Analysis of the Effectiveness of Proline Substitutions and Glycine Replacements in Increasing the Stability of Phage T4 Lysozyme

H. NICHOLSON,* D. E. TRONRUD, W. J. BECKTEL,† and B. W. MATTHEWS‡

Institute of Molecular Biology, Howard Hughes Medical Institute and Department of Physics, University of Oregon, Eugene, Oregon 97403

SYNOPSIS

It was previously shown that the two replacements Gly 77 → Ala (G77A) and Ala 82 → Pro (A82P) increase the thermostability of phage T4 lysozyme at pH 6.5. Such replacements are presumed to restrict the degrees of freedom of the unfolded protein and so decrease the entropy of unfolding [B. W. Matthews, H. Nicholson, and W. J. Becktel (1987) *Proceedings of the National Academy of Science USA* Vol. 84, pp. 6663–6667].

To further test this approach, three additional replacements—G113A, K60P and A93P—have been constructed. On the basis of model building, each of these three replacements was judged to be less than optimal because it would tend to introduce unfavorable van der Waals contacts with neighboring parts of the protein. The presence of such contacts was verified for G113A and K60P by conformational adjustments seen in the crystal structures of these mutant proteins. In the case of G113A there are backbone conformational changes of 0.5–1.0 Å in the short α-helix, 108–113, that includes the site of substitution. In the case of K60P the pyrrolidine ring shows evidence of strain. The thermal stability of each of the three variants at both pH 2.0 and pH 6.5 was found to be very close to that of wild-type lysozyme. The results suggest that the procedure used to predict sites for both Xaa → Pro and Gly → Ala is, in principle, correct. At the same time, the increase in stability expected from substitutions of this type is modest, and can easily be offset by strain associated with introduction of the alanine or proline. This means that the criteria used to select substitutions that will increase thermostability have to be stringent at least. In the case of T4 lysozyme this severely limits the number of sites. The analysis reveals a significant discrepancy between the conformational energy surface predicted for the residue preceding a proline and the conformations observed in crystal structures.

INTRODUCTION

A glycine residue has greater conformational freedom than any other amino acid. If a glycine within an unfolded polypeptide is replaced with another residue, the presence of the β-carbon will restrict the conformation of the unfolded peptide. This reduction in entropy destabilizes the unfolded form of the peptide relative to a folded conformation. In other words, substitutions of glycines with nonglycine residues at appropriately chosen sites can, in principle, be used to enhance protein stability.[1–5]

The pyrrolidine ring of a proline residue is even more restrictive. Therefore, the replacement of an amino acid of a protein with proline, if it can be accomplished without perturbing the native structure, should also stabilize the folded protein relative to the unfolded state.[1,6–8]

We previously showed,[1] and it has been confirmed by Ellman et al.[9] and Hu et al.,[10] that two replacements, Gly 77 → Ala (G77A) and Ala 82 → Pro (A82P), could be used to increase the thermostability of phage T4 lysozyme. These two examples

Biopolymers, Vol. 32, 1431–1441 (1992)
 CCC 0006-3525/92/111431-11

* Present address: Department of Chemistry and Biochemistry, Massey University, Palmerston North, New Zealand.

† Present address: Department of Biochemistry, Ohio State University, 484 West 12th Avenue, Columbus, OH 43210.

‡ To whom correspondence should be addressed.

were chosen because they best satisfied the criteria that were used to select sites for substitutions of the form Gly → Xaa or Xaa → Pro. To further explore the potential of the method, three additional replacements—G113A, K60P, and A93P—have been constructed and are the subject of this report.

Selection of Sites

Mutations of the Form Xaa → Pro. The first requirement of a site at which a proline is to be introduced is that the conformation of the polypeptide backbone should have (ϕ, ψ) values close to those characteristic of a proline. We required that a site of proline substitution must either have ϕ between $-50°$ and $-80°$, and ψ between 120° and 180°, or have ϕ between $-50°$ and $-70°$, and ψ between $-10°$ and $-50°$. These limits are consistent with the conformations observed for proline residues in well-refined protein crystal structures (see Discussion). Most of the amino acids in T4 lysozyme, an α-helical protein, satisfy the first requirement.

The second requirement relates to the need to avoid unfavorable contacts between the δ-carbon of the proline and the carbonyl carbon and the β-carbon of the residue that precedes the proline. This places restrictions on the (ϕ, ψ) values of the preceding residue, as originally discussed by Schimmel and Flory.[11] In our initial report[1] we required that the preceding residue should have (ϕ, ψ) values within the allowed range given by Schimmel and Flory.[11] As will be discussed below, however, this is unnecessarily restrictive.

The third requirement for a proline substitution is that the introduction of the pyrrolidine ring not cause unfavorable van der Waals contacts with nearby atoms in the folded protein. In practice, such contacts were evaluated with the interactive graphics program FRODO.[12]

The fourth requirement is that the removal of the residue being substituted by proline not remove stabilizing interactions from the folded protein. In practice this generally meant that a candidate site had to be relatively solvent exposed and the replaced residue did not participate in hydrogen-bonding, hydrophobic, or electrostatic interactions with other parts of the protein.

The first site selected for proline substitution (A82P) and described previously[1] met all four of the criteria described at the outset.

Table I Conformational Parameters for Sites of Substitution[a]

	Backbone Dihedral Angles							
	Wild Type				Mutant			
Mutant	ϕ_{n-1}	ψ_{n-1}	ϕ_n	ψ_n	ϕ_{n-1}	ψ_{n-1}	ϕ_n	ψ_n
K60P	−97	168	−58	−46	−86	169	−60	−45
A82P	−90	128	−67	−24	−96	123	−64	−24
A93P	−74	157	−59	−41	na	na	na	na
G77A	—	—	−67	−43	—	—	−64	−44
G113A	—	—	−71	−18	—	—	−72	−9

	Close Contacts			
Mutant	Introduced Atom	Contact Atom	Distance Based on Model Building in Wild Type (Å)	Distance Observed in Mutant Structure (Å)
K60P	C^{γ} Pro 60	C^{β} Thr 59	3.0	3.4
A82P	C^{γ} Pro 80	$O^{\gamma 1}$ Asn 81	3.2	3.3
A93P	C^{γ} Pro 93	C^{β} Asp 92	3.0	na
G77A	C^{β} Ala 77	O Ala 73	3.2	3.4
	C^{β} Ala 77	$O^{\epsilon 2}$ Glu 108	3.1	3.9
G113A	C^{β} Ala 113	O Gly 110	3.1	3.6

[a] The table gives the dihedral angles of the introduced residue (ϕ_n, ψ_n), and in the case of proline substitutions, the preceding residue (ϕ_{n-1}, ψ_{n-1}) for both wild-type lysozyme and the mutant crystal structure. No structure is available for A93P. The contact distances are given based on both model building in the wild-type structure (assuming no relaxation of the structure), and in the actual mutant crystal structure, where available.

