Highly Restricted Distributions of Hydrophobic and Charged Amino Acids in Longitudinal Quadrants of α-Helices*

(Received for publication, October 22, 1990)

Rochelle R. Torgerson§, Robert A. Lew‡, Victor E. Reyes§§, Larry Hardy¶¶, and Robert E. Humphreys**‡‡

From the Departments of §Pharmacology and **Medicine and the §Program in Molecular Medicine, University of Massachusetts Medical Center, Worcester, Massachusetts 01655

Helix formation in folding proteins is stabilized by binding of recurrent hydrophobic side chains in one longitudinal quadrant against the locally most hydrophobic region of the protein. To test this hypothesis, we fitted sequences of 247 α-helices of 55 proteins to the circular (infinite) template [OOAOOOOOA] to maximize the strip-of-helix hydrophobicity index (the mean hydrophobicity of residues in [ ] positions). These template-predicted configurations closely matched crystallographic structures in 87% of four- or five-turn helices compared. We determined the longitudinal quadrant distributions of amino acids in the template-fitted, sheet projections of α-helices with respect to the best longitudinal, hydrophobic strip on each helix and to the N and C termini, interiors, and entire helices. Amino acids Leu, Ile, Val, and Phe were concentrated in one longitudinal quadrant (p < 0.001). Lys, Arg, Asp, and Glu were not in the quadrant of Leu, Ile, Val, and Phe (p < 0.001). Significant quadrant distributions for other amino acids and for termini of the helices were also found.

α-Helices in proteins are predicted with sensitivity and efficiency from runs of aliphatic, hydrophobic amino acids only at recurrent positions (n, n + 4, n + 7, n + 11, n + 14, etc.) that form an axial hydrophobic strip when the sequence is coiled as an α-helix (1). Perutz et al. (2) originally observed in the α-helices of hemoglobin the recurrence of invariant, nonpolar residues every 36 residues, on the average, making the interior faces of the helices nonpolar. Schiffer and Edmundson (3) created the wheel projection to identify such segments with helical potential. Eisenberg et al. (4) and Finer-Moore and Stroud (5) developed methods based upon amphipathic moments to predict α-helices. Kaiser and Taylor (6) tested the function of a longitudinal hydrophobic surface on α-helices to promote folding against a hydrophobic surface.

* This work was supported by American Cancer Society Grant IM-582 (to R. E. H.) and by the Scientific Council at the University of Massachusetts Medical Center 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 1752 solely to indicate this fact.

§ Student from Williams College, Williamstown, MA and Fellow of the Juvenile Diabetes Foundation International Summer Research Program at the University of Massachusetts Medical Center.

† Fellow of the National Institutes of Health (Immunovirology Training Grant T32 AI-07272).

‡‡ 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: 508-856-5080.

** The abbreviation used is: SOHHI, strip-of-helix hydrophobicity index.
We compared the observed set of frequencies for the four quadrants to the null distribution by means of the $\chi^2$ goodness-of-fit test on three degrees of freedom and verified the results with the likelihood ratio test (14). The null probability distribution assigned $3/18$ to quadrant I and $5/18$ to quadrants II, III, and IV. To indicate possible false rejection of the null hypothesis, we distinguished results with $p$ values of 0.05, 0.01, and 0.001. With $p = 0.001$ and 50 independent tests, a type I error occurs 5% of the time using a Bonferroni correction for multiple comparisons. For each quadrant within a statistically significant distribution of frequencies, we standardized the deviation from the expected proportion, $p$, using the quantity $(\text{observed frequency} - \hat{p})/\text{S.E.}$, where S.E. was the standard error under the binomial model, $\sqrt{p(1-p)/n}$, and $n$ is the number of times the amino acid occurred. We used the standardization because the difference in proportions can be misleading. For example, if one amino acid appeared in quadrant I 20 out of 40 times while a rarer amino acid appeared in quadrant I 2 out of 4 times, both would have observed deviations of $50 - 17 = 33\%$.

