Title
Functional modulation of a protein folding landscape via side-chain distortion

Journal
Proceedings of the National Academy of Sciences of the United States of America, 109(24)

ISSN
0027-8424

Authors
Kelch, Brian A
Salimi, Neema L
Agard, David A

Publication Date
2012-06-12

DOI
10.1073/pnas.1119274109

Peer reviewed

Functional modulation of a protein folding landscape via side-chain distortion

Brian A. Kelch[1], Neema L. Salimi, and David A. Agard[2]

Howard Hughes Medical Institute and the Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94158

Edited by Brian W. Matthews, University of Oregon, Eugene, OR, and approved March 26, 2012 (received for review November 23, 2011)

Ultrahigh-resolution (<1.0 Å) structures have revealed unprecedented and unexpected details of molecular geometry, such as the deformation of aromatic rings from planarity. However, the functional utility of such energetically costly strain is unknown. The 0.83 Å structure of α-lytic protease (αLP) indicated that residues surrounding a conserved Phe side-chain dictate a rotamer which results in a ∼6° distortion along the side-chain, estimated to cost 4 kcal/mol. By contrast, in the closely related protease Streptomyces griseus Protease B (SGPB), the equivalent Phe adopts a different rotamer and is undistorted. Here, we report that the αLP Phe side-chain distortion is both functional and conserved in proteases with large pro regions. Sequence analysis of the αLP serine protease family reveals a bifurcation separating those sequences expected to induce distortion and those that would not, which correlates with the extent of kinetic stability. Structural and folding kinetics analyses of family members suggest that distortion of this side-chain plays a role in increasing kinetic stability within the αLP family members that use a large Pro region. Additionally, structural and kinetic folding studies of mutants demonstrate that strain alters the folding free energy landscape by destabilizing the transition state (TS) relative to the native state (N). Although side-chain distortion comes at a cost of foldability, it suppresses the rate of unfolding, thereby enhancing kinetic stability and increasing protein longevity under harsh extracellular conditions. This ability of a structural distortion to enhance function is unlikely to be unique to αLP family members and may be relevant in other proteins exhibiting side-chain distortions.

protein distortion | protein structure | X-ray crystallography | protein stability | protein lifetime

The physicochemical forces that determine macromolecular structure and function are generally thought to be well understood. The predominant forces include well-studied hydrogen bonding, van der Waals, and ionic interactions. By contrast, bonded terms are largely ignored as it is commonly held that bond lengths and angles are at or very close to canonical values in the native state. This view is so widely adopted that nearly all crystallographic and NMR structure determination methods include restraints to optimal values. However, distortion of covalent bonds in protein sidechains away from their low energy configurations has now been noted in several ultrahigh-resolution structures (1, 2) (Fig. 1*A*), but no functional role for the distortion has been ascribed. We examine the functional consequences of side-chain distortion in α-lytic protease (αLP), in which a 0.83 Å structure (2) revealed a surprisingly large distortion of the aromatic ring of a conserved Phe (Phe228) by almost 6° from planarity (Fig. 1*B*; about 2.6 standard deviations from the mean in Fig. 1*A*).

The αLP sub-family of serine proteases is one of the best-studied examples of a kinetically stable protein (3, 4); i.e. a protein that retains its native state through exceedingly slow unfolding rates. The folding landscape of αLP is distinct from typical, thermodynamically stable proteins in that its folding barrier is very high (corresponding to a $t_{1/2} \sim 1{,}800$ y), and the native state (N) is less stable than the unfolded forms by an unprecedented 4 kcal/mol (5). Therefore, αLP's functional stability is dictated solely by its extremely high unfolding barrier [$t_{1/2} \sim 1.2$ y for N to the transition state (TS)] rather than the overall free energy difference between its native and unfolded states (5). In αLP and other family members, kinetic stability is a preferred mechanism for stability because the high unfolding energy barrier and the reduced native state dynamics result in enhanced longevity for these proteases in harsh environments, such as pH fluctuations and high protease concentrations (6–9).

A covalently attached folding catalyst (the pro region) performs two crucial functions to facilitate αLP folding (5): It (*i*) accelerates folding by nine orders of magnitude so that folding can occur on a biologically relevant timescale and *ii*) stabilizes the native state over the unfolded state so that the reaction is pulled in the direction of the native state. The pro region is subsequently degraded, leaving the native protease in its metastable but kinetically trapped active form (10).

Based on thermodynamic, kinetic, mutational, and computational studies of αLP and its homologs (6, 8, 11–13), our group has proposed a TS structure wherein each β-barrel domain is relatively well folded but the interface between the two domains is not fully formed. Although these studies have provided a phenomenological model for the TS, an explanation for the source of the unique kinetic stability observed in αLP is still lacking. Specifically, what causes the TS to be at such a high energy and why is the N destabilized relative to the unfolded forms? Given the context, the observed distortion of Phe228 led to the suggestion that it might somehow play a role in αLP kinetic stability; perhaps by suppressing native state dynamics (2), but this hypothesis had not been experimentally tested.

By analyzing mutants of αLP as well as related family members with altered folding landscapes, we show that distortion of Phe228 correlates with the degree of kinetic stability and contributes directly to the height of the unfolding barrier by preferentially destabilizing TS. However, contrary to previous hypotheses (2), the distortion appears to not play a role in the ultracooperativity observed in αLP unfolding. This represents, to our knowledge, a previously uncharacterized functional role for side-chain distortion. Finally, our meta-analysis of previous ultrahigh-resolution structures indicates that side-chain distortion is a surprisingly common and underappreciated phenomenon that may be playing functional roles in other protein families as well.