After A82P the two most promising candidates appeared to be K60P and A93P, and we therefore proceeded to evaluate these substitutions. The replacements are, however, less than ideal. As can be seen in Table I, model building of the replacement Lys 60 → Pro predicts an unfavorable contact (3.0 Å) with the β-carbon of the preceding residue. This is also true for the replacement Ala 93 → Pro (Table I). In the case of the replacement Ala 82 → Pro there was predicted to be a somewhat close approach (3.2 Å) to the previous residue, but in this case it is to a distal side-chain atom ($O^{\gamma 1}$) rather than a β-carbon, and no backbone movement would be necessary to relieve the contact. Also, the distance (3.2

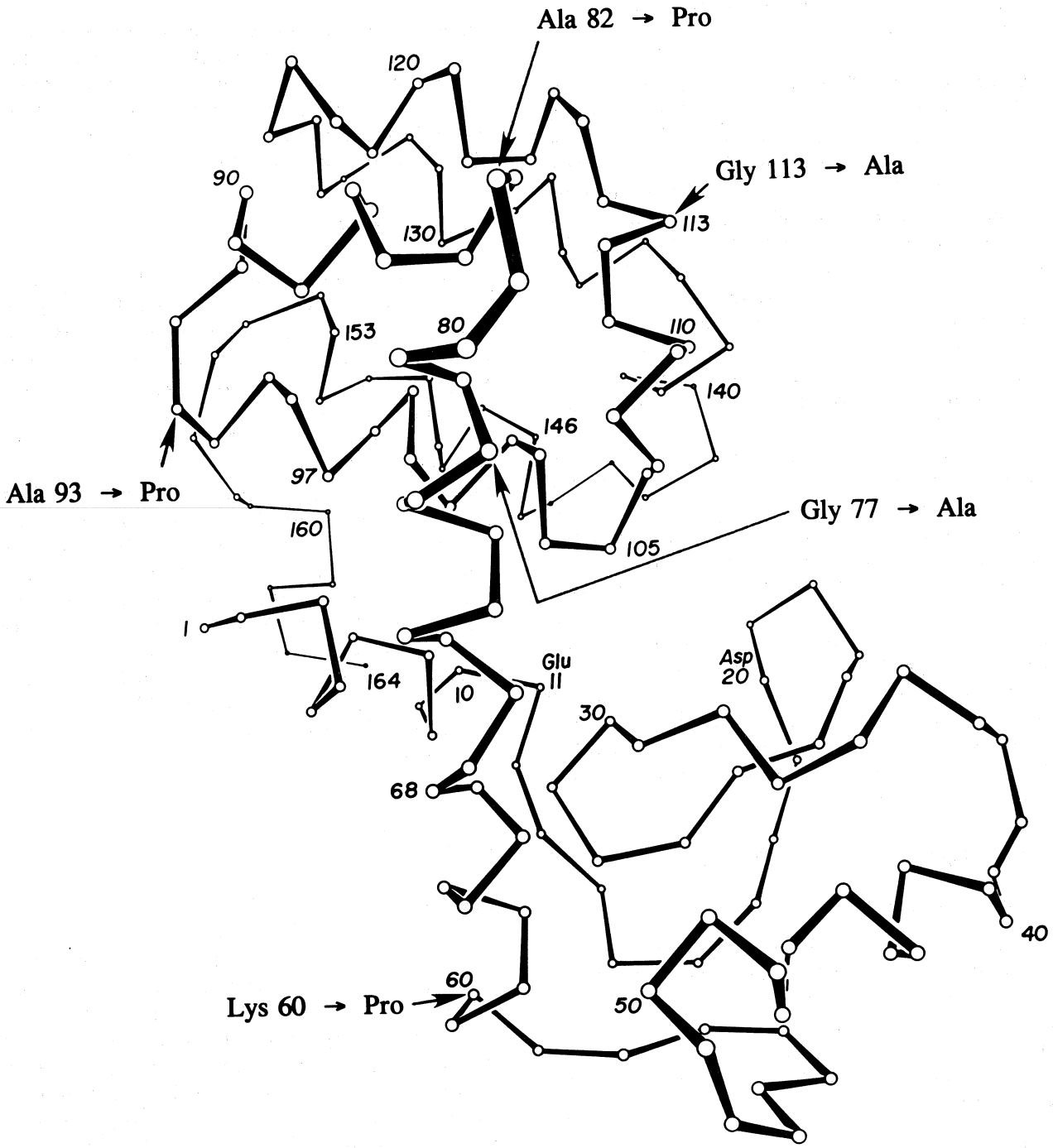

Figure 1. Backbone of T4 lysozyme showing the locations of the substitutions discussed in the text.

Å) is longer than is the case for K60P (3.0 Å) and A93P (3.0 Å). Finally, the contact in the case of A82P involves an oxygen, which has a smaller van der Waals radius than carbon. For all these reasons K60P and A93P were judged to be inferior candidates to A82P. The fraction of the side-chain surface areas of Ala 82, Lys 60, and Ala 93 accessible to solvent in wild-type lysozyme are, respectively, 100, 75, and 100%. The positions of the mutations are shown in Figure 1.

Mutations of the Form Gly → Xaa. The first requirement for the replacement of a glycine was that its (ϕ, ψ) angles be within the region of conformational space normally observed for residues that contain a β-carbon. Glycines that had conformations in the "left-handed helical" region of the Ramachandran diagram ($\phi \sim 60°$, $\psi \sim 60°$) were also excluded from consideration, although subsequent replacements of such residues suggest that this may have been unnecessarily restrictive.[13]

The second requirement was that the residue introduced in place of glycine not introduce any unacceptable van der Waals contacts into the folded protein.