### Table I

**Standardized deviation from expected frequency**

We compared the observed set of frequencies for the four quadrants to the null distribution by means of the $\chi^2$ goodness-of-fit test on three degrees of freedom and verified the results with the likelihood ratio test (14). The null probability distribution assigned $3/18$ to quadrant I and $5/18$ to quadrants II, III, and IV. To indicate possible false rejection of the null hypothesis, we distinguished results with $p$ values of 0.05, 0.01, and 0.001. With $p = 0.001$ and 50 independent tests, a type I error occurs 5% of the time using a Bonferroni correction for multiple comparisons. For each quadrant within a statistically significant distribution of frequencies, we standardized the deviation from the expected proportion, $p$, using the quantity $(\text{observed frequency} - \hat{p})/\text{S.E.}$, where S.E. was the standard error under the binomial model, $\sqrt{p(1-p)/n}$, and $n$ is the number of times the amino acid occurred. We used the standardization because the difference in proportions can be misleading. For example, if one amino acid appeared in quadrant I 20 out of 40 times while a rarer amino acid appeared in quadrant I 2 out of 4 times, both would have observed deviations of $50 - 17 = 33\%$.

| $p = 0.001$ | Asp | Leu | Ile | Leu | Val | Ile | Phe |
|-------------|-----|-----|-----|-----|-----|-----|-----|
| N terminus  |     |     |     |     |     |     |     |
| I           | $+8.5$ | $-2.1$ | $-2.2$ | $-1.9$ | $-0.5$ | $-2.3$ | $-0.8$ |
| II          | $+0.3$ | $-3.2$ | $-0.4$ | $-1.3$ | $-0.8$ | $-1.4$ | $-1.8$ |
| III         | $-3.2$ | $+6.2$ | $+4.9$ | $+4.5$ | $+4.0$ | $+5.6$ | $+4.7$ |
| IV          | $+0.1$ | $-1.2$ | $-2.7$ | $-1.7$ | $-2.8$ | $-2.3$ | $-2.3$ |

| $p = 0.01$ | Gln | Glu | Lys | Phe |
|-------------|-----|-----|-----|-----|
| Interior    |     |     |     |     |
| I           | $+1.0$ | $+0.1$ | $-0.2$ | $-1.4$ |
| II          | $+1.9$ | $+2.1$ | $+0.2$ | $-2.0$ |
| III         | $-2.3$ | $-2.9$ | $-2.8$ | $+2.3$ |
| IV          | $+1.3$ | $+0.7$ | $+2.7$ | $+0.9$ |

| $p = 0.05$ | Gln | Glu | Lys | Phe |
|-------------|-----|-----|-----|-----|
| C terminus  |     |     |     |     |
| I           | $+0.9$ | $+2.6$ | $+1.5$ | $-0.5$ |
| II          | $-0.9$ | $-0.5$ | $+1.1$ | $+0.2$ |
| III         | $-2.4$ | $-2.3$ | $-2.8$ | $-2.6$ |
| IV          | $+2.5$ | $+0.7$ | $+0.6$ | $+2.9$ |

| $p = 0.001$ | Asp | Leu | Ile | Leu | Val | Ile | Phe |
|-------------|-----|-----|-----|-----|-----|-----|-----|
| Entire helix|     |     |     |     |     |     |     |
| I           | $+5.5$ | $+0.7$ | $+1.4$ | $+0.4$ | $-3.9$ | $-2.2$ | $+0.8$ |
| II          | $+1.2$ | $+1.9$ | $+1.1$ | $+0.6$ | $-3.4$ | $-2.9$ | $+3.0$ |
| III         | $-4.8$ | $-4.4$ | $-5.1$ | $-3.8$ | $+7.7$ | $+6.1$ | $-5.5$ |
| IV          | $-1.0$ | $+1.9$ | $+2.9$ | $+2.8$ | $-1.1$ | $-1.4$ | $+1.8$ |
second and fourth quadrants; two successive C's in the template were associated with an empty first quadrant. The distributions of individual amino acids in each quadrant were determined for N and C termini, for the interior, and for the entire helix. In accord with Presta and Rose (8), the left terminus was defined as the first 4 amino acids of a helix and the right terminus as the last 4 amino acids of a helix. The amino acids between the termini constituted the interior of a helix. The template placed the amino acids in four quadrants of the sheet projection with 3, 5, and 4 symbols in quadrants I, II, III, and IV of the projection, respectively, so that the null hypothesis distribution assigned a probability of 3/18 for an amino acid being in quadrant I and of 5/18 for being in quadrants II, III, or IV. Since each terminus had four amino acids, only segments with 9 or more amino acids had nonempty interiors. Segments with 3 or fewer amino acids were excluded from analysis. For segments with 4-7 amino acids, the N and C termini overlapped. We counted the frequency of amino acids predicted to fall in the axial hydrophobic strip of quadrants I, II, III, and IV. The maximal sector angle was the absolute value of the greatest angle among those radii. The mean sector angle was the average of the absolute values for the angles between the most clockwise radius and the other radii. The maximal sector angle was the absolute value of the greatest angle among those radii. The mean sector angle was the average of the absolute values for the angles between the most clockwise radius and the other radii.