Results and Discussion

We used the wealth of information in sequence data to understand the structural basis for kinetic stability in the secreted bacterial

Author contributions: B.A.K. and D.A.A. designed research; B.A.K. performed research; N.L.S. contributed new reagents/analytic tools; B.A.K. and D.A.A. analyzed data; and B.A.K. and D.A.A. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 3URC, 3URD, and 3URE).

[1]Present address: Department of Molecular and Cell Biology, University of California, Berkeley, CA 94270.

[2]To whom correspondence should be addressed. E-mail: agard@msg.ucsf.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1119274109/-/DCSupplemental.

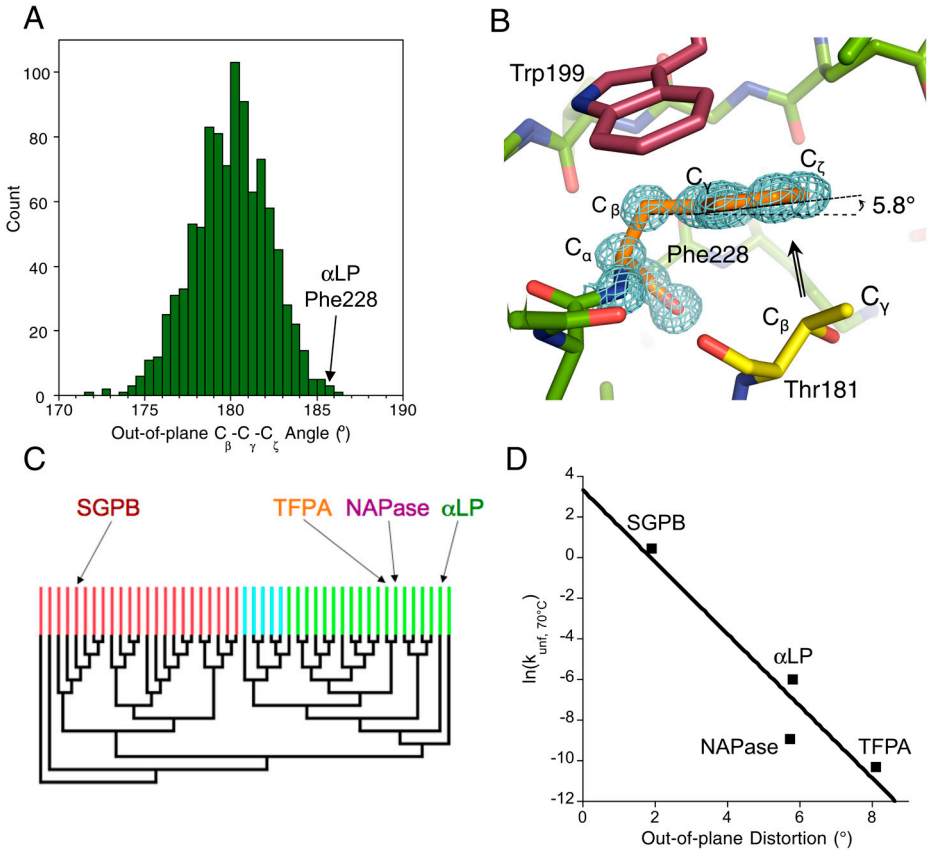

Fig. 1. Distortion of phenylalanine side-chains. (*A*) Histogram of Phe distortions present in ultrahigh-resolution (≤0.99 Å) structures in the PDB. The distortion of αLP Phe228 is nearly 3 standard deviations from the mean. (*B*) Distortion of Phe228 of αLP is caused by close contact of Thr181 [from a 0.83 Å resolution structure (2)]. (*C*) Phylogenetic analysis shows an evolutionary bifurcation in C-terminal domain packing and pro region size. Proteases utilizing a small pro region (red) have a packing arrangement similar to SGPB, while those associated with a large pro region are expected to be arranged as in αLP (green). Proteases exhibiting sequences that are a hybrid of SGPB and αLP are shown in cyan. (*D*) The unfolding barrier of the kinetically stable proteases is dependent on the distortion of residue 228. There is a strong correlation with the height of the unfolding barrier with the degree of Phe228 distortion for all homologs that have been tested.

proteases. A protease that is closely related to αLP, *Streptomyces griseus* Protease B (SGPB), requires a smaller and less effective pro region because it has a significantly reduced kinetic stability compared to αLP (9). Therefore, analysis of sequence variation patterns and structural differences between the two proteases and their close homologs should help to identify regions of the molecule responsible for the observed functional differences. A prime candidate region is the differential packing in the C-terminal domain responsible for the distortion of Phe228 in αLP (Fig. 1*B* and Fig. S1*A*; numbered by homology with chymotrypsin). In SGPB, the equivalent Phe adopts a different rotamer and is not distorted due to alternate packing in the hydrophobic core (14, 15). Additionally, two other structurally characterized proteases that share SGPB's sequence features in the C-terminal domain also place Phe228 in the same rotamer as found in SGPB (16, 17). These proteases also use a small pro region as in SGPB, suggesting that they too have reduced kinetic stability as compared to αLP. In fact, all of the proteases that use a small pro region share the same C-terminal domain packing scheme as SGPB (Fig. 1*C* and Fig. S1*B*). The correlation of the C-terminal domain packing with the functional attributes of a less powerful folding catalyst and, presumably, lower energy barriers for folding and unfolding suggests that there may be a causative link between distortion of Phe228 and kinetic stability.

