Eur Biophys J (2008) 38:13–23
DOI 10.1007/s00249-008-0348-2

ORIGINAL PAPER

Exploration of the activation pathway of Δα-Chymotrypsin with molecular dynamics simulations and correlation with kinetic experiments

Janka Mátrai · Abel Jonckheer · Eddy Joris · Peter Krüger · Eric Carpenter · Jack Tuszynski · Marc De Maeyer · Yves Engelborghs

Received: 27 February 2008 / Revised: 25 May 2008 / Accepted: 4 June 2008 / Published online: 27 August 2008

Abstract Correlating the experimentally observed kinetics of protein conformational changes with theoretical predictions is a formidable and challenging task, due to the multitude of degrees of freedom (>5,000) in a protein and the huge gap between the timescale of the kinetic event of interest (ms) and the typical timescale of computer simulations (ns). In this study we show that using the targeted molecular dynamics (TMD) method it is possible to simulate conformational changes of the ms time range and to correlate multiple simulations of single pathways with ensemble experiments on both the structural and energetic basis. As a model system we chose to study the conformational change of rat-Δα-chymotrypsin from its inactive to its active conformation. This activation process has been analyzed previously by experimental and theoretical methods, i.e. fluorescence stopped-flow spectroscopy (FSF), molecular dynamics (MD) and TMD. Inspired by the results of these studies on the wild type (WT) enzyme, several mutants were constructed to alter the conformational pathway and studied by FSF measurements. In the present work WT and mutant N18G were subjected to multiple MD and subsequent TMD simulations. We report the existence of two main activation pathways, a feature of chymotrypsin activation that has been hitherto unknown. A method to correlate the energetics of the different pathways calculated by TMD and the kinetic parameters observed by experimental methods such as FSF is presented. Our work is relevant for experimental single molecule studies of enzymes in general.

Keywords Molecular dynamics · Targeted molecular dynamics · Fluorescence stopped flow · Multiple simulations · Conformational changes · Kinetics

Electronic supplementary material The online version of this article (doi:10.1007/s00249-008-0348-2) contains supplementary material, which is available to authorized users.

J. Mátrai · A. Jonckheer · E. Joris · P. Krüger · Y. Engelborghs
Laboratory of Biomolecular Dynamics, Katholieke Universiteit Leuven, Celestijnenlaan 200 G-box2403, 3001 Leuven, Belgium

Present Address:
J. Mátrai (✉)
Centre of Gene Therapy and Transgene Technologies, Campus Gasthuisberg, KUL-VIB3, Herestraat 49 bus 912, 3000 Leuven, Belgium
e-mail: janka.matrai@med.kuleuven.be

Present Address:
P. Krüger
ALA Analytisches Labor GmbH, Charlottenstrasse 14, 52070 Aachen, Germany

E. Carpenter · J. Tuszynski
Department of Physics, University of Alberta, Edmonton, AB, Canada T6G 2J1

E. Carpenter · J. Tuszynski
Cross Cancer Institute, 11560 University Avenue, Edmonton, AB, Canada T6G 1Z2

M. De Maeyer
Laboratory of Biomolecular Modelling and BioMacS, Katholieke Universiteit Leuven, Celestijnenlaan 200G, 3001 Leuven, Belgium

Introduction

Chymotrypsin activation is used as a model system, to study the relation between the calculated structural and energetic features of the conformational change and the experimental kinetic data.

Chymotrypsinogen is physiologically activated by the proteolytic cleavage of the peptide bond between R15-I16 (Fig. 1), followed by the rotation of residue I16 to form a buried salt-bridge (Wang et al. 1985) and by further hydrolytic steps. The conformational change leading to enzymatic activation is composed of the following main steps (Fig. 1): Loop7 (lp7) and loop6 (lp6) shift sideways in a gate-opening movement, enlarging the space around segment 2 (sg2) and facilitating the movement of I16 from the solvent into the protein interior (Matrai et al. 2004; Wroblowski et al. 1997). Next to the rigid body motion of lp7 and lp6, which are coupled via H-bonds and the disulfide bridge C191-C220 (residues C191 and C220, located in lp6 and lp7, see Fig. 1), lp6 undergoes a more individual, rotary motion, linked to the outward rotation of the side chain of M192. As a result, residue G193, which is part of the substrate-binding pocket, occupies its active position with a correct orientation of its NH group. Furthermore, space is created allowing residue I16 to rotate inwards and to form the stabilizing salt bridge between its amino terminus and the side chain of D194 (Fig. 1).

The transition from an already cleaved inactive (I) form to the active (A) form can be induced in vitro by a basic to neutral pH change (Fersht and Requena 1971; Stoesz and Lumry 1978). The extreme states of this conformational change have been characterized in great detail by X-ray crystallography (Tsukada and Blow 1985; Wang et al. 1985) and by IR spectroscopy (Heremans and Heremans 1989). The kinetics of the activation process have been followed by fluorescence stopped flow (FSF) experiments (Verheyden et al. 2004) while the structural features of the transition were explored by in silico simulations (Matrai et al. 2004; Wroblowski et al. 1997).

Computational methods such as molecular dynamics (MD) can be used to investigate conformational changes. However, the activation of chymotrypsin is a millisecond time scale process, out of reach for the standard MD method currently available even using supercomputers. Several methods were developed to simulate extensive conformational changes in proteins, e.g. the transition network method (Noé et al. 2006), jumping-among-minima methods (Kitao et al. 1998), minimum energy path (Fischer et al. 2005), steered MD (Gao et al. 2002), conformational flooding (Grubmuller 1995), elevated temperature simulations (Oroguchi et al. 2007) and targeted molecular dynamics (TMD) (Kruger et al. 2001; Schlitter et al. 1993; Wroblowski et al. 1997).

TMD is among the often-used methods that investigate large-scale conformational changes and roles of individual amino acids and ligand molecules (Jacoby et al. 1996; Kamerlin et al. 2007; Kruger et al. 2001; Kuppens et al. 1999; Noy et al. 2007; Perdih et al. 2007; Yu et al. 2007). In TMD distance constraints between the corresponding

