cell base dimensions between α- and γ-PVF_2, it is not unreasonable to contemplate existence of an antipolar analogue of the γ phase (ϵ-PVF_2).

Irrespective of the situation prevailing in epitaxially grown crystals, existence of such a fifth polymorph of PVF_2 may be suggested with a certain degree of caution based upon our knowledge of solid-state transformations of $TGT\bar{G}$ chains at high temperatures.[5,6,11,12] Such transformations have been studied by Lovinger[6,11] in spherulites of the α phase, by Takahashi and Tadokoro[5] in drawn films of α-PVF_2, and by Servet and Rault[12] in uniaxially oriented samples of the δ form; in all three cases, the transformation has been found to yield chains of a $T_3GT_3\bar{G}$ conformation. Since the high-temperature solid-state transformation requires only *intra*molecular rearrangements,[5,6,11] it is reasonable to expect that the difference in chain packing between the α and δ phases might be reflected in the packing of $T_3GT_3\bar{G}$ chains after annealing, leading to two different polymorphs. One might then enquire whether or not the $TGT\bar{G} \rightarrow T_3GT_3\bar{G}$ transformation is also accompanied by dipole inversion in alternate chains—a process which may be achieved by rotation of every second chain by 180° or, as recently suggested,[13] by relatively small, intramolecular bond rotations. Intuitively, one would not expect annealing to cause dipole inversions in half of the chains. However, Takahashi and Tadokoro[5] found a reduction in the intensity of the 100 reflection during annealing of α-PVF_2, while Servet and Rault[12] reported an increase after heat treatment of the δ phase; both data are consistent with dipole inversions and may suggest, therefore, that high-temperature annealing causes the antipolar α phase to transform to the *polar* γ phase and the polar δ phase to transform to the *anti*polar ϵ phase. This somewhat surprising possibility is under investigation and will be discussed on the basis of current experiments in a subsequent publication.

Conclusions

It has been found that γ-type molecular chains pack in a unit cell with a slight monoclinic angle β (=93°), in agreement with the model of Takahashi and Tadokoro;[5] the *b*-axis length is, however, closer to the value reported by Weinhold, Litt, and Lando.[2,4] Epitaxial crystallization of PVF_2 on NaCl yields γ-phase crystals grown with one of their {110} planes parallel to the substrate. Possible existence of an antipolar analogue of γ-PVF_2 (ϵ phase) is suggested; the γ and ϵ forms would then be mutually related in the same manner as the δ and α phases.

Acknowledgment. Helpful discussions with Dr. H. D. Keith are gratefully acknowledged, as is a careful review of the manuscript.

References and Notes

(1) Hasegawa, R.; Takahashi, Y.; Chatani, Y.; Tadokoro, H. *Polym. J.* **1972**, *3*, 600.
(2) Weinhold, S.; Litt, M. H.; Lando, J. B. *J. Polym. Sci., Polym. Lett. Ed.* **1979**, *17*, 585.
(3) Lovinger, A. J.; Keith, H. D. *Macromolecules* **1979**, *12*, 919.
(4) Weinhold, S.; Litt, M. H.; Lando, J. B. *Macromolecules* **1980**, *13*, 1178.
(5) Takahashi, Y.; Tadokoro, H. *Macromolecules* **1980**, *13*, 1317.
(6) Lovinger, A. J. *J. Polym. Sci., Polym. Phys. Ed.* **1980**, *18*, 793.
(7) The goniometer stage of our microscope does not accommodate tilts larger than 60°; larger angles were obtained by tilting crystals that had fortuitously not been flat at 0° as a result of localized warping of the specimen.
(8) Lovinger, A. J., submitted to *J. Appl. Phys.*
(9) Lovinger, A. J. *Polymer*, in press.
(10) Bachmann, M. A.; Gordon, W. L.; Weinhold, S.; Lando, J. B. *J. Appl. Phys.* **1980**, *51*, 5095.
(11) Lovinger, A. J. *Polymer* **1980**, *21*, 1317.
(12) Servet, B.; Rault, J. *J. Phys. (Paris)* **1979**, *40*, 1145.
(13) Lovinger, A. J. *Macromolecules* **1981**, *14*, 225.

Method for Estimating the Configurational Entropy of Macromolecules†

Martin Karplus*
Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138
Joseph N. Kushick‡
Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138, and Department of Chemistry, Amherst College, Amherst, Massachusetts 01002.
Received July 17, 1980

ABSTRACT: A method is outlined by which computer simulations of the structure and dynamics of large molecules can be used to estimate the classical configurational entropy difference between molecular conformations. The method can be employed with any simulation technique, such as harmonic analysis, molecular dynamics, or Monte Carlo. The formulation permits the use of arbitrary coordinates (e.g., normal modes, internal or Cartesian coordinates) and makes possible the determination of the contributions of particular degrees of freedom to the entropy change. Illustrative applications are made to conformational transitions of butane and decaglycine. It is found that significant contributions to the entropy change arise from fluctuations of internal coordinates other than the single-bond torsion angles. This indicates that additional internal coordinates (principally the bond angles) have to be retained in statistical mechanical models of the conformational thermodynamics of large molecules.

I. Introduction

The conformations of macromolecules, such as proteins and nucleic acids, play an essential role in their biological function.[1] To understand the factors involved in the stability of a given conformation and in the change of one conformation to another, it is necessary to be able to evaluate the energetic and entropic contributions to the free energy. Since the configurational entropy is expected to be very important in many transformations (e.g., protein denaturation, helix–coil transition), considerable effort has been expended in developing methods for its evaluation.[2-7] Although there have been a number of studies of the helix–coil transition in peptides, little progress has been made

†Work supported in part by grants from the National Science Foundation and the National Institutes of Health.
‡Camille and Henry Dreyfus Teacher-Scholar. Permanent address: Amherst College.

0024-9297/81/2214-0325$01.25/0

in determining the configurational entropy contribution to the stability of systems like globular proteins. The formulation of a method that can be used for examining this and related problems is the primary concern of the present paper.

