

On the Unfolding of α -Lytic Protease and the Role of the Pro Region

Yoshihiko Inuzuka and Themis Lazaridis*

Department of Chemistry, City College of the City University of New York, New York, New York

ABSTRACT Molecular dynamics simulations of α -lytic protease (α LP) alone and complexed with its pro region (PRO) are performed to understand the origin of its high unfolding (and folding) barrier when it is alone and how the pro region lowers this barrier. At room temperature, α LP exhibits lower dynamic fluctuations than α -chymotrypsin. Simulation of PRO alone led to reorientation of its N terminal helix and collapse to a more compact state. A model for the uncleaved proenzyme was built and found to be stable in the time scale of the simulations. Energetic analysis suggests that the origin of strain in the uncleaved proenzyme compared with the cleaved complex is in the intramolecular backbone electrostatic interactions of the cleaved strand. In high temperature simulations, the interaction of the long beta hairpin of the enzyme with the C terminal beta sheet of PRO is among the most stable in the complex and a likely “nucleation site” for folding. In the course of unfolding, the C terminal tail of PRO is sometimes observed to intervene between the long hairpin and the aspartate loop of the enzyme, perhaps thereby lowering the energy barrier for separation of the two hairpins. Tighter interactions at the interface between the enzyme and its pro region are also occasionally observed, providing an additional mechanism for unfolding catalysis. Simulations of a mutant enzyme where the buried ion pair residues R102 and D142 were replaced by W and L, respectively, did not display any distinguishable behavior compared with the wild type. *Proteins* 2000;41:21–32.

© 2000 Wiley-Liss, Inc.

Key words: α -lytic protease; protein folding; proenzymes; kinetic control

INTRODUCTION

α -Lytic protease (α LP) is a serine protease synthesized as a pre-pro-enzyme and secreted by the bacterium *Lyso bacter enzymogenes*. The N-terminal 33-residue signal peptide (the pre sequence) serves to target the molecule to the periplasmic space. There, the signal peptide is cleaved and the resulting proenzyme folds and cleaves itself to produce the PRO- α LP complex. This complex is transported to the extracellular space, where the 166-residue pro region is degraded by proteases to give the mature, 198-residue enzyme.

Biophysical studies have shown that the pro region is necessary for proper folding of the enzyme, either in *cis*

(covalently bound) or in *trans* (provided as a separate molecule).¹ In the absence of the pro region, α LP remains in a “molten globule”-like state, with some secondary structure but little or no tertiary structure.² However, once the enzyme is folded with the help of the pro region, it remains in the native state even when the pro region is removed. This finding is an example of a protein conformation under kinetic control.^{2,3} From a combination of folding and unfolding kinetic studies, it has recently been shown that the native mature enzyme is in a metastable state: it is higher in free energy than the unfolded state but remains folded due to a high unfolding barrier.⁴

The crystal structures of the mature enzyme^{5,6} and of the pro region alone and in complex with α LP⁷ are known. α LP consists of an N terminal and a C terminal domain (residues 1–83 and 84–184, respectively), both with a β -barrel topology, and a three-turn helix (residues 185–191) near the C terminus (see Fig. 1). Figure 2 shows the nomenclature for the secondary structure elements used in the discussion below. The structure of the pro region in its own crystal is similar to that in the complex, suggesting that the pro region is an autonomously folding region. In the complex, the C terminus of the pro region is bound to the active site of the enzyme in a product-like configuration. The N-terminus of the enzyme, which was covalently bonded to the PRO C terminus, rearranges after cleavage to form part of the N terminal α LP β -barrel.

These experimental studies shed considerable light into the mechanism of folding catalysis by PRO. The pro region was found to stabilize the native state by more than 13 kcal/mol and the transition state (TS) by an even greater amount (approximately 4 kcal/mol more than the native state). Still, the microscopic origin of these energetic effects is not clear. Specifically, it is not clear what gives rise to the exceptionally high unfolding (and folding) barrier of α LP and which interactions with PRO lower this barrier. Possible answers to these questions are here sought by using molecular dynamics (MD) simulations of the mature enzyme alone (ALP), a mutant enzyme for which a buried ion pair is replaced by hydrophobic residues (ALPmut), the pro region alone (2PRO), the wild-type

Grant sponsor: National Institutes of Health; Grant number: RCMI grant; Grant sponsor: ACS PRF.

*Correspondence to: Themis Lazaridis, Department of Chemistry, City College of the City University of New York, 138th Street & Convent Avenue, New York, NY 10031. E-mail: themis@sci.cuny.cuny.edu

Received 4 January 2000; Accepted 23 May 2000

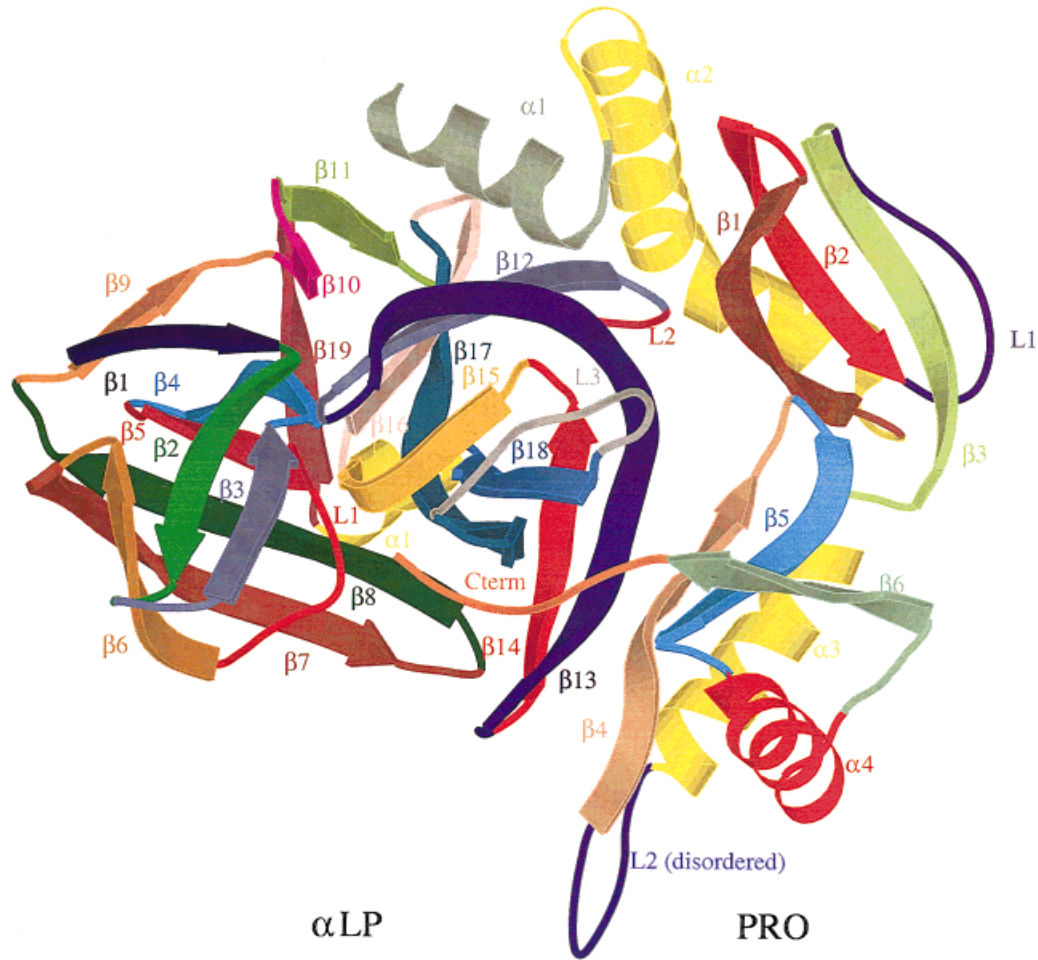


Fig. 1. Ribbon representation of the structure of the PRO- α LP complex. The labels correspond to the structural elements defined in Figure 2. The view is the same as in Figure 2 of Sauter et al.⁷