To investigate this linkage further, we determined the high-resolution structures of two αLP homologs, TFPA (6) and NAPase (8) that use a large pro region to catalyze folding and have extremely large unfolding barriers. These proteins exhibit the same core packing as αLP and thus also show significant distortion in Phe228 (Fig. S2). Analysis of side-chain distortion in Phe228 and unfolding rates of multiple homologs reveals a strong correlation between the out-of-plane distortion angle (as measured by the angle between C_β-C_γ-C_ζ; see *Materials and Methods* for calculation) and unfolding barrier height across multiple αLP homologs (Fig. 1*D*), further suggesting that side-chain distortion in Phe228 directly contributes to kinetic stability.

To test this hypothesis, we created αLP mutants that are expected to reduce the distortion of Phe228. This is particularly challenging because there is no single mutant that simply removes the Phe228 strain without disrupting other interactions. For example, mutating Phe228 to alanine would eliminate the strain, but would also create a large cavity in the hydrophobic core that would severely impact folding independent of eliminating strain. Therefore, our goal was to create multiple mutants that remove strain with minimal perturbation to other interactions, and to determine the effect of strain on folding.

Our first strategy was to modify the interactions between Phe228 and the Thr residue (Thr181) mainly responsible for inducing the distortion. Because the strain is mostly imposed by the C_β atom of Thr181 (Fig. 1*B*), the T181G mutant should relieve the strain. To account for any contributions of the C_γ methyl and hydroxyl groups of Thr181 toward stability, we made a second mutant in which the Thr residue is converted to Ala (T181A). The T181A mutant should maintain the majority of the Phe228 distortion, while removing most of the stabilizing contributions from the branching moieties of the threonine side chain. Therefore, to best quantify just the effect of strain on stability, we compare T181G with T181A. In an alternate mutational strategy, we sought to rebuild the αLP hydrophobic core so that strain is eliminated while van der Waals interactions are maximally preserved. Using SGPB as a guide (14), αLP was mutated (T181I, W199L, Q210I; hereafter called "Repack") to construct a well-packed protease that selects for the unstrained Phe228 rotamer.

To quantify the strain present in these mutants, we solved the structures of the T181G and T181A mutants at atomic resolution (both ~1.1 Å; Table S1). The structures reveal that the T181G mutation significantly alleviates strain in Phe228 (2.7° bend angle, or ~1.2 σ from the mean), while the T181A mutation results in

a mild reduction in strain (4.6° bend angle, or over 2 σ from the mean; see Fig. 1*A*). We treat the distortion of Phe228 as a Hookean spring. Using the quantum mechanical calculation of 4 kcal/mol for the strain energy in WT-αLP Phe228 (2) to obtain the spring constant, the T181A mutant results in an expected native state strain energy of ~2.4 kcal/mol and the T181G mutant results in only 0.8 kcal/mol strain energy. As discussed, focusing on the Ala to Gly change (T181A vs. T181G) provides a more precise means to differentiate the folding impact of strain removal from the contributions of Thr $O_{\gamma 1}$ and $C_{\gamma 2}$ (WT vs. T181A). Importantly, only modest changes in B factors were observed in the T181G mutant; thus the observed effects should only be due to the loss of the C_β atom and not to altered mobility in the native state.

To quantify the functional consequences of these mutations, we determined folding rates by measuring the very small fraction of active protease that refolds within a practical time frame (~1 week), and unfolding rates by denaturant titration monitored by tryptophan fluorescence (5). The rate of folding for the T181A mutant was ~7-fold slower than that of wild-type αLP (Fig. 2*B*; Table S1). Remarkably, removing the entire Thr side-chain (T181G) restored the folding rate to wild-type levels (Fig. 2*B*). By contrast, both mutants unfold significantly faster than WT-αLP (T181A ~13-fold, T181G ~35-fold; Fig. 2*C*). These changes in folding and unfolding rates were converted to $\Delta\Delta G^{\ddagger}$ and used to generate comparative free energy diagrams (Fig. S3). In the absence of confounding effects such as strain, loss of favorable interactions important for folding should destabilize both the TS and N states. As expected, the removal of the C_γ and hydroxyl groups from Thr181 (T181A) causes destabilization of both TS (1.0 kcal/mol) and N (2.4 kcal/mol). The C_γ methyl group of Thr181 is packed tightly within the hydrophobic core and the hydroxyl group coordinates an intricate hydrogen-bonding network (2), so it is not surprising that the Ala mutation significantly destabilizes N and TS, even though the strain energy in Phe228 is estimated to be reduced by ~1.6 kcal/mol. By contrast, the loss of C_β (T181G), which is clearly well packed in the native state, results in surprising energetic effects (1.1 and 0.6 kcal/mol stabilization of TS and N relative to T181A, respectively). This stabilization is unexpected because Gly increases configurational entropy of unfolded states, which should destabilize TS and N by about 0.7 kcal/mol (18) (assuming that this residue is completely ordered in the TS). This apparent contradiction can be readily understood if there were significant strain in both the TS and N of T181A that was diminished upon mutation to Gly. In support of this, the crystal structures show that the only significant effect of the T181A to T181G mutation is relief of the Phe228 side-chain distortion.

Although the mutations described above suggest that the strain from side-chain distortion plays an important role in destabilizing the αLP TS, the additional removal of stabilizing van der Waals interactions, entropic effects of mutation to Gly (18), and the mild decrease in strain in T181A (Fig. 2*A*) can mask the full effect of strain removal. The Repack mutant described above (in which the αLP core is rebuilt to match that of SGPB) is expected to eliminate these complications more effectively. A 1.5 Å structure of Repack (Table S1) shows that the residues in the C-terminal domain core superimpose with those of SGPB (Fig. 3*A*) and confirms that Phe228 is not distorted (average residual distortion of ~0.4° based on the two molecules in the asymmetric unit).

