Structure of α-lytic protease complexed with its pro region

While the majority of proteins fold rapidly and spontaneously to their native states, the extracellular bacterial protease α-lytic protease (αLP) has a $t_{1/2}$ for folding of ~2,000 years, corresponding to a folding barrier of 30 kcal mol^{-1}. αLP is synthesized as a pro-enzyme where its pro region (Pro) acts as a foldase to stabilize the transition state for the folding reaction. Pro also functions as a potent folding catalyst when supplied as a separate polypeptide chain, accelerating the rate of αLP folding by a factor of 3×10^9. In the absence of Pro, αLP folds only partially to a stable molten globule-like intermediate state. Addition of Pro to this intermediate leads to rapid formation of native αLP. Here we report the crystal structures of Pro and of the non-covalent inhibitory complex between Pro and native αLP. The C-shaped Pro surrounds the C-terminal β-barrel domain of the folded protease, forming a large complementary interface. Regions of extensive hydration in the interface explain how Pro binds tightly to the native state, yet even more tightly to the folding transition state. Based on structural and functional data we propose that a specific structural element in αLP is largely responsible for the folding barrier and suggest how Pro can overcome this barrier.

Most extracellular bacterial proteases are synthesized as pro-enzymes, and where examined, the pro regions have been shown to be necessary for production of active protease[1–3]. Similarly, the pro regions of numerous eukaryotic proteases have been shown to be required for correct folding[4,5]. Pro region-assisted folding contrasts with the maturation of mammalian digestive enzymes such as chymotrypsin, trypsin, and elastase, in which a short N-terminal zymogen peptide prevents premature enzyme activation, but does not determine the protein's ability to fold[6]. Also, pro regions facilitate protein folding through a mechanism fundamentally distinct from that of molecular chaperones. Whereas chaperones bind to unfolded or partially-folded states to block off-pathway aggregation reactions, pro regions promote on-pathway folding by interacting directly with folding transition states. αLP provides one of the best-studied and most dramatic examples of pro-mediated folding (Fig. 1): its pro region lowers the free energy of the folding transition state by 18.2 kcal mol^{-1} (refs 6–10). Unlike molecular chaperones, the αLP pro region also binds tightly to the native state and is one of the most potent αLP inhibitors known (K_i = 0.3 nM)[2,9,11]. This tight binding is, in fact, required for efficient and complete folding. In the absence of Pro, the native state of αLP is less stable than the molten globule folding intermediate by 4 kcal mol^{-1} (Fig. 1)[10]. Formation of the tightly bound Pro–αLP complex shifts the thermodynamic equilibrium in favor of folded αLP. Active protease is subsequently released from this complex by proteolytic degradation of Pro by exogenous proteases. Although thermodynamically unstable, the protease remains in its native state because it is kinetically trapped by a high barrier to unfolding[10].

The structural studies presented here address two fundamental questions: what is the structural basis for the large αLP folding barrier, and how does Pro stabilize the folding transition state to catalyze protein folding. The very tight binding of Pro to both the native state and the folding transition state indicates that the transition state complex and the native state complex must have structural elements in common. We can therefore infer features of the transition state complex (Pro•Int‡) from interactions observed to stabilize the native state complex (Pro•Nat; Fig. 1).

As a first step towards understanding how Pro acts as a foldase, we determined the structure of unbound Pro at 3.0 Å by multiwavelength anomalous dispersion[12] (Table 1), with phase improvement by solvent flattening and three-fold non-crystallographic symmetry (NCS) averaging. Pro is a 166 residue α/β protein mostly made up of two globular domains (N-domain, C-domain) connected by a nearly rigid hinge to form a novel C-shaped molecule (Fig. 2*a*). The concave surface of the 'C' is lined by a pair of three-stranded β-sheets, one from each domain, while the convex surface consists of α-helices packed against these β-sheets. The N-terminal helix (residues 5–16) adopts different conformations in two of the three NCS-related molecules in the unit cell and is absent in the third. No density is seen for the C-terminal tail (C-tail; residues 159–166), which is presumed to be disordered in unbound Pro.

Two different native state complexes were examined, one with wild-type Pro, and one with a Pro mutant (Pro-3) in which the absence of the last three residues results in a significant reduction in Pro's ability to fold the αLP intermediate (k_{cat} decreased by ~300-fold)[9]. Structures of the wild-type Pro com-