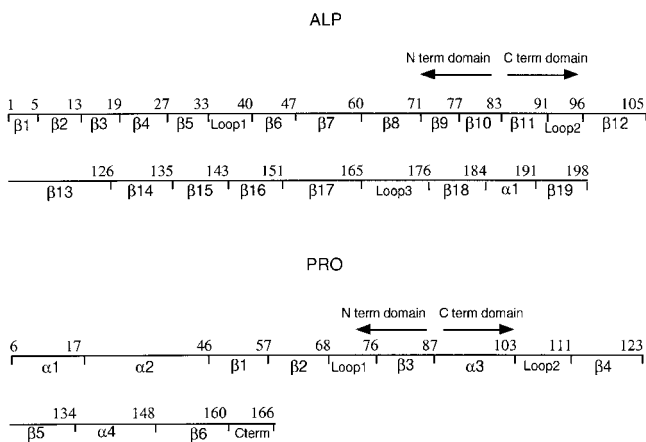


Fig. 2. Nomenclature for the secondary structure elements of α LP and PRO. Residue numbering is sequential (not by homology to chymotrypsin).

complex (4PRO+ALP), the mutant complex with a truncated PRO C-terminus (3PRO+ALP), and the uncleaved proenzyme (PROALP). Simulations were first performed

at 300 K to assess the stability of these structures in solution and subsequently at 500 K to probe the mechanism of unfolding. The simulations suggest new experiments and lead to several experimentally testable predictions.

METHODS

All simulations were performed with the CHARMM program¹⁶ and the EEF1 energy function,^{8,10} which is based on the polar hydrogen CHARMM energy (Charmm 19 parameter set)¹⁷ and includes an implicit solvation term. This function uses neutralized ionic sidechains and a linear distance dependent dielectric constant, which eliminates long range electrostatic interactions. EEF1 has been shown to give RMS deviations from experimental structures similar to those in explicit solvent simulations,⁸ thermal unfolding pathways in agreement with explicit solvent simulations,¹⁰ and to discriminate the native from misfolded protein structures.¹⁸ The same energy function was used for the calculation of the heat capacity and compactness of denatured proteins,¹⁹ forced unfolding of fibronectin domains,²⁰ and the folding thermodynamics of

a β -hairpin.²¹ No explicit water molecules or counterions were included in the simulations.

The starting point of all simulations was the available crystal structures: the mature enzyme (2ALP),⁶ the pro region (2PRO, molecule A was selected),⁷ the complex of α LP with WT PRO (4PRO),⁷ and the complex of α LP with mutant PRO lacking the last three C-terminal residues (3PRO, molecules A and C).⁷ In all cases, missing hydrogen coordinates were built with the HBUILD algorithm,²² followed by 300 steps of energy minimization with the ABNR method.¹⁶ Molecular dynamics was performed by using the Verlet algorithm with a 2-fs time step and the use of SHAKE²³ for the bonds involving hydrogens. For comparison with α LP, a previous 50-ps simulation of α -chymotrypsin with EEF1⁸ was continued for another 200 ps.

PRO contains a disordered loop (residues 103–114 in 2PRO, 105–111 in 4PRO, and 105–112 in 3PRO) that is not visible in the crystal structure. An initial conformation for this loop was generated by building the loop in an extended conformation and energy minimizing for 1,000 steps keeping the rest of the protein fixed. The whole molecule was subsequently energy-minimized for 300 steps.

Explicit solvent simulations of PRO were performed by using the stochastic boundary method.²⁴ A 36 Å radius sphere of TIP3P water molecules was overlaid on the PRO molecules and all water molecules with oxygens 2.8 Å from any protein heavy atoms were deleted. This sphere size is sufficient to provide two or more solvent layers between any protein atom and the boundary (one layer for Lys 103 NZ). The solvent was thermalized for 2.5 ps, this overlay was repeated four times, and the solvent was thermalized again for another 5 ps. Then the system was equilibrated for 25 ps at 300 K and dynamics was run for 900 ps. The water molecules between 34 and 36 Å were subjected to Langevin dynamics and the rest of the system to Newtonian dynamics.

A model for the uncleaved proenzyme was generated based on 4PRO. A bond was introduced between the C terminus of PRO and the N terminus of α LP. The β 1 strand of α LP (residues 1–5) was allowed to move during a 1,000-step energy minimization while the rest of the protein was held fixed. Then the whole protein (359 residues) was minimized for 300 steps and subjected to a 1-ns simulation at 300 K.

A model for the R102W/D142L mutant of α LP was generated based on the 2ALP crystal structure. The leucine side-chain was overlaid on that of Asp. The tryptophan side-chain was overlaid upon the Arg side-chain up to the CG atom. The remaining atoms were built based on the ideal geometry of the Trp side-chain. Subsequently, the mutant was subjected to the same protocol as the WT molecule (300 steps of ABNR minimization, 50 ps of dynamics at 300 K, and three 500-K simulations with different random number seeds).

The simulation of PRO in the crystal state was performed by using the CRYSTAL/IMAGE facility of CHARMM. The three molecules of PRO in the unit cell were treated explicitly and the neighboring unit cells were

treated as images of the primary atoms. Only atoms within 10 Å of any primary atom are considered in the construction of the image-primary nonbonded lists. First, the constructed disordered loops were allowed to move during a 1,000-step minimization keeping the rest of the system fixed. Subsequently, the whole system was energy-minimized for 300 steps, keeping the lattice parameters constant. Then a dynamics simulation was run for 550 ps.

As in previous work,^{10,25} several native contacts were monitored during the high temperature simulations to visualize unfolding pathways in detail. The contacts (hydrogen bonds and tertiary contacts) were chosen from a list of atom pairs that are closer than approximately 4 Å in the crystal structure. A contact is said to exist if the distance between the two atoms is less than 1.5 times that in the crystal structure. The contacts are grouped according to the structural elements they connect (Table I; Fig. 2).

Effective interactions between structural elements are calculated with the INTE command in CHARMM and include the direct interactions (van der Waals, electrostatics) and the amount of solvation free energy the atoms in each structural element lose due to the presence of the other structural element. “Effective energy” is the sum of intramolecular energy and the solvation free energy, i.e., it does not include any conformational entropy effects (it is the free energy for a fixed macromolecular conformation). The term “free energy” is used to denote quantities that include conformational entropy effects.

RESULTS

300-K Simulations

Mature enzyme

α LP alone was simulated at 300 K for 1 ns. The backbone root mean square deviation (bRMSD) from the crystal structure at the end of the simulation was 2.4 Å and the radius of gyration (Rg) 15.6 Å (from 15.01 Å in the crystal structure). This deviation is somewhat larger than those observed in explicit solvent simulations (typically between 1 and 2 Å, occasionally higher).⁸ The most significant change is a relative shift of the N and C terminal domains. This shift results in an increase in the distance between the catalytic triad residues, two of which (His36 and Asp63) reside in the C terminal domain and the third (Ser143) in the N terminal domain. The distance between His36 ND1 and Asp63 OD1 becomes 6.05 Å (from 2.7 in the crystal structure), that between His36 ND1 and Ser143 OG becomes 9.3 Å (from 4.52), and that between 63 OD1 and 143 OG becomes 12.9 Å (from 6.2). A similar shift occurs in the present simulation of α -chymotrypsin with the same energy function, although to a smaller extent (for example, the His57 NE to Ser195 HG distance becomes 4.553 from 2.141 Å). The origin of this shift is not clear, but it is not expected to affect the conclusions of this study, which is concerned with large scale unfolding. It may be partly due to the omission of counterions in the simulations. In the α LP crystal structure, there is a sulfate ion at the active site, which may play a stabilizing role, although it is not observed in other serine protease structures.



