(8) C. Holt, T. G. Parker, and D. G. Dalgleish, *Biochim. Biophys. Acta*, **400**, 283 (1975).
(9) S. W. Provencher, J. Hendrix, L. De Maeyer, and N. Paulussen, *J. Chem. Phys.*, **69**, 4273 (1978).
(10) B. Chu and Es. Gulari, *Macromolecules*, **12**, 445 (1979).
(11) Er. Gulari, Es. Gulari, Y. Tsunashima, and B. Chu, *Polymer*, **20**, 347 (1979).
(12) C. C. Han and F. L. McCrackin, *Polymer*, **20**, 427 (1979).
(13) J. Raczek and G. Meyerhoff, *Ber. Bunsenges. Phys. Chem.*, **83**, 381 (1979).
(14) J. C. Selser, *Macromolecules*, **12**, 909 (1979).
(15) Y. Tagami and R. Pecora, *J. Chem. Phys.*, **51**, 3293 (1969).
(16) R. Pecora and Y. Tagami, *J. Chem. Phys.*, **51**, 3298 (1969).
(17) S. R. Aragon and R. Pecora, *J. Chem. Phys.*, **64**, 2395 (1976).
(18) J. Raczek and G. Meyerhoff, *Eur. Polym. J.*, **13**, 539 (1977).
(19) B. Appelt and G. Meyerhoff, *Macromolecules*, in press.
(20) G. Meyerhoff, H. Hack, and J. Raczek, *J. Polym. Sci.*, **61**, 169 (1977).
(21) H. Hack and G. Meyerhoff, *Makromol. Chem.*, **179**, 2475 (1978).
(22) C. de Boor and J. C. Rice, "Cubic Spline Approximations at I-Fixed Knots", Computer Sciences Dept. TR20, Purdue University, April 1968.
(23) Since both types of light scattering techniques are applied in our laboratory, we prefer to name them according to the measuring process: frequency-averaging LS and frequency-analyzing LS for classical LS and dynamic LS, respectively. The latter takes dynamic effects into account.
We think that most of the other names for frequency-analyzing LS should be avoided. For instance, the terms inelastic, quasi-elastic, and elastic cannot unequivocally be connected with the two LS types. There are good reasons to call both quasi-elastic or perhaps both quasi-inelastic. Moreover, the term laser LS is not reasonable since, if MWDs are to be characterized, a laser should be used for FAV LS as well as FAN LS. Furthermore, FAN LS is not a form of spectroscopy.

# Helix End Effects in Block Copolypeptides, Proteins, and Protein–Detergent Complexes

**Wayne L. Mattice,* Govind Srinivasan, and German Santiago**

*Department of Chemistry, Louisiana State University, Baton Rouge, Louisiana 70803.*
*Received March 4, 1980*

ABSTRACT: Average configuration-dependent properties (mean-square radius of gyration, mean-square end-to-end distance, helical content, helix probability profile, and average number of amino acid residues in a helical segment) have been evaluated for unperturbed partially helical polypeptides by using two weighting schemes. The weighting schemes differ with regard to whether end effects in a helical segment are assigned in like manner to amino acid residues at each end or are instead associated with only one amino acid residue. Molecules considered are block copolypeptides and specific sequence copolypeptides with amino acid sequences corresponding to those found in 44 proteins. Without exception, helix formation in the proteins is found to exhibit greater cooperativity if amino acid residues at each end of a helical segment contribute to the end effects. The helix probability profile for proteins is also found to be much smoother with this weighting scheme. Block copolypeptides show more dramatic changes in helicity, helix probability profile, and average number of amino acid residues in a helical segment. Unperturbed dimensions for both proteins and block copolypeptides are nearly independent of the weighting scheme adopted. Modification of the statistical weights used for arginyl, histidyl, and lysyl residues can produce helicities and unperturbed dimensions in reasonable agreement with helicities and dimensions deduced from circular dichroism spectra and viscosity measurements of complexes formed by reduced bovine serum albumin and sodium dodecyl sulfate.

Matrix methods are widely used to compute statistical mechanical averages of configuration-dependent properties for unperturbed macromolecules.[1] A statistical weight matrix, $\mathbf{U}_i$, is formulated for each bond (or virtual bond) in the main chain. The configuration partition function, $Z$, for an unperturbed linear chain molecule is extracted from the sequential product of these statistical weight matrices. Appropriate modification of $\mathbf{U}_i$ permits computation of the average occupancy of a rotational state accessible to bond $i$. Dimensional properties, such as the mean-square end-to-end distance, $\langle r^2 \rangle$, and mean-square radius of gyration, $\langle s^2 \rangle$, are obtained by using supermatrices formulated from the statistical weight matrices and generator matrices appropriate for the molecule in a specified configuration.[1] These procedures can be rigorously extended to encompass treatment of branched molecules.[2,3]

Application of matrix methods to partially helical homopolypeptides is most easily achieved by using a $2 \times 2$ statistical weight matrix[4,5]

$$\mathbf{U}_i = \begin{bmatrix} 1 & \sigma s \\ 1 & s \end{bmatrix}_i \qquad (1)$$

Columns index the state of amino acid residue $i$, rows index the state of amino acid residue $i - 1$, and the order of indexing is coil (c), helix (h). The statistical weight for amino acid residue $i$ is unity if it is not in a helical state. Its statistical weight is $\sigma s$ if it initiates a sequence of helical amino acid residues and $s$ if it propagates an existing helix. The configuration partition function for a homopolypeptide containing $n$ amino acid residues is

$$Z = \text{row } (1, 0)\ \mathbf{U}_1\ \mathbf{U}_2 \ldots \mathbf{U}_n \text{ col } (1, 1) \qquad (2)$$

The serial product of statistical weight matrices can be replaced by $\mathbf{U}^n$ for a homopolypeptide. Statistical weight matrices in eq 1 can be combined with suitably constituted generator matrices in order to evaluate the mean-square unperturbed dimensions of partially helical homopolypeptides, both in the absence[5] and in the presence[6] of interchain cross-links. Asymmetry of the spatial distribution for partially helical homopolypeptides has been characterized[7] by using a priori and conditional probabilities extracted from the $2 \times 2$ statistical weight matrix in eq 1.

