*ANALYSIS OF THE SKELETAL CONFIGURATION OF CRYSTALLINE HEN EGG-WHITE LYSOZYME**

By David A. Brant and Paul R. Schimmel†

DEPARTMENT OF CHEMISTRY, UNIVERSITY OF CALIFORNIA (IRVINE), AND
DEPARTMENT OF CHEMISTRY, STANFORD UNIVERSITY

Communicated by Paul J. Flory, June 19, 1967

The configuration of the polypeptide chain with all amide groups in the planar *trans* configuration is given by specifying every dihedral angle φ_i,ψ_i defined in Figure 1,[1] whereas the conformational energy of the *random coil* can be properly resolved into contributions each of which depends upon a *single* pair of rotation angles φ_i,ψ_i for residue i.[2] Estimates of this energy for glycine, alanine, and proline residues situated within a polypeptide random coil are reported elsewhere.[2-8] Such estimates take account of bond torsional strain, van der Waals repulsions, London attractions, and electrostatic interactions between nonbonded atoms and groups separated by a single pair of angles φ_i,ψ_i.[2-8] These energy functions have been used in conjunction with statistical mechanical methods to predict successfully values of the dimensions[2, 6, 9] and of the molecular dipole moment[10] for *unperturbed* polypeptide chains of varying amino acid sequence and composition.

The energy of a residue situated within a rigid helical sequence, where each φ_i,ψ_i pair is identical, has been similarly calculated.[4, 5, 11-13] In addition to the contributions enumerated above, the helix energy includes hydrogen bond interactions, amide-amide dipolar interactions, and steric interactions between residues separated by *more than one* φ_i,ψ_i pair of rotation angles.[13] Calculations for homopolymers of L-amino acids possessing a variety of side chains reveal that for most such polymers the commonly observed *right-handed* Pauling-Corey α-helix is the *helical* configuration of lowest energy.[4, 5, 11-13]

We here compare the residue dihedral angles recently determined for crystalline lysozyme[14] with expectations based upon the above-mentioned conformational energy estimates, duly recognizing that such estimates are not strictly applicable

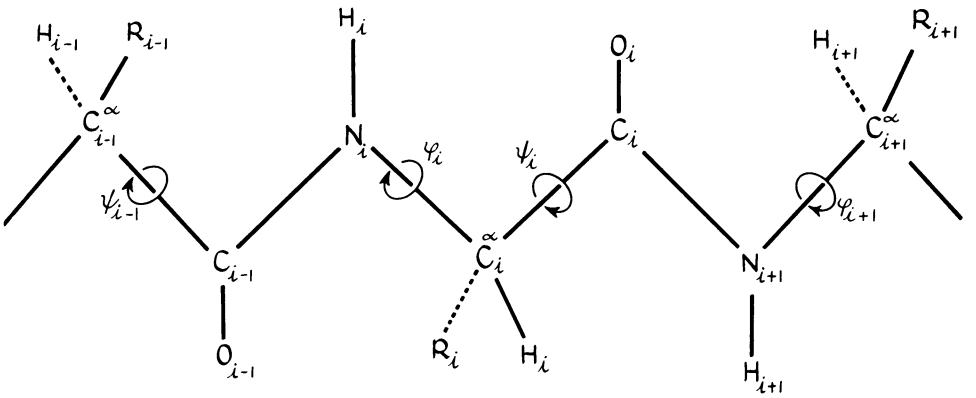

FIG. 1.—Schematic portion of the α-L-polypeptide chain in its fully extended conformation for which the dihedral rotation angles φ and ψ are zero. Positive values of φ and ψ result for right-handed rotations in accordance with conventions established in ref. 1. All amide groups are assigned to the planar *trans* conformation and subscripts denote serial order of amino acid residues.

to a nearly space-filling protein chain which is packed into an ordered crystal. Residues which occur in helical sequences are plotted on a helix map (cf. *seq.*), while those occurring in nonhelical (irregular) sequences are plotted on the random coil map appropriate for the given residue. A large proportion of the residues are found to fall close to or within a calculated low energy contour, and the relative population of residues in a given sterically allowed region qualitatively conforms with that expected from energy estimates. Deviations in the dihedral angles from the predicted values reflect on the abundance and magnitude of the unaccounted interactions which occur in the protein, and possibly also on kinetic limitations to the attainment of the equilibrium state.

