Unperturbed Dimensions for Homopolypeptides and Sequential Copolypeptides Cross-Linked via a Disulfide Bond

Wayne L. Mattice

Department of Biochemistry, Louisiana State University, Baton Rouge, Louisiana 70803. *Received January 3, 1977*

ABSTRACT: Rotational isomeric state theory, in the form appropriate for branched molecules, has been used to calculate the mean-square unperturbed radius of gyration, $\langle s^2 \rangle_0$, for cross-linked polyglycine, poly(L-alanine), poly(L-proline), poly(L-alanyl-D-alanine), poly(L-prolyl-L-prolylglycine), poly(L-prolyl-L-alanylglycine), poly(glycyl-L-alanyl-L-proline), and poly(L-alanyl-L-alanylglycine). The central amino acid residue in each polypeptide chain is replaced by the L-cysteinyl residue involved in cross-link formation. Each cross-linked molecule is considered to contain two trifunctional branch points, the α-carbon atoms of the two L-cysteinyl residues. Random flight statistics provide a poor estimate for g, defined as the ratio of $\langle s^2 \rangle_0$ for branched and linear polypeptides containing the same number of amino acid residues, for molecules of moderate molecular weight. The values of g obtained by random flight statistics and rotational isomeric state theory merge as the molecular weight becomes infinite. Deviations of g from its random flight value correlate with the size of the characteristic ratio, $\langle s^2 \rangle_0 / n_p l_p^2$, for the linear polypeptides. The number of peptide bonds is n_p, and l_p denotes the distance between neighboring α-carbon atoms. Random flight statistics perform better in estimating the change in $\langle s^2 \rangle_0$ accompanying the cross-linking of the two polypeptide chains than it does in the estimation of g.

The unperturbed dimensions of disordered homopolypeptides,[1-10] as well as random[11-13] and sequential[14] copolypeptides, have been the subject of several investigations. Rotational isomeric state theory[15,16] provides the means for relating the unperturbed dimensions to the short-range interactions operative in each polypeptide. The unperturbed dimensions of several disordered proteins, obtained from measurements in concentrated guanidine hydrochloride containing mercaptoethanol,[17,18] are in agreement with results deduced from the short-range interactions acting in simpler polypeptides.[19] The presence of mercaptoethanol is required to reduce any disulfide cross-links present in the proteins.

Cross-linking of disordered polypeptides will lead to alterations in the unperturbed dimensions. Figure 1 represents a simple example, where two polypeptide chains are joined through the formation of a disulfide bond by two cysteinyl residues. This molecule can be viewed as a "4-star," i.e., a molecule in which four branches emanate from a common locus, the cystinyl residue. The customary approach to the evaluation of the unperturbed dimensions of a "4-star" has been to apply random flight statistics. Attention is focused on a parameter, g, which is defined as the ratio of the mean square radius of gyration for the branched and unbranched molecules having the same number of bonds. The result for a "4-star" is given by eq 1, where n_j is the number of bonds in branch j and $N = \Sigma n_j$.[20,21]

$$g = N^{-3} \sum_j (3N{n_j}^2 - 2{n_j}^3) \qquad (1)$$

The applicability of eq 1 to disordered polypeptides and proteins which are crossed-linked as shown in Figure 1 is uncertain. One difficulty concerns the size which N must attain before random flight statistics can be justified. Rotational isomeric state theory[22,23] has shown that the necessary value of N may exceed 1000 for certain simple "4-stars" whose bonds possess a symmetric threefold rotation potential.[24] The direction and size of the error encountered upon the application of random flight statistics to smaller "4-stars" depends on the nature of the short-range interactions present.[24] An additional difficulty in the application of eq 1 to cross-linked proteins is a possible ambiguity in the definition of the unbranched molecule containing an identical number of bonds. This problem arises when the two polypeptide chains have different amino acid sequences.

The present objective is to use rotational isomeric state theory to obtain the unperturbed dimensions for homopolypeptides and sequential copolypeptides cross-linked in the manner shown in Figure 1. The following paper will present results obtained with a variety of cross-linked proteins.

Computations

Configuration Partition Function. The configuration partition function, Z,[25] is given by eq 2.

$$Z = {}_1\mathbf{U}_1^{(n_1)}({}_2\mathbf{U}_1 \ominus {}_3\mathbf{U}_1)[({}_2\mathbf{U}_2^{(n_2-1)}) \otimes \{{}_3\mathbf{U}_2^{(4)}({}_4\mathbf{U}_1 \ominus {}_5\mathbf{U}_1)[({}_4\mathbf{U}_2^{(n_4-1)}) \otimes ({}_5\mathbf{U}_2^{(n_5-1)})]\}] \qquad (2)$$

Each statistical weight matrix, $\mathbf{U}$, is assigned a presubscript which denotes the pertinent branch and a postsubscript which denotes the pertinent bond (branch 3) or virtual bond (branches 1, 2, 4, 5). Branch 3 is that portion of the L-cystinyl residue which lies between the two α-carbon atoms. Notation of the type ${}_1\mathbf{U}_1^{(n_1)}$ denotes the product of n_1 statistical weight matrices, commencing with ${}_1\mathbf{U}_1$.[15] The rectangular matrices denoted by ${}_2\mathbf{U}_1 \ominus {}_3\mathbf{U}_1$ and ${}_4\mathbf{U}_1 \ominus {}_5\mathbf{U}_1$ are defined in ref 22, and $\otimes$ denotes the direct product. In the present case, the expression for Z can be simplified through the application of eq 3, yielding eq 4.[25]

$${}_1\mathbf{U}_i = {}_2\mathbf{U}_j = {}_4\mathbf{U}_j = {}_5\mathbf{U}_j = [1], \quad i < n_1, j > 1 \qquad (3)$$

$$Z = {}_1\mathbf{U}_{n_1}({}_2\mathbf{U}_1 \ominus {}_3\mathbf{U}_1){}_3\mathbf{U}_2^{(4)}({}_4\mathbf{U}_1 \ominus {}_5\mathbf{U}_1) \qquad (4)$$

The elements appearing in the statistical weight matrices in eq 4 have been formulated previously.[25] The resulting configuration partition function was found to provide excellent agreement with the optical activity (and its temperature dependence) exhibited by low molecular weight derivatives of L-cystine and *S*-alkylthio-L-cysteine.

