Circular Permutation of T4 Lysozyme†

Tao Zhang,‡ Eric Bertelsen,§ Dominic Benvegnu,‖ and Tom Alber*,‡

Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, Utah 84132

Received July 26, 1993; Revised Manuscript Received September 9, 1993®

ABSTRACT: To examine the relationship between polypeptide chain synthesis and protein folding, we have constructed a circularly permuted variant of phage T4 lysozyme. The permuted protein begins at residue 37 of the wild-type sequence and ends at residue 36. The normal chain termini are joined by a six-residue linker, Ser-Gly_4-Ala. The permuted lysozyme folds efficiently and cleaves bacterial cell walls with normal specific activity. As judged by circular dichroism, UV absorbance, fluorescence, and nuclear magnetic resonance spectroscopy, the permutation causes little change in the structure of the protein. Reversible denaturation experiments show that the permutation reduces the stability of T4 lysozyme only 0.8–1.1 kcal/mol. These results demonstrate that a protein with two domains can be permuted with little change in activity, structure, and stability. The order of chain synthesis, the sequential arrangement of secondary structures, and the position of chain termini with respect to domain boundaries do not determine the protein fold.

The three-dimensional structures of proteins are overdetermined by their amino acid sequences (Richards, 1991; Heinz et al., 1992). Phylogenetic comparisons and genetic studies, for example, show that proteins can tolerate many amino acid substitutions (Perutz et al., 1965; Matthews, 1987; Lim & Sauer, 1989; Hynes et al., 1989). Physical and structural characterizations of mutant proteins have shown that substitutions are accommodated by several distinct mechanisms. Some residues apparently make small contributions to protein stability, some amino acid replacements preserve important interactions, and some substitutions cause structural adjustments that minimize destabilizing effects.

In addition to examining the roles of individual amino acids in folding, studies have been undertaken to explore the importance of the continuity of the polypeptide chain in determining three-dimensional structure. Complementary fragments of proteins such as ribonuclease (Richards, 1959), β-galactosidase (Ullmann et al., 1967), staphylococcal nuclease (Anfinsen, 1973), bacteriorhodopsin (Liao et al., 1983; Sigrist et al., 1988), HIV reverse transcriptase (Hostomsky et al., 1991), phosphoribosyl-anthranilate isomerase (PRAI)[1] (Eder & Kirschner, 1992), barnase (Sancho & Fersht, 1992), and ATCase (Powers et al., 1993) can associate and fold to give active molecules. Functional single-chain F_v antibody fragments also have been expressed (Huston et al., 1988; Pantoliano et al., 1991). In addition, constructions of active, circularly permuted forms of bovine pancreatic trypsin inhibitor (BPTI) (Goldenberg & Creighton, 1983), PRAI (Luger et al., 1989), and dihydrofolate reductase (DHFR) (Buchwalder et al., 1992) have been reported. These studies show that the ends of a protein can be moved with small effects on folding and function. For these examples, the order of synthesis of the polypeptide chain is not coupled in a unique way to protein folding.

To analyze the effects of moving the termini from one domain to the other in a multidomain protein, we have constructed a circularly permuted variant of phage T4 lysozyme (T4L). T4L has two domains linked by two connections (Weaver & Matthews, 1987) (see Figure 1). The domains are defined by residues that maintain similar relative positions in different conformations of the enzyme (Faber & Matthews, 1990). The N-terminal domain contains residues 15–60, and the so-called C-terminal domain contains residues 1–11 and 77–164. Despite the presence of two structural domains, the enzyme folds as a single cooperative unit. Here we describe the design, construction, and characterization of the permuted T4L.

MATERIALS AND METHODS

Design of the Permuted Protein. The design of the permuted protein was based on the refined, high-resolution X-ray structure of T4L (Weaver & Matthews, 1987). A linker to join the normal ends of the polypeptide chain was modeled using the program FRODO (Jones, 1978). Residues were added to the C-terminus and manually adjusted to allowed main-chain dihedral angles until the chain ends could be joined by a peptide bond. The final linker design (Figure 1) was compared to a database of protein structures (Jones & Thirup, 1986) to assure that the model was reasonable.

Construction of Extended and Permuted Lysozymes. Figure 2 shows the steps used to construct genes for the permuted T4 lysozyme (cpT4L) and for an extended T4L (xT4L) with the linker residues added to the C-terminus. The constructions were confirmed by DNA sequencing (Sanger et al., 1977). The template for these constructions was the cysteine-free, "pseudo-wild-type" lysozyme (Matsumura & Matthews, 1989), which contains the mutations C54T and C97A. The gene encoding this protein was inserted into phage

† This work was supported by a grant to T.A. from the Pew Scholarship in the Biomedical Sciences.

* To whom correspondence should be addressed.

‡ Present address: Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720.

§ Present address: Institute of Molecular Biology, University of Oregon, Eugene, OR 97403.

‖ Present address: Biocircuits Corp., 1324 Chesapeake Terrace, Sunnyvale, CA 94089.

® Abstract published in *Advance ACS Abstracts,* November 1, 1993.

[1] Abbreviations: ATCase, aspartate carbamoyltransferase; BPTI, bovine pancreatic trypsin inhibitor; C_m, midpoint concentration of chemical denaturation; CD, circular dichroism; DHFR, dihydrofolate reductase; DQF-COSY, double-quantum-filtered correlated spectroscopy; DTT, dithiothreitol, GuHCl, guanidine hydrochloride, IPTG, isopropyl β-thiogalactoside; NMR, nuclear magnetic resonance; PRAI, phosphoribosylanthranilate isomerase; T_m, midpoint temperature of thermal denaturation; cpT4L, circularly permuted T4 lysozyme; xT4L, extended T4 lysozyme.

0006-2960/93/0432-12311$04.00/0

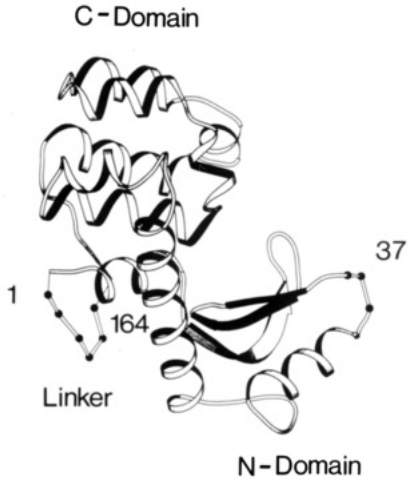


FIGURE 1: Ribbon drawing of T4 lysozyme. The normal N- and C-termini are at residues 1 and 164, respectively. The α-carbons of the designed linker sequence (residues 165–170) and the flexible, surface loop (residues 36–38) where the new termini were engineered are indicated with closed circles. The active site of the enzyme is in the cleft between the lower (N-terminal) and upper (C-terminal) domains.

m13mp18 (Yanisch-Perron et al., 1985), and site-directed mutagenesis (Kunkel, 1985) was used to create *NcoI* restriction sites at the codons for amino acids 1 and 37. This introduced three additional point mutations: N2D, P37M, and S38D (clone 1).

