A test of the "jigsaw puzzle" model for protein folding by multiple methionine substitutions within the core of T4 lysozyme

NADINE C. GASSNER, WALTER A. BAASE, AND BRIAN W. MATTHEWS

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

ABSTRACT To test whether the structure of a protein is determined in a manner akin to the assembly of a jigsaw puzzle, up to 10 adjacent residues within the core of T4 lysozyme were replaced by methionine. Such variants are active and fold cooperatively with progressively reduced stability. The structure of a seven-methionine variant has been shown, crystallographically, to be similar to wild type and to maintain a well ordered core. The interaction between the core residues is, therefore, not strictly comparable with the precise spatial complementarity of the pieces of a jigsaw puzzle. Rather, a certain amount of give and take in forming the core structure is permitted. A simplified hydrophobic core sequence, imposed without genetic selection or computer-based design, is sufficient to retain native properties in a globular protein.

The cores of globular proteins consist of buried, primarily hydrophobic, amino acids. Tight packing of the amino acid side chains (1) has led to the idea that the size and shape of the nonpolar amino acids within the core may constrain or define the overall protein fold (2, 3). This "jigsaw puzzle" model of protein folding was originally introduced by Crick (4) as a "knobs into holes" description of α-helix packing and has been elaborated by Chothia *et al.* (5), and by Alber and co-workers (6). Here the jigsaw puzzle model refers to shape complementarity (3), not to the pathway of folding (7). The model is supported by the observation that changes in the sizes and shapes of residues within the cores of proteins are usually destabilizing (8–10). Also in support of the model, the structures of α-helical coiled coils appear to be determined by the shapes of the buried side chains (6). In contrast with this view, it has been shown that alternative core sequences that lead to viable proteins could be selected by random mutagenesis for both λ-repressor (11) and T4 lysozyme (12), among others (13, 14). It is possible, however, that a limited number of combinations of amino acids are viable and that they are the ones identified by the mutagenic selection. Here we explore an approach in which there is no selection other than the sites of substitution.

MATERIALS AND METHODS

We chose methionine as a generic core-replacement residue for a combination of reasons. First, a methionine side chain occupies roughly the same volume as the frequently observed core residues leucine, isoleucine, and phenylalanine. It is, however, more flexible and can more readily adapt to occupy whatever space might be available. In this sense methionine contrasts with the rigid, predetermined shape of a piece of a jigsaw puzzle. Methionine also occurs relatively infrequently in known proteins (15). Thus multiple methionine substitutions would be expected to substantially change the composition of the core. Finally, we wondered if the intro-

Table 1. Activity and stability of methionine-substituted lysozymes

Mutant	Activity (%)	ΔT_m (°C)	$\Delta H(T_m)$ (kcal/mol)	ΔH(ref) (kcal/mol)	$\Delta\Delta G$ (kcal/mol)
WT*	100		130	115	—
I78M	70	−3.7	117	111	−1.5
L84M	104	−4.9	110	108	−1.9
L91M	96	−2.0	125	115	−0.8
L99M†	90	−1.3	134	122	−0.4
I100M	105	−4.5	125	121	−1.6
V103M	70	−3.1	117	109	−1.2
L118M	98	−1.8	130	119	−0.7
L121M	87	−2.1	129	119	−0.8
L133M	106	−1.0	128	115	−0.4
F153M†	87	−1.6	128	116	−0.6
7-Met‡	43	−14.5	96	117	−5.0
10-Met‡	≈20	−25	42	88	−7.3

The activity was determined as in ref. 28 except at 20°C in 66 mM potassium phosphate, pH 6.8. For the 10-Met mutant, a loss in activity was seen with time. Activities were also determined using lysis plates (29) and found to be in agreement with the values given in the table.