Results.—Residues 2 through 128 of hen egg-white lysozyme are listed in Table 1 along with the values of the dihedral angles given by Phillips[14] for each residue. The uncertainties in the dihedral angles have not been determined. Each residue (with exceptions noted below) has been plotted on one of the energy contour diagrams in Figures 2–4. The diagram in Figure 2 is based on calculations for a residue bearing a $-CH_2-R'$ side chain at the alpha carbon, this residue being situated within a random coil polypeptide.[6] This map gives the total residue energy calculated from interactions between atoms whose distance of separation depends *exclusively* on the single pair φ,ψ of rotation angles associated with the residue. All residues with side chains of the type $-CH_2-R'$ and those with branching at the β-carbon, i.e., valine and isoleucine, which do not occur within a helical sequence of at least five residues are plotted on the diagram in Figure 2. (The effect of β-branching is to narrow somewhat the allowed areas on this conformational energy map.[15]) Residues falling within regions I, II, or III of Figure 2 are so labeled in Table 1, while those in the saddle between minima I and III or in high energy regions are labeled *S* or *H*, respectively.

Glycine residues exclusive of those in a helical sequence of at least five residues are plotted on the similarly calculated glycine map[6] in Figure 3. Residues falling within minima of Figure 3 are designated *G* in the last column of Table 1, while residues occurring in the saddles of somewhat higher energy or in high energy regions are indicated by *S* or *H*, respectively.

Six regular (helix-like) sequences can be identified, each of which comprises five or more residues in closely similar conformations. Each of these corresponds to a configuration approximating the Pauling-Corey right-handed α-helix; the participating residues are designated α_R in Table 1. Residues within such helical sequences are plotted in Figure 4 on a helix energy map calculated for polyalanine.[13] Each helix-like sequence is identified by a distinctive symbol in Figure 4. The energy plotted represents the conformation energy of an interior residue in a long helical sequence.[13] Applicability of this diagram to the residues of lysozyme can be questioned on several grounds: the helices are indicated to be imperfect;[14] some residues are near the ends of helical sequences where their intrahelical interactions are not fully realized; some of the side chains are not of the kind $-CH_2-R'$ to which the calculation are most clearly applicable. However, refinement of the crystallographic analysis may show these sequences to be more regular than is presently suggested.[14] Moreover, the general features of Figure 4 are also shown on helix maps for polypeptides possessing a variety of side chains and for other estimates of the conformational energy.[4, 5, 11, 12]

