Microseconds Dynamics Simulations of the Outer-Membrane Protease T

Marilisa Neri,* Marc Baaden,† Vincenzo Carnevale,* Claudio Anselmi,* Amos Maritan,‡ and Paolo Carloni*
*International School for Advanced Studies and CNR National Institute for the Physics of Matter, National Simulation Center, Trieste, Italy;
†Institut de Biologie Physico-Chimique, CNRS, Paris, France; and ‡Dipartimento di fisica G. Galilei, Universita' degli studi di Padova, Padova, Italy

ABSTRACT Conformational fluctuations of enzymes may play an important role for substrate recognition and/or catalysis, as it has been suggested in the case of the protease enzymatic superfamily. Unfortunately, theoretically addressing this issue is a problem of formidable complexity, as the number of the involved degrees of freedom is enormous: indeed, the biological function of a protein depends, in principle, on all its atoms and on the surrounding water molecules. Here we investigated a membrane protease enzyme, the OmpT from *Escherichia coli*, by a hybrid molecular mechanics/coarse-grained approach, in which the active site is treated with the GROMOS force field, whereas the protein scaffold is described with a Go-model. The method has been previously tested against results obtained with all-atom simulations. Our results show that the large-scale motions and fluctuations of the electric field in the microsecond timescale may impact on the biological function and suggest that OmpT employs the same catalytic strategy as aspartic proteases. Such a conclusion cannot be drawn within the 10- to 100-ns timescale typical of current molecular dynamics simulations. In addition, our studies provide a structural explanation for the drop in the catalytic activity of two known mutants (S99A and H212A), suggesting that the coarse-grained approach is a fast and reliable tool for providing structure/function relationships for both wild-type OmpT and mutants.

INTRODUCTION

Recently, molecular dynamics (MD) studies have lead to the suggestion that enzymatic function results from a subtle interplay between chemical kinetics and molecular motion. Examples include the dihydrofolate reductase enzyme (1), whose completion of the catalytic cycle has been suggested to require coupled motions, and the proteases superfamily (2–4), whose biological function has been proposed to be affected by conformational fluctuations. These conclusions are being corroborated by a variety of experimental biophysical studies (5–10).

A major limitation of MD in addressing this issue is obviously the accessible timescale (currently up to the submicrosecond timescale). In principle, one could achieve much longer timescales by coarse-grained (CG)-potential-based MD approaches (11,12).

Low-resolution CG models provide the capability for investigating the longer time- and length-scale dynamics that are critical to many biological processes. CG models have been developed for investigating lipid membranes (13–19), proteins (20–23), and DNA (24–26). However, although CG potentials are useful to understand large-scale phenomena, they cannot describe the exquisite molecular recognition events among enzymes and their substrates, which are key for the enzymatic function.

Recently, we have presented a hybrid approach, the MM/CG model (28), with the goal to preserve the advantages of both MM and CG approaches, i.e., the necessary details associated with the biological activity and the ''large'' accessible timescale. Specifically, the amino acid residues involved in the ligand binding are treated with atomic detail (MM region) by means of a molecular mechanics force field, whereas the rest of the protein is treated at the CG level. This approach allows a fast and efficient description of the mechanical coupling between the dynamics of the active site with the enzymatic substrate (the MM region) and that of the protein environment (the CG region).

So far, MM/CG simulations have been applied to selected members of the aspartic protease family. These calculations were meant as a validation of the approach versus all-atom MD simulations (28). The MM/CG approach turned out to correctly reproduce both the local and the global features of two cytoplasmatic proteins (the aspartic proteases from HIV-1 PR (2) and β-secretase (BACE) (3)). Furthermore, the calculations reproduced the structural fluctuations of the substrate in the binding cavity, which plays an important role for the enzymatic activity (2,3). These test calculations covered the same timescale as reference MD simulations used for comparison (10 ns for HIV-1 PR and 8 ns for BACE). However, the computational cost of this scheme turned out to be about two orders of magnitude smaller than that of the corresponding all-atom MD (28).

In this article, we use the MM/CG approach to study the relevant case of the Michaelis complex of a membrane protease, the outer-membrane protease T (OmpT) (29) on the microsecond timescale.

OmpT is a defense protein expressed by Gram negative bacteria belonging to the omptin protein family (30), which has been shown to be important for the virulence of *Yersinia pestis* and clinical *Escherichia coli* isolates (31). The protein

Submitted June 28, 2007, and accepted for publication August 24, 2007.

Address reprint requests to Paolo Carloni, International School for Advanced Studies (SISSA/ISAS) and INFM-DEMOCRITOS Modeling Center for Research in Atomistic Simulation, Via Beirut 2, Trieste I-34014, Italy. Tel.: 39-040-7887-07; Fax: 9-40-7887-28; E-mail: carloni@sissa.it.

Editor: Klaus Schulten.

0006-3495/08/01/71/08 $2.00

doi: 10.1529/biophysj.107.116301

cleaves peptides preferentially between two consecutive basic amino acid residues (32,33).

The isoenzyme from *E. coli*, for which the x-ray structure is available (29), features a deep groove formed by loops L4 and L5 on one side and L1, L2, L3 on the other one (Fig. 1 *a*). This groove constitutes the active site of the enzyme. Initially, OmpT was classified as a novel-type serine protease; however, the relatively large distance between the putative catalytic S99 and H212 (~9 Å) observed in the x-ray structure (29) has suggested that the protease may work by a novel mechanism involving the D210-H212 and D83-D85 pairs. H212 and D83 groups have been suggested to activate a water molecule for the nucleophilic attack, whereas D85 and D210 contribute to polarizing the substrate scissile peptide bond (32,33). A peculiar H-bond network formed by these residues orients both the substrate and the nucleophilic water, promoting the cleavage of the peptide bond. MD simulations of the enzyme on a 10-ns timescale have further supported this scenario (34,35).

