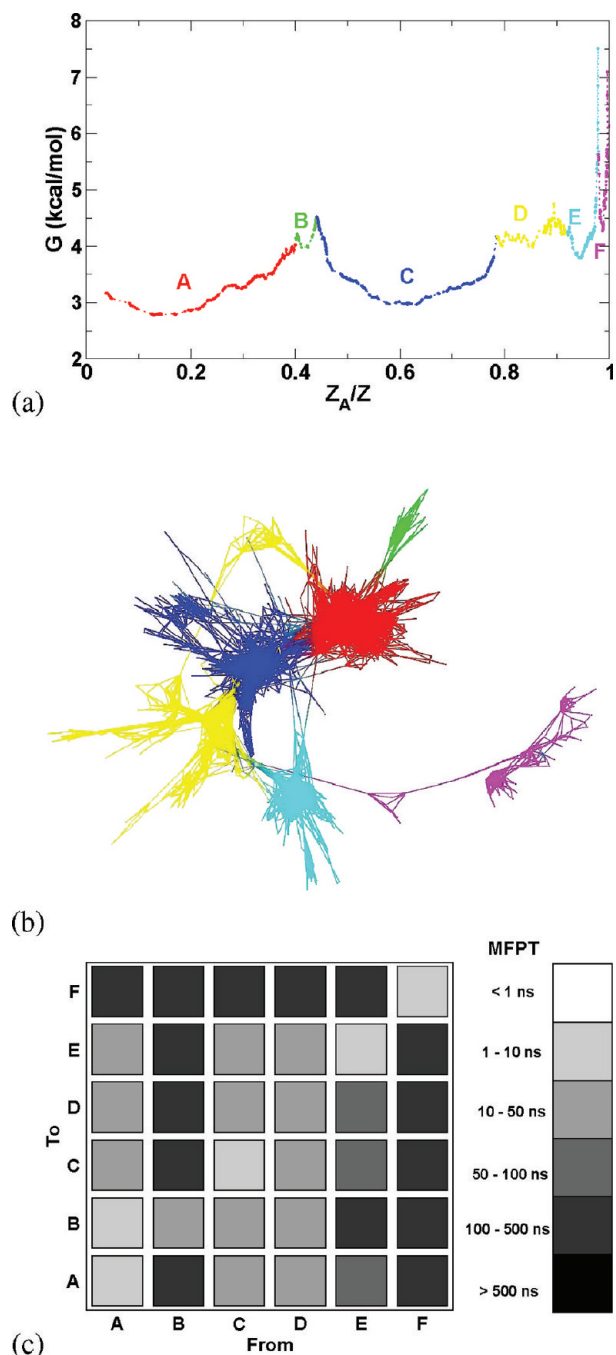
**Figure 3.** Hydrogen-bond network between the substrate and the active site of BACE. (a) Relative stability of different hydrogen bonds between BACE and substrate during BACE–reactant (red), BACE–intermediate (green), and BACE–product (blue) MD simulations. Residue names in italics indicate intra-BACE hydrogen bonds. Note that Leu(P1) indicates carbonyl, hydroxyl, and carboxy groups for the reactant, gem-diol intermediate, and biproduct complex, respectively. The criterion for hydrogen bond is a distance between the acceptor atom and the hydrogen atom smaller than 2.6 Å and the donor–H...acceptor angle larger than 130°. (b) Network of active site hydrogen bonds from the representative snapshot of the most populated node obtained from the DRMS clustering of all MD snapshots of BACE–substrate complexes. The most populated node originated from the BACE–intermediate state which is shown in the figure. The carbon atoms of the substrate are shown in magenta for clarity. Hydrogen bonds mentioned in the text are marked by dashed lines.



hydrolysis. The evolution of these interactions in the three states of the hydrolysis reaction can reveal the conformational rearrangements the enzyme and the substrate undergo during the catalytic reaction. However, a thorough analysis of the dynamics of these interactions is very challenging due to the intrinsic multidimensionality of the problem. A possible alternative is to make use of the network analysis by clustering the MD snapshots and obtaining free energy profiles by the barrier preserving cFEP method.<sup>26,27</sup> For the present purpose, where a structural evolution of the interaction between BACE and substrate is sought, the variable for clustering was chosen to be the DRMS between all atom pairs in the active site. The DRMS not only monitors the changes between BACE and substrate but also takes into account the intra-BACE and intrasubstrate fluctuations, some of which could also play a role in the enzyme activity, for example, the above-mentioned intraflap hydrogen bond.

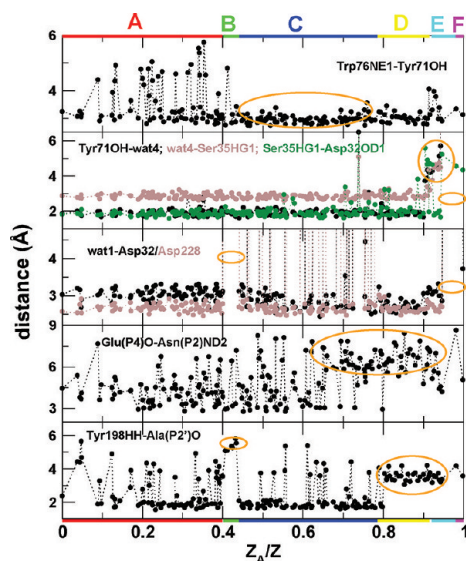
**BACE–Reactant.** The DRMS clustering involving 144 heavy atoms (10 296 distance pairs) of BACE and substrate (in the reactant state) resulted in 1240 clusters and 8804 transitions, i.e., nodes and links of the conformational space network (Table 2). The validity of the DRMS clustering was verified by examining the diffusivity of the inherent dynamics (the so-called diffusivity test<sup>33</sup>), see Figure S3 in Supporting Information. The cFEP for BACE–reactant exhibits two well-defined large basins, called A (which consists of 40% of the snapshots) and C (35% of the snapshots), and three small basins B, E, and F (Figure 4a). The unstructured region in the cFEP ( $0.78 < Z_A/Z < 0.92$ ), labeled D in Figure 4a, is an inhomogeneous basin as revealed in the conformational space network<sup>34</sup> shown in Figure 4b. From the network diagram, basins A and C are found connected, and basin B appears to be a satellite basin of A. The time scale of transitions between the basins are obtained from the MFPT values between them. The MFPT from basin A to basin B is quite short while the opposite transition is a rare event (Figure 4c). Basin F, also evident from the conformational space network, is an isolated basin. In general, the smaller basins (B, E, and F) represent the rare event snapshots which is reflected from the very large MFPT values associated with the transitions from these basins to other basins. While the off-diagonal elements of Figure 4c give the transition time scale, the diagonal elements represent the average MFPT value from the most populated node of a given basin to its other nodes. This time scale is equivalent to the relaxation time of a basin which is less than 10 ns for the two large basins (A and C) and is much slower for the inhomogeneous basins (B and D).