In accordance with our hypothesis, the Repack mutant shows a dramatic enhancement in the folding rate (~60-fold faster than wild type; Fig. 3*B*) and a smaller, but significant, increase in the unfolding rate (~13-fold faster than WT; Fig. 3*C*; Table S2). This translates into a stabilization of both TS and N by ~2.2 and 0.8 kcal/mol, respectively (Fig. S4). Remarkably, Repack captures key energetic as well as structural features of SGPB, recapitulating ~36% of the difference in folding barrier height with only ~2.5% of the sequence variation between the two proteases.

Having shown that side-chain distortion affects the folding barrier height, we next wanted to examine whether the distortion is associated with the cooperativity of unfolding. It had been previously hypothesized (2) that the strain in the Phe228 side-chain

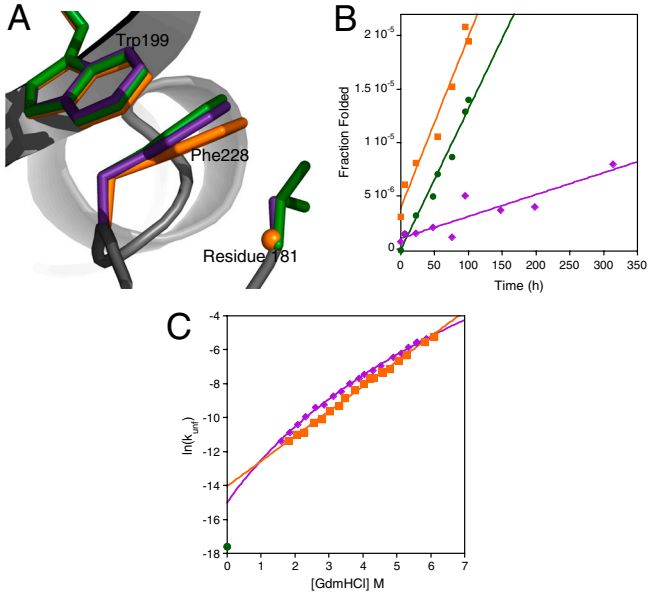

Fig. 2. Structure and folding kinetics of Thr181 mutants. (*A*) Structures of wild-type αLP (2) (green), T181A (1.08 Å; violet), and T181G (1.10 Å; orange). Phe228 distortion is still present in T181A, but absent in T181G. (*B*) Time course of folding for WT-αLP (green circles), T181A (violet diamonds), and T181G (orange squares). T181A folds ~10 times slower than WT and T181G. (*C*) Extrapolated unfolding rates for T181A (violet) and T181G (orange), compared to the unfolding rate for WT (green). T181A and T181G unfold 35- and 13-times faster than WT.

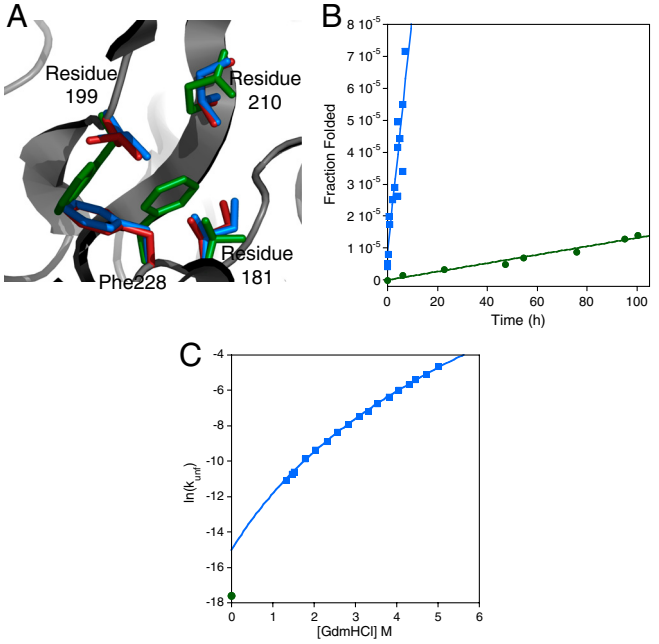

Fig. 3. Structure and folding kinetics of the Repack mutant. (*A*) C-terminal domain packing of wild-type αLP (2) (0.83 Å; green), Repack (1.5 Å; blue), and SGPB (15) (1.2 Å; red). C-terminal domain residues of Repack superpose with those of SGPB. (*B*) Time course of folding for WT-αLP (green), Repack (blue). Repack folds ~60 times faster than WT. (*C*) Extrapolated unfolding rate for Repack (blue) compared to the unfolding rate for WT (green). Repack unfolds ~13 times faster than WT-αLP.

is used to suppress native state dynamics in αLP, leading to its ultracooperative unfolding (7). To assess whether side-chain distortion suppresses αLP native state dynamics, we measured proteolysis using an autolysis assay in which the unfolding rate (measured by tryptophan fluorescence) and the loss of activity are compared across a large αLP concentration range. Proteolysis has been used as a very effective probe for measuring subglobal unfolding transitions (19–21). It has been previously established that the autolysis rate for αLP closely matches that of the global unfolding rate, indicating that autolysis is limited by global unfolding and is not due to partially unfolded forms (7), in contrast to many thermodynamically stable proteins (19, 20). If nonglobal unfolding events are populated to an appreciable degree, then the autolysis rate will be faster than the global unfolding rate. Because αLP is both an enzyme and substrate in the autolysis reaction, the autolysis rate should be especially sensitive to changes in concentration ($k_{autolysis}$ is roughly proportional to $[\alpha LP]^2$.) In addition, increasing the pH should further increase proteolytic activity by activating key catalytic residues.