The sequence encoding amino acids 1–36 was removed from clone 1 by *NcoI* digestion and ligation to produce clone 2. Clone 2 codes for a C-terminal fragment of T4L starting with Met-37.

Four overlapping oligonucleotides (Figure 3) coding for the linker sequence (Ser-Gly_4-Ala) were synthesized with an Applied Biosystems DNA synthesizer. The linker DNA started at the *MluI* site corresponding to amino acids 159–160 and ended with a termination codon followed by a *Hin*dIII restriction site. This DNA was used to replace the *MluI*–*Hin*dIII fragment at the 3′ end of a truncated T4L gene (*Eco*RI–*Hin*dIII, clone 3).

A DNA fragment encoding residues 1–36 was isolated from clone 1 by *Nco*I digestion. This fragment was inserted into the *Nco*I site in the linker DNA in clone 3 to generate clone 4.

To assemble the gene for cpT4L, the *Eco*RI–*Hin*dIII fragment in clone 2 was replaced with the *Eco*RI–*Hin*dIII fragment from clone 4. The resulting gene codes for a protein that starts at amino acid 37, proceeds through the linker to residues 1–36, and ends with the two added amino acids Met-Gly. The six bases specifying these last two amino acids were deleted by site-directed mutagenesis (Kunkel, 1985) to produce the final construct encoding cpT4L.

A gene for an extended T4L with the linker added to the normal C-terminus was constructed by replacing the *MluI*–*Hin*dIII fragment of clone 1 with the linker DNA. As for the permuted protein, the last two residues (Met-Gly) encoded by the linker DNA were removed by site-directed mutagenesis. The extended and permuted T4 lysozymes have the same amino acid composition.

Protein Expression and Purification. The genes for both xT4L and cpT4L were placed in the vector pHSe5, and the expression plasmids were used to transform *Escherichia coli* RR1 (Muchmore et al., 1989). Protein production was induced by adding isopropyl β-thiogalactoside (0.4 mM) to log-phase cultures of RR1 harboring the expression plasmid.

The yield of cpT4L was lower than xT4L in this expression system. Large amounts of permuted protein were obtained by growing the cells harboring the expression plasmid to high density in a rolling drum fermentor (Labline, Model 29500). By providing O_2 at 4 L/min, the cells were grown on an optical density at 600 nm of 1–2 before addition of IPTG to induce lysozyme production. Terrific broth (Tartof & Hobbs, 1987) was used for high-density growth, and the pH of the culture was adjusted to 7 when IPTG was added.

Large amounts of cpT4L also were produced from an alternative expression vector based on pAED2 (Doering, 1991). This plasmid contains a phage T7 promoter to drive expression of genes inserted into a polylinker site. The *Bam*HI–*Hin*dIII fragment containing both the cpT4L gene and the ribosome binding site from pHSe5 was inserted between the *Bam*HI and *Hin*dIII sites in pAED2. In this plasmid, T4 lysozyme is the second open reading frame translated from the message produced from the T7 promoter. The yield of purified cpT4L was typically 200–300 mg/L of *E. coli* BL21(DE3) cultures containing the pAED2-based plasmid.

T4 lysozymes were purified by ion-exchange chromatography (Alber & Matthews, 1987; Muchmore et al., 1989), and the purified proteins were shown to be homogeneous by SDS gel electrophoresis. Protein concentrations were determined spectrophotometrically using an extinction coefficient of 1.28 $cm^2\ mg^{-1}$ at 280 nm (Tsugita & Inouye, 1968). The N-terminal amino acid sequences of purified T4 lysozymes were determined using an Applied Biosystems 477A pulsed liquid protein sequencer.

Enzyme Assays. Lysozyme activity was measured by monitoring the change in turbidity of *E. coli* cell wall suspensions (Tsugita & Inouye, 1968) in 50 mM Tris-HCl, pH 7.4. Relative specific activities were determined from the rate of decrease of light scattering at 450 nm catalyzed by mutant and wild-type lysozymes. The enzyme activity also was assayed on petri plates as described (Alber & Matthews, 1987).

Thermodynamic Measurements. Stabilities of the purified T4 lysozymes were measured in three different ways: GuHCl denaturation, urea gradient gel electrophoresis, and differential scanning calorimetry. For GuHCl denaturation, lysozyme samples (0.05 mg/mL) in 50 mM Tris-HCl, 1 mM $MgCl_2$, and 0.1 mM DTT, pH 7.4, were incubated for 24 h at 22 °C in different concentrations of GuHCl (Pierce). GuHCl concentrations were determined by refractive index measurements (Nosaki, 1970). Unfolding was monitored by circular dichroism (CD) at 223 nm. The fraction of unfolded protein was plotted against GuHCl concentration (Pace, 1986). Relative stabilities ($\Delta\Delta G_u$) of mutant and wild-type lysozymes were determined by linear extrapolation of ΔG_u to the C_m of the wild-type protein (Pace, 1986).

The method of urea gradient gel electrophoresis has been described (Goldenberg, 1989a; Klemm et al., 1991). The buffer used in the gels and electrode solutions was 0.05 M MOPS/imidazole, pH 7.0. Protein samples (80 μL) containing 30 μg of lysozyme were mixed with 20 μL of 50% glycerol/0.2% methyl green and applied in a single band across the gel. Temperature was maintained at 22 °C during electrophoresis by circulating water through coils in the cathode solution. Unfolding was reversible under these conditions. For accurate comparisons of transition curves, samples were loaded on gels poured together in a multigel casting chamber. Gels were stained with Coomassie blue and photographed. The melting curves were digitized with a laser pen scanner (Silk Scientific) mounted on a Hewlett Packard plotter. The x axis was aligned with the urea gradient, and