In T4 lysozyme there are eleven glycines. Eight of these were eliminated from consideration by the first criterion given above. Of the remaining three, Gly 77 and Gly 113 appeared to be possible candidates for replacement with alanine, although neither was ideal (Table I). Both these residues are within an α-helix and in each case there were close approaches between the introduced β-carbon and a carbonyl oxygen in the preceding turn of the α-helix. In the case of Gly 77, the introduced β-carbon would be 3.1 Å from the side-chain carboxylate ($O^{\epsilon 2}$) of Glu 108. In wild-type lysozyme the α-carbon of Gly 77 is 9% accessible to solvent while that of Gly 113 is 87% solvent accessible.

MATERIALS AND METHODS

Mutant Lysozymes

Mutant lysozymes K60P, A93P, and G113A were prepared by oligonucleotide-directed mutagenesis.[14,15] Procedures for DNA sequencing and cloning, and protein purification, have been described.[16–18] All proteins were shown to be at least 95% pure as judged by reverse-phase high pressure liquid chromatography.

Activity

The ability of each mutant lysozyme to clear a suspension of cell walls was monitored as a decrease in absorbance at 350 nm.[19] Turbidity assays were performed at pH 7.0 in 50 m*M* Tris at 21°C. Protein concentration varied between 50 and 500 ng/mL, where activity was found to be linear with concentration.

Thermal Stability

Stability toward thermal denaturation was monitored by observing the change in CD at 223 nm.[20] The protein (20 μg/mL) was heated at the rate of 1°C/min until the protein was completely unfolded. Then the temperature was decreased to ensure that the transition was reversible.

Structural Analysis

High-resolution x-ray crystallographic analysis was carried out to investigate structural changes in the native state resulting from the mutations. Mutants K60P and G113A formed crystals in space group $P3_221$, isomorphous with the native T4 lysozyme crystal form, under conditions similar to those used to grow wild-type crystals.[21] Small crystals of A93P that appeared isomorphous with wild-type also formed under these conditions, but none was large enough for data collection. Before x-ray photography, the crystals were equilibrated with a standard mother liquor containing 1.05*M* K_2HPO_4, 1.26*M* NaH_2PO_4, 0.23*M* NaCl, and 1.4 m*M* 2-mercaptoethanol at pH 6.7. High-resolution x-ray data were collected by oscillation photography.[22,23] The crystal structures were refined by the method of conjugate gradient minimization using the TNT package of programs,[24,25] starting with the model of wild-type lysozyme.[21]

RESULTS

Activity

The mutant protein A93P has approximately the same activity (90 ± 10%) as the native enzyme, while G113A has slightly lower activity (80 ± 10%) and K60P has higher activity (175 ± 10%). Several T4 lysozyme mutants that decrease the positive charge on the protein, like K60P, have been observed to display increased activity at low salt concentrations. As has been discussed elsewhere, this effect is thought to be due to the extreme dependence of activity on salt concentrations under conditions of low salt.[26]

Table II Thermodynamic Parameters for Wild-Type and Mutant Lysozymes

Protein	T_m (°C)	ΔT_m (°C)	ΔH (kcal/mol)	ΔS (cal/° · mol)	$\Delta\Delta G$ (kcal/mol)
			Denaturation at pH 2.0		
WT[b]	41.9 ± 0.4	—	89 ± 5	282 ± 16	—
K60P	42.2 ± 0.5	0.3 ± 0.5	—	—	0.1 ± 0.15
G77A[b]	40.5 ± 0.7	−1.4 ± 0.8	85 ± 4	270 ± 13	−0.4 ± 0.2
A82P[b]	42.7 ± 0.1	0.8 ± 0.4	90 ± 5	283 ± 16	0.3 ± 0.15
A93P	42.3 ± 0.2	0.4 ± 0.4	93 ± 5	296 ± 16	0.1 ± 0.1
G113A	42.3 ± 0.4	0.4 ± 0.6	99 ± 7	314 ± 22	0.1 ± 0.15
			Denaturation at pH 6.5		
WT	64.7 ± 0.5	—	129 ± 9	381 ± 27	—
K60P	64.6 ± 0.5	−0.1 ± 0.7	—	—	0.0 ± 0.25
G77A[a]	65.6 ± 0.2	0.9 ± 0.5	125 ± 9	368 ± 27	0.4 ± 0.2
A82P[a]	66.8 ± 0.2	2.1 ± 0.5	126 ± 9	371 ± 26	0.8 ± 0.2
A93P	64.9 ± 0.4	0.2 ± 0.6	121 ± 10	359 ± 29	0.1 ± 0.2
G113A	65.3 ± 0.4	0.8 ± 0.6	120 ± 10	354 ± 30	0.3 ± 0.2

[a] Thermal unfolding at pH 2.0 was performed in 0.2*M* KCl, 10 m*M* HCl, and 10 m*M* H_3PO_4. At pH 6.5, unfolding was in 0.15*M* KCl and 20 m*M* $K_{1.5}H_{1.5}PO_4$. T_m is the temperature at the midpoint of thermal denaturation and ΔT_m is T_m(mutant) − T_m(wild type). Thermodynamic parameters were derived from a van't Hoff analysis of the reversible thermal denaturations. ΔS and ΔH are the entropy and enthalpy of denaturation evaluated at the midpoint of transition T_m. The difference between the free energy of unfolding of the mutant and wild-type proteins, $\Delta\Delta G$, was determined from the relationship $\Delta\Delta G = \Delta T_m \Delta S$, which is generally valid for small perturbations of protein stability.[39] A positive value of $\Delta\Delta G$ indicates that the mutant protein is more stable than the wild type.

[b] Data for wild-type, G77A, and A82P are from Ref. 1.

Thermal Stability

Table II summarizes the thermal denaturation parameters. Mutant proteins K60P and A93P have stabilities essentially the same as wild type at both pH 2.0 and pH 6.5. Mutant G113A is slightly more stable than wild type. At pH 6.5, near the pH of the crystal structures, the changes in ΔS and the ΔH of denaturation are both negative. A decrease in ΔS corresponds to an increase in the stability of the mutant relative to wild type. However, compensating changes in ΔH reduce the free energy relative to that expected on the basis of entropy alone. It must be stressed that the ΔS and ΔH values are subject to a rather large experimental error, and should be regarded with caution. The pH dependence of the melting temperatures of G77A and A82P have been discussed previously.[1] The thermodynamic stabilities of K60P, A93P, and G113A are relatively pH independent. In summary, none of the five mutant proteins (except G77A at pH 2.0) is significantly less stable than the native enzyme. A82P is a stabilizing replacement, and G77A and G113A slightly increase stability at pH 6.5.