We measured the rate of protease inactivation by loss of activity and/or fluorescence at 0.05 μM and 5 μM Repack-αLP at pH 8.0, close to αLP's maximal protease activity, thus increasing the potential rate of autolysis up to $\sim 10^4$-fold. However, the protease inactivation rate at high concentration matched the unfolding rate at low αLP concentration as measured by loss of tryptophan fluorescence (Fig. S5). This is an extremely robust result as the two rates are equivalent over a wide range of denaturant concentration and pH. Thus, removal of Phe228 distortion has no impact on the degree of unfolding cooperativity as measured by autolysis rates, indicating that Phe228 does not play a functionally important role in the suppression of native state dynamics in αLP as previously hypothesized. However, the inactivation and unfolding rates for the mutants are significantly increased relative to WT-αLP, demonstrating that side-chain distortion does play a significant role in increasing the functional lifetime of αLP via increasing the unfolding energy barrier.

The T181A to Gly and Repack mutations were designed as two independent strategies for alleviating strain in Phe228. In addition to the dominant effect of relieving side-chain strain, both of these mutants also lead to subtle backbone improvements in the β-sheet around Phe228 (three cross-strand H-bonds are ~0.1 Å shorter). These mutations produce coincident and unexpected thermodynamic results: the stabilization of N and TS. Because the common feature is the reduction of strain, the implication is that strain destabilizes TS and N in WT-αLP. The greater effect seen in Repack most likely results from better packing using the evolutionarily optimized SGPB core and the avoidance of the entropic consequences of the Gly mutation.

Because these mutations affect the stability of both TS and N, it is useful to quantify what fraction of the energetic effect has direct functional consequences. The relative contribution of strain to folding kinetics versus its contribution to thermodynamic stability can be quantified by Φ-value analysis ($\Phi = \Delta\Delta G^{\ddagger}/\Delta\Delta G$) (22). Φ values typically range from 0 to 1.0, and are interpreted as follows: Interactions altered by mutation have no contribution to the TS ($\Phi = 0$) or fully contribute to the TS ($\Phi = 1$), with fractional Φ values indicating a partial involvement in the TS structure. The Φ values for strain removal in both T181A to Gly and Repack mutants are remarkable: 1.9 (±0.5) and 2.7 (±0.5), respectively. Φ values are usually reported without their associated errors, but can be very susceptible to errors when the difference between $\Delta\Delta G^{\ddagger}$ and $\Delta\Delta G$ is small (23). Because our Φ values are extremely noncanonical, we propagated the errors to ensure these values are not an artifact of imprecise measurements, as interpretation of Φ values of this magnitude is not without controversy (24–26). Although not described in this context previously, such noncanonical Φ values should be a hallmark of removing interactions that contribute most significantly to kinetic stability, as they indicate interactions that affect TS more than N. Values such as these also suggest a TS structure that is highly organized, at least in the vicinity of the mutations.

The data from both Repack and the Thr181 mutations provide a consistent picture of the role of strain, confirming three key features of our hypothesis: (*i*) strain from the Phe228 distortion is indeed present in N, (*ii*) this strain also manifests itself in the TS, and (*iii*) the strain energy in TS must be ~2–3-fold greater than in N, as indicated by the Φ values (Fig. 4*A*). The excess contribution of side-chain distortion to TS directly indicates that the three-dimensional environment of Phe228 must be highly structured in the TS, although whether the degree of mechanical distortion is enhanced or if there are simply fewer compensatory interactions in the TS is not known. If the increased strain in the TS occurs entirely through deformation of the Phe, then we estimate the distortion of Phe228 (modeled as a Hookean spring) in the TS must be between 8–10°. This extends previous insights into the structure of the TS, which had only implicated regions around the domain interface as being important (6, 8). More importantly, the enhanced contribution of side-chain distortion to the TS provides a critical rationale for incorporating strain: the increased energetic differential between TS and N directly results in an increased kinetic stability (Fig. 4*A*), which, in turn, leads to a significantly extended functional lifetime.

Our results can rationalize important issues of pro-dependent folding of αLP and its homologs (4). The structure of αLP bound by its pro region showed that the pro region is a two-domain C-shaped protein that binds to and surrounds the αLP C-terminal

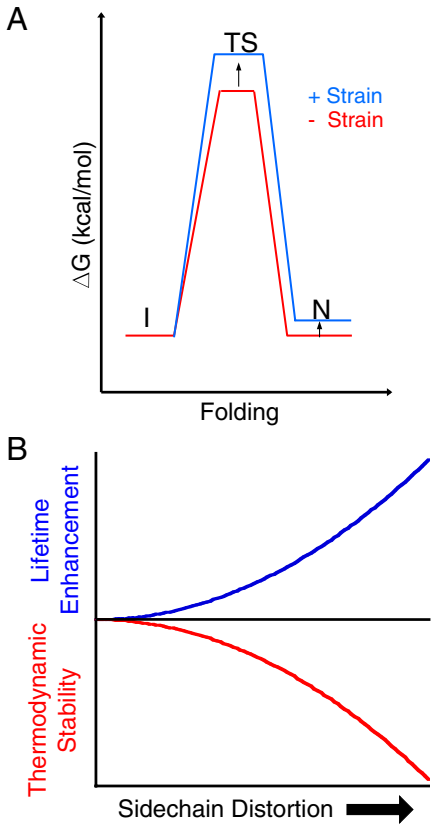

Fig. 4. Strain increases kinetic stability. (*A*) A schematic model for the effect of strain on the folding landscape of the kinetically stable proteases. Strain destabilizes the Transition State (TS) more so than the native state (N), thereby enlarging the unfolding free energy barrier and extending protease lifetime. (*B*) Schematic illustrating the differential effects of strain on kinetic versus thermodynamic stability. The energetic trade-off (modeled as a Hookean spring) between kinetic and thermodynamic stability is shown as found in αLP.

domain (27). Based on these results, it was hypothesized that much of the folding defect of αLP would be localized to the C-terminal domain. In this work, we localize a significant folding defect to the distortion of Phe228. Further, our data suggests that the C-terminal domain is highly structured in the TS, which is corroborated by the tight binding of the pro region to both TS and N (5). These data suggest significant structural similarity between TS and N. Therefore, we propose that the role of the large pro region of αLP and its close homologs is to clamp onto and organize the C-terminal domain such that the distortion of Phe228 is stabilized.