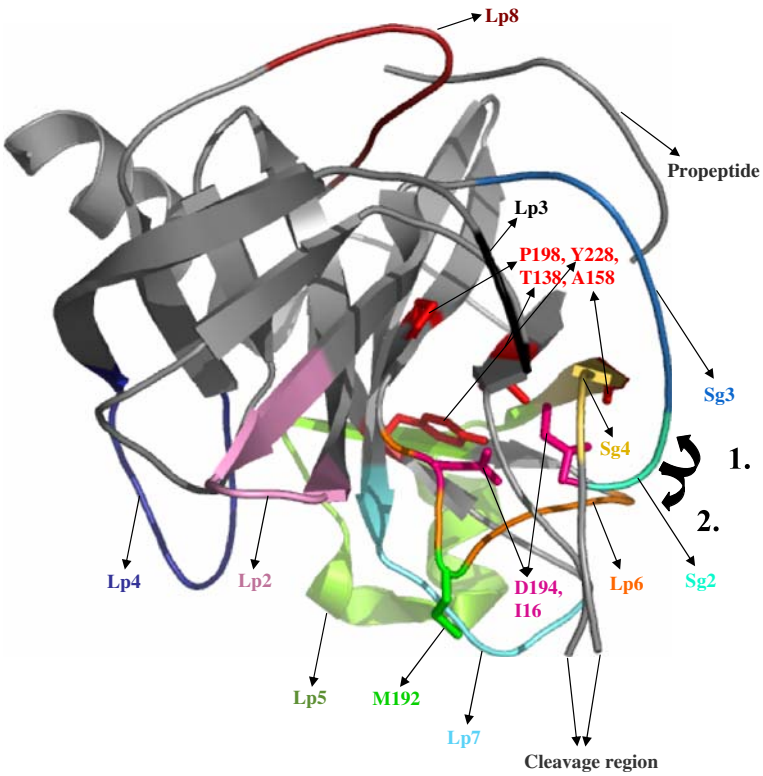

Fig. 1 The important residues and regions that are involved in the conformational change. The regions are: sg2: I16-G19, sg3: E20-W27, sg4: L153-P161, lp2: S33-H40, lp3: Q73-S77, lp4: K90-D102, lp5: I162-C182, lp6: S186-S195, lp7: S214-S221, lp8: Q113-L123, where the 'lp' and 'sg' expressions indicate the loops and segments respectively. The *black thick arrows* indicate the two main rotational directions of sg2 during the activation process. Direction 1 is the less solvent exposed route while direction 2 indicates the more solvent exposed rotational pathway. The figure is made from the 4CHA structure

atoms of two conformations are applied, in this study between the inactive and active forms of chymotrypsin, respectively. The distance between the initial and final (target) structure is reduced in a stepwise manner, allowing the modelling of different transition pathways between the two conformations (Schlitter et al. 1993; Wroblowski et al. 1997).

Inspired by previous TMD studies carried out on wild-type bovine α-chymotrypsin (Wroblowski et al. 1997) and on the rat variant with and without the α-chain (propeptide, see Fig. 1) (Matrai et al. 2004), several mutants were constructed from the latter (Verheyden et al. 2000). The effect of these mutations on the activation kinetics was measured by FSF (Verheyden et al. 2004). From this collection, an analysis with MD and TMD of the rat-Δα-chymotrypsin, WT and the N18G mutant is presented in this paper. This mutant was chosen for its interesting location in the protein (in sg2) (Fig. 1) and for its effect on the kinetics of the activation process. This mutant shows the largest difference in kinetic parameters relative to the WT (Verheyden et al. 2004). Residues in sg2, such as N18 may play a crucial role in the activation process by forming an extensive H-bonding network between sg2 and other regions of the protein e.g. via residues G187 and S189 in lp6 (Fig. 1) (Matrai et al. 2004). Other residues of sg2, especially G19, ensure its flexibility by acting as hinges (Matrai et al. 2004; Wang et al. 1985). Mutant N18G was expected to enhance the flexibility of sg2, and to accelerate the activation (Verheyden et al. 2004). The FSF experiments indeed demonstrate that N18G accelerates the activation and, very interestingly, it does so in an entropically favoured manner, since both the activation entropy ($\Delta S^{\#}$) and enthalpy ($\Delta H^{\#}$) are higher than their counterparts for the WT (Verheyden et al. 2004). In order to find out how this mutation affects the activation process, multiple MD and TMD studies were performed on both the WT and the mutant and a procedure is presented how to correlate the set of multiple pathways with the experimental data.

A discussion concerning the effect of the sequence difference between the bovine and rat proteins and the absence of the α-propeptide chain (Δα variants) was presented elsewhere (Matrai et al. 2004). The Asn residue at position 18 is conserved in the bovine and rat chymotrypsin variants (Kardos et al. 1999).

The main objectives of this study are: (1) to characterize the conformational pathways of the WT and the mutant from both the structural and energetic points of view; (2) to obtain the activation energies associated with these pathways and (3) to correlate the simulation results with the experimental data.

Materials and methods

Molecular dynamics calculations

The starting structures (active and inactive) for the simulations of WT and mutant N18G were derived from the bovine α-chymotrypsin structure (WT), pdb code [Brookhaven Protein Data Bank (Bernstein et al. 1977)], 4CHA (1.68 Å resolution, *R* value 0.234) (Tsukada and Blow 1985) and 2CGA (1.80 Å resolution, *R* value 0.170) (Wang et al. 1985) of chymotrypsin and chymotrypsinogen, respectively. The same methodology in obtaining the mutant structure was used as previously published (Matrai et al. 2004; Verheyden et al. 2000). In order to generate the chymotrypsinogen-like inactive structure that is present at high pH (∼11) (Verheyden et al. 2004), we deleted the residues Thr147, Asn148 and the complete propeptide (first 15 residues) from the original chymotrypsinogen coordinates (2CGA). The missing atoms of the newly generated C termini were added using the Gromos96 program. Polar hydrogens were added using standard geometries. All other nonpolar hydrogen atoms (CH, CH2, CH3) were included in the corresponding carbon atom. Because the coordinates of rat chymotrypsin and rat chymotrypsinogen are not available, the rat active and inactive conformations without the propeptide were modeled. A reasonable modeling of the structures was possible because of the high sequence homology between rat and bovine chymotrypsin. For modeling the rat chymotrypsin structures the MUTATE option of WHATIF was used and the placement of the amino acid chain was optimized using the option DEBUMP. The original enzyme contains three chains, i.e. the α (the propeptide), β and γ chain. In this study the propeptide was omitted since it does not play an essential role in the activation (Matrai et al. 2004).

For the WT and the N18G mutant, 25 independent MD and activation TMD simulations were carried out on an Indigo 2 Solid Impact R10000 workstation, an Apple cluster of seven Xserve G5, the supercomputer facilities of Westgrid (2004), Canada (Westgrid) and the KULeuven computer centre cluster (https://www.ludit.kuleuven.be/hpc/wiki/index.php/Main_Page#Cluster_information), all using the Gromos96 software package (Van Gunsteren et al. 1996) with the TMD module (Schlitter et al. 1993). The multiple MD simulations were performed with different random number seeds corresponding to different initial velocities. The inactive and active structures obtained after 200 ps MD simulation were used respectively in the TMD simulations as start and end structures. The TMD simulation setup was tested for different time scales from 16 ps to 10 ns. Details are summarized in the results section. Based on these tests we decided to perform

each 25 TMD run over a time span of 500 ps. Each simulation was done in the presence of explicit water ($\sim$15,000 water molecules). The structures were placed in a truncated octahedral box of SPC water (Berendsen et al. 1981) with a minimum distance of 1.05 nm between the protein and the wall of the box. The integration step was 2 fs, non-bonded (Van der Waals and electrostatic) interactions closer to 1.4 nm were included, and the non-bonded pairs were updated every ten steps.

