Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 423–427, January 1994
Biophysics

The crystal structure of a mutant protein with altered but improved hydrophobic core packing

(protein folding/protein evolution)

WENDELL A. LIM*, ALEC HODEL*, ROBERT T. SAUER†, AND FREDERIC M. RICHARDS*

*Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511; and †Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by Frederic M. Richards, September 29, 1993

ABSTRACT **The dense packing observed in protein interiors appears to be crucial for stabilizing the native structure—even subtle internal substitutions are usually destabilizing. Thus, steric complementarity of core residues is thought to be an important criterion for "inverse folding" predictive methods, which judge whether a newly determined sequence is consistent with any known folds. A major problem in the development of useful core packing evaluation algorithms, however, is that there are occasional mutations that are predicted to disrupt native packing but that yield an equally or more stable protein. We have solved the crystal structure of such a variant of λ repressor, which, despite having three larger core substitutions, is more stable than the wild type. The structure reveals that the protein accommodates the potentially disruptive residues with shifts in its α-helical arrangement. The variant is apparently more stable because its packing is improved—the core has a higher packing density and little geometric strain. These rearrangements, however, cause repositioning of functional residues, which result in reduced DNA binding activity. By comparing these results with the predictions of two core packing algorithms, it is clear that the protein possesses a relatively high degree of main-chain flexibility that must be accounted for in order to predict the full spectrum of compatible core sequences. This study also shows how, in protein evolution, a particular set of core residue identities might be selected not because they provide optimal stability but because they provide sufficient stability in addition to the precise structure required for optimal activity.**

A major goal in structural biology is to be able to predict the three-dimensional structure of a newly sequenced protein. So-called "inverse folding" methods have proved to be one of the most successful predictive schemes (1). In this approach, a recently determined sequence is compared to a library of known structures. One evaluates the likelihood that the sequence adopts one of these folds by determining whether, when arranged in that conformation, the sequence would satisfy a set of empirical criteria for well-folded proteins. One criterion thought to be of central importance is the steric complementarity of internal residues. In known structures, core residues fill almost all the available interior space with minimal geometric strain and no steric overlaps (2). Such dense packing is thought to provide many favorable van der Waals interactions as well as the exclusion of solvent, thereby maximizing hydrophobic stabilization. Consistent with this view is the finding that even subtle internal substitutions tend to be destabilizing (3–8). Several algorithms have therefore been developed to determine which sets of residues can pack together efficiently within a given backbone fold (9, 10). In these methods, the backbone is held fixed and side chains are rotated to find any sterically acceptable conformations. Such algorithms correctly predict highly homologous core sequences to be acceptable. A significant problem, however, is that there are cases in which more deviant residue combinations are predicted to be sterically incompatible with a given fold yet yield proteins equally or more stable than the wild type.

To improve packing algorithms, we must understand how these potentially disruptive residues can be tolerated in the core of a protein with minimal destabilization. Is good packing actually not required, or does the structure change to allow the new set of core residues to maximize packing density in an alternative way? If the structure does change, how does it do so in a way that maintains high stability? To address these questions, we crystallized and determined the structure of the λ repressor triple mutant Val-36 → Leu, Met-40 → Leu, Val-47 → Ile (referred to hereafter as LLI) complexed to the O_L1 operator. This mutant protein has two extra methylene equivalents inserted in the core and its core sequence has been predicted by one algorithm to be sterically incompatible with the wild-type fold (9). The LLI protein should therefore provide a good example of how potentially disruptive internal substitutions can be favorably accommodated, since despite its mutations the protein is 0.5 kcal/mol (1 cal = 4.184 J) more stable than the wild type by guanidine hydrochloride denaturation and has a t_m 4°C higher (4). Interestingly, this increased stability appears to be gained at the expense of structural rearrangement, since the protein has a 10-fold reduced affinity for operator DNA as well as a 100-fold reduced affinity for a conformation-specific monoclonal antibody (4).