Generator Matrices. The mean square unperturbed radius of gyration of the α-carbon atoms, $\langle s^2 \rangle_0$, can be obtained by application of the procedures presented in ref 23. For this purpose generator matrices must be assigned to each virtual bond in branches 1, 2, 4, and 5, and to each bond in branch 3. However, interatomic distances involving the sulfur atoms and methylene groups of the L-cystinyl residue must not be counted in the evaluation of $\langle s^2 \rangle_0$ if this quantity is to be the mean square unperturbed radius of gyration of the α-carbon atoms. For this reason the generator matrices will be formu-

$$ {}_j\mathbf{S}_i = \begin{bmatrix} 1 & {}_jm_{i-1} & 2({}_jm_{i-1})_j\mathbf{l}_i{}^T({}_j\mathbf{T}_i) & ({}_jm_{i-1})_jl_i{}^2 & ({}_jm_{i-1})({}_jm_i)_jl_i{}^2 \\ 0 & 1 & 2({}_j\mathbf{l}_i{}^T)_j\mathbf{T}_i & {}_jl_i{}^2 & ({}_jm_i)_jl_i{}^2 \\ \mathbf{0} & \mathbf{0} & {}_j\mathbf{T}_i & {}_j\mathbf{l}_i & ({}_jm_i)_j\mathbf{l}_i \\ 0 & 0 & \mathbf{0} & 1 & {}_jm_i \\ 0 & 0 & \mathbf{0} & 0 & 1 \end{bmatrix} \quad (5) $$

lated from eq 5.[23] The transformation matrix (or averaged transformation matrix), bond (or virtual bond) vector, and the transpose of the bond (or virtual bond) vector for bond (or virtual bond) i in branch j are denoted by ${}_j\mathbf{T}_i$, ${}_j\mathbf{l}_i$, ${}_j\mathbf{l}_i{}^T$, respectively. The length of virtual bond vectors is 3.8 Å. Bond lengths, bond angles, and dihedral angles in the L-cystinyl residue are specified in ref 25. The mass, m, is always zero if it is that of a sulfur atom or methylene group in the L-cystinyl residue, thereby rejecting all interatomic distances involving these groups. Unity is always used for the mass if it refers to an α-carbon atom outside the L-cystinyl residue. The mass used for the α-carbon atoms in the L-cystinyl residue is unity unless stated to the contrary. Averaged transformation matrices for L-alanyl and glycyl residues which are not followed by an L-prolyl residue are from Brant et al.[4] If the following residue is L-prolyl, the results obtained by Schimmel and Flory[12] were used instead. The averaged transformation matrix for an L-prolyl residue which is not followed by another L-prolyl residue is also from Schimmel and Flory.[12] Two averaged transformation matrices were used for an L-prolyl residue followed by another L-prolyl residue. One predicts $(\langle r^2\rangle_0/n_p l_p{}^2)_\infty$ is in excess of 100 for poly(L-proline),[5] while the other yields a limiting ratio of 19.[8] The unperturbed mean-square end-to-end distance is denoted by $\langle r^2\rangle_0$, and n_p is the number of virtual bonds of length l_p.

Equation 5 was used to formulate the generator matrices ${}_{ij}\mathbf{F}_k$ and ${}_{hij}\mathbf{F}_k$, according to the procedures summarized in eq 31–38 of ref 23. The presubscripts denote the succession of branches which constitute the pertinent chain, and the postsubscript denotes the bond (or virtual bond) in that chain. The appropriate representation for the statistical weight matrix and transformation matrix associated with a particular bond depends on the sequence of branches which constitute the chain of interest.[23] Proper attention was given to the formulation of the statistical weight matrices and transformation matrices for each bond.

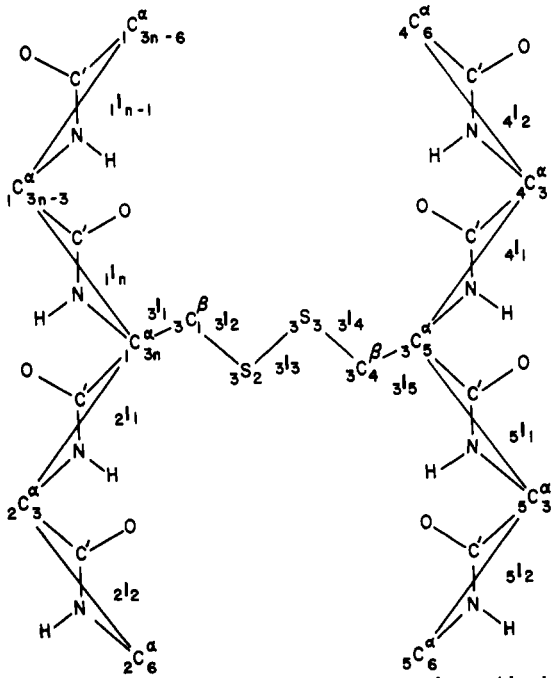

Figure 1. Diagrammatic representation of two polypeptide chains cross-linked via the disulfide bond of an L-cystinyl residue. Nonpeptide hydrogen atoms and the side chains of noncystinyl residues have been omitted.