A comparison of MM/CG and all-atom MD simulations of the enzyme in the free state shows that the MM/CG approach is equally well suited for membrane proteins (36) and suggests that fluctuations of the active site cleft may play a role for substrate recognition. Here, we extend our investigations to the Michaelis complex of the protein with a model substrate (Ala-Arg-Arg-Ala). We carried out simulations on four OmpT/ARRA Michaelis complexes (A–D), which differ for: i), the protonation state of the putative catalytic residue D83; ii), for the N- and C-terminal tails of the substrate, which are considered in the Zwitterionic form (charged state) or capped with acetyl and *N*-methyl groups (neutral state), respectively, following all-atom MD calculations on the same system.

The aim of our MM/CG simulations is twofold. First, we investigated the role of conformational fluctuations of the substrate in the active site on the microsecond timescale. Results show that both large-scale motions of the protein and the electrostatic field, evaluated on such a timescale, impact on the function of the enzyme. Second, we used the MM/CG approach to investigate the effects of mutations for the enzymatic function. We focused on protein mutants S99A and H212A, which are experimentally known to be much less efficient than the wild-type (wt) (residual activity within 0% and 4% relative to the wt (32)). Our calculations, on the ~0.1 microsecond timescale, provide a structural basis for the dramatic decrease in the catalytic activity. Because of its relatively cheap computational cost (two orders of magnitude faster than standard all-atom MD), the methodology may be helpful to investigate structure/function relationships of high-throughput site-directed mutagenesis data.

METHODS

The MM/CG model

A detailed description of the MM/CG approach is reported elsewhere (28), therefore we only summarize its principles here.

A small part of the protein (e.g., the enzymatic active site) is investigated in atomic detail, whereas the rest is treated with a CG approach using a modified Go-model (37) by only considering C_α centroids. An interface region (I) is located between the two MM and CG regions, bridging the large discontinuity between full-atom and CG descriptions. The total potential energy of the system reads:

$$V = E_{MM} + E_{CG} + E_I + E_{MM/I} + E_{CG/I} + E_{SD},$$

where the first three terms represent the interactions within the MM, CG, and I regions, respectively, whereas the fourth and fifth represent the cross-terms' potentials. The last term, E_{SD}, mimics stochastic and frictional forces acting on the system due to the solvent (38,39).

In regions MM and I, all atoms are explicitly considered and, consequently, E_{MM}, E_I, and $E_{MM/I}$ energy terms have all the same formulation (i.e., the GROMOS96 43a1 force field (40)).

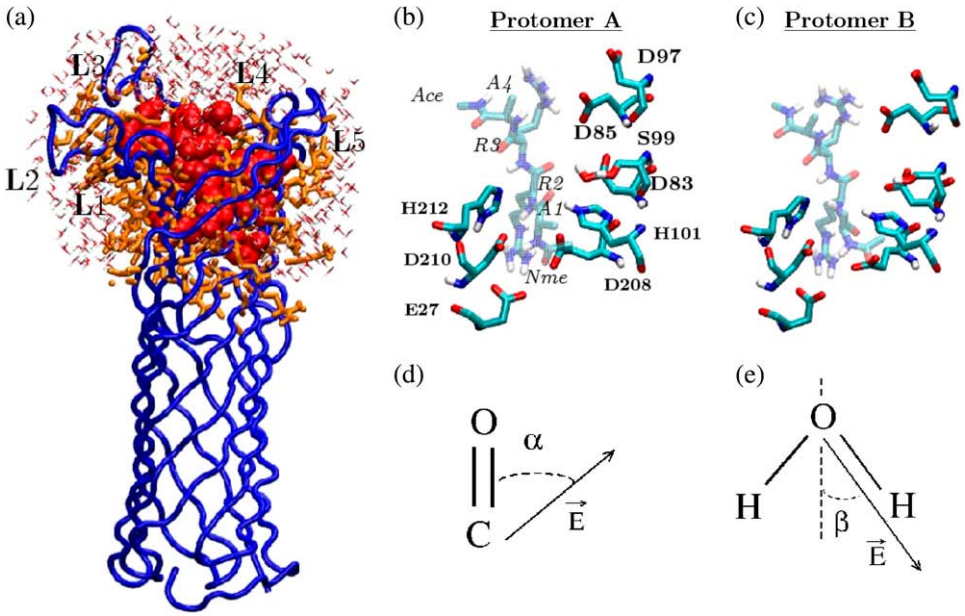

FIGURE 1 OmpT in complex with the model substrate ARRA. (*a*) Atoms treated with MD potential are depicted as red van der Waals spheres. Atoms belonging to the interface region (also treated with MD potential) are represented as orange licorice and the CG region is depicted in blue tube representation. A shell of water centered around the MM region is also shown. (*b* and *c*) Geometry of the active site of A and B, respectively, after 24 ns of MM equilibration simulations. The residues in licorice representation constitute the MM region. The substrate is depicted with a transparent effect. (*d* and *e*) Representation of the electrostatic field acting on the carbonyl carbon and on the dipole axes C_2 of the catalytic water, respectively.

E_{CG} takes the following form:

$$E_{CG} = \frac{1}{4}\sum_{i} K_b(|\mathbf{R}_i - \mathbf{R}_{i+1}|^2 - b_{ii+1}^2)^2 + \sum_{i>j} V_0[1 - \exp(-B_{ij}(|\mathbf{R}_i - \mathbf{R}_j| - b_{ij}))]^2. \quad (1)$$

The first term in Eq. 1 takes into account bonded interactions between consecutive CG (C_α) centroids, identified by the position vectors $\mathbf{R}_i$ and $\mathbf{R}_{i+1}$, and K_b is the relative bond force constant; b_{ij} is the equilibrium distance, corresponding to the native distance between CG atoms. The second term in Eq. 1 describes the nonbonded interactions between CG atoms. V_0 is the interaction well depth and B_{ij} is the modulating exponent of the Morse potential. The latter, along with V_0, have been obtained so as to reproduce features of all-atom MD calculations on a test system (28).

At the interface between the I and CG regions, bonds between consecutive C_α belonging to the I and CG regions ensure backbone connectivity. In addition, a term describing the nonbonded interactions is added: the interface atom, which is either a C_α or a C_β, interacts with the C_α in the CG region with the same potential as in Eq. 1.

A simple solvation model using explicit water molecules is also implemented: if the active site of the enzyme is solvent exposed a water drop is constructed around the MM and I regions. If a water molecule exits from the drop, its velocity is reflected toward the inside (36) to avoid evaporation.