The subscript $i$ for $\mathbf{U}$ must be retained if the treatment concerns a copolypeptide. If eq 1 is used for $\mathbf{U}_i$, a helical segment starting at amino acid residue $i$ and extending to amino acid residue $j$ will have a statistical weight given by $\sigma_i s_i s_{i+1} \ldots s_{j-1} s_j$. The burden of helix initiation is placed solely on amino acid residue $i$, while none of that burden

0024-9297/80/2213-1254$01.00/0 

is shared by the amino acid residue at the other end of the helical segment.

Amino acid residues $i$ and $j$ both contribute to end effects if the statistical weight for a helical sequence is $\sigma_i^{1/x} s_i s_{i+1} \ldots s_{j-1} s_j \sigma_j^{(x-1)/x}$. In order to explore the general consequences of the change in weighting scheme, attention will be confined to the case where $x = 2$, causing amino acid residue $i$ and $j$ to contribute in a similar manner to the end effects. No change in weighting would result for a homopolypeptide. For a copolypeptide, however, the two weighting schemes produce different results unless amino acid residues $i$ and $j$ happen to have the same value for $\sigma$. Only the modified scheme always gives a statistical weight for a helical segment which is independent of whether indexing commences at the amino or carboxyl terminal end of the polypeptide chain. Matrix schemes which assign end effects to each end of a helical segment have been proposed.[8-10] A 3 × 3 statistical weight matrix using this weighting scheme is shown in eq 3. Rows index

$$\mathbf{U}_i = \begin{array}{c} \\ \text{c(c or h)} \\ \text{hc} \\ \text{hh} \end{array} \begin{array}{c} \begin{array}{ccc} \text{c(c or h)} & \text{hc} & \text{hh} \end{array} \\ \begin{bmatrix} 1 & 0 & \sigma^{1/2}s \\ 1 & 0 & 0 \\ 0 & \sigma^{1/2}s & s \end{bmatrix}_i \end{array} \quad (3)$$

$$Z = \text{row } (1, 0, 0)\ \mathbf{U}_1\ \mathbf{U}_2 \ldots \mathbf{U}_n \text{ col } (1, 1, 0) \quad (4)$$

states of amino acid residues $i - 1$ and $i$, while columns index states of amino acid residues $i$ and $i + 1$. A helix consisting of only one helical amino acid residue is not allowed when the second element in the top row is zero. This element would be $\sigma s$ if there were to be no prejudice against helices containing only one amino acid residue.

Average properties calculated for copolypeptides may depend on whether the computation is based on the 2 × 2 or 3 × 3 statistical weight matrix. The average properties of interest here will be $\langle r^2 \rangle$, $\langle s^2 \rangle$, the fraction $f$ of amino acid residues which are helical, the probability $f_i$ the $i$th amino acid residue is helical, and the average number $\nu$ of amino acid residues in a helical segment. Consequences of statistical weight matrix size will be explored here for three situations of interest. The first concerns an AB block copolypeptide in which amino acid residues have different values for $\sigma$. The next case is specific-sequence copolypeptides (proteins) in which $\sigma$ and $s$ are those found,[11-24] or estimated, for amino acid residues in aqueous random copolypeptides. The final case is proteins whose $\sigma$ and $s$ have been modified in the manner required to account for properties of complexes formed by binding large amounts of the detergent sodium dodecyl sulfate.[25]

## Calculations

Configuration partition functions evaluated using 2 × 2 and 3 × 3 statistical weight matrices are given in eq 2 and 4, respectively. The probability of a helical state at amino acid residue $i$ is

$$f_i = Z_i' Z^{-1} \quad (5)$$

where $Z_i'$ differs from $Z$ only in substitution of $\mathbf{U}_i'$ for $\mathbf{U}_i$. Matrix $\mathbf{U}_i'$ is obtained from $\mathbf{U}_i$ by zeroing all elements in the first column. The fraction of all amino acid residues which are helical is

$$f = n^{-1} \Sigma f_i \quad (6)$$

The average number of amino acid residues in a helical segment is

$$\nu = \Sigma Z_i' / \Sigma Z_i'' \quad (7)$$

where $Z_i''$ differs from $Z$ only in substitution of $\mathbf{U}_i''$ for $\mathbf{U}_i$. Matrix $\mathbf{U}_i''$ is obtained from $\mathbf{U}_i$ by zeroing all elements except the last element in the first row.

Table I
$\sigma$ and $s$ Used in Water at 30 °C

| amino acid residue | $10^4\sigma$ | $s$ | ref[a] |
|---|---|---|---|
| Pro | 0 | 0 | 38 |
| Gly | 0.1 | 0.615 | 11 |
| Asp | 70 | 0.63 | 23 |
| Ser (Cys) | 0.1 | 0.78 | 20 |
| Asn | 0.1 | 0.806 | 19 |
| Thr | 0.1 | 0.836 | 24 |
| Lys | 1 | 0.946 | 18 |
| Tyr | 66 | 0.96 | 17 |
| Val (Ile) | 1 | 0.97 | 15 |
| Glu (Gln) | 6 | 0.97 | 16 |
| Arg | 0.01 | 1.028 | 22 |
| Ala | 8.4 | 1.056 | 12 |
| Phe (His, Trp) | 18 | 1.066 | 14 |
| Leu | 30 | 1.12 | 13 |
| Met | 51 | 1.14 | 21 |

[a] Reference for first amino acid residue listed.

Generator matrices described by Flory[1] were formulated in order to calculate $\langle r^2 \rangle$ and $\langle s^2 \rangle$. The latter is precisely defined as the mean-square radius of gyration of all $C^\alpha$ atoms in the polypeptide. Peptide bonds were maintained in the planar trans state. Averaged transformation matrices for glycyl and alanyl residues are those described by Brant et al.[26] unless amino acid residue $i + 1$ is prolyl, in which case they were taken from Schimmel and Flory.[27] Prolyl transformation matrices are from Mattice et al.[28] or from unpublished work by Ooi, depending on whether or not amino acid residue $i + 1$ is also prolyl. Alanyl transformation matrices were used for all amino acid residues other than glycyl or prolyl. The transformation matrices are summarized in eq 21–27 of ref 25. Transformation matrices used successfully reproduce measured unperturbed dimensions of several homopolypeptides,[26,28-32] sequential[33] and random[34] copolypeptides, and completely disordered proteins.[35] They also are in agreement with NMR coupling constants between the NH and $C^\alpha H$ protons of several homopolypeptides[36] and the dipole moments of several small peptides.[37]

Zimm–Bragg statistical weights for 14 of the amino acid residues in water are those determined by Scheraga and co-workers, using the host–guest approach.[11-24] Cysteinyl, glutaminyl, histidyl, isoleucyl, and tryptophanyl residues were assigned statistical weights found for related amino acid residues. Prolyl residues were assigned an $s$ of zero on steric grounds.[38] Values used at 30 °C are summarized in Table I.