Fig. 3. Structure of PRO after 800 ps of simulation at 300 K (b) compared with the initial (crystal) structure (a).

Because α LP is thought to be more rigid than other serine proteases because of its shorter loops,⁴ we examined the RMS fluctuations during the last 200 ps of the dynamics run and compared them with those from an equal-time simulation of α -chymotrypsin. The average RMS fluctuation of all atoms was 1 Å for α LP and 1.42 Å for α -chymotrypsin. Indeed, α LP seems to be overall less flexible than α -chymotrypsin on this time scale. Even if we include in the average only buried atoms (those atoms with solvent accessibility less than 0.1 Å), we obtain 0.85 Å for α LP and 1.17 for α -chymotrypsin. Thus, the dynamic difference between the two proteins is not restricted to loop regions but propagates to the buried atoms.

Pro region

The crystal structure of PRO alone contains three molecules in the unit cell, which differ mainly at the N terminal helix, which adopts different orientations in the two molecules and does not exist in the third. Molecule A of the crystal structure was chosen for the simulation, in which the N terminal helix points away from the molecule. There is a significant shift in the structure during the 800-ps simulation (see Fig. 3). The bRMSD at the end of the simulation was 8.5 Å and the $R_g = 18.3$ Å (compared with $R_g = 19.2$ Å in the crystal structure). The N terminal helix partially melts and changes orientation. It actually adopts an orientation similar to that of molecule B in the crystal structure. Hence, this change is not surprising, considering the variability concerning this helix among

the three molecules in the unit cell. More unexpected is the change in helix $\alpha 4$ in the C terminal domain. This helix partially melts and changes position and orientation with respect to the rest of the C terminal domain. In addition, the whole C terminal domain reorients and moves toward the N terminal $\alpha 2$ helix to close the cavity that exists in the crystal structure. It is possible that the conformation observed in the crystal is due to crystal packing; the collapse observed in the simulation is not possible in the crystal because in the crystal the concave surface of PRO is occupied by molecules from surrounding unit cells. The kind of open structure observed in PRO is rare in the database of known protein structures. In fact, one can hardly imagine what physical forces would rigidly keep PRO in that conformation if it has no other molecule to interact with. On the other hand, it seems an uncanny coincidence that the crystal packing forces in 2PRO would force PRO to a conformation so similar to that in another crystal (4PRO) (the bRMSD between molecule A of 2PRO and molecule C of 4PRO is approximately 2 Å). Of course one cannot rule out the possibility that the collapse is caused by deficiencies in the energy function.

To further investigate this issue, we performed two additional sets of simulations. First, we simulated one molecule of PRO in explicit solvent for 1 ns. Because large scale conformational changes, such as collapse and helix reorientation, take place in much longer time scales in explicit solvent due to solvent friction and caging effects, this time scale may not be long enough to produce the

compaction observed in the implicit solvent simulations. Indeed, the deviations observed here are much smaller than in the implicit solvent simulations (bRMSD 3.3 Å). Nevertheless, a small compaction of the protein is observed ($R_g = 18.96$ from 19.11 Å), the N terminal helix ($\alpha 1$) changes orientation by 40 degrees, and the $\alpha 4$ helix loses one turn.

The second test was to simulate PRO in the crystal. In this calculation, the three molecules of the unit cell are treated explicitly and the surrounding molecules are treated as images of the primary atoms. In the 550 ps of the simulation, we do observe shifts from the crystal structure, but they are smaller than in the simulation of PRO alone. The bRMSD for molecules A, B, and C is 5.1, 4.0, and 3.7 Å, respectively. The changes observed in this simulation are qualitatively different from those observed in the simulation of isolated PRO; no collapse of the concave surface is observed. The R_g of molecule A becomes 20.7 Å (from 19.2), of molecule B 19.8 Å (from 18.6) and of molecule C remains the same at 18.3 Å. In view of the simulation results, it would be useful to obtain additional experimental data on the conformation of PRO in solution, possibly by NMR.

Complex

The 300-K simulation results for 4PRO+ALP and 3PRO+ALP are similar. After 1 ns, the bRMSD for 4PRO+ALP was 3.3 Å for both molecules (2.7 Å for the enzyme alone and 3.4 Å for PRO alone). The R_g becomes 20.2 Å, from 20.3 Å in the crystal. 3PRO+ALP after 1 ns of simulation shows a bRMSD of 3 Å for the complex (2.4 Å for αLP alone and 3.1 Å for PRO alone). The R_g becomes 20.43 Å, from 20.17 Å in the crystal.

Uncleaved proenzyme

The model for the uncleaved proenzyme is very similar to the complex, except for the $\beta 1$ strand, which has to move out of its position in αLP to connect with the C terminus, which is approximately 24 Å away. Other than this change, the uncleaved proenzyme appears stable in the 300-K simulation. After 1 ns of simulation at 300 K the bRMSD from the starting structure is 4.3 Å (3.3 Å for the enzyme portion alone and 3.6 Å for the PRO portion alone) and the R_g 20.1 Å, slightly decreased from 20.2 Å in the starting structure. Other than a partial melting of the N terminal helix of PRO, the other secondary structure elements remain intact, although they shift a little with respect to each other producing the 4.3 Å RMSD. These RMS deviations are larger than those in the cleaved complexes 4PRO and 3PRO (see above), probably because of the destabilizing effect of the covalent link on the rest of the molecule. If we start from the final structure of the simulation of the uncleaved proenzyme and remove the bond between the enzyme and the PRO region, the $\beta 1$ strand does not return to its position in the enzyme within 200 ps, which reveals that the uncleaved proenzyme is in a local effective energy minimum.

The uncleaved proenzyme recently has been the subject of an experimental investigation⁹ and was found to have

reduced stability compared with the cleaved enzyme-pro complex. One might expect this to be due to the fact that the $\beta 1$ -Cterm strand has less favorable interactions in the uncleaved proenzyme and after cleavage the two segments relax to energetically more favorable configurations. To test this expectation, we compared the interactions of the $\beta 1$ -Cterm strand with itself and with the rest of the protein in the uncleaved enzyme and the cleaved complex. The values reported below refer to averages over the last 10 ps of the 1-ns MD simulations. Surprisingly, the effective interaction of the $\beta 1$ -C term strand with the rest of the protein is more favorable in the uncleaved proenzyme (-122.7 vs. -114.6 kcal/mol). It is the intramolecular energy of the $\beta 1$ -Cterm strand that favors the cleaved complex (-22.9 for the cleaved complex vs. -6.3 for the uncleaved proenzyme). This difference is dominated by the electrostatic term, and especially the backbone electrostatic interactions (-111 vs. -93 kcal/mol). Thus, it is the intramolecular interactions in the cleaved strand that apparently increase the stability of the cleaved complex.

The uncleaved proenzyme experimentally was found to have a smaller portion of beta structure and a slightly larger portion of alpha structure.⁹ The loss of beta structure is adequately explained by the displacement of the $\beta 1$ strand from its position in the beta barrel. That the precursor binds ANS, albeit not as strongly as "molten globule" states, can also be explained by the exposure of a hydrophobic pocket (residues ILE168, PHE192, LEU230, LEU236) when the $\beta 1$ strand is displaced. It was suggested that the slight increase in alpha helix content might be due to α -helix formation by the $\beta 1$ strand.⁹ This was not observed in the present simulations. α -helix formation by the $\beta 1$ strand would require some further unfolding of the N domain beta barrel of αLP . Although this possibility cannot be precluded, it is possible that the slight apparent increase in α -helix content is a result of uncertainties in the deconvolution of the far-ultraviolet CD spectra.