The system was relaxed using 200 steps of steepest descent energy minimization. To simulate the experimental neutral pH environment that was needed for the activation, counter-ions were applied. The acidic and basic amino acids (including His) receive a negative, respectively positive charge in the force field. In order to compensate these charges, counter-ions, i.e 21 Na^+ and 15 Cl^- ions were added resulting in a system globally without charge. A deprotonated acid (charge -1) or a protonated base (charge $+1$) corresponds for most amino acids to the situation at pH = 7. Then the system was relaxed again under the same conditions as before. During the MD simulations the reference temperature for the heat bath was set to 300 K with a coupling constant of 0.1 ps. No pressure coupling was applied. A cut-off of 8 Å for the pair-list construction and 14 Å for the long-range interactions and reaction-field calculations was used. Shake was applied for all bonds of the solute with a tolerance of 10^{-3} Å. No explicit force-field term for hydrogen bonds was included. The system was simulated using two gromos energy groups, one containing the protein and one the ions and solvent.

Targeted molecular dynamics calculations

The input parameters and conditions of the TMD simulations (Schlitter et al. 1993; Swegat et al. 1997) were as follows. The TMD method establishes a distance constraint between two structures. Two structures are necessary to perform a simulation, an initial and a final structure. The pathway between the initial and the finial structure describes the direction the simulation takes. In each step of the simulation first an unconstrained MD step is performed. After the free simulation step, the distance between the initial and target structure has to be reduced. The changes of the distance constraint between the current and the final structure were linear with time. The simulation stopped if the mean distance per atom was reduced below 0.02 nm. During the simulation the distance to the final structure (ρ) at each simulation step was reduced by $\lambda_F\rho$, where λ_F was set to 0.8. It is good practice to set λ_F close to this value because values closer to 1 would generate unrealistic static structures at decreasing ρ. This would happen because near $\rho = (3N\sigma^2)^{1/2}$, with σ representing the mean thermal standard deviation per coordinate, a strongly reduced dynamics is to be expected (Schlitter et al. 1993). The remaining parameters were set as for the MD simulations.

Data analysis

For the analysis of the structures along the activation pathways SIMLYS (Kruger and Szameit 1992), Swiss-PdbViewer 'spdbv' (Guex and Peitsch 1997) and the Brugel program package (De Maeyer et al. 1997; Delhaise et al. 1988; Desmet et al. 1992) were used. Graphics were prepared with Rasmol 2.7.1.1 (Bernstein 2000), Swiss-PdbViewer (Guex and Peitsch 1997), PyMol (DeLano 2002), Molscript (Kraulis 1991), Futuris Imager (Tereschenko 1999), and Sigma Plot 8.0 (2001). The results were analyzed for (1) main-chain changes in ϕ and ψ angles, (2) H-bond network of sg2, by monitoring the motion of residues I16, M192, G193 and D194, (3) 'best plane' projection (Diamond and Levitt 1971; Levitt 1983) and (4) estimating the activation enthalpy $DH^{\#}$. The non-bonded energy of H-bonds was calculated using the default distance, angle and atom type criteria of the Brugel program (De Maeyer et al. 1997; Delhaise et al. 1988; Desmet et al. 1992). The force field used in Brugel is the Charmm force field in a full atom model.

The correct orientation of the peptide NH group of G193, the possibility of forming H-bonds between the backbone of residue G193 and a bound inhibitor motif was checked as follows. The structures along the MD trajectories were main-chain fitted on the chymotrypsin structure taken from the pdb file 1ACB. The coordinates of the eglin c inhibitor in 1ACB were merged with the coordinates of the simulated structure and it was determined whether residue G193 could form H-bonds to the imported inhibitor.

During the analysis, unless mentioned otherwise, the 'Cα–CO–N' atoms were considered to be main chain atoms. The distances were measured between the Cα atoms. Superimposition was done by the Brugel program (De Maeyer et al. 1997; Delhaise et al. 1988; Desmet et al. 1992).

We used the 'best plane projection' to compare protein conformations generated along the pathways. A 2D representation of the conformational space is obtained by projecting the space onto the best plane through the conformations (Diamond and Levitt 1971; Levitt 1983).

Potential energy calculations

The potential energy calculations were carried out using the Brugel program (De Maeyer et al. 1997; Delhaise et al.

1988; Desmet et al. 1992). The protein potential energy (assumed to be proportional to the enthalpy H) was calculated, along the generated TMD trajectories without including explicit protein–solvent interactions. The protein–protein interaction contained the classical bonded and non-bonded energy terms. The bonded energy includes penalties for bonds, bond-angles, torsions, planarity and deformations of tetragonality. The non-bonded energy contribution includes the van der Waals, the H-bonds and the electrostatic terms. The interaction energy is only computed for pairs of atoms separated by less than the 8 Å cut-off.

Correlation with experimental kinetics

In order to correlate the calculations with the observations, the activation trajectories are considered to be parallel reactions. Since for parallel reactions the rate constants can be summed (Fersht 1999), and applying the transition state theory, we obtain:

$$k_{\text{obs}} = \sum k_{\text{i}} = k_0 \sum \exp(-\Delta H_{\text{i}}^{\#} + T\Delta S_{\text{i}}^{\#})/RT).$$

The value of the observed rate constant, k_{obs}, is mainly determined by the fastest individual reactions, i.e. those with the smallest positive activation free energy change ($\Delta G^{\#}$). Following the transition state theory, the rate constants depend on the activation free energy and not only on the activation enthalpy. We estimate the activation entropy by assuming partial enthalpy-entropy compensation, as is also observed in the experimental data (Verheyden et al. 2004) and in many biological systems. We therefore introduce a factor f that is defined as the fraction of $\Delta H^{\#}$ compensated by entropy. The weight factor for the calculation of $<\Delta H^{\#}>$ from the collection of pathways, can be obtained from the definition of $\Delta H_{\text{obs}}^{\#}$:.

$$\Delta H_{\text{obs}}^{\#} = -R.\partial \ln k_{\text{obs}}/\partial(1/T) = -R.\partial k_{\text{obs}}/k_{\text{obs}}.\partial(1/T)$$

substituting k_{obs} as defined above, and performing the algebra leads to the following definition:

$$<\Delta H^{\#}> = \frac{\sum w_{\text{i}} \Delta H_{\text{i}}}{\sum w_{\text{i}}} \quad \text{with}$$

$$w_{\text{i}} = \exp(-\Delta H^{\#} + T.\Delta S^{\#})/RT)$$
$$= \exp((-\Delta H^{\#}(1-f))/RT).$$

In which we make the simplifying assumption that the enthalpy/entropy compensation factor (f) is the same for all individual pathways.

Nomenclature

In all figures the extended version of α-chymotrypsin's nomenclature (Wroblowski et al. 1997) was used (Fig. 1). Segment II (I16-G19): sg2, segment III (E20-W27): sg3, segment IV (L153-P161): sg4, loop II (S33-H40): lp2, loop III (Q73-S77): lp3, loop IV (K90-D102): lp4, loop V (I162-C182): lp5, loop VI (S186-S195): lp6, loop VII (S214-S221): lp7, loop VIII (Q113-L123): lp8 and the cleavage region between the β and γ chains. Sg2 contains I16 and the mutation sites. Sg4 is the outermost strand in the first β-barrel (Fig. 1). Lp2, lp3, lp4 and lp8 are solvent exposed loops. In lp6 residues M192, G193, D194 and S195 can be found. Lp7 is connected with lp6 via the C191-C220 disulfide bond and via H-bonding with the side chain of S190 being part of the substrate-binding site. The β-barrels run from G25 to N125 and from P132 to Q226. The most important individual residues from the activation viewpoint are I16, M192, G193, D194 and the members of the catalytic triad: residues H57, D102 and S195.

