(14) N. G. McCrum, B. Read, and G. Williams, "Anelastic and Dielectric Effects in Polymer Solids", Wiley, New York, 1967.
(15) M. E. Baird, "Electrical Properties of Polymeric Materials", The Plastics Institute, London, 1973.
(16) R. W. Sillars, *J. Inst. Elec. Eng.*, **80**, 378 (1937).
(17) L. K. H. Van Beek, *Prog. Dielectr.*, **7**, 69 (1967).
(18) M. E. Baird, *Rev. Mod. Phys.*, **40**, 219 (1968).
(19) S. B. Dev, A. M. North, and J. C. Reid in "Dielectric Properties of Polymers", F. E. Karaz, Ed., Plenum Press, New York, 1972.
(20) W. P. Harris, *Annu. Rep. Conf. Electr. Insul. Dielectr. Phenom.*, **No. 1481**, 72 (1967).
(21) S. Robert, Report No. 66-0-333, General Electric R&D Center, Schenectady, N.Y., 1966.
(22) J. G. Berberian and R. H. Cole, *Rev. Sci. Instrum.*, **40** (6), 811 (1969).
(23) R. S. Lakes and R. A. Harper, *Rev. Sci. Instrum.*, **46** (11), 1583 (1975).
(24) M. Gordon and J. S. Taylor, *J. Appl. Chem.*, **2**, 493 (1952).
(25) I. Kirshenbaum, *J. Polym. Sci., Part A*, **3**, 1869 (1965).
(26) U. Alter and R. Bonart, *Colloid Polym. Sci.*, **254**, 348 (1976).
(27) F. P. Warner, D. S. Brown, and R. E. Wetton, *J. Chem. Soc., Faraday Trans. 2*, **72**, 1064 (1976).
(28) I. H. Hall and M. G. Pass, *Polymer*, **17**, 807 (1976).
(29) C. R. Ashcraft and R. H. Boyd, *J. Polym. Sci., Part A-2*, **14**, 2153 (1976).
(30) R. E. Wetton and G. Williams, *Trans. Faraday Soc.*, **61**, 2132 (1965).
(31) R. E. Wetton, G. S. Fielding-Russell, and K. U. Fulcher, *J. Polym. Sci., Part C*, **30**, 219 (1970).
(32) S. de Petris, V. Frosini, and E. A. Nicol, *J. Polym. Sci., Part A-2*, **15**, 1121 (1977).
(33) F. Stern and C. Weaver, *J. Phys. C*, **3**, 1736 (1970).
(34) J. R. Macdonald, *J. Chem. Phys.*, **54**, 2026 (1971).
(35) C. T. O'Konski, *J. Phys. Chem.*, **64**, 605 (1960).
(36) H. Looyenga, *Physica (Utrecht)*, **31**, 401 (1965).
(37) W. E. A. Davies, *Br. J. Appl. Phys.*, **7**, 120 (1974).

Conformational Energy Minimization in the Approximation of Limited Range Interactions

Robert L. Jernigan* and Shousun C. Szu‡

Laboratory of Theoretical Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205. *Received March 19, 1979*

ABSTRACT: A general method to determine the conformational free energy minimum is developed by approximating the total molecular free energy as the sum of contributions from independent segments, each of which can exist in several conformations. The simple algorithm to determine the best overall conformation rigorously yields the same results as those obtained by enumerating all possible combinations of independent segments. However, the method, based upon dynamic programming concepts, is much simpler because it avoids explicit generation of all possible combinations of segments. Application of the method to protein secondary structures is discussed. Development of this conformational selection scheme is independent of the method by which free energies have been calculated. An example is presented for determining relative preferences for α-helix and β-strand in a fragment of trypsin inhibitor.

The simplest considerations based on small molecules would lead to the postulation of myriad forms for large molecules; however, frequently a unique conformation is observed for many large biological macromolecules. Determining such preferred conformations for proteins has been attempted by traversing conformational space to determine the conformations of lowest free energy.[1] An exhaustive search of conformational space is usually impossible for all except the smallest molecules; therefore, the existence of numerous local minima renders the detection of a global minimum particularly difficult. Here we are attempting to approximate molecular free energies and to shorten the list of conformations in order to simplify the search for minima.

Energies are approximated by including only interactions within molecular segments. The total molecular energy is approximated as the sum of segment energies; the assumption is that the intrasegment interactions will include the majority of all molecular interactions. Application of a concept from dynamic programming permits a shortening of the number of combinations of segments which must be considered.