Investigated systems

We focused on two OmpT protomers in complex with their substrate ARRA. The simulations differ in the protonation state of D83 at the active site and in the N- and C-termini of the peptide (see Table 1). The protonated aspartate might orient the putative attacking water molecule. The neutral termini help to reduce artifacts relative to the short length of the peptide substrate. The catalytic histidine residue H212 was assumed to be δ-protonated as in Baaden and Sansom (34) (Fig. 1, *b* and *c*). Before the MM/CG simulations we equilibrated the systems A, B, C, and D via 24 ns of standard all-atom MD starting from the complex model described in Baaden and Sansom (34). S99A and H212A mutants were built by simply replacing these residues with alanine. We also performed MD simulations of a reference system (diglycine) to compare the electrostatic properties of key residues in the active site of OmpT. The systems underwent MM/CG simulations for the timescales summarized in Table 1.

TABLE 1 Systems and mutants of OmpT/ARRA complexes investigated in this work

Complex identification	Simulated time (μs)
Production runs at the MM/CG level	
A, protonated D83, neutral N- and C-termini (AceARRA NMe)	1.00
B, deprotonated D83, neutral N- and C-termini (AceARRANMe)	1.00
C, protonated D83, charged N- and C-termini (NH_3^+ ARRACOO$^-$)	0.05
D, deprotonated D83, charged N- and C-termini (NH_3^+ ARRACOO$^-$)	0.05
S99A-A	0.16
S99A-B	0.16
H212A-A	0.15
H212A-B	0.15
All-atom reference simulations	
Aeq, equilibration	0.024
Beq, equilibration	0.024
Ceq, equilibration	0.024
Deq, equilibration	0.024
Diglycine (AceGGNme)	0.015

Computational details

Preliminary MM simulations on four protomers of the OmpT-ARRA complex described in Baaden and Sansom (34) were performed for 24 ns using the Yasara software and the associated Yamber2 force field (41). Details of these equilibration simulations are provided as Supplementary Material.

Hybrid MM/CG simulations were performed using a modified version of the Gromacs 3.2.1 program (28). The enzymatic active site was treated at atomic detail with the GROMOS96 43a1 force field (40) as aforementioned (MM and I regions in Fig. 1 *a*). Approximately 400 SPC water molecules (42) were added. This constitutes a water layer of ~15 Å around the MM region (Fig. 1 *a*). This approach has been shown to accurately describe structural and dynamic features of the active site of two aspartic proteases in complex with their substrate (28) and of OmpT in the free state (36). The entire systems were composed of ~4400 particles.

The analyses of the trajectories of A and B were performed over the last 0.95 μs, that is after the equilibration phase (Supplementary Fig. 1, *a* and *b*). The analyses of the trajectories of C and D were not carried out as we will discuss later in the results section.

The leap-frog stochastic dynamics algorithm was used to integrate the equations of motion with a time step $\Delta t = 2$ fs and a friction coefficient $\gamma_i = m_i/\tau$, where $\tau = 0.5$ ps is the time constant for the coupling and m_i is the mass of the *i*th particle. A cutoff distance of 14 Å was used for the electrostatics. (This very crude assumption in the treatment of the electrostatics appears to be justified by the simplicity of the model used. Careful checks were made on energy conservation. In addition, test calculations with a longer cutoff for electrostatics (36 Å) provided very similar results to those with the shorter cutoff (data not shown)). A cutoff of 14 Å was also used for the van der Waals interactions. The pair list was updated every 10 steps. The SHAKE algorithm (43) was used to keep bonds containing hydrogens at a fixed length.

Following the 24 ns of equilibration at an all-atom MM level, the systems were relaxed by a 1-ns MM/CG simulation with positional restraints on the OmpT/ARRA complexes to minimize the energy of the solvent. Then further 1 ns with positional restraints on OmpT were performed to allow the ligand to accommodate itself inside the binding pocket under the MM/CG potentials.

Finally, we performed an atomic force-field-based MD simulation of a reference system. This is diglycine (AceGGNMe) in a periodic box of 779 water molecules. The GROMOS96 43a1 (40) and SPC (42) force fields were used for the dipeptide and water, respectively. Room conditions ($T = 300$ K, $P = 1$ bar) were achieved by coupling the system with a Berendsen thermostat (44) with $\tau = 1.0$ ps and a Berendsen barostat (44) with compressibility of 4.5 10^{-10}bar^{-1} in all three dimensions. The time step of the integration was 2 fs. Electrostatic and van der Waals interactions were calculated using a cut-off of 18 and 14 Å, respectively; 0.015 μs of trajectory were collected.

The following properties were calculated: i), large-scale motions as eigenvectors of the covariance matrix of the C_α's (45); ii), the analysis of the cosine content (46,47); iii), the electric field in the MM region, using the electrostatic term in the GROMOS96 43a1 force field (40). This analysis has been used for qualitative comparisons between different systems.

RESULTS

MM/CG of systems A and B

The structure of A and B (Fig. 1, *b* and *c*) is well maintained within the timescale investigated (1 μs): the C_α root mean-square deviations (RMSD) of A and B rise during the first ~0.05 μs and then fluctuate around an average value for the remainder of the simulations (Supplementary Fig. 1, *a* and *b*),

suggesting that this specific protein requires large sampling to equilibrate. The analysis of the cosine content (C_i) (46,47) ascertains the convergence of the simulations: C_i computed for the first 8 eigenvectors of the covariance matrix is lower than 0.5 (see Supplementary Fig. 1, *c* and *d*), meaning that the largest fluctuations are related to the potential.

A and B appear to be productive Michaelis complexes: a water molecule bridges H212 and D83, thus pointing toward the substrate carbonyl carbon. The average distance during the simulations between the catalytic water and the carbon of the scissile CO-NH bond in A and B is 3.91 (±0.35) Å and 3.50 (±0.30) Å, respectively. This water molecule is located in a position prone for a nucleophilic attack on the carbonyl group of the substrate (Fig. 2, *a* and *b*). At times a second water molecule bridges H212 and D83 and the H212-water(s)-D83 interaction is replaced by a direct H-bond between D83 and S99. Most of the key contacts at the active sites are maintained during the dynamics of both systems (Table 2; Fig. 1, *b* and *c*): i), Hδ@H212 H-bonding to Oδ@D210 or to O@D210; ii), the salt bridge between substrate R2 and E27 and D208; iii), the salt bridge between substrate R3 and D97 and D85; iv), H S99 backbone H-bonding to D83 backbone; v), D85 backbone H-bonding to D97 backbone. In addition (Table 2; Fig. 1, *b* and *c*): i), at times, S99 H-bonds to D85 and D97 and Nε@H212 H-bonds with H@R3 of the substrate; ii), Hε@H101H-bonds to either S99 or D83.

