Macfarlane, R. G. (1964), *Nature 202*, 495.
Moore, S., and Stein, W. H. (1954), *J. Biol. Chem. 211*, 908.
Smith, E. L. (1955), *Methods Enzymol. 2*, 98.
Spackman, D. H., Smith, E. L., and Brown, D. M. (1955), *J. Biol. Chem. 212*, 255.
Wagner, R. H., and Thelin, G. M. (1957), *in* Hemophilia and Hemophiloid Diseases, Brinkhous, K. M., Ed., Chapel Hill, N. C., University of North Carolina, p 3.

Dimensions of Protein Random Coils*

Wilmer G. Miller and Carol V. Goebel

ABSTRACT: The dimensions of protein random coils have been calculated for a variety of proteins of known amino acid sequence. As was anticipated from the work of Flory and coworkers, glycine and proline contribute to reducing the dimensions of random coil proteins. Branched side chains expand the chain only slightly more than unbranched side chains. Side chains represented as structured to the γ position were compared with structureless representations. It was demonstrated that the two approaches give comparable chain dimensions. The effect of sequence was investigated. Although the distribution of glycine and proline along the chain was of some importance, the mean-square end-to-end distance based on a knowledge of composition alone and assuming a randomized sequence was generally within 10% of the value obtained when sequence was taken into account. The calculated dimensions were compared with the experimental values of Tanford and coworkers determined from viscosity and osmotic pressure data in 6 M guanidine hydrochloride. The agreement is substantial with the experimentally deduced values tending to be slightly smaller than the calculated ones. An approximate accounting was made of the effect of regions of highly associated side chains within a randomly coiling molecule. The chain dimensions were found rather insensitive to reasonably large knots of associated residues. Thus it cannot be ruled out that proteins in 6 M guanidine hydrochloride have sizeable amounts of highly associated though not regularly ordered regions.

The measurement and understanding of thermodynamic and kinetic changes accompanying conformational changes in proteins has advanced to the point where very detailed models are being proposed. The reference state generally chosen is a folded or near-native conformation, giving a unique reference state for each protein. It would be desirable, if possible, to have a well-defined reference state that is common to all proteins.

Tanford *et al.* (1967a) have investigated the intrinsic viscosities and sedimentation coefficients of proteins in concentrated guanidine hydrochloride solutions and concluded "that protein polypeptide chains, in the solvent medium employed, are true random coils, retaining no elements of their original native conformation." Tanford and coworkers have presented extensive additional experimental evidence in an effort to support this conclusion (Nozaki and Tanford, 1967; Tanford *et al.*, 1967b; Lapanje and Tanford, 1967; Aune *et al.*, 1967).

The configuration of polymer random coils is a well-studied area of synthetic polymer chemistry. The theory leading to predictions of the average dimensions of polymer random coils has been quite successful in that values in agreement with experiment may be calculated for numerous synthetic polymer random coils (Volkenstein, 1963). Thus a polymer random coil as conformational reference state for proteins is attractive in that it is theoretically tractable.

The configuration of polymer random coils is well known to depend upon hindrance potentials for rotation about the main-chain bonds, as well as on bond angles, bond distances, and long-range interactions (excluded volume). Appropriate though approximate forms for the potential functions have been given for peptide-bond polymers (Brant and Flory, 1965b; DeSantis *et al.*, 1965; Ramakrishnan and Ramachandran, 1965; Leach *et al.*, 1966; Scott and Scheraga, 1966; Brant *et al.*, 1967; Schimmel and Flory, 1967, 1968). Utilizing hindrance potentials, bond angles, and bond distances (Sasisekharan, 1962; Leung and Marsh, 1958) applicable to specific types of amino acid residues, unperturbed random coil dimensions, *i.e.*, dimensions neglecting long-range interactions such as the chain crossing back on itself, have been computed for homopolymers having an alanine type (*cf. seq.*) amino side chain (Brant and Flory, 1965b; Brant *et al.*, 1967), for poly-L-proline (Schimmel and Flory, 1967), for random sequence copoly (glycine, alanine) (Miller *et al.*, 1967), and for copoly (glycine, proline), copoly (alanine, proline),

* From the Department of Chemistry, University of Minnesota, Minneapolis, Minnesota 55455. *Received June 26, 1968*. This work was aided by a U. S. Public Health Service grant (GM-15903).

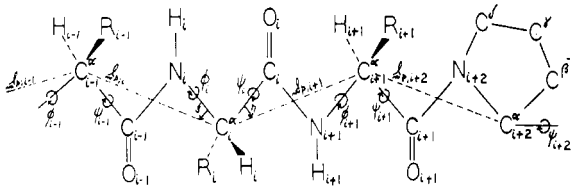

FIGURE 1: Section of a protein chain composed of L-amino acids. Virtual bond vectors l_p connecting α-carbons are shown as dashed lines. All amino acid side chains are designated R except proline.

and copoly (glycine, alanine, proline) (Schimmel and Flory, 1968).

In this communication conformation energy estimates are made for additional residues, and unperturbed random coil dimensions are computed for a variety of known-sequence proteins. The factors most strongly influencing the dimensions are discussed. The effect on the chain dimensions of strongly interacting side chains, as might be the case with hydrocarbon side chains in aqueous solution, is considered. The experimental results of Tanford and coworkers are compared and discussed with respect to the calculated values.

Mathematical Framework

The general approach, the approximations, and the mathematical framework for calculations of unperterbed random coil protein dimensions have been developed and discussed (Brant and Flory, 1965b; Brant *et al.*, 1967; Miller *et al.*, 1967; Schimmel and Flory, 1968). An abbreviated outline will be given.

A segment of a protein chain, considered to contain only L residues, is shown in Figure 1 where all amino acid side chains except proline are designated R. Assuming all amide groups are restricted to the *trans* conformation makes the C^α_i–C^α_{i+1} distance a constant and allows a chain of $n_p + 1$ amino acids to be described as a sequence of n_p virtual bonds l_p of length 3.8 Å joining consecutive α-carbon atoms. Neglecting side-chain configurations, a configuration of the protein is specified by assigning values to the dihedral angles φ and ψ defining rotations about the N–C^α and C^α–C bonds, respectively, for each residue.

A convenient measure of the unperturbed chain configuration is either the mean-square distance between the ends of the chain $\langle r^2\rangle_0$ or the mean-square radius of gyration $\langle s^2\rangle_0$, which differs numerically from $\langle r^2\rangle_0$ in long-chain random coils by a factor of 6 (Flory, 1953). Considering the virtual bonds as vectors $\mathbf{l}_p$, the end-to-end distance for a given configuration is obtained by summing the virtual bond vectors, and $\langle r^2\rangle$ is given by

$$\langle r^2\rangle = \sum_i \sum_j \langle \mathbf{l}_{p;i} \cdot \mathbf{l}_{p;j}\rangle \qquad (1)$$

where the fences denote the statistical mechanical average over-all configurations of the chain, i and j run independently from 1 to n_p, and $\mathbf{l}_{p;i}$ and $\mathbf{l}_{p;j}$ are expressed in a common coordinate system. A cartesian coordinate system is now affixed to each virtual bond with the origin taken as the α-carbon and the x axis taken coincident with $\mathbf{l}_p$. Each virtual bond is expressed in terms of its own corrdinate system. In order to obtain the end-to-end distance by summing the virtual bond vectors, however, the vectors must be expressed in a common coordinate system. To accomplish this, coordinate transformation matrices $\mathbf{T}$ are defined. Thus the matrix $\mathbf{T}_i$ transforms $\mathbf{l}_{p;i+1}$ from the coordinate system affixed to virtual bond $i + 1$ to that for virtual bond i, etc. The mean-square end-to-end distance becomes

$$\langle r^2\rangle = \sum_i \sum_j \langle \mathbf{l}_{p;i} \cdot (\mathbf{T}_i \mathbf{T}_{i+1} \ldots \mathbf{T}_{j-1} \mathbf{l}_{p;j})\rangle$$

or

$$= \sum_i \sum_j \mathbf{l}_{p;i}{}^{T} \langle \mathbf{T}_i \mathbf{T}_{i+1} \ldots \mathbf{T}_{j-1}\rangle \mathbf{l}_{p;j} \qquad (2)$$

where $\mathbf{l}_{p;i}{}^{T}$ and $\mathbf{l}_{p;j}$ are

$$[l_p\ 0\ 0] \text{ and } \begin{bmatrix} l_p \\ 0 \\ 0 \end{bmatrix}$$

respectively, expressed in matrix notation.