Method

The method presented below is a practical approach to this problem. Interatomic interactions here are to be limited to those within single regularly structured regions. All possible regions are considered; the combination of such regions with the lowest total free energy is determined. Because of the compact nature of globular proteins, an individual regular region usually includes only a small fraction of the atoms in the whole molecule. A regular region in a protein is taken to be one in which the backbone rotation angles are identical, within some permitted range of variation. Interatomic interactions which can be included here are all those within each such region, of both short and medium range.

Consider a linear molecule of N residues, each of which can exist in one of ζ regular conformational states. For this approach, it is necessary to specify k, the maximum length of such regions. This will correspond to the maximum range of interactions to be included. Further, it is assumed here that there are no interactions between separate regular regions; such interactions could be included by a simple modification of the method. In the present version of this method, it is important that the regions be relatively long, so as to include, on the average, the majority of interatomic interactions.

The free energy of a region encompassing residues $i-j$ through i is given as $F^{\xi_{i-j}\xi_{i-j+1}\cdots\xi_i}_{i-j,i}$, where the conformational states of each residue included in the region are specified by the ξ's. The present method requires that a complete set of such free energies can be calculated and is available. Calculation of these energies must involve approximations; that is the subject of other work.[2]

The formalism of the selection process is necessarily directional. The choice of a direction in which to proceed is, however, arbitrary; with independent regions, the results are independent of this direction. Here we proceed from

‡ Division of Bacterial Products, Bureau of Biologics, Food and Drug Administration, Bethesda, Maryland 20205.

residue one toward residue N; residue one is designated to be the amino chain terminus. If nucleation sites[3,4] were known, these could be designated as initiation loci. Introduction of interactions between regular regions may lead to a dependence on the starting point.

The free energy is minimized stepwise along the linear sequence; to be determined is the combination of regions with lowest total free energy from the origin to sites successively distant. The lowest free energy of residue one is given as:

$$F^*_{1,1} = \min\ (F^1_{1,1}, F^2_{1,1}, \ldots, F^\zeta_{1,1}) \tag{1}$$

The form associated with this minimum is retained along with the value of $F^*_{1,1}$ for later use in combination with regions with an amino terminus at unit 2.

The lowest free energy of sites 1 and 2 in all permitted forms is given by:

$$\begin{aligned} F^*_{1,2} = \min\ [\ & F^{1,1}_{1,2}, F^{1,2}_{1,2}, \ldots, F^{1,\zeta}_{1,2}, \\ & F^{2,1}_{1,2}, F^{2,2}_{1,2}, \ldots, F^{2,\zeta}_{1,2}, \\ & \cdot \\ & \cdot \\ & \cdot \\ & F^{\zeta,1}_{1,2}, F^{\zeta,2}_{1,2}, \ldots, F^{\zeta,\zeta}_{1,2}] \end{aligned} \tag{2}$$

If the set $F^{\xi_i \cdots \xi_j}_{i,j}$ for all possible values of the ξ's is denoted by $\mathcal{F}_{i,j}$, then eq 1 and 2 are simplified:

$$F^*_{1,1} = \min\ [\mathcal{F}_{1,1}] \tag{1'}$$

and

$$F^*_{1,2} = \min\ [\mathcal{F}_{1,2}] \tag{2'}$$

For the ith site, the minimum free energy is given by:

$$F^*_{1,i} = \min\ [\mathcal{F}_{1,i}] \text{ for } i \leq k \tag{3}$$

or

$$F^*_{1,i} = \min\ [F^*_{1,i-k} + \mathcal{F}_{i-k+1,i}] \text{ for } i > k \tag{4}$$

Note that imposing the separation at residue $i - k$ is highly arbitrary and in certain specific cases may be a very poor approximation. It is possible to proceed stepwise, using eq 3 and 4 until the carboxyl chain terminus is reached and a minimum free energy form is determined for the entire molecule.