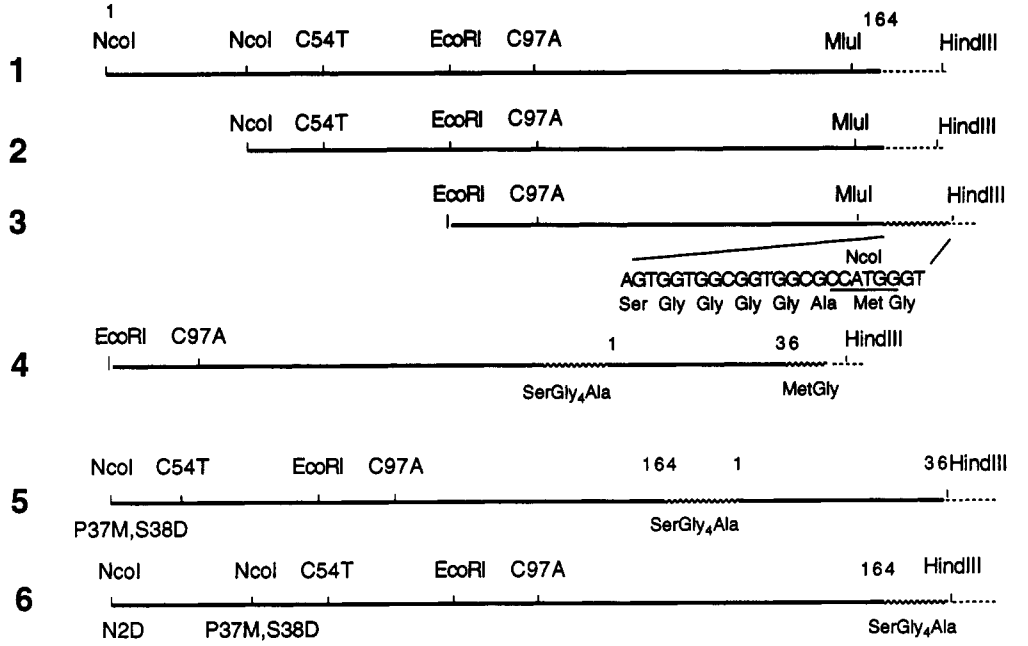


FIGURE 2: Schematic illustration of the coding sequences used to construct the permuted and extended T4 lysozymes. Sequence numbers refer to the corresponding amino acids in the extended protein, and amino acid substitutions are shown in the one-letter code. The vector for each DNA insert was phage m13mp18. Dashed lines indicate noncoding sequences. *Clone 1*: "Pseudo-wild-type" T4L gene with *Nco*I restriction sites engineered at amino acids 1 and 37. *Clone 2*: The *Nco*I restriction fragment specifying residues 1–36 was removed from clone 1. *Clone 3*: The 3′ half of the T4L gene with the added linker DNA sequence. The *Nco*I restriction site in the linker is underlined. *Clone 4*: The *Nco*I fragment of clone 1 coding for amino acids 1–36 was inserted into the *Nco*I site in the linker sequence in clone 3. *Clone 5*: Gene for the permuted T4L. The *Eco*RI–*Hin*dIII fragment in clone 3 was replaced with the insert in clone 4. Site-directed mutagenesis was employed to delete the C-terminal Met-Gly dipeptide. *Clone 6*: Gene for the extended T4L. The *Mlu*I–*Hin*dIII fragment in clone 1 was replaced with the linker DNA, and site-directed mutagenesis was used to delete the C-terminal Met-Gly dipeptide.

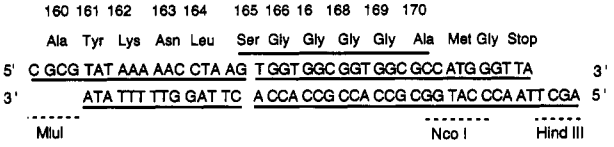


FIGURE 3: Oligonucleotides used to insert the linker between the termini of T4L. Each oligonucleotide is underlined, and the restriction sites are indicated by dashed lines. The corresponding amino acid sequence contained residues 160–164 of wt T4L and the linker sequence, Ser-Gly$_4$-Ala (165–170). The codons for the last two amino acids, Met-Gly, were included to accommodate a *Nco*I restriction site in the linker DNA. These two codons were deleted by site-directed mutagenesis at the end of construction of the genes for the extended and permuted T4L's.

the y axis was aligned with the direction of electrophoresis. For each point, the urea concentration was calculated from the x coordinate (Goldenberg, 1989a), and the fraction of unfolded protein was derived from the y coordinate (Pace, 1986). The unfolding transitions were fit to a two-state model to extract the parameters C_m, m, and ΔG_u (Hollecker & Creighton, 1982; Pace, 1986; Klemm et al., 1991).

Differential scanning calorimetry was carried out under conditions described by Kitamura and Sturtevant (1989). T4 lysozyme samples were exhaustively dialyzed at 4 °C against 20 mM KH_2PO_4, 25 mM KCl, and 0.1 mM DTT, pH 2.0–3.0. The final dialysate was used as the reference solution in the scanning calorimeter (Hart Scientific Corp., Model 4207). The dependence of ΔH on T_m was obtained by monitoring denaturation at pHs 2.0, 2.3, 2.5, 2.7, and 2.9. The heat capacity change for unfolding, ΔC_p was obtained by assuming the change of ΔH_0 with T_m is due to the heat capacity effect (Privalov, 1979):

$$\Delta H = \Delta H_0 + \Delta C_p(T - T_0)$$

ΔG versus temperature was derived by integrating the Gibbs–Helmholtz equation:

$$\frac{\partial(\Delta G/T)}{\partial T} = \frac{-\Delta H}{T^2}, \text{ with } \Delta G(T^\circ) = 0$$

Thus

$$\Delta G = \Delta H_0 - T\Delta H_0/T_0 + \Delta C_p[T - T_0 - T\ln(T/T_0)]$$

Spectroscopic Methods. Absorbance spectra were recorded with an Aviv 118DS spectrophotometer. Fluorescence emission spectra from 300 to 400 nm were recorded with a Perkin-Elmer fluorometer by exciting at 280 nm. Far-ultraviolet circular dichroism spectra were measured with an Aviv 62DS spectrometer at a protein concentration of 40 μM in 50 mM Tris, 1 mM $MgCl_2$, and 0.1 mM dithiothreitol, pH 7.4. The cell (1-cm path length) was maintained at 22 °C.

The NMR spectra of the T4 lysozymes were recorded at 20 °C with a General Electric Omega 500 spectrometer, as described previously (McIntosh et al., 1990). The samples were 1.5 mM cpT4L or 2.4 mM xT4L in 95% H_2O/5% D_2O solutions containing 100 mM potassium chloride, 30 mM potassium phosphate, and 0.01% sodium azide at pH 5.6.

RESULTS

Circular permutation requires joining the existing termini and creating new ends elsewhere in the polypeptide chain. The chain termini of T4L are approximately 10.5 Å apart in the X-ray crystal structure of wild-type (wt) lysozyme, but a high degree of disorder makes this placement uncertain (Weaver & Matthews, 1987). The distance between the ends is too large to bridge with a peptide bond, so a linker was added to make the connection.

The linker to join the ends of T4 lysozyme was designed using geometric criteria. Linker conformations that overlapped other residues in the X-ray crystal structure or that had forbidden main-chain dihedral angles were excluded. A

model could be constructed with six residues spanning the gap between the chain termini (Figure 1). To evaluate the plausibility of the model, α-carbon positions were compared to a data base of 51 known structures (Jones & Thirup, 1986). The data base contained 11 segments with root mean square deviations of 1.14–1.30 Å from the α-carbons of residues of 164–170 in the linker model. Residues 65–71 of high-potential iron protein (Freer et al., 1975) gave the best fit to the linker conformation.

On the basis of the reasoning that exposed and flexible parts of the protein are less important for structural stability (Alber & Matthews, 1987), residue 37 was chosen for the new N-terminus. Pro-37 is in the N-terminal domain in a surface loop joining a β-strand and an α-helix (Figure 1). With the exception of the normal termini, this loop has the highest main-chain crystallographic *B* values in the structure (Weaver & Matthews, 1987).