In a macromolecular system, the configurational entropy contribution separates into two parts; the first is introduced by the local fluctuations in the neighborhood of a well-defined structure (e.g., an α helix, a native globular protein) and the second arises from the larger scale conformational variations that occur in less well-defined structures (e.g., the random-coil state). Somewhat different specialized techniques are best suited to the two extremes, though the evaluation of the classical configuration integral is the most direct approach to the entropy in both cases. Because of the large number of atoms present in many macromolecules of interest, emphasis in the published methods has been on the reduction of the coordinate set to be included in the integral evaluation. The most common choice is the complete set of dihedral angle variables,[3-5] although a further reduction by the use of "virtual" bonds has also been examined.[2,8] Given such a reduced set, the contribution due to local fluctuations has most often been estimated from a normal-mode analysis, while those due to the large-scale variations have been evaluated by integrations over the configuration space with the assumption of only near-neighbor correlations among the chosen variables.

For a globular protein in its native conformation, the available methods for treating local fluctuation contribution to the entropy are not applicable. Although a normal-mode analysis of the dihedral angle motions may be feasible in principle, it is extremely difficult to apply in practice because of the magnitude of the problem. Further, energy minimization calculations suggest that restriction to the space of dihedral angle variables would not yield accurate results for such a dense system.[9] An alternative approach to determining the local fluctuations occurring in macromolecules is by molecular dynamics simulations, which have been applied to small- and medium-sized proteins,[10-12] as well as to such regular polypeptide systems as α helices.[13] In these studies all heavy atom degrees of freedom were included, although molecular dynamics simulations with reduced coordinate sets have also been performed.[14] Such simulations provide the necessary information concerning the magnitudes of the local fluctuations, but they do not directly yield values of the configuration integral nor of the configurational entropy.

In the present paper we report an approximate method for computing molecular entropies from molecular dynamics, Monte Carlo, or other simulation techniques. The principal assumption made in the method, which is based on the properties of multivariate Gaussian distribution functions, is that the intramolecular potential energy can be expressed locally as a general quadratic function of the intramolecular coordinates, with coefficients which may depend on the temperature. The potential energy function used in the simulation can be that of an isolated molecule or it can be the potential of mean force which includes the effect of solvent on the magnitude of the fluctuations. The choice of internal coordinates is arbitrary. They can be the usual set of bond lengths, bond angles, and dihedral angles, but atomic Cartesian coordinates can also be used. An attractive feature of the method is that explicit contact is made between the entropy and the fluctuations and correlations of particular internal coordinates, so that the contribution of any group of internal coordinates to the change in configurational entropy with conformation can be evaluated directly. This makes it possible to examine the effect of neglecting certain molecular degrees of freedom in evaluating the configuration integral; e.g., the contribution of degrees of freedom other than the dihedral angles can be examined in a straightforward fashion.

Section II describes the method used for evaluating the configurational entropy. Illustrative applications are given in section III. They include the trans–gauche equilibrium in *n*-butane and the α-helix–extended-chain equilibrium of decaglycine. In both cases, comparisons are made with the results of standard vibrational entropy calculations. Particular attention is given to determining the contribution to the entropy of degrees of freedom other than the dihedral angles. Also, the significance of quantum corrections to configuration entropy differences is examined. Section IV presents the conclusions.

II. Method of Calculation of Configurational Entropy

The most direct approach to the classical configurational entropy difference between conformations of a macromolecule is by evaluation of the configuration integral for each of the conformations, C

$$Q_C = \int_C e^{-U(\mathbf{r})/k_B T}\, d\mathbf{r} \tag{1}$$

where k_B is the Boltzmann constant and T the absolute temperature. The effective potential energy of the system, $U(\mathbf{r})$, is a function of the Cartesian coordinates, $\mathbf{r}$; for the given conformation (C), the range of integration over the coordinates $\mathbf{r}$ is restricted appropriately. The potential energy function $U(\mathbf{r})$ can describe the intramolecular interactions in a vacuum or it can be a potential of mean force that takes account of the solvent contribution to the molecular energy; in the examples of section III, an empirical vacuum potential is used. The kinetic energy factor which multiplies Q_C in the complete partition function depends only on the atomic masses and temperature and so can be ignored since it is independent of conformation.

To evaluate Q_C it is usually convenient to transform from atomic Cartesian coordinates to a complete set of internal coordinates plus the six external coordinates. The latter specify the position of the center of mass and the overall orientation of the molecule; integration over these six coordinates yields a constant factor $8\pi^2 V$. The transformation to internal coordinates introduces a Jacobian, J, into the configuration integral; that is, if the new internal coordinates are $\mathbf{q}'$ with

$$d\mathbf{r} = J(\mathbf{q}')\, d\mathbf{q}' \tag{2}$$

we have[4,5,15]

$$Q_C = 8\pi^2 V \int_C e^{-U(\mathbf{q}')/k_B T} J(\mathbf{q}')\, d\mathbf{q}' \tag{3}$$

where the explicit dependence of the Jacobian J on the internal coordinates $\mathbf{q}'$ is indicated. As has been shown previously,[4,5,15] the Jacobian is a simple function of the bond lengths and bond angles and is independent of the dihedral angles. This form of the configuration integral greatly simplifies its evaluation if it is possible to assume that the bond lengths and bond angles are restricted to relatively narrow ranges by the potential function $U(\mathbf{q}')$ in the exponential of the integrand, so that the equilibrium (or average) values of these variables can be assigned to the Jacobian. If we further assume that these selected values of the bond lengths and bond angles do not differ significantly in the various conformations of the molecule, the Jacobian may be regarded as independent of molecular conformation.

We now divide the full set of molecular internal coordinates (which specify the relative positions of all the

atoms) into a group of "important" coordinates (**q**) and a group of "unimportant" coordinates (**q**″). A particular internal coordinate is designated as "important" if it contributes significantly to the thermodynamics of conformational change; otherwise it is regarded as "unimportant". Such a division of the set of internal coordinates was proposed by Gō and Scheraga,[4] who designated some internal coordinates as soft variables (e.g., dihedral angles for rotation about single bonds) and others as hard variables (e.g., bond lengths). With the separation into two types of coordinates the integration over the unimportant internal coordinates (**q**″) can be performed. There results an approximate configuration integral over the important coordinates **q** of the form

$$Q_C \simeq \text{const} \int_C e^{-V(\mathbf{q})/k_B T}\, d\mathbf{q} \tag{4}$$

where the "const" includes the configuration-independent terms and the potential $V(\mathbf{q})$ is the value of the potential energy as a function of the **q**, given that the unimportant coordinates are in the neighborhood of their equilibrium values. The approximate configurational partition function, Q_C, is the one proposed and justified for use in polymer problems by Flory[2] and by Gō and Scheraga.[4] The present development differs from the work of these authors in that no explicit restriction of the important coordinates to dihedral angles has been made; in fact, we shall demonstrate the importance of other internal coordinates to the configurational entropy in section III.

