Protein Science (1998), 7:96–104. Cambridge University Press. Printed in the USA.

Subdomain interactions as a determinant in the folding and stability of T4 lysozyme

MANUEL LLINÁS AND SUSAN MARQUSEE

Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California 94720

(RECEIVED May 19, 1997; ACCEPTED October 6, 1997)

Abstract

The folding of large, multidomain proteins involves the hierarchical assembly of individual domains. It remains unclear whether the stability and folding of small, single-domain proteins occurs through a comparable assembly of small, autonomous folding units. We have investigated the relationship between two subdomains of the protein T4 lysozyme. Thermodynamically, T4 lysozyme behaves as a cooperative unit and the unfolding transition fits a two-state model. The structure of the protein, however, resembles a dumbbell with two potential subdomains: an N-terminal subdomain (residues 13–75), and a C-terminal subdomain (residues 76–164 and 1–12). To investigate the effect of uncoupling these two subdomains within the context of the native protein, we created two circular permutations, both at the subdomain interface (residues 13 and 75). Both variants adopt an active wild-type T4 lysozyme fold. The protein starting with residue 13 is 3 kcal/mol less stable than wild type, whereas the protein beginning at residue 75 is 9 kcal/mol less stable, suggesting that the placement of the termini has a major effect on protein stability while minimally affecting the fold. When isolated as protein fragments, the C-terminal subdomain folds into a marginally stable helical structure, whereas the N-terminal subdomain is predominantly unfolded. ANS fluorescence studies indicate that, at low pH, the C-terminal subdomain adopts a loosely packed acid state. An acid state intermediate is also seen for all of the full-length variants. We propose that this acid state is comprised of an unfolded N-terminal subdomain and a loosely folded C-terminal subdomain.

Keywords: circular permutation; protein folding; subdomain; T4 lysozyme

The structural elements that dictate the stability and folding of small proteins are not well understood. For large, multidomain proteins, folding appears hierarchical; individual domains are relatively autonomous with differing stabilities and independent folding (Wetlaufer, 1973; Rose, 1979; Jaenicke, 1991). Thus, the folding of these proteins involves modular assembly coupled with cooperative association. It is unclear, however, if such a modular assembly process describes the folding and stability of small, single-domain proteins.

Evidence supporting the existence of modules within small, single-domain proteins has been difficult to evaluate due to the cooperative nature of proteins. The unfolding transitions of small proteins are usually well-described by the two-state (native ↔ unfolded; N ↔ U) assumption, suggesting that small proteins fold as a cooperative unit (Kim & Baldwin, 1982). However, recent examples of subdomains within single-domain proteins (Oas & Kim, 1988; Staley & Kim, 1990) as well as hydrogen exchange studies of protein folding (Englander et al., 1996) imply that small proteins may indeed be modular. In this study, we have used the small, single-domain protein, T4 lysozyme, to investigate the potential existence and roles of such subdomains.

T4 lysozyme is an ideal system for investigating the role of subdomains. It is a small (164 amino acids) protein with a bipartite structure made up of an amino-terminal α/β lobe and a completely α-helical carboxy-terminal lobe. The lobes are connected by a long central α-helix (Weaver & Matthews, 1987) (Fig. 1). T4 lysozyme has served as a model system for extensive studies on protein folding and stability (Alber, 1989; Poteete & Hardy, 1994; Matthews, 1995, 1996). Despite its apparent dumbbell shape, equilibrium studies show that the protein is cooperatively folded in an apparent two-state (N ↔ U) transition (Elwell & Schellman, 1975). Thermodynamically, therefore, T4 lysozyme behaves as a single-domain protein, although the independence and interdependence of the two lobes remains unclear.

Despite its cooperative thermodynamic behavior, several studies suggest that, in the native state of T4 lysozyme, the N-and C-terminal lobes are somewhat independent. A comparison of the crystal struc-

Reprint requests to: Susan Marqusee, Department of Molecular and Cell Biology, 229 Stanley Hall, University of California at Berkeley, Berkeley, California 94720; e-mail: marqusee@zebra.berkeley.edu.

Abbreviations: WT*, cysteine-free T4 lysozyme; 13cpT4L*, cysteine-free circular permutation of T4 lysozyme beginning at residue 13; 75cpT4L*, cysteine-free circular permutation of T4 lysozyme beginning at residue 75; xT4L*, cysteine-free T4 lysozyme with a C-terminal linker; EPR, electron paramagnetic resonance; GdmCl, guanidinium chloride; ANS, 8-anilino-1-naphthalene sulfonate.

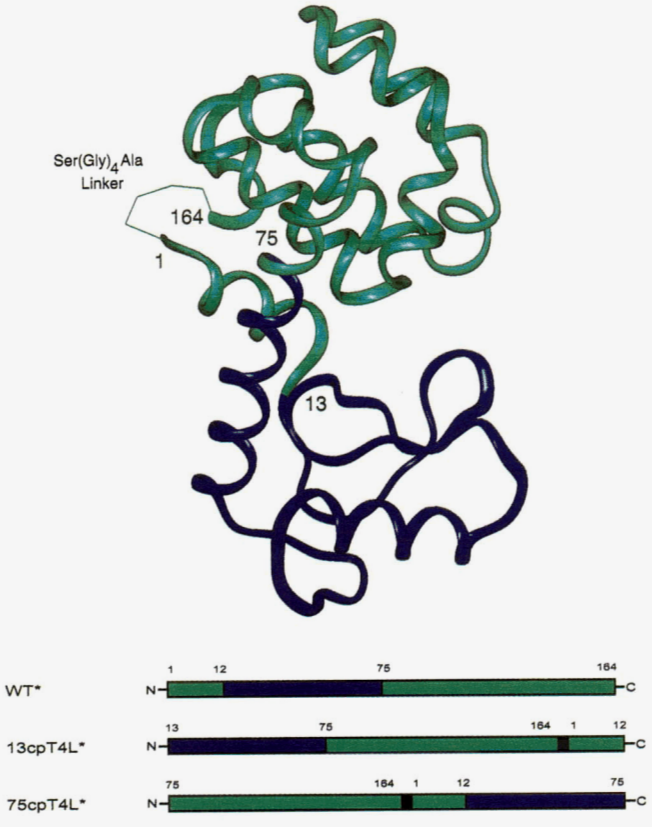


Fig. 1. Ribbon diagram and primary sequence representation of T4 lysozyme variants WT*, 13cpT4L*, and 75cpT4L*. The N-terminal subdomain (Leu 13–Val 75) is depicted in blue and the C-terminal subdomain (Val 75–Leu 164 + Met 1–Leu 13) is in green. The wild-type N and C termini (residues 1 and 164) have been joined by a linker to illustrate the connectivity in the permuted lysozymes.

tures of the mutants M6I, I3P, and wild-type shows that the angle formed between the lobes varies from 9.1° to 32.9° (Faber & Matthews, 1990; Dixon et al., 1992). NMR studies of several site-specific mutants of T4 lysozyme show changes in amide hydrogen exchange rates and proton chemical shifts that are localized only to the lobe containing the mutation (Anderson et al., 1993). More recently, EPR studies have shown that the relative motion of these lobes is important for the enzyme's biological activity (McHaourab et al., 1996, 1997). All of these observations suggest that T4 lysozyme has two relatively independent subdomains. In contrast, the early kinetic folding intermediate of T4 lysozyme shows some structure formation in both lobes, implying cooperativity between these two regions (Lu & Dahlquist, 1992). Thus, it is unclear whether T4 lysozyme is truly a single-domain protein or whether it is comprised of two smaller subdomains.

