Title

Dissection of protein structure and folding by directed mutagenesis

Journal

Faraday Discussions, 93(0)

ISSN

1359-6640

Authors

Baase, WA
Eriksson, AE
Zhang, XJ
et al.

Publication Date

1992

DOI

10.1039/fd9929300173

Peer reviewed

Faraday Discuss., 1992, **93**, 173–181

Dissection of Protein Structure and Folding by Directed Mutagenesis

Walter A. Baase, A. Elisabeth Eriksson,†Xue-Jun Zhang, Dirk W. Heinz, Uwe Sauer,‡ Michael Blaber, Enoch P. Baldwin, Joan A. Wozniak and Brian W. Matthews*

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

The lysozyme from bacteriophage T4 is being used as a model system to determine the roles of individual amino acids in the folding and stability of a typical globular protein. One general finding is that the protein is very adaptable, being able to accommodate many potentially destabilizing replacements.

In order to determine the importance of 'α-helix propensity' in protein stability, different replacements have been made within α-helical segments of T4 lysozyme. Several such substitutions of the form Xaa → Ala increase the stability of the protein, supporting the idea that alanine is a strongly helix-favouring amino acid. It is possible to engineer a protein that has up to ten alanines in succession, yet still folds and has normal activity. This illustrates the redundancy that is present in the amino acid sequence.

A number of 'cavity-creating' mutants of the form Leu → Ala have been constructed to understand better the nature of hydrophobic stabilization. The structural consequences of these mutations differ from site to site. In some cases the protein structure hardly changes at all; in other cases removal of the wild-type side-chain allows surrounding atoms to move in and occupy the vacated space, although a cavity always remains. The destabilization of the protein associated with these cavity-creating mutations also varies from case to case. The results suggest how to reconcile recent conflicting reports concerning the strength of the hydrophobic effect in proteins.

We have been using the lysozyme from bacteriophage T4 as a model system to determine the roles of individual amino acids in the folding of the protein. Such studies also provide quantitative information on the contributions that different types of interactions (hydrogen bonds, hydrophobic interactions, salt bridges *etc.*) make to the protein stability.

Tolerance to Change

One of the encouraging developments has been the relative freedom with which amino acid replacements can be introduced in a protein of interest.

To investigate the ability of T4 lysozyme to accommodate potentially destabilizing amino acid substitutions, and also to investigate the steric requirements for catalysis,

† Present address: Department of Molecular Biology, Biomedical Center, Box 590, S-751 24 Uppsala, Sweden.

‡ Present address: European Molecular Biology Laboratory, Postfach 10.2209, Meyerhofstrasse 1, 6900 Heidelberg, Germany.