At no point in the derivation of eq 4 have rigid constraints been imposed on the internal coordinates. Thus, eq 4 does not contain the metric tensor contributions associated with such constraints.[16,17]

For convenience in the subsequent development, we introduce the reduced configuration integral, $Q_q^C = Q_C(\text{const})^{-1}$, and define the configurational entropy, S_q^C, for conformation C as

$$S_q^C = \langle V \rangle / T + k_B \ln Q_q^C \tag{5}$$

Since Q_q^C is not a dimensionless quantity, the absolute value of S_q^C has no physical meaning; if the kinetic energy term were calculated and the constant factor were evaluated, the absolute entropy could be obtained. However, in what follows S_q^C will be evaluated for two conformations of a particular molecule for which the difference ΔS_q *is* a meaningful quantity. If the set of important internal coordinates, **q**, is properly selected, ΔS_q will contain all of the significant configurational contributions to the entropy difference ΔS between the two conformations. Furthermore, because the kinetic energy contribution to the partition function is treated exactly so that S_q^C is directly related to the configurational partition function, ΔS_q implicitly includes the contributions to ΔS arising from vibrational motion, rotational motion, and vibration–rotation coupling. To illustrate this last point, the relationship between ΔS_q and the vibration–rotation contributions to ΔS is outlined for the simple case of a diatomic molecule in the Appendix.

To use eq 4 and 5 in the calculation of S_q^C from the results of simulations of the fluctuations of macromolecules, we introduce the joint probability distribution function, $P(\mathbf{q})$, for the set of important coordinates, **q**

$$P(\mathbf{q}) = \frac{1}{Q_q^C} e^{-V(\mathbf{q})/k_B T} \tag{6}$$

The configurational entropy, S_q^C (eq 5), is simply related to $P(\mathbf{q})$ by

$$S_q^C = -k_B \int_C P(\mathbf{q}) \ln P(\mathbf{q})\, d\mathbf{q} \tag{7}$$

In principle, the simulation provides $P(\mathbf{q})$ so that S_q^C can be evaluated by eq 7. For systems such as globular proteins or well-structured polypeptides, in which the fluctuations important for equilibrium properties are expected to be relatively small, machine simulations can provide accurate values for $P(\mathbf{q})$. Given $P(\mathbf{q})$ from such results, the direct evaluation of the integral in eq 7 is still a formidable problem. Consequently, it is helpful to introduce some assumption concerning the form of $P(\mathbf{q})$. It is expected that the fluctuations of many of the internal coordinates are strongly correlated in macromolecules like globular proteins. A satisfactory representation of $P(\mathbf{q})$ should, therefore, not contain any assumption about the extent to which the internal coordinates are coupled; i.e., the distribution function should provide a clear link between the entropy and both the fluctuations of the internal coordinates and the correlations among them. As a first approximation, which results in an enormous simplification of eq 7, we assume that $V(\mathbf{q})/k_B T$ can be represented as a general quadratic function of the internal coordinates **q**. An arbitrary degree of coupling between the internal coordinates is included in the quadratic assumption, although triplet and higher order correlations are excluded. At low temperatures, this assumption reduces to the harmonic approximation for $V(\mathbf{q})$. More generally, the coefficients of the quadratic terms can be chosen in accordance with the calculated fluctuations of the internal coordinates for a given temperature. It is likely that the coefficients in the quadratic representation of $V(\mathbf{q})$ have a nontrivial temperature dependence. Thus, although the effective potential for coordinate fluctuations may be significantly different from that evaluated by a harmonic expansion of $V(\mathbf{q})$ about the equilibrium configuration, assumption of a temperature-dependent quadratic form may be adequate.

The above considerations suggest that a reasonable model for $P(\mathbf{q})$ is a normalized multivariate Gaussian distribution function; that is

$$P(\mathbf{q}) = \frac{1}{(2\pi)^{n/2}\sigma^{1/2}} \exp[-\tfrac{1}{2}(\mathbf{q} - \langle \mathbf{q} \rangle)\cdot\sigma^{-1}\cdot(\mathbf{q} - \langle \mathbf{q} \rangle)] \tag{8}$$

Here, n is the number of degrees of freedom in the set **q**, $\langle \mathbf{q} \rangle$ is the vector of average values of the internal coordinates, and the matrix σ contains the temperature-dependent variances (diagonal elements) and covariances (off-diagonal elements) of the internal coordinates; i.e.

$$\sigma_{ij} = \langle (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \rangle \tag{9}$$

The determinant of σ is denoted by σ. Introducing eq 8 into eq 7, we have[18]

$$S_q^C = \tfrac{1}{2} n k_B + \tfrac{1}{2} k_B \ln [(2\pi)^n \sigma] \tag{10}$$

Thus, all that is required for a calculation of S_q^C is the determinant of the covariance matrix σ. The elements of σ, the mean-square fluctuations and correlation coefficients of the various internal coordinates, are readily available in any numerical simulation. Configurational entropies can therefore be calculated directly from molecular dynamics or Monte Carlo simulations without any need to resort to a normal-coordinate analysis. The behavior of the internal coordinates in a computer simulation (e.g., the higher moments of the internal-coordinate distributions) can be used to determine the degree to which the approximation, eq 8, is valid, so that a self-consistency check is built into the procedure. Further, the approximations made in the derivation of eq 4 (e.g., that the equilibrium values of the bond lengths and bond angles do not differ significantly between molecular conformations) can be

verified from the result of the simulation. Finally, if σ is calculated for two conformations (labeled a and b) of a macromolecule, the entropy difference associated with the conformational change in the present approximation is given by the simple expression

$$\Delta S \simeq \Delta S_q = \frac{k_B}{2} \ln \frac{\sigma(b)}{\sigma(a)} \tag{11}$$