MATERIALS AND METHODS

Protein Expression and Purification. The gene for the LLI mutant protein was initially constructed in the background of residues 1–102 of λ repressor (11). A restriction fragment bearing all three mutations was ligated into a vector expressing residues 1–92 of λ repressor, the protein fragment that had been crystallized previously (12, 13). The mutant protein was expressed and purified as described (4), with an additional chromatography step on a C-4 reverse-phase column using a 0–80% acetonitrile gradient (in 0.1% trifluoroacetic acid). The purified protein in the 1–92 background showed behavior identical to that observed in the 1–102 background, as assayed by circular dichroism spectroscopy and by thermal unfolding.

Crystallization, Data Collection, and Refinement. Crystals of the mutant protein complexed to the 20-bp O_L1 operator were grown as described (13) except that 16% PEG 400 was used as a precipitant. We choose to crystallize the mutant as a cocrystal with DNA because the crystals of wild-type protein alone diffract to only 3.2 Å, whereas those of the cocrystal diffract to 1.8 Å. Moreover, within the limits of resolution, the structure of the wild-type protein complexed with DNA is identical to that of the protein alone (14). The

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

mutant complex crystallized in the same space group as the wild type ($P2_1$) with nearly identical cell constants (mutant: a = 37.15, b = 68.90. c = 56.79, β = 92.2°; wild type: a = 37.22, b = 68.72. c = 57.03, β = 92.2°). Data were collected at 0°C on a single crystal (approximately 0.3 × 0.2 × 0.1 mm) on a Xuong-Hamlin Mark II multiwire detector (15) and scaled using the program SCALEPAK (16); 99.7% of the reflections between 2.1 and 8 Å resolution were collected, and statistics on these data are given in Table 1. An initial electron density map was calculated by using phases from the wild-type structure, with the three mutant side chains truncated at their C^βs to avoid model bias. This initial model gave an R factor of 0.354. The model was subjected to conventional (positional and B factor) refinement using the program X-PLOR (17) until convergence. Mutant side chains were built into $2F_o - F_c$ maps and the model was subject to simulated annealing refinement, followed by conventional refinement. At this point, simulated annealing omit maps (18), calculated after omitting the mutated residues and 10% of the surrounding atoms, were examined. Internal side-chain conformations were carefully checked using these unbiased maps as a guide and were rebuilt if necessary. After final rounds of conventional refinement, and the addition of 183 solvent molecules, the R factor was 0.196. The rms deviations from ideal geometries in the final model are ±0.015 Å for bond lengths and ±3.0° for bond angles.

Packing Density Calculations. Packing densities are defined as

$$\text{packing density} = \frac{\text{van der Waals volume}}{\text{occupied volume}}.$$

Densities were calculated by two methods, each using a different measure of occupied volume. The Connolly method calculates the volume within the total molecular surface defined by rolling a 1.4-Å-radius probe over the ensemble of atoms (19). The second method uses a radical plane Voronoi procedure (20) to define a polyhedron associated with each atom. The total occupied volume of an ensemble of atoms is the sum of these Voronoi volumes. In both methods, we have calculated van der Waals volumes by using the Connolly procedure but with a probe radius of 0. The atomic radii listed by Richards (20) were used for all calculations.

RESULTS AND DISCUSSION

The LLI mutant protein was crystallized with the 20-bp O_L1 operator site and the structure was refined to an R factor of 0.196 at 2.1 Å resolution. As with the wild-type N-terminal domain of λ repressor (12, 13), the protein crystallizes in a complex consisting of a protein dimer and a single operator. The overall fold of the mutant protein is nearly identical to that of the wild type—when the entire structures are aligned, the main-chain rms deviation is 0.3 Å. There are, however, small but significant changes in the structure, which are discussed below.

A common problem in comparing mutant and wild-type structures is that differences tend to be small, and therefore it can be difficult to distinguish significant changes from error due to limitations of the data. Fortunately, this structure comprises two nonidentical protein monomers (the operator sequence is nonpalindromic). Although somewhat different sets of structural changes are observed in the two monomers, those changes observed in both monomers are likely to be real. The structural features discussed below are therefore limited to those that are similar in both protein monomers.