Statistical weights for amino acid residues in protein–sodium dodecyl sulfate complexes are based on a model described earlier.[25] This model recognizes that cationic,[39-44] but not anionic[45] or uncharged,[46,47] homopolypeptides have an increased tendency to form ordered structures in the presence of sodium dodecyl sulfate. Consequently, larger values for $\sigma$ and $s$ for arginyl, histidyl, and lysyl residues are used in the presence of this detergent. This model for protein–sodium dodecyl sulfate complexes has been found to account for the optical activity and transport properties of a large number of complexes.[25,48] It also correctly identifies certain proteins as forming complexes with "abnormal" hydrodynamic properties. When used with a branched molecule formalism,[2,3] it can reproduce the measured[49] dimensions of cross-linked tropomyosin in sodium dodecyl sulfate.[50]

Amino acid sequences are those reported in ref 51. For purposes of this work, uncertainties between Asn and Asp, or between Gln and Glu, were always resolved in favor of the side chain containing a carboxyl group. The few un-

Table II
Conformational Properties for $A_{50}B_{50}$ and $B_{50}A_{50}$[a]

| property | 2 × 2 matrix | | 3 × 3 matrix | |
|---|---|---|---|---|
| | $A_{50}B_{50}$ | $B_{50}A_{50}$ | $A_{50}B_{50}$ | $B_{50}A_{50}$ |
| $f$ | 0.17 | 0.34 | 0.24 | 0.24 |
| largest $f_i$ | 0.33 | 0.58 | 0.39 | 0.39 |
| $i$ at largest $f_i$ | 73 | 50 | 67, 68 | 33, 34 |
| $\nu$ | 19 | 30 | 25 | 25 |
| $\langle r^2\rangle^{1/2}$, Å | 109 | 111 | 110 | 110 |
| $\langle s^2\rangle^{1/2}$, Å | 42 | 40 | 41 | 41 |
| $\langle r^2\rangle/\langle s^2\rangle$ | 6.9 | 7.5 | 7.1 | 7.1 |

[a] $s$ is unity for A and B, while $\sigma = 10^{-4}$ and $10^{-3}$ for A and B, respectively.

identified amino acid residues in bovine serum albumin were assumed to be alanyl.

## Experimental Section

Bovine serum albumin was obtained from Sigma Chemical Co. Stock solution concentrations were determined from the absorbance at 280 nm.[52] Circular dichroism and absorbance were measured in the manner described previously.[25] Viscosity measurements for bovine serum albumin–sodium dodecyl sulfate solutions were performed with a Beckman low-shear rotating-cylinder viscometer. The number of revolutions was chosen so that the observation time for solvent was near 200 s. Viscosity measurements were also attempted by using Cannon-Ubbelohde semimicro dilution viscometers. Reproducibility with capillary viscometers was markedly inferior to that obtained with the rotating-cylinder viscometer. The difficulty with the Cannon-Ubbelohde viscometers may arise from the ease with which small bubbles form when the protein–detergent solution is forced through the capillary.

## AB Block Copolypeptide

Computations were executed for several block copolypeptides. The general pattern of the results is conveniently described by consideration of an AB copolypeptide containing 50 amino acid residues in each block. In order to bring out the consequences of the different weighting schemes, let $s$ be unity for both types of amino acid residues, while $\sigma = 10^{-4}$ for A and $10^{-3}$ for B. Values selected for $\sigma$ are well within the range found for amino acid residues in water (Table I). When the 2 × 2 statistical weight matrix is used, the statistical weight for a helical segment will be either $10^{-4}$ or $10^{-3}$, depending on whether the helix is considered to originate in the A or the B block. With the 3 × 3 statistical weight matrix, however, the statistical weight for a helical segment may be $10^{-4}$, $10^{-3.5}$, or $10^{-3}$. The smallest statistical weight is obtained if the helix is confined to the A block, while the largest statistical weight is assigned if the helix lies within the B block. The intermediate weight arises when the helical segment extends from one block into the other. Computed conformational properties are summarized in Table II, and the $f_i$ are depicted as a function of $i$ in Figure 1.

Different helicities are calculated for $A_{50}B_{50}$ and for $B_{50}A_{50}$ when the 2 × 2 statistical weight matrix is employed. The helical content calculated for $B_{50}A_{50}$ is twice that obtained for $A_{50}B_{50}$. The amino acid residue with the highest probability of being helical is near the middle of the B segment in $A_{50}B_{50}$, while it is at the A–B junction in $B_{50}A_{50}$ (Figure 1). The A block in $A_{50}B_{50}$ has a low helical content because in this copolypeptide all helices containing A amino acid residues have a statistical weight of $10^{-4}$. In $B_{50}A_{50}$, however, an A amino acid residue can occur in a helical segment that merits a statistical weight of $10^{-3}$. While $f$, $f_i$, and $\nu$ are markedly different for $A_{50}B_{50}$ and $B_{50}A_{50}$ when the 2 × 2 statistical weight matrix is used, there are only small differences in the unperturbed dimensions (Table II). Unperturbed dimensions for a polypeptide containing 100 amino acid residues are not very sensitive to $f$ in the range covered.

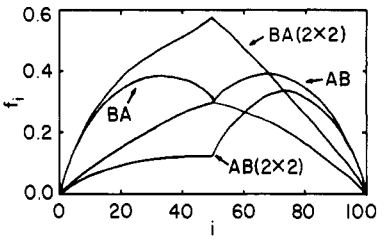


**Figure 1.** Helix probability profiles for block copolypeptides AB and BA. There are 50 amino acid residues in each block. Both amino acid residues have $s = 1$, while $\sigma$ is $10^{-4}$ and $10^{-3}$ for amino acid residues A and B, respectively. The helix probability profile was calculated twice for each block copolypeptide, using statistical weight matrices whose dimensions were 2 × 2 or 3 × 3. Profiles obtained with 2 × 2 statistical weight matrices are noted as such in the figure.

