Macromolecules **1984**, *17*, 2690–2696

Practical Estimates of the Upper Limit for the Distribution Function for Strand Lengths in Large Homopolymers Containing Intramolecular Antiparallel Sheets with Tight Bends

Wayne L. Mattice*

Department of Chemistry, Louisiana State University, Baton Rouge, Louisiana 70803

Harold A. Scheraga*

Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853.
Received January 31, 1984

ABSTRACT: Recently we described a tractable matrix formulation of the conformational partition function for a chain undergoing formation of intramolecular antiparallel sheets with tight bends. The distribution function for the number of residues per strand is among the physical properties which can be extracted from this conformational partition function. The behavior at the high end of the distribution function determines the dimensions of the smallest statistical weight matrix that may be used in the characterization of the transition. A simple relationship, which incorporates the degree of polymerization and the two statistical weights arising from end effects, provides a rapid estimate of the dimensions of the matrix. The validity of this approximation is demonstrated by a consideration of several different chains. Some of the chains considered are assigned end-effect parameters that are in qualitative agreement with those expected for linear poly(methylene).

A tractable matrix formulation[1] has been developed for the formation of intramolecular antiparallel β sheets[2] in homopoly(amino acids) having a finite degree of polymerization. Each antiparallel sheet has a statistical weight, w, given by

$$w = \delta^{n_{\text{bend}}/2}\tau^{n_b}t^{n_{b+B}} \tag{1}$$

where n_{b+B} is the number of residues in the sheet. The numbers of residues that do not, and do, have a partner in a preceding strand are denoted by n_b and n_B, respectively, and n_{bend} denotes the number of residues located at bends. Each bend consists of two residues, and residues involved in bend formation are considered to be part of the sheet. Statistical weights t, τ, and δ are assigned to $(b + B)$ and b states and bends, respectively. For example, the sheet depicted as structure e in Figure 1 contains 36 residues in four strands. It has a statistical weight of $\delta^3\tau^9t^{36}$. The matrix formulation[1] does not require that all strands in a sheet contain the same number of residues. Sheets in which contiguous strands contain unequal numbers of residues actually dominate the ensemble under many circumstances of interest,[3] including transitions giving rise to cross-β fibers[4,5] and sheets of the type found in globular proteins.[6,7]

Interactions of long range play an important role in the formation of antiparallel sheets. The sheet depicted as structure e in Figure 1 contains many pairs of adjacent interacting residues in contiguous strands. The longest range of these adjacent interacting residues bears the indices i and $i + 17$, with indexing taking place from one end of the chain to the other. In general there will be one such longest-range interaction between residues i and $i + 2j - 1$ whenever the shorter of two contiguous strands contains j residues. Our formulation[1] rigorously accounts for all intramolecular sheets containing tight bends in which the maximum number of residues permitted in a strand is I. Ultimate interest is in the behavior of computed conformational averages in the limit where I approaches n, the number of residues in the chain. The value of I required to reach this limit depends on the behavior of the distribution function for the number of residues per strand. Once I has been made large enough to capture a sufficiently large portion of this distribution function, further increases in I will have little effect.

The most rigorous means of determining the necessary size for I is by direct examination of the distribution function for the numbers of residues per strand. Considerable computer time may be required for computation of the distribution function if n and the pertinent I are large. The major goal here is to find simple, easily evaluated expressions that indicate how large I must be if the computation is to take account of those strands occurring in a significant number of conformations. Such values of I also give an indication of the maximum thickness of a sheet. This information about size is different from the average number of residues per strand, which is obtained by means of the conventional manipulation of the partition function Z.[8]

A quick estimate of the behavior of the high end of the distribution function for strand lengths is of tremendous practical benefit because of the manner in which Z is formulated. The dimensions of the statistical weight matrix, **U**, are $I(I + 3)/2 \times I(I + 3)/2$.[1] Computations are feasible even for rather large values of I, because **U** is sparse. Previous work[9] has described calculations for a chain of 1000 residues in which I was 25. The present work employs I as large as 52. The dimensions of **U** become 1430×1430 when $I = 52$. An attempt to store every element in a matrix this large would exceed the capacity of many computers. Calculations remain possible, however, because **U** is never formulated as such. Instead, the computation is performed by taking advantage of the sparseness of **U** and of summarizing statements which identify the locations of all nonzero elements.[1] The upper limit on the usable size for I is not imposed by available memory but is imposed instead by the unavoidable increase in CPU time as I and n increase.

For short chains, calculations can be performed with increasing values of I until the conformational averages converge to a limit. Many applications to proteins will fall in this category because of their rather low degree of polymerization. The approach using repetitive calculations becomes less attractive for long chains in which the statistical weights give rise to sheets containing many residues per strand. Among the circumstances where this problem may occur are investigations of chain behavior as the degree of polymerization becomes infinite. There would be a definite practical advantage to knowing in advance the

0024-9297/84/2217-2690$01.50/0

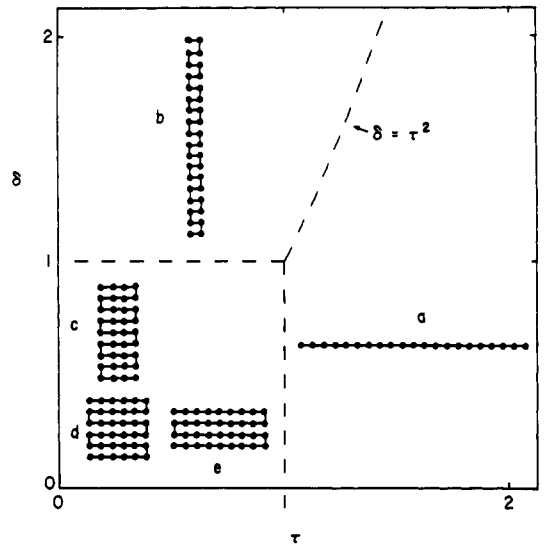

Figure 1. Diagrammatic representation of optimal intramolecular antiparallel sheets when t is infinite.

smallest assignment of I that does not exclude conformations making significant contributions to the conformational averages of interest.