A key advantage of the kinetically stable proteases compared to their thermodynamically stable homologs is their increased ability to survive in harsh environments (7). The data presented here indicate that one key component in the evolution of kinetic stability is the introduction of geometric strain into the TS via the distortion of planar, aromatic side-chains, resulting in a dramatic enhancement of functional lifetime. However, the cost of introducing strain is significant; the additional improvement in lifetime comes at an ever-increasing penalty in terms of foldability and N stability (Fig. 4*B*). Although the substantial distortion observed in N (Fig. 1*B*) can be viewed as wasted residual strain from the TS, this strategy is surprisingly efficient: 2- to 3-fold higher strain energy is realized in the TS compared to N. This compares quite favorably to the fractional contribution to kinetic stability (Φ values between 0 and 1) that would be imparted by a typical thermodynamically stabilizing interaction such as an H-bond or van der Waals interaction.

For more conventional thermodynamically stable proteins, introduction of side-chain distortion would cause significant destabilization (Fig. 4*B*), thus we expected such distortions to be relatively rare in proteins. Surprisingly, ~20% of all proteins in our dataset of ultrahigh-resolution structures display a Phe or Tyr residue with a distortion similar to or even greater than that found in αLP (Fig. 5). Thus, the functional benefits must outweigh the energetic penalties associated with these distortions, suggesting that other proteins may also use the strain energy from side-chain distortions to modulate biological activity. In the case of αLP and its homologs, strain energy is used to extend functional lifetime, but we envision side-chain distortion also being relevant for other important macromolecular functions, such as allostery and catalysis. Although the functional importance of distortion in enzyme substrates (28–30) and cofactors (31–34) has long been appreciated, and recent analysis points to conserved peptide bond distortions (35), the observation of side-chain distortion and the discovery that it can play a significant role in modulating energetic landscapes to provide biologically important advantages is quite unique. This study identifies an unanticipated challenge: the need to observe structurally subtle yet functionally significant covalent distortions to fully understand the energetic forces acting on proteins and their impact on function.

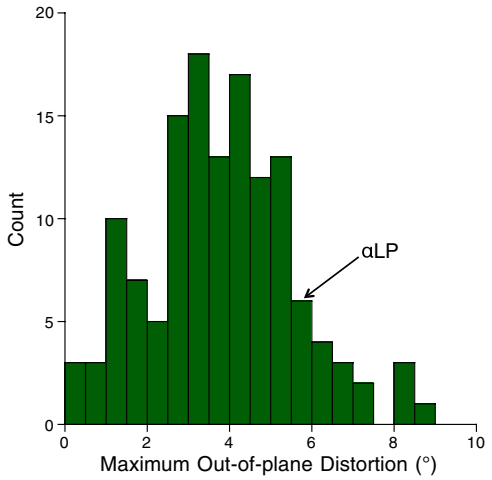

Fig. 5. Large distortions are not unique to αLP. Shown is a histogram of the largest Phe or Tyr distortion per protein from the ultrahigh resolution structures in our dataset. Approximately 20% of the proteins contained at least one Phe or Tyr residue with a distortion comparable or even larger than that of Phe228 in αLP.

Materials and Methods

Database of Phe Distortions. A database of atomic resolution structures (<1.0 Å) was built by examining the Protein Data Bank (PDB, www.pdb.org). A size threshold of 30 residues was enforced and structures with >90% sequence identity to another previously accepted member of the database were not included. Phenylalanine residues with multiple conformations were not included. The C_β-C_γ-C_ζ angle in Phe is an imperfect measure of out-of-plane ring distortion because in-plane ring distortions can also reduce the angle from its idealized value of 180°. To eliminate the effect of in-plane distortions, we developed the following protocol. For each Phe or Tyr, the side-chain is translated and rotated such that C_β is at the origin, C_γ is on the positive z axis, $C_{\delta1}(x) = C_{\delta2}(x)$, and $C_{\delta1}(y) < C_{\delta2}(y)$. The C_β-C_γ-C_ζ angle is unaffected by this transformation and measures the uncorrected distortion. Then $C_\zeta(y)$ is set equal to 0, eliminating the sideways bend, as all out-of-plane distortion comes from the value of $C_\zeta(x)$. The new C_β-C_γ-C_ζ angle measures the corrected distortion. In addition, the direction of the distortion can be defined because $C_{\delta1}$ and $C_{\delta2}$ are consistently oriented. Here, the bend angle is defined by the corrected C_β-C_γ-C_ζ if $C_\zeta(x) < 0$, and 360-C_β-C_γ-C_ζ if $C_\zeta > 0$. If $C_{\delta1}$ and $C_{\delta2}$ are assigned randomly when the crystal structure is solved, the distribution of bend angles should center at 180°, which we find to be the case. Side-chain distortion angles from structures with multiple protein copies per asymmetric unit are averaged values.