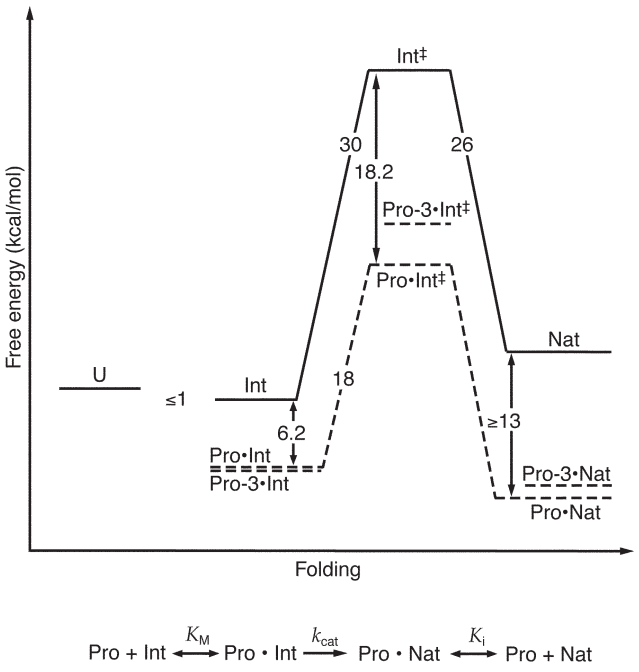

Fig. 1 Free energy diagram for the *in vitro* folding pathway of αLP, adapted from refs 9,10. Unfolded protein (U) spontaneously folds to a molten globule-like intermediate (Int), which folds at an extremely slow rate to the native state (Nat) through a high energy transition state (Int‡). Pro region (Pro) provides a catalyzed folding pathway which proceeds through a Michaelis complex (Pro•Int), a lower energy transition state (Pro•Int‡) and finally to the complex of inhibited native protease (Pro•Nat). Values shown correspond to the Michaelis constant (K_M = 9 μM) and catalytic rate (k_{cat} = 1.9 min^{-1}) measured at 4 °C and the inhibition constant (K_i = 0.3 nM) measured at 25 °C. Values for the Pro-3 mutant are also shown. Folding and unfolding barrier heights were calculated using transition state theory.

letters

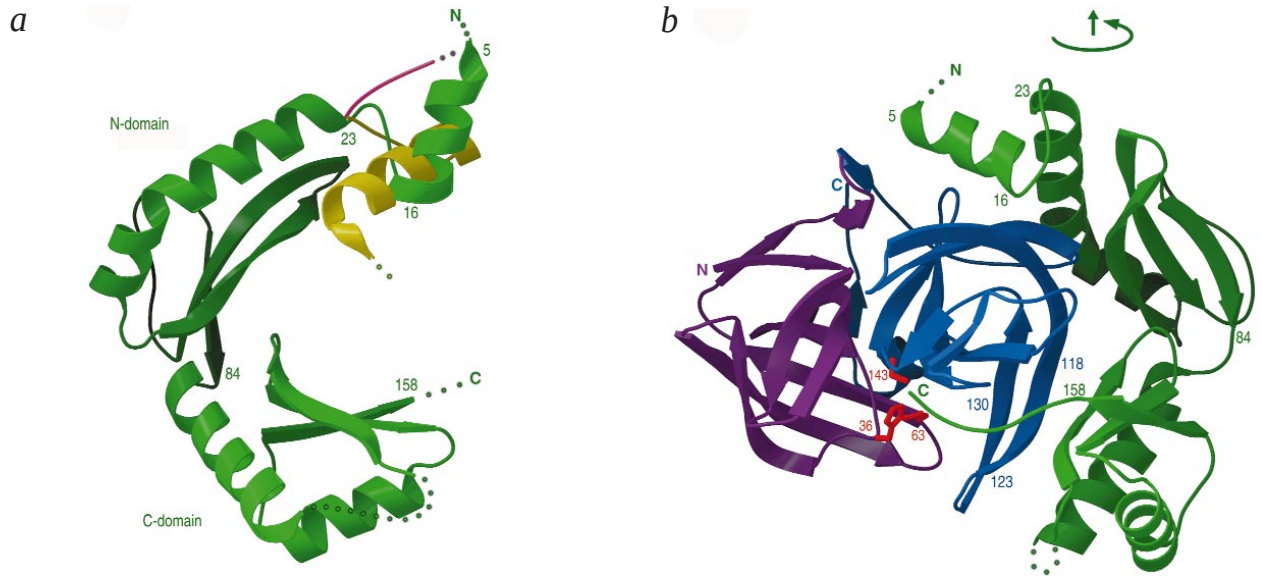

Fig. 2 ***a***, Ribbon diagram of unbound Pro. Disordered residues are indicated by dots. The N-terminal helix (residues 5–16) is positioned differently in two of the three non-crystallographic symmetry-related molecules in the unit cell (green and yellow), and is absent in the third molecule (pink). The other residues (23–158) overlap in all three molecules and are shown in green. ***b***, Complex between native α-lytic protease (αLP) and wild-type Pro. αLP's N-terminal domain (residues 1–84) and C-terminal domain (residues 85–198) are shown in magenta and blue respectively, and side chains of the catalytic triad (His 36, Asp 63, and Ser 143) are shown in red. Bound Pro is illustrated in green, with the C-tail (residues 160–166) inserted into αLP's active site. The disordered loop in Pro (residues 105–112) is likely to be the initial site of proteolytic attack required to release Pro from the Pro–αLP complex, producing active αLP. The unbound molecule shown in (*a*) can be aligned with the bound molecule in (*b*) by a 240° rotation about the vertical axis as indicated. Figs 2 and 3 were generated using the programs BobScript and Raster3D.

plex (Fig. 2*b*) and the Pro-3 complex were solved by molecular replacement at 2.4 Å and 1.8 Å respectively (Table 1). The structures of the two Pro–αLP complexes are very similar (Cα r.m.s.d. = 0.35 Å). An extensive interface (4,117 Å^2 total buried accessible surface area for the wild-type Pro complex) accounts for the tight binding that is critical for efficient folding. Pro surrounds the αLP C-terminal domain, with its C-tail in the protease active site and its N-terminus contacting the opposite side of αLP. Pro makes contacts almost exclusively with the αLP C-terminal domain, implying that the αLP C-terminal domain is the focus of Pro's folding activity. In particular, the three-stranded β-sheet in the Pro C-domain pairs with a long β-hairpin (residues 118–130) in the αLP C-terminal domain, forming a continuous five-stranded β-sheet. The structure of αLP in the complex is virtually identical to that of the free αLP (Cα r.m.s.d. = 0.29 Å). Similarly, the structure of Pro's globular domains is not significantly altered upon complex formation. The only notable changes in Pro are seen in the flexible N-terminal helix and the C-tail, both of which adopt specific conformations as they bind αLP.