Results and discussion

MD simulations

MD simulations of 200 ps were carried out on the active and inactive conformations of the WT and mutant N18G to generate equilibrated initial and final structures for the TMD simulations. During these MD runs only minor changes such as relaxation of the dihedral angles of the mutational environment, were observed (data are not shown). However, as expected from the experimental time scale (ms) and from previous MD studies (Wroblowski et al. 1997) the inactive conformation did not undergo the major conformational changes required to transform it into the active conformation [data not shown]. Figure 2 shows the 'best plane projection' (Diamond and Levitt 1971; Levitt 1983) of the conformational space explored by the

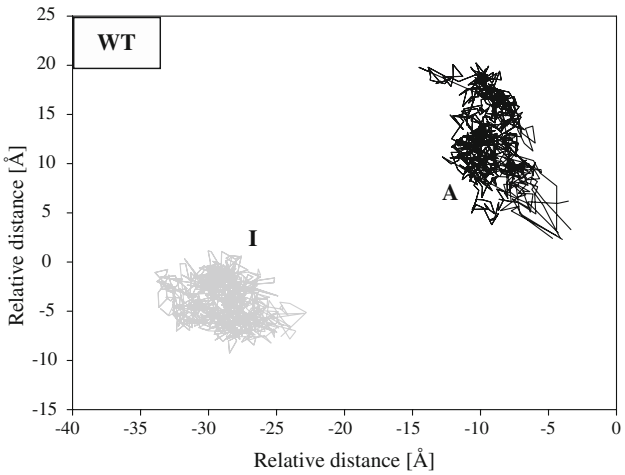

Fig. 2 Best plane projection of the inactive (**I**) and active (**A**) trajectories during the multiple MD simulations (WT)

active (Cluster A) and inactive (Cluster I) structures during the simulation of the WT. Similar results were obtained for the mutant (data not shown). Although the clusters show a slightly different spread in the two conformations, the active and inactive trajectories always visit non-overlapping regions. The absence of activation is further confirmed by analysing the I16-D194 distance, the G193 ϕ/ψ angles and G193's H-bonding pattern along the active and inactive MD trajectories. No salt-bridge formation was observed during the MD simulations of the inactive structures between I16 and D194 nor were H-bonds formed between G193 and the Eglin C inhibitor after placing it from the 1ACB structure into the active site. Among the active structures, examples where no salt-bridge is present and cases where G193 did not take part in H-bonding, could be found. The number of cases with a broken salt-bridge is limited [data are not shown], but for ±40% of the total simulation time the H-bonds were not formed between G193 and the inhibitor. However, even when the salt-bridge or the G193 H-bond was absent, residue I16 remained in the protein interior and residues M192 and G193 stayed close to their active position as deduced from G193's ϕ/ψ angles (data are not shown). These structural fluctuations of the catalytic centre might be part of the observed fluctuations in catalytic activity on single chymotrypsin molecules (Lee and Brody 2005).

Root mean square deviation (rmsd) values were calculated between the WT and the mutant at the beginning and at the end of the MD simulations (Table 1). The small mean and error values of the energy-minimized structures indicate that irrespective of the starting conditions both the active and inactive mutant conformations became equilibrated after 100 steps of energy minimization, and that the structures are not significantly altered by the mutation. After the MD simulations, the rmsd values obtained for the active conformation are smaller and the distribution around the mean is narrower than those obtained for the inactive enzymes. This demonstrates that in the active conformation the WT and its mutant are more similar to each other than in the less homogeneous inactive conformation. For the inactive conformations the largest mean rmsd and error values are found for sg2, indicating its enhanced mobility in the mutant as compared to the rest of the protein.

We also calculated the H-bond energy of sg2 along the MD trajectories. Figure 3 illustrates the dynamics of the WT/mutant active and inactive structures during the MD simulations. Similarly for both types of enzymes and for both the active/inactive conformations sg2 explores multiple minima and maxima of various heights. However, due to the solvent exposed state of sg2 in the inactive conformation the H-bond energy is less negative for the inactive structures in comparison with the active ones (Fig. 3). Due to the mutation in both conformations, but especially in the active conformation, sg2 forms fewer stabilizing H-bonds with the body of the protein in the mutant than in the WT. In the active conformation, the H-bond energy transitions are higher for the WT than in the case of the mutant. This correlates well with our observation that in the WT the formation/breakage of H-bonds is always cooperative, while in the mutant single H-bonds can be formed and broken in several consecutive steps without influencing the H-bonds formed by neighbouring residues.

TMD simulations

The final structures of the MD simulations were used as initial and target structures for the TMD simulations. To determine the optimal simulation time, several TMD simulations were carried out for a duration ranging from 16 ps up to 10 ns. Figure 4 shows the evolution of the calculated enthalpy as a function of the time span used to simulate the conversion from the inactive to the active conformation. For the short runs very high enthalpy barriers were observed which decrease when the time span of the simulation is increased. The height of the energy peaks becomes practically independent of the total simulation time, once this is equal to or longer than 500 ps.

Multiple TMD simulations were carried out for both enzyme variants as explained in the "Materials and methods" section. During all TMD simulations the I→A rearrangement took place. Since, enzymatic activation is correlated with the I16-D194 salt bridge formation, activation is followed by measuring the distance between the N-terminus of I16 and the Cγ atom of D194 (Fig. 5a). It is found that the difference between the trajectories of the mutant and WT is rather small, especially during the first

Table 1 Averaged root mean square deviation (rmsd) values between the main chain atoms of the WT and the mutant after main chain fit, all values are in Å

The compared structures	Energy minimized structures at the beginning of the 200 ps MD simulation complete main chain	Final structure of the 200 ps MD		
		Complete main chain	Main chain of sg2	Main chain without sg2
WT-N18G active	0.84 (±0.08)	1.81 (±0.42)	1.46 (±0.44)	1.82 (±0.41)
WT-N18G inactive	0.95 (±0.09)	1.90 (±0.48)	1.97 (±0.70)	1.91 (±0.46)

