Critical Functional Role of the COOH-terminal Ends of Longitudinal Hydrophobic Strips in α-Helices of T4 Lysozyme*

(Received for publication, April 28, 1992)

Dale Rennell§, Anthony R. Poteete§§, Michelle Beaulieu$$, Dennis Z. Kuo$$*, Robert A. Lew$$, and Robert E. Humphreys$$§§

From the Departments of Molecular Genetics and Microbiology, Pharmacology, Medicine, and the Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts 01655

The sensitivity of bacteriophage T4 lysozyme function to amino acid substitutions at defined positions in and around the longitudinal, hydrophobic strips of 9 α-helices was assessed after systematic replacement of each residue in the protein with a series of 13 amino acids. The hydrophobic strips were defined by identifying the longitudinal sectors in the helices with the highest mean residue hydrophobicities. Sensitivity to mutation (the percentage of replacements leading to loss of function) was calculated for each residue in the following positions: whole protein, helices, hydrophobic strips, other positions within the helices, and various positions within the hydrophobic strips as well as their extensions beyond the helices. Substitutions at positions in the hydrophobic strips led more frequently to loss of function than substitutions in the protein as a whole. One subset, the COOH-terminal hydrophobic strip residues, is apparently critical; substitutions of these residues (but not of their NH2-terminal counterparts) led at least as frequently to loss of function as substitutions of solvent-inaccessible residues, and nearly as frequently as substitutions of the most highly conserved residues.

Proteins are generally tolerant of amino acid substitutions. Studies of natural variants, as well as of proteins subjected to intensive mutagenesis, have revealed that many, possibly most, single amino acid substitutions are tolerated. Moreover, it appears that few, if any, residues in a protein are irreplaceable (1–5). If combinations of substitutions are permitted, even the hydrophobic core of a protein can be packed in many different ways (6).

Against this background of tolerance, certain positions in proteins stand out as intolerant of substitutions. These critical residues are ones whose replacement with other residues frequently results in a loss of function. For purposes of understanding and engineering proteins, it would be useful to be able to pick out such critical residues in a protein of known structure, without having to carry out systematic mutational studies. A few general rules for doing this exist. For instance, residues identified by structural or chemical data as involved in substrate binding and/or catalysis usually are critical for function. In addition, residues with buried side chains tend to be sensitive to substitutions (2, 3). Finally, residues that are conserved among members of families of related proteins are more sensitive to substitutions than non-conserved residues (7).

In an attempt to elucidate further general rules that might be of use in picking out critical residues and to understand the structural basis of those residues’ importance, we have investigated the applicability of a structure-predicting algorithm. This algorithm, described previously (8), identifies a longitudinal hydrophobic strip in a surface α-helix (see Fig. 1). Like some other methods for predicting α-helices, it is based on the observation that α-helices on the surfaces of proteins tend to be amphipathic (1). Typically, one longitudinal sector of the helix will have predominantly hydrophobic residues, which pack into the hydrophobic core of the protein.

In a previous study (9), it was observed that, in the helices of 55 proteins, Leu, Ile, Val, Phe, and Met were found at high frequency in the longitudinal hydrophobic strip, relative to other positions; they were not found at high frequency in the first “virtual” hydrophobic strip positions (identified by extension of the hydrophobic strip beyond the helices). It was suggested that the formation of the longitudinal hydrophobic strip was critical to helix propagation and that the absence of hydrophobic residues in strip positions beyond a helix dictated helix termination.

We now find that application of the hydrophobic strip-of-helix algorithm to the α-helices of T4 lysozyme picks out a small set of residues, associated with the hydrophobic strip, that are critical to function. Their apparent functional importance equals or exceeds that of the least solvent-accessible residues and nearly equals that of the residues in T4 lysozyme

* This research was supported by Grant IM-582 from the American Cancer Society, Grant AI-24083 from the National Institutes of Health, and by a grant from the scientific council of the University of Massachusetts Medical School for the Molecular Graphics Facility. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1754 solely to indicate this fact.

§ Undergraduate research fellow (from Dartmouth College) of the National Cancer Institute.