The Pro C-tail is inserted into the protease active site in a substrate-like manner (Fig. 3*a*), making the same hydrogen bonds and van der Waals contacts as previously seen with both covalently bound transition state analogs[13] and with a non-covalently bound αLP–peptide complex (ref. 36; R. Bone, unpublished data). As predicted from biochemical studies[11], Pro inhibits αLP by directly blocking the active site. Given the excellent complementarity of the Pro C-tail and the αLP binding pocket, deletion of the last three residues of Pro (Pro-3) results in a remarkably small effect on binding (five-fold increase in K_i)[9]. Our Pro-3 complex structure shows that the Pro–αLP interface remains unchanged, indicating that the small effect on K_i is not due to compensating structural rearrangements in other parts of the interface. In contrast to its modest contribution to native state binding, the C-tail has a profound effect on the rate of folding, with the Pro-3 deletion causing a ~300-fold reduction in k_{cat} (Fig. 1). Deletion of the last four residues of Pro has an even greater effect on k_{cat} (>10^6-fold reduction)[9], demonstrating that the energetic gain from binding the C-tail is used predominantly to stabilize the transition state, and not the native state. This optimal use of binding energy to promote catalysis by increasing k_{cat} is a hallmark of highly evolved enzymes[14].

The crystal structure of the complex provides clues as to how Pro could preferentially stabilize the folding transition state over the native state. Likely sites for improved Pro–αLP interactions in the folding transition state can be found in regions of the interface that show poor complementarity in the native state complex. The molecular interface contains a significant gap at the juncture of the Pro N- and C-domains which is filled with eight water molecules (Fig. 3*b*). Adjacent to this, three ordered water molecules make hydrogen bonds that bridge the β-sheet interface between the Pro C-domain and the αLP 118–130 β-hairpin (Fig. 3*b*). Ordered water molecules have been observed to function as adapters in protein–DNA complexes[15] and the quaternary interfaces of allosteric enzymes[16,17], where the waters allow two surfaces to interact in two different conformational states. The higher affinity states tend to contain fewer waters, reflecting the exclusion of water due to improved surface complementarity. We propose that in the transition state of the Pro-catalyzed folding reaction, the C-terminal domain of αLP is distorted into a conformation more complementary to Pro, expelling the bound waters. Enhanced stabilization is thus achieved by increasing the surface area of interaction and reducing the entropic cost by freeing bound waters.

While Pro binds tightly and preferentially to the folding transition state, it binds only weakly to the molten globule intermedi-