As expected, equivalent results are obtained for $A_{50}B_{50}$ and $B_{50}A_{50}$ when the 3 × 3 statistical weight matrix is used. The fraction of amino acid residues which are helical is 0.24. The highest helicity, 0.39, is obtained for the B amino acid residues located at positions 17 and 18 from the A–B junction. The 3 × 3 statistical weight matrix yields $f$, largest $f_i$, $\nu$, $\langle r^2\rangle$, and $\langle s^2\rangle$ lying between the two sets of results obtained when 2 × 2 statistical weight matrices were employed.

Significantly different $f$, $\nu$, and helix probability profiles may be obtained for block copolypeptides, depending on whether the statistical weight for a helical segment is $\sigma_i s_i s_{i+1} \cdots s_{j-1} s_j$ or $\sigma_i^{1/2} s_i s_{i+1} \cdots s_{j-1} s_j \sigma_j^{1/2}$. Computed unperturbed dimensions are much less sensitive to the difference in weighting.

## Proteins Experiencing Short-Range Interactions in Water

Proteins are specific-sequence copolypeptides in which 20 different kinds of amino acid residues are usually present. Long-range interactions play an important role in determination of their conformational properties in the biologically important "native" state. Concern here is not with the native state but rather with those conformations accessible when short-range interactions completely dominate the conformational energies. Specifically, conformations accessible to each nonhelical amino acid residue are those dictated by intrinsic torsional potentials as well as short-range van der Waals and electrostatic interactions.[26-28] Helix-forming tendencies (Table I) are either those determined in water, using the host–guest technique,[11-24] or are estimated from analogy to amino acid residues for which $\sigma$ and $s$ are known.

The average number of amino acid residues in a helical segment is not expected to be large because many amino acid residues are assigned $s < 1$. For most proteins, therefore, similar $\langle r^2\rangle$ and $\langle s^2\rangle$ values are expected from calculations using the 2 × 2 and 3 × 3 statistical weight matrices. It is not readily apparent, however, in what manner $f$, $f_i$, and $\nu$ will be affected by the choice of statistical weight matrix size. This problem is investigated here in detail for two proteins containing several hundred amino acid residues each. One protein, bovine serum albumin, contains 582 amino acid residues, 25 of which are prolyl. Since $s$ for prolyl residues has been assigned the value of zero on steric grounds,[38] their presence forces sections of the polypeptide chain to have a helical content unaffected by events elsewhere in the polypeptide chain. The upper limit for $\nu$ is 22.3 for bovine serum albumin

Table III
Tropomyosin with $\sigma$ and $s$ from Random Copolypeptides in Water

| property | 2 × 2 matrix | 3 × 3 matrix |
|---|---|---|
| $f$ | 0.19 | 0.14 |
| largest $f_i$ | 0.33 | 0.26 |
| $i$ at largest $f_i$ | 100 | 106 |
| $\nu$ | 13.4 | 15.5 |
| $\langle r^2\rangle^{1/2}$, Å | 178 | 183 |
| $\langle s^2\rangle^{1/2}$, Å | 71 | 73 |
| $\langle r^2\rangle/\langle s^2\rangle$ | 6.3 | 6.3 |

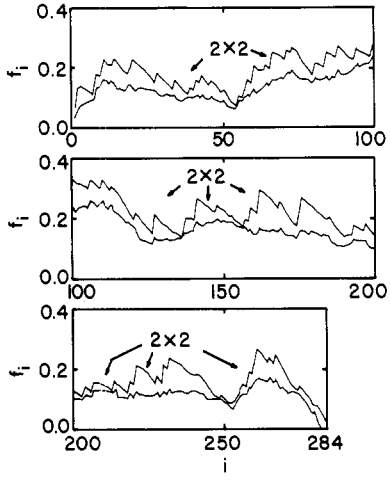


**Figure 2.** Helix probability profiles for tropomyosin computed with 2 × 2 and 3 × 3 statistical weight matrices formulated from the statistical weights in Table I. The profile obtained with 2 × 2 statistical weight matrices is noted as such in the figure.

when $s$ for prolyl residues is zero. Tropomyosin, in contrast, contains no prolyl among its 284 amino acid residues. Consequently there is no prohibition on the first and last amino acid residues being in the same helical segment in tropomyosin, and the upper limit for $\nu$ is $n$.

**Tropomyosin.** Conformational properties, calculated for tropomyosin using the $\sigma$ and $s$ values in Table I, are presented in Figure 2 and Table III. A higher helical content is calculated when the 2 × 2 statistical weight matrices are used, but helical segments contain more amino acid residues when evaluation is via the 3 × 3 statistical weight matrices. When a single unperturbed tropomyosin chain is considered, both calculations yield $f$ which lie well below the result of unity found with the tropomyosin dimer in the native state.[49,53,54] This result supports the suggestion[55] that dissociation of the tropomyosin dimer to single chains is accompanied by a dramatic reduction in helical content.

Behavior of the $f_i$ is depicted in Figure 2. Only six of the 284 amino acid residues show an increase in $f_i$ when the statistical weight matrix size changes from 2 × 2 to 3 × 3. Both calculations, however, find the helix probability is highest in the area of amino acid residues 100–115. The $f_i$ are a much smoother function of $i$ when the calculation is based on 3 × 3 statistical weight matrices. For example, when $i$ is in the range 200–250, $f_i$ is confined to 0.096–0.133 when 3 × 3 statistical weight matrices are used, but it ranges from 0.098 to 0.238 when the calculation is based on 2 × 2 matrices. The large fluctuations experienced by $f_i$ in the 2 × 2 calculation have a characteristic pattern. This pattern appears several times in the region where $i$ is 125–190. Here occur several isolated sharp jumps in $f_i$, with each jump being followed by a gradual decline extending over several amino acid residues. The nine largest jumps in this region occur at aspartyl, leucyl, methionyl,