** Fellow (from Williams College) of the National Cancer Institute (professional education training grant).

$$ To whom correspondence should be addressed: Dept. of Pharmacology, University of Massachusetts Medical School, 55 Lake Ave. North, Worcester, MA 01655. Tel.: 508-856-3327; Fax: 508-856-5090.

Fig. 1. Sheet projection of the strip-of-helix hydrophobicity template. □ positions in the linear sequence are seen to fall in one longitudinal quadrant of an α-helix. The strip-of-helix hydrophobicity index is the mean hydrophobicity of □ positions in the fitting of the repeating template pattern to a known or putative helix maximizing the index. The template may then be extended beyond the helix to define adjacent (virtual helix) ○ and □ positions.

17748

Vol. 267, No. 25, Issue of September 5, pp. 17748-17752, 1992
Printed in U.S.A.
that are the most highly conserved among phage-encoded lysozymes. These critical residues are at the COOH-terminal ends of their respective helices' hydrophobic strips. The corresponding NH2-terminal residues are significantly less sensitive. These observations are discussed in terms of an apparent pattern in the structural organization of surface α-helices in T4 lysozyme.

**EXPERIMENTAL PROCEDURES**

For each of the α-helices of T4 lysozyme found in the crystal structure (10, 11), the longitudinal hydrophobic strip-of-helix was defined by fitting a circular template, . . . [diagram with symbols representing template]. The frequency of these residues' occurrences was higher than at other positions in the protein, as one would expect. In contrast, the frequency of these residues' occurrence in the NH2- and COOH-terminal, first virtual hydrophobic positions about the helices of T4 lysozyme was not elevated above the uniform distribution found also in the larger survey of Vazquez et al. (9) (data not shown). In that study, lack of extension of a longitudinal hydrophobic strip-of-helix with Leu, Ile, Val, Phe, or Met correlated with the termination of the helix.

In the survey of Vazquez et al. (9), the crossing region of a helix was defined as the 3 residues closest to a neighboring helix in terms of minimal interhelical C, distances. It was observed that helices generally crossed through their longitudinal hydrophobic strips-of-helix. Furthermore, the smallest residue in the longitudinal hydrophobic strip-of-helix was usually in the crossing region, i.e. helices crossed through

---

1 S. Dao-Pin and L. Weaver, personal communication.
"notches" in the longitudinal hydrophobic strip-of-helix. Charged residues which occurred in the longitudinal hydrophobic strip-of-helix were excluded from crossing regions. The sides of such notches in the longitudinal hydrophobic strip-of-helix were composed of Leu, Ile, Val, Phe, and Met predominantly. It was suggested that the structure of the longitudinal hydrophobic strip influenced the rotational and longitudinal positioning of crossing helices.

The relative sensitivities to substitutions of all residues in the α-helices of T4 lysozyme are indicated in Fig. 2. Residues of the hydrophobic strip were generally sensitive; the COOH-terminal strip residues were particularly sensitive. The sensitivities to substitutions of groups of residues in T4 lysozyme, including those occupying the structural positions described above, are compared in Table I. The protein as a whole scored 16, that is 16% (328/2015) of substitutions tested were found to be deleterious. Buried residues, as a group, were more sensitive to substitutions. Loss of function resulted from 38% of substitutions for residues with side chains which have less than 12% of their surface areas accessible to solvent; this sensitivity increased to 42% if the group was restricted to those residues with completely inaccessible side chains. These observations establish a criterion for the performance of a scheme to pick out critical residues. As shown previously, residue conservation, too, is correlated with sensitivity to substitutions; 47% of substitutions in the 14 positions in T4 lysozyme that are fully conserved among 5 known bacterio-phae-encoded lysozymes resulted in loss of function (7).