Table III Crystallographic Data Collection and Refinement Statistics[a]

Protein	WT	K60P	G113A
Data collection			
Cell dimensions			
a,b (Å)	61.2	61.1	61.0
c (Å)	96.8	96.4	97.5
Unique reflections	19,200	13,581	16,682
Resolution (Å)	1.7	1.8	1.65
R_{merge} (%)	9.1	7.9	5.7
Refinement			
R (%)	16.7	17.3	16.0
$\Delta_{bond\ length}$ (Å)	0.018	0.017	0.018
$\Delta_{bond\ angle}$ (°)	2.2	2.5	2.3

[a] Unique reflections are those reflections used in refinement. R_{merge} is the agreement between intensities measured on different films. *R* is the crystallographic residual after refinement. $\Delta_{bond\ length}$ and $\Delta_{bond\ angle}$ are the rms deviations of the refined models' bond lengths and angles from their "ideal" values. The data for wild type are obtained from Ref. 40.

Structures of Mutant Proteins

The x-ray data collection and refinement statistics for mutants K60P and G113A are summarized in Table III.

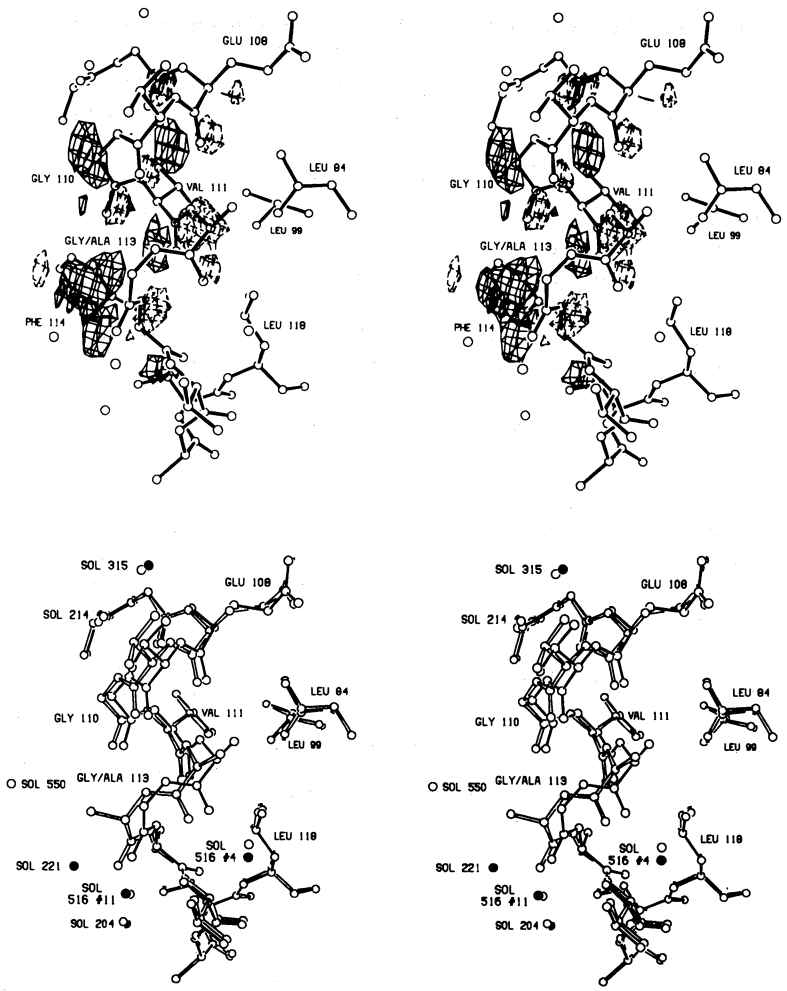

Figure 2. (a) Stereo drawing showing the difference in electron density between G113A and wild-type lysozyme. Coefficients are ($F_{G113A,obs} - F_{WT,obs}$) and the phases are calculated from the wild-type model. The map is contoured at $\pm 4\sigma$, with solid lines for positive features and broken lines for negative features. The resolution is 1.7 Å. (b) Superposition of the refined structure of G113A (open bonds) on that of wild-type (solid bonds). Prior to constructing the figure the structure of the mutant (residues 1–162) was superimposed by least squares onto that of wild type. The wild-type model corresponds to coordinate set T4167RF.[21,41] Solvent molecules in the crystal structures are drawn open in the mutant and solid in the wild-type structure.

Mutant G113A. Gly 113 is at the end of a short α-helix that includes residues 108–113. It is followed by another longer α-helix that encompasses residues 115–123. The map showing the difference in electron density between G113A and wild-type lysozyme [Figure 2(a)] has several large paired positive and negative features along the helix that extend from residues 108 to 113. These features indicate an opening or deformation of the central part of the helix toward the solvent. Superposition of the refined structure of G113A on wild-type lysozyme [Figure 2(b)] indicates that there are significant backbone movements on the solvent-exposed side of the helix, while the buried valine 111 stays relatively fixed. The largest adjustment (1.0 Å) is in the backbone of residue 113, the site of substitution. The next largest backbone shifts (0.5 Å) are at residues 109 and 110 in the preceding turn of the α-helix [Figure 2(b)].

The backbone hydrogen bonds in the vicinity of Gly 113, both intramolecular and to bound solvent, are summarized in Table IV. The Gly 113 → Ala replacement causes a lengthening of most of the hydrogen bonds within the 108–113 α-helix, but somewhat improved hydrogen bonding to solvent.