The structural characterization of the cFEP basins obtained from the DRMS clustering that involves 10 296 distance pairs is a difficult task. However, the individual cFEP basins can be characterized in terms of differences in the active site by the analysis of the chains of hydrogen bonds described earlier. The two large basins (C and A) differ in the Trp76–Tyr71 distance, i.e., while all snapshots in basin C exhibit a hydrogen bond between the side chains of Trp76 and Tyr71 (top panel in Figure 5), in nearly 30% of the snapshots of basin A this hydrogen bond is destroyed. Furthermore, the presence or absence of the catalytic water molecule in the active site provides a further distinction between these two large basins (Figure 5). During 80% of the simulation of BACE–reactant, the catalytic water molecule remained in the active site. In 2 out of 9 trajectories, this water molecule was seen to leave the active site. The satellite basin B differs from the basin A when



**Figure 4.** Network analysis of the BACE–reactant complex. (a) The cFEP<sup>27</sup> from the most populated node obtained from DRMS clustering of BACE–reactant trajectories shows different basins. (b) Conformational space network<sup>34</sup> of the 1240 nodes obtained from the DRMS clustering. The links represent 5 ps transitions between the nodes sampled during the MD runs. The layout of the nodes was obtained by employing the force-directed spring-embedder algorithm as implemented in Tulip program.<sup>63</sup> The links are colored according to the basin(s) of the connected pairs of nodes. (c) Matrix of MFPTs between the basins. The MFPT from basin *I* to basin *J* is obtained by averaging the MFPTs between  $i_1$  (the most populated node of basin *I*) to all  $n_j$  nodes of basin *J*, i.e.,  $(MFPT)_{I \rightarrow J} = \sum_{j=1}^{n_j} (1/n_j)(MFPT)_{i_1 \rightarrow j}$ .

the catalytic water molecule leaves the active site and when the hydrogen bond between Tyr198 and Ala(P2') is lost. The smaller basins E and F represent the structures where the hydrogen-bond network between Tyr71–wat4–Ser35–Asp32 is

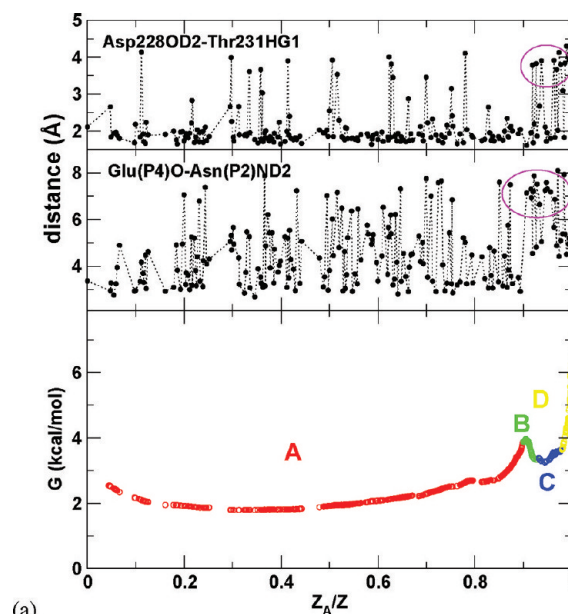


**Figure 5.** Structural characterization of the cFEP basins (marked by colored lines in the abscissa) of the BACE–reactant complex. The panels depict individual hydrogen bond distances in the representative snapshots of the nodes when arranged according to their  $Z_A/Z$  values with respect to the most populated node. Regions distinguishing different basins are highlighted by orange ellipses. For example, the ellipse in the top panel emphasizes the stability of the flap Tyr71–Trp76 hydrogen bond in basin C.

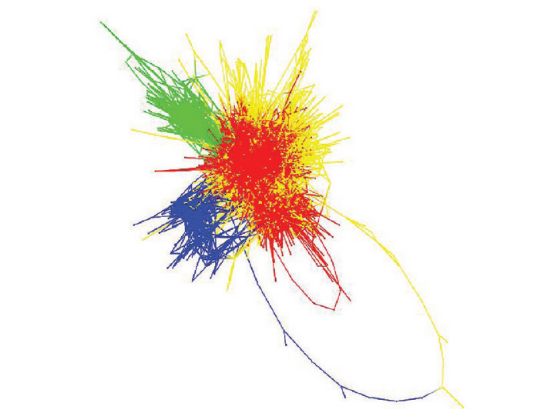
lost (second panel in Figure 5). The distances described above assume different values in different cFEP basins and have been analyzed in detail since they participate in the BACE–substrate hydrogen bond network. Other degrees of freedom associated with the substrate and BACE active site atoms might be required for a complete characterization of the cFEP basins which might be useful for the design of inhibitors that target a particular BACE conformation (as reported for another protease<sup>54</sup>).