500-K Simulations

Isolated αLP

The three simulations (ALP500A, B, C) lasted 400 ps (the first was subsequently extended to 800 ps). The extent of unfolding by 400 ps is similar in the three simulations: most native contacts are destroyed but the overall topology of the chain remains roughly native. (The native topology is gradually lost upon further simulation. The bRMSD between the 400-ps and 800-ps structures of ALP500A is 13.5 Å.) Unfolding occurs without a large expansion of the molecule (Fig. 4); the R_g at 400 ps in the three simulations is 17.7, 18.6, and 18.6 Å, respectively, compared with 15.01 Å in the crystal structure (an increase of 18–24%, smaller than that observed for CI2 with the same energy function at the same temperature¹⁰). The bRMSD after 400 ps of simulation at 500 K was 11.9 Å for the first (16.1 Å after 800 ps), 12.8 Å for the second, and 14.2 Å for the third simulation.

There are commonalities and differences in the residual structure among the three simulations, which can be best seen by examining the contact plots in Figure 5 (see Table I

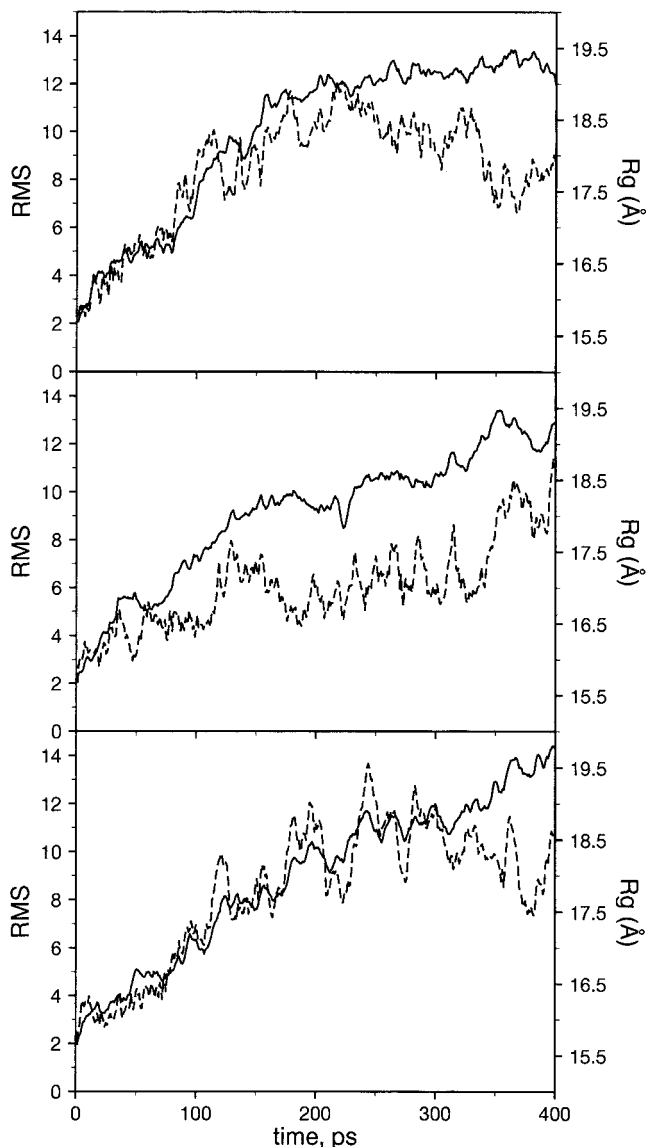


Fig. 4. Backbone root mean square deviation (bRMSD) and R_g in three independent simulations of isolated α LP at 500 K as a function of time.

for a key to Figs. 5–7). In all simulations, the β_4 – β_5 strands and partly the β_{12} – β_{13} strands remain paired after 400 ps. (After 800 ps in ALP500A, some β_{12} – β_{13} contacts persist, as well as one β_3 -loop1 and a β_{15} – β_{18} contact.) The helix remains partly folded until the end in the third simulation. Among the contacts that break early in the simulation are those between β_1 – β_9 , β_{12} – β_{16} and β_4 – β_{19} . The contacts between β_2 – β_6 , β_6 – β_7 are of relatively low stability too. The long hairpin contacts, β_{13} – β_{14} , are of intermediate stability. The contacts between this hairpin and the aspartate loop (β_8 – β_{14}) break relatively early. Also, the contact between His36 and Asp143 breaks relatively early (it also breaks in the 300-K simulation, see above). Overall, the N terminal domain tends to unfold faster than the C terminal domain.

Directly from the simulations it is not easy to say why unfolding of α LP is so slow. In terms of energy, the unfolding barrier need not be much larger than that of other proteins. According to TS theory, reduction in the unfolding rate by three orders of magnitude, corresponds to an increase in the unfolding barrier by a mere 4 kcal/mol. The unfolding barrier is created by loss of interactions that are not compensated by increase in conformational entropy. One possible source for this extra 4 kcal/mol in the unfolding barrier of α LP compared with that of other proteases like chymotrypsin, is the interaction of the β_{13} – β_{14} hairpin with the β_7 – β_8 hairpin. The effective interaction between these two elements (residues 60–62 with residues 127–129) in the native α LP structure is -11.4 kcal/mol (-8.5 van der Waals, -10.2 electrostatic, $+7.4$ solvation). The interaction between the corresponding residues in chymotrypsin (pdb code 5CHA) is -0.9 kcal/mol (-3.8 , -0.1 , $+3$). These numbers do not include any conformational entropy contribution. The conformational entropy contribution to the activation free energy is usually positive; therefore, the contribution of this interaction to the unfolding free energy will be smaller than 11.4 kcal/mol. The experimental value of 4 kcal/mol is not inconsistent with the above calculations.

Another possible source for the high unfolding barrier is the buried ion pair between Arg 102 and Asp 142. A cluster of ionic residues in subtilisin BPN' has been proposed to contribute significantly to the (un)folded barrier.¹¹ In α LP, the mutant R102H/G134S in the context of 3PRO was found to have a folding barrier ~ 3.5 kcal/mol lower than WT (unpublished data in www.msg.ucsf.edu/agard. No information is provided on the unfolding barrier.). To investigate the role of this ion pair, we created the mutant R102W/D142L and performed simulations on it. The mutant was stable at 300 K (1.95 Å bRMSD after 50 ps, compared with 1.65 Å for the WT after the same amount of simulation time). From the contact plots in Figure 5 no systematic difference in the unfolding speed at 500 K can be discerned. The contact between 102 and 142 (contact 87) disappears at 10, 142, and 400 ps in the WT simulations (average 184 ps) and at 355, 139, and 64 ps in the mutant simulations (average 186 ps), an insignificant difference. Thus, high temperature simulations cannot discern a difference in the unfolding rates or the unfolding mechanism between this mutant and the WT.