Table IV
Serum Albumin with $\sigma$ and $s$ for Random Copolypeptides in Water

| property | 2 × 2 matrix | 3 × 3 matrix |
|---|---|---|
| $f$ | 0.06 | 0.04 |
| largest $f_i$ | 0.20 | 0.13 |
| $i$ at largest $f_i$ | 399 | 347 |
| $\nu$ | 7.0 | 8.2 |
| $\langle r^2\rangle^{1/2}$, Å | 260 | 264 |
| $\langle s^2\rangle^{1/2}$, Å | 105 | 107 |
| $\langle r^2\rangle/\langle s^2\rangle$ | 6.1 | 6.1 |

or tyrosyl residues, each of which is assigned a rather large value for $\sigma$. The large $\sigma$ permits comparatively easy helix initiation at these sites. Hence these amino acid residues spend a significant amount of time as the middle amino acid residue in a chh triplet. These same amino acid residues will tend to occur at *either* end of a helical segment when the 3 × 3 matrix is used; i.e., they occur as the middle amino acid residue in both chh and hhc triplets. Since helical segments may now propagate in either direction from the amino acid residue with the large $\sigma$, $f_i$ becomes a smoother function of $i$. Assignment of the entire end effect to the amino terminal end of a helical segment also significantly affects the helix probability profiles for poly(hydroxybutyl-L-glutamine) containing 5% seryl residues.[56]

Nearly identical unperturbed dimensions are computed with the 2 × 2 and 3 × 3 statistical weight matrices (Table III), as expected. The radius of gyration measured[49] for the completely helical tropomyosin dimer in water is twice that computed for the unperturbed, partially helical isolated chains.

**Bovine Serum Albumin.** Conformational properties computed for bovine serum albumin are summarized in Table IV. The helical content is appreciably less than that computed for tropomyosin, which is not surprising since 25 of the 582 amino acid residues are prolyl in bovine serum albumin. As was the case with tropomyosin, $f$ becomes smaller, and $\nu$ becomes larger, when 3 × 3 statistical weight matrices replace the 2 × 2 matrices. Will substitution of 3 × 3 for 2 × 2 matrices decrease the calculated value of $f$ for most proteins, or are the results reported here a consequence of amino acid sequences peculiar to tropomyosin and bovine serum albumin? A detailed analysis of the helix probability profile for bovine serum albumin suggests the former position is correct. Since $s$ for prolyl residues is assigned a value of zero, helices cannot propagate through these amino acid residues. Consequently prolyl residues have the effect of making 25 different polypeptides of the bovine serum albumin chain insofar as the computation of the $f_i$ is concerned. Computed helical contents for these 25 polypeptides are collected in Table V. In *every* case a lower helical content is obtained when 3 × 3 statistical weight matrices are employed.

The above results can be rationalized as follows. A trivial effect, important only for extremely short polypeptide chains, is the prohibition of the chc triplet when 3 × 3 statistical weight matrices are used. The major difference between the two calculational methods lies in the contributions helix ends make to the statistical weight of a helical segment. When the helical segment extends from amino acid residue $i$ to amino acid residue $j$, this contribution is $\sigma_i$ or $(\sigma_i\sigma_j)^{1/2}$, depending upon whether the matrix size is 2 × 2 or 3 × 3. Will $\sigma_i$ or $(\sigma_i\sigma_j)^{1/2}$ tend to be the larger for a specific-sequence copolypeptide containing about 20 different kinds of amino acid residues? In order to estimate the relative size of these two terms for proteins in general, without regard to the composition

Table V
Helical Content (%) for Isolated Segments of Bovine Serum Albumin[a]

| segment | 2 × 2 matrix | 3 × 3 matrix | segment | 2 × 2 matrix | 3 × 3 matrix |
|---|---|---|---|---|---|
| 1-34 | 8.9 | 6.3 | 338-363 | 11.4 | 7.4 |
| 36-95 | 7.9 | 4.9 | 365-381 | 4.5 | 3.1 |
| 97-109 | 1.3 | 1.2 | 383-413 | 10.5 | 5.5 |
| 111-112 | 0.9 | 0.3 | 415-418 | 0.1 | 0.0 |
| 114-116 | 0.3 | 0.1 | 420-438 | 1.4 | 0.3 |
| 118 | 0.4 | 0 | 440-444 | 0.34 | 0.32 |
| 120-145 | 7.5 | 4.3 | 446-465 | 8.8 | 3.9 |
| 147-150 | 2.4 | 1.3 | 467-483 | 0.6 | 0.2 |
| 152-177 | 5.3 | 3.0 | 485-490 | 0.8 | 0.3 |
| 179-221 | 8.1 | 5.0 | 492-496 | 1.5 | 0.5 |
| 223-278 | 6.0 | 3.9 | 498-514 | 4.8 | 3.3 |
| 280-296 | 3.6 | 2.0 | 516-534 | 5.0 | 3.3 |
| 298-300 | 0.7 | 0.4 | 536-570 | 5.6 | 4.1 |
| 303-336 | 7.8 | 5.0 | 572-582 | 1.3 | 0.8 |

[a] Prolyl residues occur at the positions between segments.

or sequence appropriate for any particular protein, attention is focused on averages where each of the 20 amino acid residues in Table I is weighted equally. The numerical results are

$$20^{-1}\sum\sigma_k = 1.5 \times 10^{-3} \quad (8)$$

and

$$400^{-1}\sum\sum(\sigma_k\sigma_l)^{1/2} = 7.5 \times 10^{-4} \quad (9)$$

End effects tend to make the statistical weight of a helical segment smaller when 3 × 3 statistical weight matrices are used. Helix–coil transitions for homopolypeptides become increasingly more cooperative as $\sigma$ decreases.[4,5] Therefore substitution of 3 × 3 for 2 × 2 statistical weight matrices in a computation for a typical protein would be expected to decrease $f$ if the helical content was low and to increase $f$ if the helical content was large. Since the 2 × 2 statistical weight matrices yield $f < {}^1/_2$ for the amino acid sequences found in typical proteins,[25] $f$ will decrease if the computation is repeated with 3 × 3 matrices. It must be emphasized that this prediction assumes that each kind of amino acid residue is distributed in a somewhat uniform fashion along the polypeptide chain. Segregation of the amino acid residues may produce a contrary result, as illustrated by the $B_{50}A_{50}$ block copolypeptide considered above.

**Other Proteins.** Validity of the above rationalization was tested by carrying out computations of $f$ and $\nu$ for 42 additional proteins of known amino acid sequence. Figure 3 depicts the relationship of the $f$ calculated using the two different statistical weight matrices. Proteins used are identified in the legend for this figure. All $f$ are less than 0.50, and $f$ for a given protein is always smaller when the 3 × 3 statistical weight matrix is used. In contrast, $\nu$ for a particular protein is always larger when 3 × 3 matrices are employed (Figure 4). The behavior is that expected if cooperativity increases when amino acid residues at each end of a helical segment contribute to the end effects. Unperturbed dimensions for each protein are essentially independent of statistical weight matrix size.