By virtue of the absence of rotation about the C–N bonds, rotations about the single bonds connecting an α-carbon atom to the two adjoining amide groups are sensibly independent of rotations about other bonds in the chain. Thus to good approximation the potential, V, affecting rotations within residue i depends upon the variables φ_i and ψ_i (only ψ_i in proline) and the conformational energy can be expressed as

$$V(\varphi_i,\psi_i) = V_t(\varphi_i) + V_t(\psi_i) + \sum_{j,k}[(V_{c,jk}(\varphi_i,\psi_i) + V_{r,jk}(\varphi_i,\psi_i) + V_{l,jk}(\varphi_i,\psi_i)] \qquad (3)$$

where V_t is an intrinsic threefold torsional potential ascribed to each of the main-chain single bonds; $V_{c,jk}$, $V_{r,jk}$, and $V_{l,jk}$ are, respectively, terms representing coulombic interactions between the two amide groups separated by the bonds N_i–C^α_i and C^α_i–C_i, repulsive steric overlaps, and London dispersion energies. The summation in eq 3 is over all pairs of atoms whose distance of separation depends upon only φ_i and ψ_i. If the amino acid side chain is considered to have structure, side-chain conformations must be considered. The total potential will then depend also upon bond rotation angles χ, describing rotation about the side-chain bonds and defined in Figure 2c for the rotation about the C^α–C^β bond. As a result of the independence of rotation about pairs of main-chain bonds, the average of products of coordinate transformation matrices in eq 2 may be replaced by products of averages, *i.e.*

$$\langle \mathbf{T}_i \mathbf{T}_{i+1} \ldots \mathbf{T}_{j-1}\rangle = \langle \mathbf{T}_i\rangle \langle \mathbf{T}_{i+1}\rangle \ldots \langle \mathbf{T}_{j-1}\rangle \qquad (4)$$

The averaged coordinate transformation matrix for the ith residue is obtained by averaging over all values of

φ_i and ψ_i, an element being given by

$$\langle T_i \rangle = \frac{\iint T_i(\varphi_i,\psi_i) \exp[-V(\varphi_i,\psi_i)/RT]\mathrm{d}\varphi_i\mathrm{d}\psi_i}{\iint \exp[-V(\varphi_i,\psi_i)/RT]\mathrm{d}\varphi_i\mathrm{d}\psi_i} \quad (5)$$

In practice $\langle T_i \rangle$ is obtained by replacing the integral by a summation over φ_i and ψ_i at specified intervals. To each amino acid in a protein chain is assigned an independently averaged matrix $\langle \mathbf{T} \rangle$ which depends upon the amino acid side chain (R group) and upon whether or not proline follows it in the chain. The latter effect is primarily a result of different steric overlaps (Schimmel and Flory, 1968).

Substitution of eq 4 into eq 2 allows the unperturbed mean-square end-to-end distance $\langle r^2 \rangle_0$ to be expressed as

$$\langle r^2 \rangle_0 / n_p l_p^2 = 1 + (2/n_p l_p^2)[10000] \prod_{i=1}^{n_p - 1} G_i \begin{bmatrix} 0 \\ l_p \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (6)$$

where

$$G_i = \begin{bmatrix} 1 & \mathbf{l}_p^{\mathrm{T}} \langle \mathbf{T}_i \rangle & 0 \\ \mathbf{0} & \langle \mathbf{T}_i \rangle & \mathbf{l}_p \\ 0 & 0 & 1 \end{bmatrix}$$

and $\mathbf{0}$ is a 3×1 null matrix.

Inasmuch as each protein has a unique monomer sequence, enumeration of that sequence permits $\langle r^2 \rangle_0$ to be calculated. The chain dimensions may be expected to depend upon the residue sequence as well as upon the amino acid composition. In many instances, the amino acid composition of a protein is known whereas the amino acid sequence is either unknown or only partially known. In synthetic polypeptide copolymers synthesized in a manner such that the amino acid residues are incorporated into the copolymer at random, chains vary widely in monomer sequence. It has been shown (Miller *et al.*, 1967) that eq 6 may still be used, however, if a single, averaged $\langle \mathbf{T}_i \rangle$ is defined, given by

$$\langle \mathbf{T}_i \rangle = \sum_{r=1}^{x} P_r \langle \mathbf{T}_r \rangle \quad (7)$$

where P_r is the mole fraction of residues of type r with an averaged transformation matrix $\langle \mathbf{T}_r \rangle$ in a copolymer containing x types of residues. When amino acid composition is of predominant importance compared with sequence in determining chain dimensions (*cf. seq.*) eq 6 and 7 may be used for proteins of known composition but unknown sequence.

Conformational Energy Estimates and $\langle \mathbf{T} \rangle$ Matrices to Be Used in Protein Calculations

The experimental finding of Brant and Flory (1965a) that $\langle r^2 \rangle_0 / n_p l_p^2$ was the same (9 ± 1) for random coil polybenzyl-L-aspartate, sodium poly-L-glutamate, poly-L-lysine hydrobromide, and polybenzyl-L-glutamate, each in a different solvent system, indicated that a separate $\langle \mathbf{T} \rangle$ matrix may not be needed for each amino acid. Inasmuch as $C^\beta H_2$ is the only group of atoms common to the four amino acid side chains, they concluded that carbons past the β position can rotate away from steric conflicts and that these atoms need not be considered when computing the potential functions given by eq 3 (Brant and Flory, 1965b). The methylene group was treated as a structureless entity, *i.e.*, as a carbon atom with larger than normal van der Waals contact radius (r_{VDW}). Using point dipoles in the coulombic potential, V_c, exponential repulsive potentials in V_r, and $r_{VDW} = 1.85$ Å in describing the methylene group, chain dimensions consistent with experiment were calculated. Agreement with experiment was also obtained using point monopoles in V_c and $1/r^{12}$ repulsive terms in V_r (Brant *et al.*, 1967). Keeping all parameters fixed and replacing the methylene with a hydrogen atom to describe glycine lead to calculated dimensions for glycine–glutamate copolymers consistent with experiment (Miller *et al.*, 1967).

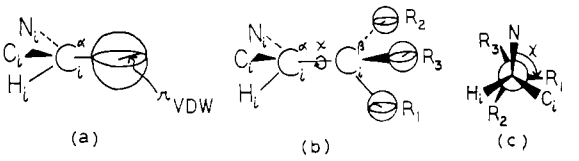

FIGURE 2: Models for treating side chains in conformational energy calculations. (a) Side chain considered structureless and represented by placing at the β position a carbon atom, or hydrogen for glycine, having a van der Waals contact radius r_{VDW}; (b) side chain structured out to the γ position with the β position occupied by a normal carbon atom of $r_{VDW} = 1.70$ Å; (c) definition of bond rotation angle χ with eclipsed $N_i - R_1$ taken as zero angle.