Stability measurements (18) were made in 0.1 M sodium chloride/1.4 mM acetic acid/8.6 mM sodium acetate, pH 5.42. The melting temperature, T_m, of WT* lysozyme was 65.3°C. ΔT_m is the change in the T_m of the mutant relative to wild type. For the single mutants, the uncertainty in ΔT_m os ±0.2°C; for the multiple mutants it is ±0.5°C. $\Delta H(T_m)$ is the enthalpy of unfolding measured at T_m. The uncertainty is ±5 kcal/mol. ΔH(ref) is the enthalpy of unfolding calculated at the reference temperature of 59°C using a constant ΔC_p of 2.5 kcal/mol·deg. $\Delta\Delta G$ is the free energy of unfolding of the mutant relative to wild type. ΔG values were computed at 59°C using a constant ΔC_p of 2.5 kcal/mol·deg. The uncertainty in $\Delta\Delta G$ is ±0.1 kcal/mol for the single mutants and ±0.4 kcal/mol for the 7-Met replacement. Because of the low value of ΔH of the 10-Met mutant, $\Delta\Delta G$ was determined at the T_m of the mutant with an estimated uncertainty of about 1 kcal/mol.
†Mutants L99M and F153M were described previously (18).
‡The 7-Methionine mutant includes the substitutions L84M/L91M/L99M/L118M/L121M/L133M/F153M. The 10-Methionine variant includes the additional substitutions I78M/I100M/V103M. The molecular masses of these proteins determined by mass spectrometry agreed with the theoretical values, suggesting little if any oxidation of the introduced methionines (data not shown).

duction of multiple, flexible, amino acids within the core of a protein might lead to the onset of molten globule characteristics (16).

All sites of substitution are buried within the carboxyl-terminal domain of T4 lysozyme, and the side chain of each residue contacts at least one other side chain of the set. The 10 single-site mutants as well as various multiple-methionine mutants were constructed (17) in cysteine-free pseudo wild-type lysozyme, hereafter identified as WT* or wild type (18). Activity and stability measurements for the 10 single mutants, together with the 7-Met and 10-Met mutants, are listed in Table 1.

RESULTS AND DISCUSSION

All variants possessed native-like properties. The thermal denaturations of the one- and seven-methionine variants are
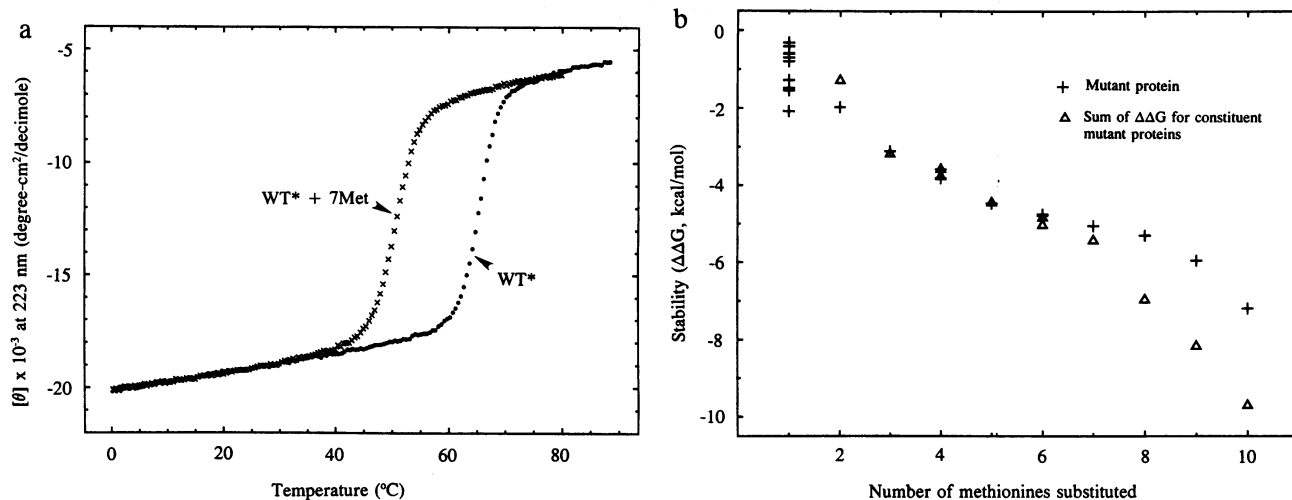
FIG. 1. (*a*) Comparison of the thermal unfolding transition of the seven-methionine mutant with that of wild-type lysozyme. (*b*) Stabilities for mutant lysozymes plotted as a function of the number of introduced methionines. The crosses show the stabilities, relative to wild-type, of the single mutants and the seven- and 10-methionine mutants listed in Table 1. Additional crosses show the stabilities of other multiple-methionine variants that have been constructed but are not described explicitly in Table 1 (unpublished results). The triangles show the sums of the stabilities of the single mutants that are combined together to obtain a given multiple mutant. Two different combinations of substitutions were used to obtain 4-Met and 6-Met lysozymes. The stabilities of both constructs are included in the figure.