TABLE 1

Residue no.	Amino acid[19]	ϕ	ψ	Conformation	Residue no.	Amino acid[19]	ϕ	ψ	Conformation
2	Val	56.0	301.4	III	66	Asp	35.0	185.2	*H*
3	Phe	106.9	337.9	III	67	Gly	250.5	179.7	*H*
4	Gly	78.9	336.0	*G*	68	Arg	45.5	200.1	*S*
5	Arg	132.7	114.5	α_R	69	Thr	59.8	269.5	—
6	Cys	142.5	116.4	α_R	70	Pro	132.0	141.4	—
7	Glu	126.8	153.4	α_R	71	Gly	98.8	171.4	*S*
8	Leu	119.1	121.4	α_R	72	Ser	127.6	313.3	III
9	Ala	126.2	129.9	α_R	73	Arg	50.8	313.8	III
10	Ala	115.1	140.4	α_R	74	Asn	60.9	261.6	III
11	Ala	140.9	120.9	α_R	75	Leu	124.4	135.1	I
12	Met	109.0	152.9	α_R	76	Cys	81.1	200.6	*S*
13	Lys	123.3	114.1	α_R	77	Asn	217.5	223.1	II
14	Arg	142.3	105.4	α_R	78	Ileu	45.3	321.4	—
15	His	116.9	176.4	α_R	79	Pro	101.9	309.4	—
16	Gly	277.1	171.6	*S*	80	Cys	165.7	117.1	α_R
17	Leu	127.8	145.9	I	81	Ser	133.5	141.0	α_R
18	Asp	104.7	330.3	III	82	Ala	127.8	137.6	α_R
19	Asn	241.4	182.4	*H*	83	Leu	127.7	140.7	α_R
20	Tyr	103.4	318.6	III	84	Leu	95.8	182.5	α_R
21	Arg	241.5	219.8	II	85	Ser	119.6	326.1	III
22	Gly	239.6	202.2	*G*	86	Ser	90.8	186.7	*S*
23	Tyr	84.7	298.1	III	87	Asp	51.9	312.3	III
24	Ser	96.9	352.2	III	88	Ileu	110.5	176.7	α_R
25	Leu	149.4	131.3	α_R	89	Thr	144.2	122.0	α_R
26	Gly	123.4	131.7	α_R	90	Ala	149.1	100.1	α_R
27	Asn	127.9	125.7	α_R	91	Ser	125.9	120.2	α_R
28	Try	120.0	138.4	α_R	92	Val	124.9	134.0	α_R
29	Val	102.1	131.5	α_R	93	Asn	113.6	150.0	α_R
30	Cys	99.3	147.0	α_R	94	Cys	108.6	128.9	α_R
31	Ala	124.7	111.9	α_R	95	Ala	128.0	141.4	α_R
32	Ala	115.4	150.0	α_R	96	Lys	117.1	128.5	α_R
33	Lys	115.2	128.8	α_R	97	Lys	104.7	162.1	α_R
34	Phe	108.2	150.6	α_R	98	Ileu	91.4	141.2	α_R
35	Glu	97.6	131.0	α_R	99	Val	102.8	169.3	α_R
36	Ser	76.7	156.1	α_R	100	Ser	66.0	184.3	*H*
37	Asn	250.9	181.4	*H*	101	Asp	97.8	123.5	I
38	Phe	292.4	153.4	*H*	102	Gly	3.5	149.2	*H*
39	Asn	95.6	292.2	III	103	Asp	84.6	181.0	*S*
40	Thr	109.8	165.3	I	104	Gly	233.8	36.7	*G*
41	Gln	93.6	138.8	I	105	Met	102.6	187.8	*H*
42	Ala	143.0	330.1	III	106	Asn	107.7	154.3	I
43	Thr	36.1	329.8	III	107	Ala	113.3	160.7	I
44	Asn	46.2	284.8	III	108	Try	71.8	257.5	III
45	Arg	76.4	315.2	III	109	Val	149.9	126.4	α_R
46	Asn	90.1	2.8	III	110	Ala	117.7	140.7	α_R
47	Thr	111.3	161.3	I	111	Try	124.0	127.2	α_R
48	Asp	86.9	216.9	III	112	Arg	123.1	112.9	α_R
49	Gly	276.5	107.2	*G*	113	Asn	122.3	134.6	α_R
50	Ser	184.9	291.2	*H*	114	Arg	76.7	163.3	α_R
51	Thr	35.7	331.9	III	115	Cys	63.6	135.6	I
52	Asp	69.0	302.6	III	116	Lys	127.4	285.1	III
53	Tyr	65.0	314.0	III	117	Gly	275.9	201.6	*G*
54	Gly	255.2	352.2	*G*	118	Thr	65.4	349.6	III
55	Ileu	133.7	144.6	I	119	Asp	92.8	262.5	III
56	Leu	71.5	189.9	*S*	120	Val	116.0	156.6	α_R
57	Gln	232.6	228.1	II	121	Gln	125.2	142.4	α_R
58	Ileu	114.9	317.7	III	122	Ala	121.2	160.6	α_R
59	Asn	92.2	342.6	III	123	Try	115.9	141.6	α_R
60	Ser	74.9	185.0	*S*	124	Ileu	82.6	176.2	α_R
61	Arg	98.8	162.5	I	125	Arg	109.7	308.9	III
62	Try	45.0	138.7	*H*	126	Gly	279.0	167.6	*S*
63	Try	93.5	145.5	I	127	Cys	103.7	321.0	III
64	Cys	26.9	330.9	III	128	Arg	92.0	283.2	III
65	Asn	83.2	329.4	III					