Radius of Gyration. The mean-square unperturbed radius of gyration of the α-carbon atoms was obtained from eq 6.

$$ \langle s^2\rangle_0 = (n_1 + n_2 + n_4 + n_5 + 2)^{-2}Z^{-1} \times [a_{12} + a_{134} + a_{135} + a_{234} + a_{235} + a_{45} - a_1 - 2a_2 - a_4 - 2a_5 - a_{13} - a_{34}] \quad (6) $$

The first six terms in the brackets evaluate all pertinent interatomic distances. In this process, however, certain of the interatomic distances are evaluated more than once. The last six terms in the brackets correct for this multiple counting. The following expressions for the terms in the brackets are written in a manner which takes full advantage of eq 3, all elements in ${}_2\mathbf{U}_1$, ${}_4\mathbf{U}_1$, and ${}_5\mathbf{U}_1$ being unity, and the sum of the elements in ${}_1\mathbf{U}_{n_1}$ being unity.

$$ a_{12} = Z_{12}\boldsymbol{F}_1{}^{(n_1+n_2)} \quad (7) $$

$$ a_{134} = {}_{134}\boldsymbol{F}_1{}^{(n_1+n_4+5)} \quad (8) $$

$$ a_{135} = {}_{135}\boldsymbol{F}_1{}^{(n_1+n_5+5)} \quad (9) $$

$$ a_{234} = {}_{234}\boldsymbol{F}_1{}^{(n_2+n_4+5)}, \quad {}_{234}m_{n_2} = 0 \quad (10) $$

$$ a_{235} = {}_{235}\boldsymbol{F}_1{}^{(n_2+n_5+5)}, \quad {}_{235}m_{n_2} = {}_{235}m_{n_2+5} = 0 \quad (11) $$

$$ a_{45} = Z_{45}\boldsymbol{F}_1{}^{(n_4+n_5)}, \quad {}_{45}m_{n_4} = 0 \quad (12) $$

$$ a_1 = {}_{12}\boldsymbol{F}_1{}^{(n_1)}{}_{134}\boldsymbol{U}_{n_1+1}{}^{(6)} \quad (13) $$

$$ a_2 = Z_{21}\boldsymbol{F}_1{}^{(n_2-1)} \quad (14) $$

$$ a_4 = Z_{45}\boldsymbol{F}_1{}^{(n_4-1)} \quad (15) $$

$$ a_5 = Z_{54}\boldsymbol{F}_1{}^{(n_5-1)} \quad (16) $$

$$ a_{13} = {}_{134}\boldsymbol{F}_1{}^{(n_1+5)}\ \mathrm{col}(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1) \quad (17) $$

$$ a_{34} = {}_{134}\mathbf{U}_{n_1}{}^{(6)}{}_{134}\boldsymbol{F}_{n_1+6}{}^{(n_4)} \quad (18) $$

The mean-square unperturbed radius of gyration of the α-carbon atoms in non-cross-linked polypeptide chains was computed using the customary methods.[16]

Results

Homopolypeptides. Consider a polypeptide chain which contains $2i + 1$ amino acid residues. The residue in the middle of the chain is an L-cysteinyl residue, while all others are either glycyl, L-alanyl, or L-prolyl residues. The L-cysteinyl residue in this chain reacts with an L-cysteinyl residue in a like chain, forming a cross-linked molecule containing $4i + 2$ amino acid residues. If random flight statistics is to be used, it will be convenient to view the cross-linked polypeptide as a "4-star" in which all four branches contain i amino acid residues. Equation 1 yields $g = \frac{5}{8}$ when $N = 4i$ and $n_j = i$. In the random flight approximation, g will be independent of the molecular weight and independent of the type of amino acid residues present in the polypeptide under consideration.

The results obtained for g using rotational isomeric state theory are shown as curves 1, 2, 4, and 5 in Figure 2. The value of g is a function of the number of amino acid residues, n, and depends upon the nature of the short-range interactions present. The result predicted by random flight statistics is the limit attained by cross-linked polyglycine (curve 1) at high molecular weight. Cross-linked poly(L-alanine) approaches

Table I
Summary of the Behavior of the Unperturbed Dimensions for Various Polypeptides, Arranged in the Order of Increasing Characteristic Ratio

	Not cross-linked		Cross-linked[a]				
Polypeptide	$(\langle s^2\rangle_0/n_p l_p^2)_\infty$	Slope[b]	g slope[c]	f_1 slope[d]	n_{min}[e]	g_{min}[e]	g_{min}/g_∞[e]
L-Ala-D-Ala	0.137[f]	10	17	−16			
Gly	0.36[g]	−1	1	−3			
Pro-Pro-Gly	0.436[f]	−4	−2	−1	42	0.604	0.966
Pro-Ala-Gly	0.471[f]	−3	−3	−1	30	0.591	0.946
Gly-Ala-Pro	0.518[f]	−6	−5	−1	30	0.564	0.903
Ala-Ala-Gly	0.578[f]	−4	−5	−1	18	0.564	0.903
Ala	1.54[g]	−15	−24	7	22	0.391	0.625
Pro	3.17[h]	−64	−60	5	34	0.359	0.574
Pro	19.3[i]	−350	−350	21	34	0.313	0.501