essentially as cooperative as wild-type, with comparable enthalpies of unfolding (Fig. 1*a*; Table 1). The 10-methionine variant unfolds cooperatively, although the enthalpy is reduced (Table 1). Activity was equal to at least 20% that of wild type, suggesting that active site structure is retained (Table 1). The aromatic circular dichroism spectra of the largest construct, a 10-methionine core variant, was comparable in shape and magnitude to the spectra of wild type. The one-dimensional NMR spectrum of the same variant had significant chemical shift dispersion (data not shown). Taken together these results strongly suggest that the 10-Met variant has a well-defined three-dimensional structure and is not a molten globule (16).

Crystals of the 7-Met variant (Table 1) were obtained and found to be isomorphous with wild-type lysozyme (18). X-ray data to a 1.9-Å resolution, 92% complete, were measured at room temperature (19, 20). A difference density map (Fig. 2*a*) showed seven well defined positive peaks corresponding to the introduction of the electron-dense sulfur at each of the sites of substitution as well as negative density where atoms were deleted.

The variant structure (Fig. 3) was refined (21, 22) to a crystallographic residual of 15.2% with bond lengths and angles within 0.18 Å and 3.0° of ideal values and was found to be very similar to wild type. The coordinates have been deposited in the Brookhaven Data Bank. The root-mean-square discrepancy of the main chain atoms within the car-boxyl-terminal domain is 0.20 Å. In the six cases in which a methionine replaced a leucine, the χ_1 and χ_2 values in the mutant were similar to those in wild type. Thus, each of the substituted methionines essentially traced the path of the residue that it replaced. For the Phe-153 → Met substitution, however, χ_1 changed by 92° to avoid a steric clash.

The crystallographic thermal factors of the side chains of the seven methionines are, on average, marginally less than the thermal factors of the amino acids that they replace (24.0 Å^2 versus 25.7 Å^2). The distal methyl groups are also well-ordered (average thermal factor 23.8 Å^2). As shown in Fig. 2*b*, the mobility of the surrounding side chains in the mutant structure is also very similar to wild type. Therefore there is no suggestion that the substitution of seven methionines leads to disorder of the hydrophobic core.

As more and more methionines are introduced into the protein, the overall stability decreases (Fig. 1; Table 1). When six or more methionines are substituted, the loss of stability is somewhat less than the sum of the constituent single replacements (Fig. 1*b*) with the discrepancy increasing to a maximum of 2.5 kcal/mol for the 10-methionine construct. This indicates that there is some relaxation in the polymethionine protein that either introduces new, favorable, interactions or relieves some of the strain associated with the single substitutions. The loss in protein stability is understandable. For each methionine replacement there is a reduction in the solvent transfer free energy (about 0.6 kcal/mol for Leu to Met) (23). Also the side chain of methionine has more degrees of freedom than do other hydrophobic core amino acids. Each methionine-to-leucine replacement at a restricted, internal, site is predicted to have an entropy cost of about 0.8 kcal/mol (24, 25). Taken together, these two factors are expected to reduce the stability of the seven-methionine mutant by about 10 kcal/mol relative to wild type. Some of the replacements may also decrease stability because of introduced strain. The actual loss in stability for the 7-Met mutant is only 5.0 kcal/mol, suggesting that the above estimate of ≈10 kcal/mol is too high.