We have used a protein dissection approach to investigate the degree of independence of the N- and C-terminal lobes of T4 lysozyme. Individual subdomains were defined by minimizing the loss of native contacts in the X-ray structure of the native protein. A cysteine-free variant of T4 lysozyme (WT*) whose structure and stability is similar to the wild-type protein was used as the parent molecule (Matsumura & Matthews, 1989). The roles of the individual subdomains were characterized by creating circular permutations of the full-length T4 lysozyme as well as characterizing the subdomains as isolated fragments.

Results

Subdomain modeling

The crystal structure of T4 lysozyme reveals a bipartite molecule with two apparent subdomains (Weaver & Matthews, 1987) (Fig. 1). To define the precise boundaries of these potential subdomains, we utilized the program CONTACTS (kindly provided by Terry Oas) on the crystal structure of wild-type lysozyme (2lzm.pdb). This program evaluates the atomic coordinates and identifies all potential van der Waals and hydrogen bonding interactions between any two defined sets of residues. Using the results from CONTACTS, we mapped the regions with maximal intrasubdomain interactions and minimal loss of possible intersubdomain interactions. The two subdomains of T4 lysozyme were defined as follows: the N-terminal subdomain consists of residues 13–75 and the C-terminal subdomain consists of residues 76–164 + 1–12 (Fig. 1). Although helix A (residues 4–10) resides at the N terminus of the protein sequence, these residues comprise part of the C-terminal hydrophobic core of T4 lysozyme and were classified as part of the C-terminal lobe or subdomain. Therefore, the C-terminal subdomain of T4 lysozyme includes the last 90 residues of the protein in addition to the first 13 residues of the protein (helix A). In a study in which he identified hydrophobic clusters in proteins, Zehfus (1995) has picked similar regions of T4 lysozyme as potential compact units.

Tolerance to intersubdomain breaks

In order to determine the independence of the subdomains, we first uncoupled them by constructing two different circularly permuted variants of T4 lysozyme, placing the new termini in the subdomain interface (see Materials and methods). In T4 lysozyme, the N and C termini are only ~10.5 Å apart, making it amenable to circular permutation (Zhang et al., 1993). One permutation, 13cpT4L*, begins at residue 13 and the second, 75cpT4L*, begins at residue 75. In both variants, the wild-type N and C termini were connected by a six-amino acid linker, $Ser(Gly)_4Ala$. This same linker was used by Zhang et al. (1993) to create a circular permutation of T4 lysozyme at residue 37. Our permutations, 13cpT4L* and 75cpT4L*, create protein sequences with contiguous arrangements of the two putative subdomains. 13cpT4L* (residues 13–164 + linker + 1–12) begins with the N-terminal subdomain followed by the C-terminal subdomain, whereas the permutation beginning with Val 75 places the C-terminal subdomain before the N-terminal subdomain (75–164 + linker + 1–75).

Starting the protein at residue 13 (13cpT4L*) yielded a protein with a structure and stability similar to the parent protein (WT*). The far-UV CD spectra of 13cpT4L* and WT* (Fig. 2) indicate that the permuted protein assumes an $\alpha+\beta$ fold resembling WT*. The unfolding transition of 13cpT4L* is cooperative by both thermal and chemical denaturation (Fig. 3). The free energy of unfolding, $\Delta G_{unf}(H_2O)$, is 11.1 kcal/mol, only marginally destabilized relative to WT* under the same conditions (14.1 kcal/mol) (Table 1). The permuted protein was also active in T4 lysozyme turbidity assays (data not shown) (Tsugita & Inouye, 1968). Together, these data demonstrate that the permutation at Leu 13 produces a stable, well-structured variant of T4 lysozyme.

Starting the protein with the C-terminal subdomain (75cpT4L*) resulted in a much greater disruption to the protein. Although

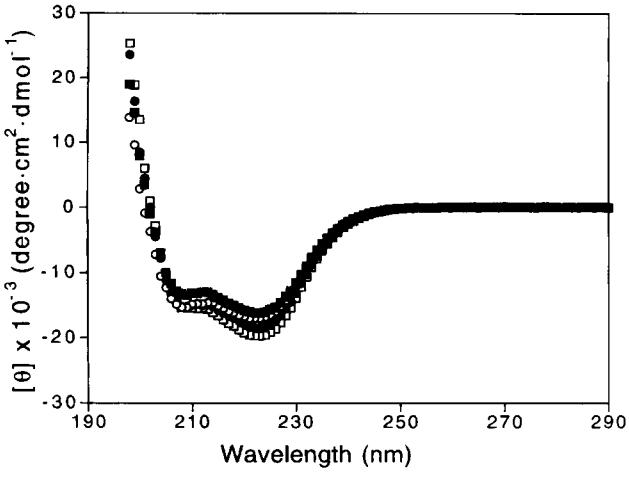


Fig. 2. Far-UV CD spectra for WT* (open squares), 13cpT4L* (closed circles), 75cpT4L* (closed squares), and H31N 13cpT4L* (open circles). All CD spectra were recorded at 4 °C with 50 μg/mL protein in 50 mM potassium phosphate, pH 5.75, in a 1.0-cm pathlength cell.

well-folded with a cooperative unfolding transition (Figs. 2, 3), this permutation was greatly destabilized; the free energy of unfolding is 4.8 kcal/mol, a destabilization of ~6 kcal/mol from 13cpT4L* (Table 1). 75cpT4L* is active, but only about 10% as active as the wild-type protein in the same turbidity assays. Therefore, although the ordering of the subdomains along the protein sequence plays an essential role in determining the stability of the protein, it only minimally affects the overall fold of the protein. This led us to question if the local interactions within each subdomain are sufficient to encode for a specific fold. In other words, do the N- and C-terminal subdomains form autonomous folding units?

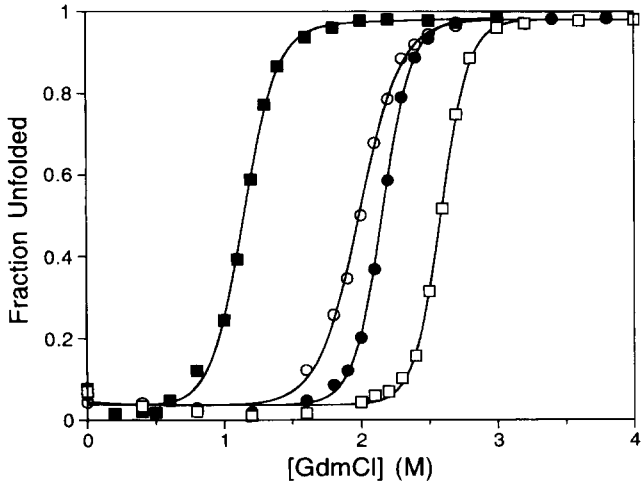


Fig. 3. Guanidine unfolding transitions for WT* (open squares), 13cpT4L* (closed circles), H31N 13cpT4L* (open circles), and 75cpT4L* (closed squares). Unfolding transitions were measured at 4 °C with 50 μg/mL protein in 50 mM potassium phosphate, pH 5.75, in a 1.0-cm pathlength cell. Raw data were converted to fraction unfolded by fitting data to a two-state model (see Materials and methods).