This result is to be compared with the usual classical expression based on separating the rotational and vibrational motion in the harmonic oscillator, rigid-rotator approximation. We have[19]

$$\Delta S = \Delta S_{rot} + \Delta S_{vib} \tag{12}$$

with

$$\Delta S_{rot} = \frac{k_B}{2} \ln \frac{I_A(b)I_B(b)I_C(b)}{I_A(a)I_B(a)I_C(a)} \tag{13}$$

and

$$\Delta S_{vib} = k_B \ln \left[\prod_{i=1}^{3N-6} \omega_i(a) / \prod_{i=1}^{3N-6} \omega_i(b) \right] \tag{14}$$

where I_A, I_B, I_C are the principal moments of inertia, N is the number of atoms, and the ω_i are the normal-mode vibrational frequencies. As is well-known and evident from eq 12, the atom masses cancel in the evaluation of ΔS. Quantitative comparisons of ΔS obtained for model systems with a harmonic potential by use of eq 11 and 12–14 are given in the following section. In accord with the discussion following eq 4 and 5, the only differences between the two for such cases arise from the breakdown of the rigid-rotator approximation and presence of vibration–rotation coupling (included in eq 11 and neglected in eq 12; see Appendix) and the variation with conformation of the Jacobian involved in the transformation from Cartesian to internal coordinates (included in eq 12 but neglected in eq 11).

III. Illustrative Calculations

In this section we describe model calculations which demonstrate the utility of the formalism developed for configurational entropy calculations. The simulation technique employed is the method of harmonic analysis.[20] This technique makes use of a quadratic potential function for $U(\mathbf{r})$ in eq 1 so that the validity of eq 8 is guaranteed; this would not be the case in, say, a molecular dynamics simulation with a more generalized potential function. Further, since for harmonic systems the method for evaluating ΔS by eq 12 is expected to be a very good approximation, we have a standard for comparison to assess directly the effect of retaining various subsets of the internal coordinates in the configurational entropy evaluation. This simplifies the determination of a sufficient set of internal coordinates for a statistical mechanical treatment of conformational equilibria.

Calculations were carried out for two test molecules, butane and the finite polypeptide chain decaglycine. Butane was selected as a very simple molecule which has more than one stable conformation; decaglycine is a more complex molecule which can form an α-helical structure stabilized by internal hydrogen bonds and which was the subject of a recent dynamical study using harmonic analysis[20] and is now being examined by a full molecular dynamics treatment.[13] The decaglycine calculation represents a first step in employing the formalism described here to treat the entropy differences associated with biologically significant conformational changes of proteins and protein structural elements.

Table I
Parameters for the Butane Intramolecular Potential (Eq 15)

K_b = 250 kcal/(mol·Å²); b_0 = 1.54 Å
K_θ = 120 kcal/(mol·rad²); θ_0 = 112°
K_ϕ = 2.0 kcal/mol
σ = 3.50 Å;[a] ϵ = 0.0897 kcal/mol[a]

[a] The parameters σ and ϵ are the usual Lennard-Jones potential parameters and are related to A and C of eq 15 by $A = 4\epsilon\sigma^{12}$ and $C = 4\epsilon\sigma^6$.

For both butane and decaglycine, the harmonic approximation to the potential function was obtained from empirical potentials that consist of a sum of bond stretching, bond angle bending, and torsional terms augmented by nonbonded interactions including van der Waals, electrostatic, and hydrogen bonding contributions. The form of the potential used is[21]

$$U = \tfrac{1}{2} \sum_{\text{bonds}} K_b(b - b_0)^2 + \tfrac{1}{2} \sum_{\text{bond angles}} K_\theta(\theta - \theta_0)^2 + \tfrac{1}{2} \times \sum_{\text{torsion angles}} K_\phi[1 + \cos(n\phi - \delta)] + \tfrac{1}{2} \sum_{\text{improper torsions}} K_\xi(\xi - \xi_0)^2 + \tfrac{1}{2} \sum_{nb \text{ pairs}} \left(\frac{A}{r^{12}} - \frac{C}{r^6} + \frac{q_i q_j}{r} \right) + \sum_{\text{H bonds}} \left(\frac{A'}{r^{12}} - \frac{C'}{r^{10}} \right) \tag{15}$$

where b is the value of a bond length, θ is the value of a bond angle, ϕ is the value of a torsional angle, ξ is the value of an improper torsional angle, and r is an interparticle distance; q_i are atomic charges; K_b, K_θ, K_ϕ, and K_ξ are force constants; b_0, θ_0, ξ_0, n, δ, A, C, A', and C' are parameters specifying the various interactions. The methyl and methylene groups were regarded as extended atoms[21] with appropriate masses, so that the only hydrogen atoms explicitly included were those attached to the nitrogen–hydrogen bond donor atoms in decaglycine. For decaglycine the parameters used in eq 15 were those of ref 20. For butane, no electrostatic or hydrogen bonding terms were included in eq 15 and the potential parameters were adjusted to give reasonable agreement with experimental and/or ab initio results for the following quantities: the equilibrium geometry in both the trans (t) and gauche (g) conformations, the energy difference between the t and g conformations, the cis barrier height, the geometry at the top of the cis barrier,[23,24] and the frequencies of all of the skeletal normal modes of *trans*-butane.[25,26] The butane parameters for eq 15 are shown in Table I, and the resulting molecular properties are compared with the experimental and ab initio values in Table II. As can be seen, the empirical potential function gives satisfactory values for most of the data, although improvements could certainly have been achieved by more extensive refinement. Of particular note in Table II is the theoretical variation of the CCC bond angle with the torsional angle and its reproduction by the empirical potential. It will be shown in what follows that the correlation between bond angle bending and the torsional angle has an important effect on the configurational entropy.