ALP-PRO complex

Three 400-ps simulations at 500 K were performed for the WT α LP-PRO complex (4PRO500A,B,C) and for the complex of α LP with a PRO mutant with the three residues at its C terminus truncated (3PRO500A,B,C). Comparison of the contact plots for unfolding of isolated α LP and of α LP complexed with 4PRO or 3PRO (Fig. 5) shows the following. The β_4 – β_5 and β_{12} – β_{13} contacts, which unfold late in isolated α LP, are somewhat less stable in 4PRO and 3PRO. For example, the β_4 – β_5 contacts are destroyed early in simulations 4PRO500B and 3PRO500B. In contrast, certain α LP contacts are more stable in the complex. These include the β_{13} – β_{14} long

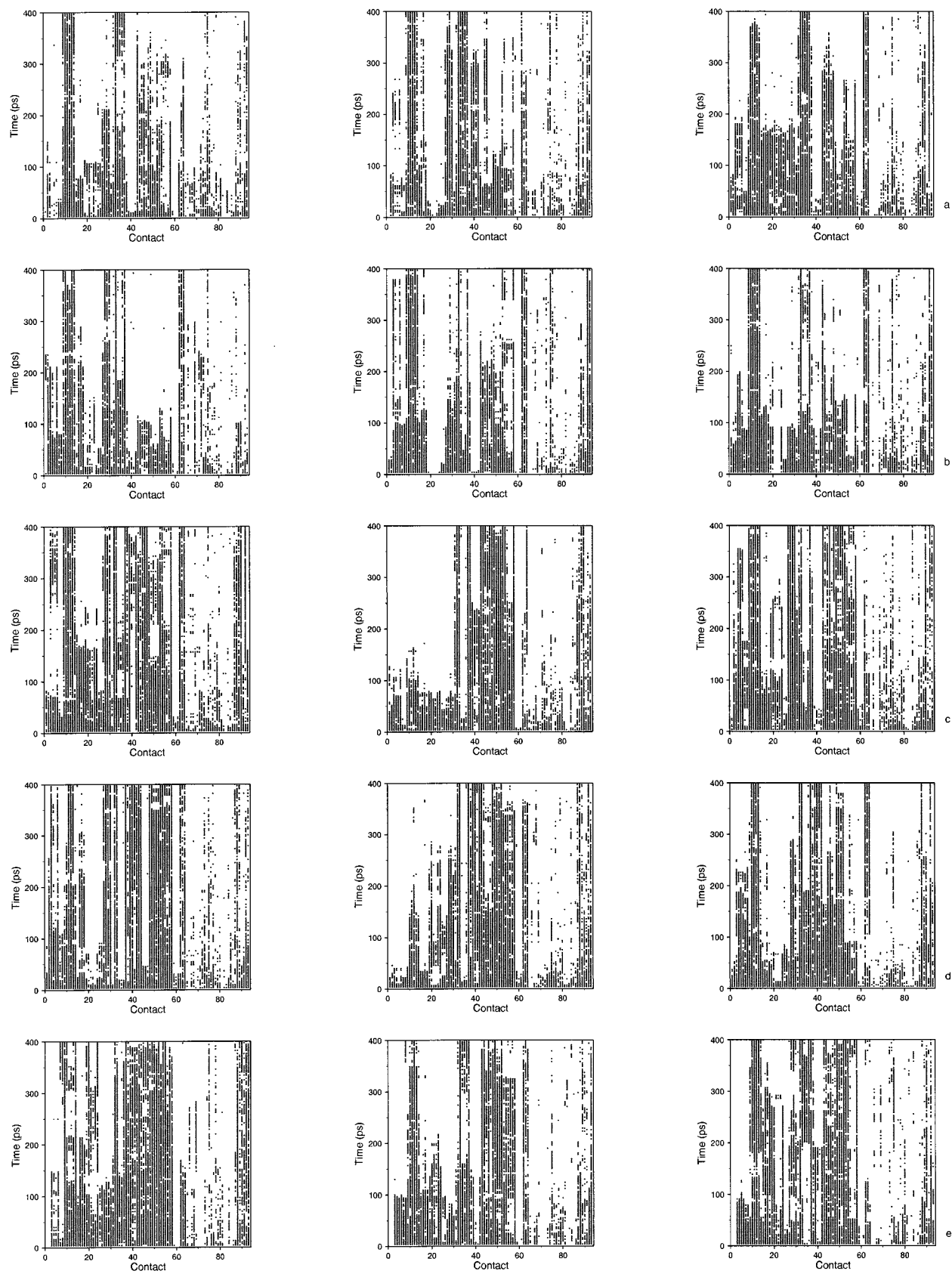


Fig. 5. Contact plots for α LP in the 500-K simulations: (a) α LP alone, (b) ion pair mutant of α LP alone, (c) α LP in 4PRO+ALP, (d) α LP in 3PRO+ALP, (e) α LP in PROALP.

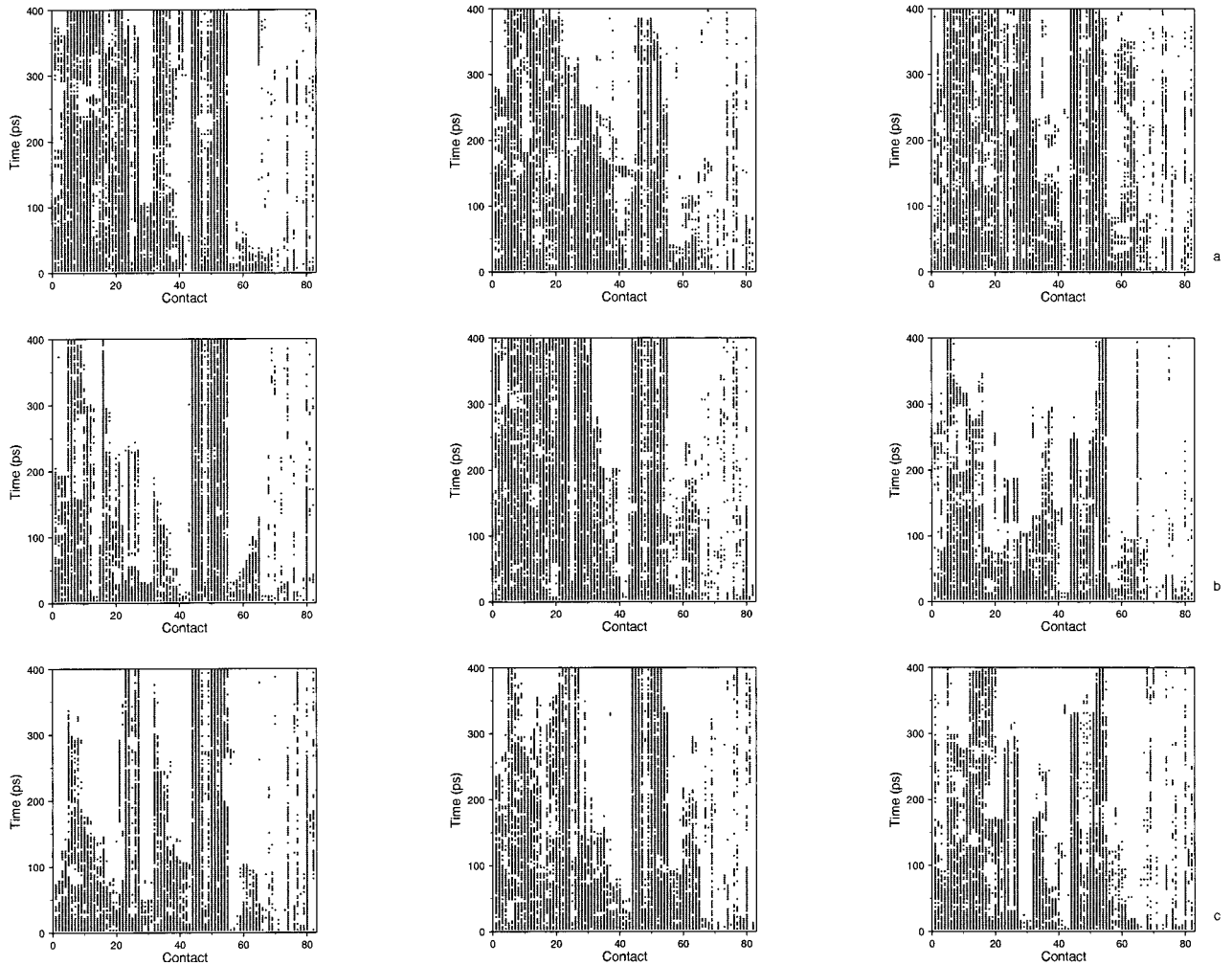


Fig. 6. Contact plots for PRO in the 500-K simulations: **(a)** in 4PRO+ALP, **(b)** in 3PRO+ALP, **(c)** in PROALP.

hairpin contacts, and the β_{14} – β_{18} , β_{17} – β_{18} , β_{12} – β_{16} contacts in the C terminal domain of α LP, which is expected because this domain interacts predominantly with PRO.