To start translation of the permuted protein, Pro-37 was changed to Met. Rearrangement of the T4L gene was facilitated by placing the methionine codons at positions 1 and 37 in engineered *Nco*I restriction sites. This required two more amino acid substitutions—Ser-38 and Asn-2 were changed to Asp. These substitutions slightly stabilize wt T4L (Nicholson et al., 1991). In addition, cysteines at positions 54 and 97 were changed to Thr and Ala, respectively. These two substitutions do not alter the activity and stability of T4L [Matsumura & Matthews, 1989], but the absence of cysteine oxidation reactions makes the mutant proteins easier to handle.

The linker sequence, Ser-Gly$_4$-Ala, was encoded by four oligonucleotides (Figure 3). Restriction sites for *Mlu*I and *Hind*III at the ends of linker DNA permitted directional replacement of the wild-type sequences. The linker DNA was added to the 3′ end of the lysozyme gene, and a restriction fragment coding for amino acids 1–36 was inserted into the *Nco*I site in the linker (Figure 2). An additional C-terminal Met-Gly dipeptide, which was included in the linker to specify a *Nco*I site (Figure 2), was removed from circularly permuted T4L (cpT4L) in the last step of the gene constructions. To permit careful comparisons to cpT4L, we also constructed a gene for an extended lysozyme (xT4L) with the six linker amino acids, Ser-Gly$_4$-Ala, added to the normal C-terminus. The permuted and extended proteins have identical amino acid compositions.

After assembly, the genes for the permuted and extended proteins were sequenced to ensure the absence of unintended mutations. The proteins were produced using two different expression plasmids: pHSe5 [containing the *tac* promoter (Muchmore et al., 1989)] and pAED2 [containing the *T7* promoter (Doering, 1991)]. In pHSe5, the permuted protein was expressed at a lower level than the wild-type and the extended proteins (data not shown). To increase the expression level, the ribosome binding site and the coding region for the permuted lysozyme were placed 3′ to the T7 promoter in the expression vector pAED2. In the resulting plasmid, the lysozyme gene was the second open reading frame translated from the transcript produced from the T7 promoter. This construct increased the yield of cpT4L by nearly 10-fold compared to the pHSe5 vector.

The extended and permuted lysozymes were purified using published methods for isolating the wild-type protein (Muchmore et al., 1989). N-Terminal amino acid sequencing confirmed that the permuted protein begins at position 37 of the normal sequence. Here, the sequence numbers of the permuted protein are derived in reference to the analogous residues in the extended lysozyme. Thus, the first residue in the permuted protein is Met37.

Table I: Enzymatic Activities and Stabilities of wt, Extended, and Permuted Lysozymes

	protein		
	wt	extended	permuted
specific activity (%)	100	200	90
GuHCl denaturation (pH 7.4)			
C_m (M)	2.2	2.2	2.0
m [kcal/(M·mol)]	5.5	6.5	5.1
$\Delta\Delta G$ (kcal/mol)	0	0	1.0
urea denaturation (pH 7.0)			
C_m (M)	6.3	6.4	5.8
m [kcal/(M·mol)]	2.0	1.9	1.7
$\Delta\Delta G$ (kcal/mol)	0	–0.2	1.1
calorimetry (pH 2.5)			
T_m (°C)	45.0[a]	44.0 (±0.6)	40.1 (±0.7)
ΔH_0 (kcal/mol)	113.6[a]	84 (±5)	68 (±4)
$\Delta\Delta G$ (kcal/mol) (45 °C)	0.0	0.3 (±0.1)	1.1 (±0.1)

[a] From Kitamura and Sturtevant (1989).

Enzymatic Activity. The specific activities of the wild-type, extended, and permuted lysozymes were compared using a quantitative turbidity assay (Tsugita & Inouye, 1968). Because the substrate employed in the assay is a heterogeneous suspension of cell walls, kinetic constants for the enzyme-catalyzed reactions cannot be determined from the results. As shown in Table I, the wild-type and permuted proteins had the same specific activity, and the extended proteins showed 2-fold higher activity. We conclude that all three purified proteins have similar specific activities.

Stability of the Permuted Lysozyme. Three different methods were used to compare the thermodynamic stabilities of the wt, permuted, and extended lysozymes (Figure 4). The GuHCl denaturation curves of the three proteins at neutral pH (Figure 4A) were obtained from CD measurements at 223 nm. The unfolding transitions of the wt and extended proteins were virtually superimposable despite the introduction of five point mutations and the linker. However, xT4L showed slightly higher cooperativity than wt T4L (m value, Table I). The permuted protein was less stable than both the extended and wt proteins, and the m value of cpT4L was similar to that of wtT4L. The stability differences ($\Delta\Delta G$) were calculated by linear extrapolation of melting curves to 2.22 M GuHCl, the C_m of the wt protein. Compared to extended and wt lysozymes, the stability of the permuted lysozyme is reduced by 1 kcal/mol (Table I).

The three proteins also were denatured at neutral pH with urea. Unfolding was monitored by urea gradient gel electrophoresis [Goldenberg, 1989a,b; Klemm et al., 1991], and the denaturation transitions were analyzed after digitizing the melting curves (Figure 4B). The C_m and m (slope) values are listed in Table I. All three proteins showed similar m values, and the extended and wt proteins had similar C_m values. In contrast, the C_m of the permuted protein was 0.5 M lower. This corresponds to a reduction of approximately 1.1 kcal/mol in the stabilization free energy ($\Delta\Delta G$).

The thermal unfolding transitions of xT4L and cpT4L were monitored by differential scanning calorimetry. The calorimetric melting curve of cpT4L at pH 2.5 is shown in Figure 4C. The reversibility of unfolding under these conditions was >90%. Calorimetric measurements of cpT4L were made at pHs 2.0, 2.3, 2.5, 2.7, and 2.9, values of ΔH were calculated from the transitions, and ΔC_p values were obtained from the variation of ΔH with temperature. Thermodynamic parameters are listed in Table I. The T_m values of the wt and extended proteins were within 1 °C, although the differences in ΔH were significant. The T_m of cpT4L was 5 °C lower than the