Table 1. *Protein stability data*

Protein	T_m[a] (°C)	$\Delta G_{unf}(H_2O)$ (kcal/mol)	C_m (M)
WT*	69.9	14.1 ± 0.8[b]	2.6
13cpT4L*	64.0	11.1 ± 0.5[b]	2.2
H31N 13cpT4L*	60.2	7.3 ± 0.4[b]	2.0
75cpT4L*	48.4	4.6 ± 0.6[b]	1.2
C-terminal subdomain	31.5	2.1 ± 0.3[c]	2.1

[a] H_2O at pH 4.5.
[b] 50 mM potassium phosphate, pH 5.75, in GdmCl.
[c] 50 mM potassium phosphate, pH 5.75, in urea.

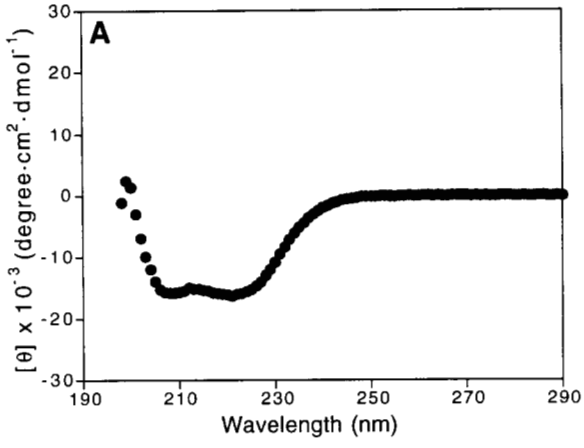
Folding of the isolated C-terminal lobe

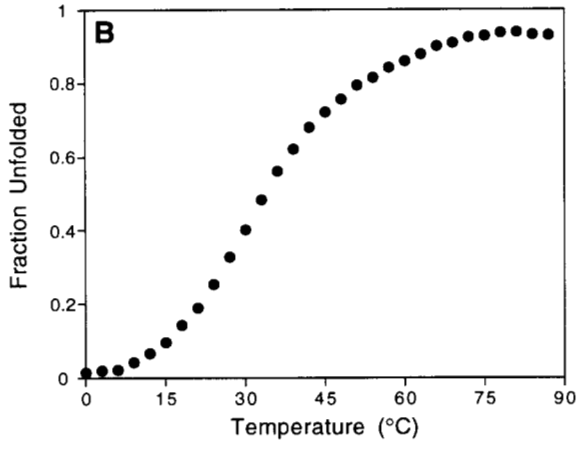
The protein fragment corresponding to the C-terminal subdomain (a start methionine followed by residues 75–164, a Ser(Gly)$_4$Ala linker, and residues 1–12) was cloned, overexpressed in *Escherichia coli*, and purified to homogeneity (see Materials and methods). Far-UV CD spectroscopy indicated that the isolated fragment folds into a predominantly α-helical structure with characteristic minima at 222 nm and 208 nm (Fig. 4A); in the full-length protein, this region is also highly helical (Weaver & Matthews, 1987). The CD signal was independent of protein concentration over the range of 25–150 μg/mL, suggesting that the isolated fragment folds as a monomer (data not shown). Thermal denaturation of the isolated subdomain monitored by the CD signal at 222 nm revealed a broad transition with lower cooperativity than is normally seen for natural proteins (Fig. 4B). The urea-induced denaturation was also broad (Fig. 4C), yielding a free energy of unfolding of 2.1 kcal/mol (Table 1). Therefore, despite our design, which preserved greater than 98% of the native contacts, including the major hydrophobic core of T4 lysozyme, the C-terminal subdomain is only marginally stable. It is, however, capable of forming an independent folding unit.

The hydrophobic-binding dye ANS was used to assess the tertiary packing of the isolated C-terminal subdomain. ANS is known to bind loosely packed hydrophobic interiors of molten globule states as well as kinetic folding intermediates of natural proteins (Kuwajima, 1989; Ptitsyn et al., 1990). At pH 7.5, no ANS binding in the C-terminal subdomain was detected (Fig. 5). As the pH was lowered, ANS was able to penetrate into the molecule and the fluorescence increased. The fluorescence increased until ~ pH 3.0, where the protein began to unfold. These data indicate that, under acidic conditions, the protein forms a loosely packed conformation reminiscent of a molten globule. NMR evidence further supports this notion: a 1D proton NMR spectrum at pH 4.5 showed limited dispersion of the peaks in both the aromatic and amide regions (data not shown). Therefore, the C-terminal subdomain adopts a marginally stable conformation at neutral pH, which is perturbed under more acidic conditions.

Folding of the N-terminal subdomain in isolation

The N-terminal subdomain (residues 13–76) encompasses the α/β lobe of T4 lysozyme. Because an expression plasmid containing a gene for this fragment showed no detectable overexpression (data





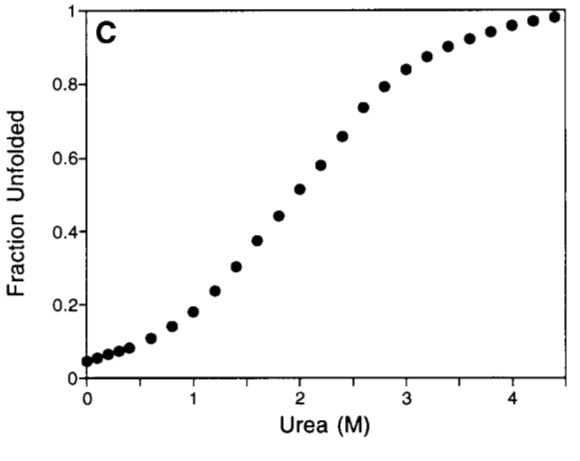


Fig. 4. C-terminal subdomain is folded and stable in isolation. **A:** Far-UV CD spectrum indicates high helical content (4 °C, 50 μg/mL protein, 50 mM potassium phosphate, pH 5.75). **B:** Reversible thermal unfolding is cooperative with a T_m of 32 °C (4.0 μM protein, pH 4.6). **C:** Urea denaturation is cooperative with a C_m of 2.1 M urea and $\Delta G_{unf}(H_2O)$ of 2.1 kcal/mol (4 °C, 50 μg/mL protein, 25 mM potassium phosphate, pH 5.75, 50 mM KCl).