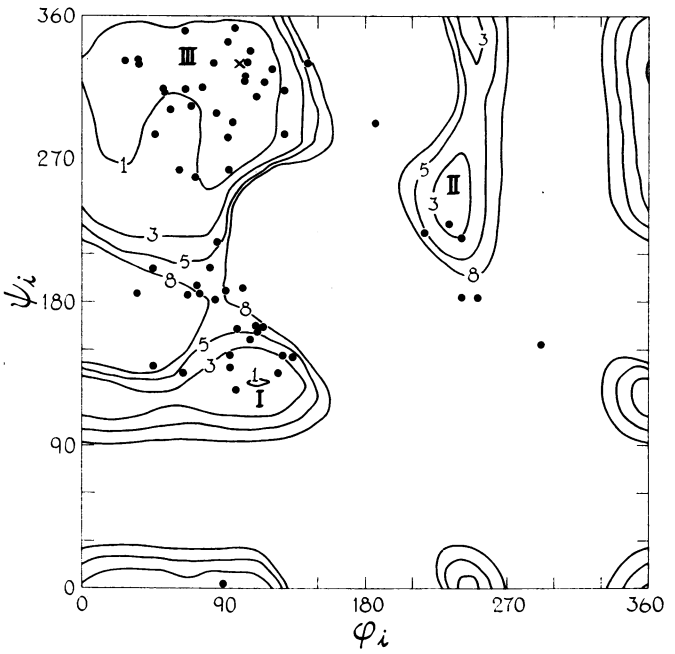

FIG. 2.—Energy contour map for a residue bearing a CH_2—R′ side chain at the alpha carbon, this residue being situated within a random coil polypeptide chain. The energy contours, numbered in kcal mole^{-1}, are drawn relative to the absolute minimum X and are calculated as described in the text (see also ref. 6). The three principal "sterically allowed" regions are labeled I, II, and III, respectively. Points represent the observed conformations of the appropriate residues in crystalline lysozyme.

Phillips has reported that residues 42–46 participate in an antiparallel pleated sheet with residues 50–54.[14] On the presumption that residues 46–50 adopt conformations which enable these two sequences to become antiparallel, the pleated sheet can be formed if residues 43–45 and 51–53 adopt approximately the conformation $\varphi_i = 35°$, $\psi_i = 322°$.[16] Inasmuch as these sequences are very short, the residues therein are most appropriately plotted in Figure 2, but their positions on the helix map in Figure 4 are indicated as well.

Neither the proline residues, 70 and 79, nor the immediately preceding residues, 69 and 78, are plotted because none of the aforementioned energy estimates is applicable to proline and its immediately *preceding* neighbor. The conformation of each of these residues is, however, within a low energy region of some apposite energy estimates.[8]

Discussion.—The relative population of residues within a "sterically allowed" region[15, 17] depends simultaneously on the energy and the "effective" accessible area of that region in the φ,ψ plane. Almost all of the residues are within or near a low energy contour. Moreover, 22 of the 61 residues plotted in Figure 2 cluster within the 1 kcal contour surrounding the absolute minimum. Some of the apparent departures from predicted conformations may be attributed to errors in the crystallographic data and in the energy estimates. In further qualitative agreement with the calculations, the numbers of residues in the three principal regions in