Table 1. X-ray data statistics

Resolution shell, Å	Completeness, %	R_{sym}*	Avg. redundancy	R factor†
8.0–3.0	99.8	0.07	5.5	0.169
3.0–2.4	99.9	0.22	4.9	0.228
2.4–2.1	99.4	0.31	3.8	0.261
All (8.0–2.1)	99.7	0.10	4.4	0.196

*$R_{sym} = (\Sigma|I - \langle I\rangle|)/(\Sigma\langle I\rangle)$.
†Crystallographic R factor $= (\Sigma|F_o - F_c|)/(\Sigma F_o)$.

The Mutant Protein Has a Slightly Altered Main-Chain Conformation. A difference distance matrix of changes in the inter-C^α distances (Fig. 1*A*) clearly shows that there are some changes in the precise main-chain conformation. The largest change observed in both monomers is an overall shift of the C-terminal half of helix 4 (residues 65–70) away from helices 2 and 3 (residues 27–54), where the three mutations are located (Fig. 1*B*). Within this general shift, the largest individual movements are 0.9 Å, observed between the C^α of Ile-68 and the C^αs of residues 36 and 40. The new and larger mutant side chains at positions 36 and 40 would sterically overlap the side chain of Ile-68 in its original position. However, as shown in Fig. 2, the protein alleviates this

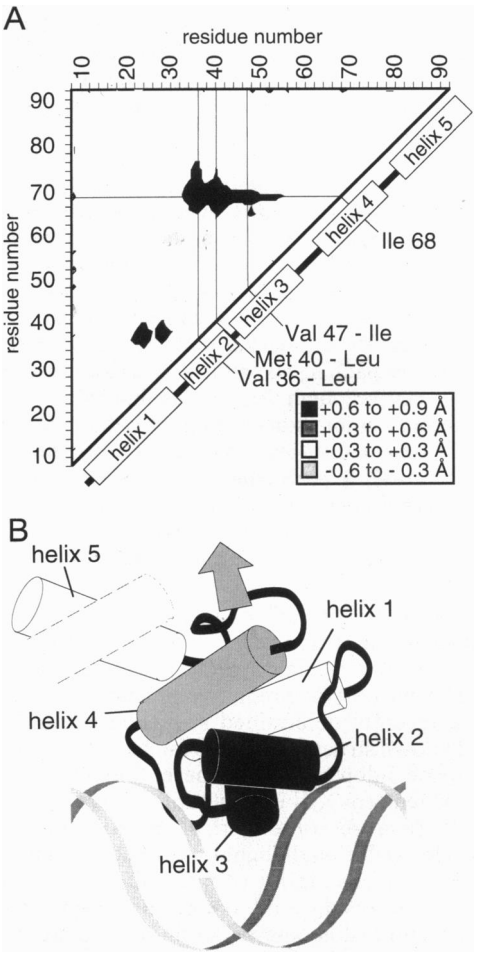

FIG. 1. Changes in the LLI mutant structure. (*A*) Difference distance matrix showing relative C^α displacements upon mutation. The plot shown is calculated for monomer 1. A positive displacement indicates a relative expansion upon mutation. Contours begin at ±0.3 Å because coordinate error is estimated at ≈0.3 Å by Luzzatti analysis (21). The largest individual shifts observed are +0.9 Å, between the C^α of residue 68 and the C^αs of both residues 36 and 40. (*B*) Cartoon indicating general main-chain shifts that occur upon mutation. A single monomer is shown bound to DNA. Dashed line indicates the position of helix 5 of the second monomer. As shown in *A*, there is a shift of the C terminus of helix 4 (light shading) away from the helix 2–3 unit (dark shading). The three mutated residues are in the helix 2–3 unit.