**BACE–Intermediate.** Employing the same procedure for DRMS clustering as described earlier, the BACE–intermediate snapshots yield a conformational space network of 2256 nodes and 25 255 links (Table 2). The minimum-cut procedure across this network, when calculated from the most populated node, yields a cFEP with a very large basin (A) which accounts for 85% of the snapshots (Figure 6a). The remaining 15% of the snapshots are clustered into a basin which has three regions (B, C, and D in Figure 6a), the barriers between which appear when the cFEP is calculated starting from one of the nodes of these smaller basins (Figure S4 in Supporting Information). The conformational space network shows the large basin A (red region in Figure 6b), as a central basin with basins B, C, and D surrounding it.

Similar to BACE–reactant, a structural characterization of the cFEP basins has also been achieved in BACE–intermediate state. The large and small basins can be differentiated according to the presence or absence of hydrogen bond between the Asp228 and Thr231 and the intrasubstrate hydrogen bond between Glu(P4) and Asn(P2) side chains (Figure 6a). While the carboxylate oxygen of Asp228 is stabilized by the Thr231 in >90% of the snapshots of the large basin, the same is true for only 35% of the snapshots of the small basins. Also, the intrasubstrate hydrogen bond interaction is absent in the small basins while it is present in almost half of the snapshots in the large basin.



(a)

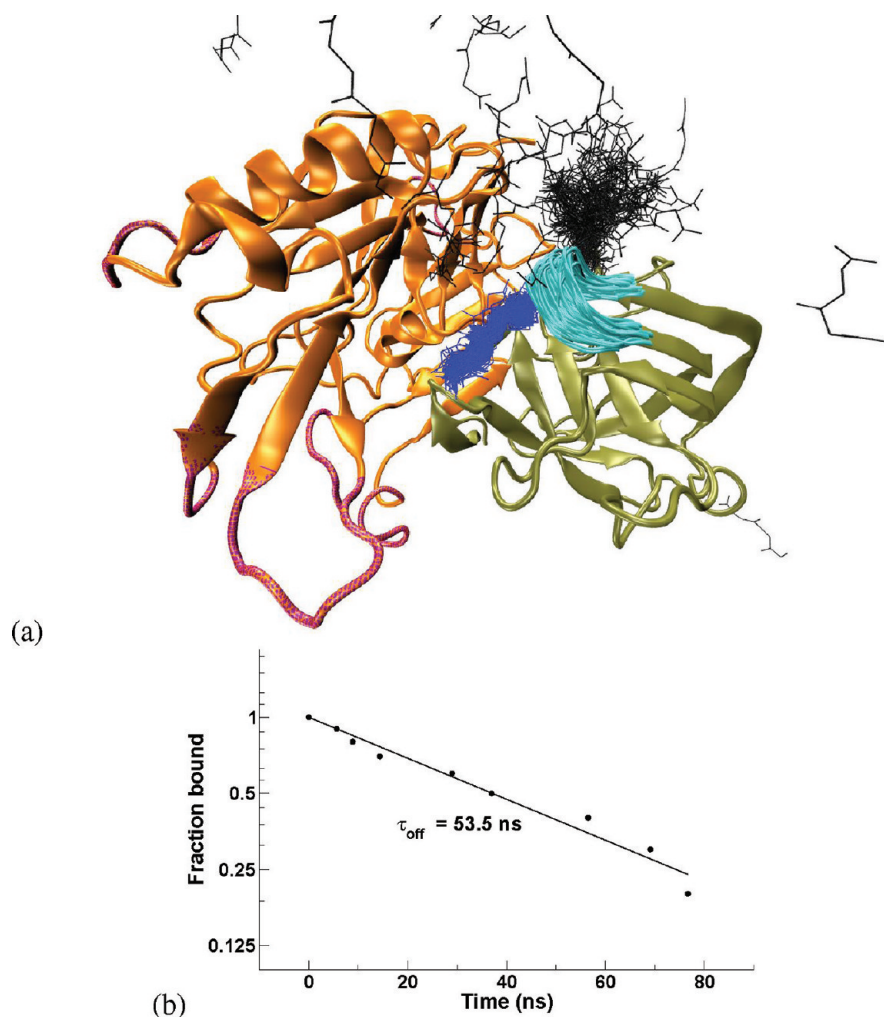


(b)

**Figure 6.** Network analysis of the BACE–intermediate complex. (a) Structural characterization of the cFEP basins and the cFEP from the most populated node obtained from the DRMS clustering of BACE–intermediate trajectories. (b) Conformational space network. The graph layout and colors are described in the caption of Figure 4b.

Compared to the BACE–reactant state, the BACE–intermediate state shows rather limited fluctuations which is reflected in the cFEP for BACE–intermediate where only one major basin is observed. This observation suggests that BACE when bound to the intermediate undergoes little conformational change which is essential for an efficient hydrolysis reaction. A highly flexible BACE–intermediate complex with multiple accessible conformations would slow down the hydrolysis reaction. On the other hand, when only one major conformation is available to the BACE–intermediate complex, the reaction is expected to progress faster by assuming optimal position and orientation of the gem diol and BACE side chains directly involved in hydrolysis.

**BACE–Product.** The NTF of the cleaved product remained bound to the BACE active site in all the MD runs. The structural stability of the NTF residues is due to the intermolecular hydrogen bonds (Figure 3) and van der Waals interactions between BACE and the hydrophobic side chains of Val(P3) and Leu(P1). In contrast, the CTF of the cleaved



**Figure 7.** Spontaneous release of the C-terminal fragment of the cleaved substrate. (a) The C-terminal fragment of the product (black lines) moves out of the BACE active site in 8 of 10 MD runs while the N-terminal fragment (blue lines) is not released in any MD run. The N- and C-terminal lobes of BACE are shown in green and orange, respectively, while the insertion regions are shown in maroon. The fluctuations in the flap (cyan) are shown by overlapping 160 snapshots equally spaced along the MD trajectories. (b) Single-exponential fitting of the cumulative distribution of unbinding of the C-terminal fragment of the cleaved product. The fitting yields a time constant of 53 ns (with  $\chi^2 = 0.0098$ ).