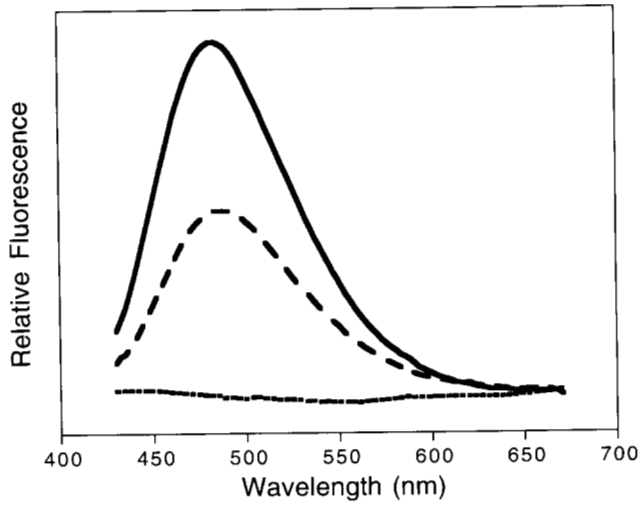


Fig. 5. ANS binding to the C-terminal subdomain: pH 7.4 (dotted line), pH 3.4 (solid line), and pH 1.2 (dashed line). All spectra were taken with a 250:1 molar ratio of ANS:protein, 2.0 μM protein in 50 mM potassium phosphate.

not shown), the fragment was synthesized chemically on an automated peptide synthesizer (see Materials and methods). Although this subdomain was also designed to retain most of the native interactions, the far-UV CD spectrum indicated that the N-terminal peptide was predominantly unfolded (Fig. 6). ANS binding was not detected at any pH, indicating that the isolated N-terminal subdomain does not form a loosely packed hydrophobic interior or molten globule-like state (data not shown). Therefore, without a contiguous C-terminal subdomain, the N-terminal subdomain appears incapable of folding. Some communication between the subdomains seems essential for the proper folding and stability of this region in the full-length protein.

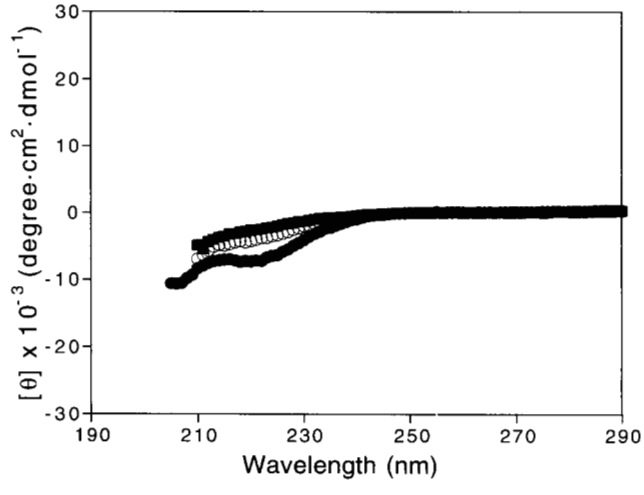


Fig. 6. Far-UV CD spectra of the isolated N-terminal subdomain (4 °C, 50 μg/mL protein, pH 4.4, 1.0-cm pathlength cell): 0.0 M GdmCl (closed circle), 0.5 M GdmCl (open circle), 1.0 M GdmCl (closed square).

The acid state of T4 lysozyme

Because the isolated C-terminal subdomain appears to form a loosely packed acid state, we examined the behavior of this region in the context of the full-length proteins. For all full-length lysozymes studied (WT*, 13cpT4L*, 75cpT4L*), a partially folded acid state was detected by both CD and ANS binding. [As a control for linker effects, an extended version of T4 lysozyme (xT4L*) (Zhang et al., 1993) was also probed.] All of the lysozymes bound ANS maximally around pH 1.2 and showed quenched fluorescence at pH 7.5 (Fig. 7). GdmCl-induced unfolding monitored by fluorescence at pH 5.75 indicated a two-state transition with an isosbestic point at ~340 nm. In stark contrast, the unfolding transition at pH 1.2 showed a clear hyperfluorescence at low concentrations of guanidinium, with a maximal intensity at 1.0 M GdmCl. At this pH, no isosbestic point was seen, indicating non-two-state unfolding (Fig. 8). These results correlate well with those for the isolated C-terminal subdomain and suggest that, in the full-length lysozymes, the C-terminal subdomain may become less well packed under acidic conditions, whereas the N-terminal subdomain selectively unfolds.

In full-length WT*, the N-terminal subdomain contains a salt bridge (H31–D70) known to stabilize the protein by 3–5 kcal/mol (Anderson et al., 1990). Disruption of this electrostatic interaction at low pH could selectively unfold the N-terminal subdomain. To test this possibility, we made the mutation H31N in 13cpT4L* and examined the fold and stability of this variant (Figs. 2, 3). If the salt bridge in the N-terminal subdomain contributes to the structure and stability of the acid state, this mutation should cause a significant destabilization of the acid state. At pH 5.75, the mutation destabilized the protein by ~3.0 kcal/mol (Table 1), which is very similar to the effect of the same mutation in WT* (Anderson et al., 1990). At low pH (pH 1.2), however, the stability and ANS binding profile of 13cpT4L* were unaffected by the mutation (data not shown). These results are consistent with our hypothesis that the low-pH conformation of T4 lysozyme is composed mainly of the C-terminal subdomain. This acid state may represent an intermediate in the folding of the protein.

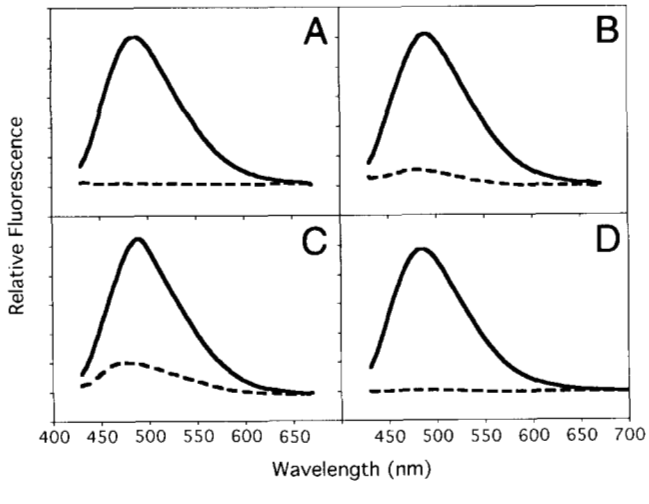


Fig. 7. ANS binding to the lysozymes at pH 1.2 (solid line) and pH 7.4 (dashed line). **A:** WT*. **B:** xT4L*. **C:** 13cpT4L*. **D:** 75cpT4L*. All spectra were taken with a 250:1 molar ratio of ANS:protein, 2.0 μM protein.