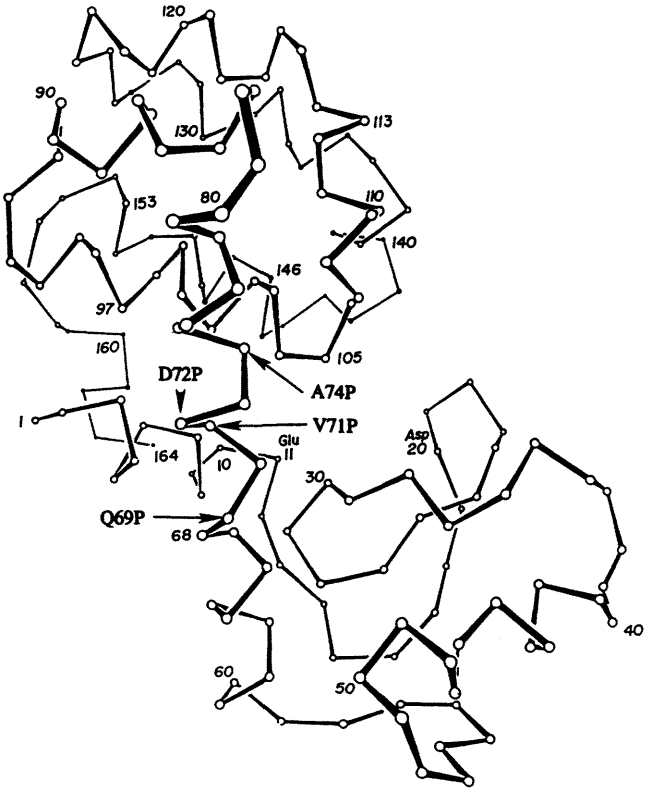


Fig. 1 Backbone of T4 lysozyme showing the locations of four proline substitutions. Mutations are identified by the one-letter code: *e.g.* V71P is Val-71 → Pro. Mutations Q69P, D72P and A74P give active, folded protein. Reproduced with permission from U. H. Sauer, S. Dao-pin and B. W. Matthews, *J. Biol. Chem.*, 1992, 2393.

proline was substituted at different sites within the long α-helix that connects the amino-terminal and carboxy-terminal domains[1] (Fig. 1). Of the four substitutions attempted, three yielded folded, functional proteins. The catalytic activities of these three mutant proteins (Q69P, D72P and A74P) are 60–90% that of wild-type. Their melting temperatures are 7–12 °C less than that of wild-type at pH 6.5.

Mutant D72P formed crystals isomorphous with wild-type allowing the structure to be determined at high resolution. In the crystal structure of wild-type lysozyme the interdomain α-helix has an overall bend angle of 8.5°. In the mutant structure the introduction of the proline causes this bend angle to increase to 14° and also causes a corresponding rotation of 5.5° of the carboxy-terminal domain relative to the amino-terminal one. Except for the immediate location of the proline substitution there is very little change in the geometry of the interdomain α-helix. The results support the view that protein structures are adaptable and can compensate for potentially destabilizing amino acid substitutions. The results also suggest that the precise shape of the active site cleft of T4 lysozyme is not critical for catalysis.[1]

Toward a Simplification of the Protein-folding Problem

Although it has long been recognized that the amino acid sequence of a protein determines its three-dimensional structure, recent work has made it clear that certain amino acids are more important than others in the folding process. At some positions,

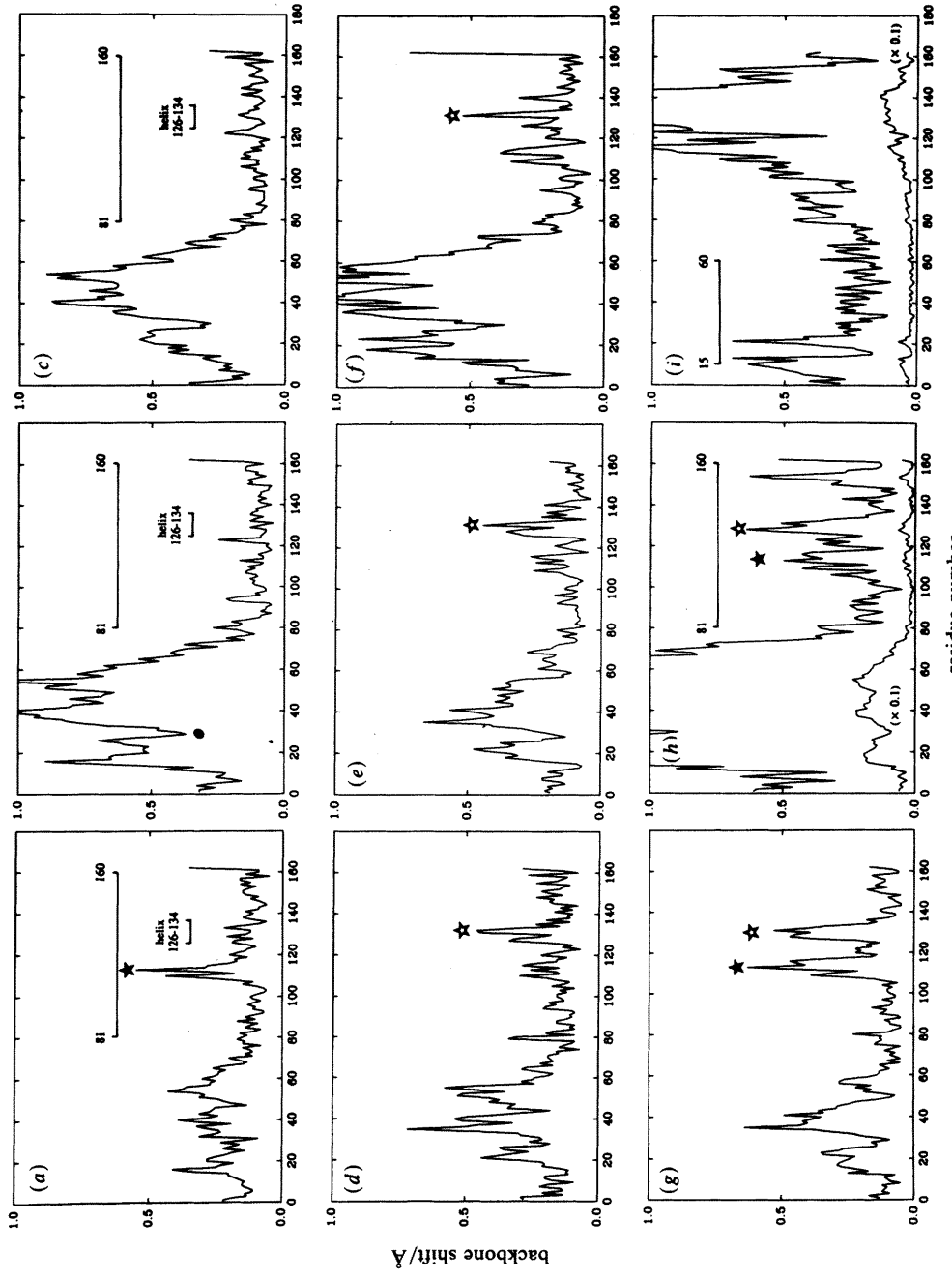
typically the solvent-exposed or mobile sites in the folded protein, amino acids can be interchanged almost at will with little apparent effect on folding or stability.[2,3] These amino acids seem to be unimportant in protein folding. On the other hand, sites at which amino acid replacements substantially destabilize folded proteins appear to be mostly restricted to the buried or rigid parts,[4] suggesting that the amino acids at these positions are important in determining the folded conformation.