[a] The middle residue in each polypeptide chain is an L-cysteinyl residue. A disulfide bond is formed by the reaction of L-cysteinyl residue on like chains. [b] $d[(\langle s^2\rangle_0/n_p l_p^2)/(\langle s^2\rangle_0/n_p l_p^2)_\infty]/d(1/n_p)$ as n_p goes to infinity. [c] $d(g/g_\infty)/d(1/n)$ as n goes to infinity. [d] $d(f_1/f_{1,\infty})/d(1/n)$ as n goes to infinity. [e] Characteristics of the minimum in a plot of g vs. n. [f] Obtained from the extrapolation of $\langle s^2\rangle_0/n_p l_p^2$ vs. $1/n_p$ to $1/n_p = 0$. [g] Reference 4. [h] Calculated as $(1/6)[(\mathbf{E} + \langle\mathbf{T}\rangle)(\mathbf{E} - \langle\mathbf{T}\rangle)^{-1}]_{11}$ from the $\gamma 2$ transformation matrix in ref 8. [i] Reference 5.

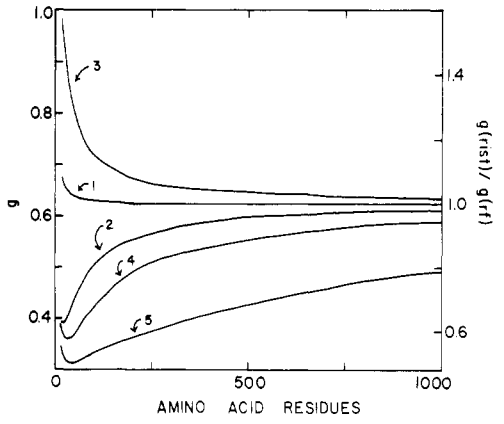

Figure 2. Dependence of g, calculated using rotational isomeric state theory, on the total number of amino acid residues in polypeptides containing one interchain disulfide cross-link. All polypeptide chains contain the same number of amino acid residues, with the middle residue in each chain being the one involved in cross-link formation. The other amino acid residues are glycyl (curve 1), L-alanyl (curve 2), L-prolyl (curves 4 and 5), or alternating L-alanyl-D-alanyl (curve 3). The averaged transformation matrix used for the L-prolyl residue followed by another L-prolyl residue was from reference 8 for curve 4 and from reference 5 for curve 5. Random flight statistics predict $g = 5/8$. The scale on the right shows the ratio of the g obtained using rotational isomeric state theory to that obtained using random flight statistics.

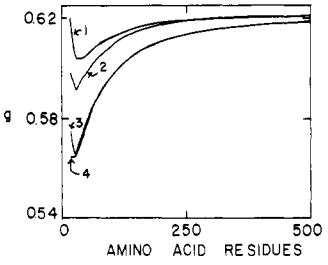

Figure 3. Dependence of g, calculated using rotational isomeric state theory, on the total number of amino acid residues in collagen-like sequential copolypeptides containing one intermolecular disulfide cross-link. The copolypeptides are poly(L-prolyl-L-prolylglycine) (curve 1), poly(L-prolyl-L-alanylglycine) (curve 2), poly(glycyl-L-alanyl-L-proline) (curve 3), and poly(L-alanyl-L-alanylglycine) (curve 4). The middle amino acid residue in each polypeptide chain is replaced by the L-cysteinyl residue involved in cross-link formation. The value of g predicted by random flight statistics corresponds to the top of the figure.

this limit, but has not quite reached it when n is 1000. An even slower approach to this limit is found with cross-linked poly(L-proline). Extrapolation of a plot of g vs. $1/n$ to $1/n = 0$ (not shown) reveals that $g_\infty = 5/8$ for cross-linked poly(L-alanine) and poly(L-proline).

The curves in Figure 2 for cross-linked poly(L-alanine) and poly(L-proline), but not cross-linked polyglycine, exhibit minima when the number of amino acid residues is in the range 22–34. Consequently the sign of dg/dn is positive upon the approach to the asymptotic limit for cross-linked poly(L-alanine) and poly(L-proline), but it is negative for cross-linked polyglycine. The characteristics of the minima are summarized in Table I.

Sequential Copolypeptides. Similar computations were carried out for several cross-linked sequential copolypeptides. One of these is obtained from the alternating copolypeptide poly(L-alanyl-D-alanine), with the middle L-alanyl residue replaced by the L-cysteinyl residue involved in cross-link formation. The results obtained from rotational isomeric state theory, shown as curve 3 in Figure 2, share with cross-linked polyglycine the absence of a minimum, but the rate of approach to the asymptotic limit is reminiscent of cross-linked poly(L-alanine).

Results for four cross-linked sequential copolypeptides which contain a glycyl residue at every third position are shown in Figure 3. These molecules are obtained from cross-linking poly(glycyl-L-alanyl-L-proline), poly(L-prolyl-L-alanylglycine), poly(L-alanyl-L-alanylglycine), or poly(L-prolyl-L-prolylglycine), with the middle amino acid residue in each polypeptide chain being replaced by the L-cysteinyl residue involved in cross-link formation. In each case a minimum is obtained, and the asymptotic limit is approached quickly. The value of g_∞ is $5/8$ for all of the sequential copolypeptides.