The site of the G113A mutation is close to a symmetry-related molecule in the crystal. Some solvent molecules are rearranged at the molecular interface,

Table IV Hydrogen-Bond Distances in the Vicinity of the Mutation Site in WT and G113A[a]

H-Bond Partners		WT Distance (Å)	G113A Distance (Å)
109 N	315 Sol[b]	3.4	3.3
109 N	214 Sol	3.8	3.2
109 $O^{\gamma 1}$	315 Sol	3.4	2.9
110 N	214 Sol	3.0	2.8
110 N	107 O	3.1	3.2
111 N	107 O	3.1	3.1
112 N	108 O	2.8	3.0
112 O	516 4 Sol	2.9	3.1
113 N	109 O	3.2	3.6
113 O	204 Sol	3.7	2.8
113 O	221 Sol	2.6	[c]
114 N	111 O	2.8	2.8
114 O	118 N	3.0	2.9
114 O	117 N	3.2	3.0

[a] The list gives all apparent hydrogen bonds involving residues 109–114 with distance involving residues 109–114 with distance 3.3 Å less in wild type or G113A.

[b] "SOL" denotes a solvent molecule. The associated numbering refers to the coordinates as deposited in the Brookhaven Data Bank.

[c] Water molecule 221 is absent in the mutant structure.

but the absence of difference electron density features and of significant coordinate shifts in the symmetry contact region shows that the structural shifts associated with the mutation do not cross the molecular boundary. A solvent molecule (no. 221) bound to wild-type lysozyme is displaced in the mutant structure by a close contact with the β-carbon of Ala 113. At a different position in the mutant structure, there is a new, somewhat disordered solvent molecule (no. 550; thermal factor of 54 $Å^2$), which is trapped between the β-carbon of Ala 113 and the symmetry-related molecule.

Structure of K60P. In contrast to the structure of G113A, the mutation of lysine 60 to proline causes few structural alterations other than the replacement of the side chain itself. The difference electron density map [Figure 3(a)] has a negative feature corresponding to the loss of the lysine side chain and a positive feature due to the addition of the pyrrolidine ring of proline 60. There is also a small negative feature associated with the displacement of solvent by the proline side chain. The negative density superimposed on the side chain of Lys 60 extends only to the δ-carbon. This is as expected because of partial disorder of the lysine side chain. In the refined structure of wild-type lysozyme[21] the thermal factors of the atoms within the side chain of Lys 60 have successive values of 28, 30, 51, 73, and 66 $Å^2$ in going from C^β to N^ζ. The superposition of the refined structure of the mutant on to wild-type lysozyme [Figure 3(b)] indicates that the respective backbone conformations are extremely similar.

In the crystal structure of K60P, the distance from the δ-carbon of the pyrrolidine ring to the β-carbon of Thr 59 is 3.4 Å. Model building suggested that this distance, in the absence of any change in structure, would be 3.0 Å (Table I). The relaxation from 3.0 Å to the observed value of 3.4 Å appears to be achieved by an exaggerated pucker in the pyrrolidine ring of the proline that moves the C^δ atom directly away from the β-carbon of Thr 59. The pyrrolidine ring is puckered in a way that is rarely observed for prolines in well-refined protein structures (data not shown). Presumably there is some backbone strain associated with the distortion of the proline ring, although the backbone torsional angle ϕ of $-60°$ is close to the average value for prolines.[27,28] Thr 59 caps the longest helix in T4 lysozyme. The backbone of this residue does not change its position in the mutant, presumably because several H bonds with the N-terminal end of the helix and the side chain of Glu 62 make it energetically costly for backbone movement to occur.

DISCUSSION

Selection of Sites

Both empirical and theoretical methods were used to evaluate the conformations of the polypeptide backbone that were most compatible with the introduction of a proline residue. (Subsequent to the analysis reported here, an independent survey of proline residues in known protein structures has been given by MacArthur and Thornton.[28])

Figure 4(a) shows the distribution of (ϕ, ψ) values observed for 221 prolines in 37 highly refined protein structures taken from the Brookhaven Data Bank.[29] The potential energy surface for a proline, calculated recently by Summers and Karplus[30] using the CHARMM force field,[31] is superimposed. The restriction of ϕ to values in the vicinity of 60° is, of course, due to the presence of the pyrrolidine ring. There is a discrepancy of 10–30° between the values of ϕ that correspond to the potential energy minima and the values that are most frequently observed. In the case of ψ, the potential energy surface has a minimum extending from $\psi \sim -60°$ to $\psi \sim 180°$, whereas prolines in proteins segregate into two distinct classes, one with $\psi \sim -30°$ and the other with $\psi \sim 150°$ [Figure 4(a)]. In order to select a site for a possible proline substitution, it was required that the polypeptide backbone have values of (ϕ, ψ) that

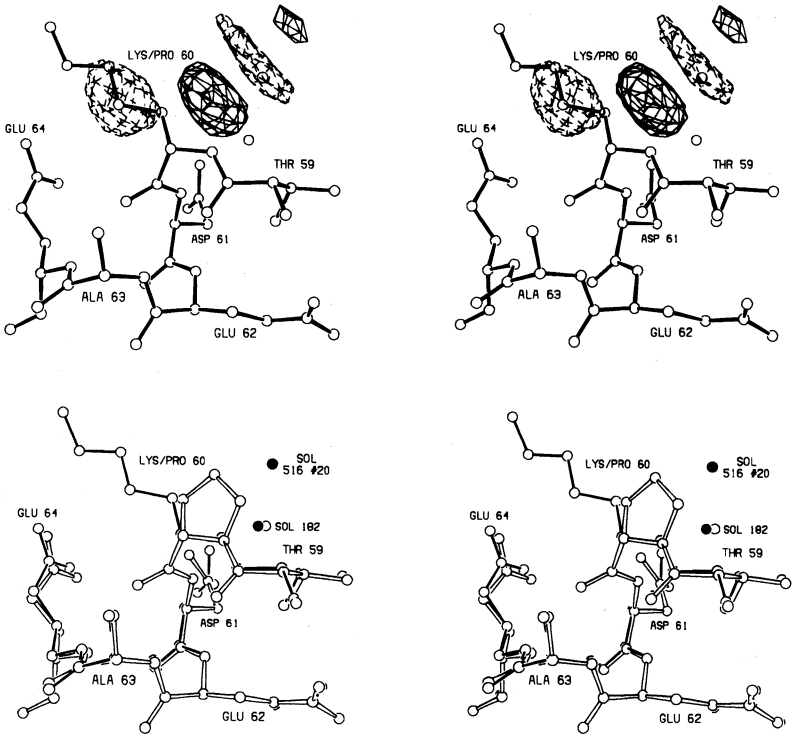

Figure 3. (a) The difference electron density map for the mutant K60P superimposed on the wild-type structure. Coefficients are ($F_{K60P,obs} - F_{WT,obs}$) and the phases are from the wild-type model. Positive contours (solid lines) and negative contours (broken lines) are drawn at $\pm 3.5\sigma$. The resolution is 1.85 Å. (b) Superposition of the refined structure of K60P (open bonds) on the refined wild-type structure (solid bonds). Solvent molecules in the wild-type structure are drawn solid.

fall within one or other of the two densely populated regions shown in Figure 4(a).