Structural fluctuations of the substrate in the binding cavity have been shown to play a functional role for cytoplasmatic proteases (4) such as HIV-1 protease (2) and BACE (3). For these enzymes it was found that the distance between the catalytic dyad and the substrate fluctuates around characteristic values corresponding to different mutual positions of the catalytic water relative to the substrate carbonyl carbon: only conformations in which the distance between the enzyme and the substrate is at a minimum turned out to be catalytically efficient (2,3). Because those motions are correlated to the large-scale motions of the proteins, the enzyme might play a role for the reaction by steering the substrate into its appropriate reactive conformation.

To ascertain whether this is also the case for OmpT, we have chosen to monitor the substrate motion within the β-barrel: i), The distance δ of the center of mass of the ARRA peptide from the center of mass of the D83-D85 pair of residues; δ is affected by the width of the cleft and, therefore, its fluctuations also modulate the position of the water inside the catalytic cleft. ii), The distance ξ of the center of mass of ARRA from the center of mass of the β-barrel (see Fig. 3, *left panel*). ξ is affected by the distance between the substrate and the catalytic water. Thus, δ and ξ can be used as suitable descriptors of the enzyme ''active'' conformations (4) to evaluate the presence of a functional mechanical coupling between the substrate and large-scale conformational fluctuations of the entire enzyme.

In A, $\langle\delta\rangle \sim 7.8\,(0.5)$ Å and $\langle\xi\rangle \sim 24.1\,(0.4)$ Å, showing a sharp Gaussian-like distribution (see Supplementary Fig. 5, *a* and *b*). In B, both quantities feature bimodal distributions: δ fluctuates from $\delta_1 = 7.3$ Å to $\delta_2 = 9.8$ Å (Fig. 3 *a*) and ξ from $\xi_1 = 25.9$ Å to ξ_2 27.5 Å (Fig. 3 *b*). Interestingly, the transitions from ξ_1 to ξ_2 and from δ_1 to δ_2 occur at the same time, suggesting that the oscillation of the cleft is correlated with the oscillation of the substrate along the axis of the β-barrel (the linear correlation between the two data sets δ and ξ is 0.60 over $\sim 10^3$ data points). No such transitions were observed in the all-atom simulations used for equilibration, where the δ/ξ space is explored to a much lower extent (see Supplementary Material for details). The presence of relevant correlations in OmpT is investigated by computing the projection of the top 10 eigenvectors of the covariance matrix on the trajectory ($\mathbf{X}\cdot\mathbf{V}_i$) (45). The correlation can be assessed in a quantitative way by analyzing the scatter plot in which the projection at each time step is plotted versus the δ and ξ distances. We find a significant correlation only with the second largest eigenvector ($\mathbf{V}_2$;

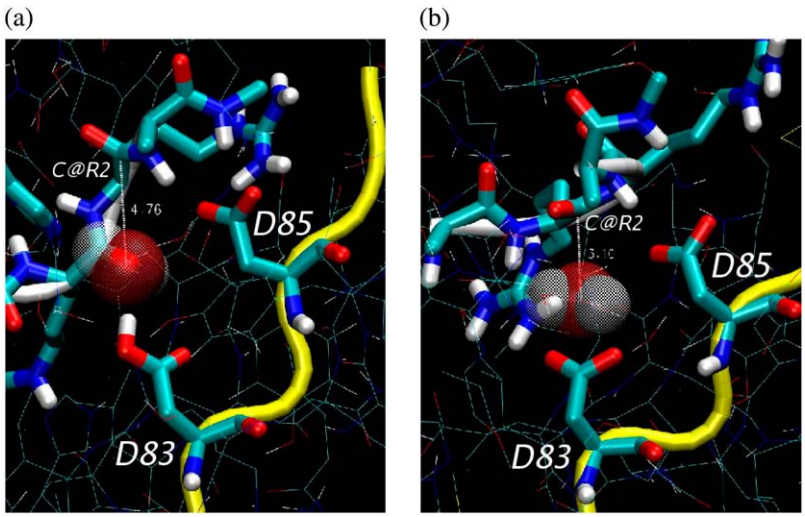

FIGURE 2 Significant snapshots of the active site of A (*a*) and B (*b*). The catalytic water is depicted in van der Waals representation with a transparent effect.

TABLE 2 MM/CG simulations of A and B, H212A-A, H212A-B, S99A-A, and S99A-B: selected MD-averaged distances (Å) within the active site of the protein