## Protein–Sodium Dodecyl Sulfate Complexes

Conformational properties of complexes formed by several proteins with sodium dodecyl sulfate have been rationalized by a model which makes the following four assumptions.[25] (a) The native structure is completely disrupted. (b) Conformational properties are dominated by short-range interactions. (c) Conformational energy surfaces for disordered amino acid residues are not affected by the detergent. (d) Helix-forming tendencies are unaffected by the detergent unless the amino acid residue's side chain is cationic, in which case there is an increase in $\sigma$ and $s$. Helicities computed at 30 °C for tropomyosin and bovine serum albumin using this model are depicted in Figure 5 as a function of $s$ for the cationic amino acid residues (arginyl, histidyl, and lysyl). The $\sigma$ for these amino acid residues is calculated from $s$ via eq 28 of ref 25. Computed helicities rise as $s$ for cationic amino acid residues increases from 1.1 to 1.9. At a given $s$, lower helicity is obtained with the 3 × 3 statistical weight matrices if $f < {}^1/_2$. Curves calculated with 2 × 2 and 3 × 3 matrices cross at $f = {}^1/_2$ for tropomyosin. When $f$ exceeds

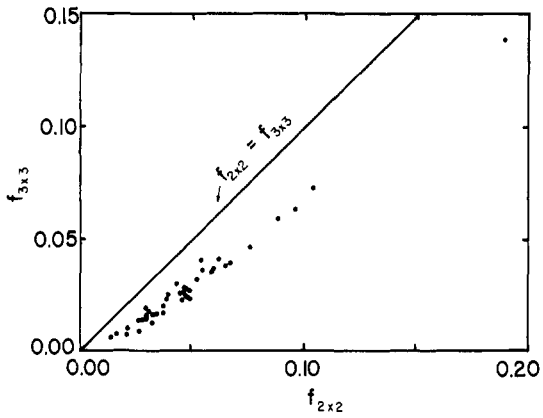


**Figure 3.** Relationship between helicities calculated for various proteins, using $\sigma$ and $s$ from Table I and either 2 × 2 or 3 × 3 statistical weight matrices. The proteins, in order of increasing $f_{3\times3}$, are ferredoxin, rubredoxin, bovine trypsin inhibitor, protease B, superoxide dismutase, subtilisin, trypsinogen, alcohol dehydrogenase, phospholipase $a_2$, papain, penicillopepsin, pepsin, trypsin inhibitor (Kunitz), ribonuclease A, chymotrypsinogen, prealbumin, cytochrome $c$, concanavalin A, flavodoxin, chicken lysozyme, elastase, high-potential iron protein, adenylate kinase, thiosulfate sulfurtransferase, thermolysin, carbonic anhydrase, glyceraldehyde-3-phosphate dehydrogenase, carboxypeptidase, thioredoxin, hemoglobin $\alpha$ chain, triose phosphate isomerase, cytochrome $c_2$, nuclease, serum albumin, lactate dehydrogenase, glucagon, myohemerythrin, hemoglobin $\beta$ chain, cytochrome $b_5$, T4 lysozyme, myoglobin, myogen, hemerythrin, and tropomyosin. Overlapping points are obtained in some instances.

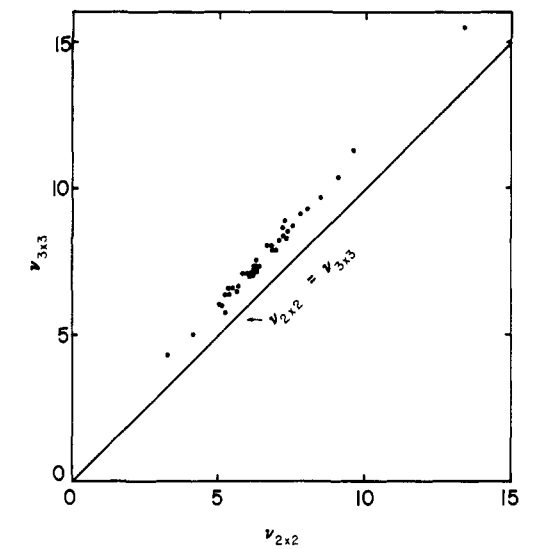


**Figure 4.** Relationship between average lengths of a helical segment calculated, using 2 × 2 and 3 × 3 statistical weight matrices, for the proteins listed in the legend for Figure 3.

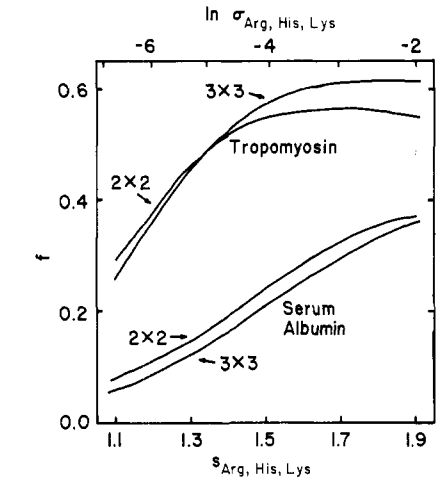


**Figure 5.** Helicities computed for tropomyosin and bovine serum albumin, using 2 × 2 and 3 × 3 statistical weight matrices formulated by using modified $\sigma$ and $s$ for amino acid residues bearing cationic side chains.

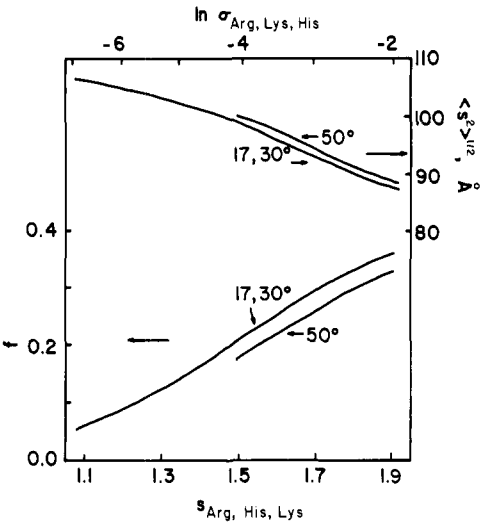


**Figure 6.** Helicities and root-mean-square radius of gyration computed for bovine serum albumin, using 3 × 3 statistical weight matrices and the indicated $\sigma$ and $s$ for amino acid residues bearing cationic side chains. The $\sigma$ and $s$ for the remaining 17 amino acid residues are those found for residues in random copolypeptides at 17, 30, and 50 °C in water (Table I and references therein).

