Contributions of all 20 amino acids at site 96 to the stability and structure of T4 lysozyme

Blaine H. M. Mooers,[1,2] Walter A. Baase,[1,2] Jonathan W. Wray,[1,2] and Brian W. Matthews[1,2]*

[1]Institute of Molecular Biology, Howard Hughes Medical Institute, University of Oregon, Eugene, Oregon 97403-1229
[2]Department of Physics, University of Oregon, Eugene, Oregon 97403-1229

Received 2 January 2009; Revised 11 February 2009; Accepted 12 February 2009
DOI: 10.1002/pro.94
Published online 23 February 2009 proteinscience.org

Abstract: To try to resolve the loss of stability in the temperature-sensitive mutant of T4 lysozyme, Arg 96 → His, all of the remaining 18 naturally occurring amino acids were substituted at site 96. Also, in response to suggestions that the charged residues Lys85 and Asp89, which are 5–8 Å away, may have important effects, each of these amino acids was replaced with alanine. Crystal structures were determined for many of the variants. With the exception of the tryptophan and valine mutants R96W and R96V, the crystallographic analysis shows that the substituted side chain following the path of Arg96 in wildtype (WT). The melting temperatures of the variants decrease by up to ~16°C with WT being most stable. There are two site 96 replacements, with lysine or glutamine, that leave the stability close to that of WT. The only element that the side chains of these residues have in common with the WT arginine is the set of three carbon atoms at the C^{α}, C^{β}, and C^{γ} positions. Although each side chain is long and flexible with a polar group at the distal position, the details of the hydrogen bonding to the rest of the protein differ in each case. Also, the glutamine replacement lacks a positive charge. This shows that there is some adaptability in achieving full stabilization at this site. At the other extreme, to be maximally destabilizing a mutation at site 96 must not only eliminate favorable interactions but also introduce an unfavorable element such as steric strain or a hydrogen-bonding group that remains unsatisfied. Overall, the study highlights the essential need for atomic resolution site-specific structural information to understand and to predict the stability of mutant proteins. It can be very misleading to simply assume that conservative amino acid substitutions cause small changes in stability, whereas large stability changes are associated with nonconservative replacements.

Keywords: T4 lysozyme; electrostatics; strain; temperature-sensitive mutant

Introduction

The temperature-sensitive mutant R96H (Figs. 1 and 2) was the first variant of T4 lysozyme to be structurally characterized.[1] It is 3–4 kcal/mol less stable than the wild-type (WT) protein depending on the conditions[1–3] but the reasons for this drop in stability have been controversial.[4–6]

In the WT protein, the side chain of Arg96 is embedded in a crevice on the surface of the C-terminal domain [Figs. 2 and 3(A)]. Its side chain is held rigidly in place by polar and nonpolar contacts with the backbone carbonyls of several residues and the aromatic

Blaine H. M. Mooers's current address is Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, PO Box 26901, BRC 466, Oklahoma City, OK 73126-0901.

Jonathan W. Wray's current address is Pharmacofore, Inc., 75 Shoreway Road, San Carlos, CA 94707.

Grant sponsor: NIH; Grant number: GM21967.

*Correspondence to: Brian W. Matthews, Institute of Molecular Biology, 1229 University of Oregon, Eugene, OR 97403-1229. E-mail: brian@uoregon.edu

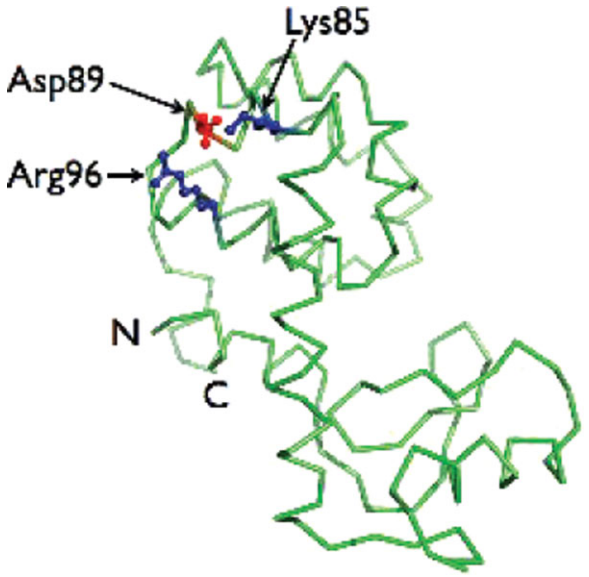

Figure 1. C^{α} trace of the WT T4 lysozyme backbone showing sites Lys85 (blue), Asp89 (red), and Arg96 (blue) on the surface of the C-terminal domain. Sites 85 and 89 were mutated to alanine to isolate the long-range interaction of these sites with site 96. Figure was rendered with PyMoL [http://www.pymol.org]

ring of Tyrosine 88. The guanidinium group of the arginine extends beyond the phenol ring of Tyr88, is partially exposed to the solvent and forms hydrogen bonds (H-bonds) with the backbone carbonyls of Tyr88 and Asp89 [Fig. 2(A)]. One of these H-bonds is lost when Arg96 is replaced with histidine in the mutant protein [Fig. 2(B)].

An analysis of the crystal structure of R96H at a resolution of 1.9 Å[4] suggested two factors that might account for most of the difference in stability between the mutant R96H and the WT protein. The first was steric strain introduced by the imidazole ring; the second was reduced hydrogen-bonding interactions with carbonyl oxygens at the C-terminus of helix D, as illustrated in Figure 2.

As discussed in the recent study,[7] the energetics of the mutant were investigated with computational methods by Tidor and Karplus[5] and Mattos *et al.*[6] The former analysis suggested that the side chains of Asp89 and Lys85 made major electrostatic contributions to the energetics of the site 96 variants.[5] These results seemed somewhat surprising in that these two side chains are solvent-exposed and their charged groups are 5–8 Å from the center of positive charge on the residue at position 96. The latter calculations suggested that neither Lys85 nor Asp89 contributes significantly to the loss of stability of the mutant protein.[6]

In this study, we have made a series of substitutions at sites 85, 89, and 96 to try to understand how both short- and long-range interactions contribute to the destabilization associated with the R96H mutant. The main contributions of the His96 N^{δ} and N^{ε} imidazole nitrogens (see Fig. 2) were evaluated by mutants that included R96N, R96Q, and R96A. Other contributions due to the Arg96 side chain were isolated with the mutants R96V, R96T, R96S, R96A, and R96G. The mutations R96K, R96D, R96E, R96L, R96I, R96F, R96Y, and R96W provided additional information. Possible long-range electrostatic interactions between Lys85, Asp89, and site 96 were isolated with alanine substitutions at sites 85 and 89. The results of these studies support the importance of short-range contributions and establish that long-range interactions between site 96 and the charged side chains of Lys85 and Asp89 are unimportant.