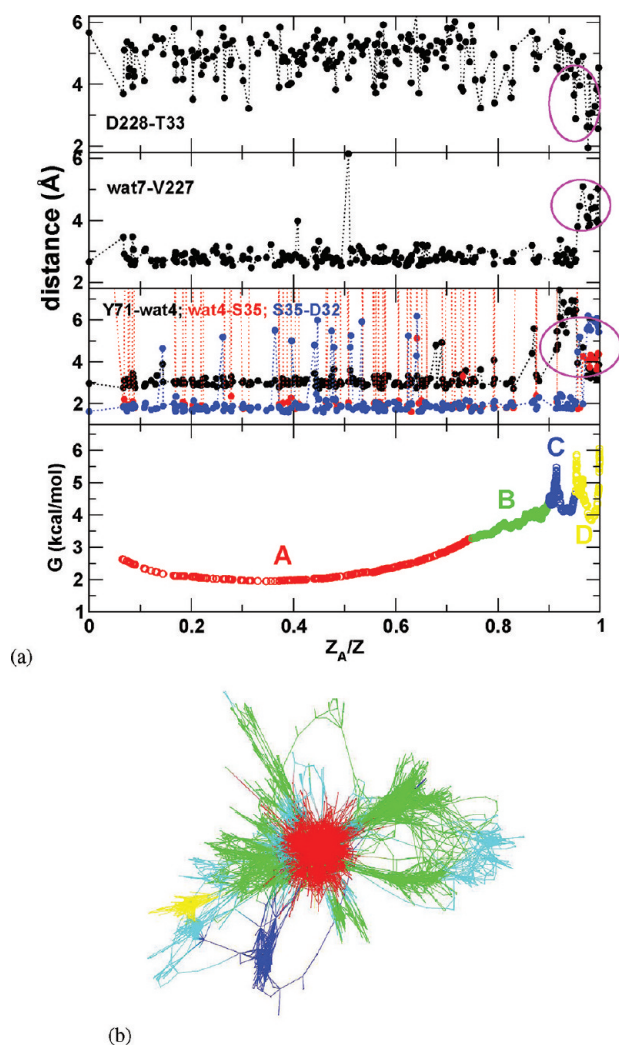
product spontaneously left the BACE binding site and migrated to the bulk water in 8 of the 10 MD runs (Figure 7a). The time dependence of the spontaneous unbinding of the CTF follows a single exponential decay with a time constant of 53 ns (Figure 7b), which suggests that unbinding is governed by a main free energy barrier. Interestingly, the release of the CTF does not require opening of the flap as can be seen from the limited flap movements (Figure 7a and Figure S6 in Supporting Information). This simulation result is in agreement with previous coarse-grained MD simulations of HIV protease where the product release was observed while the pair of flaps (each from one monomer of the dimer protease) remained in the closed form.<sup>55</sup>

In vivo, the membrane bound BACE cleaves the membrane anchored APP at the so-called  $\beta$ -site. The NTF of the cleaved APP becomes the C-terminus of the secreted soluble ectodomain. On the other hand, the CTF becomes the N-terminus of the 99-residue fragment of APP that undergoes a further hydrolysis cleavage by the  $\gamma$ -secretase which ultimately yields the  $A\beta$  peptide. Therefore, the fast dissociation of the CTF (tens of nanoseconds) facilitates further processing by  $\gamma$ -secretase.

Since the CTF of the cleaved product moves out of the protein matrix, the DRMS clustering of BACE–product complex was done by excluding the CTF of the cleaved product. Thereby, the number of atoms considered in the clustering reduces to 128 (i.e., 8128 distance pairs). The network obtained from this clustering was used for cFEP calculation as described earlier. The cFEP, when calculated from the most populated node, exhibits a large basin A which accounts for 75% of the snapshots and two small basins C and D (Figure 8a). In addition, the cFEP region termed B ( $0.78 < Z_A/Z < 0.90$ ) includes structures that form the rim of basin A according to the conformational space network (Figure 8b). Basin A can be structurally differentiated from the smaller basins by the absence of the Asp228–Thr33 side chain hydrogen bond and the presence of the hydrogen bond between Val227 and wat7. Additionally, the hydrogen-bond network Tyr71–Wat4–Ser35–Asp32 is absent in the small basins.

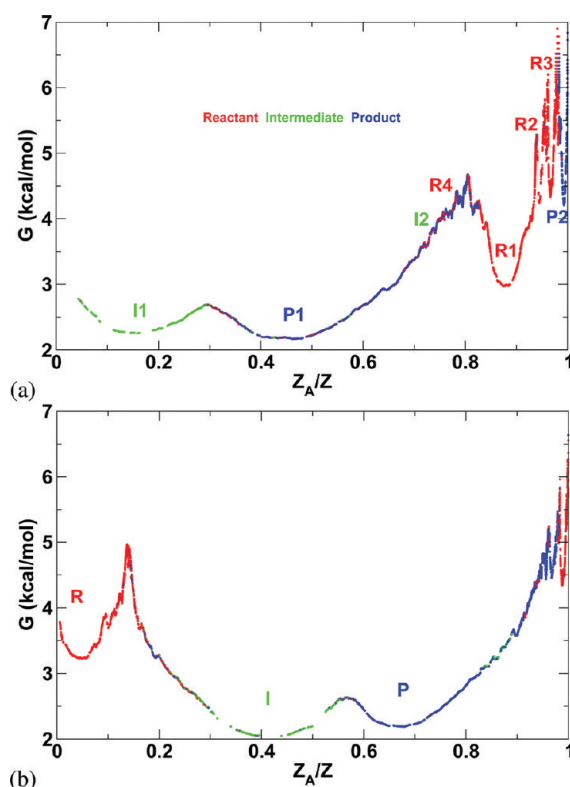
**Unified Network Analysis of BACE–Reactant, –Intermediate, and –Product Complexes.** The three forms of the peptide substrate (i.e., reactant, intermediate, and product) share the majority of their atoms so that it is possible to use the same DRMS clustering and combine the three cFEP analyses