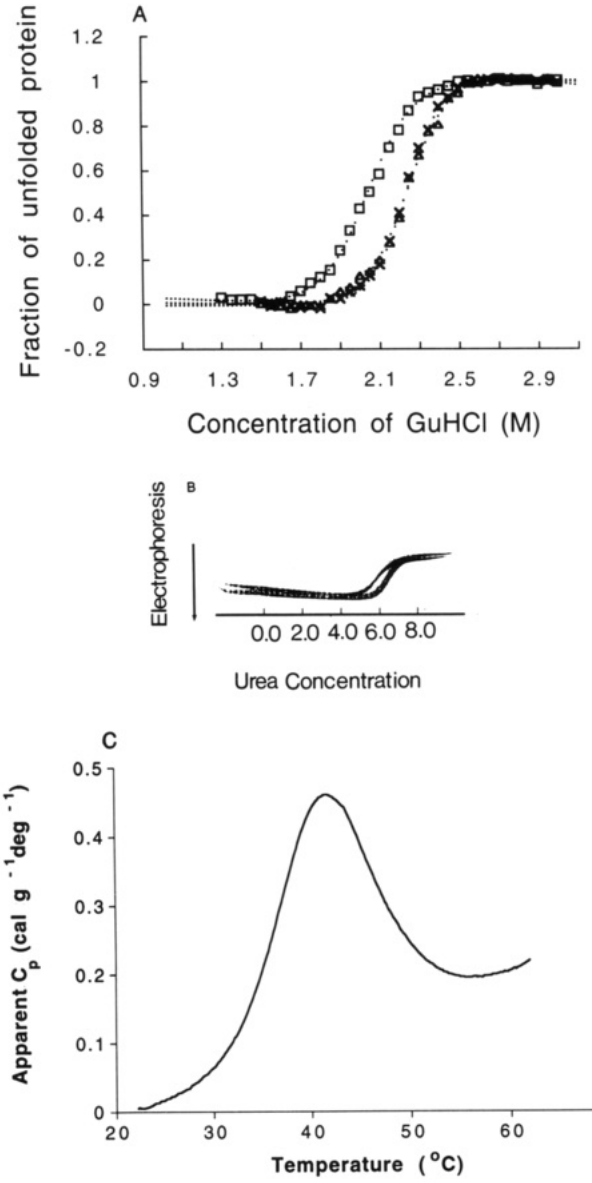


FIGURE 4: Denaturation studies of the wt, extended, and permuted lysozymes. (A) GuHCl denaturation at pH 7.4, 22 °C, monitored by circular dichroism at 223 nm. The fraction of unfolded protein is plotted as a function of GuHCl concentration. (△) wt T4L; (×) extended T4L; (□) permuted T4L. The theoretical curves calculated with the derived parameters (m and C_m) are shown as dotted lines superimposed on the data. (B) Urea gradient gel of extended (right) and permuted (left) T4 Lysozymes. Electrophoresis was carried out at pH 7.0, 22 °C, in a native polyacrylamide gel containing a transverse gradient of 0–8 M urea (Goldenberg, 1989a; Klemm et al., 1991). Samples were loaded in a single band across the gel. Because the volume of the unfolded protein is larger, it migrates more slowly than the folded protein. The smooth transition between the folded and unfolded states indicates that the unfolding transition is reversible and rapid on the time scale of electrophoresis. (C) Thermal denaturation of permuted T4L monitored by differential scanning calorimetry. The protein (7.5 mg/mL) was in 25 mM KCl, 0.1, mM DTT, and 20 mM KH_2PO_4, pH 2.5.

T_m of xT4L at pH 2.5. At this pH, values of ΔG were extrapolated to 45 °C, the T_m of wt T4L. The circularly permuted lysozyme was only 0.8 and 1.1 kcal/mol less stable than xT4L and wt T4L, respectively. In summary, the results of the chemical and thermal denaturation studies consistently indicate that the permutation only marginally destabilizes T4 lysozyme.

Structure of the Permuted Lysozyme. The secondary structures of the permuted, extended, and wt lysozymes were compared by CD spectroscopy (Figure 5). The far-UV CD

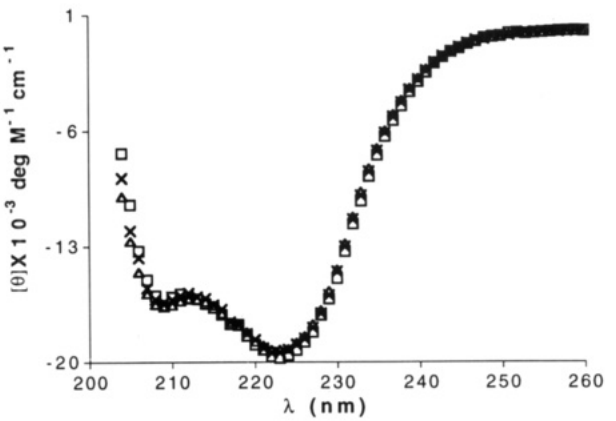


FIGURE 5: Circular dichroism spectra of the wt (△), extended (×), and permuted (□) T4 lysozymes. Spectra were recorded at a protein concentration of a 40 μM in 50 mM Tris, 1 mM $MgCl_2$, and 0.1 mM DTT, pH 7.4.

spectrum of wt T4L has distinctive minima at 223 and 208 nm (Elwell & Schellman, 1974). Both xT4L and cpT4L showed the same peaks with the same intensities.

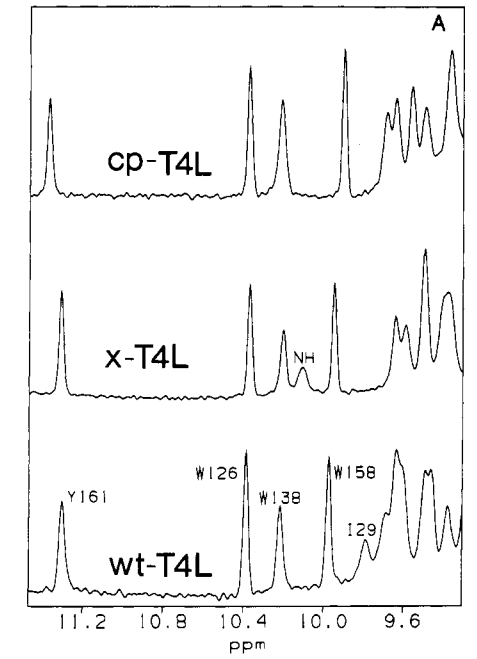
Absorbance and fluorescence spectra were recorded to compare the environments of the Tyr and Trp side chains in the three proteins. The UV absorbance and fluorescence spectra of wt T4L, xT4L, and cpT4L were essentially superimposable (data not shown).

To characterize the tertiary folds in detail, one- and two-dimensional NMR spectra of xT4L and cpT4L were compared to spectra of the wt protein (McIntosh et al., 1990). Regions of the one-dimensional ^{1}H NMR spectra recorded at 500 MHz are shown in Figure 6. Although there are only a modest number of resolved peaks, these are distinct markers for the native structure of T4L. For example, the buried hydroxyl proton of Tyr-161 and the three tryptophan indole $H^{\epsilon 1}$ protons had similar chemical shifts in all three proteins (Figure 6A). On the other hand, the chemical shifts of the amide protons of Val-94 and Ile-29 are altered in both xT4L and cpT4L compared to wt T4L, suggesting that the changes resulted from the amino acid substitutions rather than the permutation. The spectrum of xT4L contained a new peak near 10.2 ppm, presumably reflecting a downfield-shifted amide proton. The upfield region of the spectra of xT4L and cpT4L (Figure 6B) showed similar peaks at the positions of the resonances for side-chain methyl protons of Val-94 and Ala-98.

Two-dimensional ^{1}H DQF-COSY spectra of xT4L and cpT4L were recorded to resolve resonances from additonal protons in the two proteins. A representative region of the spectra containing aromatic proton cross-peaks is shown in Figure 7. These spectra are strikingly similar to each other and to the spectrum of the wt protein [cf. Figure 11 of McIntosh et al. (1990)]. On the basis of assignments of wt T4L, residues in similar environments in xT4L and cpT4L include Phe-67, Tyr-88, Phe-114, Trp-126, Trp-138, Tyr-139, Phe-153, and Trp-158 (Figure 7). Overall, the two-dimensional ^{1}H DQF-COSY spectra appeared distinctly similar, indicating that the wt, extended, and permuted proteins adopt similar structures.