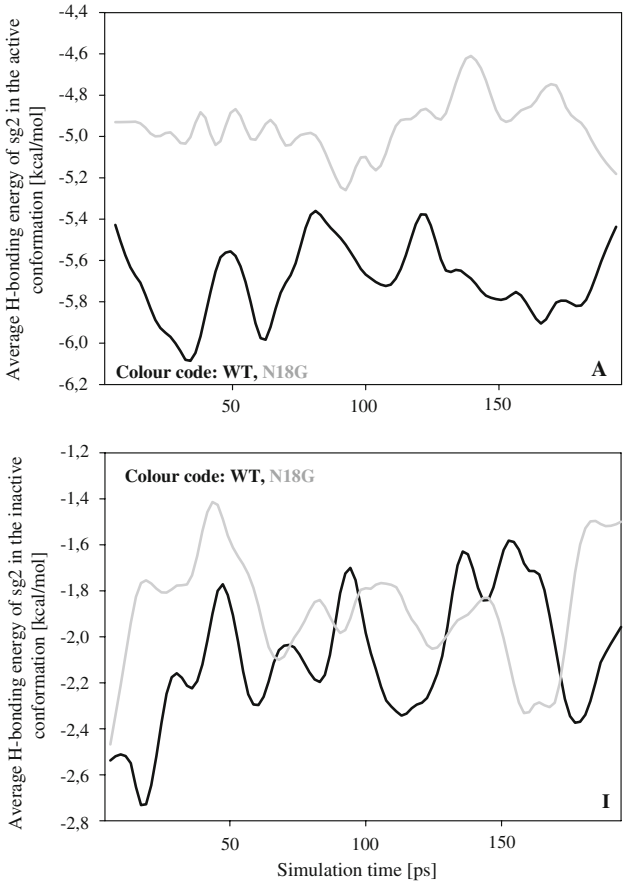

Fig. 3 The average H-bonding energy of the active (**A**) and inactive (**I**) structures during the 200 ps MD simulations

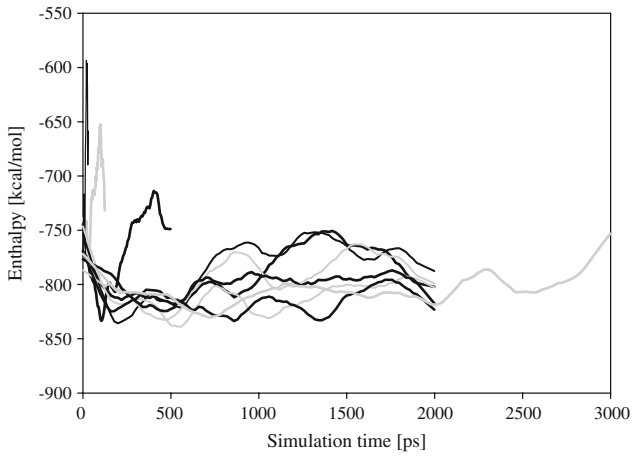

Fig. 4 TMD simulations at different timescales. The 2 ns trajectories correspond to simulations carried out between different starting and target structures and are plotted to illustrate the variety among the single simulations

half of the simulations, but for the mutant the salt bridge distance remains larger at the end of the simulation. Individual trajectories for the I16-D194 distance in WT are illustrated in Fig. 5b. Although the averaged I16-D194 distances behaved rather similarly for both enzymes, individual trajectories may show very different behaviour. Two classes of I16 rotations are observed: a gradual progression or an abrupt drop where residue I16 rotates only at the end of the simulation into the protein. Similarly to the WT, a large variety of trajectories could be found for the mutant as well (data not shown).

Figure 5c, d shows the motion of the D194 Cγ atom and I16 N-terminus during activation. These figures show the average of the distances calculated between the D194 Cγ atom of a structure at a given simulation step and the same atom in the 4CHA X-ray reference structure. The same procedure was applied for the I16 N-terminus. The distances were measured after main chain fitting of each structure of the simulation onto the 4CHA X-ray structure. These figures reveal that the motion of D194 is very similar in both enzymes. Consequently, the difference in the activation between the WT and the mutant, as monitored by the I16-D194 distance change (Fig. 5a), is due to a different behaviour of sg2 (Fig. 5d). In the mutant, this segment explores the solvent for a longer time and approaches the interior more randomly than in the WT. This behaviour is correlated with a higher H-bonding energy for the mutant during the MD and TMD simulations (Figs. 3, 6). Sg2's rotation is characterized by a smaller H-bonding enthalpy gain than in the WT (Fig. 6). On the other hand, glycine increases the flexibility, allowing sg2 to explore a larger conformational space in the mutant than in the WT (Fig. 5a, d). This may give rise to an entropically more favoured activation. The observed differences in the rotational behaviour of sg2 in the mutant and in the WT originate from the N to G mutation in sg2 (Fig. 5d) and explain in a structural way the higher activation enthalpy and entropy measured for the mutant in the FSF experiments (Verheyden et al. 2004).

Multiple pathways

In this study we performed multiple (25) simulations for both the WT and its N18G mutant. These simulations reveal that the activation process can take place along different activation pathways. These pathways are different in terms of the structural and energetic features of the activation process. The most remarkable structural difference among the single pathways is the direction of I16 rotation. During the activation I16 could rotate in two directions, something that was unnoticed by previous studies (Fig. 1). For half of the WT runs the outward rotation is chosen (14 outwards of 25 and Fig. 1), while for the mutant the preferred way is the inward rotation (6 outwards of 25 and Fig. 1). However, there are many more subtle differences as is illustrated by Fig. 5b. The most