The second requirement for the substitution of a proline relates to the need to avoid unfavorable contacts with the preceding residue. This places restrictions on the (ϕ, ψ) values of the preceding residue. These restrictions can be evaluated in two different ways. The first is to calculate the potential energy surface for a residue that is succeeded by a proline, as was initially done by Schimmel and Flory[11] and has been repeated recently by Summers and Karplus[30] using CHARMM.[31] A second, empirical method is to consider the values of (ϕ, ψ) that are observed for residues that precede prolines in well-determined protein structures. Figure 4(b) combines both results. Consistent with the energy calculation, many residues that precede prolines have "extended" conformations with backbone dihedral angles in the vicinity of ($\phi \sim -100°$, $\psi \sim 140°$). In apparent disagreement with the energy surface, however, a number of the residues that precede prolines have an essentially α-helical conformation (i.e., $\phi \sim -60°$, $\psi \sim -50°$). The energy calculation of Summers and Karplus[30] [Figure 4(b)] suggests that the α-helical conformation is disfavored by over 5 kcal/mol relative to the extended conformation. Based on a rather similar calculation, Schimmel and Flory[11] predicted that the α-helical conformation was, in fact, precluded for any residue succeeded by a proline. However, Figure 4(b) shows that approximately 9% of all residues succeeded by proline have an approximately α-helical conformation. This has also been independently noted by MacArthur and Thornton.[28] Possible reasons for this discrepancy are explored in the accompanying manuscript.[32]

Proline Substitutions

As was discussed above, there are several criteria that need to be met in order for a substitution of the form Xaa → Pro not to destabilize the folded protein structure. That an improper placement of a proline can cause drastic reduction in stability is evidenced by the fact that a number of the known destabilizing mutants of T4 lysozyme—e.g., L33P, L66P, Q69P, V71P, D72P, A74P, and L91P[33,34]—involve such substitutions. These provide examples of proline replacements that do not meet the criteria

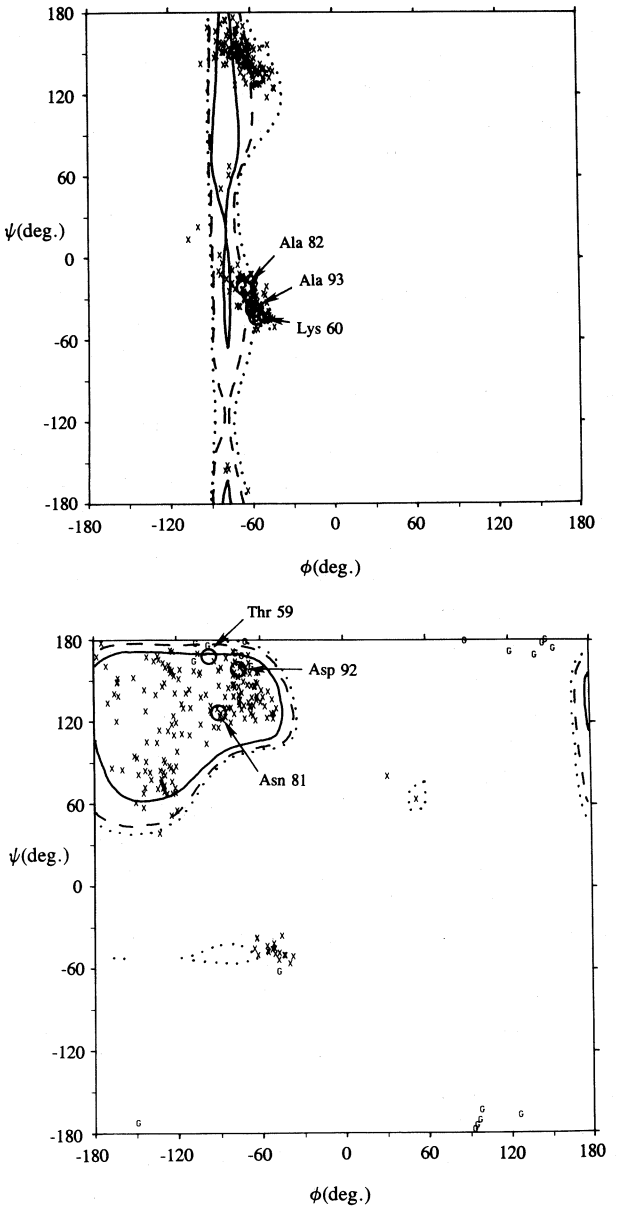

Figure 4. (a) Ramachandran angles observed for 221 prolines in 37 highly refined crystal structures. Superimposed is the energy surface for proline calculated by Summers and Karplus[30] using the CHARMM force field.[31] Continuous, broken, and dotted contours enclose regions that are within 5, 8, and 10 kcal/mol of the global minimum, respectively. Also shown are the dihedral angles of the residues in the crystal structure of wild-type lysozyme that were substituted with proline (from Table I). The basis for selection of the 37 refined protein structures is described by Nicholson et al.[13] To eliminate unreliable regions of structure it was also required that the average crystallographic thermal factor for the backbone atoms be less than or equal to 30.0 Å^2. The Brookhaven identification codes for the structures included are as follows: 2ACT, 3WGA, 2ALP, 2AZA, 5CPA, 4CHA, 3CTS, 1CRN, 1GCR, 4CYT, 2CCY, 2CYP, 2CDV, 351C, 3DFR, 2EBX, 1ECD, 4FXN, 3HHB, 2RHE, 1INS, 2LZM, 1LZ1, 1CTF, 1MBD, 1SN3, 2OVO, 1PPT, 9PAP, 1BP2, 1PCY, 3SGB, 1RN3, 5RXN, 3TLN, 2PTN, and 5PTI. A full set of references is included in Ref. 13. (b) Ramachandran angles observed for residues preceding proline in 37 well-refined crystal structures. The letter G indicates glycine residues and X indicates nonglycine. Superimposed is the potential energy surface for alanine preceding a proline[30] with the conformation of the proline defined by $\omega = -180°$, $\phi = -60°$, and $\psi = -145°$. Also shown are the (ϕ, ψ) values for the residues in wild-type T4 lysozyme preceding sites of proline substitution (from Table I).