**Figure 8.** Network analysis of the BACE-product complex. (a) Structural characterization of the cFEP basins and the cFEP from the most populated node obtained from the DRMS clustering of BACE-product trajectories. (b) Conformational space network. The graph layout and colors are described in the caption of Figure 4b.

into a unified description. As the CTF of the cleaved product leaves the protein matrix in most MD runs, DRMS clustering of the 444 977 snapshots of all MD simulations was done by including BACE active site atoms and the atoms of the NTF of the substrate, which amounts to 128 atoms and 8128 distance pairs. The resulting network (14 568 nodes and 104 857 links, Table 2) was used for cFEP calculation from the most populated node which arises from the BACE-intermediate state (Figure 9a). The cFEP features three well-separated basins, named R1, I1, and P1, corresponding to each of the three states of the substrate during the hydrolysis reaction. In addition, the cFEP also shows three small basins (R2, R3, and P2). The cFEP in the region  $0.7 < Z_A/Z < 0.8$  exhibits nodes arising from all three states due to possible overlap of nodes after the first barrier.<sup>27</sup> Figure 9b shows the cFEP profile when calculated from the most populated node of the BACE-reactant complex. The cFEP reveals forward and reverse barriers between reactant and intermediate amounting to about 2 and 3 kcal/mol, respectively. On the other hand, the barrier between intermediate and product is calculated to be 0.5 kcal/mol. In the cFEP method, the free energy of a basin



**Figure 9.** Network analysis of the combined sampling. (a) The cFEP from the most populated cluster obtained from the DRMS clustering of all BACE-substrate trajectories. Each node is colored according to the fraction of its snapshots belonging to the reactant (red), intermediate (green), and product (blue) states. (b) The cFEP calculated from the most populated node of the reactant.

corresponds to its partition function which is reflected in Figure 9b where the free energy of the intermediate state (I1 in Figure 9a) accounts for 85% of the intermediate state population while the population of the reactant and product states are distributed over multiple basins (R1 to R4, P1, and P2 in Figure 9a). Moreover, the free energy of the states in the present cFEP accounts for only the noncovalent interactions, and therefore, it is expected to be influenced when the free energy associated with the breaking and formation of covalent bonds are included. It should be pointed out that the free energy barriers in the cFEPs are typically the lower bound of the real barriers due to possible presence of shortcuts<sup>33</sup> between different parts of conformational space network. However, the effects of these shortcuts on the cFEP barriers (in particular, the first barrier) can be minimized with a clustering procedure that retains the diffusive nature of the underlying dynamics.<sup>33</sup>

Several studies have focused on the activation energy barrier of hydrolysis reaction by protease enzymes. Employing empirical methods for the calculation of the equilibrium constants for the formation of tetrahedral intermediates in solutions, it was concluded that there is very little chance of observing the accumulation of such an intermediate in the course of the hydrolysis of peptide by serine proteases.<sup>56</sup> Later on, using structural information from X-ray crystallography of HIV protease with inhibitor, a new model of the enzymatic mechanism was proposed where the proteolytic reaction was viewed as a concerted process.<sup>57</sup> On the other hand, using the

Car–Parrinello MD method for HIV protease, the activation energy barriers between reactant and intermediate and between intermediate and product were determined to be 18 and 21 kcal/mol, respectively, suggesting the latter to be the rate-determining step.<sup>58</sup> Identical values of activation energy barrier between reactant–intermediate and intermediate–product states for plasmepsin II (19.6 kcal/mol) and very similar values for HIV protease (17.2 and 16.4 kcal/mol, respectively) were obtained by employing empirical-valence bond approach.<sup>59</sup> Recently, with DFT methods for model system of BACE, these barriers were estimated at 19 and 10 kcal/mol, respectively, suggesting the formation of the intermediate to be the rate-limiting step.<sup>60</sup> Similar results were also obtained from DFT calculations of model system of  $\gamma$ -secretase.<sup>61</sup> On the contrary, a recent crystallographic study proposed the collapse of the intermediate as the rate-limiting step of hydrolysis in HIV protease.<sup>12</sup>

While an agreement on the reaction activation energy of the hydrolysis reaction by protease enzymes is yet to be reached, the barriers seen in the cFEP in the present study account for the activation energy required by the enzyme–substrate complex for their internal rearrangement in the three states of BACE–substrate complex. These realignment barriers are different from the hydrolysis reaction barriers,<sup>58–61</sup> since the former involve solely noncovalent interactions while the latter is the activation energy required to break or form appropriate covalent bonds during the progress of the reaction, which has earlier been studied by various methods<sup>58–61</sup> and lies beyond the scope of the present work. The conformational rearrangement barriers from the present study are found to be considerably smaller than the hydrolysis reaction barrier.<sup>58–61</sup>