$^1/_2$, a higher helicity is obtained with 3 × 3 statistical weight matrices. This behavior is in accord with the suggestion in the previous section that helix formation for proteins tends to display a greater cooperativity when the calculation is performed with the larger statistical weight matrices.

Figure 6 depicts $f$ and $\langle s^2 \rangle^{1/2}$ calculated for bovine serum albumin using 3 × 3 statistical weight matrices, the indicated $s$ for cationic amino acid residues, and $s$ appropriate for 17, 30, and 50 °C for the remaining amino acid residues.[11-24] Heating from 17 to 30 °C produces a decrease in $s$ for some amino acid residues (e.g., tyrosyl[17]), an increase in $s$ for other amino acid residues (e.g., valyl[15]), and no change for still others (e.g., glutamyl[16]). The net result is only a small change in $f$ and $\langle s^2 \rangle^{1/2}$, causing curves at 17 and 30 °C to be indistinguishable in Figure 4. The decrease in $s$ is dominant as the temperature rises from 30 to 50 °C, producing distinguishable curves for these two temperatures.

Circular dichroism spectra for bovine serum albumin in sodium dodecyl sulfate and a disulfide bond disrupting agent (0.0028 M dithiothreitol) have the shape expected for a partially helical polypeptide. The mean residue ellipticity at 222 nm suggests the helical content at 17–50 °C is 0.33 ± 0.03 and 0.36 ± 0.04 at sodium dodecyl sulfate concentrations of 0.018 and 0.035 M, respectively. These results are consistent with an $s$ for the cationic amino acid residues which is at the upper end of the range depicted in Figure 6, as expected from previous work.[25] Absence of a strong temperature dependence for the measured $f$ suggests that the $s$ appropriate for cationic amino acid residues in sodium dodecyl sulfate also has little temperature dependence.

Intrinsic viscosities for bovine serum albumin, measured in 0.035 M sodium dodecyl sulfate containing dithiothreitol, are constant at about 0.23 dL/g over the temperature range 17–50 °C. Conversion to a root-mean-square radius of gyration via the equation[29]

$$\langle s^2 \rangle^{1/2} = 6^{-1/2}([\eta]M/\Phi)^{1/3} \quad (10)$$

with $\Phi = 0.0021$ for $\langle s^2 \rangle^{1/2}$ in Å and $[\eta]$ in dL/g, yields a temperature-independent value of about 80 Å. This estimate is within about 10% of the unperturbed value depicted at the higher $s_{Arg,His,Lys}$ in Figure 6. Thus the model is in reasonable agreement with both the helicity and dimensions, provided expansion factors differ little from unity.

## Conclusion

Block copolypeptides may give markedly different $\nu$, $f$, and helix probability profiles when end effects are assigned to both ends of a helical segment or just to the amino terminal end. The sequence in which the blocks are connected will dictate which weighting scheme gives the larger $f$ and $\nu$. Typical proteins, in contrast, yield lower $f$ when the weighting scheme used assigns end effect to both ends of a helical segment. A smoother helix probability profile is also obtained. Helix formation in proteins appears to be more cooperative when both ends of a helical segment contribute to the end effects. A modification of this weighting scheme can account for the hydrodynamic properties and circular dichroism spectra of complexes formed by sodium dodecyl sulfate and bovine serum albumin.

**Acknowledgment.** This work was supported by the National Science Foundation (Research Grant PCM 78-22916).

## References and Notes

(1) Flory, P. J. *Macromolecules* **1974**, *7*, 381.
(2) Mattice, W. L. *Macromolecules* **1975**, *8*, 644.
(3) Mattice, W. L. *Macromolecules* **1976**, *9*, 48.
(4) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526.
(5) Miller, W. G.; Flory, P. J. *J. Mol. Biol.* **1966**, *15*, 298.
(6) Mattice, W. L. *Macromolecules* **1978**, *11*, 15.
(7) Mattice, W. L. *Macromolecules* **1980**, *13*, 506.
(8) Lifson, S.; Roig, A. *J. Chem. Phys.* **1961**, *34*, 1963.
(9) Go, N.; Lewis, P. N.; Go, M.; Scheraga, H. A. *Macromolecules* **1971**, *4*, 692.
(10) Tanaka, S.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 168.
(11) Ananthanarayanan, V. S.; Andreatta, R. H.; Poland, D.; Scheraga, H. A. *Macromolecules* **1971**, *4*, 417.
(12) Platzer, K. E. B.; Ananthanarayanan, V. S.; Andreatta, R. H.; Scheraga, H. A. *Macromolecules* **1972**, *5*, 177.
(13) Alter, J. E.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1972**, *5*, 739.
(14) Van Wart, H. E.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1973**, *6*, 266.
(15) Alter, J. E.; Andreatta, R. H.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1973**, *6*, 465.
(16) Maxfield, F. R.; Alter, J. E.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1975**, *8*, 479.
(17) Scheule, R. K.; Cardinaux, F.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 23.