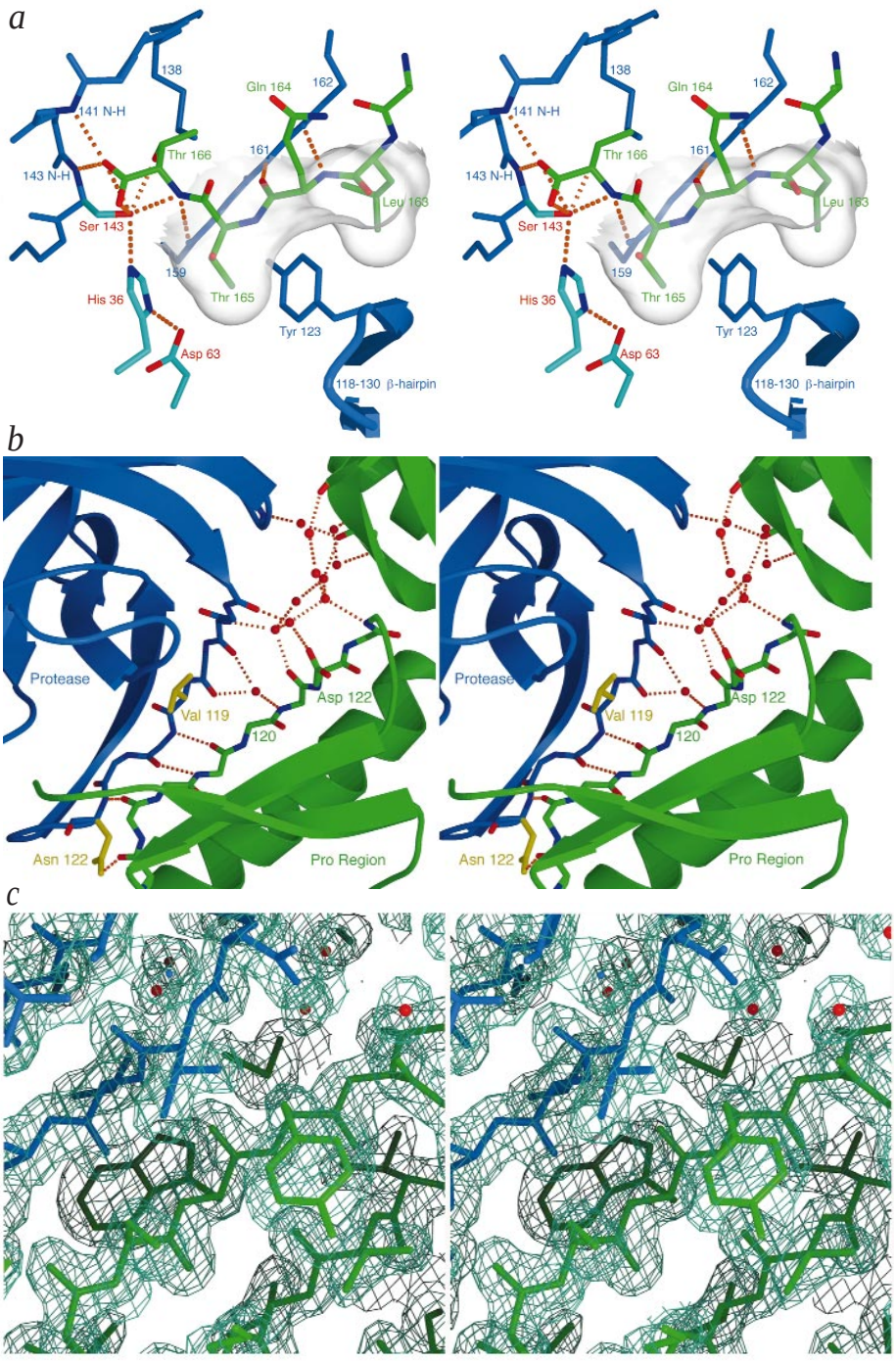

Fig. 3 ***a***, Detailed view of the C-tail of wild-type Pro (green) inserted into the active site of αLP (blue). All hydrogen bonds to the catalytic triad (His 36, Asp 63, and Ser 143) and oxyanion hole (141 and 143 amides) are shown in orange. Pro residues 164–166 form an antiparallel β-sheet with αLP residues 159–161 in the same manner as peptide substrates[13]. The Tyr 123 side chain at the end of αLP's 118–130 hairpin fits snugly into the pocket formed by Pro residues 163–165 (surface shown). ***b***, Close-up view of the highly hydrated interface between Pro-3 (green) and the C-domain of αLP (blue). Ordered water molecules are shown as red spheres and hydrogen bonds are shown in orange. A prominent feature of this interface is the five-stranded β-sheet formed by two strands of αLP and three strands of Pro. Two side chains of αLP implicated in forming the initial Michaelis complex between Pro and the molten globule folding intermediate (Val 119 and Asn 122) are indicated in yellow. ***c***, Stereo view of the 1σ contour of the $2F_o - F_c$ electron density map showing the β-sheet interface between Pro-3 (green) and αLP (blue). Water molecules are indicated in red.

ate, the substrate of the folding reaction (K_M = 9 μM)[9]. The C-tail does not contribute significantly to this step (formation of the Michaelis complex), since the Pro-3 deletion has no effect on K_M[9] (Fig. 1). In an effort to discover potential regions of Pro–αLP interaction, we had previously chosen to mutate αLP residues Val 119 to Ile and Asn 122 to Lys (Fig. 3*b*). These mutations, located in the β-sheet interface described above, have a small but significant effect on K_M[9], indicating that interactions between the αLP 118–130 β-hairpin and the Pro C-domain β-sheet occur in the earliest stages of catalyzed folding.

It is notable that this αLP hairpin interacts not only with the Pro C-domain β-sheet, but also with the Pro C-tail. Pro C-tail residues 163–165 form a hydrophobic binding pocket for αLP Tyr 123, stabilizing the end of the hairpin (Fig. 3*a*). The fact that the hairpin contains residues known to be important for formation of the Michaelis complex and also interacts with Pro residues critical for stabilization of the transition state suggests a pivotal role for the hairpin in folding.

Based on the preceding observations, we propose the following model for Pro-catalyzed folding of αLP (Fig. 4). In the folding intermediate we envision that the hairpin is solvated and making only minimal contacts with the rest of the αLP C-terminal domain. Also, the intermediate's large hydrodynamic radius[8] suggests that the N- and C-terminal domains of the protease are separated. Both these factors contribute to the intermediate's molten globule-like character. The first step in folding is the binding of the Pro C-domain to the αLP hairpin to form the five-stranded β-sheet. Productive folding then requires that the Pro C-tail bind the nascent substrate binding site. This interaction forces the hairpin to pack against the αLP C-terminal domain, with the C-tail providing a docking site and the rest of the Pro surface acting as a template to guide the hairpin into the proper position to stabilize the αLP C-terminal domain. Once αLP's C-terminal domain has folded, the N-terminal domain docks into place, establishing αLP's tightly packed core and completing the catalytic triad. For the *in vivo* case where Pro is covalently attached to the N-terminus of αLP, formation of the catalytic triad allows the intramolecular processing of the Pro-

αLP peptide bond. This leaves the C-tail of Pro in the active site as we observe, and frees the N-terminus of αLP to rearrange into its native position 24 Å away (Fig. 2*b*).

In this model, the 118–130 hairpin contributes significantly to the large folding barrier, posing a folding defect that is overcome by Pro. This idea is supported by a comparison of the sequence of αLP with those of other members of the chymotrypsin superfamily. The presence of an αLP-like hairpin is correlated with the presence of a pro region: the hairpin is highly conserved in all 13 known bacterial homologs that have pro regions (data not shown); however the hairpin is replaced by different structures or is absent altogether in bacterial, viral and mammalian family members that do not require pro regions for folding. Although the picornaviral 3C cysteine protease does contain a hairpin[18], this structural element is not homologous in sequence or structure to αLP's hairpin. Currently, the role of its short pro region in folding is unknown. The critical role of αLP's hairpin in folding is further underscored by the high degree of sequence conservation in the β-sheet segment of Pro (residues 116–133) with which it interacts. The proper positioning of a β-hairpin by a β-sheet could indeed be a general mechanism in the folding of β-structures. By contrast, the pro region-dependent folding of subtilisin BPN', an α/β protein, is thought to be facilitated by the stabilization of a pair of α-helices[19]. Unique solutions appear to have evolved in proteins with unrelated folds.