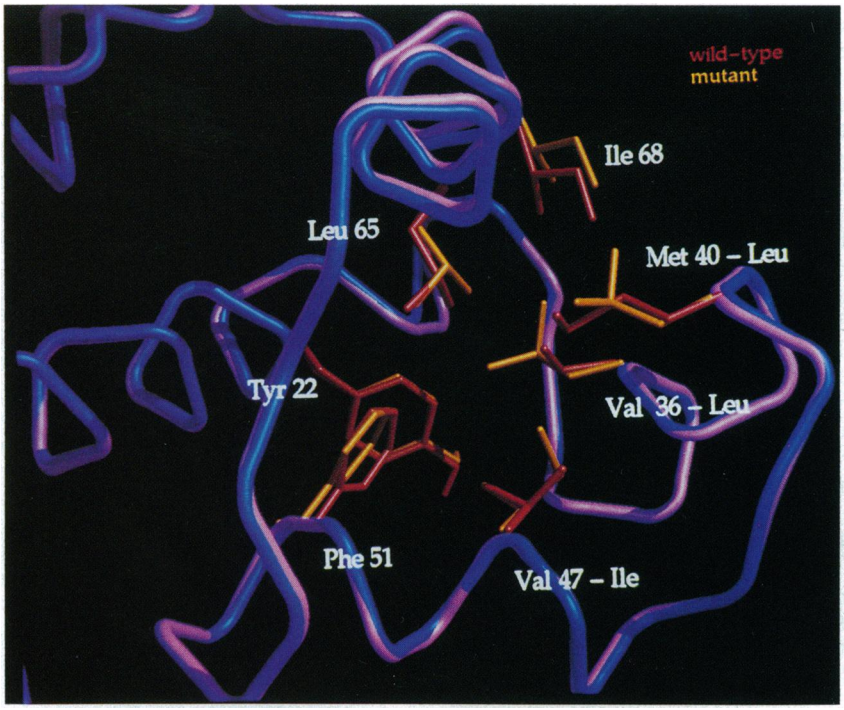

FIG. 2. Conformational shifts upon mutation, shown at atomic resolution. The internal region of monomer 2 is shown. The wild-type C^{α} trace and side chains are shown in blue and red, respectively, and the mutant C^{α} trace and side chains are shown in magenta and orange, respectively. Mutant and wild-type structures were aligned using residues 32–58. The substitutions at positions 36 and 40 clash with residues 65 and 68, forcing the small outward translation and counterclockwise rotation (as viewed here) of helix 4.

potential clash by displacing and slightly rotating helix 4 away so that residue 68 no longer points as directly into the core. The movement of helix 4 is largely a rigid body motion, as the helix itself shows no significant deformation.

These main-chain rearrangements may account for the reduced DNA and antibody affinity of the mutant protein. All of the hydrogen bonds between the protein dimer and the DNA present in the wild-type complex (12, 13) are also observed within the mutant complex, suggesting that the 10-fold reduction in affinity is not caused by the loss of any particular interaction. Most likely, the activity change is a result of the overall main-chain expansion observed in the structure. This expansion, which pushes helix 3, the DNA reading helix, away from helix 4, which is part of the dimer interface, could have two potentially deleterious effects: (*i*) the expansion increases the distance between each DNA binding helix in the two monomers by a few tenths of an angstrom, which could perturb the energy of many of the protein–DNA interactions; (*ii*) the expansion could destabilize the dimerization interaction, which would lower the overall binding equilibrium, given that under both *in vivo* and *in vitro* assay conditions the protein exists in solution largely as a monomer but binds as a dimer (22). It is not surprising that the observed displacements would reduce the affinity for the conformation-specific antibody 51F (4), since this antibody is known to recognize an epitope in helix 4 (23).

Large Side-Chain Rotations Are Not Observed. Very little of the structural relaxation that takes place in the LLI mutant occurs through side-chain rotations. A majority of nonmutated side chains remain in essentially identical conformations. More important, within the limits of resolution, almost all of the internal side chains in the mutant structure appear to adopt nearly ideal canonical rotamer conformations (9, 24–26), with the exception of Leu-40, which has a moderately strained χ_2 torsional angle. As shown in Fig. 3, the side-chain dihedral angles for the mutated residues in both monomers are essentially unaltered. The maintenance of similar dihedral angles is also observed for two of the buried residues in helix 4, Leu-65 and Ile-68, which pack directly against the mutated residues and are most directly challenged to find alternative conformations. The bulk of the observed structural relaxation therefore appears to occur via subtle changes in main-chain torsional angles resulting in an essentially rigid body shift of helix 4 relative to helices 2 and 3. In this case,