RESULTS

Predicted quadrant orientations approximated crystallographic structures well. The structural relevance of our template-fitting model of α-helices was tested by direct examination of all four- or five-turn helices in 7CAT, 5CPA, 2CYP, 4LDH, 2MBN, 1MBO, and 2SNS for alignment of the residues predicted to fall in the axial hydrophobic strip of quadrant III (8, 9). Projections of x-ray crystallographic coordinates for 22 of 28 helices demonstrated a maximal sector angle among residues in the axial hydrophobic strip of 92° and a mean sector angle of 61° (Table II). We conclude that our assignments of amino acid residues to four quadrants, based on positioning of recurrent hydrophobic residues in one

### Table I—continued

| p = 0.001 | Ile | Phe |
|-----------|-----|-----|
| I         | -3.1 | -2.4 |
| II        | -2.3 | -2.2 |
| III       | +7.8 | +5.9 |
| IV        | -2.8 | -1.6 |

| p = 0.01 | Asn | Gln |
|-----------|-----|-----|
| I         | +2.5 | +1.2 |
| II        | +0.5 | +1.5 |
| III       | -3.7 | -4.0 |
| IV        | +1.1 | +1.5 |

| p = 0.05 | Thr | Tyr |
|-----------|-----|-----|
| I         | +1.9 | +0.1 |
| II        | +1.4 | +1.4 |
| III       | -2.9 | -3.2 |
| IV        | -0.1 | +1.7 |

### Table II

| Sectors of residues in four- or five-turn, axial hydrophobic strips |
|---------------------------------------------------------------|
|                  | Maximum sector | Mean sector |
| Presta and Rose helices |
| 5CPA              | 15-28          | 66 53       |
| 74-89             | 77 43          |
| 216-230           | 77 57          |
| 1MBO              | 4-17           | 79 63       |
| 21-35             | 70 48          |
| 59-76             | 90 56          |
| 83-95             | 67 52          |
| 101-118           | 63 38          |
| Richardson and Richardson helices |
| 7CAT              | 53-67          | 124 101     |
| 258-271           | 65 42          |
| 437-450           | 149 120        |
| 470-485           | 99 62          |
| 485-500           | 59 55          |
| 5CPA              | 14-30          | 94 62       |
| 173-187           | 78 53          |
| 2CYP              | 42-55          | 63 51       |
| 103-120           | 180 94         |
| 164-177           | 75 69          |
| 4LDH              | 30-44          | 119 74      |
| 55-70             | 73 61          |
| 141-154           | 94 53          |
| 165-181           | 69 48          |
| 247-264           | 72 56          |
| 2MBN              | 20-37          | 120 65      |
| 82-98             | 137 95         |
| 100-116           | 65 42          |
| 2SNS              | 54-69          | 52 31       |
| 121-136           | 72 60          |

### Table III

| Termini of helices |
|---------------------|
| Beginnings          |
| I                   | -3.5       |
| II                  | -3.2       |
| III                 | -0.9       |
| IV                  | +7.0       |
| Endings             |
| I                   | +7.5       |
| II                  | +1.5       |
| III                 | -2.4       |
| IV                  | -5.3       |

The standardized deviations of the frequencies of first and last residues of the helices are presented for each axial quadrant of their appearances, p = 0.001.
axial strip to maximize the SOHHI, closely matched crystallographic measurements.

The hydrophobic amino acids Leu, Ile, Val, and Phe each occurred more frequently in quadrant III than in other quadrants \((p < 0.001)\). The absence of these residues from quadrants I, II, and IV was as remarkable as their presence in quadrant III, as reflected in the exceptionally large standardized deviations. This finding is consistent with the hypothesis that recurrent placement of such amino acids only in quadrant III stabilizes \(\alpha\)-helical coiling in nascent proteins and that the cooperativity of binding the longitudinal hydrophobic strip (given a hydrophobic region in the protein against which to fold) governs the formation of local structure as a helix. Hydrophobic residues Leu, Ile, Val, and Phe in other quadrants of a putative helix would presumably compete for \(\alpha\)-helix formation. Only a narrow axial hydrophobic strip appears to lead to helix formation.