The observation that many amino acids in a protein sequence can be freely replaced with little if any effect on protein stability leads one to consider the protein folding problem in a new light. What fraction of the amino acids in a given polypeptide sequence is, in fact, essential for the successful folding of the protein? Is it 75% or 50%? Could it be as low as, say, 20%? To restate the problem in a different way, could one take a known protein and replace a large number of 'non-essential' amino acids with alanine, yet still have a folded functional protein? In such a 'polyalanine protein' all non-essential side chains would have been lost; this allows one to focus on those parts of the amino acid sequence that are critical for the folding process.

As a modest first step in this direction, and also to further investigate the role of alanine as a helix-stabilizing residue, five amino acids within α-helix 126–134 (Fig. 1) were substituted by alanine, either singly or in selected combinations. Together with three alanines already present in the wild-type structure this provided a set of mutant proteins with up to eight alanines in sequence. All the variants behave normally, suggesting that the majority of residues in the α-helix are non-essential for the folding of T4 lysozyme. Of the five individual alanine substitutions, four result in a slight increase in protein stability and one, the replacement of a buried leucine with alanine, substantially decreases stability.[5,6] The results support the idea that alanine is a residue of high helix propensity. The change in protein stability observed for each of the multiple mutants is approximately equal to the sum of the energies associated with each of the constituent substitutions.

All of the variants could be crystallized isomorphously with wild-type lysozyme and, with one trivial exception, their structures have been determined at high resolution. Substitution of the largely solvent-exposed residues Asp-127, Glu-128 and Val-131 with alanine causes essentially no change in structure except at the immediate site of replacement. Substitutions of the partially buried Asn-132 and the buried Leu-133 with alanine are associated with modest ($\leq$0.4 Å) structural adjustments. The structural changes seen in the multiple mutants are essentially a combination of those seen in the constituent single replacements (Fig. 2). The different replacements therefore act essentially independently not only so far as changes in energy are concerned, but also in their effect on structure. The destabilizing replacement Leu-133 $\rightarrow$ Ala makes α-helix 125–134 somewhat less regular. Incorporation of additional alanine replacements tends to make the helix more uniform. For the penta-alanine variant a distinct change occurs in a crystal-packing contact, and the 'hinge-bending angle' between the amino- and carboxy-terminal domains changes by 3.6°. This tends to confirm that such hinge-bending in T4 lysozyme is a low-energy conformational change.