Correlations with the Limiting Characteristic Ratio. The presence or absence of a minimum in a plot of g vs. n, the depth of the minimum, and the rate of approach to g_∞ can each be related to properties of the uncross-linked polypeptides. Limiting characteristic ratios, $(\langle s^2\rangle_0/n_p l_p^2)_\infty$, for the uncross-linked polypeptides studied are shown in Table I. They span two orders of magnitude. A convenient manner of expressing the rate at which the limiting characteristic ratios are approached is as $d[(\langle s^2\rangle_0/n_p l_p^2)/(\langle s^2\rangle_0/n_p l_p^2)_\infty]/d(1/n_p)$ at $1/n_p = 0$. The pertinent curves are presented in Figure 4,

(7) W. L. Mattice and J. T. Lo, *Macromolecules*, **5**, 734 (1972).
(8) W. L. Mattice, K. Nishikawa, and T. Ooi, *Macromolecules*, **6**, 443 (1973).
(9) T. Ooi, D. S. Clark, and W. L. Mattice, *Macromolecules*, **7**, 447 (1974).
(10) D. S. Clark and W. L. Mattice, *Macromolecules*, in press.
(11) W. G. Miller, D. A. Brant, and P. J. Flory, *J. Mol. Biol.*, **23**, 67 (1967).
(12) P. R. Schimmel and P. J. Flory, *J. Mol. Biol.*, **34**, 105 (1968).
(13) S. P. Rao, D. E. Carlstrom, and W. G. Miller, *Biochemistry*, **13**, 943 (1974).
(14) W. L. Mattice and L. Mandelkern, *Biochemistry*, **10**, 1934 (1971).
(15) P. J. Flory, "Statistical Mechanics of Chain Molecules", Interscience, New York, N.Y., 1969.
(16) P. J. Flory, *Macromolecules*, **7**, 381 (1974).
(17) C. Tanford, K. Kawahara, and S. Lapanje, *J. Am. Chem. Soc.*, **89**, 729 (1967).
(18) S. Lapanje and C. Tanford, *J. Am. Chem. Soc.*, **89**, 5030 (1967).
(19) W. G. Miller and C. V. Goebel, *Biochemistry*, **7**, 3925 (1968).
(20) B. H. Zimm and W. H. Stockmayer, *J. Chem. Phys.*, **17**, 1301 (1949).
(21) T. A. Orofino, *Polymer*, **2**, 305 (1961).
(22) W. L. Mattice, *Macromolecules*, **8**, 644 (1975).
(23) W. L. Mattice, *Macromolecules*, **9**, 48 (1976).
(24) W. L. Mattice and D. K. Carpenter, *Macromolecules*, **9**, 53 (1976).
(25) W. L. Mattice, *J. Am. Chem. Soc.*, **99**, 2325 (1977).
(26) P. M. Gallop, O. O. Blumenfeld, and S. Seifter, *Annu. Rev. Biochem.*, **41**, 617 (1972).
(27) A. D. McLachlan, M. Stewart, and L. B. Smillie, *J. Mol. Biol.*, **98**, 281 (1975).
(28) P. Johnson and L. B. Smillie, *Biochem. Biophys. Res. Commun.*, **64**, 1316 (1975).
(29) L. Pauling, R. B. Corey, and H. R. Branson, *Proc. Natl. Acad. Sci. U.S.A.*, **37**, 205 (1951).
(30) W. G. Miller and P. J. Flory, *J. Mol. Biol.*, **15**, 298 (1966).

Unperturbed Dimensions of Disordered Proteins Containing an Interchain Disulfide Cross-Link

Wayne L. Mattice

Department of Biochemistry, Louisiana State University, Baton Rouge, Louisiana 70803. *Received January 3, 1977*

ABSTRACT: Mean-square unperturbed radii of gyration, $\langle s^2 \rangle_0$, have been calculated for several proteins cross-linked via an interchain disulfide bond. Thirty different polypeptide chains were used. Characteristic ratios tend to be smaller for cross-linked proteins than for the uncross-linked chains, although exceptions to this generalization do exist. Random flight statistics tend to overestimate the value of g, defined as the ratio of $\langle s^2 \rangle_0$ for the cross-linked protein to $\langle s^2 \rangle_0$ for an analogous linear polypeptide chain containing the same number of amino acid residues. The parameter f_i, defined as the ratio of the $\langle s^2 \rangle_0$ for the ith uncross-linked polypeptide chain and the cross-linked protein, is usually more accurately estimated by random flight statistics than is g. When the cross-link connects two chains of identical amino acid sequence, the values of f_i obtained via random flight statistics are within 6% of those provided by rotational isomeric state theory.

The ordered structures adopted by proteins in their native states can be completely disrupted by suitable changes in solvent composition.[1] Concentrated guanidine hydrochloride containing a disulfide bond reducing agent, such as mercaptoethanol or dithiothreitol, is the most thoroughly characterized such system.[2-5] The reducing agent serves to rupture any disulfide cross-links present, thereby permitting characterization of the disordered polypeptides by the methods customarily applied to linear polymers. Presumably the ordered structures would also be disrupted if guanidine hydrochloride were used without a disulfide bond reducing agent, but the intact cross-links would bring about alterations in the dimensions of the molecules. Our present objective is to examine the conformational consequences arising from the presence of a single intact interchain disulfide cross-link in denatured proteins. Calculations have been performed using rotational isomeric state theory[6,7] in the form appropriate for branched molecules.[8,9]

Computations

The formulation of the configuration partition function, generator matrices, and expression for the mean-square unperturbed radius of gyration of the α-carbon atoms can be found in ref 10 and 11. All amino acid residues, except glycyl, L-prolyl, and those L-cysteinyl residues involved in cross-link formation, were treated as L-alanyl residues in the formulation of the generator matrices. The $\gamma 2$ transformation matrix from ref 12 was used for those L-prolyl residues followed by another L-prolyl residue. Polypeptide chains used, along with literature citations for the amino acid sequences and the number of amino acid residues, n, are collected in Table I. Cross-linked molecules contain one interchain disulfide bond. The disulfide bonds considered do not necessarily occur in the proteins in their native states.