There will also be circumstances where it is necessary to make automatic changes in I during a series of computations. A continuous updating of I may be required during the selection of those values of the statistical weights that optimally account for measured properties for a particular polymer sample. During the optimization procedure, δ and τ may be changed in a manner which broadens the distribution function for strand lengths. If the value of I is not changed accordingly, the optimization will fail because it ignores significant sheets containing long strands.

Even an approximate estimate of the location of the high end of the distribution of strand sizes will be shown to be sufficiently accurate to serve as a useful guide in optimization procedures and to define the limits of practical calculations for long chains. Chains having two different sets of statistical weight assignments are examined for illustrative purposes. They have assignments for δ and τ which permit an assessment of the effect of the short-range interactions in linear poly(methylene) upon the types of intramolecular antiparallel sheets preferred by this chain.

Shape of Optimal Sheets as t Becomes Infinite

The weight assigned to an antiparallel sheet, relative to a weight of 1 for a disordered chain, is given by eq 1. Each strand must have at least two residues. An example of a sheet in which all strands contain two residues is depicted as structure b in Figure 1.

When t is finite, a chain of n residues may contain several antiparallel sheets of different sizes. However, the preferred chain conformation must become a single sheet of n residues if t becomes infinite while δ and τ remain finite. The shapes for such sheets, and their relative weightings, are determined by the values of δ and τ. The symbol I_∞ will be used to denote the maximum number of residues per strand in that single sheet of n residues that has the largest statistical weight as t becomes infinite. That sheet is the one for which w_{edge}

$$w_{edge} = \delta^{n_{bend}/2}\tau^{n_b} \tag{2}$$

is maximal, subject to the constraint that the sheet contain n residues. The major concern here is with the behavior of I_∞ for large n.

Four different regions of τ–δ space are distinguished in Figure 1. The first region is made up of the single point where $\delta = \tau = 1$. Here, all possible sheets of n residues have the same statistical weight. Since there is no preferred shape for the sheet as t becomes infinite, I_∞ is undefined.

The next region includes the quadrant where $\delta < 1$ and $\tau > 1$. Maximization of the expression in eq 2 requires minimization of the number of bends and maximization of the number of residues which do not have a partner in a preceding strand. This goal is achieved when the "sheet" consists of a single extended strand of n residues (structure a in Figure 1). The value of I_∞ is then n.

A third region includes the quadrant where $\delta > 1$ and $\tau < 1$. Now the preferred sheet is obtained when the number of bends is maximal and the minimum value is assigned to the number of residues which do not have a partner in a preceding strand. The optimal sheet consists of $n/2$ strands of two residues each (if n is even) or $(n - 3)/2$ strands of two residues each and a single strand of three residues (if n is odd). The value of I_∞ is 2 if n is even (structure b, Figure 1) or 3 if n is odd. The latter result is obtained because I_∞ is determined by the longest (not the average or most probable) strand in the optimal sheet.

The cases described in the preceding two paragraphs extend into the quadrant where $\delta > 1$ and $\tau > 1$. They meet at the line $\delta = \tau^2$. The origin of this border becomes apparent upon consideration of the first step in the conversion of structure a into structure b. That conversion takes a single strand of n residues (with statistical weight $\tau^n t^n$) and converts it into a sheet consisting of one strand of $n - 2$ residues and a second strand of 2 residues (with statistical weight $\delta\tau^{n-2}t^n$). The ratio of the statistical weights of these two- and one-stranded sheets is δ/τ^2. Formation of each additional strand of two residues alters the statistical weight by the same factor. The equilibrium constant for each of these processes is unity when $\delta = \tau^2$.

The fourth, and most interesting, region is provided by the quadrant where $\delta < 1$ and $\tau < 1$. The objective is still to maximize w_{edge}, but in this case this goal cannot be achieved by maximizing n_b or n_{bend}, as was done with structures a and b, respectively. Instead, we must maximize w_{edge} explicitly. Since δ and τ are associated with the edges of the sheet, the optimal sheet will be one in which edges are avoided. Candidates are provided by the sheets depicted as structures c, d, and e in Figure 1. These sheets contain $n_{bend}/2$ bends, $(n_{bend}/2 + 1)$ strands, and n_b residues per strand. The number of residues in these sheets is

$$n = n_b(n_{bend}/2 + 1) \tag{3}$$

We consider only rectangular (regular[3]) sheets here in order to minimize the perimeter (edges). Using eq 3 to eliminate $n_{bend}/2$ from eq 2, we write the expression for w_{edge} as

$$w_{edge} = \delta^{(n-n_b)/n_b}\tau^{n_b} \tag{4}$$

If n_b is now treated as a continuous variable, with δ, τ, and n held constant (implying that n_{bend} is also a continuous variable), w_{edge} will be a maximum when

$$(\partial w_{edge}/\partial n_b)_{\delta,\tau,n} = \delta^{(n-n_b)/n_b}\tau^{n_b}[\ln \tau - (n/n_b^2) \ln \delta] \tag{5}$$

is zero, and the second derivative is negative, i.e., when

$$n_b = [(\ln \delta/\ln \tau)n]^{1/2} \tag{6}$$

Equating this n_b to I_∞, the final result is

$$I_\infty = [(\ln \delta/\ln \tau)n]^{1/2} \tag{7}$$

The preferred sheet, at infinitely large t, has a small axial ratio of $\delta = \tau < 1$ (structure d, Figure 1) and a large axial ratio if $\delta < \tau < 1$ (structure e, Figure 1) or $\tau < \delta < 1$ (structure c, Figure 1). The last two structures are found in parallel-β and cross-β fibers, respectively.