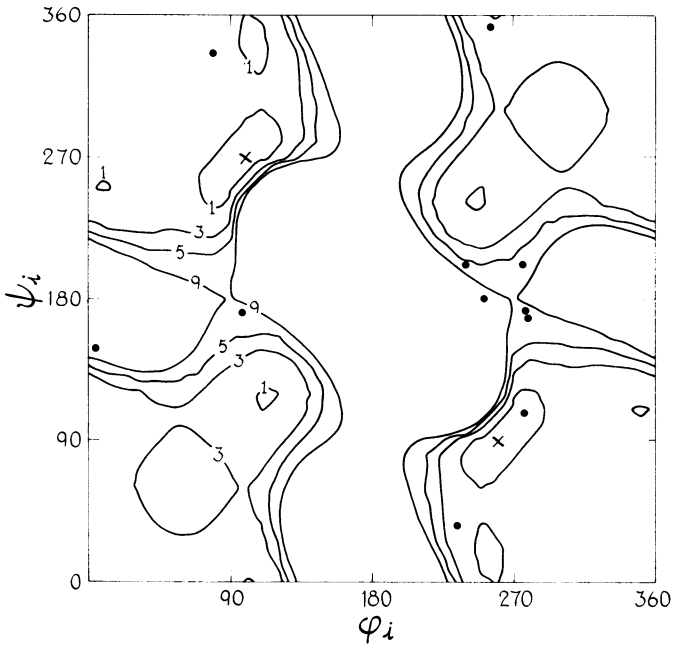

FIG. 3.—Energy contour map for a glycyl residue situated within a random coil polypeptide chain (see text and ref. 6 for details). The contours, numbered in kcal mole^{-1}, are drawn relative to the two absolute minima X of identical energy. Points represent the observed conformations of the appropriate glycyl residues of crystalline lysozyme.

Figure 2 are in the order III > I > II. Residue conformations not falling within I and III are expected to fall in the region of next lowest energy, i.e., II. The large population in the saddle region between I and III contrasts, therefore, with the sparse number in the lower energy region II. This circumstance is probably due to unaccounted compensating interactions, e.g., protein-solvent, steric overlaps between residues that are distant in the chain sequence, which are as a whole most favorable when these residues adopt conformations in the saddle region. It may also be due in part to the inexactitude of the energy estimates. In particular, the relative energy of the saddle region decreases when the constraint of constancy of structural parameters, e.g., bond angles, is relaxed.[3, 5] Alternatively, or additionally, the activation barrier for passing from I or III, the conformations adopted by virtually all residues in the random coil, to II may present a kinetic limitation to the attainment of the equilibrium state when the chain is folding (i.e., during or after its biosynthesis or upon renaturation). The activation barrier *per residue*, which depends upon the kinetic pathway followed in the φ,ψ plane, is for *all* pathways >10 kcal/mole[6] and for *most* pathways >20 kcal/mole.[6] The residues therefore adopt the next lowest energy state, the saddle region, which is kinetically accessible during the folding process. Such an interpretation must, however, be viewed as speculative.

Only 8 of the 61 residues plotted in Figure 2 occupy positions of excessively high energy; 2 of the 11 glycine residues in Figure 3 adopt significantly high energy conformations. Various explanations may be invoked to account for these anomalous