Results and Discussion

Characteristic Ratios. The characteristic ratios for cross-linked and uncross-linked polypeptide chains will be defined as $\langle s^2 \rangle_0/(n-1)l_p^2$, where $\langle s^2 \rangle_0$ is the unperturbed mean-square radius of gyration of the n α-carbon atoms in the molecule. For uncross-linked polypeptide chains this definition is equivalent to $\langle s^2 \rangle_0/n_p l_p^2$, where n_p is the number of virtual bonds. The length of the virtual bond, l_p, is 3.8 Å.[13]

Figure 1 presents the characteristic ratios obtained for the molecules considered. Characteristic ratios for the uncross-linked polypeptide chains are denoted by squares. They range from a low of 0.81 (α-trypsin C chain) to a high of 1.41 (α-tropomyosin). All but three polypeptide chains with $n \geq 100$ have a characteristic ratio which lies between 1.00 and 1.29. The exceptions are cytochrome c, globin, and α-tropomyosin. The incorporation of a few glycyl or L-prolyl residues in a poly(L-alanine) chain reduces its characteristic ratio.[14,15] Therefore the high characteristic ratio for α-tropomyosin arises because only three glycyl residues, and no L-prolyl residues, occur out of a total of 284 amino acid residues. In contrast, 16–21% of the amino acid residues in cytochrome c and globin are glycyl or L-prolyl, leading to the unusually low characteristic ratios for these proteins. The average characteristic ratio for those polypeptide chains having $n \geq 100$ is 1.13, with a standard deviation of 0.11.

Characteristic ratios for the polypeptide chains cross-linked via a disulfide bond are also shown in Figure 1. Filled circles denote cases where both polypeptide chains have the same amino acid sequence. Multiple cross-linked forms are possible

with the same number of amino acid residues as the cross-linked one. However, it would be possible to study the uncross-linked α-tropomyosin molecule, and compare its dimensions to those of the cross-linked α-tropomyosin. It will be convenient to express the results as f_i, defined in eq 19.

$$f_i = \frac{\langle s^2 \rangle \text{ for the } i\text{th uncross-linked polypeptide chain}}{\langle s^2 \rangle \text{ for the cross-linked polypeptide}} \quad (19)$$

The subscript is required to encompass cases where nonidentical polypeptide chains are cross-linked. Random flight results for f_i can be obtained from eq 1. If branches 1 and 2 constitute polypeptide chain 1,

$$f_1 = (n_1 + n_2)N^2[\sum(3Nn_j^2 - 3n_j^3)]^{-1} \quad (20)$$

When all n_j are equal, $f_1 = \frac{4}{5}$.

Results obtained via rotational isomeric state theory for the homopolypeptides and poly(L-alanyl-D-alanine) are shown in Figure 7. The presentation is as the ratio of f_1 computed via rotational isomeric state theory to the result obtained from random flight statistics. As with g, f_1 is found to be a function of molecular weight and to depend on the nature of the short-range interactions present. The result predicted by random flight statistics is attained by polyglycine in Figure 7. Higher molecular weight is required for f_1 to attain its asymptotic limit in the other cases.

The rate of approach of f_1 to the random flight value is conveniently expressed as $d(f_1/f_{1,\infty})/d(1/n)$, evaluated at $1/n = 0$. Results are summarized in Table I. It is of interest to compare the rates at which f_1 and g approach their asymptotic limits. The pertinent quantities are $d(f_1/f_{1,\infty})/d(1/n)$ and $d(g/g_\infty)/d(1/n)$, the "f_1 slope" and "g slope" of Table I. For all polypeptides other than polyglycine, f_1 nears its asymptotic limit more quickly than does g. The difference in rates exceeds an order of magnitude for poly(L-proline). This phenomenon is readily apparent from a comparison of Figures 2 and 7. Equation 20 will generally be more accurate than eq 1 when applied to cross-linked homopolypeptides and sequential copolypeptides in which both chains contain the same number of amino acid residues. The better performance of eq 20 arises because it involves comparisons of molecules in which all polypeptide chains contain the same number of amino acid residues, and hence tends to suppress any influence of a molecular weight dependence different from that arising in random flight statistics. In contrast, eq 1 entails a comparison of molecules in which the number of amino acid residues per polypeptide chain differs.

Relationship to Cross-Linked α Helices. Dimensions for cross-linked poly(L-alanine) will generally depend upon whether the individual polypeptide chains exist in the random coil state or as rigid α helices. The mean-square radius of gyration for the cross-linked helices is given by eq 21 if the

$$\langle s^2 \rangle = \frac{(\frac{1}{12})(L_1^4 + L_2^4) + L_1L_2[L_1^2F_1(F_1 - 1) + L_2^2F_2(F_2 - 1) + (\frac{1}{3})(L_1^2 + L_2^2)]}{(L_1 + L_2)^2} \quad (21)$$

cross-link acts as a free joint and the helices are represented by infinitely thin rods. Intersection of the rods occurs at fractions F_1 and F_2 from an arbitrarily chosen end of helices with lengths L_1 and L_2, respectively. The case of current interest corresponds to $L_1 = L_2$ and $F_1 = F_2 = \frac{1}{2}$, yielding $\langle s^2 \rangle = L_1^2/12$. This combination of L_1, L_2, F_1, and F_2 produces the same value of $\langle s^2 \rangle$ for any orientation of the helices, including the case where they are colinear. The values of f_1 and g are unity and $\frac{1}{4}$, respectively.