Equation 7 is an oversimplified approximate expression. Objections can be raised on several counts. Among them are the following: (i) Equation 5 is obtained by treating the number of residues in a strand as a continuous variable. In contrast, the formulation of the conformational partition function assumes that the number of residues in a strand must be an integer selected from 2, 3, ..., n. (ii) While there may be a single optimal sheet (given by eq 7) as t becomes infinite, that single sheet will not necessarily dominate the behavior of the ensemble. Other sheets comprised of n residues, but arranged in different shapes, may retain significant contributions. Some of these sheets will contain strands in which the number of residues exceeds the value specified by eq 7. Therefore eq 7 may underestimate the upper limit of the number of residues per strand in the important conformations as t becomes infinite. (iii) The usual interest is not in the behavior as t becomes infinite, but rather in the manner in which sheets are formed in the transition region, where t is finite. Even if the sheet content is high (i.e. $0.9 < f_{b+B} < 1.0$), the ensemble for a very long chain is likely to be dominated by conformations in which there are several distinct sheets separated by a few residues in the disordered state due to the entropy gain. Since the number of residues per sheet will then be smaller than n, eq 7 may overestimate the upper limit for the number of residues per strand in the important conformations.

Objection i may be of little consequence for long chains. While the remaining two objections individually retain significance at large n, it is apparent that they will act in opposing directions. Objection ii suggests that eq 7 provides an underestimate, while objection iii suggests that it yields an overestimate, of the value of I_∞ which must be employed in computations designed to produce an accurate characterization of the transition. If these two effects exert comparable influences, eq 7 may provide useful guidance in treatments of intramolecular antiparallel sheets in long chains. Before we pursue this possibility, eq 7 will be separated into two predictions which are subject to indendent tests:

$$I_\infty = bn^{1/2} \tag{8}$$

for given δ and τ, which merely asserts that I_∞ becomes proportional to the square root of n for large chains, and

$$b^2 = \ln \delta / \ln \tau \tag{9}$$

which relates the proportionality constant to the edge effects in the sheet of given n.

Equation 8 has the same functional form as the familiar $\langle r^2 \rangle^{1/2} = ln^{1/2}$, where $\langle r^2 \rangle^{1/2}$ is the root-mean-square end-to-end distance for a freely-jointed chain comprised of bonds of length l. Pursuing the analogy, the proportionality constant b plays the role of an effective "bond length". When $\delta = \tau$, b is 1 and $bn^{1/2}$ is the length of the side of a square sheet (structure d in Figure 1). The length of a side is identical with the end-to-end distance when the sheet contains an even number of strands. The optimal sheet is not square if $\delta \neq \tau$, and the length of a side is no longer identical with the end-to-end distance. For the purpose of identifying I_∞, our interest is then in the length of the short side of the rectangle when $\tau < \delta < 1$ ($b < 1$, structure c in Figure 1) and of the long side of the rectangle when $\delta < \tau < 1$ ($b > 1$, structure e in Figure 1).

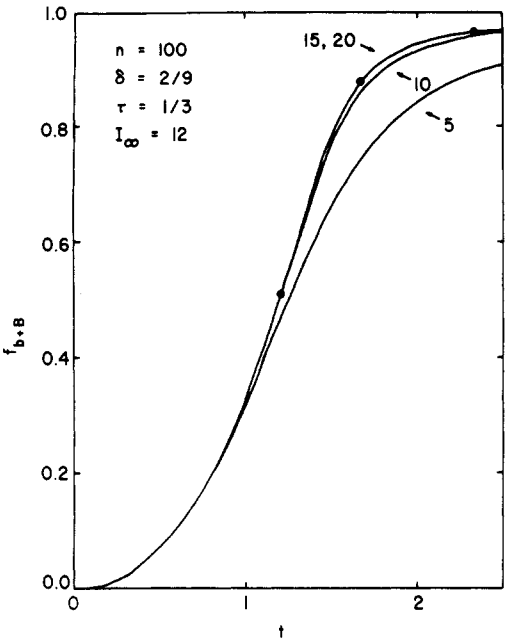

Figure 2. Formation of intramolecular antiparallel sheets, evaluated at the indicated values of I, when $n = 100$, $\delta = 2/9$, and $\tau = 1/3$. The value of I_∞ of eq 7 is 12.

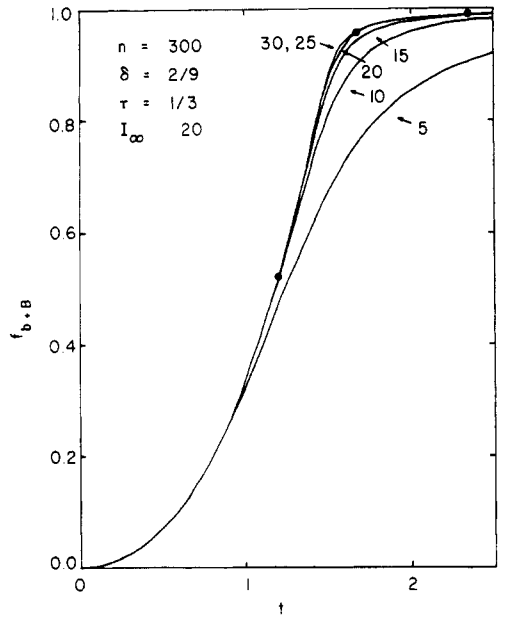

Figure 3. Formation of intramolecular antiparallel sheets, evaluated at the indicated values of I, when $n = 300$, $\delta = 2/9$, and $\tau = 1/3$. The value of I_∞ of eq 7 is 20.