DISCUSSION

We have engineered a circularly permuted variant of T4L that is enzymatically active and thermodynamically stable. The most remarkable aspect of these results is that cpT4L folds at all. This is surprising because genes for circularly permuted proteins have not been identified in the sequence



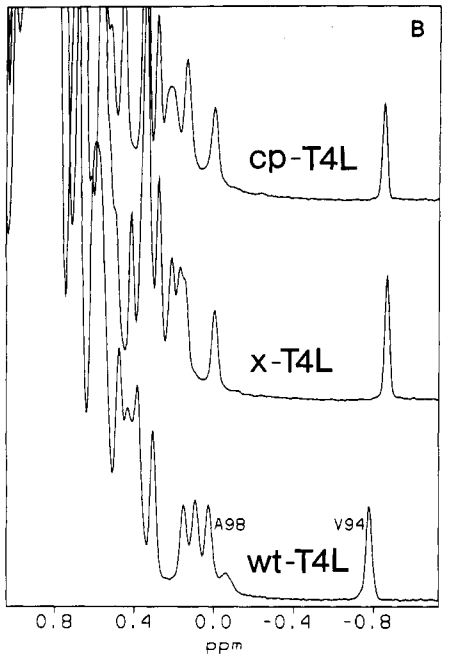
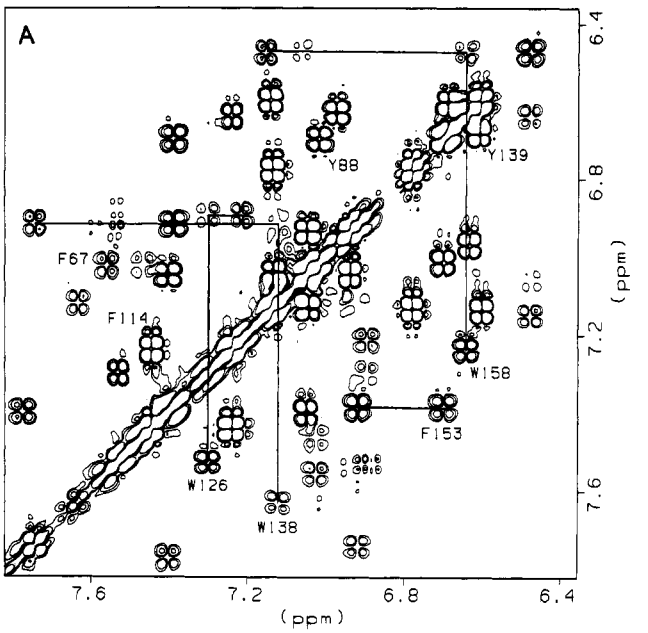


FIGURE 6: One-dimensional NMR spectra of the wt, extended, and permuted lysozymes. Spectra were recorded at 500 MHz in 95% H_2O/5% D_2O solutions containing 50 mM K_2HPO_4, 100 mM KCl, and 0.01% sodium azide, pH 5.6 at 20 °C. Similar chemical shifts of individual protons reflect similarities of local tertiary structures. (A) The downfield portions of the one-dimensional ^{1}H-NMR spectra of the wt (bottom), extended (middle), and permuted (top) proteins. The resonances of the hydroxyl H^η proton of Y161, the H^N proton of Ile29, and the indole $N^{\epsilon 1}$ protons of Trp-126, -138, and -158 are labeled in the spectrum of the wt protein (McIntosh et al., 1990). An unassigned amide proton resonance is indicated as NH in the spectrum of the extended protein. (B) The upfield regions of the one-dimensional ^{1}H-NMR spectra of the wt (bottom), extended (middle), and permuted (top) T4 lysozymes. The resonances from the methyl protons of Val-94 and Ala-98 are indicated.



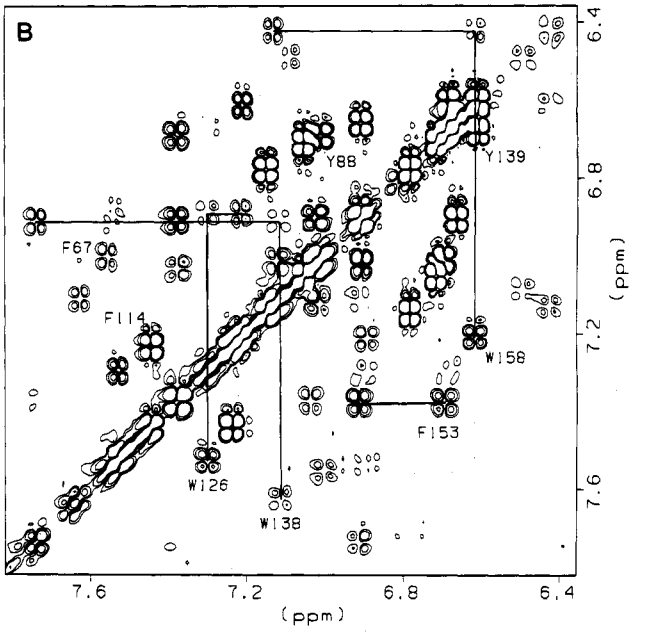


FIGURE 7: Representative regions of the DQF-COSY spectra recorded at 500 MHz showing aromatic proton resonances of extended T4L (A) and circularly permuted T4L (B). The samples contained 2.4 and 1.5 mM extended and permuted lysosymes, respectively. Resonances of side-chain protons in Phe-67, Tyr-88, Phe-114, Trp-126, Trp-138, Tyr-139, Phe-153, and Trp-158 are indicated. These tentative assignments were based on the positions of analogous, assigned peaks in the DQF-COSY spectrum of wt T4L (McIntosh et al., 1990). The unassigned resonances in the spectra of extended and permuted T4L also show striking similarities.

data base. As with other proteins whose sequences have been permuted by design, including BPTI (Goldenberg & Creighton, 1983; Goldenberg, 1989b), PRAI (Luger et al., 1989), DHFR [Buchwalder et al., 1992], and ATCase (Yang & Schachman, 1993), the structure and activity of cpT4L indicate that the three-dimensional structure of a protein does not depend on a unique order of polypeptide synthesis. This implies that the enumeration of amino acid sequences that can specify a given fold (Ponder & Richards, 1987; Lim & Sauer, 1989; Bowie et al., 1991; Jones et al., 1992) should be expanded to encompass functional permutations.

Because permuted proteins with little change in stability and activity can be designed and the N- and C-termini of many proteins are close (Thornton & Sibanda, 1983), it is curious that genes that code for permuted sequences have not been discovered in the sequence data base. The absence of obvious permutations among natural sequences could have several causes. A trivial possibility, for example, is that proper searches for permuted sequences remain to be done. Alternatively, permutations may be obscured if they are ancient or if subsequent genetic events such as point mutations, insertions,

and deletions are much more rapid. It is also possible, despite hypothetical proposals (Cunningham et al., 1979; Li & Graur, 1991), that organisms may lack efficient mechanisms for generating permutations that survive selection.