Results

Protein stability

To probe the short-range interactions between the arginine side chain at site 96 and its immediate surroundings, we substituted all the natural amino acids at this site. The stabilities of these variants are summarized in Table I. The original temperature-sensitive mutant R96H is one of the more destabilizing. Substitutions with proline, tryptophan, phenylalanine, and tyrosine cause the greatest loss in stability. At pH 5.35, introduction of the short, polar aspartic acid or asparagine residues at site 96 is significantly more deleterious than truncation of the side chain to alanine. Lysine has the most similar side chain to arginine, and it is not surprising that R96K is almost as stable as WT. The stability of R96Q, however, is also similar to WT, showing that it is not essential to maintain a positive charge. Going one step further, the introduction of a negatively charged glutamate at site 96 is moderately destabilizing.

To isolate possible long-range electrostatic interactions between site 96 and the charged residues Lys85 and Asp89, the alanine mutants K85A, D89A, K85A/R96H, and D89A/R96H were also made. These show that removal of the charged groups at sites 85 and 89 cause very modest changes in stability in the context of either the WT protein or the R96H mutant (Table I).

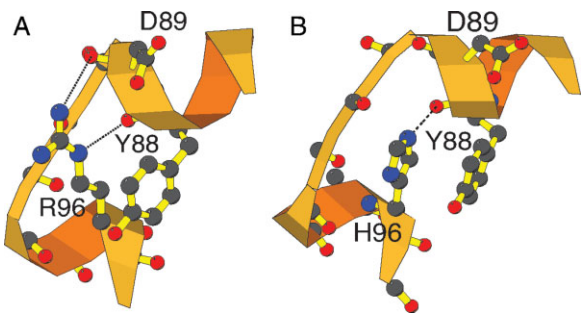

Figure 2. (A) Close-up of Arg96 and its surroundings in the wild-type protein. (B) Cartoon of site 96 in mutant R96H.

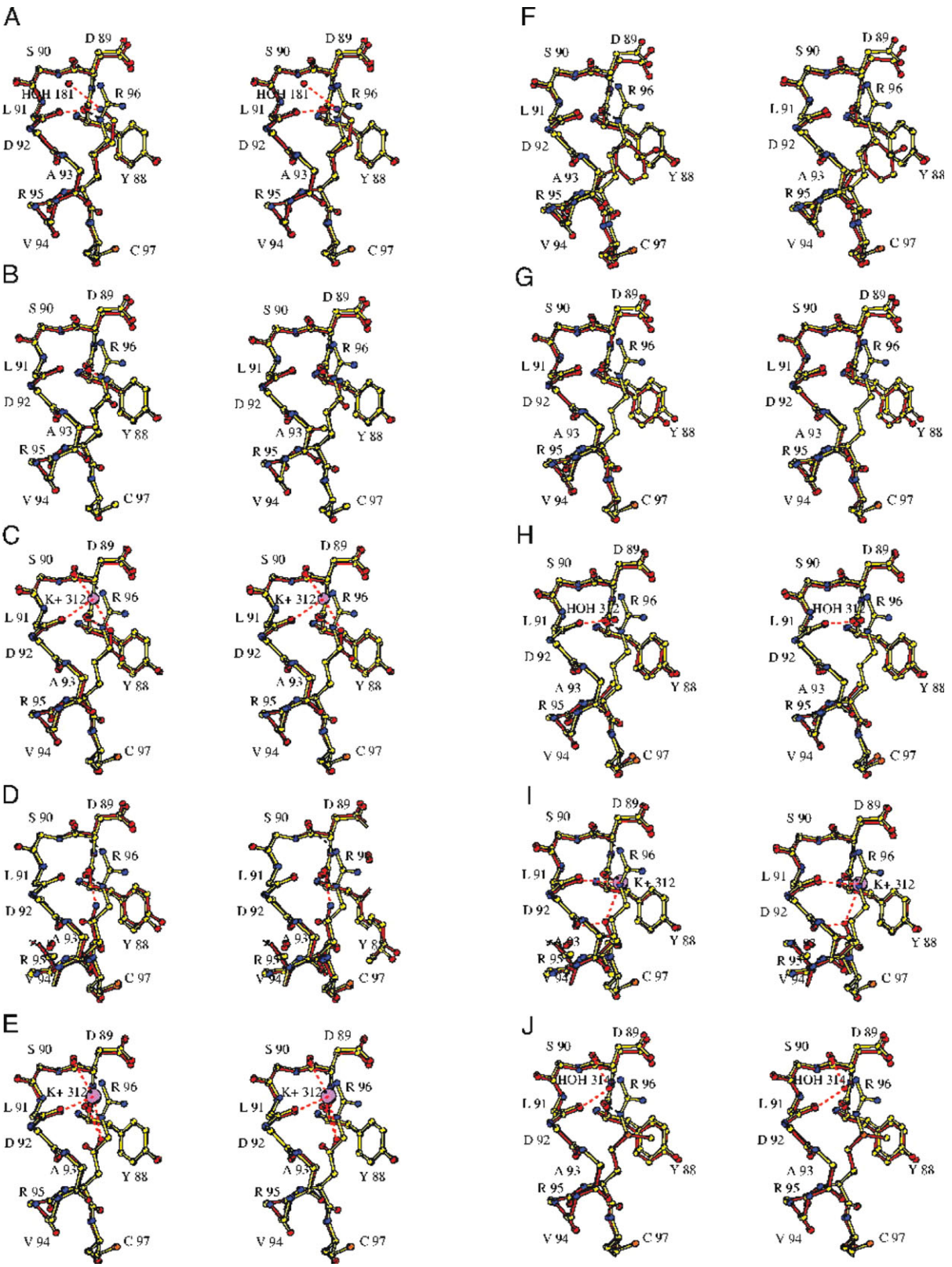

Figure 3. Stereopairs comparing the structures of mutant proteins (red) with WT lysozyme (yellow) in the vicinity of site 96. The one-letter code is used for the amino acids. A pink sphere indicates a presumed potassium ion. All structures are isomorphous with WT. (A) R96K, (B) R96Q, (C) R96E, (D) R96N, (E) R96D, (F) R96Y, (G) R96G, (H) R96A, (I) R96S, and (J) R96M. The structure of R96A has a BME molecule bound to Cys97 (not shown).