for selection of sites described in the Introduction. In an example of an inverse substitution, Pro 86 in T4 lysozyme was replaced with a number of different amino acids and found, in all cases, to slightly destabilize the protein.[35]

Although other effects may be important, the differences in stabilities of the three selected proline mutants A82P, K60P, and A93P correlate with the degree to which unfavorable van der Waals contacts are introduced by the substitutions. The mutant A82P is the most stabilizing. When a proline is model built into wild-type lysozyme at position 82, the only potentially unfavorable contact is 3.2 Å between the δ-carbon and the side-chain oxygen of the preceding residue, Asn 81 (Table I). In the crystal structure of A82P this distance is 3.3 Å. Model building of the other two proline substitutions, K60P and A93P, predicts close contacts of 3.0 Å between the α-carbon and the β-carbon of the preceding residue (Table I). In the K60P crystal structure this close contact relaxes to an observed value of 3.4 Å. In order to achieve this relaxation the pyrrolidine ring of Pro 60 is distorted. Although no crystal structure was obtained in the case of mutant A93P, inspection of the wild-type model indicates that a situation similar to K60P may pertain. In the case of K60P and A93P, not only are the potential van der Waals contacts more severe than in the case of A82P, they also involve main-chain (or at least β-carbon) contacts, which are energetically more costly to relax than contacts with distal side-chain atoms, as is the case with A82P (cf. Ref. 23).

The contacts discussed above between the proline and the residue that precedes it can also be evaluated in terms of the (ϕ, ψ) values of the preceding residue [Figure 4(b)]. For the residues preceding Lys 60 and Ala 93 in wild-type lysozyme, the (ϕ, ψ) values are close to the boundary between the "allowed" and "disallowed" regions indicative of the potential bad contact discussed above. In the case of Ala 82, however, the preceding residue has (ϕ, ψ) values nearer the center of the "allowed" region.

Glycine Replacements

The replacements of both Gly 77 and Gly 113 of T4 lysozyme with alanine give proteins that have melting temperatures, at pH 6.5, that are about 0.8°C higher than wild type (Table II). This is about half of the theoretically expected increase (in the absence of any offsetting factors).[36] At pH 2.0, however, G77A has a melting temperature 1.4°C lower than wild type, and G113A is about the same as wild type (Table II). Gly 77 and Gly 113 are located within α-helices (Figure 1). Model building of both the G77A and G113A structures, based on the refined structure of wild-type lysozyme, predicted unfavorable steric contacts between the introduced β-carbon and oxygen atoms within the preceding turn of the α-helix (Table I). In the observed crystal structures of G77A[1] and G113A (this work), conformational changes were in fact observed within the α-helices that included the substituted residue. In the case of G113A the structural changes are very obvious [Figure 2(b)]. As a result of these changes, the contact distance between the β-carbon of Ala 113 and the carbonyl oxygen of Gly 110 relaxes from 3.1 Å in the model based on wild-type lysozyme to 3.6 Å in the observed crystal structure of G113A. In the case of mutant G77A[1] the close contact relaxes from 3.2 Å in the model to 3.4 Å in the refined crystal structure. Because the mutations Gly 77 → Ala and Gly 113 → Ala perturb the native structure of wild-type lysozyme, it suggests that these replacements introduce strain. One cannot equate the amount of structural distortion with the amount of strain energy. Indeed, the ability of a protein to relax may be a measure of its ability to avoid the introduction of strain.[25] Perhaps the structural changes associated with the G113A mutation are larger than those associated with G77A because the 108–113 helix is relatively mobile whereas the 60–79 helix is more rigid. Backbone atoms within the 108–113 α-helix have higher crystallographic thermal factors than most other parts of the backbone [see Figure 1(b) of Ref. 30]. This suggests that this α-helix is only weakly coupled to the rest of the protein core, and can, therefore, be displaced relatively easily.

The van der Waals contacts between the β-carbon of an introduced alanine and the carbonyl oxygens in the preceding turn of the α-helix are an intrinsic property of an α-helix, and therefore need to be considered at any site within an α-helix at which a glycine to alanine replacement is contemplated. As was the case with the proline substitutions, model building based on the refined wild-type structure predicted in advance that certain close contacts would occur, and subsequent determination of the mutant crystal structures verified the prediction.

In several examples, "inverse" substitutions of the form Xaa → Gly have been shown to slightly decrease the stability of T4 lysozyme.[13,37]

Summary

A total of five mutants in phage T4 lysozyme have been constructed to test the proposition that the folded structure of a protein can be stabilized by appropriately designed mutations of the form Xaa → Pro or Gly → Ala. Such replacements are expected to reduce the entropy of the unfolded state.[1,36] Of the five replacements, three—namely A82P, G77A, and G113A (at pH 6.5)—can be regarded as successful. Sturtevant and co-workers[10] have recently confirmed by differential scanning calorimetry that each of these mutant proteins is approximately 0.5 kcal/mol more stable than the wild-type protein at pH 2.5. Glycine to alanine replacements, especially within α-helices, have also been effective in increasing the thermostability of other proteins.[2–4]

It can be estimated that the reduction in entropy associated with a glycine to alanine replacement corresponds to a free energy change of 0.8 kcal/mol[36] (see also Ref. 5). For an alanine to proline replacement, the estimated value from considerations of entropy alone, and ignoring other factors, is about 1.4 kcal/mol.[1,36] Yun et al.[38] have recently given a substantially lower estimate of 0.5 kcal/mol. In the glycine and proline case the effect is modest, and can easily be offset if the mutation either removes a favorable interaction from the wild-type protein or introduces an unfavorable interaction in the mutant structure. For this reason, the criteria used to select sites for Xaa → Pro or Gly → Ala replacements that will increase protein stability have to be very stringent.