For both the t and g conformations of butane, initial atomic coordinates were selected near the expected equilibrium geometry and energy minimization was performed on the multidimensional potential surface until the local minimum energy geometry was achieved. A normal-coordinate analysis was performed to obtain the normal modes and corresponding frequencies.[20,22] These results were then used to calculate the entropy difference for the butane conformational change (gauche → trans) both by the standard method, eq 12, and by the new configurational entropy technique (eq 11). The decaglycine

Table II
Butane Potential Results[a,b]

	eq 15	exptl	theor
trans-gauche energy difference	−0.744	−0.726* (24) −0.89 (28)	−0.822* (24)
cis barrier	4.76	5.3 (24)	4.98* (24)
gauche dihedral angle	71.3	65.4* (24) 62 (28)	66.5* (24)
trans CCC bond angle	112.0		112.2 (23)
gauche CCC bond angle	112.7		113.1 (23)
cis CCC bond angle	115.8		115.3 (23)
trans skeletal normal frequencies	1045	1087 (25), 1058 (26)	
	1003	970 (25, 26)	
	902	960 (25), 835 (26)	
	437	408 (25), 432 (26)	
	406	380 (25), 365 (26)	
	120	140 (25)	
gauche skeletal normal frequencies	1037		
	966		
	864		
	600		
	420		
	148		

[a] References to sources are given in parentheses; asterisks represent averages over a range of reported results. [b] All energies in kcal/mol, all angles in degrees, and all frequencies in cm^{-1}.

calculations were carried out in a manner parallel to the butane calculations. The conformations studied were the α-helix structure[20] and the fully extended structure in which all the backbone dihedral angles are near 180°, which is a local minimum. When the molecule reached the minimum energy geometry in the extended (ext) conformation, its energy was 58.39 kcal/mol above the energy of the α-helix (α) conformation. The major factors in the energy difference were the loss of hydrogen bond stabilization and less favorable nonbonded interactions in ext-decaglycine. Since in solution the loss of internal hydrogen bonding would be compensated at least in part by hydrogen bonds to solvent, the energy results are valid only in a vacuum.

To apply eq 11 to these harmonic systems, the covariance matrix for each conformation was obtained from the harmonic analysis results.[20] The displacement of any internal coordinate, q_i, from its mean value, $\langle q_i \rangle$, can be expressed as a linear combination of the normal coordinates (Q_k) by the equation

$$q_i - \langle q_i \rangle = \sum_{k=1}^{3N-6} \alpha_{ik} Q_k \tag{16}$$

The equal-time correlation function for any two normal coordinates has the form

$$\langle Q_k Q_l \rangle = \delta_{kl} \frac{k_B T}{\omega_k^2} \tag{17}$$

Given eq 16 and 17, the elements of the covariance matrix (eq 9) can be calculated from

$$\sigma_{ij} = \langle (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \rangle = k_B T \sum_{k=1}^{3N-6} \frac{\alpha_{ik}\alpha_{jk}}{\omega_k^2} \tag{18}$$

This procedure can be carried out for any subset of internal coordinates. Since ΔS is already known within the rigid-rotator, harmonic oscillator approximation from eq 12, we can directly determine the internal coordinates that are important for the configurational entropy.

Quantum Corrections. Prior to examining the classical entropy results, it is of interest to determine the quantum corrections to the classical limit. Since the vibrational contribution to the entropy is most likely to deviate from the classical result, we restrict the comparison to S_{vib} and ΔS_{vib}. The quantum calculations were done by introducing the normal-mode frequencies into the standard entropy formula.[19] The quantum mechanical and classical values of S_{vib} and ΔS_{vib} for butane and decaglycine are listed in Table III. The classical calculations of *absolute* vibrational entropies give poor results in all cases. For butane, all of whose normal frequencies lie above 120 cm^{-1} (Table II), there is also a significant (~17%) difference between the classical and quantum calculations of ΔS_{vib}. As one proceeds to larger molecules, the principal contributions to ΔS_{vib} come from changes in the low-frequency modes (some of which are shown for decaglycine in Table III), so that a classical treatment is considerably better. In the case of decaglycine, which is still not a large molecule from a biochemical viewpoint, the classical approximation to ΔS_{vib} deviates from the quantum result by only 6.5%.

Table III
Classical and Quantum Vibrational Entropies[a]

	quantum	classical
	Butane	
$S_{vib}(g)$	4.428	0.058
$S_{vib}(t)$	5.223	0.989
ΔS_{vib}	0.795	0.931
	Decaglycine	
$S_{vib}(\alpha)$	211.97	70.17
$S_{vib}(ext)$	286.31	149.25
ΔS_{vib}	74.34	79.08

Ten Lowest Decaglycine Frequencies (cm^{-1})

α helix		extended	
48.53	28.09	16.05	6.39
46.89	26.40	14.70	5.94
43.96	24.39	12.47	4.60
40.09	21.65	11.39	3.32
34.23	12.91	10.26	1.24

[a] All entropy values at 300 K in units of cal/(mol·deg).

Butane. The configurational entropy differences between *trans*- and *gauche*-butane obtained from eq 11 and 12 are shown in Table IV.

As can be seen ΔS_{vib} and ΔS_{rot} have similar magnitudes and opposite signs. In general, for larger molecules ΔS_{rot} is expected to be much less important than ΔS_{vib}. For ΔS_q,

Table IV
Entropy Differences (Trans Minus Gauche) for Butane[a,b]

ΔS_q (eq 11)	ΔS (eq 12)	contribution
ΔS_ϕ		0.222
ΔS_θ		0.065
ΔS_b		0.019
$\Delta S_{\phi,\theta}$		0.342
$\Delta S_q = \Delta S_{\phi,\theta,b}$		0.377
	ΔS_{vib}	0.931
	ΔS_{rot}	−0.560
	ΔS	0.371

[a] All values in units of cal/(mol·deg). [b] The internal-coordinate symbols correspond to those used in eq 15; also, see text.

we have made use of its direct expression in term of internal coordinates to separate out the contributions from the various types, that is, the set of bond lengths (b), bond angles (θ), and the torsional angle (ϕ), as well as combinations of the different types. As expected, the torsional degree of freedom (ϕ) dominates all other internal coordinates in the conformational entropy difference. However, a calculation of ΔS_q which regards ϕ as the only important coordinate gives a result ΔS_ϕ which is only 60% of the overall ΔS. Most of the difference between ΔS and ΔS_ϕ is attributable to the two bond angles; if the torsional and bond angle bending coordinates are together regarded as the important coordinates, the resulting configurational entropy $\Delta S_{\phi,\theta}$ is more than 90% of the overall ΔS. The bond lengths make a much smaller contribution than the bond angles, though they are not completely negligible in this case. When all six internal coordinates are included in ΔS_q, the resulting value, $\Delta S_{\phi,\theta,b}$, is slightly different from $\Delta S_{vib} + \Delta S_{rot}$. This small residual difference must be attributed in this harmonic system to the approximations introduced in the derivation of eq 4 and to the rigid-rotator approximation used in evaluating ΔS_{rot}.