To construct cpT4L, five amino acid substitutions (N2D, P37M, S38D, C54T, and C97A) were introduced, and a linker of six residues (Ser-Gly_4-Ala) was added to join the normal chain termini. The linker was designed to be compatible with the X-ray crystal structure of T4L and to conform to accepted structural principles. Nonetheless, an alternative permuted sequence containing three more glycines in the middle of the linker also was found to be catalytically active (data not shown). Although the permuted lysozyme containing the nine-residue linker was not characterized in detail, the finding that the linker can be varied is consistent with studies showing that loops in proteins often tolerate insertions and amino acid substitutions (Bashford et al., 1987; Pascarella & Argos, 1992).

An important aspect of the work described here is that the sequence changes needed to construct cpT4L also were engineered into wt T4L to generate xT4L. The small differences in the properties of xT4L and the wt protein reflect the effects of the sequence changes. On the other hand, because xT4L and cpT4L have the same amino acid composition, the differences in the stabilities and enzymatic activities of cpT4L and xT4L provide a measure of the effects of the permutation itself.

A key difference in the properties of xT4L and cpT4L is the reduction in stability of the permuted protein by $\sim$1 kcal/mol (Figure 4 and Table I). This difference is surprisingly small. A comparable reduction in thermodynamic stability is often associated with single amino acid substitutions that remove a single methyl group or hydrogen-bonding substituent (Fersht, 1987; Matthews, 1987; Shortle, 1992; Sondek & Shortle, 1992; Eriksson et al., 1993).

The reduced stability of cpT4L cannot arise simply from differences in the hydration properties of the amino acids in the unfolded polypeptide chain, because the extended and permuted proteins contain the same amino acids. This suggests that the change in thermodynamic stability is associated with structural changes in the unfolded state and/or the native state. The folded structures of xT4L and cpT4L must differ, for example, at the new termini and in the linker region. Residues 36 and 37 must move apart when the peptide bond between them is broken, and the C-terminus of the linker is constrained to be near Met-1 only in the permuted lysozyme.

Overall, however, the structures of wt T4L, xT4L, and cpT4L appear to be extremely similar. The comparable specific activities of the three proteins (Table I) provide indirect evidence for close structural homology of their active sites. The active site of T4L includes a large cleft between the domains (Figure 1) and a peptide binding surface on the C-terminal domain (Weaver & Matthews, 1987; Dixon et al., 1992). Thus, a substantial part of the protein surface is unaffected by the permutation. In addition, a hinge motion of the two domains is required for catalytic activity (Faber & Matthews, 1990). This conformational change is apparently accommodated by the linker sequence and the permutation.

More direct comparisons of the structures of the wt, extended, and permuted lysozymes were made using spectroscopic methods. The far-UV CD spectra of the three proteins were extremely similar (Figure 5), indicating that the permutation had little effect on the secondary structure of T4L. The extension and permutation also preserved the environments of the three Trp and six Tyr side chains, as judged by absorbance and fluorescence emission spectra.

NMR spectroscopy provides one of the most sensitive probes of tertiary structure. Even small structural changes induced by mutations, however, can produce substantial changes in proton chemical shifts. As a result, chemical shift differences are poor indicators of the magnitudes of the associated structural changes. Similarities in chemical shifts, on the other hand, provide strong evidence for similar structures. Common features in the NMR spectra of wt, xT4L, and cpT4L tentatively can be interpreted in structural terms based on the assignments of the NMR spectra of wt T4L (McIntosh et al., 1990). A more rigorous accounting of the structural changes caused by the five amino acid substitutions and the permutation will require independent assignment of the spectra of xT4L and cpT4L.

With this caveat in mind, the NMR spectra provide evidence for extensive structural identities. The proton resonances shown in Figures 6 and 7, for example, suggest that the "C-terminal domains" (amino acids 1–11 and 77–164) in wt T4L, xT4L, and cpT4L are extremely similar. Residues buried in the core of this domain include Val-94, Ala-98, Phe-114, Trp-126, Trp-138, Tyr-139, Phe-153, Trp-158, and Tyr-161. Val-94 and Ala-98 contact the side chains of Trp-158 and Phe-153, respectively, and Tyr-161 packs behind the catalytic Glu-11. In addition, Phe-67, which packs into the "N-terminal domain", is in a similar environment in the three proteins. The similar resonances assigned to Trp-158 and Tyr-161 are especially noteworthy, because these residues are close to the normal C-terminus (residue 164). Thus, even three residues away from the added linker, the structures of xT4L and cpT4L closely match each other and the structure of the wt protein.

The similarities between xT4L and cpT4L emphasize that the sequence of a two-domain protein can be permuted with little effect on structure and function. The chain ends were moved from the C-terminal domain in the wt and extended proteins to the N-terminal domain in the permuted lysozyme. In addition, the normal N-terminus precedes a helix and the β-sheet in T4L, but the new N-terminus at residue 37 follows the β-sheet. In this case, therefore, the fold of the protein is not determined by the position of the termini with respect to the domain boundaries or the order of the secondary structure elements.

This conclusion has encouraging implications for protein design, because it suggests that there is considerable latitude in the choice of locations for the chain termini. In this regard, analysis of globular protein structures showed that terminal regions are generally flexible and accessible to solvent (Thornton & Sibanda, 1983). Chain termini in native proteins also show a statistical preference to be near each other. These patterns are reinforced by the growing number of active, circularly permuted proteins with ends introduced into surface loops. More systematic constructions of permuted proteins will permit a critical assessment of the limitations on the locations of chain termini.

Terminal regions of several proteins, including cytochrome *c* (Roder et al., 1988), myoglobin (Hughson et al., 1990), ribonuclease A (Udgaonkar & Baldwin, 1990), barnase (Serrano et al., 1992), and hen lysozyme (Radford et al., 1992), have been implicated in the early steps in protein folding. Early involvement of chain ends may provide several advantages. Folding may be blocked, for example, until synthesis of the polypeptide chain is completed. In addition, pairing the ends of the chain at the outset could provide the maximum restrictions on subsequent folding steps. Kinetic studies of folding of circularly permuted proteins will allow direct tests of these proposals.

ACKNOWLEDGMENT

We are indebted to Lawrence McIntosh for recording and analyzing the NMR spectra and for critical reading of the manuscript. We are grateful to T. Alwyn Jones for providing FRODO and structure comparison algorithms. We also thank Aron Branscomb for constructing an intermediate used in the genetic rearrangements, R. Schackmann of the Utah Cancer Center Protein/DNA Core Facility (supported by NIH Grant ZA42014) for oligonucleotide synthesis and N-terminal protein sequencing, and the University of Utah Center for Biopolymers at Interfaces for use of the calorimeter and digital scanner.