T4 lysozyme has 9 standard α-helices, which account for 59% of its amino acid residues; and 2 other helices as well, described as αII and "distorted α" structures, which we did not evaluate (10, 11). In terms of sensitivity to substitutions, α-helical residues were typical of the protein as a whole; 16% of substitutions were deleterious. Within this group, though, residues designated as belonging to the hydrophobic strip were more sensitive (26%), while non-hydrophobic strip residues were less (13%). Within the hydrophobic strips, the NH2-terminal residues scored 26%, while the COOH-terminal residues scored an extraordinarily high 44%. The smallest residues in the hydrophobic strips scored 31%. The NH2- and COOH-terminal "shoulders" (those strip residues on either side of the smallest) scored 14 and 31%, respectively. The COOH-terminal and virtual COOH-terminal strip positions together were much more sensitive to substitutions than were their NH2-terminal counterparts. Residues in the hydrophobic strip were generally sensitive to substitutions; COOH-terminal strip residues were extremely sensitive.

The character of acceptable substitutions at the hydrophobic strip positions partly mirrored their restricted amino acid compositions. For instance, Leu and Phe substitutions were universally tolerated at the COOH termini of hydrophobic strips (not shown).

An apparent exception to the strip-of-helix template predictive rules actually supported the general theory of the functional importance of that strip in stabilizing the helix, and thus in protein function. In one α-helix, consisting of residues 3–10, the COOH-terminal residue, Asp10, was found to be especially sensitive to substitutions. Asp13 is in a position in which it would have been designated part of the hydrophobic strip if it were not charged. Inspection of the structure in this particular case was instructive; the side chain of Asp13 is solvent-inaccessible and participates in a buried salt bridge.

Sensitivity scores of small groups of amino acid residues could be biased by the limited compositions of the groups. For example, if the Phe residues in T4 lysozyme were especially critical, then a small group that happened to include Phe might be expected on this basis to have a high sensitivity score. To compensate for such a bias, we compared the sensitivity scores of each group identified above with an “expected” score based on the group’s amino acid composition. For example, the group of 9 COOH-terminal hydrophobic strip residues consists of 4 Leu, 2 Phe, 1 Ile, 1 Val, and 1 Trp. The expected sensitivity score of this group was 24%, the highest such value found (Table I). However, the actual score of COOH-terminal hydrophobic strip residues exceeded the expected score by 20. The previously identified groups of functionally critical residues, buried and conserved, exhibited scores 18–25 above their respective expected values. Two groups of residues, corresponding to the smallest residues in the hydrophobic strip, and the virtual extensions of the COOH termini of the hydrophobic strips exhibited scores 13 above their expected values. The latter of these was a special case, for reasons probably having little to do with structural principles; the group included Glu11, the key catalytic residue of T4 lysozyme. Without Glu11, this group’s score would have been only 5 above its expected value. Similarly, the group of smallest residues in the hydrophobic strip was skewed by inclusion of Trp138, which, although large, was counted as a member of this group by virtue of being the only (and thus, coincidentally, the smallest) residue in the hydrophobic strip of its short α-helix. Without Trp138, the score of the group of break residues would have been only 8 above its expected value. All the other groups deviated from their expected scores by 11 or less. A statistical test (described under “Experimental Procedures”; see Table I) indicates that almost all subgroups of the hydrophobic strip residues were significantly more

| Table I |

| Sensitivities of positions in T4 lysozyme to amino acid substitutions |
|---------------------------------------------------------------|
| **Group**          | **No. of residues** | **Substitutions (n)** | **Score** | **Residues represented** | **Expected score** | **Difference p** |
|--------------------|---------------------|-----------------------|-----------|-------------------------|-------------------|-----------------|
| All residues       | 163                 | 2015                  | 16        | All                     | (16)              | 0               |
| >88% buried        | 50                  | 621                   | 38        | ACDGILMNRTVWY            | 20                | 18              | <0.001          |
| 100% buried        | 24                  | 302                   | 42        | ACDGILMNTVY              | 19                | 23              | <0.001          |
| Conserved          | 14                  | 172                   | 47        | ADEGHTVY                 | 22                | 25              | <0.001          |
| α-Helical          | 97                  | 1200                  | 16        | All but H                | 15                | 19              | <0.001          |
| Hydrophobic strip  |                     |                       |           |                         |                   |                 |
| NH2-terminal       | 25                  | 307                   | 26        | ACFIKLVW                 | 19                | 7               | <0.001          |
| COOH-terminal      | 9                   | 110                   | 26        | AIKLW                    | 18                | 8               | <0.05           |
| Smallest           | 9                   | 111                   | 44        | FVLJVW                   | 24                | 20              | <0.0001         |
| NH2 shoulders      | 3                   | 36                    | 14        | All but H                | 16                | 19              | <0.001          |
| COOH shoulders     | 7                   | 86                    | 31        | KL                       | 20                | 16              | <0.01           |
| Virtual NH2-terminal| 8                   | 97                    | 10        | FGLRST                   | 15                | 5               | <0.001          |
| Virtual COOH-terminal| 9                  | 112                   | 31        | EGILRST                  | 18                | 13              | <0.001          |
| Non-strip          | 72                  | 880                   | 13        | All but C and H          | 15                | 2               | <0.05           |
sensitive to substitution than one would have expected based simply on their amino acid compositions. A residue’s designation as a member of any group other than COOH-terminal hydrophobic strip residues, though, was significantly less informative than its solvent accessibility in predicting sensitivity to substitutions.