Single and multiple Xaa $\rightarrow$ Ala substitutions were also constructed[7] in the α-helix in bacteriophage T4 lysozyme comprising residues 39–50 (Fig. 1). The individual substitutions of the solvent-exposed residues Asn-40, Ser-44 and Glu-45 with alanine slightly increase the thermostability of the protein (Fig. 3). Replacement of Asp-47 and Lys-48 with alanine slightly decrease stability. The replacements of the partially buried side chain of Lys-43 and the completely buried side chain of Leu-46 result in a significant decrease in the thermostability. The crystal structures of the latter two variants reveal that the mutations result in the creation of hydrophobic cavities within or on the surface of the protein. In the case of the charged amino acids Glu-45 and Lys-48, the changes in melting temperature indicate that the putative salt bridge between these two residues contributes essentially nothing to the stability of the protein.



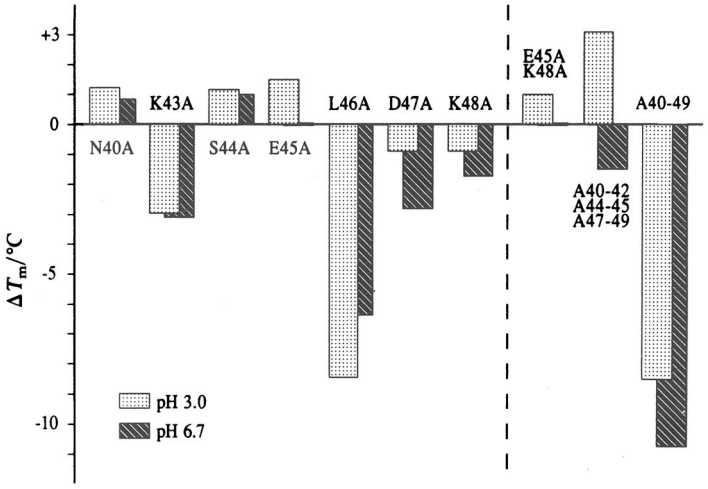


Fig. 3 Histogram showing the differences in melting temperatures (ΔT_m) between pseudo wild-type ($\Delta T_m = 0$) and mutant lysozymes at pH 3.0 (stippled bars) and pH 6.7 (striped bars). A positive value of ΔT_m indicates that the mutant is more stable than pseudo wild-type. Residues 41, 42 and 49 are alanine in wild-type T4 lysozyme

The variant with alanines substituted at ten consecutive positions (A40–49) folds normally, has activity essentially the same as wild-type, and has a decrease in thermostability (−10.7 °C in melting temperature; −2.5 kcal mol^{-1} in free energy of unfolding). The ability of this protein to fold clearly demonstrates the redundancy that is present in the amino acid sequence. The crystal structure of this polyalanine mutant displays no significant change in the main-chain atoms of the helix when compared with the wild-type structure. The melting temperature of the lysozyme in which the 'crucial' amino acids Lys-43 and Leu-46 are retained, and positions 40, 44, 45, 47 and 48 are substituted with alanine (*i.e.* A40–42/A44–45/A47–49), is increased by 3.1 °C relative to wild-type at pH 3.0, but reduced by 1.6 °C at pH 6.7.

The study provides further evidence that the replacement of solvent-exposed residues within α-helices with alanines may be a general way to increase protein stability. The general approach may permit a simplification of the protein-folding problem by retaining only amino acids proven to be essential for folding and replacing the remainder with alanine.