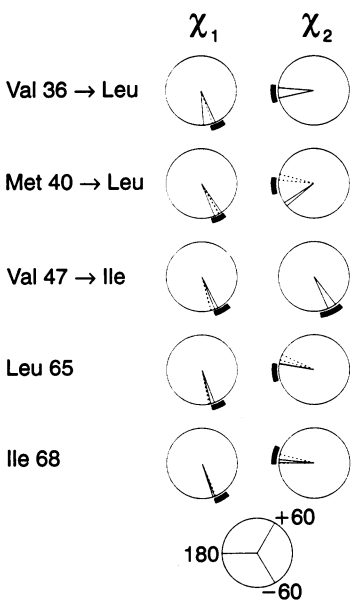

FIG. 3. Side-chain dihedral angles of mutated residues and other surrounding internal residues. The angles observed in the two mutant monomers are shown with a solid line; those observed in the two wild-type monomers are shown with a dashed line. The arcs outside of each circle show the mean torsional angle, ±1 SD, for the closest canonical rotamer as described by Ponder and Richards (9). To allow for direct comparison to leucine and isoleucine, the χ_1 angle for valine has been calculated using the $C^{\gamma 2}$ instead of $C^{\gamma 1}$.

movement of entire secondary structure elements is apparently energetically preferable to side-chain rotation as a means to escape steric overlap, despite the fact that such a movement involves considerably more atoms.

The Mutant Protein Has Improved Internal Packing. Most internal mutations in proteins are destabilizing (3–8, 27), presumably caused, in part, by the disruption of native packing interactions. Eriksson *et al.* (27) have shown that when large core side chains are replaced with smaller ones, the loss in stability appears to correlate both with a loss in hydrophobicity and with a loss in internal packing efficiency. Smaller internal substitutions are generally found to cause structural rearrangements that minimize the size of the created cavities (5, 27–29). Reduced packing efficiency also appears to contribute to the destabilization caused by replacing smaller side chains with larger ones. Despite having greater hydrophobicity, these larger substitutions are often found to cause packing lesions as a result of sterically required rearrangements (5–8). In addition, larger internal substitutions can introduce energetically unfavorable close contacts and torsional strain (8).

How are hydrophobicity and packing affected by the λ repressor mutations discussed here? In terms of hydrophobicity alone, the substitutions should theoretically stabilize the protein by 2.1 kcal/mol (30). However, because of the rearrangement that occurs in the structure, several buried residues become slightly more solvent exposed. These include Leu-40 (one of the mutant residues), Ile-68 (which, as described previously, is rotated out to avoid overlap with the mutant residues), and Leu-31. By using the atomic solvation parameters of Eisenberg and McLachlan (31), the loss of hydrophobic stabilization as a result of the new surface area exposed is calculated to balance any gain expected from the increased hydrophobicity of the substitutions alone.

We have calculated the packing density in the cores of both the LLI mutant and the wild-type protein in two ways as described in Table 2. By both methods, we find that despite the expansion of the protein main chain, the mutant protein appears to have a core that is packed as well as or slightly better than the wild-type core. The maintenance of a high packing density, in combination with the lack of significant geometric strain, is unusual in a protein with internal substitutions and may account for the equally unusual fact that this protein is more stable than the wild type.

Evaluation of Predictive Packing Algorithms. Because of its unusual properties, the LLI mutant of λ repressor provides a challenging test of computational methods that use packing criteria to enumerate internal residue combinations consistent with a particular protein fold (9, 10). In general, these methods have proven to be overly restrictive. For example, the LLI mutant sequence clearly is compatible with the native fold, but it has been predicted to be unacceptable by the "tertiary template" algorithm of Ponder and Richards (9). This algorithm uses a library of canonical side-chain rotamers to search for alternative side-chain combinations that can pack efficiently within a fixed main-chain scaffold without torsional or steric strain. Several of the principles underlying this approach appear to be valid, since in the LLI mutant structure we find that the protein does indeed maximize packing density while minimizing torsional strain. However, the algorithm of Ponder and Richards predicts the LLI mutant sequence to be unacceptable because leucine side chains cannot be inserted in canonical rotational conformations at both positions 36 and 40 without resulting in steric overlaps with neighboring residues. At the molecular level, this prediction is in part correct, since the mutant leucine side chains do adopt nearly canonical rotamers. However, they act to push away the segments of structure with which they might potentially overlap. Clearly, the assumption that the main chain stays fixed is incorrect and severely limiting.