It is encouraging that the availability of a well-refined crystal structure makes it possible to predict substitutions that increase the thermostability of the protein. On the other hand it is disappointing that the number of sites for successful application of the method, at least in T4 lysozyme, seems to be limited.

We are most grateful to Eric Anderson and Drs. James Hurley and Enoch Baldwin for helpful discussions. We also thank Joan Wozniak and Dr. Larry Weaver for technical assistance and advice, Dr. Walter Baase for help, and Dr. J. A. Schellman for providing facilities for the thermodynamic measurements. This work was supported

in part by grants from the NIH (GM21967) and from the Lucille P. Markey Charitable Trust.

REFERENCES

1. Matthews, B. W., Nicholson, H. & Becktel, W. J. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 6663–6667.
2. Hecht, M. H., Sturtevant, J. M. & Sauer, R. T. (1986) *Proteins Struct. Funct. Genet.* **1,** 43–46.
3. Imanaka, T., Shibazaki, M. & Takagi, M. (1986) *Nature* **324,** 695–697.
4. Ganter, C. & Plückthun, A. (1990) *Biochemistry* **29,** 9395–9402.
5. Chakrabarlty, A., Schellman, J. A. & Baldwin, R. L. (1991) *Nature* **351,** 586–588.
6. Suzuki, Y. Oishi, K., Nakano, H. & Nagayama, T. (1987) *Appl. Microbiol. Biotechnol.* **26,** 546–551.
7. Watanabe, K., Chishiro, K., Kitamura, K. & Suzuki, Y. (1991) *J. Biol. Chem.* **266,** 24287–24294.
8. Yutani, K., Hayashi, S., Sugisaki, Y. & Ogasahara, K. (1991) *Proteins Struct. Funct. Genet.* **9,** 90–98.
9. Ellman, J., Mendel, D. & Schultz, P. G. (1992) *Science* **255,** 197–200.
10. Hu, C.-Q., Kitamura, S., Tanaka, A. & Sturtevant, J. M. (1992) *Biochemistry* **31,** 1643–1647.
11. Schimmel, P. R. & Flory, P. J. (1968) *J. Mol. Biol.* **34,** 105–120.
12. Jones, T. A. (1982) in *Crystallographic Computing,* Sayre, D., Ed., Oxford University Press, Oxford, pp. 303–317.
13. Nicholson, H., Söderlind, E., Tronrud, D. E. & Matthews, B. W. (1989) *J. Mol. Biol.* **210,** 181–193.
14. Zoller, M. J. & Smith, M. (1984) *DNA* **3,** 479–488.
15. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. (1987) *Methods Enzymol.* **154,** 367–382.
16. Alber, T. & Matthews, B. W. (1987) *Methods Enzymol.* **154,** 511–533.
17. Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E. & Dahlquist, F. W. (1989) *Methods Enzymol.* **177,** 44–73.
18. Poteete, A. R., Daopin, S., Nicholson, H. & Matthews, B. W. (1991) *Biochemistry* **30,** 1425–1432.
19. Tsugita, A. (1971) in *The Enzymes,* Vol. 5, 3rd ed., Boyer, P. D., Ed., Academic Press, New York, pp. 343–411.
20. Becktel, W. J. & Baase, W. A. (1987) *Biopolymers* **26,** 619–623.
21. Weaver, L. H. & Matthews, B. W. (1987) *J. Mol. Biol.* **193,** 189–199.
22. Rossmann, M. G. (1979) *J. Appl. Cryst.* **12,** 225–239.
23. Schmid, M. F., Weaver, L. H., Holmes, M. A., Grütter, M. G., Ohlendorf, D. H., Reynolds, R. A., Remington, S. J. & Matthews, B. W. (1981) *Acta Cryst. A* **37,** 701–710.
24. Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987) *Acta Cryst. A* **43,** 489–503.
25. Daopin, S., Alber, T., Baase, W. A., Wozniak, J. A. & Matthews, B. W. (1991) *J. Mol. Biol.* **221,** 647–667.
26. Daopin, S., Sauer, U., Nicholson, H. & Matthews, B. W. (1991) *Biochemistry* **30,** 7142–7153.
27. Balasubramanian, R., Lakshminarayanan, A. V., Sabesan, M. N., Tegoni, G., Venkatesan, K. & Ramachandran, G. N. (1971) *Int. J. Protein Res.* **3,** 25–33.
28. McArthur, M. W. & Thornton, J. M. (1991) *J. Mol. Biol.* **218,** 397–412.
29. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, D. F. Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112,** 535–542.
30. Summers, N. L. & Karplus, M. (1990) *J. Mol. Biol.* **216,** 991–1016.
31. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983) *J. Comp. Chem.* **4,** 187–217.
32. Hurley, J. H., Mason, D. A. & Matthews, B. W. (1992) *Biopolymers,* **32,** 1443–1447.
33. Alber, T., Dao-Pin, S., Nye, J. A., Muchmore, D. C. & Matthews, B. W. (1987) *Biochemistry* **26,** 3754–3758.
34. Sauer, U. H., Daopin, S. & Matthews, B. W. (1992) *J. Biol. Chem.* **267,** 2393–2399.
35. Alber, T., Bell, J. A., Daopin, S., Nicholson, H., Wozniak, J., Cook, S. & Matthews, B. W. (1988) *Science* **239,** 631–635.
36. Nemethy, G., Leach, S. J. & Scheraga, H. A. (1966) *J. Phys. Chem.* **70,** 998–1004.
37. Bell, J. A., Becktel, W. J., Sauer, U., Baase, W. A. & Matthews, B. W. (1992) *Biochemistry,* **31,** 3591–3596.
38. Yun, R. H., Anderson, A. & Hermans, J. (1991) *Protein Struct. Funct. Genet.* **10,** 219–228.
39. Becktel, W. J. & Schellman, J. A. (1987) *Biopolymers* **26,** 1859–1877.
40. Matsumura, M., Wozniak, J. A., Daopin, S. & Matthews, B. W. (1989) *J. Biol. Chem.* **264,** 16059–16066.
41. Bell, J. A., Wilson, K. P., Zhang, X.-J., Faber, H. R., Nicholson, H. & Matthews, B. W. (1991) *Protein Struct. Funct. Genet.* **10,** 10–21.

Received January 29, 1992
Accepted April 13, 1992