From the foregoing discussion it would seem that the remaining conformational energies and corresponding $\langle \mathbf{T} \rangle$ matrices needed for protein dimension calculations may be approached with some confidence. Unless otherwise stated the parameters for torsional potentials, van der Waals contact radii, atomic polarizabilities, partial charges, etc., given by Brant *et al.* (1967) will be used throughout. Proline is considered not to follow the residue under consideration unless stated explicitly. The high molecular weight limit of the characteristic ratio, $(\langle r^2 \rangle_0 / n_p l_p^2)_\infty$, for a polymer composed of a single type of side chain is used to test sensitivity to variable parameters. It is also a qualitative indication of the effect of that side chain on protein dimensions with more extended chains having larger values of the characteristic ratio. The characteristic ratio for a protein with completely free rotation about φ and ψ is 1.93.

Previous calculations of $\langle \mathbf{T} \rangle$, except in the case of proline, have been made considering the side chain to be structureless. A representation structured out to the γ position, Figure 2b, will also be considered and compared with the structureless treatment. A side-chain carbon atom to which two or more hydrogen atoms are appended is designated as unbranched at that point, a branching point existing at a carbon with one or no hydrogens appended.

TABLE I: $\langle \mathbf{T} \rangle$ Matrices (See Eq I for Matrix Representation).

Structureless Representation	a_{11}	a_{12}	a_{13}	a_{21}	a_{22}	a_{23}	a_{31}	a_{32}	a_{33}	$(\langle r^2 \rangle_0 / n_r l_p^2)_\infty$
A, Glycine[a]	0.36	−0.077	0.00	−0.092	−0.37	0.00	0.00	0.00	−0.12	2.0
B, Glycine preceding proline[b]	0.47	0.20	0.00	0.019	−0.0021	0.00	0.00	0.00	0.038	
C, Alanine type[a]	0.51	0.20	0.59	−0.046	−0.61	0.21	0.65	−0.23	−0.30	9.3
D, Alanine type preceding proline[b]	0.50	0.14	0.81	−0.031	−0.73	0.098	0.67	−0.084	−0.32	
E, Branched side chain	0.55	0.27	0.62	−0.0055	−0.67	0.25	0.66	−0.21	−0.39	12.9
Structured Representation										
F, Proline (equal minima[b])	0.33	0.18	0.34	−0.61	−0.013	0.30	0.64	−0.54	0.11	
G, Proline (unequal minima[b])	0.070	−0.46	−0.83	−0.28	0.81	−0.45	0.95	0.23	−0.069	
H, Alanine	0.41	0.060	0.63	−0.19	−0.52	0.13	0.69	−0.29	−0.18	8.0[c]
I, $C^\beta H_2$-($C^\gamma H_2$R); $-C^\beta H_2$-(C^γHRR′)	0.39	0.044	0.65	−0.25	−0.56	0.15	0.73	−0.35	−0.21	8.9
J, Serine	0.40	0.056	0.64	−0.21	−0.53	0.14	0.70	−0.31	−0.19	8.4
K, Valine; isoleucine	0.32	−0.067	0.70	−0.37	−0.60	0.058	0.78	−0.40	−0.23	10.7
L, Threonine	0.34	−0.044	0.71	−0.33	−0.60	0.076	0.76	−0.37	−0.22	10.6

[a] Brant *et al.* (1967). [b] Schimmel and Flory (1968). [c] $(\langle r^2 \rangle_0 / n_r l_p^2)_\infty$ for structured side chains was taken as the value for a chain with $n_p = 500$. The difference between the true limit and that reported here is $\leqslant 0.1$.

All calculations to obtain $\langle \mathbf{T} \rangle$ through use of eq 5 were performed at 10° intervals of φ and ψ. A temperature of 298°K was assumed.

Glycine. The choice of parameters in the potential function affects the details of the potential map describing a glycine residue but the elements of $\langle \mathbf{T} \rangle$ are relatively insensitive. The characteristic ratio for the homopolymer, depending only upon $\langle \mathbf{T} \rangle$, is consequently insensitive and has a value of 2.0 ± 0.2 over a wide range in choice of parameters. The $\langle \mathbf{T} \rangle$ matrix of Brant *et al.* (1967) (Table I,A; see eq I for matrix representation) will be utilized.

$$\langle \mathbf{T} \rangle = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad (1)$$

Proline and the Residue Immediately Preceding Proline (Schimmel and Flory, 1968.) The potential for a proline isolated from other proline residues shows two sharp minima located at $\psi = 125$ and 325°. The initial choice of parameters used by Schimmel and Flory resulted in minima of about equal energy and a $\langle \mathbf{T} \rangle$ designated as F in Table I. Variation of parameters over reasonable values indicated that chain dimensions changed significantly with no recourse to experiment available as a guide in choosing parameters. When comparing calculated to experimental values factors which reduce the calculated dimensions become of interest (*cf. seq.*). The potential function giving a minimum at $\psi = 125°$ 3 kcal lower than the one at $\psi = 325°$ leads to the $\langle \mathbf{T} \rangle$ given in Table I,G. This potential lowers the chain dimensions and will be taken as an extreme representation of an isolated proline. These two models are referred to as proline equal minima and proline unequal minima representations.

A proline adjacent to another proline is not represented by F or G as different steric conflicts arise (Schimmel and Flory, 1967). Among the known-sequence proteins on which calculations were performed proline occurs 258 times out of a total of 5762 residues. Of these only two pairs of consecutive prolines occur. The occurrence of consecutive prolines is so rare that either F or G will be used to describe the average configuration of any proline residue.

A proline succeeding a glycine results in different steric overlaps for glycine than if any other residue succeeds glycine. The glycine potential map and $\langle \mathbf{T} \rangle$ matrix differ from those in the previous section. The $\langle \mathbf{T} \rangle$ matrix in this case is given by B in Table I.

Any residue other than glycine or proline preceding a proline also will have difierent steric conflicts. For alanine preceding proline the $\langle \mathbf{T} \rangle$ matrix D, Table I, was calculated by Schimmel and Flory. Although this is not expected to be correct for a branched side chain, such as valine, we have elected to use it for all residues other than glycine or proline preceding proline on the finding that the frequency of occurrence is low enough to have negligible effect.

Alanine and Amino Acids Unbranched at the β-Carbon. Several potential maps resulting from different expressions for interactions and based on a structureless

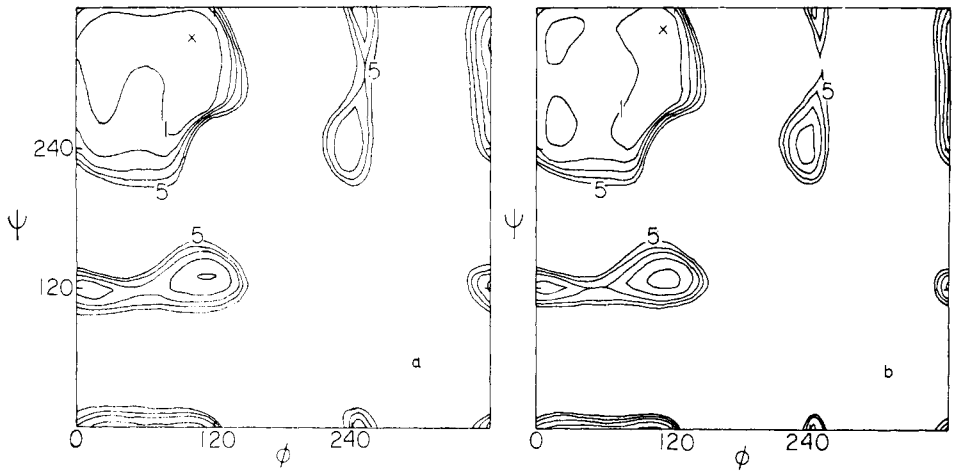

FIGURE 3: Contour map for the conformational energy of an alanine residue calculated (a) from a structureless side chain representation (Brant *et al.*, 1967) and (b) from a structured representation. Contours are given at 1-kcal intervals up through 5 kcal above the energy minimum (×).