(18) Dygert, M. K.; Taylor, G. T.; Cardinaux, F.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 793.
(19) Matheson, R. R.; Nemenoff, R. A.; Cardinaux, F.; Scheraga, H. A. *Biopolymers* **1977**, *16*, 1567.
(20) van Nispen, J. W.; Hill, D. J.; Scheraga, H. A. *Biopolymers* **1977**, *16*, 1587.
(21) Hill, D. J.; Cardinaux, F.; Scheraga, H. A. *Biopolymers* **1977**, *16*, 2447.
(22) Konishi, Y.; van Nispen, J. W.; Davenport, G.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 1264.
(23) Kobayashi, Y.; Cardinaux, F.; Zweifel, B. O.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 1271.
(24) Hecht, M. H.; Zweifel, B. O.; Scheraga, H. A. *Macromolecules* **1978**, *11*, 545.
(25) Mattice, W. L.; Riser, J. M.; Clark, D. S. *Biochemistry* **1976**, *15*, 4264.
(26) Brant, D. A.; Miller, W. G.; Flory, P. J. *J. Mol. Biol.* **1967**, *23*, 47.
(27) Schimmel, P. R.; Flory, P. J. *J. Mol. Biol.* **1968**, *34*, 105.
(28) Mattice, W. L.; Nishikawa, K.; Ooi, T. *Macromolecules* **1973**, *6*, 443.
(29) Brant, D. A.; Flory, P. J. *J. Am. Chem. Soc.* **1965**, *87*, 2788.
(30) Mattice, W. L.; Mandelkern, L. *J. Am. Chem. Soc.* **1971**, *93*, 1769.
(31) Mattice, W. L.; Lo, J. T. *Macromolecules* **1972**, *5*, 734.
(32) Clark, D. S.; Mattice, W. L. *Macromolecules* **1977**, *10*, 369.
(33) Mattice, W. L.; Mandelkern, L. *Biochemistry* **1971**, *10*, 1934.
(34) Miller, W. G.; Brant, D. A.; Flory, P. J. *J. Mol. Biol.* **1967**, *23*, 67.
(35) Miller, W. G.; Goebel, C. V. *Biochemistry* **1968**, *7*, 3925.
(36) Tonelli, A. E.; Bovey, F. A. *Macromolecules* **1970**, *3*, 410.
(37) Flory, P. J.; Schimmel, P. R. *J. Am. Chem. Soc.* **1967**, *89*, 6807.
(38) Schimmel, P. R.; Flory, P. J. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *58*, 52.
(39) Sarkar, P. K.; Doty, P. *Proc. Natl. Acad. Sci. U.S.A.* **1966**, *55*, 981.
(40) Grourke, M. J.; Gibbs, J. H. *Biopolymers* **1967**, *5*, 586.
(41) Satake, I.; Yang, J. T. *Biochem. Biophys. Res. Commun.* **1973**, *54*, 930.
(42) Satake, I.; Yang, J. T. *Biopolymers* **1975**, *14*, 1841.
(43) Mattice, W. L.; Harrison, W. H. *Biopolymers* **1976**, *15*, 559.
(44) McCord, R. W.; Blakeney, E. W.; Mattice, W. L. *Biopolymers* **1977**, *16*, 1319.
(45) Fasman, G. D.; Lindblow, C.; Bodenheimer, E. *Biochemistry* **1964**, *3*, 155.
(46) Lotan, N.; Yaron, A.; Berger, A.; Sela, M. *Biopolymers* **1965**, *3*, 625.
(47) Igou, D. K.; Lo, J. T.; Clark, D. S.; Mattice, W. L.; Younathan, E. S. *Biochem. Biophys. Res. Commun.* **1974**, *60*, 140.
(48) Mattice, W. L. *Biopolymers* **1979**, *18*, 225.
(49) Holtzer, A.; Clark, R.; Lowey, S. *Biochemistry* **1965**, *4*, 2401.
(50) Mattice, W. L. *Macromolecules* **1977**, *10*, 516.
(51) Dayhoff, M. O. "Atlas of Protein Sequence and Structure", Vol. 5, 1972, Suppl. 1, 1973, Suppl. 2, 1976, Suppl. 3, 1978.
(52) Wetlaufer, D. B. *Adv. Protein Chem.* **1962**, *17*, 378.
(53) Cohen, C.; Szent-Gyorgyi, A. *J. Am. Chem. Soc.* **1957**, *79*, 248.
(54) Chao, Y. H.; Holtzer, A. *Biochemistry* **1975**, *14*, 2164.
(55) Crimmins, D.; Holtzer, A., personal communication.
(56) Hughes, L. J.; Andreatta, R. H.; Scheraga, H. A. *Macromolecules* **1972**, *5*, 187.

# Surface Characterization of Methylsilsesquioxane–Phenylsilsesquioxane Copolymers

**José M. Sosa**†

*Chemistry Department, New Mexico State University, Las Cruces, New Mexico 88003.*
*Received March 4, 1980*

ABSTRACT: Mixtures of methyl- and phenyltrichlorosilanes have been hydrolyzed to obtain cross-linked polysiloxanes, often referred to as polysilsesquioxanes. Glassy or amorphous materials are obtained, depending on the reaction conditions used. Infrared spectroscopy, scanning electron microscopy, and gas chromatography have been employed to study the surface characteristics of these materials. The infusible solids that are obtained are thermally stable up to 400 °C. Surface areas for the materials range from 0.02 to 5 $m^2/g$, depending on the reaction conditions. Swelling experiments show that the materials are very tightly cross-linked networks. The soft materials having surface areas around 5–10 $m^2/g$ are suitable as adsorbents for gas–solid chromatography.

## Introduction

The chemistry of linear poly(dimethylsiloxanes) and poly(methylphenylsiloxanes) has been studied extensively. Highly cross-linked poly(methylphenylsiloxanes), often referred to as poly(methylphenylsilsesquioxanes), are less widely known. Polysilsesquioxanes are generally prepared by hydrolysis of organotrichlorosilanes. Gilliam and coworkers[1] described the resinous products obtained by the cohydrolysis of dimethylchlorosilane and methyltrichlorosilane in the early 1940s. Cross-linked poly(alkylsiloxanes) were first made by Ladenberg in 1872; however, Kipping[3] is credited with proposing the siloxane bond and the formation of linear and cross-linked polymers in 1914. Patnode and Wilcock[4] recognized that linear, soluble compounds are formed during the hydrolysis of methyltrichlorosilanes in certain solvents. Brown and co-workers,[5–7] after extensive studies, reported that hydrolysis of organotrichlorosilanes gave very complex structures. Slinyakova and Niemark[8] reported that few investigations dealing with methods of obtaining gels from organosilicon compounds appear in the Russian literature prior to 1962. Many of the studies dealing with silsesquioxanes up to about 1970 involved the synthesis and characterization of low and high molecular weight polymers.[9–11] Andrianov and his group, however, have investigated extensively many aspects of different silsesquioxanes, including their morphology,[12] conformations,[13] thermal decomposition,[14] hydrodynamic and optical properties,[15] and synthesis.[16]

After the introduction of surface-modified silica gels in the field of chromatography, research groups evaluated polysilsesquioxanes as adsorbents for gas–solid and liquid–solid chromatography. Although poly(phenylsilsesquioxane)[17] and poly(octylsilsesquioxane)[18] have been reported to be useful as adsorbents, poly(methyl-, -phenyl-, and -methylphenylsilsesquioxane) copolymers have been

†Current address: Chemicals Division, Vulcan Materials Co., Wichita, KS 67277.

0024-9297/80/2213-1260$01.00/0 