The finding that 10 core residues in T4 lysozyme can be replaced with methionine supports the overall importance of the hydrophobic effect in protein folding. At the same time, the results show that the interaction between the core residues is not strictly comparable with the precise spatial complementarity of the pieces of a jigsaw puzzle. Rather, a certain amount of give and take in forming the core structure is permitted. This is in contrast to α-helical coiled-coils where changes in the shape of hydrophobic residues can lead to different packing arrangements (6).

The observation that methionines substituted at various internal sites remain well ordered suggests that seleno-methionine, or telluro methionine, introduced in this fashion should be suitable for MAD phasing (26). The apparent lack of oxidation of core sites, presumably due to reduced oxygen accessibility, may aid in the use of such oxygen-sensitive analogs.

Previous studies have shown that multiple alanine replacements can be made on the surface of T4 lysozyme, at least within α-helices, with little change in structure or stability (27). The present study shows that multiple replacements with a
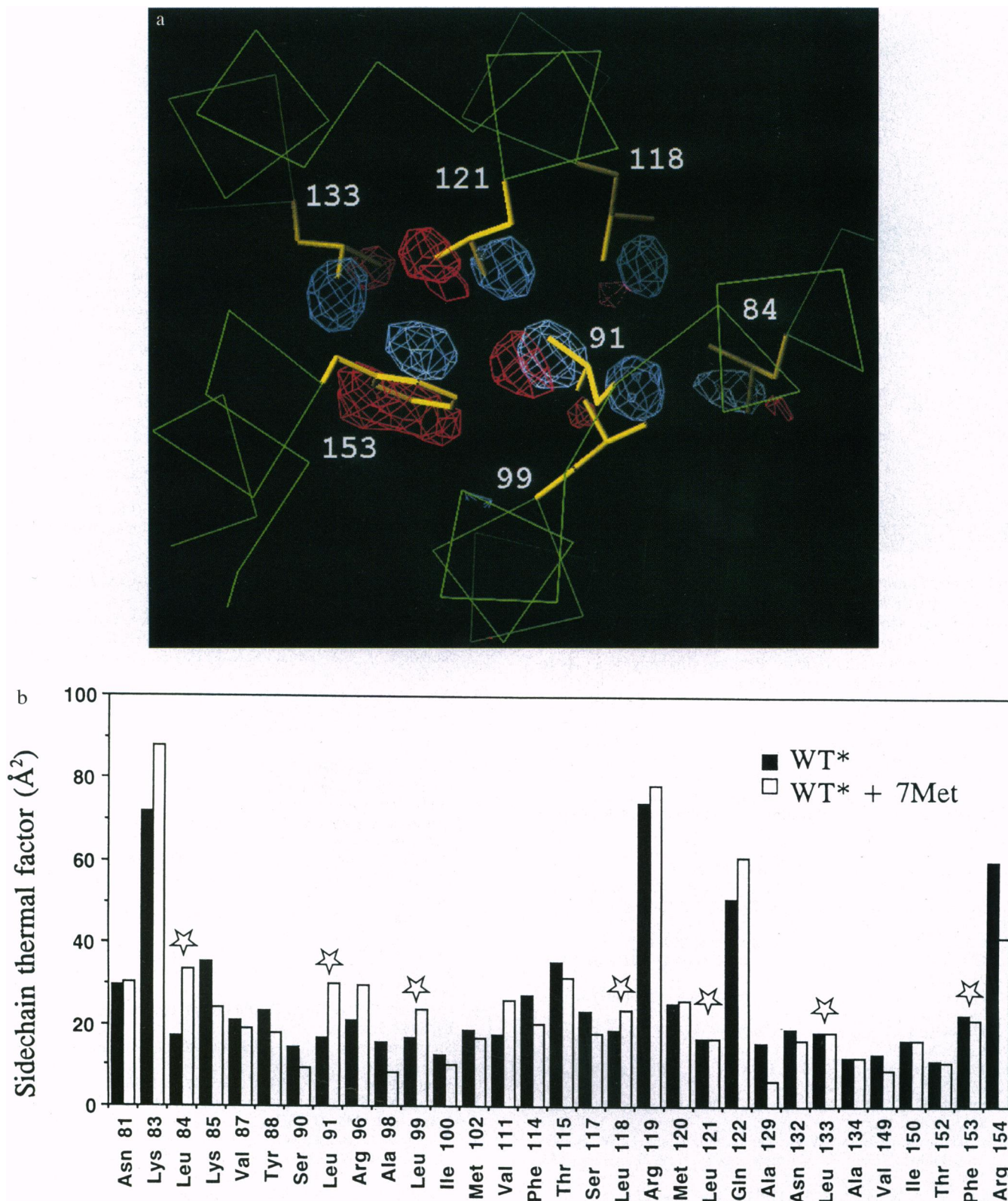
FIG. 2. (*a*) Map showing the difference in electron density between the seven-methionine mutant and wild-type T4 lysozyme. Coefficients ($F_{mut} - F_{WT^*}$) where F_{mut} and F_{WT^*} correspond to the observed structure amplitudes of the mutant and wild-type structures. Phases from the refined structure of WT* lysozyme (18). Resolution is 1.9 Å. Blue contours representing positive density are drawn at $+3\sigma$ where σ is the root-mean-square density throughout the unit cell. Red contours (negative density) are drawn at -3σ. Superimposed