A major simplification is introduced by reducing the set of conformations to those for regions consisting of only regularly repeating conformations. The equations above are modified as follows: The set $\mathcal{F}_{i,j}$ now consists of only ζ forms since all $\xi_i = \xi_{i+1} = \ldots = \xi_j$. Additional terms arise because breaks between the regular conformations must be permitted at each site. This overcomes, in one sense, the objection noted above about the break site in eq 4. Equations 3 and 4 become

$$F^*_{1,i} = \min\ [\mathcal{F}_{1,i}, (F^*_{1,1} + \mathcal{F}_{2,i}), (F^*_{1,2} + \mathcal{F}_{3,i}), \ldots, (F^*_{1,i-1} + \mathcal{F}_{i,i})] \quad \text{for } i \leq k \tag{5}$$

or

$$F^*_{1,i} = \min\ [(F^*_{1,i-k} + \mathcal{F}_{i-k+1,i}), (F^*_{1,i-k+1} + \mathcal{F}_{i-k+2,i}), \ldots, (F^*_{1,i-1} + \mathcal{F}_{i,i})] \quad \text{for } i > k \tag{6}$$

There will be at most k terms, as grouped in the parentheses, on the right-hand side of eq 6. The example to follow demonstrates one difficulty with this formulation; we have not excluded adjacent regions of the same conformation. This can be corrected simply by requiring that adjacent regions be in different conformations. It complicates the formulation because now instead of retaining a single $F^*_{1,i}$ there will be ζ of them, corresponding to one for each of the possible conformations of residue i.

$$(F^*_{1,1})_{\xi_1} = F^{\xi_1}_{1,1} \quad \text{for } \xi_1 = 1, \ldots, \zeta \tag{7}$$

$$(F^*_{1,2})_{\xi_2} = \min\ [F^{\xi_1\xi_2}_{1,2}, ((F^*_{1,1})_{\xi_1 \neq \xi_2} + F^{\xi_2}_{2,2})] \quad \text{for } \xi_2 = 1, \ldots, \zeta \tag{8}$$

If $\mathcal{G}^{\xi_2}_{1,1}$ is the set $(F^*_{1,1})_{\xi_1=j}$, for all j except the value of ξ_2, then the general forms are:

$$(F^*_{1,i})_{\xi_i} = \min\ [F^{\xi_i}_{1,i}, (\mathcal{G}^{\xi_i}_{1,1} + F^{\xi_i}_{2,i}), (\mathcal{G}^{\xi_i}_{1,2} + F^{\xi_i}_{3,i}), \ldots, (\mathcal{G}^{\xi_i}_{1,i-1} + F^{\xi_i}_{i,i})] \quad \text{for } i \leq k \tag{9}$$

or

$$(F^*_{1,i})_{\xi_i} = \min\ [(\mathcal{G}^{\xi_i}_{1,i-k} + F^{\xi_i}_{i-k+1,i}), (\mathcal{G}^{\xi_i}_{1,i-k+1} + F^{\xi_i}_{i-k+2,i}), \ldots, (\mathcal{G}^{\xi_i}_{1,i-1} + F^{\xi_i}_{i,i})] \quad \text{for } i > k \tag{10}$$

where $\mathcal{G}^{\xi_i}_{1,i-j}$ is the set of $(F^*_{1,i-j})_{\xi_{i-j}}$, except the one for which $\xi_{i-j} = \xi_i$.

Application

As an example, we consider residues 2 to 7 of bovine pancreatic trypsin inhibitor. Only two regular conformations are considered, α-helix and β-strand. All possible positions and lengths of such regions are considered. Free energies were calculated by a method described in ref 2. In this approximate method, charges are placed on ionized side chain atoms; partial charges are assigned to polar side chain and backbone atoms. Backbone – side chain electrostatic energies are calculated for each segment. An additional positive energy is added for each α-helical residue in the form of a parameter representing all backbone–backbone interactions. The value of this helical parameter used to obtain the energies here was 2.3 kcal/mol of residues for a dielectric constant of unity. All such possible segments and their associated free energies are presented in Table I. Magnitudes of the free energies appear quite large because they represent free energies times the dielectric constant. Specification of typical dielectric constants would yield much smaller free energies.

Application of eq 5 and 6 to the data in Table I gives, successively

$$F^*_{2,2} = 0.0 \text{ for } \beta_2$$

$$F^*_{2,3} = -6.5 \text{ for } \beta_2\alpha_3$$

$$F^*_{2,4} = -6.9 \text{ for } \beta_2\alpha_{3\text{-}4}$$

$$F^*_{2,5} = -7.8 \text{ for } \beta_2\alpha_{3\text{-}5}$$

$$F^*_{2,6} = -8.8 \text{ for } \beta_2\alpha_{3\text{-}6}$$

$$F^*_{2,7} = -12.2 \text{ for } \beta_2\alpha_{3\text{-}6}\alpha_7$$

The last line is the best conformation for this protein fragment with this method, but it is peculiar since two adjacent independent helical regions were chosen. This conformation is actually $\beta_2\alpha_{3\text{-}7}$ whose energy of −5.5 is much less favorable than −12.2; numerous other combinations of regions yield energies more favorable than −5.5.