Table I. *Stabilities of Mutant Lysozymes*

	Unfolding at pH 5.35			Unfolding at pH 3.05		
Protein	ΔT_m (°C)	ΔH (kcal/mol)	$\Delta\Delta G$ (kcal/mol)	ΔT_m (°C)	ΔH (kcal/mol)	$\Delta\Delta G$ (kcal/mol)
WT	0.0	136	0.0	0.0	142	0.0
R96K	−0.2	138	0.0	−0.7	134	−0.5
R96Q	−1.4	143	−0.3	−1.2	134	−0.7
R96A	−5.1	117	−2.0	−4.3	124	−2.0
R96V	−6.4	113	−2.4	−4.8	98	−2.5
R96S	−7.0	118	−2.6	−6.6	121	−2.8
R96E	−7.0	133	−2.5	−4.1	122	−1.9
R96G	−7.1	120	−2.6	−7.5	112	−3.2
R96M	−7.1	114	−2.7	−5.2	101	−2.5
R96T	−7.6	124	−2.8	−7.2	111	−3.1
R96C	−7.7	100	−2.9	−6.6	105	−2.9
R96I	−7.9	117	−2.9	−5.6	94	−2.8
R96N	−8.0	125	−3.0	−8.0	116	−3.3
R96H	−8.3	121	−3.1	−9.4	113	−3.8
R96L	−8.6	114	−3.2	−7.2	88	−3.2
R96D	−9.5	124	−3.5	−5.2	120	−2.3
R96F	−11.5	108	−4.2	−11.2	73	−4.2
R96W	−12.8	95	−4.5	−11.0	73	−4.2
R96Y	−13.2	99	−4.7	−12.3	73	−4.5
R96P	−15.5	103	−5.5	−16.6	85	−6.0
K85A	−1.0	126	−0.6	0.9	134	0.1
D89A	−1.3	133	−0.5	−1.3	135	−0.7
K85A/R96H	−9.8	122	−3.6	−8.3	109	−3.5
D89A/R96H	−10.2	119	−3.8	−11.2	87	−4.3

The site 96 mutants are ranked according to their melting temperature at pH 5.35. ΔT_m is the change in melting temperature relative to WT, which is 66.6°C at pH 5.35 and 53.2°C at pH 3.05 under the conditions specified in Materials and Methods. ΔH is the enthalpy of unfolding at T_m. $\Delta\Delta G$ is the change in the free energy of unfolding relative to WT [$\Delta G°$(mutant) − $\Delta G°$(WT)]. Uncertainties in ΔT_m are about ±0.2°C, in ΔH about ±5%, and in $\Delta\Delta G$ about 0.15–0.4 kcal/mol. More negative $\Delta\Delta G$ values correspond to proteins that are of lower stability.

Crystal structures

Twelve of the 18 new single mutants at site 96 gave crystals that diffracted X-rays to high resolution (Table II) (i.e., including WT and R96H, 14 of the possible 20 structures could be determined). Ten of the crystal structures were isomorphous with WT and were determined by molecular substitution. They were determined at room temperature to be consistent with earlier work done by Weaver *et al.*[4] on the mutant R96H. R96W crystallized in space group $P2_1$ with four molecules in the asymmetric unit and R96V crystallized in space group $P2_12_12_1$ with four molecules in the asymmetric unit. X-ray data for both R96W and R96V required synchrotron radiation and had to be determined at 100 K.

The X-ray data for the four alanine mutant proteins K85A, D89A, K85A/R96H, and D89A/R96H were measured at room temperature to be consistent with earlier work done by Weaver *et al.*[4] on the mutant R96H. The four mutants crystallized isomorphously with WT and gave X-ray diffraction to high resolution (Table II). Structure determination and refinement of K85A and D89A (not shown) show that the truncation of either side chain causes essentially no change relative to the structure of the WT protein. Likewise, the structures of K85A/R96H and D89A/R96H (not shown) are essentially identical to that of the mutant protein R96H.

Discussion

Long-range electrostatic interactions

In the initial simulation analysis of the R96H mutant by Tidor and Karplus,[5] it was concluded that the charged groups on the side chains of Lys85 and Asp89, which are 5–8 Å from the side chains at site 96, contribute substantially to the destabilization of the mutant. Specifically, the electrostatic influence of Lys85 was calculated to stabilize the WT relative to R96H by 3.3 kcal/mol, whereas Asp89 was estimated to be destabilizing by 6.5 kcal/mol. In subsequent continuum calculations in which effects of solvent screening were included,[6] the estimated effects of Lys85 and Asp85 were very substantially reduced (from 3.3 to 0 kcal/mol and from 6.5 to 0.4 kcal/mol). At the same time, however, the side chain of Asp92, which contributed less than 0.5 kcal/mol to the relative stability of WT in the initial calculations, increased to 1.6 kcal/mol in the latter.

As shown in Table I, replacement of Lys85 by alanine in the context of the WT protein reduces stability slightly at pH 5.35 ($\Delta T_m = -1$°C) and increases stability by the same amount at pH 3.05. The replacement of Asp89 by alanine reduces stability slightly at both pH values. If the Lys 85 → Ala substitution is made in the context of R96H, i.e., K85A/R96H, the overall loss in stability at pH 5.35 is −3.6 kcal/mol. This is

Table II. *X-Ray Data Collection and Refinement Statistics*

	Cell dimensions		Resolution	R_{merge}	Completeness	R	Δ_{bond}	Δ_{angle}	PDB
	a,b (Å)	c (Å)	(Å)	(%)	(%)	(%)	(Å)	(°)	code
Mutants at site 96									
R96A	61.0	97.1	23.2–1.95	8.0	83	21.1	0.013	1.4	3C7Y
R96D	61.0	97.5	29.1–1.63	7.5	99	18.6	0.018	1.3	3C8Q
R96E	61.0	97.3	19.6–1.68	8.2	88	20.4	0.018	1.3	3C8S
R96G	61.1	97.8	30.5–1.8	6.7	93	17.4	0.010	1.1	3C8R
R96K	60.9	97.6	18.5–1.77	7.4	95	18.7	0.014	1.8	3C7W
R96M	60.9	97.5	22.2–1.73	7.0	88	20.0	0.011	1.4	3CDV
R96N	60.9	97.2	15.7–1.63	7.6	92	19.5	0.011	1.2	3CDT
R96Q	61.0	97.6	23.2–1.7	8.0	89	19.9	0.013	1.0	3CDR
R96S	60.9	97.4	15.4–1.68	8.1	82	19.6	0.012	1.0	3CDO
R96Y	60.4	97.0	19.8–2.0	4.8	90	19.9	0.010	1.2	3C80
R96V	[a]	[a]	23.8–1.87	5.9	99	18.0	0.012	1.2	3CDO
R96W	[a]	[a]	18.8–1.53	4.4	92	19.4	0.006	0.9	3FI5
Alanine mutants at sites 85 and 89									
K85A	61.0	97.3	22.2–1.85	8.0	96	19.0	0.013	1.0	3C81
D89A	61.0	97.1	22.2–1.84	6.6	95	18.3	0.012	1.3	3C83
K85A/R96H	60.9	96.8	13.2–1.68	5.3	93	18.5	0.012	1.2	3C82
D89A/R96H	60.9	96.6	11.4–1.67	5.9	95	18.7	0.011	1.1	3C7Z

R_{merge} gives the average difference in agreement between the independently measured diffraction intensities. R is the crystallographic residual for the refined model. R_{free} was not used for the mutants isomorphous with wildtype, since the refinement protocol is well established. For the nonisomorphous R96V structure R_{free} was 0.234; for R96W R_{free} = 0.261. Δ_{bond} and Δ_{angle} are the average discrepancies of the bond lengths and bond angles from expected stereochemistry.