Chain Length Dependence of I_∞

Figures 2–4 depict the calculated course of sheet formation when $\delta = 2/9$ and $\tau = 1/3$. The values of n are 100, 300, and 1000 in Figures 2–4, respectively. Figures 2 and 3 present several curves in which the value of I is successively incremented by 5. Increments of 5 and 10 were used in Figure 4.

In Figure 2, the curves calculated by using $I = 5$ and $I = 10$ are readily distinguishable. A much smaller change is seen upon increasing I from 10 to 15. No further change is detectable on the scale used in this figure when I is increased to 20. Consequently, the smallest value of I which provides an accurate characterization of the transition is greater than 10, but not greater than 15. The value of I_∞ specified by eq 7 is 12.

Calculations employing $I = 15$ and $I = 20$ yield distinguishable results if the number of residues is increased to

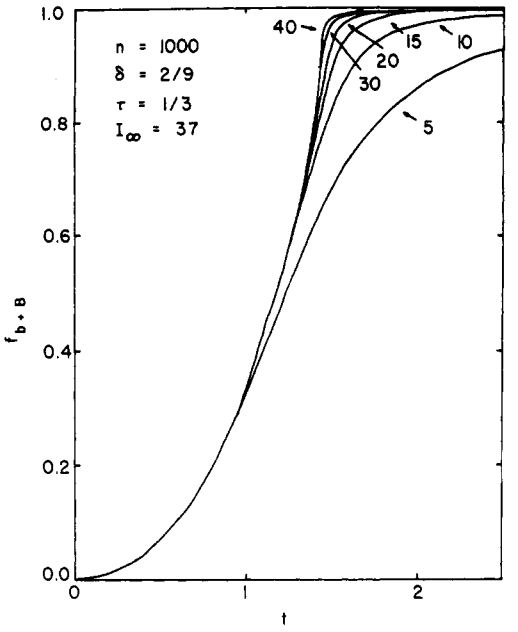

Figure 4. Formation of intramolecular antiparallel sheets, evaluated at the indicated values of I, when $n = 1000$, $\delta = 2/9$, and $\tau = 1/3$. The value of I_∞ of eq 7 is 37.

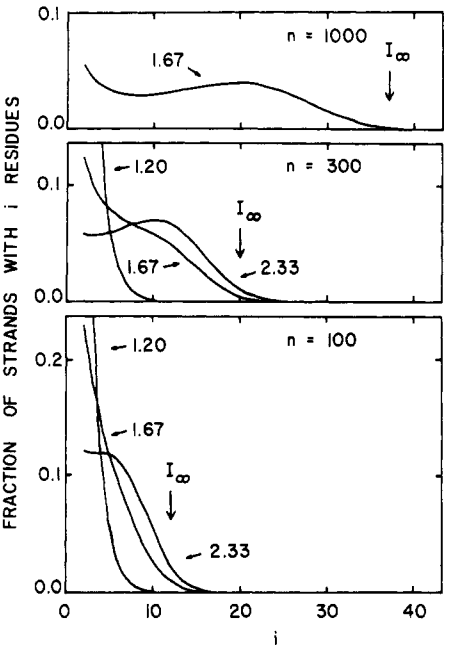

Figure 5. Normalized distribution functions for the number of residues per strand at the indicated values of t and n when $\delta = 2/9$ and $\tau = 1/3$. The values of t are indicated by filled circles on the transition curves depicted in Figures 2–4.

300 (Figure 3). Very little additional change is seen if I is increased to 25 or 30. The different responses to changes in I in Figures 2 and 3 cannot be attributed to end effects because the same values of δ and τ were employed. The value of I_∞ specified by eq 7 changes solely because of the change in n. This equation yields $I_\infty = 20$ when n is 300.

If the chain contains 1000 residues, curves computed by using $I = 20$ and 30 are distinguishable from one another (Figure 4). Curves using I of 30 and 40 can be distinguished in a limited region near $t = 1.5$. The value of I_∞ specified by eq 7 is now 37. Figures 2–4 demonstrate that eq 7 provides useful guidance to the size of the smallest value of I that characterizes the transition. Since neither δ nor τ has been changed, the results are more accurately described as a verification of the predictive capability of eq 8 at large n.

The basis for the effect of changes in I in Figures 2–4 is brought out by an examination of normalized distribution functions for the number of residues per strand. Illustrative distribution functions are depicted in Figure 5. These distribution functions were calculated by using the values of t denoted by the filled circles placed on the transition curves in Figures 2–4. The high ends of the distribution functions extend to larger strands as t increases at constant n or as n increases at constant t. The values of I_∞ specified by eq 7 are shown in Figure 5. For the longest chain, I_∞ provides a useful identification of the high end of the distribution function when the transition is essentially complete ($f_{b+B} = 0.99$). The accuracy of the identification is not as great with the shorter chains. For a chain of 300 residues, the tail at the high end starts to slip past I_∞ when f_{b+B} is about 0.95.