Figure 8 presents the ratio of $\langle s^2 \rangle_0$ for cross-linked poly(L-alanine) in the random coil and α-helical states. Each helix is assigned a length of 1.50 Å per amino acid residue.[29] Equivalent information is shown for linear poly(L-alanine). Linear molecules of low degree of polymerization have larger

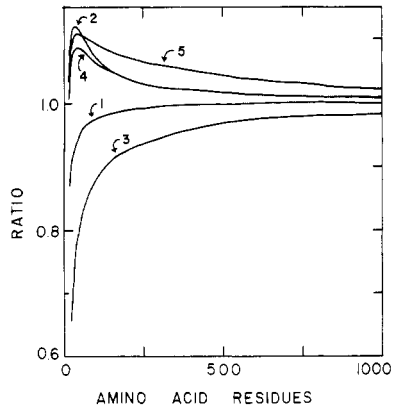

Figure 7. Ratio of the values of f_1 obtained from rotational isomeric state theory and using random flight statistics for cross-linked polyglycine (curve 1), poly(L-alanine) (curve 2), poly(L-alanyl-D-alanine) (curve 3), and poly(L-proline) (curves 4 and 5) as a function of the number of amino acid residues in the cross-linked polypeptide. All polypeptide chains contain the same number of amino acid residues, with the middle residue in each case being the one involved in cross-link formation. The averaged transformation matrix used for the L-prolyl residue followed by another L-prolyl residue was from ref 8 for curve 4 and from ref 5 for curve 5. Random flight statistics predict $f_1 = \frac{4}{5}$.

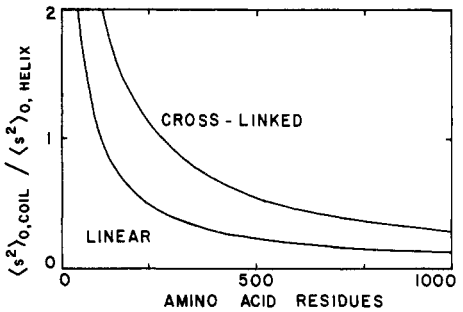

Figure 8. Ratio of $\langle s^2 \rangle_0$ for random coil and helical poly(L-alanine). The cross-link exists between the central amino acid residue (L-cysteinyl) in each polypeptide chain.

dimensions as random coils than as α helices, while the reverse holds true at high degree of polymerization.[30] The same trends occur for cross-linked poly(L-alanine). Comparison of results for molecules containing the same total number of L-alanyl residues reveals that $\langle s^2 \rangle_{0,\text{coil}}/\langle s^2 \rangle_{0,\text{helix}}$ is larger by a factor of 2.0–2.5 for the cross-linked polypeptide. If the comparison is instead made between molecules in which the individual polypeptides chains are of the same degree of polymerization, $\langle s^2 \rangle_{0,\text{coil}}/\langle s^2 \rangle_{0,\text{helix}}$ is only 20–25% larger for the cross-linked poly(L-alanine).

Acknowledgment. Supported in part by National Science Foundation Grant BMS 72-02416 A01 and in part by a fellowship from the John Simon Guggenheim Memorial Foundation.

References and Notes

(1) D. A. Brant and P. J. Flory, *J. Am. Chem. Soc.*, **87**, 663 (1965).
(2) D. A. Brant and P. J. Flory, *J. Am. Chem. Soc.*, **87**, 2788 (1965).
(3) D. A. Brant and P. J. Flory, *J. Am. Chem. Soc.*, **87**, 2791 (1965).
(4) D. A. Brant, W. G. Miller, and P. J. Flory, *J. Mol. Biol.*, **23**, 47 (1967).
(5) P. R. Schimmel and P. J. Flory, *Proc. Natl. Acad. Sci. U.S.A.*, **58**, 52 (1967).
(6) W. L. Mattice and L. Mandelkern, *J. Am. Chem. Soc.*, **93**, 1769 (1971).

(7) W. L. Mattice and J. T. Lo, *Macromolecules*, **5**, 734 (1972).
(8) W. L. Mattice, K. Nishikawa, and T. Ooi, *Macromolecules*, **6**, 443 (1973).
(9) T. Ooi, D. S. Clark, and W. L. Mattice, *Macromolecules*, **7**, 447 (1974).
(10) D. S. Clark and W. L. Mattice, *Macromolecules*, in press.
(11) W. G. Miller, D. A. Brant, and P. J. Flory, *J. Mol. Biol.*, **23**, 67 (1967).
(12) P. R. Schimmel and P. J. Flory, *J. Mol. Biol.*, **34**, 105 (1968).
(13) S. P. Rao, D. E. Carlstrom, and W. G. Miller, *Biochemistry*, **13**, 943 (1974).
(14) W. L. Mattice and L. Mandelkern, *Biochemistry*, **10**, 1934 (1971).
(15) P. J. Flory, "Statistical Mechanics of Chain Molecules", Interscience, New York, N.Y., 1969.
(16) P. J. Flory, *Macromolecules*, **7**, 381 (1974).
(17) C. Tanford, K. Kawahara, and S. Lapanje, *J. Am. Chem. Soc.*, **89**, 729 (1967).
(18) S. Lapanje and C. Tanford, *J. Am. Chem. Soc.*, **89**, 5030 (1967).
(19) W. G. Miller and C. V. Goebel, *Biochemistry*, **7**, 3925 (1968).
(20) B. H. Zimm and W. H. Stockmayer, *J. Chem. Phys.*, **17**, 1301 (1949).
(21) T. A. Orofino, *Polymer*, **2**, 305 (1961).
(22) W. L. Mattice, *Macromolecules*, **8**, 644 (1975).
(23) W. L. Mattice, *Macromolecules*, **9**, 48 (1976).
(24) W. L. Mattice and D. K. Carpenter, *Macromolecules*, **9**, 53 (1976).
(25) W. L. Mattice, *J. Am. Chem. Soc.*, **99**, 2325 (1977).
(26) P. M. Gallop, O. O. Blumenfeld, and S. Seifter, *Annu. Rev. Biochem.*, **41**, 617 (1972).
(27) A. D. McLachlan, M. Stewart, and L. B. Smillie, *J. Mol. Biol.*, **98**, 281 (1975).
(28) P. Johnson and L. B. Smillie, *Biochem. Biophys. Res. Commun.*, **64**, 1316 (1975).
(29) L. Pauling, R. B. Corey, and H. R. Branson, *Proc. Natl. Acad. Sci. U.S.A.*, **37**, 205 (1951).
(30) W. G. Miller and P. J. Flory, *J. Mol. Biol.*, **15**, 298 (1966).