The influence of the bond angles on the configurational entropy change with conformation derives from two effects. One is attributed to fluctuations of the bond angles and the other to coupling between the fluctuations of the bond angles and the torsional angle. A calculation of ΔS_q which retains *only* the two bond angles as the important coordinates gives a result ΔS_θ which measures the importance of the first effect; the difference between $(\Delta S_\phi + \Delta S_\theta)$ and $\Delta S_{\phi,\theta}$ measures the importance of the coupling. The data of Table IV show that these two effects are of approximately equal importance and that they together account for more than 30% of the total ΔS. Thus, if the potential functions used are accurate, the bond angle fluctuations play an important role in the conformational equilibrium of butane and should be retained along with the torsional angle in a statistical mechanical treatment.

Additional insight into the results of Table IV can be obtained by examination of the covariance matrices for *gauche*- and *trans*-butane with the torsional and the bond angles as the set **q** (all values are in deg^2)

$$\sigma(g) = \begin{pmatrix} 173.90 & -6.06 & -6.06 \\ -6.06 & 7.89 & -0.19 \\ -6.06 & -0.19 & 7.89 \end{pmatrix} \quad (19)$$

$$\sigma(t) = \begin{pmatrix} 217.37 & 0.00 & 0.00 \\ 0.00 & 8.15 & 0.00 \\ 0.00 & 0.00 & 8.15 \end{pmatrix} \quad (20)$$

In eq 19 and 20, σ_{11} is the mean-square fluctuation of the torsional angle and σ_{22} and σ_{33} are the mean-square fluctuations of the two equivalent bond angles. Equation 19 shows a correlation coefficient $(\sigma_{12}/(\sigma_{11}\sigma_{12})^{1/2})$ of −0.16 between ϕ and either bond angle in *gauche*-butane; there is no correlation in *trans*-butane. The physical origin of the difference is that as the torsional angle in *gauche*-butane becomes smaller, the bond angles tend to open so as to reduce the van der Waals repulsion between the end methyl groups. The entropy of *gauche*-butane is thereby reduced relative to what the entropy would be if the torsion and bond angles were independent. In *trans*-butane, where the internal coordinates are uncorrelated since the nonbonded interaction is very small, there is only the independent bond angle contribution to the overall ΔS.

Table V
Entropy Differences
(Extended Minus α Helix) for Decaglycine[a]

ΔS_q (eq 11)	ΔS (eq 12)	contribution
$\Delta S_{\phi,\psi}$		39.40
$\Delta S_{\omega,\xi}$		6.10
ΔS_θ		0.07
ΔS_b		0.03
$\Delta S_{\phi,\psi,\omega,\xi}$		54.88
$\Delta S_{\phi,\psi,\theta}$		43.22
$\Delta S_{\omega,\xi,\theta}$		7.92
$\Delta S_{\phi,\psi,b}$		39.59
$\Delta S_q = \Delta S_{\phi,\psi,\omega,\xi,\theta}$		75.88
	ΔS_{vib}	79.08
	ΔS_{rot}	2.02
	ΔS	81.10

[a] All values in cal/(mol·deg).

In a recent study of the properties of model alkanes in which only the torsional angle was included, the equilibrium distribution of butane molecules in the trans and gauche conformations was computed by a direct integration of the one-dimensional configuration integral.[27] At 298 K in vacuo, the ratio of trans molecules to either form (+ or −) of gauche molecules was found to be 3.71, which implies that the free-energy difference is −777 cal/mol. The energy difference was taken to be −700 cal/mol, so that the entropy difference is $\Delta S = (1/T)(\Delta U - \Delta A) =$ 0.259 cal/(mol·deg). This value is in approximate agreement with the present result for the purely torsional contribution to the configurational entropy difference (ΔS_ϕ) and is considerably below the calculated value of the total entropy difference.

A recent electron diffraction investigation of gas-phase *gauche*- and *trans*-butanes reported a free-energy difference (trans minus gauche) of −497 ± 220 cal/mol.[29] This result, if combined with the range of published values of the energy difference listed in Table II (−700 to −900 cal/mol), implies that the entropy of the gauche conformation is higher than that of the trans conformation, in apparent disagreement with the present calculations. However, the uncertainty in the experimental free-energy value is such that it is difficult to draw definite conclusions.

Decaglycine. The results of the decaglycine harmonic analysis were used to calculate $\Delta S = \Delta S_{vib} + \Delta S_{rot}$ and ΔS_q for the transition of the molecule from α-helical to the extended conformation. In determining ΔS_q several choices of the set of "important" internal coordinates were used. The values obtained are shown in Table V. The internal coordinates were grouped into four categories; they are the soft dihedral angles for torsion about N–C_α and C–C_α bonds (ϕ,ψ), the dihedral angles for torsion about peptide bonds (ω), and the "improper" torsion angles representing out-of-plane motions of amide hydrogen and carbonyl oxygen atoms (ξ), bond angles (θ), and bond lengths (b). In such a listing, specification of all the tor-

sional and bond angles of decaglycine which appear in the potential function (eq 15) is redundant in that the values of some bond angles are uniquely determined by other bond angles and dihedral angles. Since the covariance matrix is singular if the set **q** is overcomplete, the redundant bond angles θ were deleted before the evaluation of the configurational entropies. Comparisons verified that the calculated results are independent of which redundant coordinates are removed. The total number of decaglycine angular coordinates included is 107, the number of internal degrees of freedom (162) minus the number of bonds (55).