Approximate Statistical Weights for Simple Chains with Interdependent Symmetric Threefold Rotation Potentials

With the utility of eq 8 established, attention now turns to an examination of eq 9. An assessment of its validity requires computations of sheet formation using different values of δ and τ. A satisfactory approach would simply be to select various values of δ and τ at random, subject only to the requirement that they be less than unity. The transition obtained when $n = 300$ and $\delta = \tau = 0.3$ has been described earlier.[1] Equation 7 yields $I_\infty = 17$ in this case, which is in harmony with the behavior of both f_{b+B} at various I and the tail at the high end of the distribution function near the completion of the transition (Figures 9 and 4, respectively, from ref 1). While this result supports the assignment of the proportionality constant by means of eq 9, it is not entirely convincing because I_∞ is only slightly different from that employed in Figure 3 of the present paper, in which the value of δ differs from that used previously. A more severe test would be provided by a chain in which $\ln \delta / \ln \tau$ is much different from that obtained when $\delta = 2/9$, $\tau = 1/3$ or $\delta = \tau = 0.3$.

The calculations would take on greater interest if the selected values of δ and τ might bear a resemblance to those expected for real chains. Further assessment of the usefulness of eq 9 will be preceded by the development of approximate relationships between δ and τ and the statistical weights commonly used[10,11] to describe the conformational statistics of unperturbed simple chains with interdependent symmetric threefold rotation potentials. Residues will be taken as chain atoms, and it will be assumed that individual strands adopt the fully extended (all trans) conformation.

The edge effect denoted by τ occurs n, 2, 4, 6, and 9 times in the statistical weights for the sheets depicted as structures a–e, respectively, in Figure 1. The value assigned to this statistical weight depends on the difficulty involved in forming an isolated extended strand. It should be influenced by the ease with which gauche states can be populated in the disordered chain. If the disordered chain contains predominantly trans states and relatively few gauche states, there should be comparatively little difficulty in forming a segment of isolation extended chain. On the other hand, if the preferred conformation in the disordered chain is gauche, the chain will exhibit greater resistance to the formation of an isolated extended strand. If the bonds were subject to independent symmetric threefold rotation potentials, a useful approximation would be

$$\tau = 1/(1 + 2\sigma) \qquad (10)$$

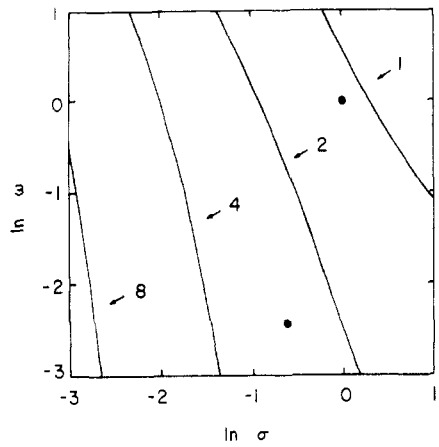

Figure 6. Behavior of $[\ln \delta/\ln \tau]^{1/2}$ when τ and δ are defined by eq 11 and 13, respectively. See text for meaning of filled circles.

where the trans state is assigned a statistical weight of 1, and σ is the weight for a gauche state. In the limit where σ approaches zero, the "disordered" state is predominantly trans, τ approaches 1, and there is no special penalty to be paid upon formation of an isolated extended strand. When σ becomes very large, most of the bonds occupy gauche states in the disordered chain, τ is very small, and a severe penalty is assessed to each isolated extended strand. If there is no preference for trans or gauche states in the disordered chain, eq 10 yields $\tau = 1/3$. It was this value that was employed in Figures 2–5.

Equation 10 should be modified if the rotation potentials become interdependent. This modification can be achieved in a simple fashion as

$$\tau = 1/\lambda_1 \tag{11}$$

$$\lambda_1 = [1 + \sigma(1 + \omega) + \{[1 - \sigma(1 + \omega)]^2 + 8\sigma\}^{1/2}]/2 \tag{12}$$

where ω is the penalty for g^+g^- or g^-g^- sequences (e.g., the weight of tg^+g^-t relative to tttt is $\sigma^2\omega$) and λ_1 is the largest eigenvalue of the statistical weight matrix.[10,11] It can be viewed as the contribution made to the conformational partition function by placing one additional bond in a long disordered chain. That contribution would have been the factor $(1 + 2\sigma)$ if the rotations were independent, i.e., if $\omega = 1$.

In order to estimate δ, it is assumed that bends require two successive gauche states of opposite sign. In the spirit of eq 11, the estimate of δ is

$$\delta = 2\sigma^2\omega/\lambda_1^2 \tag{13}$$

where recognition is given to the possibility that the bend might be g^+g^- or g^-g^+, and the denominator must now account for the contribution to the conformational partition function made by two bonds. If all states are equally probable in the disordered chain (i.e., if $\sigma = \omega = 1$), the value of δ is 2/9. It was this value which was used in Figures 2–5.

With these estimates of τ and δ, we return to eq 9, which focuses attention on $[\ln \delta/\ln \tau]^{1/2}$. Figure 6 presents a contour diagram showing the behavior of this term, as functions of σ and ω, where eq 11 and 13 provide the necessary definitions. The term $\sigma^2\omega$ vanishes as σ and/or ω approach zero, but λ_1 remains positive under these conditions. Consequently, large values of $[\ln \delta/\ln \tau]^{1/2}$, and hence long strands, are favored by small σ and small ω (arising from small values of δ). This conclusion is reasonable because a considerable price must be paid to form a bend if $\sigma^2\omega$ is small. If such a chain is forced into an intramolecular antiparallel sheet, that sheet should contain long strands.

Two filled circles occur in Figure 6. One is placed where $\sigma = \omega = 1$, corresponding to a disordered chain in which all conformations are equally weighted. The other filled circle occurs at $\sigma = 0.54$, $\omega = 0.088$. This combination of statistical weights provides a good description of unperturbed linear poly(methylene) chains.[10] Inserting these parameters for poly(methylene) in eq 11 and 13, we obtain $\tau = 0.54$ and $\delta = 0.015$. When compared with the chain having $\sigma = \omega = 1$, there has been a reduction in the difficulty in forming an isolated extended strand (τ has become larger because the trans state is preferred by disordered poly(methylene)), but a much greater penalty must be paid to make a bend (δ has become much smaller because disordered poly(methylene) contains few successive pairs of gauche states of opposite sign). The value of $[\ln \delta/\ln \tau]^{1/2}$ is now much larger than it was when $\delta = 2/9$, $\tau = 1/3$ or $\delta = \tau = 0.3$. We now inquire whether eq 9 provides a satisfactory description of the value of I_∞ which should be used in characterization of the coil-to-sheet transition in poly(methylene) chains. After the validity of the expression has been assessed, we shall address the simplications for poly(methylene).

Dependence of I_∞ on Edge Effects

Figures 7 and 8 illustrate the formation of intramolecular antiparallel sheets when $\delta = 0.015$ and $\tau = 0.54$. Figures 7 and 8 employ $n = 100$ and 400, respectively. Several different curves, computed at increments of 5 for I, are depicted in Figure 7. Curves using $I = 15$ and 20 are readily distinguishable in Figure 7, even though they were indistinguishable in Figure 2. The difference in behavior must be attributed to the change in assignments of edge effects because both figures use $n = 100$. Figure 7 shows that the curve with $I = 25$ is distinguishable from that using $I = 20$, $I = 25$ is barely distinguishable from $I = 30$, and $I = 30$ and 35 are indistinguishable. The value of I_∞ specified by eq 7 is 26. Consequently, eq 9 provides a useful quick estimate of the proportionality constant.

Distinguishable curves extend to much larger values of I in Figure 8. Here, the curve with $I = 50$ is barely distinguishable from that with $I = 40$ in the region near $t = 2.2$. Equation 7 predicts $I_\infty = 52$.

Figure 9 shows several normalized distribution functions for the strand lengths when $\delta = 0.015$ and $\tau = 0.54$. The values of t used are denoted by filled circles in Figures 7 and 8. Equation 7 provides a useful guide to the location of the tail at the high end of the distribution function near the completion of the transition when $n = 400$. A small portion of the distribution function slips past I_∞ near the end of the transition when n is only 100; i.e., eq 7 is a better approximation at larger values of n.

Implications for Poly(methylene)

The change in statistical weights from $\delta = \omega = 1$ to $\sigma = 0.54$, $\omega = 0.088$ causes the sheets formed at large t to have much longer strands (cf. Figures 5 and 9). The major cause is a marked decrease in the ease with which sharp bends are formed, although there is also a contribution from an increase in the ease of forming isolated segments of extended chain. Large sheets are produced in the final portion of the transition. Comparison of Figures 7 and 8 shows that this part of the transition becomes increasingly more dominant, and moves to smaller t, as n increases. As n becomes infinite, the chain should exhibit a highly cooperative formation of large sheets at values of t slightly larger than 1. This sharp transition would be preceded by noncooperative formation of much smaller sheets when t is less than unity. Distribution functions for the chain with $n = 400$ in Figure 9 suggest that the various large

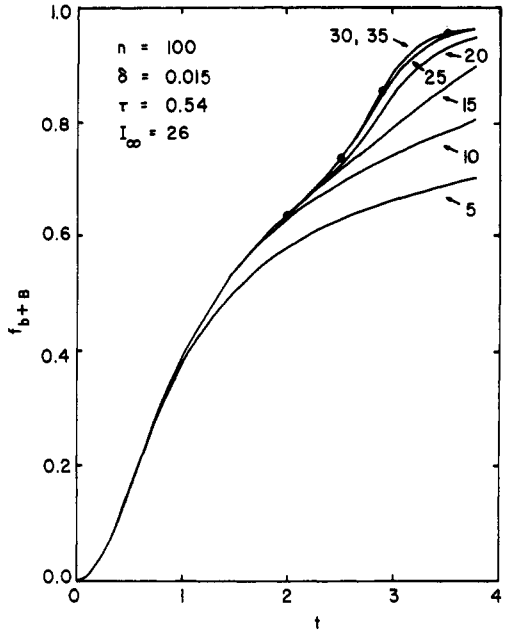

Figure 7. Formation of intramolecular antiparallel sheets, evaluated at the indicated values of I, when $n = 100$, $\delta = 0.015$, and $\tau = 0.54$. The value of I_∞ of eq 7 is 26.

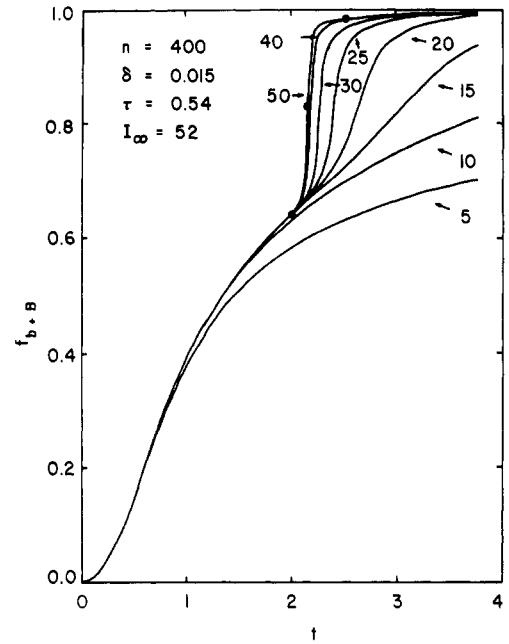

Figure 8. Formation of intramolecular antiparallel sheets, evaluated at the indicated values of I, when $n = 400$, $\delta = 0.015$, and $\tau = 0.54$. The value of I_∞ of eq 7 is 52.

sheets would contain strands of many different lengths.

In assessing the relevance of Figures 7–9 for the crystallization of a single poly(methylene) chain in dilute solution, careful attention must be given to restrictions imposed by the nature of the sheets assumed in the present model. Without exception, tight bends connect all contiguous strands in every sheet. The model and partition function were formulated in this fashion in recognition of the important role played by β-bends in the antiparallel sheets formed by proteins.[12] Since such chains can be forced into a sheet at sufficiently large values of t, no significance must be attributed to the fact that the chains form sheets with adjacent reentry. The model used to obtain this ensemble permits no other option.

A legitimate conclusion having relevance for poly(methylene) derives from the qualitative changes seen in δ and τ upon going from $\sigma = \omega = 1$ to $\sigma = 0.54$, $\omega = 0.088$.

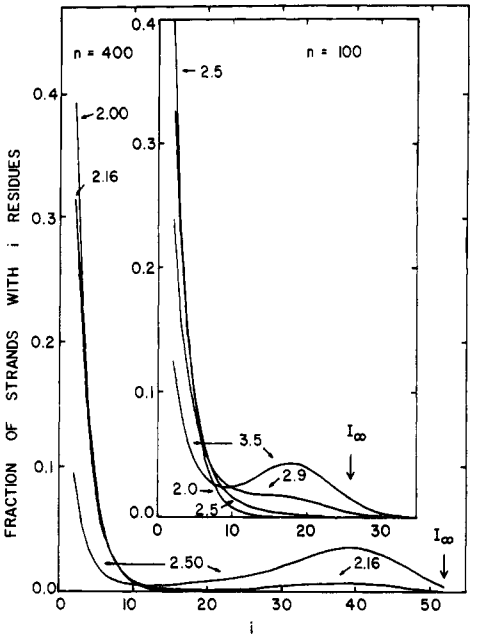

Figure 9. Normalized distribution functions for the number of residues per strand at the indicated values of t and n when $\delta = 0.015$ and $\tau = 0.54$. The values of t are indicated by filled circles on the transition curves depicted in Figures 7 and 8.

The bend statistical weight, δ, falls by more than an order of magnitude when one takes account of the conformational preferences of the poly(methylene) chain. Since δ is still greater than zero, tight bends are not completely disallowed and sheets containing such bends will dominate the present ensemble at sufficiently large values of t. However, a real poly(methylene) chain might utilize alternative means of crystallization, such as that envisioned in "switchboard" models, which avoid the penalty arising from the δ associated with each tight bend. The alternative means of connecting two crystalline subchains need only to be able to compete with δ, which has been rendered small by the short-range properties of the real poly(methylene) chain. Tight bends are more likely to occur in chains having large values of δ.

Since all chains in the present ensemble behave independently, the behavior depicted in Figures 7–9 should not be interpreted as that expected for bulk-crystallized poly(methylene). It applies to formation of intramolecular antiparallel sheets in a single chain in dilute solution.

Acknowledgment. This research was performed while W.L.M. was on sabbatical leave at Cornell University during 1983–1984. W.L.M. gratefully acknowledges sabbatical support from the National Institutes of Health (Senior National Research Service Award 1 F33 AM-07296-01). This research was also supported by research grants to H.A.S. from the National Institutes of Health (AM-08465) and from the National Science Foundation (PCM79-20279) and to W.L.M. from the National Science Foundation (PCM81-18197).

References and Notes

(1) Mattice, W. L.; Scheraga, H. A. *Biopolymers* **1984**, *23*, 1701.
(2) Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **1953**, *39*, 253.
(3) Mattice, W. L.; Lee, E.; Scheraga, H. A. *Can. J. Chem.*, in press.
(4) Astbury, W. T.; Dickinson, S.; Bailey, K. *Biochem. J.* **1935**, *29*, 2351.
(5) Geddes, A. J.; Parker, K. D.; Atkins, E. D. T.; Beighton, E. *J. Mol. Biol.* **1968**, *32*, 343.
(6) Sternberg, M. J. E.; Thornton, J. M. *J. Mol. Biol.* **1977**, *110*, 285.
(7) Richardson, J. S. *Nature (London)* **1977**, *268*, 495.

(8) Poland, D.; Scheraga, H. A. "Theory of Helix–Coil Transitions in Biopolymers"; Academic Press: New York, 1970.
(9) Mattice, W. L.; Scheraga, H. A. *Polym. Prepr., Div. Polym. Chem., Am. Chem. Soc.* **1984**, *25*, 276.
(10) Abe, A.; Jernigan, R. L.; Flory, P. J. *J. Am. Chem. Soc.* **1966**, *88*, 631.
(11) Flory, P. J. *Macromolecules* **1974**, *7*, 381.
(12) Némethy, G.; Scheraga, H. A. *Q. Rev. Biophys.* **1977**, *10*, 239.