Application of eq 9 and 10 yields regions of strictly alternating conformations:

$$(F^*_{2,2})_\alpha = 2.3;\ (F^*_{2,2})_\beta = 0.0$$

$$(F^*_{2,3})_\alpha = -6.5 \text{ for } \beta_2\alpha_3;\ (F^*_{2,3})_\beta = -2.8 \text{ for } \beta_{2\text{-}3}$$

$$(F^*_{2,4})_\alpha = -6.9 \text{ for } \beta_2\alpha_{3\text{-}4};\ (F^*_{2,4})_\beta = -6.5 \text{ for } \beta_2\alpha_3\beta_4$$

$$(F^*_{2,5})_\alpha = -7.8 \text{ for } \beta_2\alpha_{3\text{-}5};\ (F^*_{2,5})_\beta = -7.0 \text{ for } \beta_2\alpha_{3\text{-}4}\beta_5$$

$$(F^*_{2,6})_\alpha = -8.8 \text{ for } \beta_2\alpha_{3\text{-}6};\ (F^*_{2,6})_\beta = -7.8 \text{ for } \beta_2\alpha_{3\text{-}5}\beta_6$$

$$(F^*_{2,7})_\alpha = -11.2 \text{ for } \beta_2\alpha_{3\text{-}5}\beta_6\alpha_7;\ (F^*_{2,7})_\beta = -10.0 \text{ for } \beta_2\alpha_{3\text{-}6}\beta_7$$

Table I
All Regular α and β Conformations in Residues 2-7 of Bovine Pancreatic Trypsin Inhibitor

no.	residues	conformation	energy, kcal/mol
1	2-2	α	2.3
2	2-2	β	0.0
3	2-3	α	−5.2
4	2-3	β	−2.8
5	2-4	α	−5.6
6	2-4	β	−3.7
7	2-5	α	−6.2
8	2-5	β	−4.5
9	2-6	α	−7.3
10	2-6	β	−4.9
11	2-7	α	−3.0
12	2-7	β	−5.4
13	3-3	α	−6.5
14	3-3	β	−2.3
15	3-4	α	−6.9
16	3-4	β	−3.3
17	3-5	α	−7.8
18	3-5	β	−4.1
19	3-6	α	−8.8
20	3-6	β	−4.4
21	3-7	α	−5.5
22	3-7	β	−5.0
23	4-4	α	2.3
24	4-4	β	0.0
25	4-5	α	4.4
26	4-5	β	−0.06
27	4-6	α	6.6
28	4-6	β	−0.1
29	4-7	α	9.8
30	4-7	β	−0.7
31	5-5	α	2.0
32	5-5	β	−0.1
33	5-6	α	4.2
34	5-6	β	−0.2
35	5-7	α	1.6
36	5-7	β	−1.0
37	6-6	α	2.3
38	6-6	β	0.0
39	6-7	α	−2.1
40	6-7	β	−1.4
41	7-7	α	−3.4
42	7-7	β	−1.2