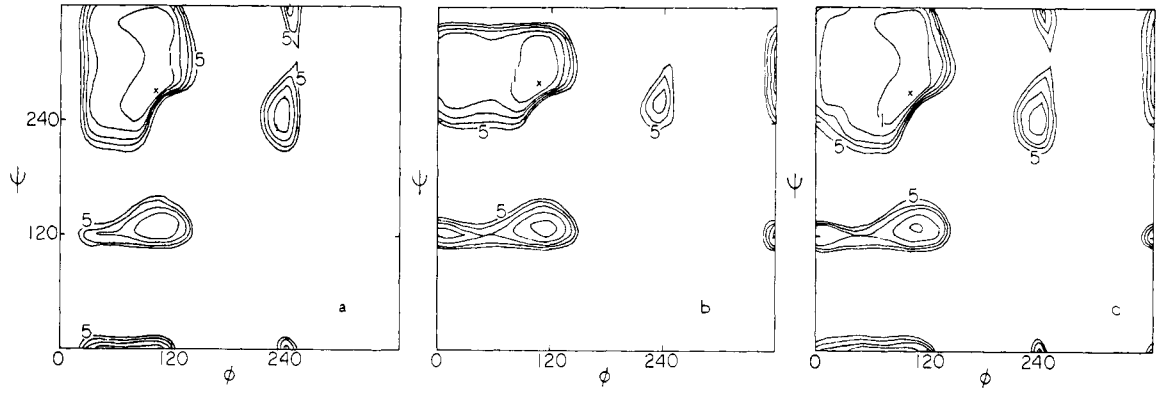

FIGURE 4: Contour map for a $C^{\beta}H_2(C^{\gamma}H_2)$ side chain using a structured representation (a) $\chi = 300°$, (b) $\chi = 180°$, and (c) composite contour map with each point corresponding to rotamer (χ = 300, 180, or 60°) of lowest energy.

side chain representing $C^{\beta}H_3$ or $C^{\beta}H_2$ have been presented (Brant and Flory, 1965b; Brant *et al.*, 1967). Although the general features are quite similar, they differ in detail and the resulting $\langle \mathbf{T} \rangle$ matrices give unperturbed dimensions varying by a greater amount than is the case with glycine. A representative $\langle \mathbf{T} \rangle$ is given in Table I,C. The corresponding potential map is given by Figure 3a.

A structured representation of the alanine side chain was considered next. The contact radius was taken as 1.70 Å for the β-carbon, and 1.10 Å for each of the hydrogen atoms. The staggered conformations corresponding to $\chi = 60$, 180, or 300° were assumed to be of lowest energy. The potential calculations were carried out maintaining χ constant at 60°. The resulting potential map and $\langle \mathbf{T} \rangle$ matrix are given in Figure 3b and Table I, respectively. Introduction of the structured representation results in a rise in the potential when $\psi = 300°$ or $\varphi = 50°$, due to steric overlaps involving the γ-hydrogens. This effect is cumulative, making the most significant difference in the region centered about $\psi = 300°$, $\psi = 50°$. The conformational energy in this region could be reduced somewhat by relaxing the restriction of a fixed χ as was done by Ooi *et al.* (1967) when investigating the conformational energy of polypeptide helices. To do so results in a marginal gain, changing elements in $\langle \mathbf{T} \rangle$ a smaller amount than does changing the method of treating the nonbonded interactions.

The amino acids methionine, lysine, arginine, and glutamic acid have $-C^{\beta}H_2-C^{\gamma}H_2R$ side chains while leucine, histidine, aspartic acid, phenylalanine, tyrosine, and tryptophan are branched at the γ-carbon. A structured model for these side chains similar to that for alanine was considered, except that R_1 (Figure 2b) was taken to be a carbon atom with $r_{VDW} = 1.85$ Å, *i.e.*, a structureless methylene. The staggered conformations were assumed to be of lowest and arbitrarily equal torsional energy. The potential maps with χ held constant at 300 and 180° are shown in Figure 4a,b. The minimum in the $\chi = 180°$ map is only 0.4 kcal above the minimum in the $\chi = 300°$ map. Values of 9.64, 7.33, and 19.1 for $(\langle r^2 \rangle_0 / n_p l_p^2)_\infty$ were obtained when $\chi = 300$, 180, and 60°, respectively. The conformational energies with $\chi = 300$ or 180° are lower than when $\chi = 60°$ for all values of ψ and φ except when $\psi = 350 \pm 20°$ and simultaneously $\varphi = 10 \pm 20°$. A composite map resulting from combining maps for the three χ values to give a map of lowest conformational energy is shown in Figure 4c and

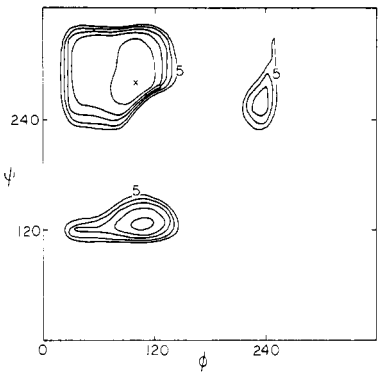

FIGURE 5: Contour map for valine or isoleucine using a structured representation. $R_1 = R_2 = C^{\gamma}H_2$; $R_3 = H$; $\chi = 180°$.

$\langle \mathbf{T} \rangle$ in Table I,I. This map is quite similar to that for structured alanine. Thus the assumption of Brant and Flory is an excellent one. If the restriction of only three allowed values of χ is relaxed, the significantly populated range of conformations, *i.e.*, those with conformational energy less than 5 kcal above the minimum, is expanded. In computing $\langle \mathbf{T} \rangle$ and hence the averaged conformation our error in not doing this is compromised considerably, however, by the fact that each conformation in Figure 4c was counted only once when computing $\langle \mathbf{T} \rangle$. As is clear from Figure 4a,b, there are values of φ and ψ where more than one side-chain rotamer is essentially equally probable. If $\langle \mathbf{T} \rangle$ were obtained by proper averaging over χ as well as φ and ψ, these regions would receive additional weight. The two factors tend to cancel each other.

Serine differs from the other residues unbranched at the β position in that it has a γ-oxygen rather than an γ-carbon. Representing the hydroxyl as an oxygen atom with slightly expanded van der Waals radius, r_{VDW} = 1.55 Å, a composite map analogous to $-C^{\beta}H_2-C^{\gamma}H_2R$ was generated. As one would anticipate, the map is intermediate in appearance between Figures 3b and 4c. The resulting $\langle \mathbf{T} \rangle$ matrix (Table I,J) leads to a characteristic ratio intermediate between H and I. Although serine (and threonine) can hydrogen bond intramolecularly, which would invalidate this calculation, insufficient knowledge is available to include such interactions in a meaningful way or to decide whether or not such interactions are important.

