efficiency of the two methods for a system with stronger EV interactions, we have simulated a single-chain model with EV, which mimics, to some extent, a dense, many-chain system. Prior to the simulation, part of the lattice has been filled randomly with "obstacles"; i.e., in addition to the EV interaction, a bead is not allowed to lie on a lattice site which is occupied by an obstacle. The results showed that the efficiency of our method, as compared to the conventional one, increased with increasing the fraction of the obstacles on the lattice. For example, for $N = 33$ and a fraction of 0.2, calculation of a certain value for $\langle d^2 \rangle$ (eq 8) required ~ 3 times less computer time with our method than the conventional one. When this fraction was further increased to 0.3 and 0.4, our method yielded $DN \sim 0.1$ and 0.02, respectively, whereas with the conventional method the center of mass did not move at all (i.e., $D = 0$) even for very long runs. This gives reason to believe that the present method will be more efficient also for dense, many-chain systems. In the future we intend to use our method for checking de Gennes'[29] reptation hypothesis in dense entangled many-chain systems.

Acknowledgment. I thank Professor K. Binder for valuable discussions.

References and Notes

(1) Verdier, P. H.; Stockmayer, W. H. *J. Chem. Phys.* **1962**, *36*, 227.
(2) Verdier, P. H. *J. Chem. Phys.* **1966**, *45*, 2122.
(3) Kranbuehl, D. E.; Verdier, P. H. *J. Chem. Phys.* **1972**, *56*, 3145.
(4) Verdier, P. H. *J. Chem. Phys.* **1973**, *59*, 6119.
(5) Verdier, P. H.; Kranbuehl, D. E. *Polym. Prepr., Am. Chem. Soc., Div. Polym. Chem.* **1976**, *17*, 148.
(6) Kranbuehl, D. E.; Verdier, P. H. *J. Chem. Phys.* **1979**, *71*, 2662.
(7) Gény, F.; Monnerie, L. *J. Polym. Sci.* **1979**, *17*, 131, 147.
(8) Lax, M.; Brender, C. *J. Chem. Phys.* **1977**, *67*, 1785.
(9) Birshtein, T. M.; Gridnev, V. N.; Gotlib, Yu. Ya.; Skvortsov, A. M. *Vysokomol. Soedin., Ser. A* **1977**, *19*, 1398.
(10) Heilmann, O. J. *Kgl. Danske Vid. Selsk. Mat. Fys. Medd.* **1968**, *37*, 2.
(11) Heilmann, O. J.; Rotne, J. *J. Stat. Phys.* **1982**, *27*, 19.
(12) Gurler, M. T.; Crabb, C. C.; Dahlin, D. M.; Kovac, J. *Macromolecules* **1983**, *16*, 398.
(13) Bishop, M.; Kalos, M. H.; Frisch, H. L. *J. Chem. Phys.* **1979**, *70*, 1299.
(14) Ceperley, D.; Kalos, M. H.; Lebowitz, J. L. *Phys. Rev. Lett.* **1978**, *41*, 313.
(15) Ceperley, D.; Kalos, M. H.; Lebowitz, J. L. *Macromolecules* **1981**, *14*, 1472.
(16) Baumgärtner, A. *J. Chem. Phys.* **1980**, *72*, 871.
(17) Rouse, P. E. *J. Chem. Phys.* **1953**, *21*, 1273.
(18) Orwoll, R. A.; Stockmayer, W. H. *Adv. Chem. Phys.* **1969**, *15*, 305.
(19) Iwata, K.; Kurata, M. *J. Chem. Phys.* **1969**, *50*, 4008.
(20) de Gennes, P.-G. *Macromolecules* **1976**, *9*, 587.
(21) Jasnow, D.; Moore, M. A. *J. Phys. (Paris)* **1977**, *67*, 4608.
(22) Hilhorst, H. J.; Deutch, J. M. *J. Chem. Phys.* **1975**, *63*, 5153.
(23) Boots, H.; Deutch, J. M. *J. