Assignment of side-chain conformation using adiabatic energy mapping, free energy perturbation, and molecular dynamic simulations

THOMAS M. FRIMURER,¹ GÜNTHER H. PETERS,² MORTEN DAHL SØRENSEN,^{3,4}
JENS JØRGEN LED,³ AND OLE HVILSTED OLSEN¹

¹MedChem Research IV, Novo Nordisk Park, Novo Nordisk A/S, DK-2760 Måløv, Denmark

²Chemistry Department III, H.C. Ørsted Institute, University of Copenhagen, DK-2100 Copenhagen Ø, Denmark

³Chemistry Department V, H.C. Ørsted Institute, University of Copenhagen, DK-2100 Copenhagen Ø, Denmark

(RECEIVED June 5, 1998; ACCEPTED September 15, 1998)

Abstract

NMR spectroscopic analysis of the C-terminal Kunitz domain fragment ($\alpha 3(\text{VI})$) from the human $\alpha 3$ -chain of type VI collagen has revealed that the side chain of Trp21 exists in two unequally populated conformations. The major conformation (M) is identical to the conformation observed in the X-ray crystallographic structure, while the minor conformation (m) cannot structurally be resolved in detail by NMR due to insufficient NOE data. In the present study, we have applied: (1) rigid and adiabatic mapping, (2) free energy simulations, and (3) molecular dynamic simulations to elucidate the structure of the m conformer and to provide a possible pathway of the Trp21 side chain between the two conformers. Adiabatic energy mapping of conformations of the Trp21 side chain obtained by energy minimization identified two energy minima: One corresponding to the conformation of Trp21 observed in the X-ray crystallographic structure and solution structure of $\alpha 3(\text{VI})$ (the M conformation) and the second corresponding to the m conformation predicted by NMR spectroscopy. A transition pathway between the M and m conformation is suggested. The free-energy difference between the two conformers obtained by the thermodynamic integration method is calculated to 1.77 ± 0.7 kcal/mol in favor of the M form, which is in good agreement with NMR results. Structural and dynamic properties of the major and minor conformers of the $\alpha 3(\text{VI})$ molecule were investigated by molecular dynamic. Essential dynamics analysis of the two resulting 800 ps trajectories reveals that when going from the M to the m conformation only small, localized changes in the protein structure are induced. However, notable differences are observed in the mobility of the binding loop (residues Thr13–Ile18), which is more flexible in the m conformation than in the M conformation. This suggests that the reorientation of Trp21 might influence the inhibitory activity against trypsin, despite the relative large distance between the binding loop and Trp21.

Keywords: adiabatic mapping; conformational change; essential dynamics; free energy simulations; Kunitz type inhibitor $\alpha 3(\text{VI})$

Protein activity is often accompanied by conformational changes, which can involve small structural changes of individual side chains or relatively large displacements of secondary structural elements such as loops or helices (Derewenda et al., 1994). These structural movements are often important in the activation of the enzyme providing access of the substrate to the interior active site (Case & Karplus, 1979; Northrup et al., 1982; Derewenda et al., 1994). These motions are generally studied using spectroscopic techniques, such as fluorescence or NMR spectroscopy. The latter

technique has been extensively used to study the solution structure of the Kunitz domain.

The C-terminal Kunitz domain fragment from the human $\alpha 3$ -chain of type VI collagen shows 40–50% sequence identity to a variety of serine protease inhibitors of the Kunitz-type (Chu et al., 1990) and has 43% sequence identity to the bovine pancreatic trypsin inhibitor (BPTI) and to the precursor protein of Alzheimer β -amyloid protein (APPI) (Kitaguchi et al., 1988). Structural information about the C-terminal Kunitz domain from human type VI collagen ($\alpha 3(\text{VI})$) has been obtained by X-ray crystallography and NMR spectroscopy (Arnoux et al., 1995; Zweckstetter et al., 1996; Sørensen et al., 1996, 1997). In general, the structures determined by the various techniques are very similar and resemble the well-known BPTI structure (Berndt et al., 1992). Recently, the NMR solution structure of the 76-residue complete C5 domain

Reprint requests to: Ole Hvilsted Olsen, MedChem Research IV, Novo Nordisk Park, Novo Nordisk A/S, DK-2760 Måløv, Denmark; e-mail: oho@novo.dk.

⁴Present address: Department of Chemical Research, Leo Pharmaceutical Products, DK-2750 Ballerup, Denmark.

(Zweckstetter et al., 1996), the X-ray crystallographic structure (Arnoux et al., 1995) and the NMR solution structure (Sørensen et al., 1997) of the 7-64 fragment were published. Zweckstetter et al. (1996) found that the complete C5 domain exists in exchanging multiple conformations, and that there are three main patches of residues that have multiple conformations; one in the central β -sheet, one in the loop around the Cys20⁽¹⁴⁾–Cys44⁽³⁸⁾ disulfide bond, and one in the vicinity of the Cys11⁽⁵⁾–Cys61⁽⁵⁵⁾ disulfide bond (notation according to Sørensen et al. (1996)). These multiple conformations have been further investigated by NMR (Sørensen et al., 1996, 1997), and it was observed that the origin of the multiple conformations of the central β -sheet is a reorientation of the Trp21 side chain. The two unequally populated distinct side-chain conformations of Trp21 are referred to as the major (*M*) and the minor (*m*) conformations, respectively. This isomerization between the *M* and the *m* conformations of the Trp21 side chain is slowly exchanging on the NMR time scale. A similar process has also been suggested to occur in APPI, which is another Kunitz type protein (Heald et al., 1991). Based on the NMR intensities of the corresponding signals in the two conformations, it has been determined that the population of the *m* conformation is 6.4% of the *M* conformation corresponding to $\Delta G_{(m \rightarrow M)} = 1.65$ kcal/mol. Due to the lack of sufficient nuclear Overhauser effects (NOEs) corresponding to the minor conformation, detailed structural data are lacking for this conformation. Similarly, fluorescence data of the reorientation of the Trp21 side chain are not sufficiently resolved due to the presence of an another Trp side chain, Trp34. Both residues are located in the antiparallel β -string and are only 8.5 Å apart.

Precise knowledge of the structure of both *M* and *m* conformations of the $\alpha 3$ (VI) molecule, and the basis for the conformational change are indispensable for the understanding of the inhibitory action of this protein. We therefore have used various computational methods to elucidate in more detail the conformational changes occurring during transition. Our investigation is based on (1) rigid and adiabatic mapping, (2) thermodynamic integration (TI) method, and (3) molecular dynamics (MD). The unique combination of these methods allow us not only to identify the structure of the *m* conformation, but also to calculate the relative populations and free energies of the Trp21 side chain in the two conformers and even predict a transition pathway between the two conformers.

Rigid mapping is the simplest approach to identify favorable conformations of amino acid side chains and to calculate their relative energies. In this methodology, the side chain is rotated at equally spaced angular intervals, while holding the surrounding atoms fixed. Favorable conformations are identified on the basis of steric and energetic criteria. Relaxation of the system can be considered in adiabatic maps, where the system additionally is subjected to reasonably extensive energy minimization (EM) (Shih et al., 1985; Wade & McCammon, 1992; Joseph-McCarthy et al., 1995). The shortcoming of this approach is that population densities cannot be deduced from the mapping procedure and the results are restricted to defining the transition pathway and low energy side-chain conformations. Population densities of the two conformations are estimated from relative free energy difference between rotational isomers of the Trp21 side chain. Relative free energies (ΔG) are usually calculated by applying free energy perturbation (FEP) or TI methods. A review of the different methods and a discussion about their accuracy and limitations can be found in several references (Fleischman & Zichi, 1991; Mazor & Pettitt, 1991; Mitchell & McCammon, 1991; Pearlman & Kollman, 1991;

Straatsma & McCammon, 1991; Hodel et al., 1993; Pearlman, 1993). Basically, the main concern focuses on the influence of statistical errors due to lack of adequate sampling of the conformational space. Determination of a reliable estimate of the free energy for a given change can often be difficult even though the free energy equation is relatively simple. In particular, this will be the case for a large conformational change, where the average value of ΔG may converge very slowly. Both rigid mapping and free energy calculations in the present set-up do not provide information about protein dynamics and how internal motions are affected by the *M* to *m* transition. To elucidate this question, we have performed MD simulations to analyze the motions of the $\alpha 3$ (VI) molecule. In particular, we have studied the mobility of the binding loop in the two conformers, which may be of importance for biological activity. Experimental data show that typical motions of peptide loops occur on the time scale of 10^{-9} to 10^{-1} s (Wade et al., 1993, 1994; Philippopoulos et al., 1995) and that the reorientation of the Trp21 side chain are slowly exchanging on the millisecond to second time scale (Sørensen et al., 1996), which at the present time are not accessible with standard MD simulations. Therefore, we have analyzed the concerted motions in $\alpha 3$ (VI) (using the essential dynamics (ED) analysis method) with the Trp21 side chain in the major and minor conformations, respectively. The ED analysis allows to extract the essential motions in the protein (filtering out Gaussian motions (Amadei et al., 1993)) and may provide new insight in the dynamics of the protein. Furthermore, motions in the major conformer may be correlated to the onset of the transition from the *M* to the *m* conformation.

Results and discussion

We have investigated the structural and energetic properties of $\alpha 3$ (VI) with the Trp21 side chain in the *M* and *m* conformation, respectively. The study focuses on elucidating the transition pathway and on correlating the orientation of the Trp21 side chain to the flexibility of the protein and in particular to the flexibility of the binding loop.

Pathway from the *M* to *m* conformation

The energetically favorable conformations of the Trp21 side chain has been investigated by adiabatic mapping. The resulting adiabatic iso energy contour map is shown in Figure 1A. This map clearly identifies two energy minima centered at approximately ($\chi 1 = -65$, $\chi 2 = -100$) corresponding to the position found in the crystal and solution structures and ($\chi 1 = -80$, $\chi 2 = 80$). The identified low energy side-chain conformations (referred to as major ($\chi 1 = -65$, $\chi 2 = -100$) and minor ($\chi 1 = -80$, $\chi 2 = 80$) conformations) are preferred over any other combinations of $\chi 1$ and $\chi 2$ due to unfavorable interactions between the Trp21 side chain and the surrounding. To examine structural features during the transition from *M* to *m*, root-mean-square (RMS) fluctuations were calculated for backbone and side-chain atoms with respect to the crystal structure. The backbone RMS data plotted in Figure 1B as a function of $\chi 1$ and $\chi 2$ indicate that only minor changes in backbone conformation are present in the major and minor conformations of the Trp21 side chain. The RMS data obtained for only side-chain atoms are more disordered indicating that the transition from *M* to *m* causes structural changes in the vicinity of Trp21 (data not shown). In agreement with the results obtained in the adiabatic map, the major and minor conformations are identi-

Fig. 1. **A:** Adiabatic iso energy contour map for the rotation of the Trp21 side chain in the $\alpha 3(\text{VI})$ molecule. The contour levels are separated by 20 kcal/mol. Energies represent the sum of Van der Waals and electrostatic interaction energies between the Trp21 side chain and the rest of the protein (given with respect to the energy minimized crystal structure). **B:** RMS data of the backbone atoms as function of $\chi 1$ and $\chi 2$ are calculated with respect to the crystal structure. The contour levels are separated by 0.03 Å.

fied as local minima located at the same $\chi 1$ and $\chi 2$ dihedral angles. The corresponding major and minor conformations of the Trp21 side chain are shown in Figure 2. The transition from *M* to *m*, schematically illustrated in Figure 2B, follows the pathway indicated in the adiabatic map (Fig. 1A). Visual inspection of the mapping procedure revealed that the most mobile residues were Arg32, Gly46, and Ser47 all located around the Trp21 side chain. The total interaction energy between the Trp21 side chain and all other atoms calculated along the transition pathway is displayed in Figure 3 as a function of the $\chi 2$ dihedral angle. The barrier height between the two conformations is 25 kcal/mol. It is interesting to note that the relative energy difference between the *M* and *m* conformations is calculated to 1.6 kcal/mol in favor to the *M* conformation. This observation is in good agreement with the result of a recent NMR study (Sørensen et al., 1996). In this study, the experimentally determined population of the minor conformation was found to be $6.4 \pm 0.2\%$, which corresponds to a relative free energy differences of 1.65 kcal/mol.

Population densities of the conformers

The two identified energy minimum conformations of the Trp21 side chain were used as initial structures in the free energy calculations to estimate the population densities. The perturbation of *M* to *m* was carried out with a dual topology representation, which involves all 11 Trp21 side-chain atoms (see Fig. 4). Figure 5 shows the variation of the free energy difference as a function of simulation time for each λ simulation when going from the *M* to the *m* conformation. To check the convergence of each λ simulations, the mean force $\langle \dots \rangle$ in Equation 2 was calculated over successive 10 ps and compared. Convergence for each λ -simulation was obtained after 100 ps consisting of 30 ps equilibration and 70 ps

production run. The total length of the free energy simulation covering the path from $\lambda = 0$ to $\lambda = 1$ was equal to 1,100 ps. The error estimation was calculated using the so-called block method (Young, 1962) implemented in the CHARMM software package (1992). The relative free energy difference between the *M* and *m* conformations was calculated to be 1.77 ± 0.70 kcal/mol, which is in excellent agreement with the relative energy difference obtained from the adiabatic map and the NMR study. Extending the equilibration period for each λ value did not change (within the statistical error) the calculated free energy. As shown in Figure 5, the total free energy difference converges to a stationary value of 1.77 ± 0.70 kcal/mol, whereas the two components, electrostatic (Elec) and van der Waals (vdW) interaction energies, diverge. This is due to the fact that the two components are correlated. This is consistent with previous observations that the components of the free energy can converge very slowly (Shi et al., 1993; Pearlman, 1994). Comparison of the free energy difference 1.77 kcal/mol with the (enthalpy) energy difference 1.6 kcal/mol obtained from the adiabatic map indicates that there is a little entropy effect in the transition from *M* to *m*. The rationale for decomposition of the total free energy into separate contributions has been questioned because the total free energy is a state function, while the individual components of the total free energy are not (Boresch et al., 1994; Mark & van Gunsteren, 1994; Boresch & Karplus, 1995).

Chemical shift calculations

To further support our findings, we have estimated the chemical shift values for protons using the procedure by Asakura et al. (1993). The structures used in these calculations were taken from the MD simulations described below and an averaged structure was calculated from the different trajectories. Experimental data

Fig. 2. A: Ribbon trace of the crystal structure of $\alpha 3(\text{VI})$ with the Trp21 side chain in the major and minor conformations. Disulfide bridges are colored in yellow. The protein consists of two helices (residue Thr2–Lys7 and Ser47–Ala56), an antiparallel β -sheet (Ile18–Asp24, Ser29–Tyr35), three loop regions centered at the residues Cys14, Asn26, and Cys38 and three disulfide bridges (5–55, 14–38, and 30–51). **B:** Trp21 conformations selected along the transition pathway, going from the major (numbered 1; upper left) to the minor (numbered 8; lower right) conformations.

Fig. 3. The energy along the pathway from *M* to *m* as a function of the χ^2 dihedral angle. The two conformations differ by less than 1.6 kcal/mol in favor to the *M* conformation. The barrier separating *M* and *m* is ~ 25 kcal/mol. The numbers below the curve refer to the snapshots in Figure 2B, which are equally distributed along the transition pathway.

show that several protons including Gln48(H β 2), Ser47(H α), and Arg32(H γ 1, H γ 2), which are all located in the vicinity of the Trp21 ring (within 3.5 Å) have very different chemical shifts. Inspection of the α 3(VI) crystal structure reveals that the low value of the chemical shift (0.48 ppm) experimentally determined for Gln48(H β 2) (Sørensen et al., 1996) is due to the position of this proton directly above the Trp21 ring, making it highly sensitive to the ring current effect. In the present study, the chemical shift values for Gln48(H β 2) were calculated to -0.33 ppm (major conformation) and 1.63 ppm (minor conformation) corresponding to a chemical shift difference of 1.96 ppm. This should be compared to the experimentally obtained chemical shift difference $\Delta\delta = 1.53$ ppm (Sørensen et al., 1996). A similar tendency was observed for other chemical shifts observed in the vicinity of Trp21.

Fig. 4. The perturbation of the Trp21 (*M*) to Trp21 (*m*) in the dual topology representation. Dihedral, van der Waals, and electrostatic interaction energies are considered in the hybrid Hamiltonian (*H*) of the growing and shrinking procedure. Structural parts involved in the perturbation are indicated by dashed lines.

Protein dynamics

Protein motions occurring in the *m* and *M* conformers were deduced from MD simulations, which were carried out for 800 ps. The stability of trajectories as checked by computing several geometric quantities; i.e., radius of gyration (R_G), RMS displacements, surface accessibility, and the number of internal hydrogen bonds. For both systems, these quantities were stable throughout the simulations (data not shown). In both systems, the radius of gyration and the average number of intermolecular hydrogen bonds were $1.03 \text{ nM} \pm 0.05 \text{ nM}$ and 36, respectively. The number of hydrogen bonds is quantitatively similar to the 36 intermolecular hydrogen bonds found by Arnoux et al. (1995). The RMS for the trajectory remained close to the starting structure of the α 3(VI) molecule in the two conformations. The RMS of the C α atoms are 1.2 Å for the major and the minor simulation. The total average surface accessibility decreases slightly ($3,982 \rightarrow 3,824 \text{ Å}^2$), which is probably a combination of the decrease in flexibility and the decrease in the average surface accessibility of the Trp21 side chain. This residue is nearly always found to be buried in the hydrophobic interior of proteins as it is predominantly nonpolar in nature (Goldsack, 1976). However, the polar nitrogen in the tryptophan indole ring system allows hydrogen bonding interactions to be made with other residues or with solvent molecules. A surface accessibility calculation along the two MD trajectories (for the *M* and *m* simulations) shows that the accessibility of the Trp21 side chain decreases from $53 \rightarrow 40 \text{ Å}^2$. This reduction in accessibility is restricted to the polar atom in the indole ring system. In the major conformation, the N-H of the indole group is more exposed to the water, probably explaining the preference for this conformation.

The resulting trajectories were further analyzed to compute the total energy difference between the conformers and to extract the concerted motions occurring in the major and minor conformations. The interaction energy averaged over the molecular dynamics trajectory between the Trp21 side-chain atoms and all other

Fig. 5. Accumulated $\Delta G(M \rightarrow m)$ for all of the free energy simulations (i.e., $\lambda = 0 \rightarrow 1$ in steps of $\Delta\lambda = 0.1$) versus simulation time. An equilibration period of 30 ps was considered for each λ value, followed by a production run of 70 ps, which was used to estimate the error bars. Energies are: Vdw = Van der Waals, Elec = electrostatic, and Int = internal contributions.

atoms was calculated for the two side-chain conformations. The energy difference between the *M* and *m* conformers was equal to 1.06 kcal/mol in favor of the *M* conformation. This value should be compared to the result of the adiabatic mapping of 1.60 kcal/mol. Different values are expected, since MD simulations sample the configurational space more thoroughly. The difference in protein dynamics was studied by performing a combined analysis of the two concatenated trajectories (van Aalten et al., 1995). Differences in the essential subspace can be studied by projecting the separate trajectories onto the “combined” eigenvectors. The average of the projections and the mean-square fluctuations in these projections as a function of eigenvector indices are shown in Figure 6. The average projection shows differences in equilibrium structure, whereas the mean-square fluctuations can be used to study differences in dynamics. Small differences are observed at several eigenvectors (see Fig. 6). The absolute value of the four most important eigenvectors as a function of residue numbers is shown in Figure 7.

The structural changes described by eigenvector 1 (see Figs. 6A, 7A) can be attributed to a small backbone conformational change of the (Asp24–Thr27) and (Phe45–Gln48) regions (see Fig. 2B). Visualization of the $C\alpha$ motions indicates that fluctuations in the (Phe45–Gln48) loop are correlated to the motions in the (Asp24–Thr27) segment. In the minor conformation the indole group of the Trp21 side chain interacts with the region (Phe45–Gly46–Ser47–Gln48), which forms a part of the cavity spanned by Trp21. These interactions result in a slight conformational change in this region. Hydrogen bonds between the side chain of Asn43 and Tyr23 and the backbone atoms of Phe45 and Trp21 as well as the Cys30–Cys51 disulfide bond imposes tight interactions between the (Asp24–Thr27) and (Phe45–Gln48) regions. Motions occurring in the (Phe45–Gln48) region are therefore transmitted to the (Asp24–Thr27) segment and vice versa.

For larger eigenvector indices (>1) differences are mainly observed in fluctuations. Average projection and the mean-square fluctuations are nearly identical for the second eigenvector (see

Fig. 6) indicating that the overall structure and fluctuations along this eigenvector are similar for the two conformers. Relatively large fluctuations are seen along eigenvector 2 in the binding loop (Thr13–Ile18) and the glycine rich region (Gly36–Gly37–Cys38–Gly39–Gly40) (see Fig. 7B). Eigenvectors 3 and 5 are interesting, since the mean-square fluctuations indicate that motions along these eigenvectors are predominantly found either in the major or in the minor conformers. Fluctuations in the major conformation are predominantly seen along eigenvector 3. A less flexible binding region is observed along this eigenvector (Fig. 7C), whereas the flexible glycine rich region Gly36–Gly37–Cys38–Gly39–Gly40 and the correlated motions between Phe45–Gly46–Ser47–Gln48 and Asp24–Pro25–Asn26–Thr27 are still pronounced as indicated in Figure 7A.

Eigenvector 5 describes fluctuations occurring predominantly in the minor conformation. Comparison of the absolute value of eigenvectors 3 and 5 (Fig. 7C,D) indicates that fluctuations in the binding loop (Thr13–Ile18) are enhanced along eigenvector 5 (i.e., more pronounced in the minor conformation). Furthermore, the N terminal region is more flexible in the *M* than in the *m* conformer. It is interesting to note that the binding region of $\alpha 3(VI)$ has a higher degree of flexibility in the minor conformation than in the major conformation, which may have an effect on the inhibition of trypsin. Mobility of the binding loop has previously been suggested to play an important role during trypsin binding (Sørensen et al., 1997). The difference in mobility of the binding loop in the major and minor conformations are somewhat surprising, as Trp21 is relatively distant to the binding loop. However, this also shows that concerted motions in a protein can be transmitted over a relatively large distance.

Sørensen et al. (1996) suggested that the disulfide bridge formed between Cys30 and Cys51 could exist in a major and minor conformation identical to the isomerization of the disulfide bridge of Cys14–Cys38 observed in the Kunitz type BPTI molecule (Otting et al., 1993). In the present study, we do not observe any correlation between the Trp21 reorientation and the conformation of the

Fig. 6. A: Average values of the projections of the extracted vectors obtained from the $C\alpha$ coordinate covariance matrix of the concatenated trajectories of $\alpha 3(VI)$ in the major and minor conformations as function of the eigenvector index. **B:** Mean-square fluctuations of the values of the projections of the extracted vectors obtained from the $C\alpha$ coordinate covariance matrix of the concatenated trajectories of $\alpha 3(VI)$ in the major and minor conformations as function of the eigenvector index. Eigenvectors are ordered by the size of their corresponding eigenvalues; i.e., the first eigenvector is the eigenvector with the largest eigenvalue.

Fig. 7. Absolute value of the components of the eigenvectors (A) 1, (B) 2, (C) 3, and (D) 5 obtained from the $C\alpha$ coordinate covariance matrix of the concatenated trajectories of $\alpha 3(VI)$ in the major and minor conformations as function of residue numbers. The dashed lines in A indicate disulfide bridges between cysteines. The binding loop and the glycine rich region are indicated in B. In the upper right corner of Figure 7A–D, the concatenated trajectory (the first 800 ps represent $\alpha 3(VI)$ in M conformation, the last 800 ps $\alpha 3(VI)$ in minor conformation) has been projected onto the respective eigenvectors to show that the average value of projection Figure 6A and mean-square fluctuations Figure 6B make a distinction between structural differences and dynamic mode. From these plots it is seen that the first eigenvector represents a structural difference between the M and m state, which is not dynamically connected, while the other are dynamic modes, some (nr 3) in the M conformation, some (nr 5) in the minor conformation, and some (nr 2 and 4) in both.

Cys30–Cys51 disulfide bridge. The conformation of this disulfide bridge was stable throughout both MD simulations.

In conclusion, we have identified the minor conformation of the Trp21 side chain in the $\alpha 3(VI)$ molecule. Evidence for this conclusion is substantiated by the calculated free energy $\Delta G_{(M \rightarrow m)} = 1.77 \pm 0.7$ kcal/mol and chemical shift differences in the two conformations. The most pronounced effect of the reorientation of the Trp21 indole ring was seen for Gln48 ($H\beta 2$), where $\Delta\delta(M \rightarrow m) = 1.96$ ppm. This is in qualitative agreement with NMR data, where a value of 1.53 ppm was measured (Sørensen et al., 1996). Essential dynamics analysis of the concatenated trajectories extracted from the simulations of the major and minor conformers shows that the reorientation of the Trp21 side-chain triggers multiple conformational changes in the central β -sheet, which is in accordance with a recently conducted experimental study on the $\alpha 3(VI)$ molecule (Zweckstetter et al., 1996). Visualization of the motions

along different eigenvectors reveals that in the minor conformation the binding loop (Thr13–Ile18) has a relatively higher degree of fluctuations than in the M conformation. This suggest that the reorientation of the Trp21 side chain could have an effect on the inhibition of trypsin.

Methods

Adiabatic mapping

An adiabatic iso energy contour map was constructed by successively constraining the $\chi 1$ and $\chi 2$ dihedral angles to a grid point (from 180° to -180° , in steps of 5°) using a dihedral harmonic potential with a barrier height of 1,000 kcal/mol and energy-minimizing each of these conformations. All other protein atoms were free to move. The CHARMM22 molecular modeling package

with polar hydrogen atoms (Brooks et al., 1983; CHARMM, 1992) was applied. $73 * 73 = 5,329$ independent conformations were generated and analyzed. Each conformation was generated from the same starting structure (the crystal structure, Arnoux et al., 1995) and subjected to energy minimization (150 steps of the steepest descent (SD) and 50 steps of the conjugated gradient (CONJ) methods). Energies were calculated with a 10 Å cut-off radius for the nonbonded interactions and all the calculations were done a distance-dependent dielectric constant to take into account implicitly the effect of the solvent (Weiner et al., 1984; Brooks et al., 1988). (In addition the calculations were performed without a distance-dependent dielectric constant. The result of this calculation identifies exactly the same low energy conformations of the Trp21 side chain (data not shown) and gives qualitatively the same results as the calculation performed with a distance-dependent dielectric constant.)

Free energy calculations

The TI method is a free energy simulation technique allowing calculation of the free energy difference between an initial state i and a final state f , represented by two different Hamiltonians. The two states are coupled by a control variable λ (also called a reaction coordinate) in such a way that for $\lambda = 0$ the system is represented by a Hamiltonian corresponding to the initial state and for $\lambda = 1$ by a Hamiltonian corresponding to the final state. Variation of λ between 0 and 1 allows an interconversion between the two states. The calculation is based on the definition of a hybrid Hamiltonian, which represents the initial state i and the final state f and is given by

$$H(\mathbf{q}, \mathbf{p}, \lambda) = H_0(\mathbf{q}, \mathbf{p}) + (1 - \lambda)^n H_i(\mathbf{q}, \mathbf{p}, \lambda) + \lambda^n H_f(\mathbf{q}, \mathbf{p}, \lambda) \quad (1)$$

where n describes the degree of nonlinearity for the thermodynamic pathway (in this study we apply $n = 1$). The 3N-dimensional vectors $\mathbf{q}$ and $\mathbf{p}$ represent the position and conjugate momenta of all the particles. H_i and H_f are Hamiltonians describing the atoms that are involved in the perturbation and are present in the initial and final state, respectively. H_0 represents all remaining atoms. The free energy change is calculated using Equation 2. An analytic expression for the derivative of the Hamiltonian (Equation 1) with respect to λ is evaluated at each time step of the simulation and integrated over the change in λ . The free energy between the two states i and f is given by

$$\Delta G = \int_0^1 \langle \partial H / \partial \lambda \rangle_\lambda d\lambda. \quad (2)$$

In the simulation, ΔG is approximated by a sum over ensemble averages for the derivative of the Hamiltonian with respect to λ at discrete values of λ . It is assumed that the intermediate states are not too dissimilar from each other, i.e., conformational differences are small, which can be accomplished by choosing a sufficient small $\Delta\lambda$ value (Elofsson et al., 1993). The average quantities $\langle \dots \rangle_\lambda$ in Equation 2) are calculated for a number of windows spanning from $\lambda = 0$ to $\lambda = 1$. The accumulated free energy change is numerically determined using the quadrature method. In this study, the relative free energy difference between the two conformations of the Trp21 side chain is calculated by slowly growing/creating the m conformation (with a dependence on λ as given in

Equation 1) and at the same time annihilating the M conformation. Atoms added to the system are denoted *product atoms* and the ones that are deleted are denoted *reactant atoms*. The simulation procedure for calculating free energy differences is illustrated in Figure 4. The DONT option for both reactant and product for the bond stretching and bending terms was applied in this study. This command implies that the specified internal energy term for the reactant or product atoms are ignored. This means that the specified interactions are not factored by λ^n or $(1 - \lambda)^n$ and do not contribute to the value of $V(\text{react})$ or $V(\text{product})$ within the perturbation terms in the Hamiltonian. The interaction is, however, computed in full and treated as part of H_0 (Equation 1). These terms are generally uncoupled from the interactions of interest and it appears that their exclusion, even with the resulting nonphysical Hamiltonian, does not significantly effect the relative free energies of solvation or drug/enzyme binding. This option has been successfully used to avoid unrealistic motions of atoms when λ are close to 0 or 1 (Elofsson et al., 1993).

The TI method was applied to calculate the relative free energy difference between the two conformers. The perturbation from M to m was carried out with a dual topology representation, which involves all 11 Trp21 side chain atoms (see Fig. 4). This was the minimum number of atoms that had to be changed to carry out the perturbation. Inclusion of more atoms would increase the possible structural difference between the two conformations, and consequently would increase the computational requirements for sampling the configurational space. The average quantity $\langle \dots \rangle$ in Equation 2 was calculated by performing a series of independent simulations for different λ values ranging from 0 to 1. The use of independent trajectories is computationally attractive, since it offers the advantage to control the equilibration and convergence of each λ simulation independently. Eleven independent simulations were performed at fixed λ -values; $\lambda = 0 \rightarrow 1$ in steps of $\Delta\lambda = 0.1$. The integration was performed applying a cubic spline interpolation.

An important question, which has led to several theoretical and experimental studies (Fersht, 1988; McCammon & Straatsma, 1992; Gilquin et al., 1993; Kollman, 1993), is the effect of water-solute interactions in free energy calculations. It appears that the organization, e.g., structural factors like hydrogen bond patterns or dynamic properties of the water molecules around the isomerization state, appears crucial in determining the population of different conformations (Gilquin et al., 1993). To take solvent molecules into account, the Trp21 side chain was solvated in the center of an equilibrated TIP3 water sphere (Jørgensen et al., 1983). All water molecules within 2.4 Å of any nonhydrogen protein atom were removed. Atoms located within 13 Å from the Trp21 residue were allowed to move freely. A so-called extended wall region (van Gunsteren et al., 1993) was applied from 13 to 16 Å, where the motions of the atoms were restrained to their initial positions by a harmonic potential with a force constant of 5 kcal/mol/Å². All atoms in the remaining part of the molecule were fixed at their original positions. The water molecules were subjected to 200 steps of SD and 200 steps of CONJ energy minimization while the protein atoms were fixed at their crystal structure coordinates. The total system was then heated. The final temperature of 303 K was obtained after 6 ps. A time step of 0.001 ps was employed. The thermalization of the system was followed by an initial MD equilibration run of 150 ps using a time step of 0.002 ps. The atom-pair list was updated every 0.01 ps. The final structure was used as a starting point for calculating the free energy difference between the major and minor conformations. The calculations were performed

using the program package CHARMM (Brooks et al., 1983) with free energy software implemented as described by Fleischman and Brooks (1987). We used the standard CHARMM parameter set PARM 22, which does not contain any specific energy terms for hydrogen bonds and only includes polar hydrogens.

Molecular dynamics; essential dynamics analysis

The coordinate of the high resolution crystal structure of the $\alpha 3$ (VI) molecule (Arnoux et al., 1995) was obtained from the Brookhaven Protein Data Bank (Bernstein et al., 1977) (entry code 1knt). Molecular dynamic simulations were performed for the $\alpha 3$ (VI) molecule with the Trp21 side chain confined to the minimum energy conformation of the M or m isomers to investigate the static and dynamic properties, and to compare the concerted motions occurring in the two conformers. The minimum-energy conformations (M and m conformers) of the Trp21 side chain obtained from the adiabatic map were used as starting points for the molecular dynamics simulations. The simulations were performed using the CHARMM22 force field, which includes polar hydrogens. Periodic boundary conditions with a box size of $56 \times 45 \times 45 \text{ \AA}^3$ were applied in both simulations. One Cl^- ion was added to the box to neutralize the system at $\text{pH} = 7$. The system was heated over a period of 6 ps to the final temperature of 300 K. The nonbonded pair list was updated every 10 steps, and the nonbonded and electrostatic interactions were truncated at 8 and 10 $\text{\AA}$, respectively. The SHAKE algorithm (Ryckaert et al., 1977) was applied to constrain the bond lengths to their equilibrium positions, and the equation of motion were solved using the Verlet algorithm. Simulations were performed at 300 K with a time step of 2 fs. Examination of the molecular structures and analysis of the trajectories were carried out using the essential dynamics technique (Amadei et al., 1993) provided within the modeling program WHAT IF (Vriend, 1990).

The essential dynamics method is based on the diagonalization of the covariance matrix built from atomic fluctuations in a trajectory from which overall translation and rotation have been removed (Amadei et al., 1993; van Aalten et al., 1995; Peters et al., 1996). The diagonalization yields a set of eigenvectors and corresponding eigenvalues, where the vectors describe correlated positional changes of groups of coordinates in the 3N conformational space. The eigenvalues indicate the total mean-square fluctuation along these directions. In general, it has been shown for several proteins (Amadei et al., 1993; Peters et al., 1996) that only the eigenvectors with large corresponding eigenvalues are important to describe the overall motion of the protein, and that less than the first 50 eigenvectors represents about 90% of the total motion of the protein. Projection of the trajectory onto a certain eigenvector allows to study the time dependence of the motion along this eigenvector, and can be used to generate a visualization of this motion (Peters et al., 1996). A useful method for comparing the essential dynamics of two simulations on similar systems is the so-called combined analysis (van Aalten et al., 1995). In this method, the individual trajectories fitted to the same reference structure are concatenated, and a covariance matrix is constructed. The combined trajectory is then projected onto the resulting eigenvectors, and the properties of these projections identify motions along the trajectory. There are two main quantities to be studied: (1) the differences in average projection, indicating structural difference between the simulated structures, and (2) differences in the mean-square fluctuations of the projections, indicating differences in the

dynamics along that direction. Here, this method is applied to the concatenated trajectories of M and m simulations.

References

- Amadei A, Linssen AMB, Berendsen HJC. 1993. Essential dynamics of proteins. *Proteins Struct Funct Genet* 17:412–425.
- Arnoux B, Mérieux K, Saludian P, Norris F, Norris K, Björn S, Olsen O, Petersen L, Ducruix A. 1995. The 1.6 $\text{\AA}$ structure of Kunitz-type domain from the $\alpha 3$ chain of human type VI collagen. *J Mol Biol* 246:609–617.
- Asakura T, Williamsson M, Magn L. 1993. Empirical comparisons of models for chemical-shift calculation in proteins. *J Magn Reson Ser B* 101:63–71.
- Berndt KD, Günter P, Orbons LPM, Wüthrich K. 1992. Determination of a high-quality structure of the bovine pancreatic trypsin inhibitor and comparison with tree crystal structures. *J Mol Biol* 227:757–775.
- Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rogers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer based archival file for macromolecular structure. *J Mol Biol* 112:535–542.
- Boresch S, Archontis G, Karplus M. 1994. Free energy simulations: The meaning of the individual contributions from a component analysis. *Proteins Struct Funct Genet* 20:25–33.
- Boresch S, Karplus M. 1995. The meaning of component analysis: Decomposition of the free energy in terms of specific interactions. *J Mol Biol* 254:801–807.
- Brooks BR, Bruccoleri RE, Olafsen BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comp Chem* 4:187–217.
- Brooks CL III, Karplus M, Pettitt BMA. 1988. Theoretical perspective of dynamics, structure and thermodynamics. *Proteins Adv Chem Phys* 71.
- Case DA, Karplus M. 1979. Dynamics of ligand binding to heme proteins. *J Mol Biol* 132:343–368.
- CHARMM Molecular Modeling Software Package, Release 3.3. 1992. Waltham, Massachusetts: Molecular Simulations Inc.
- Chu ML, Zhang R, Pan TC, Stokes D, Conway D, Kuo HJ, Glanville R, Meyer U, Mann K, Deutzmann R, Timpl R. 1990. Mosaic structure of globular domains in the human type VI collagen $\alpha 3$ chain: Similarity to Von Willibrand factor, fibronectin, actin, salivary proteins and aprotinin type protease inhibitors. *EMBO J* 9:385–393.
- Derewenda U, Swenson L, Wei Y, Green R, Kobos PM, Joerger R, Haas MJ, Derewenda ZS. 1994. Conformational lability of lipases observed in the absence of an oil-water interface: Crystallographic studies of enzymes from the fungi *Humicola lanuginosa* and *Rhizopus delema*. *J Lip Res* 35:524–534.
- Elofsson A, Kulinski T, Rigler R, Nielson L. 1993. Site specific point mutation changes specificity: A molecular modeling study by free energy simulations and enzyme kinetics of the thermodynamics in ribonuclease T₁ substrate interactions. *Proteins Struct Funct Genet* 17:1661–175.
- Fersht AR. 1988. Relationships between apparent binding energies measured in site-directed mutagenesis experiments and energetics of binding and catalysis. *Biochemistry* 27:1577–1580.
- Fleischman SH, Brooks CL. 1987. Thermodynamics of aqueous solvation: Solution properties of alcohols and alkanes. *J Chem Phys* 87:3029–3037.
- Fleischman SH, Zichi D. 1991. A. Free energy simulations of methane solvation. A study of integrand convergence properties using thermodynamic integration. *J Chem Phys* 88:2617–2622.
- Gilquin B, Perahia D, Dive V. 1993. A free-energy study of a bacterial collagenase inhibitor. *Int J Pept Protein Res* 43:384–392.
- Goldsack DE. 1976. Relation of the hydrophobicity index to the thermal stability of homologous proteins. *Biopolymers* 9:247–252.
- Heald SL, Tilton RF, Hammond LJ, Lee A, Bayney RM, Kamarck ME, Ramabhadran TV, Dreyer RN, Davis G, Unterbech A, Tamburini PP. 1991. Sequential NMR resonance assignment and structure determination of the Kunitz-type inhibitor domain of the Alzheimer's β -amyloid precursor protein. *Biochemistry* 30:10467–10478.
- Hodel A, Simonson T, Fox AO, Brünger AT. 1993. Conformational substrates and uncertainty in macro-molecular free energy calculations. *J Phys Chem* 97:3409–3417.
- Jørgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulation liquid water. *J Chem Phys* 79:926–935.
- Joseph-McCarthy D, Petsko GA, Karplus M. 1995. Use of a minimum perturbation approach to predict TIM mutant structures. *Protein Eng* 8:1103–1115.
- Kitaguchi N, Takahashi Y, Tokushima Y, Shiojiri S, Ito H. 1988. *Nature* 331:530–532.

- Kollman PA. 1993. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem Rev* 93:2395–2417.
- Mark AE, van Gunsteren WF. 1994. Decomposition of the free energy of a system in terms of specific interactions. Implications for theoretical and experimental studies. *J Mol Biol* 240:167–176.
- Mazor M, Pettitt BM. 1991. Convergence of the chemical potential in aqueous simulations. *Mol Simul* 6:1–4.
- McCammon JA, Straatsma TP. 1992. Alchemical free energy simulation. *Ann Rev Phys Chem* 43:407–416.
- Mitchell MJ, McCammon JA. 1991. A. Free energy difference calculations by thermodynamic integration—difficulties in obtaining a precise value. *J Comp Chem* 12:271–275.
- Northrup SH, Pear MR, Lee CY, McCammon JA, Karplus M. 1982. Dynamical theory of activated processes in globular proteins. Dynamical theory of activated processes in globular proteins. *Proc Nat Acad Sci USA* 79:4035–4039.
- Otting G, Liepinsh E, Wüthrich K. 1993. Disulfide bond isomerization in BPTI and BPTI(G36S): NMR study of correlated mobility in proteins. *Biochemistry* 32:3571–3582.
- Pearlman DA. 1993. Determining the contributions of constraints in free energy calculations: Development, characterization, and recommendations. *J Comp Phys* 98:8946–8957.
- Pearlman DA. 1994. Free energy derivatives: A new method for probing the convergence problem in free energy calculations. *J Comp Chem* 15:105–123.
- Pearlman DA, Kollman PA. 1991. The overlooked bond-stretching contributions. *J Chem Phys* 94:4532–4545.
- Peters GH, van Aalten DMF, Edholm O, Toxvaerd S, Bywater R. 1996. Dynamics of proteins in different solvent systems: Analysis of essential motion in lipases. *Biophys J* 71:2245–2255.
- Philippopoulos M, Xiang Y, Lim C. 1995. Identifying the mechanism of protein loop closure: A molecular dynamics simulation of the *Bacillus stearothermophilus* LDH loop in solution. *Protein Eng* 8:565–573.
- Ryckaert J-P, Ciccotti G, Berendsen HJC. 1977. Numerical integration of the Cartesian equation of motion of a system with constraints: Molecular dynamics of n-alkanes. *J Comput Phys* 23:327–335.
- Shi Y-Y, Mark AE, Wang C, Huang F, Berendsen HJC, van Gunsteren WF. 1993. Can the stability of protein mutants be predicted by free energy calculations? *Protein Eng* 6:289–295.
- Shih HH-L, Brady J, Karplus M. 1985. Structure of proteins with single-site mutations: A minimum perturbation approach. *Proc Natl Acad Sci USA* 82:1697–1700.
- Sørensen MD, Bjørn S, Norris K, Olsen O, Petersen L, James LT, Led JJ. 1997. Solution structure and backbone dynamics of the human α 3-chain type VI collagen C-terminal Kunitz domain. *Biochemistry* 36:10439–10450.
- Sørensen MD, Kristensen SM, Bjørn S, Norris K, Olsen O, Led JJ. 1996. Elucidation of the origin of multiple conformations of the human α 3-chain type VI collagen C-terminal Kunitz domain: The reorientation of the Trp21 ring. *J Biomol NMR* 8:391–403.
- Straatsma TP, McCammon JA. 1991. Multiconfiguration thermodynamic integration. *J Chem Phys* 95:1175–1188.
- Young HD. 1962. *Statistical treatment of experimental data*. New York: McGraw-Hill.
- van Aalten DMF, Amadei A, Linssen ABM, Eijssink VGH, Vriend G. 1995. The essential dynamics of thermolysin-conformation of the hinge-bending motion and comparison of simulations in vacuum and water. *Proteins Struct Funct Genet* 22:45–54.
- van Gunsteren WF, Beutler TC, Fraternali F, King PM, Mark AE, Smith PE. 1993. Computation of free energy in practice: Choice of approximations and accuracy limiting factors. In: van Gunsteren WF, Weiner PK, Wilkinson AJ, eds. *Computer simulations in biological systems*. Leiden, The Netherlands: Escon Science Publishers.
- Vriend G. 1990. WHAT IF: A molecular modelling and drug design program. *J Mol Graph* 8:52–56.
- Wade CR, Davies ME, Luty BA, Madura DJ, McCammon JA. 1993. Gating of the active site of triose phosphate isomerase: Brownian dynamics simulations of flexible peptide loops in the enzyme. *Biophys J* 64:9–15.
- Wade CR, Luty BA, Demchuk E, Madura DJ, Davies ME, Briggs JM, McCammon JA. 1994. Simulation of enzyme-substrate encounter with gated active-sites. *Nature Struct Biol* 1:63–67.
- Wade CR, McCammon A. 1992. Binding of an antiviral agent to a sensitive and a resistant human rhinovirus. Computer studies with sampling of amino acid side-chain conformations. *J Mol Biol* 225:679–696.
- Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S Jr, Weiner P. 1984. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J Am Chem Soc* 106:765–784.
- Zweckstetter M, Czisch M, Mayer U, Chu M-L, Zinth W, Timpl R, Holak TA. 1996. *Structure* 4:195–209.