	A	B	H212A-A	H212A-B	S99A-A	S99A-B
Hδ@H212–Oδ@D210	2.6 (±0.8)	2.5 (±0.7)	–	–	7.0 (±3.5)	2.8 (±0.7)
Hδ@H212–O@D210	5.2 (±0.8)	4.3 (±1.2)	–	–	7.0 (±2.1)	5.0 (±0.8)
Nα@H212–H@R2	4.7 (±0.7)	5.0 (±1.6)	–	–	6.4 (±2.2)	5.0 (±0.7)
Nα@H212–C@R2	4.4 (±0.5)	6.2 (±1.4)	–	–	6.5 (±1.9)	4.4 (±0.7)
Nα@H212–Oδ@D83	5.8 (±1.0)	8.0 (±1.5)	–	–	11.3 (±1.8)	7.9 (±1.4)
Cζ@R2–Cδ@E27	4.8 (±0.5)	5.0 (±0.8)	6.7 (±2.4)	5.0 (±0.7)	6.2 (±1.7)	6.3 (±1.4)
Cζ@R2–Cγ@D208	5.0 (±0.7)	5.2 (±0.9)	9.0 (±2.3)	8.4 (±0.3)	7.6 (±2.1)	6.6 (±0.7)
Cζ@R2–Cγ@D210	8.5 (±0.7)	8.0 (±0.9)	10.1 (±3.0)	4.0 (±0.4)	9.5 (±1.6)	6.8 (±1.2)
Cζ@R3–Cγ@D85	6.0 (±1.0)	6.5 (±1.3)	9.7 (±1.6)	7.0 (±1.7)	9.4 (±1.4)	6.7 (±1.2)
Cζ@R3–Cγ@D97	7.0 (±1.5)	8.0 (±1.8)	11.6 (±2.1)	9.0 (±1.6)	11.8 (±3.4)	15.0 (±3.1)
Hα@H101–Oγ@S99	3.1 (±0.9)	4.2 (±1.2)	4.8 (±1.0)	3.0 (±0.7)	–	–
Oδ@D97–Hγ@S99	5.5 (±1.3)	6.6 (±2.1)	3.5 (±2.3)	8.2 (±2.6)	–	–
Hδ@D83–O@R2	4.5 (±1.1)	–	5.5 (±1.8)	–	7.5 (±1.4)	–
Hδ@D83–Oδ@D97	8.8 (±1.2)	–	7.9 (±1.3)	–	4.0 (±1.2)	–
Hα@H101–Oδ@D83	3.7 (±1.0)	3.5 (±1.1)	4.6 (±2.0)	3.1 (±0.9)	5.0 (±1.1)	3.5 (±0.9)
H@D83–O@S(A)99	2.0 (±0.2)	2.0 (±0.3)	2.0 (±0.2)	2.0 (±0.2)	2.2 (±0.4)	4.0 (±0.6)
O@D83–H@S(A)99	2.0 (±0.3)	2.4 (±0.5)	2.5 (±0.6)	2.0 (±0.2)	2.6 (±0.7)	2.5 (±0.9)
H@D85–O@D97	2.1 (±0.4)	2.3 (±0.3)	2.3 (±0.7)	2.4 (±0.7)	3.0 (±1.2)	5.0 (±1.8)
O@D85–H@D97	2.0 (±0.4)	2.6 (±1.0)	2.1 (±0.4)	2.7 (±0.7)	3.5 (±1.6)	5.3 (±1.7)

Standard deviations are reported in parenthesis.

Fig. 3, a' and b'). In fact, two mostly populated regions corresponding to δ_1 and δ_2 and to ξ_1 and ξ_2, are visible.

We conclude that $\mathbf{V}_2$ induces relevant variations in the relative distances of the active site, showing high correlations with both distances. In contrast, no correlations are found with the first eigenvector $\mathbf{V}_1$, which represents the first most dominant large-scale motion. Because the cosine content associated to $\mathbf{V}_1$ is relatively larger ($C_1 \sim 0.5$, whereas $C_2 \sim 0.1$; see Supplementary Fig. 1 *d*), $\mathbf{V}_1$ might not represent a real ''coherent'' motion, but, rather, a random diffusion of less structured parts of the protein.

The large-scale motion described by $\mathbf{V}_2$ mostly affects the solvent-exposed loops embracing the active site, as found in previous work (36). In addition, the loops L2, L3, L4, and L5 as well as the substrate are more mobile in system B with respect to A (Supplementary Fig. 4 *a*). This is consistent with the fact that the substrate is significantly anticorrelated to the motion of the loops in system B (Supplementary Fig. 4 *b*).

Thus, large-scale fluctuations in B allow a well-defined motion of the substrate within the catalytic cleft concerted with that of loops, similarly to what is found for several proteases (4), such as HIV-1 PR (2) and BACE (3).

Our approach also allows to investigate the polarization of the reactants relative to a reference system (here the Gly-Gly dipeptide) in water. To this aim, we monitor the electric field along the substrate carbonyl group C@R2 of the substrate and the catalytic water C_2 axis (Fig. 1, *d* and *e*), which has been proposed to be the nucleophile agent for the protein.

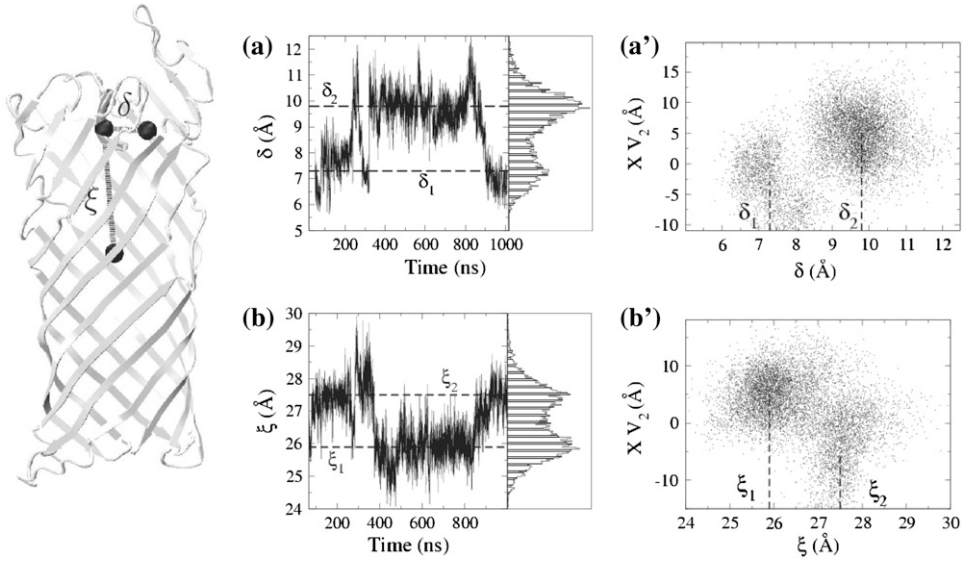

FIGURE 3 (*Left of the panel*) Cartoon representation of the OmpT/ARRA complex. Dashed lines depict δ and ξ distances defined in the text. (*Right of the panel*) Time evolution in B of δ (*a*) and ξ (*b*) parameters (defined in the text), respectively. (*Right*) Plot of $\mathbf{X}\cdot\mathbf{V_2}$ vs. δ (a') and ξ (b') distances for B.