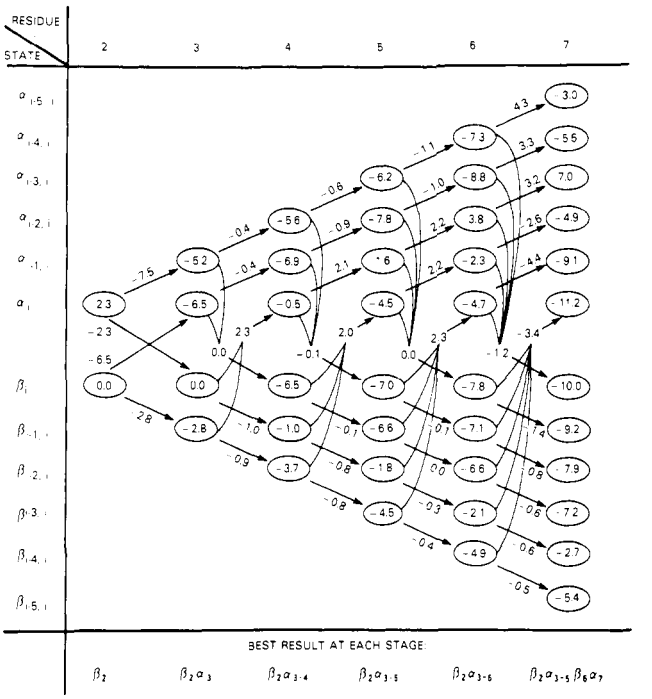

Figure 1. Selection scheme. Each column represents the *i*th stage in the selection which corresponds to the portion of the molecule from residue 2 through residue *i*. Subscripts on the row indices are the terminal residues of the last secondary region. The circled numbers are the best cumulative energies which include the conformation indicated by the row index. Numbers on the arrows represent increments of energy for adding one residue in the indicated conformation. Only two types of addition are permitted: growth of a helix or β-strand or change from either of these conformations to the other. Data are taken from Table I. Upon changing from one conformation to the other, a choice is made among all previous forms; this is represented by the multiple arrows leading to one circled number. The assumed independence of secondary regions permits a choice of the minimum energy conformation to be made rigorously at each such step. The choice of the best conformation from all values listed in the column for each stage is given in the bottom row. In this example, only in the last stage has minimization changed the results of decisions at the previous stages.

A final choice must be made between the two forms on the last line corresponding to the best conformations for the fragment with the carboxyl terminal residue in either α-helix or β-strand form. From the free energies given, the best choice is $\beta_2\alpha_{3-5}\beta_6\alpha_7$. The nature of this method is such that inclusion of successive residues may modify choices made previously for residues in shorter fragments; thus "long-range" effects may appear.

The present methods represent an application of techniques of multistage decision methods which are usually termed dynamic programming.[5] Results corresponding to the last selection scheme have been arranged in Figure 1 to make clearer the separation into separate stages. Data were taken directly from Table I. The method provides a significant shortening over consideration of all combinations of secondary regions by discarding some unfavorable combinations at each stage. Thus, on the figure when multiple lines lead to one circle only the lowest energy value of previous choices is retained.

Discussion

The present method provides the formalism for minimizing the sum of secondary structure free energies. Such a method has its most obvious application to determine good preliminary sets of secondary forms. As an example, the method of Ptitsyn and Finkelstein[6] required such a determination. In the form given here, a computer program to determine the best secondary forms has been developed and could be used with most secondary structure determining methods. Use of the present selection scheme would avoid some of the arbitrary features of these methods. Any method which yields free energies with nearest neighbor or higher order interresidue interdependences can be combined suitably with the present selection scheme. Use of methods to calculate energies without any interactions between residues would yield, with the present selection scheme, the trivial result of the best choice for each individual residue which can be determined independently of all other residues.

Breaks between separate secondary regions represent barriers through which there are no interactions. This is clearly contrary to the physical situation. However, propagation of the secondary structure regions on either side beyond the break is unfavored. An improvement beyond the examples given above would consist of also including the conformations corresponding to various types of turns.

Certain reductions are possible. For example, those regions of higher free energy, relative to those in the same conformation which are completely spanned by the first, could be discarded. The shorter more favorable forms would always be chosen. Following such a condensation,

it is sometimes possible to locate complete breaks among the remaining conformations. There may be boundaries between two adjacent residues which are never included within a single region. This permits treatment of such molecular fragments in a completely independent way.

Application of this reduction to the conformations in Table I shortens the list by eliminating forms 3, 5, 7, 9, 11, 21, 25, 26, 27, 28, 29, 30, 33, 35, 36, and 39. The original 42 forms are reduced to 26 in this way; however, there are no independent fragments since conformation 12, which spans the whole fragment, is retained. If alternation of conformations were not imposed, then the list could be shortened further by retaining only the most favored conformation for each specific region, e.g., by comparing the free energy of α_{3-5} from Table I with that of β_{3-5} and eliminating α_{3-5}.