Fig. 2 'Shift plots' showing the displacement of the backbone atoms of each mutant relative to wild-type lysozyme. Each mutant structure was superimposed on wild-type so as to minimize the root-mean-square discrepancy between the respective backbone atoms in the carboxy-terminal domains (residues 81–160). For each amino acid the value plotted is the average (*i.e.* rms) discrepancy between the corresponding backbone atom (C^{α}, C^{β}, O and N) in the mutant and wild-type structure. The single mutation Leu-133 → Ala causes backbone shifts in the vicinity of residues 109–114 and similar shifts are seen in all mutants that include this replacement (solid stars). Similarly, the mutation Asn-132 → Ala causes changes in the vicinity of residues 126–134 which are seen in all variants that include N132A (open stars). (*a*) L133A, (*b*) D127A/E128A, (*c*) E128A/V131A, (*d*) V131A/N132A, (*e*) E128A/V131A/N132A, (*f*) D127A/E128A/V131A/N132A, (*g*) E128A/V131A/N132A/L133A, (*h*) D127A/E128A/V131A/N132A/L133A, (*i*) D127A/E128A/V131A/N132A/L133A, superposition based on the amino-terminal domains (residues 15–60). The large shift in Thr-21 is associated with a change in crystal contact between the backbone of Thr-21 and the side-chain of Trp-126. Reproduced with permission from X-J. Zhang, W. A. Baase and B. W. Matthews, Multiple alanine replacements within α-helix 126–134 of T4 lysozyme have independent, additive effects on both structure and stability, *Protein Sci.*, 1992, in the press

Hydrophobic Stabilization in Proteins

It is generally agreed that the hydrophobic effect is the major factor in stabilizing the folded structures of glubular proteins. Until recently, it has also been generally agreed that the strength of the hydrophobic effect, *i.e.* the energy of stabilization provided by the transfer of hydrocarbon surfaces from solvent to the interior of a protein, is *ca.* 25–30 cal mol^{-1} $Å^{-2}$.[8,9]

Some recent studies using site-directed mutagenesis and protein denaturation have, however, suggested that the strength of the hydrophobic effect might be much higher. In a typical experiment a hydrophobic residue within the core of a protein is substituted by a smaller hydrophobic residue and the resultant change in the stability of the folded *vs.* the unfolded (or denatured) form of the protein is taken as a measure of the difference between the hydrophobic stabilization provided by the two amino acids. Such experiments carried out with different proteins[10–14] or at different sites within the same protein[15] have, however, given variable results. Shortle, for example, replaced each of the leucines in staphylococcal nuclease with alanine and found that the decrease in free energy of protein folding ranged from 1.6 to 5.8 kcal mol^{-1}.[15] The latter value corresponds to stabilization of *ca.* 80 cal mol^{-1} $Å^{-2}$, a value about four times that estimated from solvent transfer experiments.[8,9,16–18] The reason for this discrepancy has not been resolved and remains the subject of debate.

A principal difficulty in addressing this problem has been the lack of relevant structual data. How does a protein structure respond when a bulky hydrophobic residue such as leucine is replaced by a smaller residue such as alanine? Does the protein structure remain essentially unchanged[12,13] or is there structural rearrangement to avoid the creation of a cavity?[19,20] If cavities are created do they contain solvent?[12,13]

To address these questions six 'cavity-creating' mutants Leu-46 → Ala (L46A), L99A, L118A, L121A, L133A and F153A have been constructed within the hydrophobic core of phage T4 lysozyme.[21] The substitutions decrease the stability of the protein at pH 3.0 by different amounts, ranging from 2.7 kcal mol^{-1} for L46A and L121A to 5.0 kcal mol^{-1} for L99A. The double mutant L99A/F153A was also constructed and decreases stability by 8.3 kcal mol^{-1}. All variants were crystallized and their structures determined at high resolution.

The structural consequences of the mutations differ from site to site. In the case of L99A, for example, the protein structure hardly changes at all. For F153A, however, both side-chain and backbone shifts up to 0.8–1.0 Å were observed. In every case removal of the wild-type side chain allowed some of the surrounding atoms to move toward the vacated space but a cavity always remained, ranging in volume from 24 $Å^3$ for L46A to 150 $Å^3$ for L99A. Solvent molecules were never observed in any of these cavities.