[a] Wild-type lysozyme crystallizes in space group $P3_221$ with cell dimensions $a = b = 61.2$ Å, $c = 96.8$ Å. R96W has space group $P2_1$ with cell dimensions $a = 73.80$ Å, $b = 56.11$ Å, $c = 85.24$ Å, and $\beta = 106.47°$. R96V has space group $P2_12_12_1$ with cell dimensions $a = 57.82$ Å, $b = 99.32$ Å, and $c = 123.01$ Å.

essentially equal to the sum of the stability loss arising from the constituent single mutants (−0.6 to 3.1 = −3.7 kcal/mol). The same is true, within experimental error, for the double mutant D89A/R96H. Thus, removal of charge at sites 85 or 89 has very little effect on the stability of either WT or R96H lysozyme. The finding that the changes in $\Delta\Delta G$ are additive for substitutions at, respectively, sites 85 and 96, or 89 and 96, indicates that there is little if any interaction between either of these pairs of residues. In other words, there is no experimental evidence that either Lys85 or Asp89 contributes significantly to the destabilization of the temperature-sensitive mutant R96H.

Contributions to stability at site 96

In dissecting the contributions to stability at site 96, it is convenient to group the site 96 mutants according to their effects on stability. First, we consider the "stabilizing" variants, R96K and R96Q, that have stabilities close to that of WT. Then, we consider the least stable variants, including the original temperature-sensitive mutant R96H. We believe that the consideration of these two groups allows one to identify the key features that either stabilize or destabilize this family of proteins. Finally, we discuss the remaining "intermediate stability" variants.

Stabilizing variants. Because the two mutants R96K and R96Q have stabilities close to that of WT, they must maintain or replace all of the favorable features exemplified by Arg96. What do WT, R96K and R96Q have in common? A proximal element that is strictly conserved is the carbon chain that extends from C^{α} to C^{γ}. In addition, they all have at least one distal H-bond from the site 96 side chain to the rest of the protein, although the details are different. In the case of WT, the guanidinium group of Arg96 makes two H-bonds, the first to the backbone carbonyl oxygen of Asp89 and the second to the backbone carbonyl oxygen of Tyr88 (see Fig. 2). Neither Lys96 in R96K [Fig. 3(A)] nor Gln96 in R96Q [Fig. 3(B)] makes a H-bond to the backbone carbonyl of Asp89, showing that maintenance of this interaction is not essential. Gln96 in R96Q does maintain a H-bond to the carbonyl oxygen of Tyr88 (2.7 Å, N^{ε}...C—O angle of 146°, O...N^{ε}—C^{δ} angle 120°). In contrast, Lys96 in R96K lacks this H-bond but does make one to the backbone carbonyl oxygen of Leu91 (3.1 Å). Because R96Q is close in stability to WT, it shows that a positively charged side chain is not essential at site 96. In contrast, the lysine side chain in R96K carries a positive charge as does Arg96 in WT, but its mode of hydrogen bonding to the protein is completely different. This shows that a degree of adaptability is possible. To achieve full stabilization at site 96, one clearly needs the three carbon atoms within the proximal part of the side chain. In the distal region, one needs at least one direct H-bond to the protein but no particular H-bond is required.

In the above context, one can compare the situation for R96E. The crystallographic refinement of this mutant suggests that there is a potassium cation

bound by the side chain of Glu96 and the backbone carbonyl oxygens of Asp89 and Leu91 [Fig. 3(C)]. (If a water molecule or sodium ion is tested at this site, the crystallographic B-factors are lower than those of the liganding atoms, which is unreasonable. It is also known that a rubidium ion can bind at this site.[8]) One of the carboxylic oxygens of the side chain of Glu96 is 3.3 Å from the backbone carbonyl of Tyr88. This is 0.6 Å greater than the corresponding distance for Gln96 in R96Q. It suggests that Gln96 makes a good H-bond to the carbonyl of Tyr88 but Glu96 is, at best, only weakly hydrogen bonded. Given that the intrinsic pK_a of glutamic acid is 4.3–4.5,[9] this is as expected. In the crystal, at pH 6.7, Glu96 is presumably largely deprotonated and, therefore, incapable of donating a H-bond to a backbone carbonyl group. It is also the case that the R96E mutation is destabilizing and somewhat more so at pH 5.35 than at pH 3.05 (Table I). It would be expected that as the pH increases from 3.05 to 5.35, the Glu96 side chain would become more negatively charged. To the extent that this occurs, Glu96 can no longer H-bond to the carbonyl of Tyr88. Whether it is the development of the negative charge or the loss of the H-bond that causes this pH-dependent loss in stability is an open question. In any event it is deleterious to have a negative charge at site 96.

Least stable variants. The most destabilizing variants are R96F, R96W, R96Y, and R96P (Table I). It is also convenient to consider R96D, R96L, R96N, and R96H as part of this set. R96H, the original temperature-sensitive mutant, is discussed in the accompanying article (Ref. 7). The first thing to be noted is that this group does not include R96A or R96G. In other words, the deletion of the nonpolar C^{β} and C^{γ} atoms, together with the loss of the H-bond with the carbonyl oxygen of Tyr88, is not of itself sufficient to maximally destabilize the protein. Some other nonfavorable factor must come into play. The crystal structure of R96N is shown in Figure 3(D). Because the side chain of Asn96 is short, polar atoms replace those that are nonpolar in WT. The crystallographic analysis does not immediately allow one to distinguish between the amide and carbonyl groups of the asparagine side chain. Nonetheless, inspection of B-factors with the side chain refined in the alternative configurations, and consideration of the hydrogen-bonding potential, suggests that the amide group is hydrogen bonded to the carbonyl oxygen of Tyr88. The $N^{\delta 2}$...O—C angle is 130°, the C^{γ}—$N^{\delta 2}$...O angle is 110°, and the distance is 3.0 Å. Conversely, the carbonyl oxygen of the Asn96 side chain is directed toward the hydrophobic core of the C-terminal domain with no apparent H-bond partner. This is clearly unfavorable[10] and contributes to the reduced stability of this mutant.