The results obtained by the present method are very similar to those obtained by the authors, using another selection scheme.[2] That "conformation stability" selection can be applied to the set of conformations in Table I. This method follows: (1) all conformations are placed in order by their associated free energies, (2) the most favorable conformation is chosen, (3) the next most favorable form is chosen if it does not include residues previously chosen, and (4) the third step is repeated until every residue is assigned a conformation. Application to the data in Table I yields −12.2 for $\beta_2\alpha_{3-6}\alpha_7$ or −10.0 for $\beta_2\alpha_{3-6}\beta_7$, depending on whether or not the condition of alternation is imposed. A comparison with the previous examples indicates that there is a tendency for the free energy minimization to yield somewhat shorter, less regular forms than does the conformation stability scheme. The energy minimization scheme would be expected to reflect somewhat more directly any inadequacies in free energy calculations because it does not depend so much on the effects of averaging over longer segments. Results with both methods are usually approximately the same; this provides an indication that the conformation stability selection gives conformations which closely approximate minimum free energy forms. Both methods yield the most significant feature of the present fragment, namely, the short helix. Experimental X-ray results[7] indicate that the correct form is approximately $\beta_2\alpha_{3-6}\beta_{7-10}$.

The present energy minimization method yields a global minimum, within the approximation of representing the total free energy as the sum of free energies of independent secondary structure regions. It should be possible to impose inter-region energy calculations at each selection step in the energy minimization method. This might eliminate the problem of choosing too many single independent residue conformations and would incorporate some of the tertiary interactions; however, it may introduce a dependence on the selection pathway.

References and Notes

(1) G. Nemethy and H. A. Scheraga, *Q. Rev. Biophys.*, **10**, 239 (1977).
(2) R. L. Jernigan, S. Miyazawa, and S. C. Szu, to be published.
(3) R. R. Matheson, Jr., and H. A. Scheraga, *Macromolecules*, **11**, 819 (1978).
(4) V. I. Lim, *Dokl. Akad. Nauk SSSR*, **222**, 1467 (1975).
(5) D. T. Phillips, A. Ravindran, and J. J. Solberg, "Operations Research: Principles and Practice", Wiley, New York, 1976, pp 419–468.
(6) O. B. Ptitsyn and A. V. Finkelstein, *Biophysics*, **15**, 785 (1970).
(7) R. J. Feldmann, "Atlas of Macromolecular Structure on Microfiche", Structure AM 3.6.1.1.1 Tracor Jitco Inc., Rockville, Md. 1977.

Infinite-Dilution Oscillatory Flow Birefringence Properties of Four-Armed Star Polystyrene

Alan L. Soli and John L. Schrag*

Department of Chemistry and Rheology Research Center, University of Wisconsin, Madison, Wisconsin 53706.

Received July 27, 1979

ABSTRACT: The oscillatory flow birefringence properties (characterized by the magnitude S_m and relative phase angle θ_s of the complex mechanooptic coefficient S^*) of solutions containing what were thought to be four-armed polystyrene stars (equal length arms linked with $SiCl_4$) with narrow distribution number average molecular weights of 2.34×10^5 and 8.58×10^5 have been measured over a frequency range from 10^0 to 2×10^3 Hz at several temperatures. The solvent was Aroclor 1248, lot KM 502, a chlorinated biphenyl. The data have been extrapolated to obtain infinite dilution properties which are compared with the predictions of the Zimm–Kilb theory. The fits obtained indicate that the samples have substantial amounts of functionality polydispersity; the supposedly four-arm molecules apparently consist of a mixture of two-, three-, and four-armed structures, with the major component having three arms. Also, the values of the hydrodynamic interaction parameter required to generate theoretical fits to the data are essentially the same as those required to fit infinite dilution oscillatory flow birefringence and viscoelastic properties of linear polystyrenes in Aroclor solvents but are substantially different from those obtained for star polystyrene molecules from infinite dilution viscoelasticity studies. The precision obtainable with the oscillatory flow birefringence technique suggests that it may be a useful method for detecting small amounts of long-chain branching.

There have been several studies of the frequency dependence of the oscillatory flow birefringence (OFB) and viscoelasticity (VE) of linear polymers in solutions sufficiently dilute to enable extrapolation to obtain infinite dilution properties.[1-9] The theory of Zimm[10] has been shown to be in very good agreement at low frequencies with the above data when exact eigenvalues obtained by the method of Lodge and Wu[11,12] are used. Branched polymers have also been of interest since the long time end of the relaxation spectrum is affected substantially by overall chain topology. Several viscoelastic studies have been reported for polymers with a variety of types of branching, some of which have included extrapolations to obtain infinite dilution properties.[13-19] Again these data have agreed quite well at low frequencies with the Zimm–Kilb theory[20]

* Address correspondence to this author at the Department of Chemistry.

0024-9297/79/2212-1159$01.00/0