Pro regions solve an important problem for proteins that must survive in harsh environments rich in proteases[5,6]. These proteins need to have highly cooperative unfolding barriers to suppress transient unfolding which would render them susceptible to proteolysis. A consequence of the large unfolding barrier is a large barrier to folding as well. The co-evolution of a foldase (pro region) with the barrier provides an irreversible pathway to a kinetically-stabilized native state. The structures presented here suggest how such a pathway arises and provide a basis for further mutational studies to probe this phenomenon.

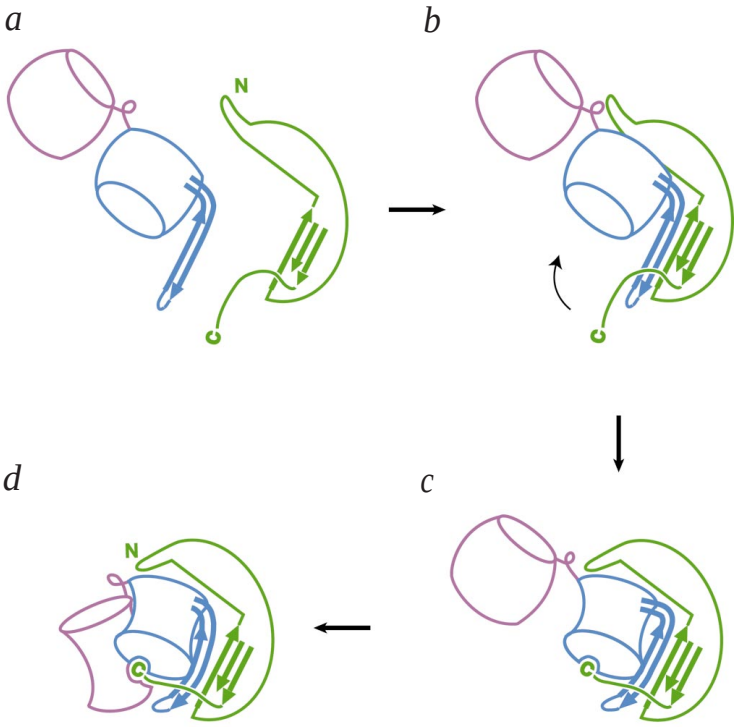

Fig. 4 Proposed mechanism for Pro-assisted folding of αLP. ***a***, In the molten globule folding intermediate the two domains of αLP are separated (shown as expanded barrels), and the αLP 118–130 hairpin is not packed against the rest of the αLP C-terminal domain. ***b***, The Pro C-domain interacts with the 118–130 hairpin of the intermediate, forming a five-stranded β-sheet. ***c***, The Pro C-tail binds to the nascent active site, positioning the hairpin and leading to folding of the αLP C-terminal domain. ***d***, αLP's N-terminal domain folds upon docking with the C-terminal domain. Color scheme is as in Fig. 2.

Methods

Structure determination of the unbound Pro. Pro was expressed from the plasmid pT7Pro[11], which encodes the residues MP followed by the 166-amino acid Pro sequence[20]. Metabolic labeling with selenomethionine[12] was accomplished using the methionine-deficient *E. coli* strain B834 (DE3) pLysS (Novagen) grown in defined LeMaster medium[21] supplemented with 25 mM MOPS, pH 7.3, and 1× Kao & Michayluk vitamin solution (Sigma #K-3129). Protein was renatured from inclusion bodies[9], and further purified by hydrophobic interaction chromatography using a Bio-Gel Phenyl-5PW column (Bio-Rad). Electrospray ionization mass spectroscopy revealed that the N-terminal methionine was removed, and that selenomethionine incorporation into the two remaining methionine residues was ~80%. This level of incorporation was sufficient for the solving the structure, although it was later determined that the incomplete labeling was due to the use of Kao & Michayluk vitamin solution, which causes the bacteria to be partially non-auxotrophic for methionine (A. Shiau, unpublished results). Crystals were grown at 7 °C by equilibrating a 14 μl drop containing 9 mg ml^{-1} Pro, 0.34 M lithium sulfate, 2.3% glycerol, 2 mM dithiothreitol, 2.7 mM MES, and 0.45x polybuffer (1× buffer = 15 mM citric acid, 30 mM phosphoric acid, 26 mM orthoboric acid, 0.155 M sodium hydroxide, adjusted to pH 8.0 with HCl), against a 1.5 ml well containing 0.8 M lithium sulfate, 5% glycerol, and 1× polybuffer, pH 8.0. After 37 h, 1 μl of diluted seed stock from pre-existing crystals was added to aid nucleation. After three months of growth, a single crystal of dimensions 0.3 × 0.1 × 0.1 mm was briefly transferred to 0.8 M lithium sulfate, 25% glycerol, 1× polybuffer (pH 8.0), and flash cooled in liquid nitrogen. Multiwavelength anomalous dispersion data were collected with inverse beam geometry at beamline X4A of the National Synchrotron Light Source (Brookhaven, New York). Location of the direct beam on Fuji BAS-2000 image plates was facilitated by the program GETBEAM (Marian Szebenyi, Cornell High Energy Synchrotron Source). Data were reduced with the DENZO/SCALEPACK package[22], and local scaling corrections[23] were applied to Bijvoet and dispersive data. Merged intensity measurements were adjusted[24] to enforce the assumption that all reflections are positive, with no adjustment being necessary for strong reflections. All data were included in subsequent calculations, without any I/σ(I) cutoff. After the heavy atom model was refined with SHARP[25], it was found that the unit cell contained three Pro molecules related by a non-crystallographic 3_1 screw axis parallel to the [111] axis. Electron density for the N-terminus (including methionine-12) is missing for one of the molecules, with the consequence that only five out of six selenium atoms are present in the model. Phases were improved by non-crystallographic symmetry averaging with the program DM[26]. Also, since the N-termini (residues 1–24) were unrelated by NCS symmetry, SOLOMON[27] was used separately to improve the density maps through solvent flattening. Secondary structural elements were identified in part with the program ESSENS[28]. An initial atomic model was refined by simulated annealing using a maximum likelihood target function[29] and torsion angle dynamics[30], while later rounds of refinement employed a conventional least-squares target function and