An almost identical situation occurs in the structure of R96D but with the addition of a presumed potassium counter-ion [Fig. 3(E)]. Again, part of the polar side chain is forced into a hydrophobic pocket. Relatively speaking, the R96D mutant is destabilized substantially more at pH 5.35 than at pH 3.05 (Table I). Given that the intrinsic pK_a of aspartic acid is 3.9–4.0,[9] the pH range 3.05–5.35 is that within which the side chain tends to become negatively charged. The presence of a negative charge in a relatively hydrophobic pocket explains the deleterious effect of this mutation at pH 5.35.

It is not surprising that the introduction of proline is very deleterious. Site 96 is located within helix E (residues 93–106). Substitution of proline at this site not only deletes most of the Arg96 side chain but also introduces severe steric clashes within the α-helix, which, at minimum, will disrupt the helical hydrogen bonding of Asp92 and Ala93.

Crystals of the proline mutant could not be obtained, but the structure of the low-stability variant R96W could be determined (see Fig. 4) and shows directly that strain can play an important role. Even though the indole nitrogen of a tryptophan is essentially isosteric with the N^{ε} atom of a histidine, the Trp96 and His96 side chains adopt completely different conformations. (This may be a further indication that the H-bond between His96 and the carbonyl oxygen of Tyr88 is weak.) Presumably, because of the bulk of the Trp96 side chain, and the associated conformational changes, R96W crystallizes in a different form (Table II). The rotameric state of the tryptophan (chain A, $\chi^1 = -172°$, $\chi^2 = 23°$) differs substantially from that of Arg96 ($\chi^1 = -81°$, $\chi^2 = 179°$). This aligns the indole ring at about 90° to that of the arginine. This alignment avoids a steric clash with the backbone at Leu91 and Asp92 and requires that the side chain of Tyr88 swing toward Asp89 by about 40°. Asp89 in turn has a large displacement along with shifts within helix D (residues 82–90). The rotation of Tyr88 changes χ^1, χ^2 from 189°, 88° to 151°, −97°. According to the survey of Blaber *et al.*,[11] 63% of tyrosines in helices have χ^1, χ^2 values of 178(8)°, 82(13)° where the number in parentheses is the standard deviation. The rotamer conformation of Tyr88 in WT is therefore unstrained and very common. In contrast, the value of χ^1 of Tyr88 in the R96W mutant is almost four standard deviations away from the low-energy minimum. These changes also come at the cost of leaving the indole nitrogen N^{ε} as an unsatisfied H-bond donor.

In the case of R96Y, part of the substituted tyrosine side chain mimics the geometry of Arg96 [Fig. 3(F)]. The χ^1 value of Tyr96 is 77° and χ^2 is 51°. In the 100 refined protein structures analyzed by Blaber *et al.*,[11] only one of 56 tyrosines within helices and none of 116 phenylalanines within helices had this conformation. In addition, the $C^{\delta 2}$ atom of Tyr96 is 3.6 Å from the backbone carbonyl of Tyr88 and prevents this carbonyl hydrogen bonding to water. The solvent molecule is forced to move away and H-bond only to the backbone carbonyls of Asp89 and Leu91. The unsatisfied hydrogen-bonding potential of the backbone carbonyl of Tyr88 also contributes to the loss of stability of R96Y. Although we do not have a

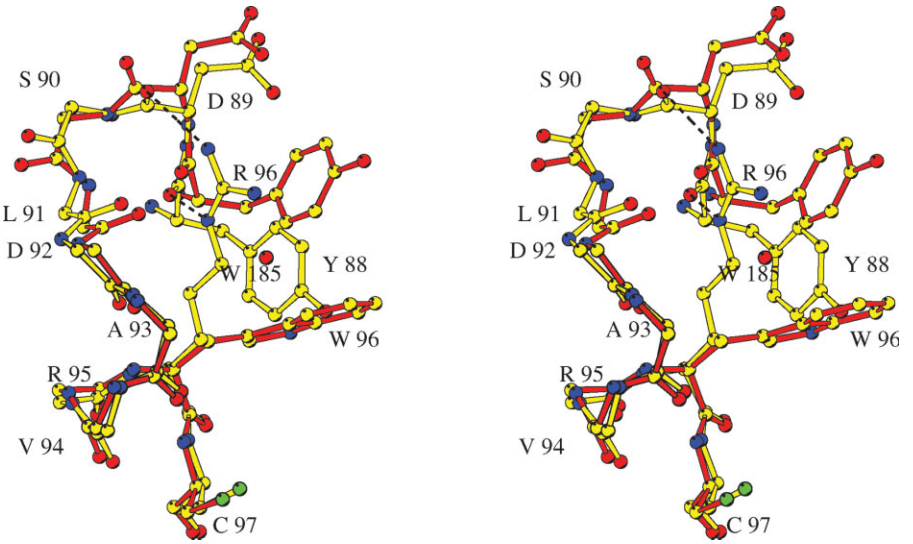

Figure 4. Comparison of the structure of chain A of the nonisomorphous structure of R96W with WT. All conventions are as in Figure 3.

crystal structure for R96F, the factors that destabilize R96Y would also be expected to apply to the phenylalanine substitution.

Variants of intermediate stability. The variants that have intermediate stability can broadly be grouped into two classes. (1) Those that largely truncate the Arg96 side chain. (2) Those that replace Arg96 with a side chain that is nonoptimal in shape and/or chemistry.

The "truncation" variants include R96G, R96A, and R96S. The structures of these three are shown, respectively, in Figure 3(G–I). Even the complete removal of the Arg96 side chain, as in R96G [Fig. 3(G)], has very little effect on the structure of the protein. This is also true for R96A. In the case of R96S [Fig. 3(I)], the hydroxyl of the serine side chain ($\chi^1 = -58°$) is directed in such a way that it H-bonds to the carbonyl oxygen of Asp92 (distance 2.8 Å, C^β—O^γ...O angle 113°, O^γ...O—C angle 129°). The variants that remove or substantially shorten the side chain at site 96 and expose the carbonyl oxygens of Tyr88, Arp89, and Leu91 to solvent create one or more solvent-binding sites. Typically, a water molecule H-bonds to all three of these carbonyl groups. In the case of R96S, this solvent molecule moves somewhat to make a H-bond (2.9 Å) with the introduced Ser96 side chain. It retains H-bonds with the backbone carbonyls of Tyr88 and Leu91 but not Asp89 (4.4 Å).