Among the PRO structural elements (Fig. 6), most stable during the unfolding simulations are α_2 and the β_4 – β_5 – β_6 beta sheet in the C terminal domain, which interacts with the C terminal domain of α LP. Least stable are α_4 and, less so, α_3 . α_1 and the beta sheet of the N terminal domain are of intermediate stability. Among the contacts between α LP and PRO, the hydrogen bonds between β_{13} of α LP and β_4 of PRO (the last three contacts in Fig. 7), which extend the β -sheet of α LP, are most stable and often remain until the end of the simulation. It is very difficult to detect any statistically significant (and rationalizable) global differences in the contact plots between the 3PRO and 4PRO simulations. Contacts between β_{12} – β_{16} tend to be more stable in 3PRO, but this is of doubtful significance.

One might have expected faster unfolding of α LP when complexed with the pro region, and faster unfolding when

it is complexed with 4PRO rather than 3PRO. This is not obvious in the simulations. Differences in room temperature unfolding rates may not be easily visible in the high temperatures used in these simulations. Instead, we can only observe the unfolding pathways and identify possible differences that could account for the slower unfolding of uncomplexed α LP and the role of the PRO region. In that regard, some interesting observations were made by visual examination of the unfolding trajectories. Most instructive is a frame around 200 ps in the 4PRO500A simulation, shown in Figure 8. In this frame, the C terminus of PRO is located between the long hairpin (β_{13} – β_{14}) and the aspartate loop (β_7 – β_8) of α LP. A similar configuration, although not quite as striking, was observed in 4PRO500C but not in 4PRO500B. In the 3PRO simulations, the C terminus is shorter and interacts with the C terminus to a much smaller extent. It is possible that this mediating interaction of the C terminus with the two hairpins may facilitate unfolding by lowering the barrier for their separation. This mechanism may not be observed in all simulations, be-

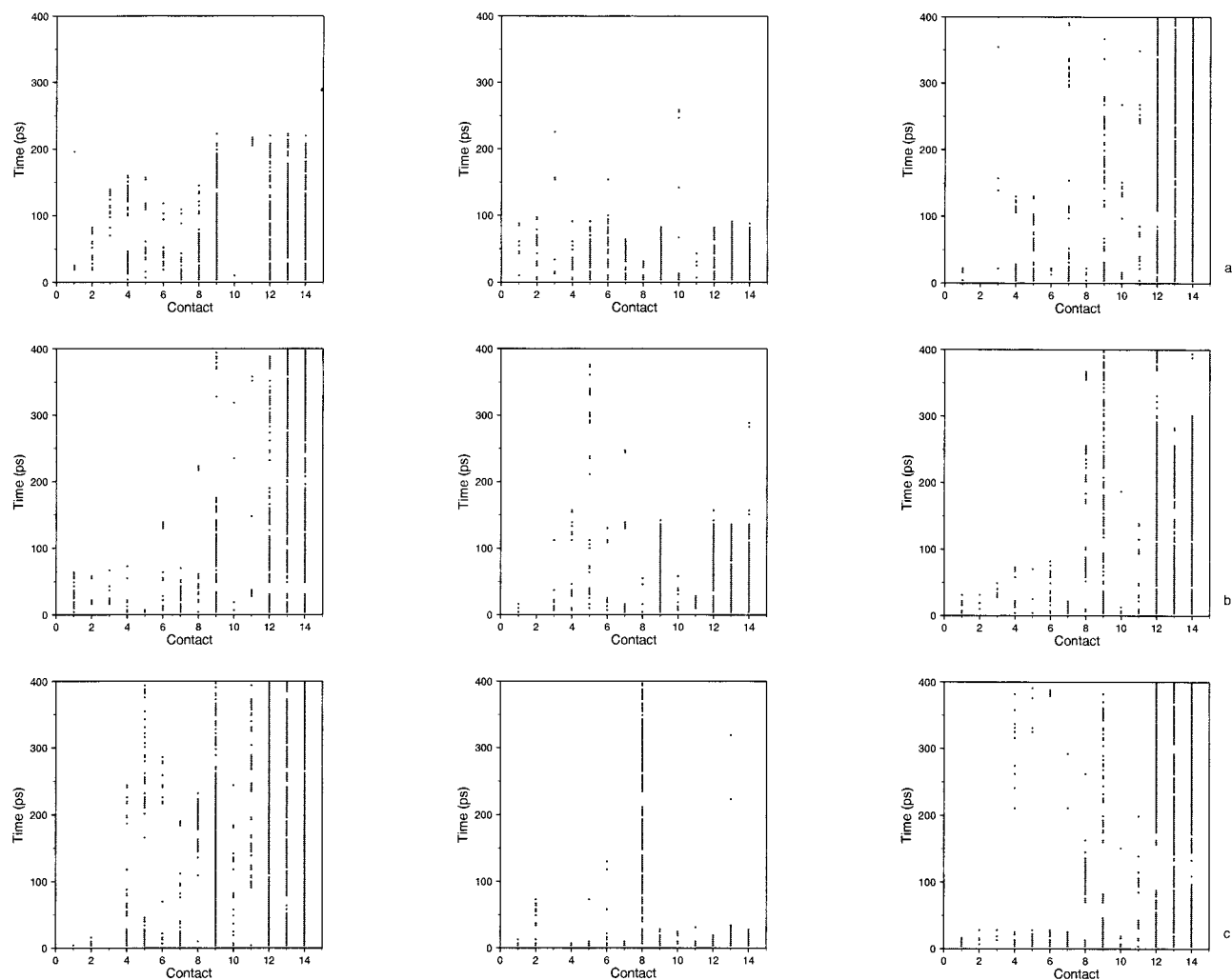


Fig. 7. α LP-PRO contact plots in the 500-K simulations: **(a)** in 4PRO+ALP, **(b)** in 3PRO+ALP, **(c)** in PROALP.

cause high temperature could make available alternative unfolding pathways that are not populated significantly at room temperature.

The strongest interaction of the PRO C terminus with α LP in this structure is a hydrogen bond between PRO:Gln164 CO and α LP:Asn 62 NH (distance between O and H 1.84 Å). Also, PRO:Leu 163 CO hydrogen bonds to α LP:Ala 124 NH (distance 2.64 Å), PRO:Thr 166 HG hydrogen bonds to α LP:Arg 64 CO (distance = 2.225 Å), and PRO:Gln 164 side-chain with α LP:Pro 60 CO (distance = 2.018 Å). The latter interaction may be relevant to the experimental observation that substitution of Gln 164 by Ala increases the folding activation free energy by 0.8 kcal/mol.¹²

The experimental data suggest that PRO must bind the TS more tightly by ~ 4 kcal/mol than the native state. If the role of the C terminus is to help the separation of the β 13– β 14 hairpin from the β 7– β 8 loop, one would expect it to bind more tightly to α LP in the TS than in the native state. The effective interaction of the PRO C terminus

(residues 160–166) with α LP in the minimized native state is -56 kcal/mol (-59 , -41 , $+44$). If we take the 200-ps frame from the first 500-K simulation, perform 100-ps MD at room temperature, and then minimize for 300 steps, the same effective interaction is -45 (-37 , -37 , $+30$). Thus, the effective interaction is not as strong as in the native state. However, these values do not include conformational entropy contributions. The conformational entropy cost of the interaction may be higher in the N state than in the putative TS. For example, in the putative TS, there can be many conformations with an effective interaction of -45 kcal/mol. In that case, the free energy of the interaction could be lower in the TS, despite that the effective energy is higher. That the conformational entropy cost of the PRO C term interaction with native α LP is high is suggested by the fact that truncation of the C terminus does not lead to a large reduction in the affinity of PRO for α LP,¹² which means that the favorable effective energy of the interaction is counterbalanced by an unfavorable protein entropy contribution.