Table 1 Statistics for data collection, phase determination and refinement

Unit cell contents	*bis*-Selenomethionyl Pro			αLP (M158A variant) + Pro-3		αLP (M158A variant) + Pro
Space group	P1			$P2_1$		P1
Unit cell	a = 47.6 Å, b = 54.7 Å, c = 64.0 Å			a = 61.2 Å, b = 101.0 Å		a = 53.8 Å, b = 61.9 Å, c = 72.6 Å
	α = 104.0°, β = 102.5°, γ = 91.9°			c = 72.1 Å, β = 109.6°		α = 109.4°, β = 99.2°, γ = 102.4°
Asymmetric unit	3 molecules			2 complexes		2 complexes
λ(Å)	0.97938 (edge)	0.97915 (peak)	0.9690 (remote)	1.080	1.54	1.54
d_{min} (Å)	3.0	3.0	3.0	1.8	2.4	2.4
Measurements	40,128	40,123	39,904	176,771	78,512	104,427
Unique reflections	22,803	22,800	22,778	73,741		31,304
Completeness (%)[1]	93.6 (95.0)	93.6 (95.0)	93.5 (94.9)	96.5 (95.5)		95.9 (93.9)
R_{sym} (I) (%)	4.7 (23.7)	4.7 (23.8)	5.3 (22.4)	8.6 (21.0)		7.1 (37.1)
<I>/<σ(I)>	14.2 (2.7)	14.8 (2.7)	12.7 (2.6)	16.9 (3.6)		17.5 (3.4)
Mean figure of merit[2]	0.55 (0.29)					
R_{cullis} (dispersive)[3]	0.44 (0.59)	0.45 (0.55)	–			
R_{cullis} (anomalous)[4]	0.80 (1.04)	0.73 (1.10)	0.81 (1.03)			
Phasing power (dispersive)[5]	3.68 (2.23)	4.30 (2.75)	–			
Phasing power (anomalous)	2.02 (1.04)	2.41 (1.21)	1.88 (0.98)			
R_{cryst} / R_{free}[6]		20.2% / 34.2%		20.6% / 22.9%		19.6% / 24.9%
R.m.s. deviation from ideality (bonds / angles)		0.014 Å / 1.8°		0.005 Å / 1.2°		0.007 Å / 1.2°

[1]Values in parentheses are statistics for the highest resolution bin.
[2]Mean figure of merit using data from all three MAD wavelengths.
[3]R_{cullis} (dispersive) = <phase integrated lack of closure>/<|F_{obs} – F_{obs} (remote λ)|>
[4]R_{cullis} (anomalous) = <phase integrated lack of closure> / <|F_{obs} (+) – F_{obs} (–)|>
[5]Phasing power = < [|F_{calc} (selenium model)| / phase-integrated lack of closure] >
[6]$R_{cryst}/R_{free} = \Sigma_{hkl} |F_{obs} - F_{calc}| / \Sigma_{hkl} F_{obs}$, where R_{cryst} is calculated for a randomly chosen 90% of the data used for refinement, and R_{free} is calculated for the remaining 10% of the data. For selenomethionyl Pro, the model was refined against data from the remote wavelength. R-values are calculated on data in the resolution range from 6 Å to the high resolution limit.

dynamics[31–33]. NCS restraints were used to relate the N-domain (residues 25–86) and C-domain (residues 87–158) portions of symmetry mates; however residues 55–60 and 133–135 were unrestrained. Due to missing electron density the three N-termini are absent up to residues 4, 3 and 14 respectively; and the final model also lacks residues 103–114 and 159–166. Relevant crystallographic statistics are presented in Table 1.

Structure determination of the Pro–αLP complex. Two steps were taken to prevent proteolysis of Pro by αLP during crystallization: crystals of the complex were prepared using a protease variant (M158A) with ~100-fold lower enzymatic activity[34], and a protease inhibitor was added. For the Pro-3 complex, αLP was inhibited with 4-(2-aminoethyl) benzenesulfonylfluoride (AEBSF). This complex crystallized in spacegroup $P2_1$ at 7 °C from 10–15% PEG 8000, 125 mM ammonium sulfate, and 50 mM MES, pH 6.25. The final electron density map clearly shows the AEB-sulfonyl adduct attached to αLP's catalytic serine. The adduct is big enough to fill the ~10 Å space between the catalytic serine and the Pro-3 C-terminus. For the wild type Pro complex, αLP was inhibited with a partially-characterized, in-house synthesized benzene sulfonyl fluoride derivative (A. Shiau, pers. comm.). This complex crystallized in space group P1 at room temperature from 20% PEG 8000, 150 mM ammonium sulfate, and 100 mM sodium cacodylate, pH 6.5. There was no evidence for the inhibitor in omit maps, and the Pro C-terminus clearly occupies the site where the inhibitor was expected to be located. This is consistent with the biochemical observation that the C-terminus of wild-type Pro is able to displace AEB-sulfonyl adducts over time[35].