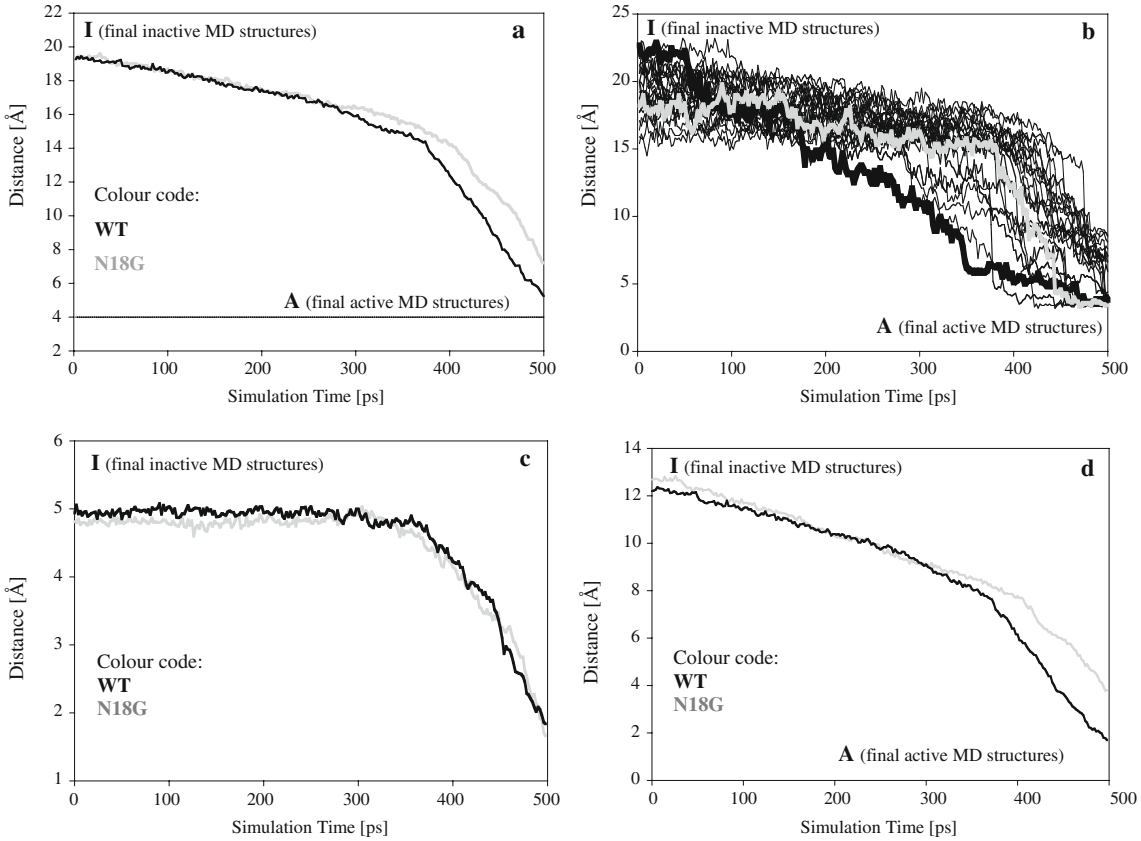

Fig. 5 **a** Average I16 N-terminus—D194 Cγ distance as seen by TMD. The *thin black line* shows the distance in the active X-ray structure (4CHA), **b** the individual I16 N-terminus—D194 Cγ distances as simulated by TMD, WT. Two of the 25 trajectories are indicated separately. They represent two main types of rotations of I16: a gradual (*black bold line*) rotation and an abrupt change (*grey bold line*) where residue I16 locates for a long time far away from D194 and then suddenly rotates into its vicinity in a one-step like manner, **c** D194 Cγ atom motion as simulated by TMD, averaged over the 25 activation simulations. Measured to D194 Cγ in the 4CHA active structure after main chain fitting onto the 4CHA structure for each simulation step, **d** I16 N-terminus motion as simulated by TMD, averaged over the 25 activation simulation. Measured to I16 N-terminus in the 4CHA active structure after main chain fitting the TMD structures of each simulation step onto 4CHA. The colour code is as follows: *black colour* indicates the WT and *grey colour* has been used for mutant N18G in figures **a**, **c** and **d**

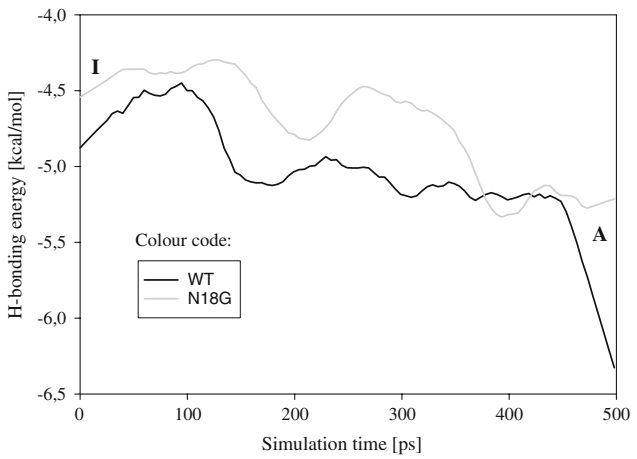

Fig. 6 Averaged H-bonding energy of sg2 with the rest of the protein in the WT and in the mutant during the 25 activation TMD simulations. Letter ***I*** indicates the inactive structures and letter ***A*** the active structures

important events of the activation process can occur at different time instants along the activation reaction coordinate and the duration of an event can also differ as is illustrated for the I16 and D194 movement in Fig. 5b. Furthermore, the order of the elementary activation events also shows variation and we observed that events that were thought to be tightly coupled in time turned out be less coupled in reality. For instance we observed cases when I16 has already rotated into the cleft, but only the segment carrying M192 (sg7, Fig. 1) has shifted. M192 itself was still buried and D194 did not rotate into the cleft during several simulation steps. As another example we could observe that M192 has already left the protein interior at a very early stage of the simulation while I16 rotated into the cleft at the end. These structural differences among the pathways are correlated with energy features that exhibit a similarly large diversity in terms of the number of

activation barriers, their height and position along the activation reaction coordinate.

Comparison with experimental results

TMD is a well-established method developed to study large-scale conformational changes in proteins and DNA complexes as well as the individual behaviour and function of single amino acids (Jacoby et al. 1996; Kamerlin et al. 2007; Kruger et al. 2001; Kuppens et al. 1999; Noy et al. 2007; Perdih et al. 2007; Yu et al. 2007). A recent study (Yu et al. 2007) reveals that despite the TMD constraint the obtained trajectories and protein behaviour are realistic both on the domain and individual residue level. The results described in that study could be correlated, at least on a structural basis, with the outcome of numerous independent methods such as the normal mode analysis, TRP fluorescence, FRET and electron microscopy. A rigorous correlation between the energetic features of the calculated pathways and the available experimental kinetic parameters might be difficult to achieve due to the fact that (a) the FSF experiments were carried out in bulk solution while each simulation was done on a single molecule, (b) 25 simulations still represents a rather small (Fersht 2002; Rao and Caflisch 2004) statistical ensemble and so the calculated activation pathways only explore a tiny subset of the phase space available to the system. Considering, however, that other research groups (Yu et al. 2007), as well as we in this study, found significant correlation between the simulated activation and the experimentally established parameters on a structural basis (see above), we are confident that the simulated trajectories are realistic and that we can correlate the energy parameters of these trajectories with the kinetic FSF parameters.

Experimentally, activation and reaction enthalpy changes are obtained from bulk kinetics as a function of temperature by applying the TS theory. The reaction and activation enthalpy changes are predominantly originating from the potential energy and solvent interactions. The simulations were carried out in explicit water, we calculated the potential energy along the reaction pathway to estimate the protein enthalpy. Since the conformational change occurs in a sequence of many consecutive steps, the highest activation enthalpy peak along the pathway is considered to define the rate-limiting step of the sequence, and its activation barrier is taken as the determinant for the pathway. In this way we estimated the reaction and activation enthalpy ($\Delta H^{\#}$) for each trajectory. The variation of $\Delta H^{\#}$ among the different pathways is quite large. To correlate the calculated $\Delta H^{\#}$ values with the ensemble average obtained by the experiment, the individual $\Delta H^{\#}$ values have to be averaged using a proper weight factor. We propose to consider the different calculated pathways as potentially acceptable parallel pathways. In a process with many parallel pathways the overall rate constant can be obtained from the weighted sum of the individual rate constants (see "Materials and methods") and therefore the pathways having very large $\Delta H^{\#}$ will not be highly populated. Hence we calculate the average $\Delta H^{\#}$ from the individual values, using the weight factor (w_i):

$$w_i = \exp(-\Delta H^{\#} + T.\Delta S^{\#})/RT)$$

obtained from the transition state theory ("Materials and methods").

Since it is very difficult to estimate the activation entropy from the TMD calculations, we assume that enthalpy–entropy compensation occurs, as shown previously (Verheyden et al. 2004), and estimate $T\Delta S^{\#}$ as a fraction f of $\Delta H^{\#}$. In this way we obtain an average $<\Delta H^{\#}>$ as a function of f (Fig. 7). As can readily be seen, the weighted average $<\Delta H^{\#}>$ varies around the experimental values for WT and mutant N18G, 21.3 and 31.1, respectively. At low f values the simulated $<\Delta H^{\#}>$ values are lower than the experimental values due to the large impact of the low individual $\Delta H^{\#}$ barriers. When $0.8 < f < 0.9$ the difference between the WT and the mutant is still present but the simulated and FSF measured $\Delta H^{\#}$ values are closer to each other due to the impact of the higher $\Delta H^{\#}$ but statistically more populated pathways. The fact that the value of f is higher than the experimental value (0.21 for WT and 0.47 for the mutant), indicates that the calculated $\Delta H^{\#}$ values are probably overestimated and that f is also used as a kind of scaling factor.

Conclusions and future outlook

The calculations presented here show that the activation enthalpy and entropy parameters measured by FSF for the

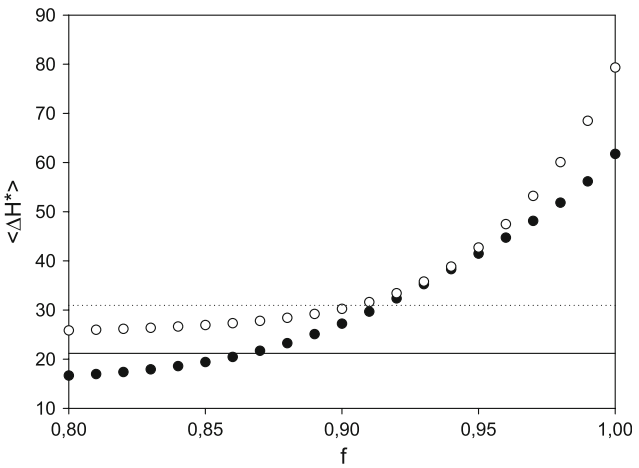

Fig. 7 Weighted average calculated activation enthalpy $<\Delta H^{\#}>$ values are shown for the WT (*filled circle*) and the N18G mutant (*open circle*) as a function of factor f. The *black* and *dotted lines* show the experimental $\Delta H^{\#}$ values for the WT and the mutant, respectively

WT and mutant enzymes could be successfully simulated and explained using both a structural and energetic basis.

Considering the structural features of the simulated pathways we demonstrated that the rather simple conformational change of chymotrypsin activation can be described in great detail as a collection of parallel pathways, each of which is a series of sequential steps at the single molecule level. The large diversity of these pathways demonstrates that the behaviour of the individual residues can be very different along the multiple activation pathways. Indeed, as shown in Fig. 5b residue I16 for instance can enter the interior either in a gradual or in a more abrupt manner. Analyzing the individual pathways (data not shown) we found that the exact, currently known chronology of the activation in WT (Matrai et al. 2004; Wroblowski et al. 1997) is not always present. Events such as the rotation of I16, the M192 flip, the repositioning of G193 and finally the orientation of D194 towards the binding cleft, are not always nicely consecutive, implying that the activation might occur in many different ways.

We found prior to the application of the TMD method through the use of MD simulations that the onset of the alterations of the structural and energetic features of the activation as compared to WT can already be seen at the MD level (Table 1; Figs. 3, 5a, c, d). However, activation did not take place in any of the simulations as followed by the behaviour of residues I16, M192, G193 and D194.

The H-bonding network around the mutation site has a significant impact on the activation. The H-bonds formed by sg2 guide and facilitate the rotation of this segment. The mutation introduced in sg2 influences the guiding force around the segment and consequently its rotational behaviour via changes in this H-bond system (Figs. 3, 5d, 6). The analysis of the decreasing distance between I16 and D194 and the motion of D194 Cγ atoms and the I16 N-terminus shows that the differences between the activation of WT and mutant N18G are due to the mutation introduced in sg2 and is correlated with structural parameters (Fig. 5a, b, d).

The altered motion of sg2 in the two species is also reflected by the differences in its rotation into the protein interior. To the best of our knowledge our study shows for the first time that in WT for sg2 two equally populated rotational directions are present, while in the mutant sg2 preferably takes the 'classical' less solvent exposed direction, already described in previous publications (Matrai et al. 2004; Wroblowski et al. 1997).

All the structural features associated with sg2's rotation, i.e. the motion of the I16 N-terminus, the formed H-bonds and the rotational pathway, illustrate that in the mutant the stabilizing enthalpic interactions are smaller than in the WT. The motion of I16 N-terminus demonstrates that for the mutant form a larger conformational space is explored and that the N-terminus rotates later and in a more random fashion into the protein interior than in the WT (Fig. 5a, d), indicating a larger $\Delta S^{\#}$ value for the mutant than for the WT. Considering that for the mutant a larger activation enthalpy and entropy was obtained by FSF (Verheyden et al. 2004), these results correlate well the simulations and the FSF experiment on a structural basis.

Besides the structural and H-bond analysis of the activation trajectories the activation enthalpy was estimated as well. We found that the multiple pathways vary in terms of activation enthalpy, from close to the experimental value to much larger. Considering these as parallel pathways and using a semi-empirical method that takes into account the effect of entropy compensation we could correlate the weighted average $\Delta H^{\#}$ from the multiple pathway calculations with the observed $\Delta H^{\#}$ obtained from solution kinetics (Verheyden et al. 2004). This result clearly underlines the importance of executing multiple random simulations.

Our results indicate that TMD is an appropriate tool to simulate conformational changes and to correlate multiple simulations of single pathways with ensemble experiments on both structural and energetic basis, provided (1) the simulation time is sufficiently long, (2) enough runs with different initial conditions are executed and (3) the experimentally examined variables differs sufficiently for the different enzymes. The results in this study might be further improved by performing more simulations, on a longer time scale (1 or 2 ns) and by incorporating configurational and solvation entropy terms into the energy calculations.

The variability that we calculate for the different pathways is consistent with the experimental observations made on single enzymes. It is to be expected that the individual motions explored by the different loops in the active site will contribute to fluctuations in enzymatic activities, as has been observed in single molecule work on chymotrypsin (Lee and Brody 2005).

Acknowledgments Janka Mátrai was supported by a scholarship from the Katholieke Universiteit Leuven (OE/00/20 and OE/01/07) and from the Fund for Scientific Research Flanders. Gert Verheyden was a research fellow of the same Fund. Janka Mátrai thanks MIT-ACS for scientific and financial support during her stay in Canada. Janka Mátrai thanks to George Matrai for his valuable contribution to the correction of this manuscript. The authors thank Prof. W. J. Rutter for providing the pTRAP vector with the rat-chymotrypsin gene. Further support was received in the pursuance of the contract P5/33: "Protein structure and function in the post-genomic era" and the program P6/19: "Interaction involved in folding, function and supramolecular assemblages" in the framework of the Belgian federal program "Interuniversity Attraction Poles -Phase V." Financial support from the impuls financing KULeuven BioMacS is greatly appreciated.

References

Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) Interaction models for water in relation to protein hydration. In: Pullman B (ed) Intermolecular forces. D. Reidel Publishing Company, Dordrecht, pp 331–342

Bernstein HJ (2000) Recent changes to RasMol, recombining the variants. Trends Biochem Sci 25:453–455

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) The protein data bank: a computer-based archival file for macromolecular structures. J Mol Biol 112:535–542

De Maeyer M, Desmet J, Lasters I (1997) All in one: a highly detailed rotamer library improves both accuracy and speed in the modelling of sidechains by dead-end elimination. Fold Des 2:53–66

DeLano (2002) The PyMOL Molecular Graphics System

Delhaise P, Bardiaux M, De Maeyer M, Prevost M, Vanbelle D, Donneux J, Lasters I, Vancustem E, Alard P, Wodak SJ (1988) The Brugel Package—toward computer-aided-design of macromolecules. J Mol Graph 6:219

Desmet J, De Maeyer M, Hazes B, Lasters I (1992) The Dead End Elimination theorem and its use in protein side chain positioning. Nature 356:539–542

Diamond R, Levitt M (1971) A refinement of the structure of lysozyme. Biochem J 125(4):92P

Fersht A (1999) The basic equations of enzyme kinetics. W.H. Freeman and Company, New York

Fersht AR (2002) On the simulation of protein folding by short time scale molecular dynamics and distributed computing. Proc Natl Acad Sci USA 99:14122–14125

Fersht AR, Requena Y (1971) Equilibrium and rate constants for the interconversion of two conformations of α-chymotrypsin. The existence of a catalytically inactive conformation at neutral pH. J Mol Biol 60:279–290

Fischer S, Windshugel B, Horak D, Holmes KC, Smith JC (2005) Structural mechanism of the recovery stroke in the myosin molecular motor. Proc Natl Acad Sci USA 102:6873–6878

Gao M, Lu H, Schulten K (2002) Unfolding of titin domains studied by molecular dynamics simulations. J Muscle Res Cell Motil 23:513–521

Grubmuller H (1995) Predicting slow structural transitions in macromolecular systems: conformational flooding. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics 52:2893–2906

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723

Heremans L, Heremans K (1989) Raman spectroscopic study of the changes in secondary structure of chymotrypsin: effect of pH and pressure on the salt bridge. Biochim Biophys Acta 999:192–197

Jacoby E, Kruger P, Schlitter J, Roper D, Wollmer A (1996) Simulation of a complex protein structural change: the T <–> R transition in the insulin hexamer. Protein Eng 9:113–125

Kamerlin SC, Rucker R, Boresch S (2007) A molecular dynamics study of WPD-loop flexibility in PTP1B. Biochem Biophys Res Commun 356:1011–1016

Kardos J, Bodi A, Zavodszky P, Venekei I, Graf L (1999) Disulfide-linked propeptides stabilize the structure of zymogen and mature pancreatic serine proteases. Biochemistry 38:12248–12257

Kitao A, Hayward S, Go N (1998) Energy landscape of a native protein: jumping-among-minima model. Proteins 33:496–517

Kraulis PJ (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 24

Kruger P, Szameit A (1992) Simlys Version 2.0 Com Phys Commun 72:265–268

Kruger P, Verheyden S, Declerck PJ, Engelborghs Y (2001) Extending the capabilities of targeted molecular dynamics: simulation of a large conformational transition in plasminogen activator inhibitor 1. Protein Sci 10:798–808

Kuppens S, Diaz JF, Engelborghs Y (1999) Characterization of the hinges of the effector loop in the reaction pathway of the activation of ras-proteins. Kinetics of binding of beryllium trifluoride to V29G and I36G mutants of Ha-ras-p21. Protein Sci 8:1860–1866

Lee AI, Brody JP (2005) Single-molecule enzymology of chymotrypsin using water-in-oil emulsion. Biophys J 88:4303–4311

Levitt M (1983) Molecular dynamics of native protein. II. Analysis and nature of motion. J Mol Biol 168:621–657

Matrai J, Verheyden G, Kruger P, Engelborghs Y (2004) Simulation of the activation of alpha-chymotrypsin: analysis of the pathway and role of the propeptide. Protein Sci 13:3139–3150

Noé F, Krachtus D, Smith JC, Fischer S (2006) Transition networks for the comprehensive characterization of complex conformational change in proteins. J Chem Theory Comput 2:840–857

Noy A, Perez A, Laughton CA, Orozco M (2007) Theoretical study of large conformational transitions in DNA: the B <–> A conformational change in water and ethanol/water. Nucleic Acids Res 35:3330–3338

Oroguchi T, Ikeguchi M, Ota M, Kuwajima K, Kidera A (2007) Unfolding pathways of goat alpha-lactalbumin as revealed in multiple alignment of molecular dynamics trajectories. J Mol Biol

Perdih A, Kotnik M, Hodoscek M, Solmajer T (2007) Targeted molecular dynamics simulation studies of binding and conformational changes in *E. coli* MurD. Proteins 68:243–254

Rao F, Caflisch A (2004) The protein folding network. J Mol Biol 342:299–306

Schlitter J, Engels M, Kruger P, Jacoby E, Wollmer A (1993) Targeted molecular-dynamics simulation of conformational change—application to the T[–]R transition in insulin. Mol Simul 10:291–308

SPSS Inc (2001) SigmaPlot version 8.0, Chicago IL

Stoesz JD, Lumry RW (1978) Refolding transition of alpha-chymotrypsin: pH and salt dependence. Biochemistry 17:3693–3699

Swegat W, Krueger P, Schlitter J (1997) TMD-implementation in GROMOS96. KUL, Belgium, pp 1–11

Tereschenko A (1999) Futuris Imager 2.6

Tsukada H, Blow DM (1985) Structure of alpha-chymotrypsin refined at 1.68 A resolution. J Mol Biol 184:703–711

Van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG (1996) Biomolecular simulation: the GROMOS96 manual and user guide. vdf Hochschulverlag AD an der ETH Zürich, Zürich

Verheyden G, Volckaert G, Engelborghs Y (2000) Expression of chymotrypsin(ogen) in the thioredoxin reductase deficient mutant strain of *Escherichia coli* AD494(DE3) and purification via a fusion product with a hexahistidine-tail. J Chromatogr B Biomed Sci Appl 737:213–224

Verheyden G, Matrai J, Volckaert G, Engelborghs Y (2004) A fluorescence stopped-flow kinetic study of the conformational activation of alpha-chymotrypsin and several mutants. Protein Sci 13:2533–2540

Wang D, Bode W, Huber R (1985) Bovine chymotrypsinogen A X-ray crystal structure analysis and refinement of a new crystal form at 1.8 A resolution. J Mol Biol 185:595–624

Westgrid (2004) What is Grid Computing?

Wroblowski B, Diaz JF, Schlitter J, Engelborghs Y (1997) Modelling pathways of alpha-chymotrypsin activation and deactivation. Protein Eng 10:1163–1174

Yu H, Ma L, Yang Y, Cui Q (2007) Mechanochemical coupling in the myosin motor domain. II. Analysis of critical residues. PLoS Comput Biol 3:e23