is the structure of the carboxyl-terminal domain of WT* lysozyme with the backbone shown in green and the substituted side chains in yellow. Crystals were obtained using ≈2M phosphate solutions, ≈pH 6.7 (18). (*b*) Comparison of the thermal factors of side chains within the core of the 7-methionine mutant (open bars) with those of wild-type lysozyme (solid bars). The figure includes the seven residues that were changed to methionine (marked with stars) as well as all residues within 4 Å of the substituted amino acids. The amino acids are identified as in the WT* structure.
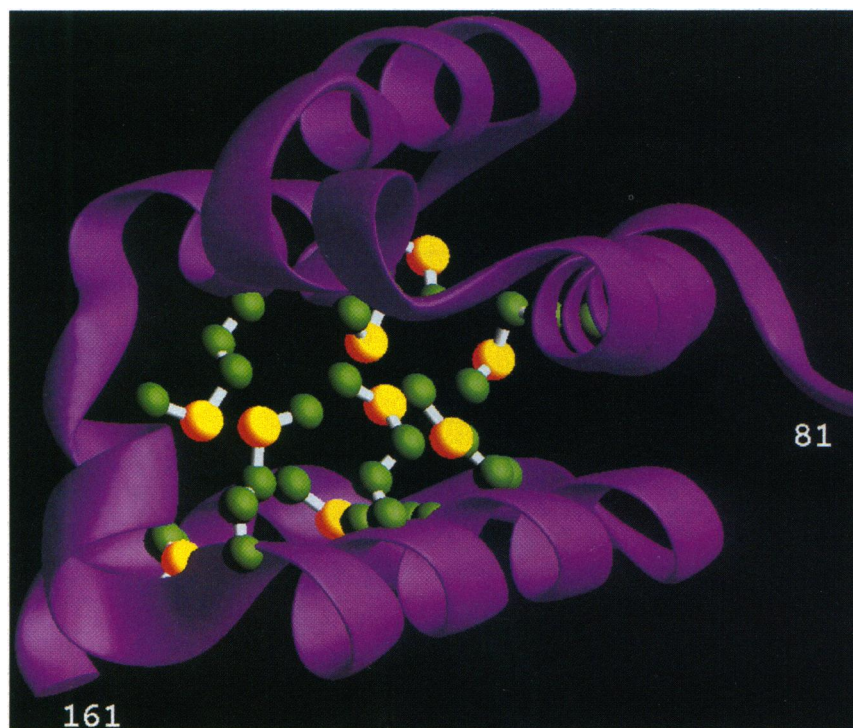
FIG. 3. Residues 81–161 of the carboxyl-terminal domain of T4 lysozyme. The figure shows the distribution of methionines within the hydrophobic core of the molecule. It illustrates the structure of the mutant in which seven methionines have been introduced genetically and includes two additional methionines that are present in the native protein. The methionine side chains are shown in green with the sulfur atoms in yellow. The carboxyl-terminal domain of T4 lysozyme contains a single, completely buried, methionine (Met-102), and two more (Met-106 and -120) that are about 80% buried. In addition, six leucines (Leu-84, -91, -99, -118, -121, and -133), two isoleucines (Ile-78 and -100), one phenylalanine (Phe-153), and one valine (Val-103) were chosen for substitution with methionine.

single type of amino acid are possible in the core as well, albeit with a progressive loss of stability. It suggests that it may be possible to replace the overall amino acid sequence of a protein with a much simpler sequence based on a subset of the 20 naturally occurring amino acids. Perhaps this may be a way to simplify the protein folding problem.