**DISCUSSION**

The importance of large hydrophobic residues in stabilizing protein structure is generally recognized (see Ref. 18 for a review). It has been confirmed and measured in recent studies of mutant proteins (for examples, see Refs. 19–21). We therefore expected and found that residues identified as parts of the hydrophobic strips of α-helices are, on the average, more sensitive to substitutions than others in the protein. Indeed, the pattern of sensitivity to substitutions of residues in a protein appears to be a useful predictor of secondary structure (19).

The unexpected result of our study was the identification of a set of especially critical, hydrophobic amino acid residues in T4 lysozyme: COOH-terminal hydrophobic strip residues in α-helices. A number of substitutions for any of the 9 members of this group resulted in loss of function. The minimum number of deleterious substitutions (out of 12 or 13 tested at each position) in this group was 4 (see Table II). The significance of this group of residues is apparent when one considers that only 74 of the 163 tested positions in T4 lysozyme were sensitive to any substitutions; only 43 were sensitive to 4 or more. Thus, the probability of picking, at random, a group of 9 residues in which all are sensitive to 4 or more substitutions is approximately 0.000006. The apparent functional significance of residues identified in this manner equals or exceeds that of residues identified on the basis of solvent inaccessibility and approaches that of the set of most highly conserved residues.

This test of the relative functional importance of groups of amino acid residues in T4 lysozyme was nearly bias-free. The same set of 13 amino acids was tested at essentially every position. Of course, a consequence of the use of 13 particular amino acids was that at some positions (59), 13 substitutions were tested, while at others (104), 12 were tested. A perfect test would examine the effects of placing all 20 amino acids at each position. On the other hand, the set of 13 amino acids represented in the collection of amber suppressors employed (Gly, Ala, Ser, Cys, Pro, Leu, Gln, Glu, His, Phe, Tyr, Lys, and Arg) contained members of most of the generally recognized classes of amino acids. The use of amber suppression as the means of effecting single amino acid substitutions introduces complexities into the interpretation of the data. Amber suppressors differ in efficiency, and the efficiency of suppression varies among amber (UAG) codons due to what are called context effects. On the other hand, all amber mutants in this particular collection formed large, healthy plaques on at least 1 amber suppressor strain (on several, in almost all cases); conversely, all of the amber suppressor strains plated many of the amber mutants well. The interpretation of amber suppression patterns was discussed in greater detail by Rennell et al. (3).

Examination of the T4 lysozyme molecule reveals a structural correlate of the functional importance of COOH-terminal hydrophobic strip residues. These residues are more tightly packed than NH₂-terminal hydrophobic strip residues. The average side-chain solvent inaccessibility of COOH-terminal hydrophobic strip residues is 99%; that of NH₂-terminal hydrophobic strip residues is 79%. Moreover, the average temperature factors of atoms in these two groups of residues are 15.6 and 20.3, respectively. These observations are based on 7 of T4 lysozyme’s 9 α-helices. The other 2 are omitted because one (residues 93–106) runs through the interior of the large domain of T4 lysozyme and is thus not a surface α-helix, while the other (residues 137–141) has only 1 residue that fits the description of a hydrophobic strip residue (as defined by either the algorithm or structural criteria), and so its NH₂-terminal and COOH-terminal hydrophobic strip residues are the same.