Valine, Isoleucine, and Threonine. A structureless model for a β-branched side chain consists of increasing the van der Waals contact radius of the structureless entity representing the side chain over that used in describing a structureless unbranched side chain. As r_{VDW} is increased the calculated characteristic ratio increases, having values of 12.9, 17.5, and 25.1 as r_{VDW} is increased to 1.95, 2.05, and 2.15 Å, respectively. The total area on the potential map less than 5 kcal above the minimum is progressively reduced as r_{VDW} is increased. In contrast to a methylene group where an equivalent contact radius may be deduced from X-ray data there seems little basis for choosing an equivalent contact radius for a branched β-carbon. In addition no experimental data on the dimensions of branched polypeptides are available as a guide in choosing r_{VDW} for a β-carbon representing in entirety a branched side chain. The $\langle \mathbf{T} \rangle$ given in Table I,E corresponds to a contact radius only 0.1 Å greater than that used in representing a methylene group. Although this may be an unrealistically small increase, it leads to an average chain dimension somewhat larger than our structured model for a branched side chain.

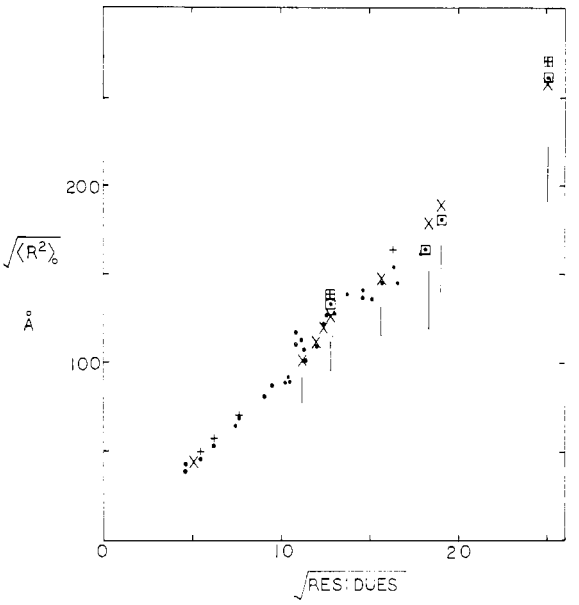

FIGURE 6: The calculated unperturbed root-mean-square end-to-end distance of random coil proteins using structured, proline equal minima (●) or structureless, proline equal minima (+) side-chain representation. Squares indicate sequence is unknown and calculation performed assuming a randomized sequence. Also shown are experimentally deduced unperturbed (.) and perturbed (×) dimensions of proteins in 6 M guanidine hydrochloride (Tanford *et al.*, 1966; Lapanje and Tanford, 1967).

A side-chain representation for valine and isoleucine structured to the γ position results from taking $R_1 = R_2 = C^{\gamma}H_2$ and $R_3 = H$, and considering R_1 and R_2 as a carbon atom with r_{VDW} = 1.85 Å. If χ is taken as 180°, the steric conflicts are fewer and the potential energy lower for most values of φ and ψ compared with similar calculations for χ = 60 or 300°. The potential map with χ taken to be 180° is shown in Figure 5, the corresponding $\langle \mathbf{T} \rangle$ in Table I,K. For reasons cited earlier a proper averaging over φ, ψ, and χ is unlikely to change $\langle \mathbf{T} \rangle$ significantly. Comparison of the calculations for structureless *vs.* structured representation of unbranched side chains leads to the conclusion that treating the γ position as structured will change the elements of $\langle \mathbf{T} \rangle$ very little.

In a structured representation of threonine one of the γ positions is occupied by a hydroxyl rather than a methylene as in valine. Since the van der Waals radius representing a hydroxyl is taken to be 0.30 Å smaller than that representing a methylene, one would expect less steric conflicts than those which occur in valine but more than those which ocur in $C^{\beta}H_2(-C^{\gamma}H_2R)$. Calculations restricted to χ equal to 300, 180, or 60° indicate that each of the three rotamers is energetically

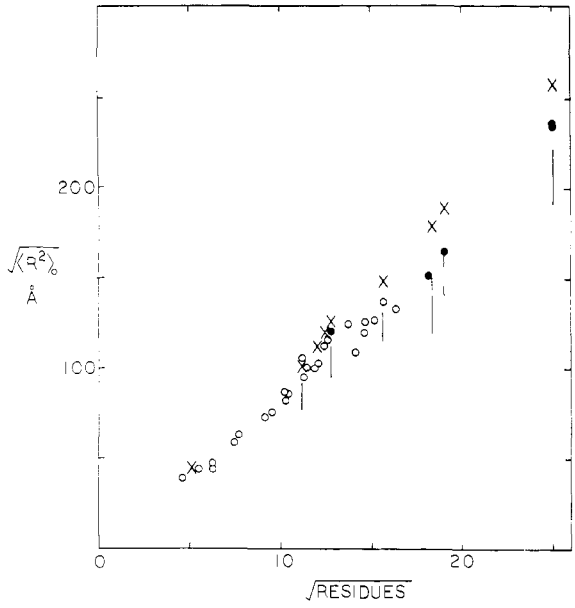

FIGURE 7: Analogous to Figure 6 with calculated values (○) based on structured side chains but using the unequal potential minima model for proline. Filled circles indicate assumption of random sequence.

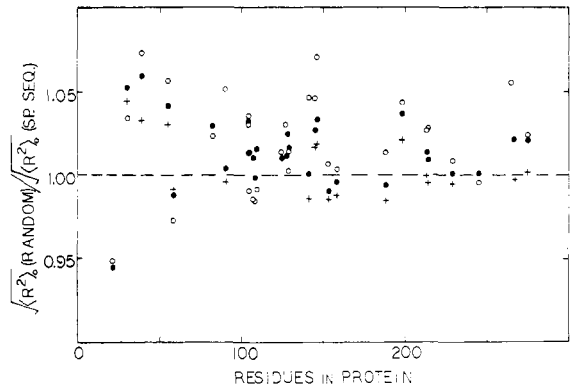

FIGURE 8: Effect of sequence on protein dimensions. The ratio of the unperturbed root-mean-square end-to-end distance of a "randomized sequence" to a specific sequence protein of the same composition is plotted for a variety of known-sequence proteins. Side chains treated as structured with equal proline minima (●), structured with unequal proline minima (○), or structureless with equal proline minima (+).

preferred in some region of the φ–ψ map. A composite lowest energy map and $\langle \mathbf{T} \rangle$ were generated in a manner analogous to those for unbranched side chains. The contour map more closely resembles valine than serine. The $\langle \mathbf{T} \rangle$ (Table I,L) leads to a characteristic ratio only slightly less than the value for valine.

Calculated Chain Dimensions of Protein Random Coils. In the last 10 years an impressive number of amino acid sequences have been elucidated, in many cases being proteins of the same functionality isolated from different organisms. The amino acid sequence for all proteins on which calculations were performed was obtained from Dayhoff and Eck (1968). The following proteins were selected from this compilation: papain (papaya), lysozyme (chicken), ribonuclease (bovine, rat, and Tl-fungus), subtilisin (*B. subtilis*), tryptophan synthetase α (*E. coli*), chymotrypsinogen A (bovine), trypsinogen B (bovine), trypsin inhibitor, hemoglobin (human α, β, γ, and Δ; plus five other sources), myoglobin (sperm whale), ferredoxins (clostridium), azurin, cytochrome (nine sources), coat protein of TMV (vulgare and dahlemense), corticotropin (four sources), β-lipotropin (sheep), growth hormone (human), insulin (a variety of A and B chains), and two Bence–Jones proteins (κ-human AG and λ-human SH). This selection comprises 51 proteins, totaling 5762 amino acid residues.