As for the substrate, the field of A does not show any preferential direction with respect to the C=O bond: the angle α between the field and the vector identified by the C=O bond (Fig. 1 *d*) is spread within 0° and 180°: the standard deviation value is large around its average value ($\langle\alpha\rangle = 125°(44°)$), and not much smaller than that of the reference system here investigated, for which the field is completely isotropic ($\langle\alpha\rangle = 100°(59°)$). In contrast, in B the field is partially aligned with the C=O bond (Fig. 1 *d*) with far lower spread ($\langle\alpha\rangle = 140°(10°)$). We conclude that the carbonyl carbon is expected to be more electrophilic and thus reactive than in A and in aqueous solution.

Similar conclusions can be drawn for the catalytic water: the field in A is orthogonal to the water C_2 axis (Fig. 1 *e*, the average angle β between the field and the C_2 axis of the catalytic water is $\sim$104°(14°)); hence it is not expected to significantly affect the nucleophilic power of the water molecule. Notice that in this case the spread is much smaller than that of the reference system ($\langle\beta\rangle\sim93°(60°)$). In B, in contrast, the angle β shows a bimodal distribution, either orthogonal ($\langle\beta\rangle\sim90°(12°)$), or aligned to the C_2 axis ($\langle\beta\rangle\sim20°(8°)$). Thus, in the latter case the field in B might render the catalytic water more nucleophilic than in A and in aqueous solution.

The fluctuation of the β-angle between β_1 and β_2 values in B is driven by $\mathbf{V}_2$ (see Supplementary Fig. 6). This means that the functional oscillations of the substrate allow the catalytic water molecule to take a well-defined geometric configuration, able to stabilize a more negative charge distribution on the oxygen and resulting in an enhanced nucleophilicity of the oxygen atom itself.

We conclude that the polarization effect may be used by the enzyme to enhance the electrophilicity of the carbonyl carbon and the nucleophilicity of the water oxygen. The same calculation in the timescale typical of all-atom MD simulations ($\sim$50 ns) shows no preferential directions of the field in both systems (the α and β distributions are spread between 0° and 180°). This is somewhat confirmed in the all-atom equilibration simulations where no significant spread in the α and β distributions is reached (see Supplementary Material, Supplementary Fig. 11) due to insufficient sampling.

MM/CG of systems C and D

The Michaelis complex in C and D disrupts already after 10 ns (Supplementary Fig. 2, *a* and *b*). The NH_3^+ terminal groups, which formed salt bridges with D210 and E27 at the beginning of the dynamics, rotate and form a salt bridge with D83 and an H-bond with Nε@H212 (Supplementary Fig. 2, *a* and *b*). Consequently, the carbonyl carbon, which undergoes the nucleophilic attack, moves away from the putative catalytic dyad composed by H212 and D83 (see Supplementary Fig. 2, *a* and *b*), whereas the oxygen of such a group flips, pointing toward the D210 and H212 dyad. Such a nonproductive Michaelis complex, in which there is no putative nucleophilic agent in the close proximity of the carbonyl carbon, is maintained for an additional $\sim$0.04 μs, after which we decided to stop the simulation. We conclude that the presence of the charged termini in addition to the two positive arginines affects the structure of the Michaelis complexes in both C and D.

MM/CG of H212A and S99A mutants

Here we compare MM/CG simulations of H212A and S99A OmpT, which show a residual activity ranging within 0% and 4% (32), with respect to wt. We ran the simulations on the mutants for protomers A and B for a shorter timescale than that of the wt (0.15 μs), as we are solely interested in constructing structural models.

In H212A, the mutation disrupts the H-bonds with Nε@H212 and the amide group of the substrate, which contributes to maintain the position of the substrate inside the enzymatic cleft. In the wt, R2 and R3 of the substrate form salt bridges with E27 and D208 and with D85 and D97, respectively. These two residues rotate in both A and B and their side chains face the solvent (Table 1 and Supplementary Fig. 7, *a* and *b*). At the end of the simulation, the carbonyl carbon, which is cleaved in the wt enzyme, moves away from the putative catalytic dyad H212-D83 and the cleft is filled by water.

In S99A, the replacement of S99 with alanine disrupts the S99-Hε@His-101 H-bond (Supplementary Fig. 8). The H atom of the same histidine H-bonds to D83 in A and B.

In A, the breaking of the S99-H101 H-bond causes a rearrangement of H101 and consequently D83, which is protonated, rearranges and H-bonds to D97 (Table 2 and Supplementary Fig. 8 *a*). Consequently, the water-mediated interaction between the proton of D83 and the substrate is lost, and the latter fluctuates allowing a rotation of the side chain of R2 of the substrate. As a result, R2 H-bonds with D210 and the catalytic residue H212 is permitted to move further apart from the active site (Table 2). This causes a drastic change of the substrate configuration and its partial detachment (Supplementary Fig. 8 *a*).

In B, D83, which is ionized, forms a stable H-bond with Nε@H101. However, because of the lack of S99-D97 H-bonding, D97 moves away, in turn causing the loss of H-bond interactions between the backbone of D85 and D97 (Supplementary Fig. 8 *b*). As a result, the salt bridge between D85, D97, and R3 is lost, allowing the side chain of the latter residue to rotate and become solvated. In this case, a partial detachment of the substrate occurs (Table 2; Supplementary Fig. 8 *b*).

DISCUSSION

MM/CG simulations have been used to investigate the fluctuations of OmpT in complex with its substrate ARRA

on the microsecond timescale as well as the effect of key point mutations at the active site. We have focused on four systems (A–D), which differ for the protonation state of D83 and the charge of the substrate (AceARRANMe or NH_3^+ ARRA COO^-), which have first been simulated by all-atom MD simulations on a shorter timescale.

The complexes with substrates with charged tails (C and D) evolved to nonproductive Michaelis complexes, because the charged tail groups interact with the residues among the catalytic cleft causing a distortion of the ARRA peptide and its detachment from the binding pocket. Interestingly the preliminary 24-ns all-atom simulations did not show a disruption of the Michaelis complex, which points to a less efficient exploration of conformational space compared to the MM/CG approach, where the systems are described by smoother potential energy surfaces. Obviously, our investigation does not rule out the presence of longer substrates with charged tail groups because those substrates may have their termini located outside the catalytic cleft.