The foregoing observations point to a feature in the structural organization of T4 lysozyme; surface α-helices are more tightly packed into the hydrophobic core at their COOH-terminal ends than at their NH₂-terminal ends. We hypothesize that this structural feature accounts for the high sensitivity of T4 lysozyme function to substitutions specifically at the COOH-terminal ends of the hydrophobic strips. The observation of such a structural feature in T4 lysozyme raises the question of whether other proteins are similarly organized. A small-scale survey suggests that this polarity in the anchoring of α-helices is not general. We compared the average temperature factors of atoms in the NH₂ and COOH termini of the hydrophobic strip residues of 43 α-helices in 5 proteins (thermolysin, sperm whale myoglobin, erythrocrucorin, carboxypeptidase A, and beef catalase). None of these proteins exhibited as great a polarity as T4 lysozyme in this measure; in 4 cases out of 5, the COOH-terminal residues had slightly higher average temperature factors than the NH₂-terminal residues (data not shown).

Rules for picking out critical residues are not completely reliable. Many substitutions are tolerated at buried and conserved positions, as well as at COOH-terminal hydrophobic strip positions. When examined in detail, proteins with unexpectedly tolerated substitutions turn out to have altered their structures in subtle ways, creating a context into which the altered residue fits (for a review, see Matthews (22)). It is apparent that the acceptability of amino acid substitutions is governed by details of structural context which cannot be captured in a simple scalar quantity like solvent accessibility. The foregoing considerations may limit, but do not rule out, the utility of simple formulas in identifying key structural determinants of function. Even so, using a structure-predicting algorithm is perhaps not an obvious strategy for studying a protein for which a high resolution, crystallographically determined structure is available. However, at least in this case, the algorithm was helpful in picking out a pattern (a class of critical residues) that the investigators did not notice by studying the structure in its full complexity. The usefulness of certain simplifying schemes, such as topology diagrams, in picking out structural relationships is well established (23).

This analysis introduces potentially severe restraints on the engineering of therapeutic proteins to decrease their immunogenicity (24) or of vaccines to increase potency or broaden the range of major histocompatibility complex restriction (25). A scavenger site (S site) motif identified with the strip-of-helix hydrophobicity template (26) may lie in or

**Table II**

Groups of functionally critical residues in T4 lysozyme

| Fully conserved residues (number of deleterious substitutions) |
|---------------------------------------------------------------|
| Glu¹ (12) Gly¹⁻² (12) Tyr¹⁻³ (3) Thr⁻³ (9) Gly¹⁻¹ (11) His¹⁻¹ (4) |
| Thr⁻¹⁻² (3) Asp⁻³⁻⁶ (5) Ala⁻³⁻⁶ (8) Leu⁻³⁻⁶ (3) Gly¹⁻⁷ (3) Leu¹⁻¹⁷ (3) |
| Ala¹⁻⁷ (5) Trp¹⁻¹⁰ (10) |
| COOH-terminal hydrophobic strip |
| Leu¹⁻³ (5) Leu¹⁻⁶ (6) Ile¹⁻³ (5) Val¹⁻⁴ (4) Phe²⁻⁴ (4) Leu¹⁻¹ (5) |
| Leu¹⁻¹ (6) Trp¹⁻¹⁰ (10) Phe²⁻⁴ (4) |

*Data from Rennell et al. (5).
near the T cell-presented epitope (T site) which may be identified by allele-specific motifs (27–31). Decreasing the hydrophobicity of residues in positions of the longitudinal hydrophobic strip may decrease scavenging or processing of an excised fragment bearing a T site. Likewise, increasing the hydrophobicity of such residues in engineered immunogens may increase potency and possibly broaden the apparent range of major histocompatibility complex restriction, if priming of low responders leads to a memory cell population capable of clinical protection upon later challenge with wild type antigen (25). Our current study shows, however, that the residues one would change to alter potency of T cell presentation of an S site-associated T site are the very residues which are required to maintain structure for the function of the protein. In the case of a vaccine, such loss of function may be irrelevant. But in the case of a therapeutic protein (enzyme, lymphokine, hormone), loss of T cell immunogenicity carries a greater likelihood to alter local structure in protein folding. To attack this issue, toward the engineering of still functional proteins with aborted T cell immunogenicity, one would want to discriminate biophysical differences between folding and docking of helices in nascent proteins, and the scavenging of excised fragments during antigen processing.