The axial distributions of other amino acids were determined. The four charged amino acids Lys, Arg, Asp, and Glu were uniformly absent from the axial hydrophobic strip. Asn and Gln occurred less frequently in the axial hydrophobic strip. At the N and C termini of helices, Cys was more frequently in the axial hydrophobic strip. Averaged over entire helices, Cys was found more often in the axial hydrophobic strip. Other amino acids, including Trp and Tyr specifically, showed no axial preferences.

The quadrants for the terminations of helices are presented in Table III. A helix was more likely to start in quadrant IV with 3 residues preceding the hydrophobic one in quadrant III. The N terminus was more likely to be an untethered loop. The last amino acid was more likely to be in quadrant I, two amino acids after the hydrophobic axial strip, and the helix was less likely to end in the hydrophobic strip.

**DISCUSSION**

We conclude that the template with the highest SOHHI predicts the quadrant orientations of amino acids in most \(\alpha\)-helices well and that Leu, Ile, Val, and Phe occur almost solely within the axial hydrophobic strip. The predominance of Leu, Ile, Val, and Phe in one longitudinal, hydrophobic strip in \(\alpha\)-helices is as striking as their absence from the other quadrants. This distribution supports the hypothesis that helical coiling is determined by stabilization of the helix against a hydrophobic region by recurrent hydrophobic residues in an axial strip. Presumably, the presence of hydrophobic residues in other axial quadrants can compete for helix formation, for example by forming \(\beta\)-pleated sheets if strictly alternating hydrophobicity is present in a segment facing a locally hydrophobic region of a nascent protein. Bowie et al. (10) and Matouschek et al. (11) have shown that hydrophobic residues along one side of a helix can insert in a hydrophobic core to stabilize protein structure.

The principles underlying both this study and a related analysis of coiling of a series of synthetic peptides that varied in the strengths of their axial hydrophobic strips (7) are applicable to the design of functional proteins and to the prediction of protein structure. Utilization of these principles might also aid the engineering of T cell-presented sequences either to decrease immunogenicity in proteins administered for diagnostic or therapeutic purposes or to increase immunogenicity and broaden major histocompatibility complex range for vaccine presentation (12, 13).

**Acknowledgment**—We thank Marion Bonin for typing the manuscript.

**REFERENCES**

1. Reyes, V. E., Phillips, L., Humphreys, R. E., and Lew, R. A. (1989) *J. Biol. Chem.* 264, 12854-12858
2. Perutz, M. F., Kendrew, J. C., and Watson, H. C. (1965) *J. Mol. Biol.* 13, 669-678
3. Schiffer, M., and Edmundson, A. B. (1967) *Biophys. J.* 7, 121-135
4. Eisenberg, D., Weiss, R. M., and Terwilliger, T. C. (1984) *Proc. Natl. Acad. Sci. U. S. A.* 81, 140-144
5. Finer-Moore, J., and Stroud, R. M. (1984) *Proc. Natl. Acad. Sci. U. S. A.* 81, 155-159
6. Kaiser, R. T., and Taylor, J. W. (1987) *Methods Enzymol.* 154, 473-487
7. Lu, S., Reyes, V. E., Lew, R. A., Anderson, J., Mole, J., Humphreys, R. E., and Ciardelli, T. (1990) *J. Immunol.* 145, 899-904
8. Presta, L. G., and Rose, G. D. (1988) *Science* 240, 1632-1641
9. Richardson, J. S., and Richardson, D. C. (1988) *Science* 240, 1648-1652
10. Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A., and Sauer, R. T. (1990) *Science* 247, 1306-1310
11. Matouschek, A., Kellis, J. T., Serrano, L., Bycroft, M., and Fersht, A. R. (1990) *Nature* 346, 440-445
12. Reyes, V. E., Fowlie, E. J., Lu, S., Phillips, L., Chin, L. T., Humphreys, R. E., and Lew, R. A. (1990) *Mol. Immunol.* 27, 1021-1027
13. Goldschmidt, T. G., Lu, S., Reyes, V. E., Lam, V., Torgerson, R. R., Nguyen, Q. V., Lew, R. A., and Humphreys, R. E. (1991) *The Vaccines* (Brown F., Chanock, R., Ginsburg, H. S., and Lerner, R., eds), Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, in press
14. Kendall, M., and Stuart, G. A. (1961) *The Advanced Theory of Statistics*, Vol. III, pp. 224-261, Griffin, London