The "nonoptimal" mutants for which crystal structures could be obtained are R96V and R96M. R96V crystallized in space group $P2_12_12_1$ with four molecules in the asymmetric unit, all with very similar structures. One of these is shown in Figure 5. The C^α and C^β atoms of Val96 superimpose on the C^α and C^β of Arg96 in WT lysozyme. The $C^{\gamma 2}$ atom of the valine projects "outward" and causes the side chain of Tyr88 to move about 1.6 Å. The average conformational angles of Val96 in the four mutant structures are ($\chi^1 = 173°$, $\chi^2 = 88°$) compared with ($\chi^1 = 189°$, $\chi^2 = 88°$)

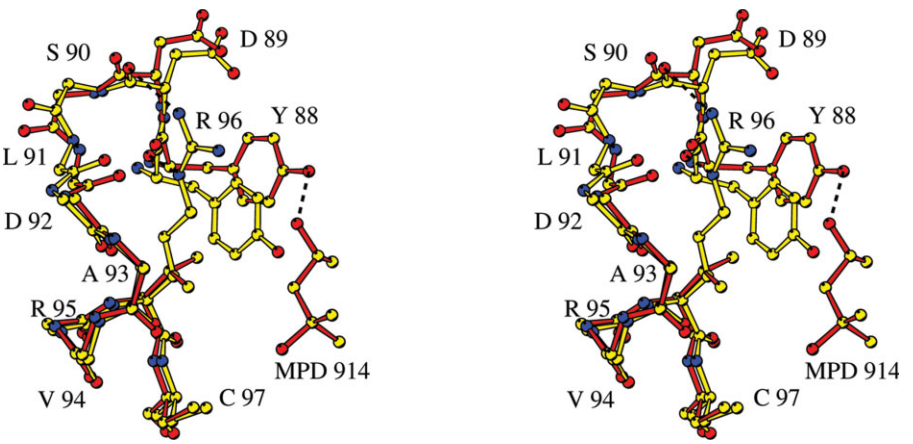

Figure 5. Comparison of the structure of chain A of the nonisomorphous structure of R96V with WT. All conventions are as in Figure 3.

in WT. The planes of the peptide bonds flanking each side of the C^{α} atom of Tyr88 move away from planarity by 7–11°. The R96V mutation also created a specific binding site for 2-methyl-3,4-pentanediol (MPD), seen for all four molecules in the asymmetric unit. The MPD molecule straddles the crevice at site 96, is hydrogen bonded to the phenolic oxygen of Tyr88, and partly occludes the hydrophobic side chain of Val96 from solvent (see Fig. 5).

In the case of R96M [Fig. 3(J)], the C^{β}, C^{γ}, and S^{δ} atoms closely follow the path traced out by C^{β}, C^{γ}, and C^{δ} of Arg96 in WT, preserving some of the favorable hydrophobic interactions. The terminal methyl group of Met96, however, is directed into the solvent. As is the case for a number of the other substitutions, a solvent molecule appears to H-bond to the backbone carbonyl oxygens of Tyr88, Asp89, and Leu91.

Overall, the characteristics of all of the substitutions at site 96 are consistent with the following principles.

1. To achieve maximal stability, the proximal part of the site 96 side chain should have carbon atoms at the C^{α}, C^{β}, and C^{γ} positions. In addition, the distal part of the site 96 side chain should include a polar function, either positively charged or neutral. This polar group must be capable of forming H-bonds to at least one of the three backbone carbonyl groups that surround the site 96 site. (The remaining H-bond potential of these groups is satisfied by solvent.)
2. To be maximally destabilizing a mutation must not only eliminate the above favorable components but must also introduce an unfavorable element such as strain or unsatisfied hydrogen bonding.

Implications for energy calculations

The results also have implications for methods used to estimate protein stability. In particular, the changes in stability and structure associated with the different substitutions at site 96 (Table I) would be impossible to predict using simplistic (i.e., nonatom-based) energy terms based on side chain similarity or on assumed pairwise interactions between amino acids that are close together in the 3D structure. For example, if one simply assumed that electrostatic interactions are dominant, it would be predicted that the Lys and His substitutions at site 96, which retain the positive charge of Arg96, would be most stable, whereas Glu and Asp, which reverse the charge, would be most destabilizing. This is clearly not the case (Table I). (As noted, e.g., by Pace *et al.*,[12] possible changes to the unfolded state always need to be kept in mind.) Likewise, although most believe that the hydrophobic effect is dominant in protein folding, this does not, alone, account for the effects of the substitutions at site 96. For example, R96A ($\Delta T_m = -5.1$°C) is more stable than R96L ($\Delta T_m = -8.6$°C) and R96W ($\Delta T_m = -12.8$°C), even though the latter have the potential for larger hydrophobic contributions. It has also been suggested that hydrogen bonding may be more significant than hydrophobic interactions in stabilizing folded proteins.[13] Against this idea, Arg96 in WT lysozyme makes two H-bonds to the rest of the protein, whereas only one is made in R96K [Fig. 3(A)] and R96Q [Fig. 3(B)]. R96K ($\Delta T_m = -0.2$°C) and R96Q ($\Delta T_m = -1.4$°C) are, however, almost as stable as WT. R96H ($\Delta T_m = -8.3$°C) and R96N ($\Delta T_m = -8.0$°C) also make a single H-bond but are much less stable than R96K and R96Q. Finally, one could consider the entropy cost associated with localizing the side chain. This is largest for the long side chains Arg (−1.88 kcal/mol), Lys (−1.89 kcal/mol), and Gln (−1.73 kcal/mol), which have the largest number of rotatable bonds.[14] Entropy alone would predict that WT, R96K and R96Q would be the least stable variants whereas they actually have the highest stability.

It can also be misleading to assume that conservative amino substitutions tend to cause small changes in stability, whereas larger stability changes are associated with nonconservative replacements. Asn and Gln, for example, are similar, but R96N and R96Q differ substantially in stability. Conversely, Gly and Met are dissimilar but R96G and R96M have identical stability.

Likewise, conservation of structure is no guarantee for conservation of stability. The tyrosine at site 96 has a conformation similar to the WT arginine [Fig. 3(F)] but is very destabilizing ($\Delta T_m = -13.2$°C). In contrast, a valine at the same site is more disruptive (see Fig. 5) but less destabilizing ($\Delta T_m = -6.4$°C).