Double-Stranded Helix of Xanthan: Dimensional and Hydrodynamic Properties in 0.1 M Aqueous Sodium Chloride

Takahiro Sato, Takashi Norisuye,* and Hiroshi Fujita

Department of Macromolecular Science, Osaka University, Toyonaka, Osaka 560, Japan.
Received June 1, 1984

ABSTRACT: Sedimentation velocity data were obtained for 12 samples of xanthan (sodium salt) in 0.1 M aqueous NaCl at 25 °C, in which this polysaccharide dissolves as a double helix. For two of these samples, light scattering and viscosity data in the same solvent were obtained and combined with the previous data for other samples. The limiting sedimentation coefficients, radii of gyration, and intrinsic viscosities as functions of weight-average molecular weight M_w showed the double helix of xanthan in 0.1 M aqueous NaCl to be almost completely rigid below and semiflexible above $M_w \sim 3 \times 10^5$. Analysis of these data in terms of the known theories for rods and wormlike chains yielded 0.47 ± 0.02, 2.4 ± 0.3, and 120 ± 20 nm for the pitch h (per main-chain glucose residue), diameter d, and persistence length of the xanthan double helix, respectively. These h and d values agreed with those reported for crystalline xanthan, confirming the previous conclusion that the double-helical structure of the polysaccharide in 0.1 M aqueous NaCl is essentially the same as that in the crystalline state.

Introduction

Xanthan is an ionic, extracellular polysaccharide produced by the bacterium *Xanthomonas campestris*. It consists of a main chain of β-1,4-linked D-glucoses and three sugar side chains attached to every other main-chain residue.[1,2] In recent papers,[3,4] we concluded from light scattering and viscosity measurements that xanthan (sodium salt) dissolves in 0.1 M aqueous sodium chloride (NaCl) as rodlike dimers having essentially the same double-stranded helical structure as that proposed by Okuyama et al.[5] for crystalline xanthan.

To confirm this conclusion from another angle the limiting sedimentation coefficient was measured on 12 samples of xanthan sodium salt in 0.1 M aqueous NaCl. Further, light scattering and viscosity measurements were made on two of these samples in the same solvent, since they had not been treated in our previous work.[3,4]

In this paper, all the data from the present and previous experiments were combined to redetermine the pitch, diameter, and rigidity (expressed in terms of the persistence length) of the xanthan double helix in 0.1 M aqueous NaCl.

Experimental Section

Samples. Twelve fractionated, purified samples of xanthan (X4-5, X5-6, X5-8, X3-5, X9-3, X7-3b, X6-3-7, X6-4-4, X10-4, X6-4-7, X8-3-5, and X8-3-8) were chosen from our stock for sedimentation velocity determination. Among these, samples X5-8 and X6-4-7 were the objects of new light scattering and viscosity measurements. Sample X4-5 was extracted directly from a commercial sample (Kelco Keltrol), while the others were all obtained by sonicating the commercial sample.[3,4] Each of these samples was converted to an almost completely substituted Na salt, and 0.1 M aqueous NaCl solutions of the polysalts were prepared by the method[3] described previously. The degree of pyruvation (DS_{pyr}) ranged from 0.32 to 0.37 with no systematic variation with the molecular weight of the sample.[6] The ratios of the z-average to weight-average molecular weights M_z/M_w determined from sedimentation equilibrium data[3] for samples X6-4-4, X8-3-5, and X8-3-8 were 1.1_5, 1.1_4, and 1.1_6, respectively.

Light Scattering. Light scattering on samples X5-8 and X6-4-7 in 0.1 M aqueous NaCl at 25 °C was studied with a Fica 50 light scattering photometer in an angular range θ from 22.5 to 150°. Vertically polarized incident light of 436- or 546-nm wavelength was used. The experimental procedures and data analysis employed were the same as those described previously.[3]

As was found in our previous study,[3] Na salt xanthan dissolves in cadoxen [tris(ethylenediamine)cadmium dihydroxide] as single flexible chains. Light scattering measurement was also made on samples X5-8 and X6-4-7 in this solvent to check whether, as found previously,[3] M_w(in 0.1 M NaCl)/M_w(in cadoxen) for either sample comes close to 2.

Sedimentation Velocity. Sedimentation velocity measurements on all 12 Na salt xanthan samples in 0.1 M aqueous NaCl at 25 °C were carried out in a Beckman Model E ultracentrifuge, using a Kel-F 30-mm single-sector cell. The rotor speed chosen was 44 000 or 48 000 rpm. Sedimentation coefficients s for a series of initial polymer mass concentrations c_0 were determined by the usual peak method and analyzed by the equation $s^{-1} = s_0^{-1}(1 + k_s c_0)$ to obtain the limiting sedimentation coefficient s_0 and the constant k_s.

Table I
Results from Light Scattering Measurements on Na Salt Xanthan Samples in 0.1 M Aqueous NaCl at 25 °C

sample	$M_w \times 10^{-4}$	$A_2 \times 10^4$/(cm^3 mol g^{-2})	$\langle S^2 \rangle^{1/2}$/ nm	M_w (in 0.1 M NaCl) / M_w (in cadoxen)
X4-5	740	3.17	378	2.08
X5-6	394	2.91	257	2.01
X5-8	256	3.08	208	1.83
X3-5	142	3.99	142	1.91
X9-3	99.4	3.79	108	2.02
X7-3b	60.3	4.35	74.8	2.04
X6-3-7	36.2	4.77	50.8	2.05
X6-4-4	24.0	5.02	36.3	1.97
X10-4	20.9	5.34	34.0	1.90
X6-4-7	16.4	5.64	25.1	1.89
X8-3-5	11.2	6.18	16.6	2.47
X8-3-8	7.40	6.80	$10_{.8}$	2.13

Results

Light Scattering Data. Figures 1 and 2 illustrate, respectively, the concentration dependence of $(Kc/R_0)^{1/2}$

0024-9297/84/2217-2696$01.50/0