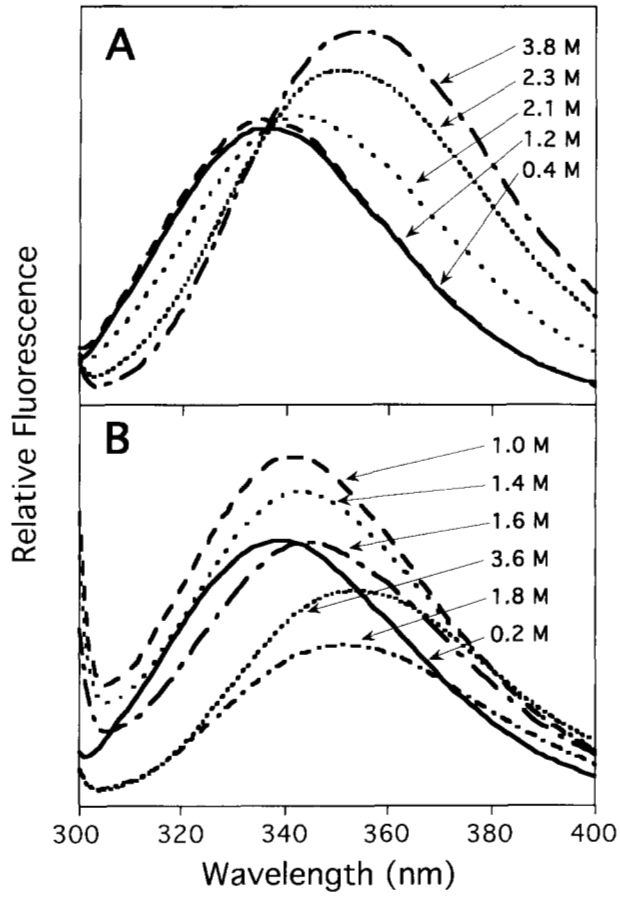


Fig. 8. Fluorescence spectra of 13cpT4L* at varying GdmCl concentrations. **A:** pH 5.75. **B:** pH 1.2 (4 °C, 50 μg/mL protein, 50 mM potassium phosphate).

Discussion

Subdomain assembly as a model for folding

Experimentally, it has been difficult to investigate the potential modular nature of small, single-domain proteins. Unlike the folding of large multidomain proteins, which appear to fold in a hierarchical process, most single-domain proteins undergo highly cooperative two-state folding transitions with no evidence for other stable conformations (Kim & Baldwin, 1982). Here, we have found evidence for an autonomous subdomain within the small, single-domain protein T4 lysozyme.

Previous studies on fragments of T4 lysozyme have focused on small isolated units of secondary structure that are only structured in the presence of extremely stabilizing solvents such as trifluoroethanol (McLeish et al., 1993, 1994; Najbar et al., 1995). We have taken advantage of the bipartite structure of T4 lysozyme to examine larger subdomains and the coupling between them. Our results indicate that intersubdomain interactions can be key determinants for the structure and stability of single-domain proteins. Isolated subdomains, therefore, serve as valuable models for studying the features necessary to give a particular protein its fold.

Autonomous folding of the C-terminal subdomain

The isolated C-terminal subdomain was designed to retain a majority of its native contacts, including the major hydrophobic core of T4 lysozyme. Indeed, the sequence of the C-terminal subdomain of T4 lysozyme is sufficient for folding into a highly helical conformation, which unfolds cooperatively. Despite being well folded, the stability of this subdomain is only 2.1 kcal/mol, ~12 kcal/mol less stable than the full-length protein. Our ANS fluorescence results indicate that the hydrophobic interior of this subdomain is sequestered from solvent. Without a detailed structure determination, however, it remains unclear how native-like this fold is.

Folding of the N-terminal subdomain is dependent on the C-terminal subdomain

Although this segment of the protein (residues 13–77) forms a compact α/β fold in the intact protein, the isolated fragment never significantly populates any specific conformation. The N-terminal subdomain is therefore not an autonomous folding unit and does not form a well-defined structure in isolation. Furthermore, addition of the C-terminal subdomain (1:1) does not result in any increase in structure (data not shown). Although this is not completely unexpected due to our minimization of intersubdomain interactions, it strongly suggests that covalent connectivity along the interconnecting C-helix between the subdomains is important for the stability of the N-terminal lobe of T4 lysozyme.

Importance of intersubdomain connectivities for subdomain folding

The isolated subdomain results highlight the importance of the communication between the subdomains in the intact protein. This result is further emphasized by the circular permutations. Both permutations studied, 13cpT4L* and 75cpT4L*, alter the connectivity between the subdomains, and both show native-like folds. There is, however, greater than 6 kcal/mol difference in their stabilities, suggesting that backbone intersubdomain connectivity is more important to the stability of the protein than to the specificity of the T4 lysozyme fold.

Because both circularly permuted lysozymes are active, it is reasonable to assume that their structures are very similar to the wild-type protein, with expected local perturbations near the new termini. It is therefore particularly surprising that the permutation 13cpT4L* shows considerable activity, because the essential catalytic residue, Glu 11, is only one residue away from the new C terminus (Shoichet et al., 1995). 75cpT4L*, which does not place new termini near the active site, is actually less active than 13cpT4L*. In this variant, the break occurs in the long C-helix, which connects the two subdomains. This central helix has been shown to be quite tolerable to mutation: even mutants containing helix-breaking substitutions, such as proline, form folded, active, but significantly destabilized proteins (Rennell et al., 1991; Sauer et al., 1992). Interestingly, a hinge-bending motion, which is crucial to the activity of T4 lysozyme, occurs near residue 75 in the C-helix. Therefore, the decrease in activity for 75cpT4L* may result from a loss in the hinge bending due to the increased flexibility in this region.

Alber and coworkers first noted that permutation of T4 lysozyme can affect the stability, but not the overall fold of the protein when they created a permutation starting at residue 37 (Zhang et al., 1993). Proline 37 is in an exposed loop of T4 lysozyme, and therefore the ability of the backbone to accommodate this disruption is not surprising. Our permutations, however, are directed between the structural lobes of T4 lysozyme. Both are more destabilizing than permutation at Pro 37, especially in the case of 75cpT4L*, whose new termini are placed in the central C-helix. However, the conclusion is similar in that the ends of a protein can be moved within a protein and minimally perturb the fold while having a dramatic effect on the stability of the protein. In the *de novo* design of stable proteins, therefore, the location of the chain termini is an important consideration and needs to be investigated in more detail. This is particularly important in the case of non-autonomous folding units, as seen here for T4 lysozyme, where the coupling between subdomains is essential to provide a framework for proper stability and folding.

The C-terminal subdomain as an intermediate

In our study, various lines of experimental evidence indicate that there is an observable intermediate state of T4 lysozyme that exists at low pH. This acid state resembles a classic molten globule in two ways: (1) it contains a high degree of secondary structure, and (2) it binds ANS. This acid state is seen for all of the lysozyme variants in this study. Interestingly, the transition to this low-pH state occurs at higher pHs for 13cpT4L* and 75cpT4L* than for WT*, implying that the native state is selectively destabilized relative to the intermediate in the permuted proteins. Our data suggest that this acid intermediate is comprised of a predominantly unfolded N-terminal subdomain with a loosely packed C-terminal subdomain.