Table 2. Packing density of mutant and wild-type cores

Unit	Wild type	Mutant	Δ packing density, %
Connolly method			
Core (monomer 1)	0.867	0.869	+0.2
Core (monomer 2)	0.847	0.859	+1.2
Monomer 1	0.797	0.797	+0.0
Monomer 2	0.799	0.810	+1.1
Richards method			
Core (monomer 1)	0.784	0.798	+1.4
Core (monomer 2)	0.761	0.787	+2.6
Monomer 1	0.724	0.728	+0.4
Monomer 2	0.722	0.746	+2.4

Packing densities were calculated as described. The units over which the calculations were performed are defined as follows: core includes the three mutated residues (residues 36, 40, and 47) as well as all residues within 3.75 Å (residues 31, 32, 35, 37, 39, 41, 46, 48, 51, 64, and 68); monomer is the entire protein monomer, excluding poorly defined terminal residues (includes residues 7–85).

In contrast, an algorithm that correctly predicts the LLI sequence to be acceptable is that of Lee and Levitt (10). This algorithm also begins with the assumption that the main chain will stay fixed. A simulated annealing procedure is then used to search all side-chain rotational conformations (not limited to canonical rotamers) for a mutated packing unit as well as a few nonmutated residues in a surrounding "molten zone." The energies are evaluated by using a torsional and van der Waals potential. By this method, the LLI sequence was found to be acceptable, and in fact to be more stable than wild type, but only if two of the mutated side chains adopt strained torsional angles. In the predicted structure, the χ_1 angle of both residues 36 and 47 deviates from the closest canonical values for leucine (2) by more than 2 SD. In this case, although the overall prediction of acceptability is correct, prediction of the molecular mechanism of relaxation is incorrect. Experimentally, the mutant side chains generally adopt unstrained conformations, and relaxation occurs largely through main-chain movements.

With respect to the general goal of designing or predicting protein structure, this and related studies (32) suggest that it will be difficult to fully understand the role of packing interactions in specifying protein structure without first understanding the range of motions available to main-chain segments. Nonetheless, these findings are encouraging since they lend support to some assumptions, such as the canonical rotamer approximation, that can greatly simplify predictive methods.

These types of packing algorithms might be improved by considering a particular fold as an ensemble of slightly different main-chain structures. For example, a fold such as that of the λ repressor might have several low energy main-chain conformations available to it, and the protein may adopt the one that, among other requirements, allows the specific set of core residues to pack well. It is noteworthy that if the experimentally determined LLI mutant main-chain structure is used in the algorithm of Ponder and Richards instead of the wild-type structure, then the LLI sequence, as well as several other sequences incorrectly predicted to be unacceptable, is judged to be acceptable. The precise fold observed for the LLI mutant may therefore represent one of a few alternative low energy main-chain structures.

Implications for the Evolution of Protein Cores. In contrast to external residues directly involved in ligand or substrate binding, a protein's interior residues are often thought of as playing a basic structural role and not a direct functional role. Thus, the selective pressures on core residues are often perceived as requirements for maintenance of high stability. This mutant protein, however, illustrates how the identity of

core residues can directly exert as significant an effect on function as external residues. The particular set of core residues and the packing restrictions placed on them can subtly alter the relative three-dimensional placement of binding residues so as to dramatically reduce activity. Clearly, the wild-type λ repressor core residues do not yield maximal stability and packing. Presumably, however, this set of residues has been selected because it optimizes both the need for high stability and the need for the precise high-affinity DNA binding structure.