As shown in Fig. 4(*a*) and (*b*), there is an approximate relationship between the increase in cavity size associated with a given replacement and the reduction in protein stability. Five of the replacements in the figure are of the form Leu → Ala. We have also included the Phe-153 → Ala replacement because a phenylalanine side chain is approximately the same size as a leucine and is expected to have approximately the same hydrophobic strength.[17,18] The inclusion of the double mutant L99A/F153A in Fig. 4 provides an example where the overall cavity volume is large (207 $Å^3$; equivalent to a 5.9 Å cube) and the protein is quite unstable. We take this double mutant to be approximately representative of a protein in which two leucines have each been replaced with alanine. Subject to these and other assumptions,[21] Fig. 4(*a*) and (*b*) suggest that the decrease in protein stability associated with a Leu → Ala replacement consists of a constant-energy term of *ca.* 1.9 kcal mol^{-1} plus a second energy term that depends on the size of the cavity created by the substitution. The magnitude of the constant energy term corresponds with values of 1.7–1.9 kcal mol^{-1} for the difference in hydrophobicity

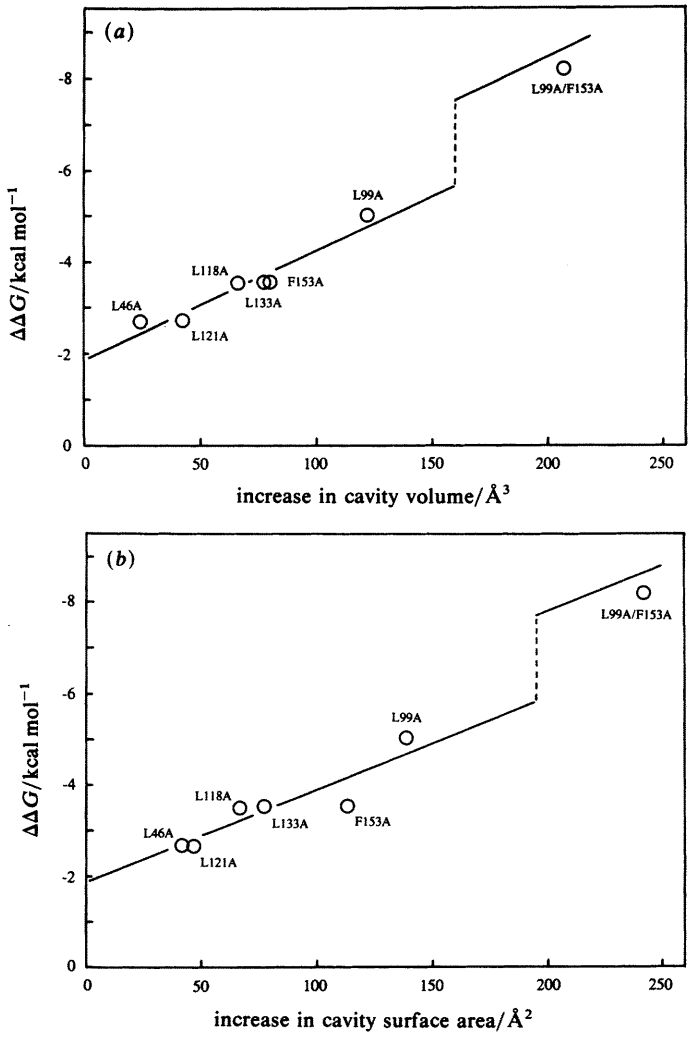


Fig. 4. (*a*) Change in the free energy of unfolding of mutant lysozymes relative to wild-type plotted as a function of the cavity volume created by the amino acid substitution(s). The straight line was fitted by least squares to the data for the six single mutants and has the equation

$$\Delta\Delta G = a + b\,\Delta V$$

where $a = -1.9$ kcal mol^{-1}, $b = -0.024$ kcal mol^{-1} Å^{-3} and ΔV is the increase in cavity volume. In the vicinity of the double mutant an additional -1.9 kcal mol^{-1} is added to $\Delta\Delta G$ to reflect the fact that in this case both a leucine and a phenylalanine (which is here regarded as equivalent to a second leucine) have been replaced with alanine. (*b*) Change in the free energy of unfolding of mutant lysozymes relative to wild-type plotted as a function of the increase in cavity surface area created by the amino acid substitution(s). The equation of the straight line that best fits the data for the single mutants is