The suggestion of a single domain forming the acid state is not unique to T4 lysozyme, but is reminiscent of two very well-characterized structural homologues of T4 lysozyme: α-lactalbumin and hen egg white lysozyme. The structures of all three of these "single-domain" proteins contain an alpha and a beta subdomain. The only region of the α-lactalbumin acid state that shows protection to amide hydrogen exchange is the α-domain (Schulman et al., 1995), and a fragment corresponding to this domain folds when oxidized (Peng & Kim, 1994). In addition, the folding pathways elucidated for α-lactalbumin and hen egg white lysozyme indicate that the α-domain plays a key role in folding (Radford et al., 1992; Kuwajima, 1996). Similarly, we suggest that the structure in the acid state of T4 lysozyme arises from the C-terminal α-domain and, by analogy, this region of the protein may serve as a scaffold upon which the N-terminal β-domain folds.

Kinetic refolding studies on T4 lysozyme, however, demonstrate that specific regions in both the N- and C-terminal lobes of the protein form structure early (Lu & Dahlquist, 1992). Therefore, the folding of T4 lysozyme involves either a rapid collapse to a non-native structure, or the two lobes fold independently at similar times. A kinetic analysis of the various different circular permutations of T4 lysozyme would allow a direct elucidation of the importance of the subdomain connectivity to the folding pathway.

Summary

In the subdomain model of protein folding, small regions of local structure form early, followed by association to form the final structure of the protein. In this study, we have investigated the relationship between potential subdomains of T4 lysozyme. Although our data do not test this model directly, we have found that

the isolated C-terminal subdomain forms a stable region of the protein. The N-terminal subdomain, on the other hand, is unstable in isolation. In the context of the native protein, the C-terminal subdomain appears to be crucial for conferring stability to the N-terminal subdomain, presumably through the C-helix, which provides a communication link between the subdomains of the protein.

Materials and methods

Plasmid construction and protein purification

*13cpT4L**

Circular permutation of proteins can generally be performed on any protein whose N and C termini lie close together by linking the ends and placing new termini within the protein (Heinemann & Hahn, 1995). For T4 lysozyme, a method has been developed by which a gene for any circular permutation can be easily generated (T. Alber, pers. comm.). The procedure is described in detail in Figure 9. Using this method, a gene encoding 13cpT4L* (residues 13–164, a Ser(Gly)$_4$Ala linker, and residues 1–12) was cloned and ligated into pCC101, a derivative of the T7 overexpression vector pAED4 (plasmid details available upon request). The new vector, pML107, was transformed into BL21 (DE3)/plysS, a T7 expression strain (NOVAGEN). Cells were grown in Luria Broth with 200 μg/mL ampicillin to an optical density of 0.5–0.8, at which point the cells were induced with 1.0 mM isopropyl-β-D-thiogalactopyranoside. After 3-h growth, cells were harvested by centrifugation. Cell pellets were resuspended in 10 mM Tris, pH 8.0/0.1 mM EDTA/20 mM NaCl and lysed by sonication. After DNase I (bovine pancreas grade II BM) treatment for 1 h at 4 °C, the soluble fraction was dialyzed against 50 mM Tris, pH 8.0. The dialysate was then purified over a CM-Sepharose column (Pharmacia CCL-6B-100) using a linear gradient of 50 mM Tris, pH 8.0/10 mM NaCl against 50 mM Tris, pH 8.0/300 mM NaCl (Poteete et al., 1991). The protein was further purified on a Vydac semi-preparative C_{18} reversed-phase column (250 mm × 10 mm) using a linear gradient on a Shimadzu LC10A series HPLC system.

*xT4L**

This extended variant of WT* has been reported previously (Zhang et al., 1993). The construct contains the entire WT* lysozyme sequence followed by the C-terminal Ser(Gly)$_4$Ala linker. The extended wild-type sequence was PCR cloned from a plasmid carrying WT*, pHSe5 (a gift from T. Alber), with one of the primers containing the linker sequence. The PCR product was digested and ligated into pCC101 (as above). The resulting vector, pML108, was transformed into BL21 (DE3)/plysS. Cells contain-

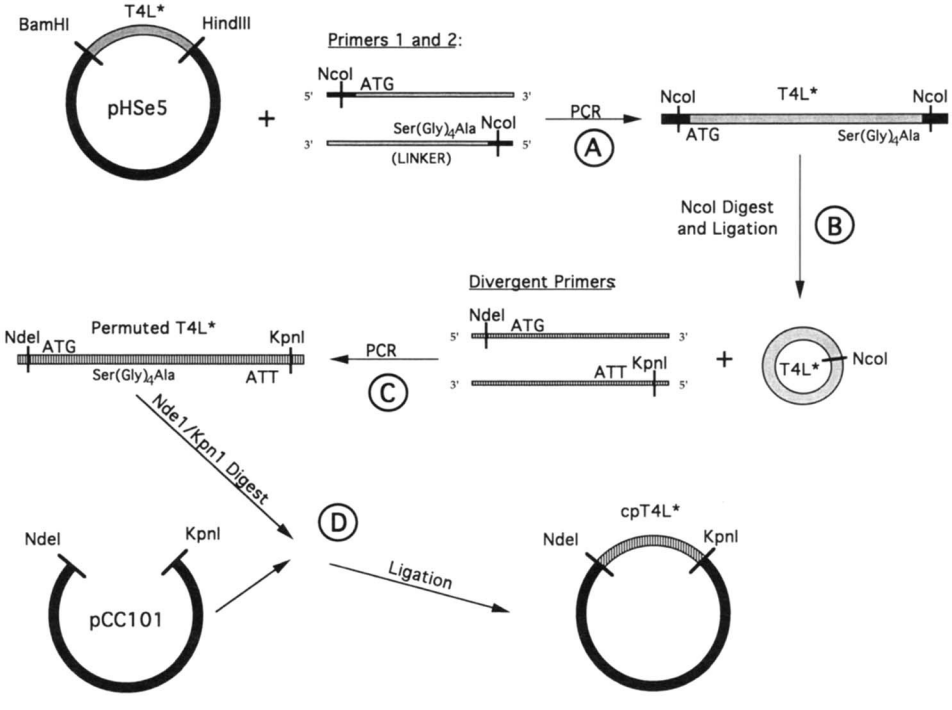


Fig. 9. Methodology for cloning of any circular permutation of T4 lysozyme. **A:** By PCR, an extended WT* T4 lysozyme sequence is produced from a plasmid (pHSe5) carrying the WT* gene (C54T, C97A) (a gift from T. Alber). The product is flanked on either side by *Nco* I restriction endonuclease sites and encodes the WT* sequence plus 18 extra bases encoding a C-terminal Ser(Gly)$_4$Ala linker. The use of *Nco* I sites introduced an N2D mutation previously shown to have no significant effect on the stability of the wild-type protein (Nicholson et al., 1991). **B:** Digestion of this DNA fragment with *Nco* I followed by ligation resulted in a 510-base pair mini-circle of DNA encoding only WT* plus the C-terminal linker conjoined to the Met 1 of the gene. **C:** Divergent PCR primers (one containing an *Nde* I site the other containing a *Kpn* I site) complementary to the new desired ends for a permuted lysozyme were then used to PCR clone a gene encoding all of the WT* amino acids as well as the linker. **D:** Ends of the PCR product were digested with *Nde* I and *Kpn* I and were ligated into pCC101, a derivative of the T7 overexpression vector pAED4 (plasmid details available upon request).

ing this plasmid were cultured and purified as described for 13cpT4L*.