An often-used choice for the set of important internal coordinates for polypeptides and proteins is the set of soft dihedral angles (ϕ,ψ). Table V shows unequivocally that a configurational entropy calculation based only on these angles gives a value ($\Delta S_{\phi,\psi}$) which is less than half the value of ΔS. Only when the set of important coordinates is expanded to include ($\phi,\psi,\omega,\xi,\theta$) does the configurational entropy difference become satisfactory ($\Delta S_q = 0.94\Delta S$). It appears therefore that the only internal coordinates which may be safely neglected are the bond lengths and that all other degrees of freedom contribute significantly to the entropy. From Table V, it is clear that it is not the contribution of the bond angles themselves which is important (i.e., ΔS_θ contributes only 0.07 cal/(mol·deg) to ΔS_q) but it is the correlation between (ϕ,ψ) and (θ) that is important; the latter, corresponding to $\Delta S_{\phi,\psi,\theta}$, contributes 3.75 cal/(mol·deg) relative to ($\Delta S_{\phi,\psi} + \Delta S_\theta$). Even more striking is the general coupling between the set (ϕ,ψ,ω,ξ) and (θ) made evident by a comparison of ($\Delta S_{\phi,\psi,\omega,\xi} + \Delta S_\theta$) vs. $\Delta S_{\phi,\psi,\omega,\xi,\theta}$. For extended decaglycine, the contributions of the various groups of internal coordinates to S_q^C are additive, which indicates that there is no correlation among the coordinates. By contrast, α-helical decaglycine, which is stabilized by hydrogen bonds and nonbonded forces, exhibits strong correlation effects through nonadditive contributions of the internal coordinates to S_q^C.

IV. Conclusions

The method proposed for configurational entropy calculations from any simulation method for macromolecules has been shown to yield accurate results for test systems. Its generality in terms of choice of coordinates suggests that it is ideal for large molecules where localized internal motions play an essential role in the thermodynamic properties. Of importance in the results for butane and decaglycine is the demonstration that internal coordinates other than the dihedral angles (i.e., bond angles) contribute significantly for folded systems where the internal motions are strongly correlated. Our current effort is aimed at configurational entropy calculations based on molecular dynamics simulations of globular proteins and related systems.

Acknowledgment. Ronald M. Levy provided essential assistance in the study of decaglycine and kindly provided the results of his calculations of the harmonic dynamics. We have benefited from numerous discussions with R. M. Levy and also with Steve Arick, Bernard Brooks, Mark Fromowitz, and Toshiko Ichiye. J.N.K. acknowledges the support provided by an Amherst College Trustee Faculty Fellowship.

Appendix

In this section we demonstrate that the difference of the configurational entropy, ΔS_q, between two conformations of a diatomic molecule implicitly includes contributions from both vibrational and rotational motion. For the development we suppose that the two molecular conformations are labeled a and b, that the bond-stretching potentials for both conformations are quadratic with force constants $K_s(a)$ and $K_s(b)$, and that the equilibrium internuclear separations are $r_0(a)$ and $r_0(b)$, respectively. In the harmonic oscillator, rigid-rotor (HORR) approximation the individual contributions to the overall entropy difference, $\Delta S = S(b) - S(a)$, between the conformations are, from eq 13 and 14

$$\Delta S_{vib} = k_B \ln \frac{K_s^{1/2}(a)}{K_s^{1/2}(b)} \tag{A1}$$

and

$$\Delta S_{rot} = k_B \ln \frac{r_0^2(b)}{r_0^2(a)} \tag{A2}$$

so that

$$\Delta S = \Delta S_{vib} + \Delta S_{rot} = k_B \ln \frac{K_s^{1/2}(a) r_0^2(b)}{K_s^{1/2}(b) r_0^2(a)} \tag{A3}$$

In the configuration integral approach, we start with the classical molecular partition function in Cartesian coordinates; that is

$$Q = \int dx_1 \ldots dz_2\, dp_{x1} \ldots dp_{z2} \exp\left[-\left(\frac{p_1^2}{2m_1} + \frac{p_2^2}{2m_2} + \frac{1}{2}K_s(r - r_0)^2 \right) / k_B T \right] \tag{A4}$$

where $\mathbf{p}_i$ is the momentum of particle i, $r^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2$ and Q is evaluated for the two molecular conformations with appropriate values of K_s and r_0. We then proceed as in the derivation of eq 3; i.e., the integration over momenta yields the conformation-independent factor $(2\pi m_1 k_B T)^{3/2}(2\pi m_2 k_B T)^{3/2}$, the six Cartesian coordinates are transformed to five external coordinates and the internal coordinate r, the integration over the external coordinates yields another conformation-independent factor, and we are left with

$$Q = C' \int_0^\infty dr\, r^2 e^{-K_s(r-r_0)^2/2k_B T} \tag{A5}$$

This equation is exact; the Jacobian factor of eq 3 here has the simple form $J(r) = r^2$.

The integral in eq A5 is easily expanded by means of the variable change $s = \gamma(r - r_0)$, where $\gamma = (K_s/2k_B T)^{1/2}$; we have

$$Q = \frac{C'}{\gamma^3}\left[(\gamma r_0)^2 \int_{-\gamma r_0}^\infty ds\, e^{-s^2} + 2\gamma r_0 \int_{-\gamma r_0}^\infty ds\, s e^{-s^2} + \int_{-\gamma r_0}^\infty ds\, s^2 e^{-s^2} \right] \tag{A6}$$

On physical grounds, one expects that the HORR approximation should hold when the force constant is large. Since the configurational expression for the partition function includes all contributions, the HORR results should be recovered from eq A6 in the limit as $\gamma r_0 \to \infty$. The limiting behavior of Q is easily found to be

$$\lim_{\gamma r_0 \to \infty} Q = \frac{C' r_0^2}{\gamma} \int_{-\infty}^\infty ds\, e^{-s^2} = C'\left(\frac{2\pi k_B T}{K_s} \right)^{1/2} r_0^2 \tag{A7}$$

and eq A3 follows immediately. To the degree that the exact value of Q in eq A6 differs from the limiting value in eq A7, the configurational integral includes corrections to the HORR approximation; in the present case, they

correspond to including the fact that the moment of inertia depends on the internal coordinate r.