## CONCLUSIONS

Alzheimer's disease is characterized by accumulation of A $\beta$  peptides which originate from the hydrolysis of APP catalyzed by the aspartic protease BACE.<sup>3</sup> Peptide bond hydrolysis by aspartic proteases progresses via a general acid–base reaction mechanism with a gem-diol intermediate.<sup>12</sup> Here, we have investigated the conformational dynamics and free energy surface of BACE during its enzymatic action on an octapeptide substrate by modeling the BACE–substrate complex in three states of the catalytic cycle. Multiple independent explicit solvent MD simulations (for a total time of about 2.2  $\mu$ s) have been carried out for each of the three states: the octapeptide reactant, gem-diol intermediate, and cleavage products. The conformational space network<sup>34</sup> and cFEP analysis of the trajectories<sup>27</sup> suggest the following main conclusions. First, the overall plasticity of BACE is essentially the same irrespective of the state of the substrate. On the other hand, the gem-diol intermediate shows the most stable network of hydrogen bonds in the active site. This is in agreement with the empirical valence bond study of substrate hydrolysis by plasmepsin II where the enzyme was shown to stabilize the tetrahedral intermediate by 10 kcal/mol more compared to the stabilization of the reactant,<sup>59</sup> although the anionic form of the tetrahedral intermediate studied in ref 59 was later shown to be less stable compared to the neutral gem-diol intermediate,<sup>62</sup> which is the form of the tetrahedral intermediate considered in the present work. The recent quantum mechanical study of the BACE–substrate model system<sup>60</sup> considers only a few active site residues of BACE and thereby ignores the stabilizing interactions due to the extended network of hydrogen bonds discussed in the present work. This

network of contacts between the substrate and BACE residues (in the flap and active site) is likely to be essential for efficient hydrolysis.

Second, the cFEP analysis of the BACE–reactant complex provides evidence of multiple free energy basins which can be discriminated by the presence or absence of a restricted set of interactions like the intraflap hydrogen bond between the side chains of Tyr71 and Trp76 and/or the hydrogen bond between the catalytic aspartates and the catalytic water molecule. In addition, the hydrogen bonds in the active site involving, in particular, the BACE residues Ser35, Thr72, Gln73, Tyr198, Thr232, Arg307, and Arg321 are expected to play a key role in the substrate binding and hydrolysis reaction which should be further explored by appropriate mutation studies. Third, the P4–P1 residues (NTF) form more persistent hydrogen bonds with BACE than the P1'–P4' residues (CTF) in all three states of the substrate. The NTF of the product did not leave the BACE active site within the simulation time scale of about 80 ns because of favorable intermolecular hydrogen bonds and van der Waals interactions. In contrast, the BACE–product state simulations led to spontaneous release of the CTF of the product with a single-exponential time dependence and a time constant of about 50 ns. The release of the CTF did not require the opening of the flap. The fast dissociation of the CTF of the product is consistent with the transmembrane location and orientation of APP and its further processing by the membrane-bound  $\gamma$ -secretase.

Finally, the cFEP analysis of the combined MD trajectories provides an estimation of the free energy barriers related to conformational rearrangements involving only noncovalent interactions. While a barrier of 2 and 3 kcal/mol is found between reactant and intermediate for forward and backward processes, respectively, a smaller barrier (0.5 kcal/mol) is obtained for the conversion of the intermediate to the product. It should be pointed out that these rearrangement barriers are, by definition, different from the hydrolysis reaction barrier which is the activation energy associated with breaking or formation of covalent bonds.<sup>58–61</sup> Thus, the noncovalent rearrangement barriers are much smaller than the previously estimated reaction barriers but are not negligible.

## ASSOCIATED CONTENT

### Supporting Information

Supplementary figures (Figures S1–S15) and complete citation for refs 2, 41, and 44. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Authors

\*E-mail: [s.mishra@bioc.uzh.ch](mailto:s.mishra@bioc.uzh.ch); Tel: +41 44 635 55 21; Fax: +41 44 635 68 62 (S.M.).

\*E-mail: [caflisch@bioc.uzh.ch](mailto:caflisch@bioc.uzh.ch); Tel: +41 44 635 55 21; Fax: +41 44 635 68 62 (A.C.).

### Funding

S.M. acknowledges financial support from the Novartis Foundation and the Forschungskredit of the University of Zurich.

## ACKNOWLEDGMENTS

The MD simulations were carried out on the Schrödinger cluster at the Informatikdienste of the University of Zurich.

## ■ ABBREVIATIONS

A $\beta$ , amyloid beta; APP, amyloid precursor protein; BACE,  $\beta$ -secretase; cFEP, cut-based free energy profile; CTF, C-terminal fragment; DRMS, distance root-mean-square deviation; MD, molecular dynamics; MFPT, mean free passage time; NTF, N-terminal fragment; RMSD, root-mean-square deviation; RMSF, root-mean-square fluctuation.