Four of the six proteins whose unperturbed dimensions have been determined from experimental data by Tanford and coworkers have not had their sequences completely elucidated. Inasmuch as we can perform calculations by the random sequence assumption if the amino acid composition is known, aldolase (rabbit muscle), pepsin (bovine), β-lactoglobulin, and serum albumin (bovine and human) were selected. These five proteins total 2106 residues.

The unperturbed root-mean-square end-to-end distances for the selected list of proteins were computed using eq 6 for known sequence proteins and eq 6 and 7 for the other five. Employing $\langle \mathbf{T} \rangle$ matrices (H,I,J), (K,-L), (B,D,F), and A, grouped here according to type of side chain, results in the calculated dimensions given in Figure 6. These calculations make the fullest use of the structured side-chain representation. In several instances, *e.g.*, the cytochromes, the source of the protein made little difference on the dimensions. Such cases are represented by a single point on the graph. If G is substituted for F and the calculations are repeated, the chain dimensions, shown in Figure 7, are systematically lowered. These calculations differ from the previous ones only in the representation of the proline residues.

Calculations based on C, E, (B,D,F), and A were performed. Some of the results are shown in Figure 6. Using G instead of F for proline results in changes very similar to those found for such a substitution using structured side chains.

The amino acid sequence as well as composition may be expected to affect the chain dimensions. The composition of our selected list of proteins varies widely from protein to protein. If, however, the amino acids from the 51 known sequence proteins were pooled, 8.5% would be glycine, 4.5% proline, 16% β branched, and 71% β unbranched. Inasmuch as the β-unbranched residues are dominant in number, the effect of various residues may be inferred by generating a reference chain for each protein composed of β-unbranched residues, followed by a systematic replacement with each type of residue at the natural occurring positions for that residue in the protein chain. An alanine residue was chosen to represent the β-unbranched homopolymer. After calculation of the homopolymer dimensions, glycine was placed at the appropriate positions in the chain, followed by proline implacement, and finally by the full complement of types of residues. The four sets of calculations given in Table II are representative of the results. Several generalizations can be made. (1) The presence of glycine lowers the chain dimensions significantly,

TABLE II: Effect of Various Types of Residues on $\langle r^2 \rangle_0^{1/2}$.

Protein	Subtilisin (*B. subtilis*)	α-Hemoglobin (human)	β-Lipotropin (sheep)	Ribonuclease (bovine)
No. of Residues	275	141	90	124
Glycine (%)	11.4	5.0	8.9	2.4
Proline (%)	5.1	5.0	5.6	3.2
I. H[a]	177.3[b]	125.7	99.2	117.5
II. H,A	147.1	116.7	90.7	111.8
III. H,A,(B,D,F)	144.1	115.1	85.5	111.3
IV. (H,I,J),A, (B,D,F,),(K,L)	145.0	117.2	86.9	113.3
V. H,A,(B,D,G)	133.6	102.6	75.8	105.0
VI. (H,I,J),A, (B,D,G),(K,L)	132.7	102.3	75.8	105.8
VII. C	191.4	135.6	106.9	126.8
VIII. C,A	157.3	125.4	97.3	120.3
IX. C,A,(B,D,F)	135.5	123.4	91.0	119.5
X. C,A,E, (B,D,F)	154.1	124.2	91.0	119.1

[a] Letters refer to the $\langle \mathbf{T} \rangle$ in Table I used in calculation of $\langle r^2 \rangle_0^{1/2}$. [b] $\langle r^2 \rangle_0^{1/2}$ in ångstroms.

as can be seen by comparing I–II or VII–VIII. The effect depends on both the amount of glycine present as well as its location along the chain. (2) The effect of proline depends upon the potential functions used in obtaining $\langle \mathbf{T} \rangle$ for proline. Comparing II–III and II–V, proline is seen to have a small effect in lowering the dimensions if the proline equal minima $\langle \mathbf{T} \rangle$ is utilized, but an effect comparable with glycine if the proline unequal minima $\langle \mathbf{T} \rangle$ is employed. (3) The branched side chains tend to force the polymer chain to expand slightly.

The effect of sequence was investigated by comparing the chain dimensions of the known-sequence proteins with the dimensions calculated on the random-sequence assumption. The results are shown in Figure 8. For the majority of the proteins $\langle r^2 \rangle_0^{1/2}$ calculated on the basis of a randomized copolymer sequence is within ±5% of the calculations which take the sequence into account, irrespective of the methods used to represent the residues. The majority of the values using the known sequence are less than corresponding randomized-sequence values.

Discussion

Factors Influencing the Calculated Chain Dimensions. The effect of glycine in reducing the chain dimensions was anticipated from the glycine–alanine random-sequence calculations of Miller *et al.* (1967). The similar effect of proline was likewise anticipated by the random copolymer calculations of Schimmel and Flory (1968). The influence of branched side chains in expanding the polymer over the unbranched side-chain values is small compared with glycine's influence in reducing the dimensions. It is tempting to explain these results in terms of flexibility or freedom of rotation about the main chain bonds. This view is erroneous. While glycine has more φ–ψ space accessible to it than any other residue, proline has the least. Yet the presence of either one reduces the average dimensions. When the conformational energy $V(\varphi_i,\psi_i)$ for a particular type of residue is equal to $V(-\varphi_i,-\psi_i)$ over all φ_i and ψ_i, the influence of that residue will be to lower the chain dimensions. This will be the result irrespective of the freedom of rotation. Thus if two methyl groups are attached to the α-carbon, instead of two hydrogens as in glycine, there is serious steric conflict for most values of φ_i and ψ_i. The calculated dimensions for the homopolymers composed of these residues are nearly identical, however, and close to the value corresponding to a chain with free rotation about φ and ψ. When the α-carbon is an asymmetric center, *i.e.* for all naturally occurring amino acids other than glycine, $V(\varphi_i,\psi_i) \neq V(-\varphi_i,-\psi_i)$. As the significantly populated φ–ψ space available to such a residue is restricted, the effect that residue will have on the chain dimensions depends mostly as to which areas in φ–ψ space the residue is restricted.

The effect of the position of the various amino acids in the protein chain is primarily a question of where glycine and, depending upon the representation, the proline residues are located. If, for example, glycine residues are placed near the ends of the chain in a glycine–alanine copolymer, they are less effective in reducing $\langle r^2 \rangle_0$ than if they are placed in the middle of the chain. A calculation of $\langle r^2 \rangle_0$ assuming a randomized sequence will fall in between. Thus it is not surprising to find some of the calculated values presented in Figures 6 and 7 to be larger than that calculated for the corresponding randomized sequence, and some to be smaller. The finding that the random sequence assumption leads to a root-mean-square distance within 5% of the specific sequence

values puts in effect an error limit on the dimensions of protein chains calculated from a knowledge of amino acid composition alone.

The validity of the calculated dimensions rests ultimately with the conformational energy estimates. As mentioned earlier the $\langle \mathbf{T} \rangle$ for glycine is rather insensitive to the potential functions. The $\langle \mathbf{T} \rangle$ matrix for β-unbranched residues, numerically dominant in proteins, is considerably more sensitive. Undetermined parameters in the potential functions were chosen to yield dimensions in agreement with experimental values for β-unbranched homopolymers, thus removing the major uncertainties. The β-branched residues affect the chain dimensions very little, and conformational energy uncertainties are of minor importance. Proline, however, is sensitive to the conformational energy treatment and no experimental data exist as a guide line in choosing parameters. Thus the treatment of proline accounts for the major uncertainty in the calculated dimensions. It should be noted that this uncertainty is greater than the error which results from basing calculations on the random sequence assumption.