Unperturbed Dimensions of Disordered Proteins Containing an Interchain Disulfide Cross-Link

Wayne L. Mattice

Department of Biochemistry, Louisiana State University, Baton Rouge, Louisiana 70803. *Received January 3, 1977*

ABSTRACT: Mean-square unperturbed radii of gyration, $\langle s^2 \rangle_0$, have been calculated for several proteins cross-linked via an interchain disulfide bond. Thirty different polypeptide chains were used. Characteristic ratios tend to be smaller for cross-linked proteins than for the uncross-linked chains, although exceptions to this generalization do exist. Random flight statistics tend to overestimate the value of g, defined as the ratio of $\langle s^2 \rangle_0$ for the cross-linked protein to $\langle s^2 \rangle_0$ for an analogous linear polypeptide chain containing the same number of amino acid residues. The parameter f_i, defined as the ratio of the $\langle s^2 \rangle_0$ for the ith uncross-linked polypeptide chain and the cross-linked protein, is usually more accurately estimated by random flight statistics than is g. When the cross-link connects two chains of identical amino acid sequence, the values of f_i obtained via random flight statistics are within 6% of those provided by rotational isomeric state theory.

The ordered structures adopted by proteins in their native states can be completely disrupted by suitable changes in solvent composition.[1] Concentrated guanidine hydrochloride containing a disulfide bond reducing agent, such as mercaptoethanol or dithiothreitol, is the most thoroughly characterized such system.[2-5] The reducing agent serves to rupture any disulfide cross-links present, thereby permitting characterization of the disordered polypeptides by the methods customarily applied to linear polymers. Presumably the ordered structures would also be disrupted if guanidine hydrochloride were used without a disulfide bond reducing agent, but the intact cross-links would bring about alterations in the dimensions of the molecules. Our present objective is to examine the conformational consequences arising from the presence of a single intact interchain disulfide cross-link in denatured proteins. Calculations have been performed using rotational isomeric state theory[6,7] in the form appropriate for branched molecules.[8,9]

Computations

The formulation of the configuration partition function, generator matrices, and expression for the mean-square unperturbed radius of gyration of the α-carbon atoms can be found in ref 10 and 11. All amino acid residues, except glycyl, L-prolyl, and those L-cysteinyl residues involved in cross-link formation, were treated as L-alanyl residues in the formulation of the generator matrices. The $\gamma 2$ transformation matrix from ref 12 was used for those L-prolyl residues followed by another L-prolyl residue. Polypeptide chains used, along with literature citations for the amino acid sequences and the number of amino acid residues, n, are collected in Table I. Cross-linked molecules contain one interchain disulfide bond. The disulfide bonds considered do not necessarily occur in the proteins in their native states.

Results and Discussion

Characteristic Ratios. The characteristic ratios for cross-linked and uncross-linked polypeptide chains will be defined as $\langle s^2 \rangle_0/(n-1)l_p^2$, where $\langle s^2 \rangle_0$ is the unperturbed mean-square radius of gyration of the n α-carbon atoms in the molecule. For uncross-linked polypeptide chains this definition is equivalent to $\langle s^2 \rangle_0/n_p l_p^2$, where n_p is the number of virtual bonds. The length of the virtual bond, l_p, is 3.8 Å.[13]

Figure 1 presents the characteristic ratios obtained for the molecules considered. Characteristic ratios for the uncross-linked polypeptide chains are denoted by squares. They range from a low of 0.81 (α-trypsin C chain) to a high of 1.41 (α-tropomyosin). All but three polypeptide chains with $n \geq 100$ have a characteristic ratio which lies between 1.00 and 1.29. The exceptions are cytochrome c, globin, and α-tropomyosin. The incorporation of a few glycyl or L-prolyl residues in a poly(L-alanine) chain reduces its characteristic ratio.[14,15] Therefore the high characteristic ratio for α-tropomyosin arises because only three glycyl residues, and no L-prolyl residues, occur out of a total of 284 amino acid residues. In contrast, 16–21% of the amino acid residues in cytochrome c and globin are glycyl or L-prolyl, leading to the unusually low characteristic ratios for these proteins. The average characteristic ratio for those polypeptide chains having $n \geq 100$ is 1.13, with a standard deviation of 0.11.

Characteristic ratios for the polypeptide chains cross-linked via a disulfide bond are also shown in Figure 1. Filled circles denote cases where both polypeptide chains have the same amino acid sequence. Multiple cross-linked forms are possible