In contrast, the complexes with the substrates with neutral tails provided productive Michaelis complexes and both remained stable over 1-μs dynamics simulations. The two complexes (*A* and *B* in Fig. 1, *b* and *c*) exhibit different protonation states of D83: in A, the aspartic acid is protonated whereas in B it is ionized. The systems are characterized by significant differences in the electrostatic polarization of the reactants. In B, the active site polarizes both the catalytic water and the carbonyl carbon, rendering the first more nucleophilic and the second more electrophilic relative to a reference system in water (diglycine). In contrast, in A no significant polarization is observed and the electric field acting on the carbonyl carbon bond and on the catalytic water is very similar to that calculated for the reference system diglycine in water solution.

System B is also characterized by large conformational fluctuations of the substrate triggered by global large-scale motions (see Fig. 3, *a′*and *b′*), which populate significantly different conformations (see Fig. 3, *a* and *b*, and Supplementary Fig. 3). This contrasts to A in which the substrate fluctuates around a well-defined conformation (see Supplementary Figs. 3 *a* and 5 *a* and *b*). The large-scale fluctuations of complex B might have functional implications as found in the aspartyl proteases superfamily (2–4). However, quantum chemical calculations on the reaction mechanism are required to discriminate between the two protomers and further address this issue.

We have next used the MM/CG approach on H212A and S99A mutants (for both A and B protomers) to provide the structural basis for the much lower activity of these mutants relative to wt. This result, which is easy to rationalize for the H212A mutant, as it is part of the putative catalytic dyad, is rather intriguing for S99A, which is not involved directly in the catalytic cleft. Our simulations suggest that in H212A the ARRA peptide detaches spontaneously in both systems due to the loss of the H-bond interaction between Nε@H212 and the amide group of the substrate, present in the wt. S99A, indeed, allowed a similar detachment of the peptide due to the disruption of the geometry of the active site in both systems, in spite of the fact that this residue is not located at the active site. In A the loss of the H-bond interaction between S99 and D83 allows the rotation of the latter residue causing the breaking of interactions between D83 and the substrate. In contrast, in B the detachment of the peptide is due to the loss of the salt bridge between R3 and D85-D97 after the breaking of the S99-D97 interaction. We conclude that not only first-shell H-bond interactions (such as those formed by H212), but also second shell H-bonding (such as that of S99) play an important role for the stability of the geometry of the active site. Removing any of those stabilizing interactions may cause a high instability of the active site and therefore a reduced activity.

In conclusion, our MM/CG approach emerges as a useful tool to investigate microsecond simulations of enzymes, which is presently difficult with all-atom MD. Our conclusions about the electric field acting on the reactants and about the motion of the substrate inside the catalytic cleft cannot be drawn if one focuses on the typical timescale of all-atom MD (0.01–0.1 μs). This suggests that this approach, within its limitations deriving from the use of a coarse-grained model for modeling the most of the protein in solution, may provide useful information—complementary to all atom MD—on phenomena occurring on relatively long timescales. In addition, it may be useful for computational molecular biology, allowing one to test the effect of point mutations via a computationally affordable method. Thus, MM/CG, by allowing to run more numerous and longer simulations than all-atom MD, is expected, on the one hand to improve our confidence in the results, and on the other one it may strengthen the interaction between molecular biology experiments and simulations.

SUPPLEMENTARY MATERIAL

To view all of the supplemental files associated with this article, visit www.biophysj.org.

The authors thank Alessandra Magistrato and Michele Cascella for discussions.

We acknowledge financial support by MURST-FIRB and IIT.

REFERENCES

1. Rod, T. H., J. L. Radkiewicz, and C. L. Brooks III. 2003. Correlated motion and the effect of distal mutations in dihydrofolate reductase. *Proc. Natl. Acad. Sci. USA*. 100:3954–3959.
2. Piana, S., P. Carloni, and M. Parrinello. 2002. Role of conformational fluctuations in the enzymatic reaction of HIV-1 protease. *J. Mol. Biol.* 319:567–583.
3. Cascella, M., C. Micheletti, U. Rothlisberger, and P. Carloni. 2005. Evolutionarily conserved functional mechanics across pepsin-like and retroviral aspartic proteases. *J. Am. Chem. Soc*. 127:3734–3742.