## ■ REFERENCES

- (1) Bermejo-Pareja, F., Benito-Leon, J., Vega, S., Medrano, M. J., and Roman, G. C. (2008) Incidence and subtypes of dementia in three elderly populations of central Spain. *J. Neurol. Sci.* 264, 63–72.
- (2) Vassar, R., et al. (1999)  $\beta$ -secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *Science* 286, 735–741.
- (3) Sinha, S., and Lieberburg, I. (1999) Cellular mechanisms of  $\beta$ -amyloid production and secretion. *Proc. Natl. Acad. Sci. U. S. A.* 96, 11049–11053.
- (4) Cole, S. L., and Vassar, R. (2007) The Alzheimer's disease  $\beta$ -secretase enzyme, BACE1. *Mol. Neurodegener.* 2, 22.
- (5) Selkoe, D. J. (2001) Alzheimer's disease: Genes, proteins, and therapy. *Physiol. Rev.* 81, 741–766.
- (6) Ghosh, A., Shin, D., Downs, D., Koelsch, G., Lin, X., Ermolieff, J., and Tang, J. (2000) Design of potent inhibitors for human brain memapsin 2 ( $\beta$ -secretase). *J. Am. Chem. Soc.* 122, 3522–3523.
- (7) John, V., Beck, J., Bienkowski, M., Sinha, S., and Heinrichson, R. (2003) Human  $\beta$ -secretase (BACE) and BACE inhibitors. *J. Med. Chem.* 46, 4625–4630.
- (8) Grueninger-Leitch, F., Schlatter, D., Kueng, E., Nelboeck, P., and Doebeli, H. (2004) Substrate and inhibitor profile of BACE and comparison with other mammalian aspartic proteases. *J. Biol. Chem.* 277, 4687–4693.
- (9) Malamas, M., Erdei, J., Gunawan, I., Turner, J., Hu, Y., Wagner, E., Fan, K., Chopra, R., Olland, A., Bard, J., Jacobsen, S., Magolda, R. L., Pangalos, M., and Robichaud, A. J. (2010) Design and synthesis of 5,5'-disubstituted aminohydantoin as potent and selective human  $\beta$ -secretase (BACE1) inhibitors. *J. Med. Chem.* 53, 1146–1158.
- (10) Northrop, D. B. (2001) Follow the protons: A low-barrier hydrogen bond unifies the mechanisms of the aspartic proteases. *Acc. Chem. Res.* 34, 790–797.
- (11) Dunn, B. M. (2002) Structure and mechanism of the pepsin-like family of aspartic peptidases. *Chem. Rev.* 102, 4431–4458.
- (12) Das, A., Mahale, S., Prashar, V., Bihani, S., Ferrer, J., and Hosur, M. (2010) X-ray snapshot of HIV-1 protease in action: Observation of tetrahedral intermediate and short ionic hydrogen bond SIHB with catalytic aspartate. *J. Am. Chem. Soc.* 132, 6366–6373.
- (13) Mesecar, A. D., Stoddard, B. L., and Koshland, D. E. (1997) Orbital steering in the catalytic power of enzymes: Small structural changes with large catalytic consequences. *Science* 277, 202–206.
- (14) Falke, J. (2002) A moving story. *Science* 295, 1480–1481.
- (15) Benkovic, S., and Hammes-Schiffer, S. (2003) A perspective on enzyme catalysis. *Science* 301, 1196–1202.
- (16) Malabanan, M., Amyes, T., and Richard, J. (2010) A role for flexible loops in enzyme catalysis. *Curr. Opin. Struct. Biol.* 20, 702–710.
- (17) Bhabha, G., Lee, J., Ekiert, D., Gam, J., Wilson, I., Dyson, H., Benkovic, S., and Wright, P. (2011) A dynamic knockout reveal that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* 332, 234–238.
- (18) Min, W., Xie, S., and Bagchi, B. (2008) Two-dimensional reaction free energy surfaces of catalytic reaction: Effects of protein conformational dynamics on enzyme catalysis. *J. Phys. Chem. B* 112, 454–466.
- (19) Hummer, G. (2004) From transition paths to transition. *J. Chem. Phys.* 120, 516–523.
- (20) Best, R. B., and Hummer, G. (2005) Reaction coordinates and rates from transition paths. *Proc. Natl. Acad. Sci. U. S. A.* 102, 6732–6737.
- (21) Ma, A., and Dinner, A. R. (2005) Automatic method for identifying reaction coordinates in complex systems. *J. Phys. Chem. B* 109, 6769–6779.
- (22) Hegger, R., Altis, A., Nguyen, P. H., and Stock, G. (2007) How complex is the dynamics of peptide folding? *Phys. Rev. Lett.* 98, 028102.
- (23) Yang, S., Banavali, N. K., and Roux, B. (2009) Mapping the conformational transition in Src activation by cumulating the information from multiple molecular dynamics trajectories. *Proc. Natl. Acad. Sci. U. S. A.* 106, 3776–3781.
- (24) Vreede, J., Juraszek, J., and Bolhuis, P. G. (2010) Predicting the reaction coordinates of millisecond light-induced conformational changes in photoactive yellow protein. *Proc. Natl. Acad. Sci. U. S. A.* 107, 2397–2402.
- (25) Ford, L. R., and Fulkerson, D. R. (1956) Maximal flow through a network. *Can. J. Math.* 8, 399–404.
- (26) Krivov, S. V., and Karplus, M. (2004) Hidden complexity of free energy surfaces for peptide (protein) folding. *Proc. Natl. Acad. Sci. U. S. A.* 101, 14766–14770.
- (27) Krivov, S. V., and Karplus, M. (2006) One-dimensional free-energy profiles of complex systems: progress variables that preserve the barriers. *J. Phys. Chem. B* 110, 12689–12698.
- (28) Krivov, S. V., Muff, S., Cafilisch, A., and Karplus, M. (2008) One-dimensional barrier-preserving free-energy projections of a beta-sheet miniprotein: New insights into the folding process. *J. Phys. Chem. B* 112, 8701–8714.
- (29) Muff, S., and Cafilisch, A. (2009) Identification of the protein folding transition state from molecular dynamics trajectories. *J. Chem. Phys.* 130, 125104.
- (30) Pellarini, R., Schuetz, P., Guarnera, E., and Cafilisch, A. (2010) Amyloid fibril polymorphism is under kinetic control. *J. Am. Chem. Soc.* 132, 14960–14970.
- (31) Huang, D., and Cafilisch, A. (2011) The free energy landscape of small molecule unbinding. *PLoS Comput. Biol.* 7, e1002002.
- (32) Scalco, R., and Cafilisch, A. (2011) Equilibrium distribution from distributed computing (simulations of protein folding). *J. Phys. Chem. B* 115, 6358–6365.
- (33) Krivov, S. V., and Karplus, M. (2008) Diffusive reaction dynamics on invariant free energy profiles. *Proc. Natl. Acad. Sci. U. S. A.* 105, 13841–13846.
- (34) Rao, F., and Cafilisch, A. (2004) The protein folding network. *J. Mol. Biol.* 342, 299–306.
- (35) Hong, L., Koelsch, G., Lin, X., Wu, S., Terzyan, S., Ghosh, A., Zhang, X., and Tang, J. (2000) Structure of the protease domain of memapsin 2 ( $\beta$ -secretase) complexed with inhibitor. *Science* 290, 150–153.
- (36) Martyna, G., Tobias, D., and Klein, M. (1994) Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* 101, 4177–4189.
- (37) Feller, S., Zhang, Y., Pastor, R., and Brooks, B. R. (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* 103, 4613–4621.
- (38) Steinbach, P. J., and Brooks, B. R. (1994) New spherical-cut-off methods for long-range forces in macromolecular simulation. *J. Comput. Chem.* 15, 667–683.
- (39) van Gunsteren, W., and Berendsen, H. (1977) Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys.* 34, 1311–1327.
- (40) Phillips, J., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781–1802.
- (41) MacKerell, A. D. Jr., et al. (1998) All atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102, 3586–3616.
- (42) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935.