REFERENCES

Alber, T., & Matthews, B. W. (1987) *Methods Enzymol. 154*, 511–533.
Alber, T., Sun, D. P., Nye, J. A., Muchmore, D. C., & Matthews, B. W. (1987) *Biochemistry 26*, 3754–3758.
Anfinsen, C. B. (1973) *Science 181*, 223–230.
Bashford, D., Chothia, C., & Lesk, A. M. (1987) *J. Mol. Biol. 196*, 199–216.
Bowie, J. U., Luthy, R., & Eisenberg, D. (1991) *Science 253*, 164–170.
Buchwalder, A., Szadkowski, H., & Kirschner, K. (1992) *Biochemistry 31*, 1621–1630.
Cunningham, B. A., Hemperly, J. J., Hopp, T. P., & Edelman, G. M. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 3218–3222.
Dixon, M. M., Nicholson, H., Shewchuk, L., Baase, W. A., & Matthews, B. W. (1992) *J. Mol. Biol. 227*, 917–933.
Doering, D. S. (1991) Ph.D. Thesis, Massachusetts Institute of Technology.
Eder, J., & Kirschner, K. (1992) *Biochemistry 31*, 3617–3625.
Elwell, M., & Schellman, J. (1974) *Biochim. Biophys. Acta 386*, 309–323.
Eriksson, A. E., Baase, W. A., & Matthews, B. W. (1993) *J. Mol. Biol. 229*, 747–769.
Faber, H. R., & Matthews, B. W. (1990) *Nature 348*, 263–266.
Fersht, A. R. (1987) *Trends Biochem. Sci. 12*, 301–304.
Freer, S. T., Alden, R. A., Carter, C. W., Jr., & Kraut, J. (1975) *J. Biol. Chem. 250*, 46–54.
Goldenberg, D. P. (1989a) in *Protein Structure: A Practical Approach* (Creighton, T. E., Ed.) pp 225–250, IRL Press, Oxford.
Goldenberg, D. P. (1989b) *Protein Eng. 2*, 493–495.
Goldenberg, D. P., & Creighton, T. E. (1983) *J. Mol. Biol. 165*, 407–413.
Heinz, D. W., Baase, W. A., & Matthews, B. W. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 3751–3755.
Hollecker, M., & Creighton, T. E. (1982) *Biochim. Biophys. Acta 701*, 395–404.
Hostomsky, Z., Hostomska, Z., Hudson, G. O., Moomaw, E. W., & Nodes, B. R. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 1148–1152.
Hughson, F. M., Wright, P. E., & Baldwin, R. L. (1990) *Science 249*, 1544–1548.
Huston, J. S., Levinson, D., Mudgett-Hunter, M., Tai, M. S., Novotny, J., Margolies, M. N., Ridge, R. J., Bruccoleri, R. E., Haber, E., & Crea, R. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 5879–5883.
Hynes, T. R., Kautz, R. A., Goodman, M. A., Gill, J. F., & Fox, R. O. (1989) *Nature 339*, 73–76.
Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992) *Nature 358*, 86–89.
Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268–272.
Jones, T. A., & Thirup, S. (1986) *EMBO J. 5*, 819–822.
Kitamura, S., & Sturtevant, J. M. (1989) *Biochemistry 28*, 3788–3792.
Klemm, J. D., Wozniak, J. A., Alber, T., & Goldenberg, D. P. (1991) *Biochemistry 30*, 589–594.
Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 488–492.
Li, W.-H., & Graur, D. (1991) *Fundamentals of Molecular Evolution*, Sinauer Associates Inc., Sunderland, MA.
Liao, M. J., London, E., & Khorana, H. G. (1983) *J. Biol. Chem. 258*, 9949–9955.
Lim, W. A., & Sauer, R. T. (1989) *Nature 339*, 31–36.
Luger, K., Hommel, U., Herold, M., Hofsteenge, J., & Kirschner, K. (1989) *Science 243*, 206–210.
Matsumura, M., & Matthews, B. W. (1989) *Science 243*, 792–794.
Matthews, B. W. (1987) *Biochemistry 26*, 6885–6888.
McIntosh, L. P., Wand, A. J., Lowry, D. F., Redfield, A. G., & Dahlquist, F. W. (1990) *Biochemistry 29*, 6341–6362.
Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E., & Dahlquist, F. W. (1989) *Methods Enzymol. 177*, 43–73.
Nicholson, H., Anderson, D. E., Sun, D. P., & Matthews, B. W. (1991) *Biochemistry 30*, 9816–9828.
Nosaki, Y. (1970) *Methods Enzymol. 26*, 43–54.
Pace, C. N. (1986) *Methods Enzymol. 131*, 266–280.
Pantoliano, M. W., Bird, R. E., Johnson, S., Asel, E. D., Dodd, S. W., Wood, J. F., & Hardman, K. D. (1991) *Biochemistry 30*, 10117–10125.
Pascarella, S., & Argos, P. (1992) *J. Mol. Biol. 224*, 461–471.
Perutz, M. F., Kendrew, J. C., & Watson, H. C. (1965) *J. Mol. Biol. 13*, 669–678.
Ponder, J. W., & Richards, F. M. (1987) *J. Mol. Biol. 193*, 775–791.
Powers, V. M., Yang, Y. R., Fogli, M. J., Ponder J. Richards (1987) & Schachman, H. K. (1993) *Protein Sci. 2*, 1001–1012.
Privalov, P. L. (1979) *Adv. Protein Chem. 33*, 167–241.
Radford, S. E., Dobson, C. M., & Evans, P. A. (1992) *Nature 358*, 302–307.
Richards, F. M. (1991) *Sci. Am. 264*, 54–57, 60–63.
Richards, F. M., & Vithayathil, P. J. (1959) *J. Biol. Chem. 234*, 1459–1465.
Roder, H., Elove, G. A. & Englander, W. (1988) *Nature 335*, 700–704.
Sancho, J., & Fersht, A. R. (1992) *J. Mol. Biol. 224*, 741–774.
Sanger, F., Nickelsen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Serrano, L., Matouschek, A., & Fersht, A. R. (1992) *J. Mol. Biol. 224*, 847–859.
Shortle, D. (1992) *Q. Rev. Biophys. 25*, 205–250.
Sigrist, H., Wenger, R. H., Kislig, E., & Wuthrich, M. (1988) *Eur. J. Biochem. 177*, 125–133.
Sondek, J., & Shortle, D. (1992) *Proteins: Struct., Funct., Genet. 13*, 132–140.
Tartof, K. D., & Hobbs, C. A. (1987) *Bethesda Res. Lab. Focus 9*, 12.
Thornton, J. M., & Sibanda, B. L. (1983) *J. Mol. Biol. 167*, 443–460.
Tsugita, A., & Inouye, M. (1968) *J. Biol. Chem. 243*, 391–397.
Udgaonkar, J. B., & Baldwin, R. L. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8197–8201.
Ullmann, A., Jacob, F., & Monod, J. (1967) *J. Mol. Biol. 24*, 339–343.
Weaver, L. H., & Matthews, B. W. (1987) *J. Mol. Biol. 193*, 189–199.
Yang, Y. R., & Schachman, H. K. (1993) *Proc. Natl. Acad. Sci. U.S.A.* (in press).
Yanisch-Perron, C., Vieira, J., & Messing, J. (1985) *Gene 33*, 103.