Sequence Analysis. From a BLAST search of the National Center for Biotechnology Information database using the αLP sequence as query, 47 homologous protease sequences were retrieved. The sequences were aligned with ClustalW and a phylogenetic tree constructed using MacVector. Proteases were categorized based on sequence covariation patterns expected to determine the rotamer of Phe228 (2).

Cloning/Protein Production. αLP mutants were made using Quik-Change Site-Directed mutagenesis (Stratagene), and protein was expressed and purified according to published protocols (36).

Folding/Unfolding. Uncatalyzed folding of αLP was performed as described in Sohl et al. (5), except that the total protein concentration was ~4 μM and the assay was calibrated within each timepoint using standards of no protein and 25 pM WT-αLP to improve the precision of each measurement. Unfolding was measured by loss of tryptophan fluorescence (excitation 283 nm, emission 320 nm) as described (13). The unfolding rate of αLP at 0 °C was determined by extrapolation from data of Jaswal et al. (13). The denaturant binding model was used to model the unfolding behavior of the Repack and T181A mutants because it provides an accurate empirical fit of the curved data. Its use has been validated for WT-αLP, where it was established that the observed curvature was not due to transition state movement (13), but rather due to electrostatic effects from the ionic denaturant guanidine, as linear extrapolations from the nonionic denaturant urea yield the same rate constant (37). It is unknown why certain mutants have curvature in their unfolding denaturant titration, but it has been seen in multiple cases. Rate constants were converted to ΔG using standard transition state theory. Although the appropriate "preexponential factor" for an unfolding or folding reaction is controversial, $\Delta\Delta G^{\ddagger}$ is unaffected by the choice of a preexponential term.

Autolysis assays were performed at 0 °C and pH 8.0 (10 mM Tris). Unfolding was measured at 50 nM protein using intrinsic tryptophan fluorescence (283 nm, 322 nm). Inactivation at 5 μM was measured by loss of proteolytic activity toward the synthetic substrate Succinate-Ala-Pro-Ala-paranitroanilide as previously described (7) and by tryptophan fluorescence.

Crystallization/Structural Analysis. The crystallization of T181A and T181G αLP mutants was performed as described for WT-αLP (2). Crystals were in space group $P3_221$ with one molecule per asymmetric unit. However, the Repack mutant could only be crystallized under slightly different conditions (addition of 10 mM $CuCl_2$) and in a different space group $P6_122$ with two αLP molecules per asymmetric unit. Diffraction data were collected at the Advanced Light Source, Beamline 8.2.2, and processed in HKL2000 (38). For the Repack

mutant, the structure was solved using Molecular Replacement with WT-αLP (with residues 181, 199, 210, and 228 as alanine) as a starting model. The T181A and T181G mutants were solved with a starting model of WT-αLP (2) with residues 228 and 181 as alanine or glycine, depending on the mutant. Initial refinement was performed in CNS (39), and anisotropic B-factors were modeled using REFMAC (40). Structural alignments were made using Combinatorial Extension (41).

The distortion of Phe228 in TFPA has been described previously (6). The 1.85 Å structure of NAPase was reported previously (8), in which distortion of Tyr228 in NAPase could be detected even with planarity restraints during refinement (~4.2°). For this work, planarity restraints were removed in CNS (39), and the distortion in the two copies of NAPase in the asymmetric unit were averaged together. Distortion of Phe 228 in SGPB was measured using PDB ID code 2QAA (15).

ACKNOWLEDGMENTS. We thank Drs. J. Weissmann, L. Rice, M. Stratton, and C. Fuhrmann for comments on the manuscript and members of the Agard lab for helpful discussions and Dr. Cori Ralston at Advanced Light Source Beamline 8.2.2 for help with data collection. B.A.K. was supported by a Howard Hughes Medical Institute Predoctoral Fellowship and N.L.S. by a National Science Foundation Predoctoral Fellowship. This work was supported by the Howard Hughes Medical Institute.