Comparison of the Calculated with Experimental Results. Tanford and coworkers have determined chain dimensions, given in Figures 6 and 7, for proteins in 6 M guanidine hydrochloride using viscosity and osmotic pressure data. They conclude that in this solvent system proteins behave as true polymer random coils. Our calculated values may be compared with experimental data analyzed in three ways: perturbed dimensions calculated from $\langle r^2 \rangle^{1/2} = (M[\eta]/\Phi)^{1/3}$ and not corrected for nonideality, the unperturbed dimensions obtained entirely from viscosity data through use of the Kurata–Stockmayer (1963) equation, and the unperturbed dimensions obtained using osmotic second virial coefficient corrections. Lapanje and Tanford (1967) found the latter two methods to give similar results.

The values in Figure 6 calculated using a proline equal minima $\langle \mathbf{T} \rangle$ are seen to fall around the perturbed experimental values and somewhat above the unperturbed values. When the proline unequal minima values are compared (Figure 7), the calculated values fall near the upper limit for the unperturbed results. In Table III the slopes from the various plots are tabulated. It should be noted that a least-squares line through the calculated points does not go through the origin. From theory for high molecular weight homopolymers $\langle r^2 \rangle_0 / n_p l_p^2$ is constant and the high molecular weight data must extrapolate through the origin. For heteropolymers of variable composition and monomer sequence, there is no *a priori* reason for extrapolation through the origin.

Comparing either slopes or absolute values the unperturbed dimensions deduced by Lapanje and Tanford are in fair agreement with the calculated values, although the calculated values tend to be somewhat larger. Before concluding that these calculations confirm that proteins in 6 M guanidinde hydrochloride are indeed true polymer random coils, several aspects of the dilute solution properties of copolymers should be considered. Consider first a random coil homopolymer dissolved in a good solvent, followed by successive reduction of the solvent power. Numerous experimental studies show

TABLE III: Slope of $\langle r^2 \rangle_0^{1/2}$ *vs.* (Number of Residues)$^{1/2}$.

	Experimental (Å) (Tanford and Coworkers)
Viscosity (uncorrected)	10.7 ± 0.5[a,b]
Viscosity (corrected)	8.4 ± 0.5[b]
Viscosity, second virial coefficient	7.7 ± 0.6[b]
Viscosity, second virial coefficient	8.7 ± 0.5[c]
	Calculated (Å)
Calculated using proline, equal minima	10.0 ± 0.7
Calculated using proline, unequal minima	9.1 ± 0.7

[a] Not corrected to unperturbed conditions. [b] As given by Tanford and coworkers. [c] Obtained by not assuming origin as a fixed point through which data must extrapolate.

that as the solvent power is reduced, $\langle r^2 \rangle$ steadily decreases. When the solvent power is reduced far enough, the polymer structure collapses and the polymer eventually precipitates. Consider next a copolymer composed of two types of monomeric units. In general a given solvent will have a different solvent power for each monomer. What will be the chain configuration if the solvent is a good solvent for one of the monomers, but a nonsolvent for the other? The viscosity? The second virial coefficient? The unperturbed dimension?

One of the most studied copolymers from the configurational viewpoint is the styrene–methyl methacrylate copolymer (Kotaka *et al.*, 1966; Froelich and Benoit, 1966; Stockmayer *et al.*, 1955). With random copolymers the unperturbed dimensions appear to be slightly larger than the linear average of the dimensions of the parent homopolymers, irrespective of whether the solvent has similar or very unequal solvent power for the two monomers. With block copolymers in cyclohexane, a solvent having similar solvent power for each of the monomers, the unperturbed dimensions are fairly close to the linear average of the homopolymer dimensions. At a given temperature the virial coefficients for the block and random copolymers in cyclohexane are different. In 2-ethyoxyethanol, a nonsolvent for polystyrene and a poor solvent for polymethyl methacrylate, the viscosities of block copolymers are anomalous and it is thought likely that the styrene end of the polymer is collapsed (Kotaka *et al.*, 1966). The relationship between viscosity and chain dimensions has not been demonstrated. Of the five polypeptides we tested, 6 M guanidine hydrochloride at 25° is a fair solvent for polylysine and polyglutamate but a nonsolvent for polyalanine, polyphenylalanine, and polytyrosine. It is reasonable then to assume that homopolymers of amino acids having largely hydrocarbon side chains will see 6 M guanidine hydrochloride as a nonsolvent. The deduction of chain dimensions from viscosity and virial coefficient data using equations

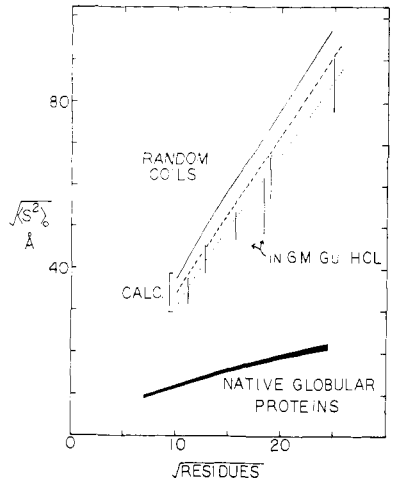

FIGURE 9: The root-mean-square radius of gyration calculated for proteins in native, globular conformation, for random coil proteins (—), and random coil proteins containing 10% (- - -) and 20% (· · ·) of their residues in "collapsed" structures. The experimental $\langle r^2\rangle_0$ values of Tanford and coworkers were converted into $\langle s^2\rangle_0$ by dividing by 6, and are shown as vertical lines.

derived for homopolymers is of unproven reliability. It seems entirely possible that sections of the protein chain rich in hydrocarbon side chains may be collapsed or highly associated. The optical rotation data of Tanford *et al.* (1967b) clearly indicate that there is little if any regularly ordered structure such as an α-helical section in this solvent system but give little information as to how much nonregularly ordered structure exists.

It is surprising how little an effect partial collapse might have on the chain dimensions. Such a chain could be visualized as a string with knots. To simulate and determine the effect of a partially collapsed structure chain dimensions were calculated for a typical random coil protein using eq 6 except that in certain regions in the chain a specified number of residues were forced to take conformations generating a very compact structure. The results are shown in Figure 9, where the root-mean-square radii of gyration are plotted as a function of chain length assuming 0, 5, and 10% of the residues are forced into each of two "knots." Here we have made use of the fact that for chains with considerable random coil behavior $\langle r^2\rangle/\langle s^2\rangle = 6$. Also shown in Figure 9 are the experimentally derived dimensions of Tanford and coworkers, similarly transformed to $\langle s^2\rangle_0^{1/2}$, as well as the calculated radii of gyration for native globular proteins which we assume to have approximately spherical shape. Two facts stand out. First, a fair amount of "knotted" (collapsed or nonregularly ordered) structure will not affect the chain dimensions significantly. Secondly, even though the interpretation of dimensions from viscosity data may be open to some uncertainty, proteins in guanidine hydrochloride have much more expanded structures than native proteins.

Specific binding of guanidine hydrochloride to the polymer could also affect the chain dimensions, but in a manner not easily predicted. Although guanidine hydrochloride is known to interact with some proteins (Noelken and Timasheff, 1967), whether this is domain or specific site binding is unknown.