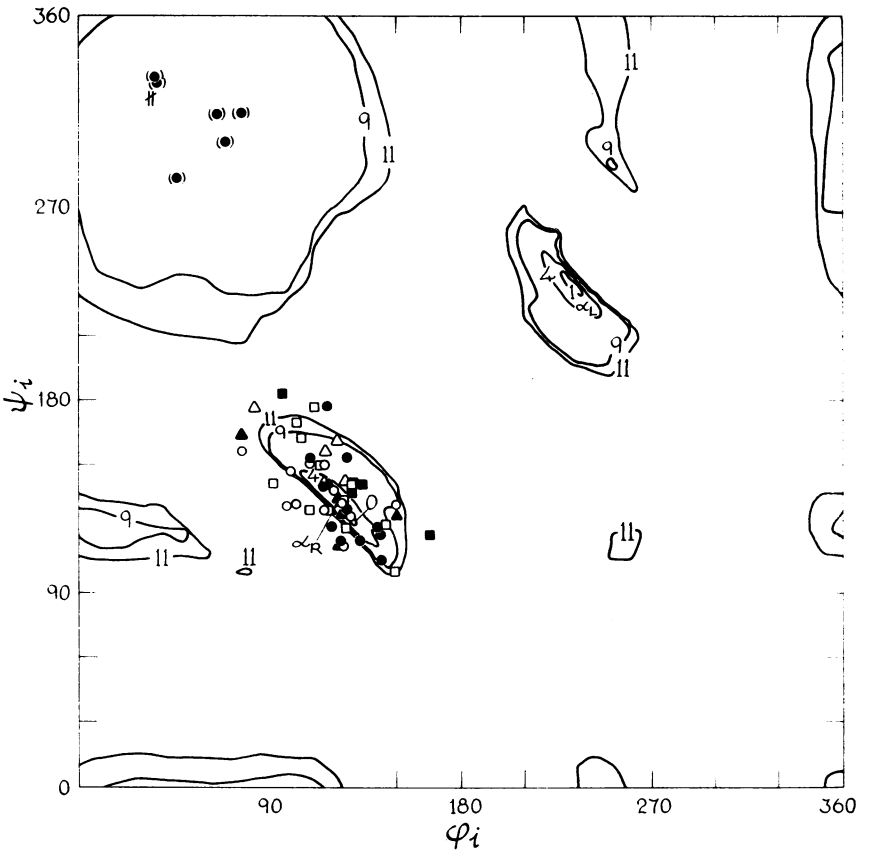

FIG. 4.—Energy contour map for a residue situated within a long helical sequence (see text and ref. 13 for details). The contours, numbered in kcal mole^{-1}, are drawn relative to the zero contour occurring at approximately $\varphi = 120$, $\psi = 130$. Coordinates of the right-handed α-helix, of the left-handed α-helix, and of the pleated sheet conformation are designated α_R, α_L, and $\mathcal{F}$, respectively. Symbols represent the observed conformations of residues occurring in regular (helix-like) sequences of crystalline lysozyme. Filled circles, filled squares, and filled triangles designate residues 5–15, 80–84, and 109–114, respectively; open circles, open squares, and open triangles designate residues 25–36, 88–99, and 120–124, respectively. The pleated sheet conformations of residues 43–45, 51–53 are designated by filled circles in parentheses; these pleated sheet residues are also plotted in Fig. 2.

conformations. For example, serine 50 occurs in the antiparallel pleated sheet region which can be formed only if the configuration of residues 46–50 permits the two participating chain segments to adopt the proper mutual orientation.

The residues classified as helical, plotted in Figure 4, scatter irregularly about the coordinates $\varphi = 122°$, $\psi = 133°$ of the perfect right-handed α-helix.[16] If this pattern is accurate, the occurrence of extensive hydrogen bonding within the helix is problematical. Refinement of the data may, however, contract this distribution. That no helical sequences occur which differ appreciably from α_R is consistent with the predicted higher energy of other helical configurations. Finally, the close adherence of the six pleated-sheet residues to the conformation of the perfect antiparallel pleated sheet is noteworthy.

While some of the physical characteristics of random coil and helical polypeptides can be deduced from the aforementioned conformational energy estimates, the preceding analysis demonstrates that such energy functions, though useful, are inadequate representations for a residue situated within a nearly space-filling protein. The high packing density of the native protein residues has no counterpart in other macromolecules.[18] It has been pointed out that conformations meeting the stringent demands of being space-filling may be exceedingly rare, or even unique.[18] The abundance and magnitude of the incalculable interactions which occur in a space-filling protein precludes, however, an *a priori* prediction of the configuration of lowest free energy.