$$\Delta\Delta G = c + d\,\Delta A$$

where $c = -1.9$ kcal mol^{-1}, $d = -0.020$ kcal mol^{-1} Å^{-2} and ΔA is the increase in cavity surface area. In the vicinity of L99A/F153A an additional -1.9 kcal mol^{-1} has been added to $\Delta\Delta G$ to reflect the fact that the double mutant is roughly equivalent to two leucine to alanine replacements. Reproduced with permission from A. E. Eriksson, W. A. Baase, X-J. Zhang, D. W. Heinz, M. Blaber, E. P. Baldwin and B. W. Matthews, *Science*, 1992, **255**, 178. Copyright 1992 by the AAAS

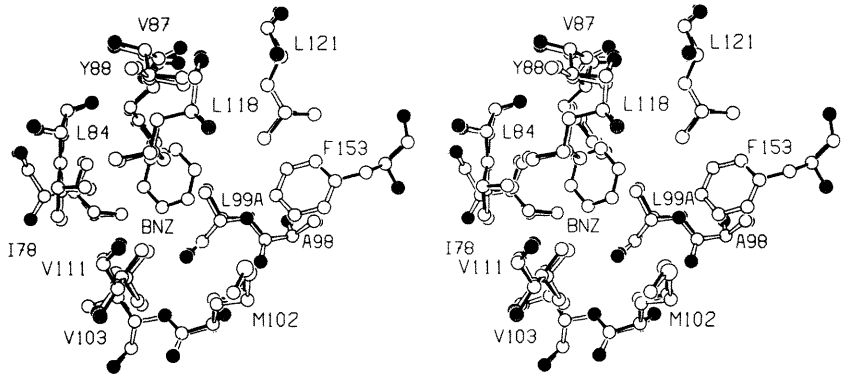


Fig. 5 Superposition of the benzene-containing mutant lysozyme structure L99A (solid bonds) on the mutant structure in the absence of ligand (open bonds). Every side chain shown in this figure is itself completely inaccessible to solvent

of leucine and alanine estimated by transfer from water to ethanol,[15] octanol[17] or *N*-methylacetamide.[18] A value of 1.9 kcal mol^{-1} for leucine relative to alanine corresponds to *ca.* 25 cal mol^{-1} Å^{-2}. The cavity-dependent energy term, at least for the Leu → Ala (or Phe → Ala) replacements in T4 lysozyme, is 24 cal mol^{-1} Å^{-3} (see also ref. 21 for an alternative estimate). Expressed in terms of area the energy cost is 20 cal mol^{-1} for each Å^2 of cavity surface created.

This suggests a way to reconcile the different values for the hydrophobic strength obtained, on the one hand, by solvent-transfer experiments and, on the other hand, by directed mutagenesis. Fig. 4(*a*) and (*b*) suggest that the change in energy associated with the replacement of a buried leucine with an alanine consists of two parts. The first part is a constant and is presumed to depend only on the identities of the two amino acids being compared, in this instance leucine (or phenylalanine) and alanine. Physically this energy term can be considered as the difference in energy required to desolvate (*i.e.* transfer from solvent to protein interior) a leucine relative to an alanine. The second part of the change in protein stability associated with the Leu → Ala replacements [Fig. 4(*a*), (*b*)] depends on the context within the three-dimensional structure and the way in which the protein structure adjusts in response to the substitution. One can imagine two extreme situations. In one case a Leu → Ala replacement is constructed and the protein structure remains completely unchanged (*cf.* L99A). In this situation the size of the created cavity is large and the mutant (Ala) protein is maximally destabilized. In the other extreme, the protein structure relaxes in response to the Leu → Ala substitution, fills the space occupied by the leucine side chain, and so avoids the formation of any cavity whatsoever. In this case, and in the absence of any other energy terms that might come into play, the decrease in energy of the mutant protein relative to wild type would reduce to the constant energy term described above, *i.e. ca.* 1.9 kcal mol^{-1}.