In summary, the calculated values for unperturbed protein random coils are in reasonable agreement with experimental values for protein dimensions in 6 M guanidine hydrochloride. This must indicate that the rotational isomeric approach considering only short-range interactions is applicable and that proteins in 6 M guanidine hydrochloride are behaving substantially as polymer random coils. Whether or not they retain any nonregularly ordered elements of their native conformation is not likely to be settled from hydrodynamic or other methods yielding average molecular dimensions.

References

Aune, K., Salahuddin, A., Zarlengo, M., and Tanford, C. (1967), *J. Biol. Chem. 242*, 4486.

Brant, D. A., and Flory, P. J. (1965a), *J. Am. Chem. Soc. 87*, 2788.

Brant, D., and Flory, P. (1965b), *J. Am. Chem. Soc. 87*, 2791.

Brant, D., Miller, W., and Flory, P. (1967), *J. Mol. Biol. 23*, 47.

Dayhoff, M., and Eck, R. (1968), Atlas of Protein Sequence and Structure 1967–1968, Silver Spring, Md., The National Biomedical Research Foundation.

DeSantis, P., Giglio, E., Liquori, A. M., and Ripamonti, A. (1965), *Nature 206*, 456.

Flory, P. J. (1953), Principles of Polymer Chemistry, Ithaca, N. Y., Cornell University, p 428.

Froelich, D., and Benoit, H. (1968), *Makromol. Chem. 92*, 224.

Kotaka, T., Ohnuma, H., and Murakami, Y. (1966), *J. Phys. Chem. 70*, 4099.

Kurata, M., and Stockmayer, W. H. (1963), *Fortschr. Hochpolymer Forschung. 3*, 196.

Lapanje, S., and Tanford, C. (1967), *J. Am. Chem. Soc. 89*, 5030.

Leach, S., Némethy, G., and Scheraga, H. (1966), *Biopolymers 4*, 369.

Leung, Y. C., and Marsh, R. E. (1958), *Acta Cryst. 11*, 17.

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

Noelken, M. E., and Timasheff, S. N. (1967), *J. Biol. Chem. 242*, 5080.

Nozaki, Y., and Tanford, C. (1967), *J. Am. Chem. Soc. 89*, 742.

Ooi, T., Scott, R. A., Vanderkooi, G., and Scheraga, H. A. (1967), *J. Chem. Phys. 46*, 4410.

Ramakrishnan, C., and Ramachandran, G. (1965), *Biophys. J. 5*, 909.

Sasisekharan, V. (1962), *in* Collagen, Ramanathan, N., Ed., New York, N. Y., Wiley, p 39.

Schimmel, P., and Flory P. J. (1967), *Proc. Natl. Acad. Sci. U. S. 58*, 52.

Schimmel P., and Flory P. (1968), *J. Mol. Biol. 34*, 105.

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

Stockmayer, W. H., Moore, L. D., Fixman, M., and

Epstein, B. N. (1955), *J. Polym. Sci. 16*, 517.
Tanford, C., Kawahara, K., and Lapanej, S. (1966), *J. Biol. Chem. 241*, 1921.
Tanford, C., Kawahara, K., and Lapanje, S. (1967a), *J. Am. Chem. Soc. 89*, 729.
Tanford, C., Kawahara, K., Lapanje, S., Hooker, Jr., T., Zarlengo, M., Salahuddin, A., Aune, K., and Takagi, T. (1967b), *J. Am. Chem. Soc. 89*, 5023.
Volkenstein, M. V. (1963), Configurational Statistics of Polymeric Chains, New York, N. Y., Interscience.

The Substructure of the Myosin Molecule. Production and Properties of the Alkali Subunits*

Dixie W. Frederiksen and Alfred Holtzer

ABSTRACT: The reaction that produces macromolecular fragments from myosin at moderately high pH (~11.0) has been studied in some detail. The fragments so produced have been separated and studied by means of light scattering, viscometry, optical rotatory dispersion, and ultracentrifugation in a variety of both benign and denaturing solvents.

The investigation shows that myosin consists of four polypeptide chains, two small globular chains of average mass near 30,000 amu, which are held in the globular head of native myosin by secondary forces, and two large chains of average mass near 220,000 amu, which form all of the helical tail and a major part of the globular head of native myosin. Neither the light nor the heavy subunit has biological activity, *i.e.*, adenosine triphosphatase activity or actin-combining ability. Upon back-titration to neutral pH, however, the subunits recombine with resulting full recovery of actin binding power and more than 50% recovery of enzymatic activity.

Recent work on the molecular structure of myosin has resolved controversies and brought substructural details into sharper focus. Numerous investigators have attacked the problem of the molecular mass. The result, in most cases, falls between 470,000 and 530,000 amu (Laki and Carroll, 1955; Holtzer and Lowey, 1956, 1959; Lowey and Cohen, 1962; Holtzer *et al.*, 1962; Mueller, 1964; Tonomura *et al.*, 1966; Richards *et al.*, 1967), which undoubtedly represents satisfactory agreement. Thus, it seems wise to use 500,000 amu. Earlier difficulties that led to controversy have been reviewed (Geiduschek and Holtzer, 1958; Holtzer *et al.*, 1962; Mueller, 1964). The molecular length also seems established at near 1600 Å, the molecule being very asymmetric (Holtzer and Lowey, 1956, 1959; Holtzer *et al.*, 1962; Rice, 1961, 1962; Zobel and Carlson, 1963; Huxley, 1963).

Great impetus was provided by the discovery that brief enzymatic digestion of myosin produces, essentially, two macromolecular fragments: light meromyosin and heavy meromyosin (Gergely, 1950, 1953; Mihalyi and Szent-Györgyi, 1953; Mihalyi, 1953). Heavy meromyosin possesses the parental ATPase[1] activity and actin binding capacity whereas light meromyosin is similar to myosin in its solubility (Szent-Györgyi, 1953). A hint of the arrangement of meromyosins in myosin was first obtained by comparing the experimental angular distribution of light scattered by the myosin molecule with the theory for various models; agreement was only obtained for a top-heavy molecule, *i.e.*, with the heavy meromyosin on one end (Holtzer and Rice, 1957). Determination of the molecular weights and relative amounts of the meromyosins established that each myosin molecule comprises one molecule of each meromyosin (Lowey and Holtzer, 1959). Characterization of the meromyosins provided length estimates showing that light meromyosin and heavy meromyosin are asymmetric and are colinearly joined in myosin (Szent-Györgyi, 1953; Geiduschek and Holtzer, 1958).

Purification of light meromyosin,[2] the finding that it is approximately fully α helical, and determination of its intrinsic viscosity led to the proposal that light meromyosin is a two-chain, α-helical, coiled coil (Szent-Györgyi *et al.*, 1960). Detailed studies of hydrodynamic prop-

* From the Department of Chemistry, Washington University, St. Louis, Missouri. *Received April 25, 1968.* This investigation was supported by U. S. Public Health Service Research Grant RG-5488 from the Division of General Medical Sciences, U. S. Public Health Service. During this investigation D. F. held National Institutes of Health Predoctoral Fellowship 5-Fl-GM-30,352.

[1] See *Biochemistry 5*, 1445 (1966).

[2] Some distinction should be made between earlier preparations of light meromyosin, which contained many contaminants, and the highly α-helical protein ("light meromyosin, fraction 1") which can be obtained from the cruder preparations by alcohol precipitation (Szent-Györgyi *et al.*, 1960.) It will be clear from the context which is meant.