*WT**

The wild-type sequence was PCR cloned from a vector containing the WT* sequence, pHSe5 (a gift from T. Alber). The PCR product was digested and ligated into pCC101 (as above). The resulting vector, pML109, was transformed into BL21 (DE3)/plysS. Cells containing this plasmid were cultured and purified as described for 13cpT4L*.

*75cpT4L**

This circular permutation was created by the PCR methodology described for 13cpT4L* and starts with residues 75–164, a Ser(Gly)$_4$Ala linker, and ends with residues 1–75. The gene was subcloned into the *Nde* I/*Kpn* I sites of pCC101 to give pML112. This plasmid was then transformed into BL21 (DE3)/plysS cells. Cultures were grown and induced as described above. The cell pellets were lysed in 10 mM Tris, pH 8.0/0.1 mM EDTA/20 mM NaCl by sonication. After DNase I (grade II BM) treatment for 1 h at 4 °C and centrifugation, the protein was found to be in inclusion bodies. The insoluble pellet was washed with 50 mM Tris, pH 8.0/1.0 mM EDTA/1% (v/v) Nonidet P-40/1% (w/v) deoxycholic acid followed by sonication to aid in solubilizing lipid membranes (Staley & Kim, 1992). Following centrifugation, the pellet was resuspended in 0.2 M Tris, pH 8.0/6.0 M GdmCl. The denatured protein was initially purified over a G-200 Sephadex column (Pharmacia) followed by reverse-phase HPLC purification as described for 13cpT4L*.

*H31N 13cpT4L**

pJKHO101 was created by the Kunkel method of site-directed mutagenesis (Sambrook et al., 1989). When overexpressed in BL21 (DE3)/plysS cells, the protein was insoluble in inclusion bodies. The protein was purified as described for 75cpT4L*.

C-terminal subdomain

This subdomain of T4 lysozyme consists of an initiator methionine plus Val 75–Leu 164 from WT* followed by the Ser(Gly)$_4$Ala linker ending in Met 1–Leu 13 from WT*. Divergent primers were used to PCR clone this fragment from a clone of cpT4L (cpT4L is a circularly permuted T4 lysozyme, which was a gift from T. Alber) (Zhang et al., 1993). The *Nde* I/*Eco*R I-containing product was partially digested and ligated into pCC101. The resulting plasmid, pML101, was transformed into BL21 (DE3)/plysS cells. Cultures were grown and induced as described for 13cpT4L*. The protein was purified from inclusion bodies, as described for 75cpT4L*.

N-terminal subdomain

The N-terminal subdomain encompasses Leu 13–Arg 76. Complete synthesis of this peptide was performed on an Applied Biosystems 431 automated peptide synthesizer using standard FMOC chemistry. For purification, 20 mg of crude peptide were dissolved in 1.0 mL 50% acetonitrile followed by the addition of 2.5 mL water. Purification was accomplished by reverse-phase HPLC on a Vydac semipreparative C_{18} column (250 mm × 10).

All genes used in this study were sequenced by the Sanger method. The purities of all expressed proteins were determined to be greater than 95% by silver-stained SDS-PAGE (data not shown). Confirmation of the expected product molecular weight was also verified by electrospray ionization mass spectrometry. Protein concentrations were determined by UV absorption spectroscopy at 280 nm. For the C-terminal and N-terminal subdomains, extinction coefficients of 21,750 M^{-1} cm^{-1} and 4,350 M^{-1} cm^{-1}, respectively, were used (Edelhoch, 1967; Brandts & Kaplan, 1973). For all full-length lysozymes, the value of 23,424 M^{-1} cm^{-1} was used (Tsugita & Inouye, 1968). The numbering system of residues used in this paper corresponds to residues 1–164 of WT*. All of the fragments and full-length proteins carry an N2D mutation except in the case of the WT* protein.

CD spectrometry

Far-UV CD experiments were performed on an Aviv 62DS spectrophotometer equipped with a Peltier temperature-controlled sample holder. Unfolding/refolding experiments were performed by monitoring the CD signal at 222 nm as a function of temperature or denaturant. Chemical denaturation experiments were performed at 4 °C using protein samples at 50 μg/mL in 50 mM potassium phosphate at either pH 5.75 or pH 1.2. Samples at varying guanidine concentrations were allowed to equilibrate at least 15 h before taking readings. For thermal denaturation, 3-min equilibration times were used with 1-min signal averaging at 222 nm. Thermal denaturation was generally greater than 90% reversible, except for protein samples in water, for which completely reversible thermal melts were observed. All unfolding data were fit using a two-state model and linear extrapolation (Santoro & Bolen, 1988) using the program Kaleidagraph (Synergy Software, Reading, Pennsylvania).

Fluorescence

Denaturation experiments were performed on a Perkin Elmer LS-50B luminescence spectrophotometer with a 1.0-cm pathlength at 4 °C. The sample temperature was controlled by a methanol/water bath. Data was taken by exciting at 295 nm and collecting emission data every 0.5 nm between 300 and 400 nm with the slit widths at 5.0.

ANS fluorescence

The hydrophobic dye ANS was used to monitor the accessibility of hydrophobic regions to solvent. Measurements were performed on a Perkin Elmer MPF-44B fluorescence spectrophotometer. Protein concentrations were all at 2.0 μM with ANS at 500 μM. The excitation wavelength was 405 nm, with emission spectra collected from 430–650 nm using a 0.4-cm pathlength cell.

Activity assay

A turbidity assay, which measures the degradation of crude cell walls, was used to detect activity (Tsugita & Inouye, 1968). For each assay, 50 pmol of protein were added to 1.0 mL of a 0.6-mg/mL suspension of lyophilized NapIV cells dissolved in 50 mM Tris, pH 7.5/1.0 mM EDTA. The decrease in turbidity was monitored at 450 nm every 10 s for 2 min.

Acknowledgments

This work was supported by a grant from the NIH (GM50945) and the Arnold and Mabel Beckman Foundation. We thank D. King for the syn-

thesis of the N-terminal subdomain and for mass spectrometric analysis; A. Glazer and J. Kirsch for use of their fluorimeters; T. Handel for use of the NMR spectrometer; G. Spudich for purification of the 75cpT4L* protein used in this study; J. Kho for cloning and purification of the H31N 13cpT4L* protein. We especially thank T. Alber for various lysozyme clones and his insight into this project. We thank the members of the Marqusee and Handel laboratories for discussion and critical reading of the manuscript.

References

Alber T. 1989. Mutational effects on protein stability. *Annu Rev Biochem 58*:765–798.

Anderson DE, Becktel WJ, Dahlquist FW. 1990. pH-induced denaturation of proteins: A single salt bridge contributes 3–5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry 29*:2403–2408.