Summary.—Observed residue conformations in crystalline hen egg-white lysozyme are compared with predictions based upon estimates of backbone conformational energy. Each residue is plotted on an energy contour diagram appropriate for the given residue and according to its occurrence in a random coil or helix-like (regular) sequence of lysozyme. Almost all residues fall within a predicted low energy domain and the relative number of units in a particular "sterically allowed" region qualitatively conforms with the theoretical expectations. Residues occurring in regular sequences cluster about the coordinates $\varphi_i = 122°$, $\psi_i = 133°$ of the right-handed alpha helix, the *helical* configuration predicted to be of lowest energy. Departures from predictions based on energy estimates are interpreted to reflect on the abundance and magnitude of the unaccounted interactions which occur in the protein, and possibly also on kinetic limitations to attainment of the state of equilibrium.

The authors are grateful to Professor P. J. Flory for his encouragement and criticism of this work.

* This research was made possible through funds provided by the Directorate of Chemical Sciences, U.S. Air Force Office of Scientific Research contract no. AF49(638)1341, and U.S. Public Health Service grant no. GM13908.

† Present address: Departments of Biology and Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

[1] Edsall, J. T., P. J. Flory, J. C. Kendrew, G. Nemethy, G. N. Ramachandran, and H. A. Scheraga, *Biopolymers*, **4**, 121 (1966).

[2] Brant, D. A., and P. J. Flory, *J. Am. Chem. Soc.*, **87**, 2791 (1965).

[3] Gibson, K. D., and H. A. Scheraga, *Biopolymers*, **4**, 709 (1966).

[4] Scott, R. A., and H. A. Scheraga, *J. Chem. Phys.*, **45**, 2091 (1966).

[5] Ramachandran, G. N., C. M. Venkatachalam, and S. Krimm, *Biophys. J.*, **6**, 849 (1966).

[6] Brant, D. A., W. G. Miller, and P. J. Flory, *J. Mol. Biol.*, **23**, 47 (1967).

[7] Schimmel, P. R., and P. J. Flory, these PROCEEDINGS, **58**, 52 (1967).

[8] Schimmel, P. R., and P. J. Flory, in preparation.

[9] Miller, W. G., D. A. Brant, and P. J. Flory, *J. Mol. Biol.*, **23**, 67 (1967).

[10] Schimmel, P. R., and P. J. Flory, in preparation.

[11] De Santis, P., E. Giglio, A. M. Liquori, and A. Ripamonti, *Nature*, **206**, 456 (1965).

[12] Ooi, T., R. A. Scott, G. Vanderkooi, R. F. Epand, and H. A. Scheraga, *J. Am. Chem. Soc.*, **88**, 5680 (1966).

[13] Brant, D. A., in preparation.

[14] Phillips, D. C., these PROCEEDINGS, **57**, 484 (1967).

[15] Leach, S. J., G. Nemethy, and H. A. Scheraga, *Biopolymers*, **4**, 369 (1966).

[16] Schellman, J. A., and C. Schellman, in *The Proteins*, ed. H. Neurath (New York: Academic Press, 1964), 2nd ed., vol. 2, p. 24.

[17] Ramakrishnan, C., and G. N. Ramachandran, *Biophys. J.*, **5**, 909 (1965).

[18] Flory, P. J., "Configurational statistics of polypeptides," a chapter to be published in *Conformation of Biopolymers*, ed. G. N. Ramachandran.

[19] The amino acid sequence is taken from Canfield, R. E., and A. K. Liu, *J. Biol. Chem.*, **240**, 1997 (1965). Another sequence, reported by Jollès, J., J. Jauregui-Adell, I. Bernier, and P. Jollès, *Biochim. Biophys. Acta.*, **78**, 668 (1963), is quite similar, although a few discrepancies exist.