1. Richards, F. M. (1974) *J. Mol. Biol.* **82,** 1–14.
2. Ponder, J. W. & Richards, F. M. (1987) *J. Mol. Biol.* **193,** 775–791.
3. Behe, M. J., Lattman, E. E. & Rose, G. D. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 4195–4199.
4. Crick, F. H. C. (1953) *Acta Crystallogr.* **6,** 689–697.
5. Chothia, C., Levitt, M. & Richardson, D. (1981) *J. Mol. Biol.* **145,** 215–250.
6. Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. (1993) *Science* **262,** 1401–1407.
7. Harrison, S. C. & Durbin, R. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4028–4030.
8. Kellis, J. T., Nyberg, K., Sali, D. & Fersht, A. R. (1988) *Nature (London)* **333,** 784–786.
9. Shortle, D., Stites, W. E. & Meeker, A. K. (1990) *Biochemistry* **29,** 8033–8041.
10. Eriksson, A. E., Baase, W. A., Zhang, X-J., Heinz, D. W., Blaber, M., Baldwin, E. P. & Matthews, B. W. (1992) *Science* **255,** 178–183.
11. Lim, W. A. & Sauer, R. T. (1991) *J. Mol. Biol.* **219,** 359–376.
12. Baldwin, E. P., Hajiseyedjavadi, O., Baase, W. A. & Matthews, B. W. (1993) *Science* **262,** 1715–1718.
13. Pielak, G. J., Auld, D. S., Beasley, J. R., Betz, S. F., Cohen, D. S., Doyle, D. F., Finger, S. A., Fredericks, Z. L., Hilgen-Willis, S., Saunders, A. J. & Trojak, S. K. (1995) *Biochemistry* **34,** 3268–3276.
14. Axe, D. D., Foster, N. W. & Fersht, A. R. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 5590–5594.
15. Klapper, M. H. (1977) *Biochem. Biophys. Res. Commun.* **78,** 1018–1024.
16. Kuwajima, K. (1989) *Proteins* **6,** 87–103.
17. Poteete, A. R., Dao-pin, S., Nicholson, H. & Matthews, B. W. (1991) *Biochemistry* **30,** 1425–1432.
18. Eriksson, A. E., Baase, W. A. & Matthews, B. W. (1993) *J. Mol. Biol.* **229,** 747–769.
19. Hamlin, R. (1985) *Methods Enzymol.* **114,** 416–452.
20. Zhang, X-J. & Matthews B. W. (1993) *J. Appl. Cryst.* **26,** 457–462.
21. Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987) *Acta Crystallogr. A* **43,** 489–503.
22. Tronrud, D. E. (1996) *J. Appl. Cryst.* **29,** 100–104.
23. Fauchére, J-L. & Pliška, V. (1983) *Eur. J. Med. Chem.* **18,** 369–375.
24. Creamer, T. P. & Rose, G. D. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 5937–5941.
25. Sternberg, M. J. E. & Chickos, J. S. (1994) *Protein Eng.* **7,** 149–155.
26. Hendrickson, W. A. (1991) *Science* **254,** 51–58.
27. Blaber, M., Baase, W. A., Gassner, N. & Matthews, B. W. (1995) *J. Mol. Biol.* **246,** 317–330.
28. Zhang, X-J., Baase, W. A., Shoichet, B. K., Wilson, K. P. & Matthews, B. W. (1995) *Protein Eng.* **8,** 1017–1022.
29. Streisinger, G., Mukai, F., Dreyer, W. J., Miller, B. & Horiuchi, S. (1961) *Cold Spring Harbor Symp. Quant. Biol.* **26,** 25–30.