4. Carnevale, V., S. Raugei, C. Micheletti, and P. Carloni. 2006. Convergent dynamics in the protease enzymatic superfamily. *J. Am. Chem. Soc.* 128:9766–9772.
5. Fersht, A. R. 1999. Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding. W. H. Freeman, NY.
6. Eisenmesser, E. Z., D. A. Bosco, M. Akke, and D. Kern. 2002. Enzyme dynamics during catalysis. *Science.* 295:1520–1523.
7. Benkovic, S. J., and S. Hammes-Schiffer. 2003. A perspective on enzyme catalysis. *Science.* 301:1196–1202.
8. Daniel, R. M., R. V. Dunn, J. L. Finney, and J. C. Smith. 2003. The role of dynamics in enzyme activity. *Annu. Rev. Biophys. Biomol. Struct.* 32:69–92.
9. Luo, J., and T. C. Bruice. 2004. Anticorrelated motions as a driving force in enzyme catalysis: the dehydrogenase reaction. *Proc. Natl. Acad. Sci. USA.* 101:13152–13156.
10. Wolf-Watz, M., V. Thai, K. Henzler-Wildman, G. Hadjipavlou, E. Z. Eisenmesser, and D. Kern. 2004. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat. Struct. Mol. Biol.* 11:945–949.
11. Tirion, M. M. 1996. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 77:1905–1908.
12. Noguti, T., and N. Go. 1982. Collective variable description of small-amplitude conformational fluctuations in a globular protein. *Nature.* 296:776–778.
13. Shelley, J. C., M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, and M. L. Klein. 2001. A coarse grain model for phospholipid simulation. *J. Phys. Chem. B.* 105:4464–4470.
14. Shelley, J. C., M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, and M. L. Klein. 2001. Simulation of phospholipids using a coarse grain model. *J. Phys. Chem. B.* 105:9792–9875.
15. Lyubartsev, A. P. 2005. Multiscale modelling of lipids and lipid bilayers. *Eur. Biophys. J.* 35:53–61.
16. Izvekov, S., and G. A. Voth. 2005. A multiscale coarse-graining methods for biomolecular systems. *J. Phys. Chem. B.* 109:2469–2473.
17. Brannigan, G., L. C. L. Lin, and F. L. H. Brown. 2006. Implicit solvent simulation models for biomembranes. *Eur. Biophys. J.* 35:104–124.
18. Shi, Q., S. Izvekov, and G. A. Voth. 2006. Mixed atomistic and coarse-grained molecular dynamics: simulation of a membrane bound ion channel. *J. Phys. Chem. B.* 110:15045–15048.
19. Ayton, G. S., and G. A. Voth. 2007. Multiscale simulation of transmembrane proteins. *J. Struct. Biol.* 157:570–578.
20. Tozzini, V. 2005. Coarse-grained model for proteins. *Curr. Opin. Struct. Biol.* 15:144–150.
21. Ding, F., W. Guo, N. V. Dokholyan, E. I. Shakhnovich, and J. E. Shea. 2005. Reconstruction of the Src-Sh3 protein domain transition state ensemble using multiscale molecular dynamics simulations. *J. Mol. Biol.* 350:1035–1050.
22. Gohlke, H., and M. F. Thorpe. 2006. A natural coarse graining for simulating large biomolecular motion. *Biophys. J.* 91:2115–2120.
23. Chu, G. W., and G. A. Voth. 2006. Coarse-grained modelling of the actin filament derived from atomistic-scale simulations. *Biophys. J.* 90:1572–1582.
24. Villa, E., A. Balaeff, L. Mahadevan, and K. Schulten. 2004. Multiscale method for simulating protein-DNA complexes. *Multiscale Model. Simul.* 2:527–553.
25. Sun, J., Q. Zhang, and T. Schlick. 2005. Electrostatic mechanism of nucleosomal array folding revealed by computer simulation. *Proc. Natl. Acad. Sci. USA.* 102:8180–8185.
26. Villa, E., A. Balaeff, and K. Shulten. 2005. Structural dynamics of the lac repressor-DNA complex revealed by a multiscale simulation. *Proc. Natl. Acad. Sci. USA.* 102:6783–6788.
27. Reference deleted in proof.
28. Neri, M., C. Anselmi, M. Cascella, A. Maritan, and P. Carloni. 2005. Coarse-grained model of proteins incorporating atomistic detail of the active site. *Phys. Rev. Lett.* 95:218102.
29. Vandeputte-Rutten, L., R. A. Kramer, J. Kroon, N. Dekker, M. R. Egmond, and P. Gros. 2001. Crystal structure of the outer membrane protease OmpT from Escherichia coli suggests a novel catalytic site. *EMBO J.* 20:5033–5039.
30. Mangel, W. F., D. L. Toledo, M. T. Brown, K. Worzalla, M. Lee, and J. J. Dunn. 1994. Omptin: an Escherichia coli outer membrane proteinase that activates plasminogen. *Methods Enzymol.* 244: 384–399.
31. Sodeinde, O. A., Y. V. Subrahmanyam, K. Stark, T. Quan, Y. Bao, and J. D. Goguen. 1992. A surface protease and the invasive character of plague. *Science.* 258:1004–1007.
32. Kramer, R. A., N. Dekker, and M. R. Egmond. 2000. Identification of active site serine and histidine residues in Escherichia coli outer membrane protease OmpT. *FEBS Lett.* 468:220–224.
33. Dekker, N., R. C. Cox, R. A. Kramer, and M. R. Egmond. 2001. Substrate specificity of the integral membrane protease OmpT determined by spatially addressed peptide libraries. *Biochemistry.* 40:1694–1701.
34. Baaden, M., and M. S. P. Sansom. 2004. OmpT: molecular dynamics simulations of an outer membrane enzyme. *Biophys. J.* 87:2942–2953.
35. Tai, K., M. Baaden, S. Murdock, B. Wu, M. H. Ng, S. Johnston, R. Boardman, H. Fangohr, K. Cox, J. W. Essex, and M. Sansom. 2007. Three hydrolases and a transferase: comparative analysis of active-site dynamics via the BioSimGrid database. *J. Mol. Graph. Model.* 25:896–902.
36. Neri, M., C. Anselmi, V. Carnevale, A. V. Vargiu, and P. Carloni. 2006. Molecular dynamics simulations of outer-membrane protease T from *E. coli* based on a hybrid coarse-grained/atomistic potential. *J. Phys. Condens. Matter.* 18:S347–S355.
37. Go, N., and H. A. Scheraga. 1976. On the use of classical statistical mechanics in the treatment of polymer chain conformations. *Macromolecules.* 9:535–542.
38. Doi, M. 1996. Introduction To Polymer Physics, 1st Ed. Oxford Science Publications, Oxford, UK.
39. van der Spoel, D., E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen, and H. J. C. Berendsen. 2004. Gromacs User Manual version 3.2. GROMACS Development Team, Groningen, The Netherlands.
40. van Gunsteren, W. F., S. R. Billeter, A. A. Eising, P. H. Hünenberg, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi. 1996. Biomolecular Simulation: The GROMOS96 Manual and User Guide. Hochschulverlag AG an der ETH Zurich, Zurich, Switzerland.
41. Krieger, E., T. Darden, S. B. Nabuurs, A. Finkelstein, and G. Vriend. 2004. Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins.* 57:678–683.
42. Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. 1981. Interaction models for water in relation to protein hydration. *In* Intermolecular Forces. B. Pullman, editor. Reidel, Dordrecht, The Netherlands. 331–342.
43. Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23: 327–341.
44. Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.
45. Amadei, A., A. B. M. Linssen, and H. J. C. Berendsen. 1993. Essential dynamics of proteins. *Proteins.* 17:412–425.
46. Hess, B. 2000. Similarities between principal components of protein dynamics and random diffusion. *Phys. Rev. E.* 62:8438–8448.
47. Hess, B. 2002. Convergence of sampling in protein simulations. *Phys. Rev. E.* 65:031910.