1. Kang BS, Devedjiev Y, Derewenda U, Derewenda ZS (2004) The PDZ2 domain of syntenin at ultra-high resolution: Bridging the gap between macromolecular and small molecule crystallography. *J Mol Biol* 338:483–493.
2. Fuhrmann CN, Kelch BA, Ota N, Agard DA (2004) The 0.83 Å resolution crystal structure of alpha-lytic protease reveals the detailed structure of the active site and identifies a source of conformational strain. *J Mol Biol* 338:999–1013.
3. Baker D, Agard DA (1994) Kinetics versus thermodynamics in protein folding. *Biochemistry* 33:7505–7509.
4. Cunningham EL, Jaswal SS, Sohl JL, Agard DA (1999) Kinetic stability as a mechanism for protease longevity. *Proc Natl Acad Sci USA* 96:11008–11014.
5. Sohl JL, Jaswal SS, Agard DA (1998) Unfolded conformations of alpha-lytic protease are more stable than its native state. *Nature* 395:817–819.
6. Kelch BA, Agard DA (2007) Mesophile versus thermophile: Insights into the structural mechanisms of kinetic stability. *J Mol Biol* 370:784–795.
7. Jaswal SS, Sohl JL, Davis JH, Agard DA (2002) Energetic landscape of alpha-lytic protease optimizes longevity through kinetic stability. *Nature* 415:343–346.
8. Kelch BA, et al. (2007) Structural and mechanistic exploration of acid resistance: Kinetic stability facilitates evolution of extremophilic behavior. *J Mol Biol* 368:870–883.
9. Truhlar SME, Cunningham EL, Agard DA (2004) The folding landscape of Streptomyces griseus protease B reveals the energetic costs and benefits associated with evolving kinetic stability. *Protein Sci* 13:381–390.
10. Cunningham EL, Agard DA (2004) Disabling the folding catalyst is the last critical step in alpha-lytic protease folding. *Protein Sci* 13:325–331.
11. Truhlar SME, Agard DA (2005) The folding landscape of an alpha-lytic protease variant reveals the role of a conserved beta-hairpin in the development of kinetic stability. *Proteins* 61:105–114.
12. Salimi NL, Ho B, Agard DA (2010) Unfolding simulations reveal the mechanism of extreme unfolding cooperativity in the kinetically stable alpha-lytic protease. *PLoS Comput Biol* 6:e1000689.
13. Jaswal SS, Truhlar SME, Dill KA, Agard DA (2005) Comprehensive analysis of protein folding activation thermodynamics reveals a universal behavior violated by kinetically stable proteases. *J Mol Biol* 347:355–366.
14. Huang K, Lu W, Anderson S, Laskowski M, James MN (1995) Water molecules participate in proteinase-inhibitor interactions: Crystal structures of Leu18, Ala18, and Gly18 variants of turkey ovomucoid inhibitor third domain complexed with Streptomyces griseus proteinase B. *Protein Sci* 4:1985–1997.
15. Lee T-W, James MNG (2008) 1.2A-resolution crystal structures reveal the second tetrahedral intermediates of streptogrisin B (SGPB). *Biochim Biophys Acta* 1784:319–334.
16. Sielecki AR, et al. (1979) Protein structure refinement: Streptomyces griseus serine protease A at 1.8 A resolution. *J Mol Biol* 134:781–804.
17. Kitadokoro K, Tsuzuki H, Okamoto H, Sato T (1994) Crystal structure analysis of a serine proteinase from Streptomyces fradiae at 0.16-nm resolution and molecular modeling of an acidic-amino-acid-specific proteinase. *Eur J Biochem* 224:735–742.
18. Matthews BW, Nicholson H, Becktel WJ (1987) Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc Natl Acad Sci USA* 84:6663–6667.
19. Park C, Marqusee S (2004) Probing the high energy states in proteins by proteolysis. *J Mol Biol* 343:1467–1476.
20. Park C, Marqusee S (2005) Pulse proteolysis: A simple method for quantitative determination of protein stability and ligand binding. *Nat Methods* 2:207–212.
21. Rupley J (1967) Susceptibility to attack by proteolytic enzymes. *Methods Enzymol*.
22. Matouschek A, Kellis JT, Serrano L, Fersht AR (1989) Mapping the transition state and pathway of protein folding by protein engineering. *Nature* 340:122–126.
23. de los Rios MA, et al. (2006) On the precision of experimentally determined protein folding rates and phi-values. *Protein Sci* 15:553–563.
24. Ozkan SB, Bahar I, Dill KA (2001) Transition states and the meaning of Phi-values in protein folding kinetics. *Nat Struct Biol* 8:765–769.
25. Ventura S, et al. (2002) Conformational strain in the hydrophobic core and its implications for protein folding and design. *Nat Struct Biol* 9:485–493.
26. Cho J-H, Raleigh DP (2006) Denatured state effects and the origin of nonclassical phi values in protein folding. *J Am Chem Soc* 128:16492–16493.
27. Sauter NK, Mau T, Rader SD, Agard DA (1998) Structure of alpha-lytic protease complexed with its pro region. *Nat Struct Biol* 5:945–950.
28. Belasco JG, Knowles JR (1980) Direct observation of substrate distortion by triosephosphate isomerase using Fourier transform infrared spectroscopy. *Biochemistry* 19:472–477.
29. Patt SL, Dolphin D, Sykes BD (1973) The detection of substrate distortion by lysozyme: An application of NMR to the study of enzyme substrate reactions. *Ann N Y Acad Sci* 222:211–219.
30. Zou J, Kleywegt G, StÅhlberg J, Driguez H (1999) Crystallographic evidence for substrate ring distortion and protein conformational changes during catalysis in cellobiohydrolase Ce16A from Trichoderma reesei. *Structure* 7:1035–1045.
31. Li D, Stuehr DJ, Yeh S-R, Rousseau DL (2004) Heme distortion modulated by ligand-protein interactions in inducible nitric-oxide synthase. *J Biol Chem* 279:26489–26499.
32. Lyubimov AY, Heard K, Tang H, Sampson NS, Vrielink A (2007) Distortion of flavin geometry is linked to ligand binding in cholesterol oxidase. *Protein Sci* 16:2647–2656.
33. Maes EM, Roberts SA, Weichsel A, Montfort WR (2005) Ultrahigh resolution structures of nitrophorin 4: Heme distortion in ferrous CO and NO complexes. *Biochemistry* 44:12690–12699.
34. Olea C, Boon EM, Pellicena P, Kuriyan J, Marletta MA (2008) Probing the function of heme distortion in the H-NOX family. *ACS Chem Biol* 3:703–710.
35. Berkholz DS, Driggers CM, Shapovalov MV, Dunbrack RL, Karplus PA (2011) Nonplanar peptide bonds in proteins are common and conserved but not biased toward active sites. *Proc Natl Acad Sci USA* 109:449–453.
36. Mace JE, Wilk BJ, Agard DA (1995) Functional linkage between the active site of alpha-lytic protease and distant regions of structure: Scanning alanine mutagenesis of a surface loop affects activity and substrate specificity. *J Mol Biol* 251:116–134.
37. Jaswal SS (2000) Thermodynamics, Kinetics, and Landscapes in Alpha-Lytic Protease: A Role for Pro Regions in Kinetic Stability. (Univ of California, San Francisco).
38. Otwinowski Z (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
39. Brünger AT, et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54:905–921.
40. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240–255.
41. Shindyalov IN, Bourne PE (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng* 11:739–747.