A single $P2_1$ crystal containing Pro-3/αLP was flash-cooled in a –170 °C nitrogen gas stream, and X-ray diffraction data were collected on an R-AXIS IIC image plate detector using CuKα radiation. Another data set from the same crystal was subsequently collected on a Mar 18 cm detector at beamline 7-1 of the Stanford Synchrotron Radiation Laboratory. A self-Patterson map revealed the presence of two-fold translational NCS symmetry[36]. A cross rotation search in X-PLOR[33] with the wild-type αLP model gave a clear solution. The two correctly positioned αLP molecules were subject to rigid-body and positional refinement, and a σ_A-weighted[37] $2F_o$ - F_c map was calculated. After solvent flattening in SOLOMON[27], density corresponding to Pro was clearly visible and permitted manual docking of the Pro model. Following rigid-body refinement of both αLP and Pro, the model was improved with consecutive rounds of manual rebuilding, simulated annealing, B-factor and positional refinement. The maximum likelihood target function[29] was used throughout, and the two complexes in the asymmetric unit were refined independently. The final model consists of all 198 residues of αLP, residues 6–163 of Pro in one monomer and residues 3–163 in the other, 321 water molecules, and a covalent AEBS adduct to the Oγ of Ser 143 of the protease in each complex. Pro residues 104–112 are missing for both monomers and presumed to be disordered. All residues are within allowed regions of the Ramachandran plot (not shown), as calculated by PROCHECK[26].

The structure of the wild-type Pro–αLP complex in the P1 crystal form was solved by molecular replacement using the Pro-3–αLP structure as a search model. Diffraction data were collected on an R-AXIS IIC image plate detector using CuKα radiation. A self-rotation function had previously established the presence of two-fold NCS in this crystal form[36]. The orientations of the two complexes and their relative positions within the unit cell were found using rotation and translation searches. The structure was refined in a similar fashion as above. Total buried accessible surface area with and without interfacial waters (4,117 Å^2 and 3,349 Å^2 respectively) was calculated using GRASP using a probe radius of 1.4 Å.

Coordinates. Atomic coordinates and structure factors have been deposited with the PDB for unbound Pro region (2pro & r2prosf), Pro-3/αLP complex (3pro & r3prosf), and wild type Pro–αLP complex (4pro & r4prosf).

Acknowledgments
This work was supported by the Howard Hughes Medical Institute (HHMI). N.K.S. was supported in part by a Damon Runyon-Walter Winchell postdoctoral fellowship, T.M. by an HHMI predoctoral fellowship, and S.D.R. by an NIH training grant. We thank D. King for mass spectroscopy analysis; C. Ogata (NSLS),

M. Soltis and H. Bellamy (SSRL) for beamline support; P. David for logistical help; C. Wilson for the occasional use of his Raxis II; A. Shiau and A. Derman for assistance with data collection; and A. Derman, S. Gillmor, A. Shiau and J. Sohl for critical comments on the manuscript. Some data were collected at SSRL, which is operated by the Department of Energy, Office of Basic Energy Sciences. The SSRL Biotechnology Program is supported by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and by the Department of Energy, Office of Biological and Environmental Research.

Nicholas K. Sauter[1], Ted Mau[1,2], Stephen D. Rader[2] and David A. Agard[2]

[1]These authors contributed equally to this work. Howard Hughes Medical Institute, [2]the Graduate Group in Biophysics and the Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California 94143-0448, USA.

Correspondence should be addressed to D.A.A. *email: agard@msg.ucsf.edu*

Received 15 July 1998; accepted 22 September, 1998.