The need to take into account the specific environment and to use potential functions based on the detailed atomic structure is obvious. Even here, however, there is room for improvement. The early calculations of Tidor and Karplus[5] correctly showed that the $C{-}C^{\alpha}{-}C^{\beta}$ angle at His96 in the R96H mutant is not significantly distorted. On the other hand, their suggestion that distant electrostatic interactions with Lys85 and Asp89 make major contributions to the loss of stability in R96H is not substantiated by the K85A and D89A mutants ($\Delta T_m \sim 1$°C; Table I). In the more recent simulations of Mattos *et al.*,[6] the calculated electrostatic contributions of Lys85 and Asp89 were very much reduced, consistent with the experimental data.

At the same time, there are major changes in a number of the other energy components, some of which are nonintuitive. For example, the largest electrostatic energy term identified by Mattos *et al.*[6] is that contributed by the carbonyl group of Asp92 (3.68 kcal/mol). This carbonyl group is involved in an intrahelical H-bond whose geometry is essentially invariant in all of the site 96 variants (see Fig. 3). A calculated interaction favoring the mutant presumably occurs because the positive charge on the side chain of His96 is closer to the carbonyl group than that of Arg96 [Fig. 1(A) of Ref. 7]. This difference is calculated to contribute 3.68 kcal/mol to the relative stability of the mutant, but overall the mutant is measured to be 3.1 kcal/mol less stable than WT

(Table I). Therefore, there have to be other offsetting effects. The next largest energy term is calculated to come from the carbonyl group of residue 96 and to stabilize WT relative to R96H by 2.76 kcal/mol.[6] This carbonyl group participates in an essentially identical intrahelical H-bond in all the structures and does not appear to interact with the side chain at site 96. It is not obvious why this carbonyl group should substantially stabilize WT relative to R96H.

Methods

Mutagenesis, protein expression, and purification

All of the mutant proteins were constructed in the WT background that includes cysteines at sites 54 and 97. The single alanine mutants at sites 85 and 89 as well as the double mutants with R96H were made by site-directed mutagenesis using the method of Kunkel.[15] The 18 single mutants at site 96 were made by two-stage polymerase chain reaction (PCR)[16] using the gene for the WT T4 lysozyme as the template. The BamHI/HindIII-digested PCR products were ligated into the vector pHS1403. The DNA sequences of the new constructs were confirmed by automated methods incorporating the PCR (Perkin-Elmer ABI PRISM 377 DNA sequencer). The vectors were transformed into *E. coli* RR1 cells for overexpression. All of the new mutant proteins were overexpressed as soluble protein and were purified by standard methods.[17–19] The molecular masses of the mutant proteins were checked with an ABI Biosystems Voyager-DE MALDI/TOF mass spectrometer. The buffer used for protein storage was 0.1*M* sodium phosphate pH 6.5, 0.55*M* NaCl, and 0.02% NaN_3. All of the mutants were catalytically active.

Thermal unfolding

Circular dichroism-monitored thermal stability data were collected at 223 nm initially using a JASCO model J-720 spectropolarimeter and the JASCO PTC-348W thermal control system.[20] Later data were collected using a JASCO model J-810 spectropolarimeter and the JASCO PDF-425S. For data at pH 5.35, the buffer was 100 m*M* sodium chloride, 1.4 m*M* acetic acid, 8.6 m*M* sodium acetate, with protein concentrations of 0.01–0.03 mg/mL as determined from optical density at 280 nm.[21] (It should be noted that Na^+ and K^+ were included in the crystallizing solution and the latter ion appeared to bind to the Asp, Glu, and Ser mutants at site 96.) Data at pH 3.05 were in 20 m*M* glycine/glycine–HCl, 1 m*M* Na_2H_2 EDTA. Unfolding profiles were analyzed by means of the two-state model to determine the temperature of melting (T_m) and the van't Hoff enthalpy at the melting temperature (ΔH).[22] Values of T_m and ΔH were used to calculate ΔG° at 58°C (pH 5.35) and 44°C (pH 3.05) by means of an integrated form of the Gibbs–Helmholtz equation[2] assuming a ΔCp of 2.5 kcal/mol-K at pH 5.35 and 1.8 kcal/mol-K at pH 3.05. $\Delta\Delta G$ values were computed as ΔG°(mutant) − ΔG°(WT).

Crystallization

Ten of the site 96 mutants were crystallized in space group $P3_221$ isomorphously with the WT protein in 1.8–2*M* K/Na phosphate buffers, pH 6.5–7.1, containing ~50 m*M* oxidized β-mercaptoethanol (HEDS) and 50 m*M* reduced β-mercaptoethanol (BME) as previously described.[23] The alanine mutants K85A, D89A, K85A/R96H, and D89A/R96H also crystallized isomorphously to the WT protein in 2*M* K/Na phosphate buffers.[20,23] Before use, the crystals were equilibrated and stored in a standard mother liquor solution of 2.3*M* K/Na phosphate pH 6.7, 0.25*M* NaCl, and 1 μL/mL HEDS. R96V crystallized in space group $P2_12_12_1$ from 200 m*M* Na/K phosphate pH 6.7, 35% MPD, 50 m*M* BME, and 50 m*M* HEDS. R96W crystallized in space group $P2_1$ from 100 m*M* sodium citrate pH 5.6, 20% PEG4000, 10% isopropanol, 50 m*M* BME, and 50 m*M* HEDS. The initial protein concentration in the hanging drop was 10 mg/mL in all cases except R96W where it was 3.5 mg/mL.

X-ray data collection

The crystals of mutants that were isomorphous with WT generally grew to large sizes (~0.5 mm in each direction), which were suitable for mounting in glass capillaries. X-ray data for these crystals were collected at room temperature using a San Diego multiwire system[24] or, in the case of R96G, with an R-AXIS IV system. The nonisomorphous crystals were generally quite small, which precluded data collection at room temperature. They were mounted in rayon cryoloops and cooled to 100 K in a nitrogen gas stream for data collection with synchrotron radiation and a MAR image plate at SSRL beamlines 7-1 and 9-1. Cryodata were processed using HKL2000[25] for R96V and XDS[26] for R96W.

Structure determinations

The structures of R96V and R96W were determined by molecular replacement using the program EPMR[27] for R96V and MolRep[28] for R96W. The remaining structures, isomorphous with WT, were determined by molecular substitution using the coordinates of WT[29] as the starting model.

The isomorphous crystal structures were refined using TNT[30,31] following the procedures described.[20] The nonisomorphous crystal structures were refined using REFMAC[32,33] in the case of R96V and REFMAC followed by PHENIX[34] in the case of R96W. Weak noncrystallographic symmetry restraints were retained during the refinement of R96V as the structure had four similar protein molecules in the asymmetric unit. The Xfit molecular graphics module of XtalView was used for model rebuilding.[35] Figures of molecular structures were rendered with MOLSCRIPT[36] except where otherwise indicated.