TABLE I. Native Contacts in Figures 5–7[†]

ALP			
Contact		Contact	
1	$\beta 1$ - $\beta 9$ h. bonds	65–69	catalytic triad
2	$\beta 1$ - $\beta 10$ h. bonds	70–72	$\beta 1$ - $\beta 6,9,10$ tertiary
3–6	$\beta 2$ - $\beta 3$ h. bonds	73–74	$\beta 2$ - $\beta 3$ tertiary
7–8	$\beta 2$ - $\beta 6$ h. bonds	75–77	$\beta 4$ - $\beta 5,16,19$ tertiary
9–14	$\beta 4$ - $\beta 5$ h. bonds	78	$\beta 5$ - $\beta 8$ tertiary
15–18	$\beta 6$ - $\beta 7$ h. bonds	79–81	L1- $\beta 7,8$ tertiary
19–24	$\beta 7$ - $\beta 8$ h. bonds	82	$\beta 8$ - $\alpha 1$ tertiary
25–26	$\beta 8$ - $\beta 14$ h. bonds	83–86	$\beta 11$ - $\beta 13,18,19$ tertiary
27–30	$\beta 10$ - $\beta 11$ h. bonds	87	102-142 tertiary
31–38	$\beta 12$ - $\beta 13$ h. bonds	88	$\beta 12$ - $\beta 16$ tertiary
39–42	$\beta 12$ - $\beta 16$ h. bonds	89–90	$\beta 13$ - $\beta 14$ tertiary
43–48	$\beta 13$ - $\beta 14$ h. bonds	91	$\beta 14$ - $\beta 17$ tertiary
49–52	$\beta 14$ - $\beta 18$ h. bonds	92–93	$\beta 15$ - $\beta 18$ tertiary
53–58	$\beta 17$ - $\beta 18$ h. bonds		
59–61	$\beta 19$ - $\beta 4$ h. bonds		
62–64	$\alpha 1$ h. bonds		
PRO			
Contact		Contact	
1–4	$\alpha 1$ h. bonds	66–67	$\alpha 2$ - $\beta 1$ tertiary
5–22	$\alpha 2$ h. bonds	68–71	$\alpha 2$ - $\beta 2$ tertiary
23–27	$\beta 1$ - $\beta 2$ h. bonds	72	$\alpha 2$ -L1 tertiary
28–31	$\beta 2$ - $\beta 3$ h. bonds	73	$\alpha 2$ - $\beta 3$ tertiary
32–43	$\alpha 3$ h. bonds	74–76	$\beta 1$ - $\beta 2,3,4$ tertiary
44–51	$\beta 4$ - $\beta 5$ h. bonds	77	$\beta 3$ - $\beta 4$ tertiary
52–55	$\beta 5$ - $\beta 6$ h. bonds	78–79	$\alpha 3$ - $\beta 4$ tertiary
56–65	$\alpha 4$ h. bonds	80–82	$\beta 4$ - $\beta 5, \alpha 4, Ct$ tertiary
ALP-PRO			
Contact (ALP, PRO)		Contact (ALP, PRO)	
1	$\beta 10$ - $\alpha 1$	9	$\beta 13$ - $\beta 4$
2	$\beta 16$ - $\alpha 1$	10	L3- $\beta 6$
3–5	$\beta 13$ - $\alpha 1$	11	$\beta 13$ -Cterm
6–7	$\beta 12$ - $\alpha 1,2$	12–14	$\beta 14$ - $\beta 4$ h. bonds
8	$\beta 14$ -L2		

[†]A detailed list is available from the authors upon request.

The results for the uncleaved proenzyme are very similar to those for the cleaved complex, other than a few contacts involving the $\beta 1$ strand, which is displaced in the proenzyme. In these simulations, the C terminus of PRO is not observed to interact closely with the long hairpin and the aspartate loop. The mechanism by which the covalently bonded PRO region facilitates folding and unfolding may be different from that in the cleaved complex. It may facilitate folding by “tethering” the N terminal domain of αLP close to the C terminal domain and unfolding by destabilizing the N state (see above), rather than mediating the separation of the two hairpins.

The interface between αLP and PRO is not very tight and is populated by several buried water molecules. It has been proposed that, in the TS, the bound water molecules are expelled, leading to a tighter interaction of PRO with αLP .⁷ To test this idea, we examined the unfolding trajectories of 4PRO+ALP, 3PRO+ALP, and PROALP for evidence of more tight binding between the C terminal domains of PRO and αLP and in particular αLP $\beta 13$ and

PRO $\beta 4$. Because we do not include any explicit water molecules in the simulations, the kinetics of this process is not realistic, i.e., formation of additional hydrogen bonds between αLP and PRO is more facile than in reality, because no buried water molecules need to be expelled. Nevertheless, we can obtain valid information on the thermodynamics of the process. We monitored the distances between three potentially hydrogen bonding pairs: $\alpha LP:N118O$ -PRO:V126N, $\alpha LP:N118N$ -PRO:V126O, $\alpha LP:K117N$ -PRO:V126O. It was found that hydrogen bonds were occasionally formed between the above pairs. At 12 ps into the 4PRO500A simulation, hydrogen bonds form between the first and third of the above pairs. The interaction between residues 114–121 in αLP and 123–129 in PRO is -31.5 kcal/mol at 12 ps, compared with -12.5 kcal/mol at 1 ps. This stronger interaction between $\alpha LP:\beta 13$ and PRO: $\beta 4$ is accompanied by a weaker interaction between $\beta 13$ and $\beta 14$ of αLP : the interaction between αLP residues 114–121 and 128–135 is -27.5 kcal/mol at 12 ps vs. -36.6 kcal/mol at 1 ps. Thus, the improvement of



Fig. 8. Snapshot at 200 ps of the first 500-K simulation of 4PRO+ALP showing the PRO C terminus between two α LP hairpins.

interactions between α LP and PRO may lead to deterioration of interactions within α LP and, thereby, destabilization of α LP and facilitation of unfolding.

DISCUSSION

The key questions in α LP folding are (1) why does α LP fold and unfold much more slowly than other proteins, (2) how does the pro region facilitate folding and unfolding. A related question that may help to answer the first two is (3) what is the TS for folding/unfolding. Because the native state of mature α LP is not thermodynamically favored,⁴ slow folding should not come as a surprise. The more “natural” question is why is unfolding so slow despite being thermodynamically favored.

The main findings of the present work, which are largely consistent with ideas proposed based on observations of the available crystal structures,⁷ are as follows. (1) The interaction of PRO: β 4 with α LP: β 13 is one of the most stable in the high temperature simulations. This finding makes it a likely “nucleation” site for folding, a structural feature whose presence is likely in the TS for folding. This interaction was found to occasionally become stronger during the simulation due to additional hydrogen bonds between β 4 and β 13. Because this stronger interaction correlated with a weaker interaction between α LP strands β 13 and β 14, it is possible that this interaction helps unravel α LP. This idea could be experimentally tested by mutations in α LP or PRO that expel the buried water molecules at their interface and prevent closer approach between the two beta strands. (2) During the simulation of WT PRO-ALP the PRO C terminus was occasionally observed to intervene between the long hairpin (β 13– β 14)

and the aspartate loop (β 7– β 8) of α LP. This finding suggests that the role of the C terminus may be to lower the barrier for separation of these two hairpins. This is probably more important than the tightening of the interface, because truncation of four residues from the PRO C terminus leads to abolishment of unfolding catalysis.¹² If the role of the PRO C terminus is to “pry apart” the two α LP hairpins, one might ask whether this role could be fulfilled by small peptides of the same sequence (or of any sequence, because this effect does not seem to be very sequence specific¹²). One issue here is the entropic cost of the interaction. A “loose” peptide at low concentration has many options for where to bind and pays a high entropy cost upon binding to a specific region of the enzyme. The same peptide attached to PRO bound to α LP has far fewer options (its “effective concentration” is high) and loses less entropy upon binding to the native or to the transition state. Only at very high peptide concentrations might any unfolding catalysis be observed. Such an experiment might be worth considering. (3) Mutation of the buried ion pair to a pair of hydrophobic residues did not lead to any observable systematic change in the unfolding mechanism or speed. This finding could mean that the buried ion pair does not contribute significantly to the unfolding barrier. It is possible, though, that the high temperature simulations cannot discern subtle differences in room temperature unfolding rates. (Unfortunately, we cannot calculate room temperature unfolding rates by such simulations.) We must also keep in mind the approximations involved in the treatment of ionic residues in EEF1,⁸ which may affect the barrier for breaking buried ion pair interactions.