1. Zhu, X., Ohta, Y., Jordan, F. & Inouye, M. *Nature* **339,** 483–484 (1989).
2. Baker, D., Silen, J.L. & Agard, D.A. *Proteins Struct. Funct. Genet.* **12,** 339–344 (1992).
3. Strausberg, S., Alexander, P., Wang, L., Schwarz, F. & Bryan, P. *Biochemistry* **32,** 8112–8119 (1993).
4. Winther, J.R. & Sorensen, P. *Proc. Natl. Acad. Sci. USA* **88,** 9330–9334 (1991).
5. Baker, D., Shiau, A.K. & Agard, D.A. *Curr. Opin. Cell Biol.* **5,** 966–970 (1993).
6. Sohl, J.L. & Agard, D.A., In *Intramolecular Chaperones and Protein Folding* (eds Shinde, U. & Inouye, M.) 61–83 (R.G. Landes Co., Austin, Texas; 1995).
7. Silen, J.L. & Agard, D.A. *Nature* 341**,** 462–464 (1989).
8. Baker, D., Sohl, J.L. & Agard, D.A. *Nature* **356,** 263–265 (1992).
9. Peters, R.J. *et al. Biochemistry*, **37**,12058–12067 (1998).
10. Sohl, J.L., Jaswal, S.S. & Agard, D.A. *Nature*, **in the press** (1998).
11. Sohl, J.L., Shiau, A.K., Rader, S.D., Wilk, B.J. & Agard, D.A. *Biochemistry* **36,** 3894–3902 (1997).
12. Leahy, D.J., Hendrickson, W.A., Aukhil, I. & Erickson, H.P. *Science* **258,** 987–991 (1992).
13. Bone, R., Shenvi, A.B., Kettner, C.A. & Agard, D.A. *Biochemistry* **26,** 7609–7614 (1987).
14. Fersht, A. *Enzyme structure and mechanism* (W.H. Freeman, New York; 1985).
15. Gewirth, D.T. & Sigler, P.B. *Nature Struct. Biol.* **2,** 386–394 (1995).
16. Schirmer, T. & Evans, P.R. *Nature* **343,** 140–145 (1990).
17. Royer, W.E., Jr., Pardanani, A., Gibson, Q.H., Peterson, E.S. & Friedman, J.M. *Proc. Natl. Acad. Sci. USA* **93,** 14526–14531 (1996).
18. Allaire, M., Chernaia, M.M., Malcolm, B.A. & James, M.N.G. *Nature* **369,** 72–76 (1994).
19. Gallagher, T., Gilliland, G., Wang, L. & Bryan, P. *Structure* **3,** 907–914 (1995).
20. Silen, J.L., McGrath, C.N., Smith, K.R. & Agard, D.A. *Gene* **69,** 237–244 (1988).
21. LeMaster, D.M. & Richards, F.M. *Biochemistry* **24,** 7263–7268 (1985).
22. Otwinowski, Z. & Minor, W. *Methods Enzymol.* **276,** 307–326 (1997).
23. Hendrickson, W.A., Smith, J.L., Phizackerly, R.P. & Merritt, E.A. *Proteins Struct. Funct. Genet.* **4,** 77–88 (1988).
24. French, S. & Wilson, K. *Acta Crystallogr. A* **34,** 517–525 (1978).
25. de La Fortelle, E. & Bricogne, G. *Meth. Enz.* **276,** 472–494 (1997).
26. Collaborative Computational Project, Number 4 *Acta Crystallogr. D* **50,** 760–776 (1994).
27. Abrahams, J.P. & Leslie, A.W.G. *Acta Crystallogr. D* **52,** 30–42 (1996).
28. Kleywegt, G.J. & Jones, T.A. *Acta Crystallogr. D* **53,** 179–185 (1997).
29. Pannu, N.S. & Read, R.J. *Acta Crystallogr. A* **52,** 659–668 (1996).
30. Rice, L.M. & Brünger, A.T. *Proteins Struct. Funct. Genet.* **19,** 277–290 (1994).
31. Brünger, A.T., Kuriyan, J. & Karplus, M. *Science* **235,** 458–460 (1987).
32. Brünger, A.T., Krukowski, A. & Erickson, J. *Acta Crystallogr. A* **46,** 585–593 (1990).
33. Brünger, A.T. *X-PLOR version 3.1: a system for X-ray crystallography and NMR* (Yale Univ. Press, New Haven, Connecticut; 1992).
34. Bone, R., Silen, J.L. & Agard, D.A. *Nature* **339,** 191–195 (1989).
35. Mau, I.-F.T. Ph. D. Thesis, University of California, San Francisco (1998).
36. Rader, S.D. Ph.D. Thesis, University of California, San Francisco (1996).
37. Read, R.J. *Acta Crystallogr. A* **42,** 140–149 (1986).

Efficient replication between non-hydrogen-bonded nucleoside shape analogs

DNA polymerase enzymes make an error only once per 10^4–10^5 initial nucleotide insertions during DNA replication. Most currently held models of this high fidelity cite the hydrogen bonds between complementary pyrimidines and purines as a critical controlling factor. Testing this has been difficult, however, since standard molecular strategies for blocking or removing polar hydrogen-bonding groups cause changes to size and shape as well as hydrogen bonding ability. One answer to this problem is the use of nonpolar molecules that mimic the shape of natural DNA bases. Here we show that a non-hydrogen-bonding shape mimic for adenine is replicated efficiently and selectively against a nonpolar shape mimic for thymine. The results establish that hydrogen bonds in a base pair are not absolutely required for efficient nucleotide insertion. This adds support to the idea that shape complementarity may play as important a role in replication as base–base hydrogen bonds.

Both geometric effects and base–base hydrogen bonding have long been thought to be important factors in DNA replication fidelity[1–9]. To test whether there is an absolute requirement for hydrogen bonding in base pair replication, we synthesized nonpolar shape analogs of two naturally paired bases, and we examined the properties of this non-hydrogen-bonded pair as substrates for the Klenow fragment of *Escherichia coli* DNA Polymerase I. A difluorotoluene deoxynucleoside (F, Fig. 1*a*) has recently been shown to be a nearly exact isosteric mimic for natural thymidine nucleoside (refs 10–13; E.T.K., D. Barsky, & M. Colvin, unpublished data). To complement this analog in shape, we synthesized a non-hydrogen-bonding analog of deoxyadenosine (Z, Fig. 1*a*), utilizing 4-methylbenzimidazole to replace the adenine base[14,15]. Thermal denaturation studies indicate that this analog shows no selective preference for pairing among the four natural bases, consistent with its nonpolar nature[12].

We examined base pair replication properties of this analog in a synthetic primer–template duplex with Z in the template strand (Fig. 1*a*), using a Klenow fragment mutant lacking 3'-exonuclease activity. Single-nucleotide insertions of each of the four natural nucleotides as well as dFTP and dZTP opposite this template Z were first examined qualitatively by separate addition of individual nucleotides (Fig. 2). Relatively efficient insertion of dFTP is seen, and among the four natural nucleotides there appears to be significant insertion of dTTP, but very little of any other. Comparison of the same single-nucleotide additions opposite a template adenine base in the same context shows efficient insertion only of dTTP and less efficient insertion of dFTP (Fig. 2*a*). Thus, the initial results showed qualitatively good selectivity for insertion of a nonpolar thymine shape analog (F) opposite the nonpolar adenine analog (Z), and with apparent efficiency considerably greater than observed for common base mismatches.

The nucleoside triphosphate analog of Z (dZTP, Fig. 1*a*) was synthesized as a possible enzymatic nucleotide substrate to examine the converse situation, using dZTP as a non-hydrogen-bonding analog of dATP. Results showed selective elonga-