References and Notes

(1) Privalov, P. L. *Adv. Protein Chem.* **1979**, *33*, 167.
(2) Flory, P. J. "Statistical Mechanics of Chain Molecules"; Wiley: New York, 1969.
(3) Gō, N.; Scheraga, H. A. *J. Chem. Phys.* **1969**, *51*, 4751.
(4) Gō, N.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 535.
(5) DeVoe, H. In "Structure and Stability of Macromolecules"; Timosheff, S. N., Fasman, G. D., Eds.; Marcel Dekker: New York, 1969; Chapter 1.
(6) Hagler, A. T.; Stern, P. S.; Sharon, R.; Becker, J. M.; Naider, F. *J. Am. Chem. Soc.* **1979**, *101*, 6842.
(7) Meirovitch, H. *Chem. Phys. Lett.* **1977**, *45*, 389.
(8) Levitt, M.; Warshell, A. *Nature (London)* **1975**, *253*, 694.
(9) van Gunsteren, W.; Karplus, M., submitted to *Macromolecules*.
(10) Karplus, M.; McCammon, J. A. *CRC Crit. Rev. Biochem.*, in press.
(11) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature (London)* **1977**, *267*, 585.
(12) Northrup, S. H.; Pear, M. R.; McCammon, J. A.; Karplus, M.; Takano, T. *Nature (London)*, in press.
(13) Perahia, D.; Levy, R. M.; Karplus, M., in preparation.
(14) McCammon, J. A.; Northrup, S. H.; Karplus, M.; Levy, R. M. *Biopolymers*, in press.
(15) Herschbach, D. R.; Johnson, M. S.; Rapp, D. *J. Chem. Phys.* **1959**, *31*, 1652.
(16) Fixman, M. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 3050.
(17) Chandler, D.; Berne, B. *J. Chem. Phys.* **1979**, *71*, 5386.
(18) Hogg, R. V.; Craig, A. T. "Introduction to Mathematical Statistics", 3rd ed.; Macmillan: New York, 1970.
(19) McQuarrie, D. A. "Statistical Mechanics"; Harper and Row: New York, 1976.
(20) Levy, R. M.; Karplus, M. *Biopolymers* **1979**, *18*, 2465.
(21) Gelin, B. R.; Karplus, M. *Biochemistry* **1979**, *18*, 1256.
(22) Gelin, B. R.; Karplus, M. *J. Am. Chem. Soc.* **1975**, *97*, 699.
(23) Radom, L.; Pople, J. A. *J. Am. Chem. Soc.* **1970**, *92*, 4786.
(24) Peterson, M. R.; Csizmadia, I. G. *J. Am. Chem. Soc.* **1978**, *100*, 6911.
(25) Pitzer, K. S. *J. Chem. Phys.* **1940**, *8*, 711.
(26) Szasz, G. J.; Sheppard, N.; Rank, D. H. *J. Chem. Phys.* **1948**, *16*, 704.
(27) Levy, R. M.; Karplus, M.; McCammon, J. A. *Chem. Phys. Lett.* **1979**, *65*, 4.
(28) Durig, J. R.; Compton, D. A. C. *J. Phys. Chem.* **1979**, *83*, 265.
(29) Bradford, W. F.; Fitzwater, S.; Bartell, L. S. *J. Mol. Struct.* **1977**, *38*, 185.

Loop Entropy of the Triple Helix

F. Y. Chen and H. W. Huang*

Physics Department, Rice University, Houston, Texas 77001. *Received August 11, 1980*

ABSTRACT: We calculate the loop entropy in an infinitely long triple-stranded chain and estimate its contribution to the helix–coil transition of collagen by comparing theory with experimental data. We find that the temperature dependence of helix fraction together with that of specific heat provides a rather stringent test for a simple model. By comparing our model (with loops) and the one-sequence model (without loops) with experiment, we find that the formation of loops is responsible for a substantial fraction of heat absorption of collagen during its helix–coil transition.

The calculation of loop entropy in triple-stranded helices is itself an interesting statistics problem. It also has practical applications to helix–coil transitions of collagen (and its analogues). In this paper we will try to calculate the loop entropy in an infinitely long triple-stranded chain and estimate its contribution to the helix–coil transition of collagen by comparing theory with experimental data.

The statistics of loop formation in double-stranded helices has been much studied,[1-4] but the triple-stranded problem is much more complicated. In order to obtain a manageable formula for our final result, we ignore the effect of excluded volume,[3] assume a Gaussian distribution for the end-to-end extension of a loop,[1] and use a semi-numerical method to carry out a double summation. Experimental data include the heat absorption and optical activity of tropocollagen measured by Privalov and Tiktopulo.[5] We will find that the temperature dependence of helix fraction together with that of specific heat provides a rather stringent test for a simple model. By comparing our model with the one-sequence model, we can see that the formation of loops is responsible for a substantial fraction of heat absorption of collagen during its helix–coil transition.

Theory

Consider, as a statistical mechanics problem, a homogeneous triple-stranded chain. Each strand is a sequence of basic units. A triple-helix unit consists of three basic units, one from each component strand. When a triple-helix unit becomes unbound, it is called a coil state. At a given temperature, the configuration of a triple-stranded chain is an alternate sequence of helical and loop states. A loop sequence consists of three unbound component strands of various length (number of basic units) attached to two helical sequences at the beginning and at the end.[6] In the limit of infinitely long chain, dissociation is neglected. It is then convenient to use the method of sequence-generating functions[7] to calculate the partition function. We write the sequence-generating function for helical states as[7]

$$V(x) = \sum_{l=1}^{\infty}(S/x^3)^l = S/(x^3 - S) \tag{1}$$

where S is the partition function for each triple-helix unit and can be written in the form

$$S = \exp\{\Delta H_u(T - T_c)/RTT_c\} \tag{2}$$

Here ΔH_u and $\Delta H_u/T_c$ are, respectively, the changes of enthalpy and entropy from a *free* coil state (see below) to a helical state. x is a parameter which will be used later to calculate the partition function. The sequence-generating function for loop states is

$$U(x) = \beta\sum_{l=1}^{\infty}u_l/x^l \tag{3}$$

where u_l is the partition function of a loop with a total strand length l (the total number of basic units in the three strands) and β is the border factor between a loop state and a helical state. If one writes u_l as

0024-9297/81/2214-0332$01.25/0