Protein data bank accession numbers

The coordinates for the crystal structures described here have been deposited in the Protein Data Bank. The codes are listed in Table II.

Acknowledgments

The authors thank Hong Xiao, Leslie Gay, and Andy Fields for help with making and crystallizing the mutant proteins, Joel Lindstrom for assistance with the thermal melts, Doug Juers for collecting X-ray data for R96V, Larry Weaver for helpful discussions, and the user support staff at SSRL for their assistance.

References

1. Grütter MG, Hawkes RB, Matthews BW (1979) Molecular basis of thermostability in the lysozyme from bacteriophage T4. Nature 277:667–669.
2. Hawkes R, Grütter MG, Schellman J (1984) Thermodynamic stability and point mutations of bacteriophage T4 lysozyme. J Mol Biol 175:195–212.
3. Kitamura S, Sturtevant JM (1989) A scanning calorimetric study of the thermal denaturation of the lysozyme of phage T4 and the Arg 96 → His mutant form thereof. Biochemistry 28:3788–3792.
4. Weaver LH, Gray TM, Grütter MG, Anderson DE, Wozniak JA, Dahlquist FW, Matthews BW (1989) High-resolution structure of the temperature-sensitive mutant of phage lysozyme. Arg96 → His Biochemistry 28:3793–3797.
5. Tidor B, Karplus M (1991) Simulation analysis of the stability mutant R96H of T4 lysozyme. Biochemistry 30: 3217–3228.
6. Mattos C, Cohen JD, Greene DF, Tidor B, Karplus M (2004) X-ray structural and simulation analysis of a protein mutant: the value of a combined approach. Proteins 55:733–742.
7. Mooers BHM, Tronrud DE, Matthews BW. Evaluation at atomic resolution of the role of strain in destabilizing the temperature-sensitive T4 lysozyme mutant Arg96 → His. Protein Sci 18:863–870.
8. Mooers BHM, Matthews BW (2004) Use of an ion binding site to bypass the 1000-atom limit to structure determination by direct methods. Acta Cryst D 60:1726–1737.
9. Creighton TE (1992) Proteins: structures and molecular properties, 2nd ed. New York: W.H. Freeman and Company, p 6.
10. Blaber M, Lindstrom JD, Gassner N, Xu J, Heinz DW, Matthews BW (1993) Energetic cost and structural consequences of burying a hydroxyl group within the core of a protein determined from Ala → Ser and Val → Thr substitutions in T4 lysozyme. Biochemistry 32:11363–11373.
11. Blaber M, Zhang X-J, Lindstrom JD, Pepiot SD, Baase WA, Matthews BW (1994) Determination of α-helix propensity within the context of a folded protein. Sites 44 and 131 in bacteriophage T4 lysozyme. J Mol Biol 235:600–624.
12. Pace CN, Alston RW, Shaw KL (2000) Charge-charge interactions influence the denatured state ensemble and contribute to protein stability. Protein Sci 9:1395–1398.
13. Takano K, Scholtz JM, Sacchettini JC, Pace CN (2003) The contribution of polar group burial to protein stability is strongly context-dependent. J Biol Chem 278:31790–31795.
14. Doig AJ, Sternberg MJE (1995) Side-chain conformational entropy in protein folding. Protein Sci 4:2247–2251.
15. Kunkel TA (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc Natl Acad Sci USA 82:488–492.
16. Landt O, Grunert H-P, Hahn U (1990) A general method for rapid site-directed mutagenesis using the polymerase chain reaction. Gene 96:125–128.
17. Alber T, Dao-pin S, Nye JA, Muchmore DC, Matthews BW (1987) Temperature-sensitive mutations of bacteriophage T4 lysozyme occur at sites with low mobility and low solvent accessibility in the folded protein. Biochemistry 26:3754–3758.
18. Muchmore DC, McIntosh LP, Russell CB, Anderson DE, Dahlquist FW (1989) Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. Methods Enzymol 177:44–73.
19. Poteete AR, Sun D-P, Nicholson H, Matthews BW (1991) Second-site revertants of an inactive T4 lysozyme mutant restore activity restructuring the active site cleft. Biochemistry 30:1425–1432.
20. Eriksson AE, Baase WA, Matthews BW (1993) Similar hydrophobic replacements of Leu99 and Phe153 within the core of T4 lysozyme have different structural and thermodynamic consequences. J Mol Biol 229:747–769.
21. Elwell M, Schellman JA (1975) Phage T4 lysozyme. Physical properties and reversible unfolding. Biochim Biophys Acta 386:309–323.
22. Zhang X-J, Baase WA, Shoichet BK, Wilson KP, Matthews BW (1995) Enhancement of protein stability by the combination of point mutations in T4 lysozyme is additive. Protein Eng 8:1017–1022.
23. Lipscomb LA, Gassner NC, Snow SD, Eldridge AM, Baase WA, Drew DL, Matthews BW (1998) Context-dependent protein stabilization by methionine-to-leucine substitution shown in T4 lysozyme. Protein Sci 7:765–773.
24. Hamlin R (1985) Multiwire area X-ray diffractometers. Methods Enzymol 114:416–452.
25. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 276:307–326.
26. Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J Appl Cryst 26:795–800.
27. Kissinger CR, Gehlhaar DK, Fogel DB (1999) Rapid automated molecular replacement by evolutionary search. Acta Cryst D 55:484–491.
28. Vagin A, Teplyakov A (1997) MOLREP: an automated program for molecular replacement. J Appl Cryst 30: 1022–1025.
29. Bell JA, Wilson KP, Zhang X-J, Faber HR, Nicholson H, Matthews BW (1991) Comparison of the crystal structure of bacteriophage T4 lysozyme at low, medium, and high ionic strengths. Proteins 10:10–21.
30. Tronrud DE, Ten Eyck LF, Matthews BW (1987) An efficient general-purpose least-squares refinement program for macromolecular structures. Acta Cryst A 43:489–501.
31. Tronrud DE (1997) TNT refinement package. Methods Enzymol 277:306–319.
32. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Cryst D 53:240–255.
33. Murshudov GN, Vagin AA, Lebedev A, Wilson KS, Dodson EJ (1999) Efficient anisotropic refinement of macromolecular structures using FFT. Acta Cryst D 55: 247–255.
34. Adams PD, Grosse-Kunstleve RW, Hung L-W, Iorger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC (2002) PHENIX: building new software for automated crystallographic structure determination. Acta Cryst D 58:1948–1954.
35. McRee DE (1999) XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. J Struct Biol 125:156–165.
36. Kraulis PJ (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. J Appl Cryst 24:946–950.