At present, it is not possible in atomistic simulations to prove that a particular structure or ensemble of structures is the TS for folding. It is reasonable to assume that the TS is an ensemble of conformations where a combination of local distortions, perhaps different distortions in each member of the ensemble, are sufficient to lead the protein with equal probability in the direction of folding or unfolding. Candidates for such local distortions in α LP are: the separation of β 13–14 from β 7– β 8, the unraveling of β 1, unraveling of β 14 from β 13, etc. A structure where only one of these distortions takes place is probably not a member of the TS ensemble because each distortion alone is not sufficient to push the protein in the direction of unfolding. Facilitation of any of these local distortions would catalyze unfolding. The short loops and tight packing of α LP¹³ could make such distortions energetically more costly than in other proteins.

Other findings from the simulations, some of which can be tested experimentally, include the following. (1) Dynamic fluctuations are smaller in α LP than in the related serine protease α -chymotrypsin. This is not restricted to external loops but holds for buried atoms as well. (2) The main source of instability in the uncleaved proenzyme is the intramolecular energy of the Cterm- β 1 strand. Thus, a mutated enzyme with residues inserted between Cterm and β 1 might be more stable than the WT proenzyme. One caveat here is that insertion of residues would lead to an increase in the conformational entropy cost upon folding

that might counterbalance the gain from relief of bad intramolecular interactions and also to a possible loss of favorable interactions with the rest of the protein. (3) The structure of PRO alone in solution is more collapsed than in the crystal structure.

It should be noted that the data on α LP require only minor adjustments to the "thermodynamic hypothesis" of protein folding and do not invalidate theoretical efforts to predict protein structures by minimizing an effective energy function. The transient native state of the proenzyme seems to be a thermodynamic free energy minimum. Furthermore, that the native mature α LP is a metastable state does not mean that it does not correspond to the lowest valley in its effective energy hypersurface. It simply means that this well is not deep enough to make native α LP thermodynamically stable; i.e., it is not deep enough to counterbalance the unfavorable conformational entropy loss upon folding. Still, the native conformation probably has effective energy lower than any other *single* conformation. The intermediate state, to which α LP folds spontaneously under folding conditions, may well be a compact denatured state, like that observed for mutants of staphylococcal nuclease.¹⁴

Pro region-dependent folding behavior is observed in several other proteases, of which most extensively studied is subtilisin BPN'.¹⁵ Simulations of subtilisin and its complex with its pro region are currently under way to determine possible similarities and differences with α LP.

ACKNOWLEDGMENTS

We thank Prof. D. Agard for providing the unpublished coordinates of PRO and the PRO-ALP complex. C.C.N.Y. received an RCMI grant from NIH, and T.L. received an ACS PRF award. The following freely available programs were used, thankfully, for analysis, visualization, and plotting: ASGL (A. Šali, guitar.rockefeller.edu/asgl/asgl.html), MOLMOL (K. Wüthrich, www.mol.biol.ethz.ch/wuthrich/software/molmol), VMD (K. Schulten, www.ks.uiuc.edu/Research/vmd), MOLSCRIPT (P. Kraulis, www.avatar.se/molscript), and RASMOL (R. Sayle, www.umass.edu/microbio/rasmol).

REFERENCES

- Silen JL, Agard DA. The α -lytic protease pro-region does not require a physical linkage to activate the protease domain in vivo. *Nature* 1989;341:462–464.
- Baker D, Sohl JL, Agard DA. A protein folding reaction under kinetic control. *Nature* 1992;356:263–265.
- Baker D, Agard DA. Kinetics versus thermodynamics in protein folding. *Biochemistry* 1994;33:7505–7509.
- Sohl JL, Jaswal SS, Agard DA. Unfolded conformations of α -lytic protease are more stable than its native state. *Nature* 1998;395:817–819.
- Brayer GD, Delbaere LTJ, James MNG. Molecular structure of the α -lytic protease from *Myxobacter 495* at 2.8 Å resolution. *J Mol Biol* 1979;131:743–775.
- Fujinaga M, Delbaere LTJ, Brayer GD, James MNG. Refined structure of α -lytic protease at 1.7 Å resolution. *J Mol Biol* 1985;184:479–502.
- Sauter NK, Mau T, Rader SD, Agard DA. Structure of α -lytic protease complexed with its pro region. *Nat Struct Biol* 1998;5:945–950.
- Lazaridis T, Karplus M. Effective energy function for proteins in solution. *Proteins* 1999;35:133–152.
- Anderson DE, Peters RJ, Wilk B, Agard A. α -lytic protease precursor: characterization of a structured folding intermediate. *Biochemistry* 1999;38:4728–4735.
- Lazaridis T, Karplus M. "New view" of protein folding reconciled with the old through multiple unfolding simulations. *Science* 1997;278:1928–1931.
- Bryan P, Alexander P, Strausberg S, Schwarz F, Lun W, Gilliland G, Gallagher DT. Energetics of folding subtilisin BPN'. *Biochemistry* 1992;31:4937–4945.
- Peters RJ, Shiau AK, Sohl JL, Anderson DE, Tang G, Silen JL, Agard DA. Pro region C-terminus: protease active site interactions are critical in catalyzing the folding of α -lytic protease. *Biochemistry* 1998;37:12058–12067.
- Baker D. Metastable states and folding free energy barriers. *Nat Struct Biol* 1998;5:1021–1024.
- Shortle D. The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J* 1996;10:27–34.
- Ruan B, Hoskins J, Bryan PN. Rapid folding of Ca-free subtilisin by a stabilized pro-domain mutant. *Biochemistry* 1999;38:8562–8571.
- Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J Comput Chem* 1983;4:187–217.
- Neria E, Fischer S, Karplus M. Simulation of activation free energies in molecular systems. *J Chem Phys* 1996;105:1902–1921.
- Lazaridis T, Karplus M. Discrimination of the native from misfolded protein models with an energy function including implicit solvation. *J Mol Biol* 1999;288:477–487.
- Lazaridis T, Karplus M. Heat capacity and compactness of denatured proteins. *Biophys Chem* 1999;78:207–217.
- Paci E, Karplus M. Forced unfolding of fibronectin type 3 modules: an analysis by biased MD simulations. *J Mol Biol* 1999;288:441–449.
- Dinner AR, Lazaridis T, Karplus M. Understanding β -hairpin folding. *Proc Natl Acad Sci USA* 1999;96:9068–9073.
- Brunger AT, Karplus M. Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison. *Proteins* 1988;4:148–156.
- van Gunsteren WF, Berendsen HJC. Algorithms for macromolecular dynamics and constraint dynamics. *Mol Phys* 1977;34:1311–1327.
- Brooks CL III, Karplus M. Deformable stochastic boundaries in molecular dynamics. *J Chem Phys* 1983;79:6312–6325.
- Lazaridis T, Lee I, Karplus M. Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin. *Protein Sci* 1997;6:2589–605.