Anderson DE, Lu J, McIntosh LP, Dahlquist FW. 1993. The folding, stability and dynamics of T4 lysozyme: A perspective using nuclear magnetic resonance. In: Clore GM, Gronenborn AM, eds. *NMR of proteins*. Boca Raton, Florida: CRC Press. pp 258–304.

Brandts JF, Kaplan LJ. 1973. Derivative spectroscopy applied to tyrosyl chromophores. Studies on ribonuclease, lima bean inhibitors, insulin, and pancreatic trypsin inhibitor. *Biochemistry 12*:2011–2024.

Dixon MM, Nicholson H, Shewchuk L, Baase WA, Matthews BW. 1992. Structure of a hinge-bending bacteriophage T4 lysozyme mutant, Ile3 → Pro. *J Mol Biol 227*:917–933.

Edelhoch H. 1967. Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry 6*:1948–1954.

Elwell M, Schellman J. 1975. Phage T4 lysozyme. Physical properties and reversible unfolding. *Biochim Biophys Acta 386*:309–323.

Englander SW, Sosnick TR, Englander JJ, Mayne L. 1996. Mechanisms and uses of hydrogen exchange. *Curr Opin Struct Biol 6*:18–23.

Faber HR, Matthews BW. 1990. A mutant T4 lysozyme displays five different crystal conformations [see comments]. *Nature 348*:263–266.

Heinemann U, Hahn M. 1995. Circular permutation of polypeptide chains: Implications for protein folding and stability. *Prog Biophys Mol Biol 64*:121–143.

Jaenicke R. 1991. Protein folding: Local structures, domains, subunits, and assemblies. *Biochemistry 30*:3147–3161.

Kim PS, Baldwin RL. 1982. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu Rev Biochem 51*:459–489.

Kuwajima K. 1989. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins Struct Funct Genet 6*:87–103.

Kuwajima K. 1996. The molten globule state of alpha-lactalbumin. *FASEB J 10*:102–109.

Lu J, Dahlquist FW. 1992. Detection and characterization of an early folding intermediate of T4 lysozyme using pulsed hydrogen exchange and two-dimensional NMR. *Biochemistry 31*:4749–4756.

Matsumura M, Matthews BW. 1989. Control of enzyme activity by an engineered disulfide bond. *Science 243*:792–794.

Matthews BW. 1995. Studies on protein stability with T4 lysozyme. *Adv Protein Chem 46*:249–278.

Matthews BW. 1996. Structural and genetic analysis of the folding and function of T4 lysozyme. *FASEB J 10*:35–41.

McHaourab HS, Lietzow MA, Hideg K, Hubbell WL. 1996. Motion of spin-labeled side chains in T4 lysozyme. Correlation with protein structure and dynamics. *Biochemistry 35*:7692–7704.

McHaourab HS, Oh KJ, Fang CJ, Hubbell WL. 1997. Conformation of T4 lysozyme in solution. Hinge-bending motion and the substrate-induced conformational transition studied by site-directed spin labeling. *Biochemistry 36*:307–316.

McLeish MJ, Nielsen KJ, Najbar LV, Wade JD, Lin F, Doughty MB, Craik DJ. 1994. Conformation of a peptide corresponding to T4 lysozyme residues 59–81 by NMR and CD spectroscopy. *Biochemistry 33*:11174–11183.

McLeish MJ, Nielsen KJ, Wade JD, Craik DJ. 1993. A peptide corresponding to the N-terminal 13 residues of T4 lysozyme forms an alpha-helix. *FEBS Lett 315*:323–328.

Najbar LV, Craik DJ, Wade JD, Lin F, McLeish MJ. 1995. CD and NMR determination of the solution structure of a peptide corresponding to T4 lysozyme residues 38–51. *Biochim Biophys Acta 1250*:163–170.

Nicholson H, Anderson DE, Dao-pin S, Matthews BW. 1991. Analysis of the interaction between charged side chains and the alpha-helix dipole using designed thermostable mutants of phage T4 lysozyme. *Biochemistry 30*:9816–9828.

Oas TG, Kim PS. 1988. A peptide model of a protein folding intermediate. *Nature 336*:42–48.

Peng ZY, Kim PS. 1994. A protein dissection study of a molten globule. *Biochemistry 33*:2136–2141.

Poteete AR, Hardy LW. 1994. Genetic analysis of bacteriophage T4 lysozyme structure and function. *J Bacteriol 176*:6783–6788.

Poteete AR, Sun DP, Nicholson H, Matthews BW. 1991. Second-site revertants of an inactive T4 lysozyme mutant restore activity by restructuring the active site cleft. *Biochemistry 30*:1425–1432.

Ptitsyn OB, Pain RH, Semisotnov GV, Zerovnik E, Razgulyaev OI. 1990. Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Lett 262*:20–24.

Radford SE, Dobson CM, Evans PA. 1992. The folding of hen lysozyme involves partially structured intermediates and multiple pathways [see comments]. *Nature 358*:302–307.

Rennell D, Bouvier SE, Hardy LW, Poteete AR. 1991. Systematic mutation of bacteriophage T4 lysozyme. *J Mol Biol 222*:67–88.

Rose GD. 1979. Hierarchic organization of domains in globular proteins. *J Mol Biol 134*:447–470.

Sambrook J, Maniatis T, Fritsch EF. 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Santoro MM, Bolen DW. 1988. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry 27*:8063–8068.

Sauer UH, San DP, Matthews BW. 1992. Tolerance of T4 lysozyme to proline substitutions within the long interdomain alpha-helix illustrates the adaptability of proteins to potentially destabilizing lesions. *J Biol Chem 267*:2393–2399.

Schulman BA, Redfield C, Peng ZY, Dobson CM, Kim PS. 1995. Different subdomains are most protected from hydrogen exchange in the molten globule and native states of human alpha-lactalbumin. *J Mol Biol 253*:651–657.

Shoichet BK, Baase WA, Kuroki R, Matthews BW. 1995. A relationship between protein stability and protein function. *Proc Natl Acad Sci USA 92*:452–456.

Staley JP, Kim PS. 1990. Role of a subdomain in the folding of bovine pancreatic trypsin inhibitor. *Nature 344*:685–688.

Staley JP, Kim PS. 1992. Complete folding of bovine pancreatic trypsin inhibitor with only a single disulfide bond. *Proc Natl Acad Sci USA 89*:1519–1523.

Tsugita A, Inouye M. 1968. Purification of bacteriophage T4 lysozyme. *J Biol Chem 243*:391–397.

Weaver LH, Matthews BW. 1987. Structure of bacteriophage T4 lysozyme refined at 1.7 Å resolution. *J Mol Biol 193*:189–199.

Wetlaufer DB. 1973. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci USA 70*:697–701.

Zehfus MH. 1995. Automatic recognition of hydrophobic clusters and their correlation with protein folding units. *Protein Sci 4*:1188–1202.

Zhang T, Bertelsen E, Benvegnu D, Alber T. 1993. Circular permutation of T4 lysozyme. *Biochemistry 32*:12311–12318.