In order to explain the observed changes in protein stability caused by cavity-creating mutations it is not necessary to suggest that the hydrophobic effect should be counted twice, once for the residue removed and once for the cavity created.[12,13] Neither is it necessary to argue that the strength of the hydrophobic effect needs to be revised from the accepted value of *ca.* 25–30 cal mol^{-1} Å^{-2} to a new value of 43–47 cal mol^{-1} Å^{-2} [19,20,22]

Ligand Binding with cavities

We have shown by crystallographic (Fig. 5) and thermodynamic analysis that the cavity created by the replacement Leu-99 → Ala in T4 lysozyme is large enough to bind benzene

and that ligand binding increases the melting temperature of the protein by 5.7 °C at pH 3.0.[23] Benzene does not, however, bind to the cavity created by the Phe-153 → Ala replacement. This shows that cavities can be engineered within proteins and suggests that such cavities might be tailored to bind specific ligands. The binding of benzene at an internal site 7 Å from the molecular surface also illustrates the dynamic nature of proteins, even in crystals.

The excellent technical assistance of Sheila Pepiot and Joel Lindstrom is gratefully acknowledged. This work was supported in part by grants to B.W.M. from the National Institutes of Health (GM21967) and the Lucille P. Markey Charitable Trust.

References

1 U. H. Sauer, S. Dao-pin and B. W. Mathews, *J. Biol. Chem.*, 1922, **267**, 2393.
2 M. F. Perutz and H. Lehmann, *Nature* (*London*), 1968, **219**, 902.
3 J. F. Reidhaar-Olson and R. T. Sauer, *Science*, 1988, **241**, 53.
4 T. Alber, S. Dao-pin, J. A. Nye, D. C. Muchmore and B. W. Matthews, *Biochemistry*, 1987, **26**, 3754.
5 X-J. Zhang, W. A. Baase and B. W. Matthews, *Biochemistry*, 1991, **30**, 2012.
6 X-j. Zhang, W. A. Baase and B. W. Matthews, *Protein Sci.*, in the press.
7 D. W. Heinz, W. A. Baase and B. W. Matthews, *Proc. Natl. Acad. Sci. USA*, in the press.
8 C. Chothia, *Nature* (*London*), 1974, **248**, 338.
9 F. M. Richards, *Annu. Rev. Biophys. Bioeng.*, 1977, **6**, 151.
10 M. Matsumura, W. J. Becktel and B. W. Matthews, *Nature* (*London*), 1988, **334**, 406.
11 M. Matsumura, J. A. Wozniak, S. Dao-pin and B. W. Matthews, *J. Biol. Chem.*, 1989, **264**, 16059.
12 J. T. Kellis, K. Nyberg, D. Sali and A. R. Fersht, *Nature* (*London*), 1988, **333**, 784.
13 J. T. Kellis, K. Nyberg and A. R. Fersht, *Biochemistry*, 1989, **28**, 4914.
14 W. S. Sandberg and T. C. Terwilliger, *Proc. Natl. Acad. Sci. USA*, 1991, **88**, 1706.
15 D. Shortle, W. E. Stites and A. K. Meeker, *Biochemistry*, 1990, **29**, 8033.
16 Y. Nozaki and C. Tanford, *J. Biol. Chem.*, 1971, **246**, 2211.
17 J-L. Fauchere and V. Pliska, *Eur. J. Med. Chem.-Chim. Ther.*, 1983, **18**, 369.
18 S. Damodaran and K. B. Song, *J. Biol. Chem.*, 1986, **261**, 7220.
19 K. A. Sharp, A. Nicholls, R. F. Fine and B. Honig, *Science*, 1991, **252**, 106.
20 K. A. Sharp, A. Nicholls, R. Friedman and B. Honig, *Biochemistry*, 1991, **30**, 9686.
21 A. E. Eriksson, W. A. Baase, X-J. Zhang, D. W. Heinz, M. Blaber, E. P. Baldwin and B. W. Matthews, *Science*, 1992, **255**, 178.
22 L. R. DeYoung and K. A. Dill, *J. Phys. Chem.*, 1990, **94**, 801.
23 A. E. Eriksson, W. A. Baase, J. A. Wozniak and B. W. Matthews, *Nature* (*London*), 1992, **355**, 371.

Paper 1/06407J; *Received* 20*th December*, 1991