Acknowledgments—We thank Dr. Larry Hardy for discussions and for reading the manuscript and Marion Bonin for typing the paper.

REFERENCES
1. Perutz, M. F., Kendrew, J. C., and Watson, H. C. (1965) J. Mol. Biol. 13, 669-678
2. Reidhaar-Olson, J. F., and Sauer, R. T. (1988) Science 241, 53-57
3. Rennell, D., Bouvier, S. E., Hardy, L. W., and Potente, A. R. (1991) J. Mol. Biol. 222, 67-88
4. Loeb, D. D., Swanstrom, R., Evertt, L., Manchester, M., Stamper, S. E., and Hutchison, C. A. (1989) Nature 340, 397-400
5. Kleina, L. G., and Miller, J. H. (1990) J. Mol. Biol. 212, 295-318
6. Lin, W. A., and Sauer, R. T. (1988) Nature 339, 51-36
7. Potente, A. R., Rennell, D., and Bouvier, S. E. (1992) Proteins Struct. Funct. Genet. 13, 36-40
8. Torgerson, R. R., Lew, R. A., Reyes, V. E., Hardy, L., and Humphreys, R. E. (1991) J. Biol. Chem. 266, 5021-5024
9. Vanperr, S. R., Kuo, D. Z., Bositis, C. M., Hardy, L. W., Lew, R. A., and Humphreys, R. E. (1992) J. Biol. Chem. 267, 7406-7410
10. Remington, S. J., Anderson, W. F., Owen, J., Ten Eyck, L. F., Grainger, C. T., and Matthews, B. W. (1979) J. Mol. Biol. 118, 81-98
11. Weaver, L. H., and Matthews, B. W. (1978) J. Mol. Biol. 193, 189-199
12. Winnnot, F., Botstein, D., and Miller, J. H. (1979) J. Bacteriol. 137, 433-439
13. Normanly, J., Masson, J.-M., Kleina, L. G., Abelos, J., and Miller, J. H. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 6540-6552
14. Normanly, J., Kleina, L., Masson, J.-M., Abelos, J., and Miller, J. H. (1990) J. Mol. Biol. 213, 719-728
15. Kleina, L. G., and Miller, J. H. (1990) J. Mol. Biol. 212, 295-318
16. McClain, W. H., and Foss, H. (1988) Science 241, 1864-1870
17. Hardy, L. W., and Potente, A. R. (1991) Biochemistry 30, 9467-9463
18. Rose, D. G., Geshelowitz, A. R., Lessor, G. J., Lee, R. H., and Zehfus, M. (1985) Science 229, 834-838
19. Bowie, J. U., and Sauer, R. T. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 2152-2156
20. Matsumura, M., Becktel, W. J., and Matthews, B. W. (1988) Nature 334, 406-410
21. Shortle, D., Stites, W. E., and Meeker, A. K. (1990) Biochemistry 29, 8033-8041
22. Matthews, B. W. (1987) Biochemistry 26, 6885-6888
23. Branden, C., and Tooze, J. (1991) Introduction to Protein Structure, pp. 20-21, Garland Publishing, Inc., New York
24. Lu, S., Reyes, V. E., Torgerson, R. R., Lew, R. A., and Humphreys, R. E. (1991) Trends Biochem. 9, 238-242
25. Lu, S., Reyes, V. E., Bositis, C. M., Goldschmidt, T. G., Lam, V., Torgerson, R. R., Corradin, G., Leuscher, I., and Maryanski, J. L. (1991) J. Immunol. 146, 1230-1240
26. Hill, C., Hayball, J. D., Allison, A. A., and Rothbard, J. B. (1991) J. Immunol. 147, 186-197
27. Largent, J., Sette, A., Sidney, J., Appella, E., Ehrhardt, C., Fuchs, S., and Adorini, L. (1991) J. Immunol. 147, 198-204