- (43) Brooks, B., Brucoleri, R., Olafson, B., States, D., Swaminathan, S., and Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4, 18–217.
- (44) Brooks, B., et al. (2009) CHARMM: the biomolecular simulation program. *J. Comput. Chem.* 30, 1545–1614.
- (45) Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD - Visual Molecular Dynamics. *J. Mol. Graphics* 14, 33–38.
- (46) Seeber, M., Felling, A., Raimondi, F., Muff, S., Friedman, R., Rao, F., Caffisch, A., and Fanelli, F. (2011) Wordom: A user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *J. Comput. Chem.* 32, 1183–1194.
- (47) Andreeva, N., and Rumsh, L. D. (2001) Analysis of crystal structures of aspartic proteinases: On the role of amino acid residues adjacent to the catalytic site of pepsin-like enzymes. *Protein Sci.* 10, 2439–2450.
- (48) Ghosh, A. K., Devasamudram, T., Hong, L., DeZutter, C., Xu, X., Weerasena, V., Koelsch, G., Bilcer, G., and Tang, J. (2005) Structure-based design of cycloamide-urethane-derived novel inhibitors of human brain memapsin 2. *Bioorg. Med. Chem. Lett.* 15, 15–20.
- (49) Gorfe, A. A., and Caffisch, A. (2005) Functional plasticity in the substrate binding site of beta-secretase. *Structure* 13, 1487–1498.
- (50) Hong, L., and Tang, J. (2004) Flap position of free memapsin 2 ( $\beta$ -Secretase), a model for flap opening in aspartic protease catalysis. *Biochemistry* 43, 4689–4695.
- (51) Patel, S., Vuillard, L., Cleasby, A., Murray, C. W., and Yon, J. (2004) Apo and inhibitor complex structures of BACE ( $\beta$ -secretase). *J. Mol. Biol.* 343, 407–416.
- (52) Levy, Y., and Caffisch, A. (2003) Flexibility of monomeric and dimeric HIV-1 protease. *J. Phys. Chem. B* 107, 3068–3079.
- (53) Prasad, B. V., and Suguna, K. (2002) Role of water molecules in the structure and function of aspartic proteinases. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* 58, 250–259.
- (54) Ekonomiuk, D., Su, X., Ozawa, K., Bodenreider, C., Lim, S., Otting, G., Huang, D., and Caffisch, A. (2009) Flaviviral protease inhibitors identified by fragment-based library docking into a structure generated by molecular dynamics. *J. Med. Chem.* 52, 4860–4868.
- (55) Trylska, J., Tozzini, V., Chang, C., and McCammon, J. A. (2007) HIV-1 protease substrate binding and product release pathways explored with coarse-grained molecular dynamics. *Biophys. J.* 92, 4179–4187.
- (56) Fastrez, J. (1983) On the stability of tetrahedral intermediates within the active sites of serine and cysteine proteases. *Eur. J. Biochem.* 135, 339–341.
- (57) Jaskolski, M., Tomasselli, A., Sawyer, T., Staples, D., Heinrikson, R., Schneider, J., Kent, S., and Wlodawer, A. (1991) Structure at 2.5 Å resolution of chemically synthesized human immunodeficiency virus type 1 protease complexed with a hydroxyethylene-based inhibitor. *Biochemistry* 30, 1600–1609.
- (58) Piana, S., Bucher, D., Carloni, P., and Roethlisberger, U. (2004) Reaction mechanism of HIV-1 protease by hybrid Car-Parrinello/classical MD simulations. *J. Phys. Chem. B* 108, 11139–11149.
- (59) Bjelic, S., and Åqvist, J. (2006) Catalysis and linear free energy relationships in aspartic proteases. *Biochemistry* 45, 7709–7723.
- (60) Barman, A., Schuerer, S., and Prabhakar, R. (2011) Computational modeling of substrate specificity and catalysis of the  $\beta$  secretase (BACE1) enzyme. *Biochemistry* 50, 4337–4349.
- (61) Singh, R., Barman, A., and Prabhakar, R. (2009) Computational insights into aspartyl protease activity of presenilin 1 (PS1) generating Alzheimer amyloid  $\beta$ -peptides (A $\beta$ 40 and A $\beta$ 42). *J. Phys. Chem. B* 113, 2990–2999.
- (62) Carnevale, V., Raugei, S., Piana, S., and Carloni, P. (2008) On the nature of the reaction intermediate in the HIV-1 protease: a quantum chemical study. *Comput. Phys. Commun.* 179, 120–123.
- (63) Auber, D. (2003) Tulip: A huge graph visualisation framework, in *Graph Drawing Softwares, Mathematics and Visualization* (Mutzel, P., and Juenger, M., Eds.) pp 105–126, Springer-Verlag, Berlin.