We thank R. Fox and A. Friedman for help in data collection as well as E. Baldwin, L. Beamer, C. Pabo, E. Kim, P. Harkins, M. Jacobs, A. Brünger, B. Matthews, K. Earle, and G. Kellogg for assistance and advice. This work was supported by National Institutes of Health Grants GM-22778 and AI-15706. W.A.L. is a Helen Hay Whitney Fellow and A.H. is a Howard Hughes Predoctoral Fellow.

1. Bowie, J. U., Lüthy, R. & Eisenberg, D. (1991) *Science* **253,** 164–170.
2. Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng.* **6,** 151–176.
3. Sandberg, W. S. & Terwilliger, T. C. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 1706–1710.
4. Lim, W. A., Farruggio, D. C. & Sauer, R. T. (1992) *Biochemistry* **31,** 4324–4333.
5. Varadarajan, R. & Richards, F. M. (1992) *Biochemistry* **31,** 12315–12327.
6. Dao-pin, S., Alber, T., Baase, W. A., Wozniak, J. A. & Matthews, B. W. (1991) *J. Mol. Biol.* **221,** 647–667.
7. Hurley, J. H., Baase, W. A. & Matthews, B. W. (1992) *J. Mol. Biol.* **224,** 1143–1159.
8. Karpusas, M., Baase, W. A., Matsumura, M. & Matthews, B. W. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 8237–8241.
9. Ponder, J. W. & Richards, F. M. (1987) *J. Mol. Biol.* **193,** 775–791.
10. Lee, C. & Levitt, M. (1991) *Nature (London)* **352,** 448–451.
11. Lim, W. A. & Sauer, R. T. (1991) *J. Mol. Biol.* **219,** 359–376.
12. Jordan, S. R. & Pabo, C. O. (1988) *Science* **242,** 893–899.
13. Beamer, L. J. & Pabo, C. O. (1992) *J. Mol. Biol.* **227,** 177–196.
14. Pabo, C. O. & Lewis, M. (1982) *Nature (London)* **298,** 443–447.
15. Hamlin, R. (1985) *Methods Enzymol.* **114,** 416–452.
16. Otwinowski, Z. O. (1993) SCALEPAK (Yale University, New Haven, CT), Version 1.0.
17. Brünger, A. T., Kuriyan, J. & Karplus, M. (1987) *Science* **235,** 458–460.
18. Hodel, A., Kim, S. H. & Brünger, A. T. (1992) *Acta Crystallogr. A* **48,** 851–858.
19. Connolly, M. L. (1983) *Science* **221,** 709–713.
20. Richards, F. M. (1974) *J. Mol. Biol.* **82,** 1–14.
21. Luzzatti, V. (1952) *Acta Crystallogr.* **5,** 802–810.
22. Sauer, R. T., Jordan, S. & Pabo, C. O. (1990) *Adv. Protein Chem.* **40,** 1–61.
23. Breyer, R. M. & Sauer, R. T. (1989) *J. Biol. Chem.* **264,** 13355–13360.
24. Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thorton, J. M. (1992) *Proteins Struct. Funct. Genet.* **12,** 345–364.
25. McGregor, M. J., Islam, S. A. & Sternberg, M. J. (1987) *J. Mol. Biol.* **198,** 295–310.
26. Dunbrack, R. L., Jr., & Karplus, M. (1993) *J. Mol. Biol.* **230,** 543–574.
27. Eriksson, E. A., Baase, W. A., Zhang, X.-J., Heinz, D. W., Blaber, M., Baldwin, E. P. & Matthews, B. W. (1992) *Science* **255,** 178–183.
28. Katz, B. & Kossiakoff, A. A. (1990) *Proteins Struct. Funct. Genet.* **7,** 343–357.
29. McRee, D. E., Redford, S. M., Getzoff, E. D., Lepock, J., Hallewell, R. A. & Tainer, J. A. (1990) *J. Biol. Chem.* **265,** 14234–14241.
30. Fauchere, J.-L. & Pliska, V. (1983) *Eur. J. Med. Chem.* **18,** 369–375.
31. Eisenberg, D. & McLachlan, A. D. (1986) *Nature (London)* **319,** 199–203.
32. Baldwin, E. P., Hajiseyedjavadi, O., Baase, W. A. & Matthews, B. W. (1993) *Science*, in press.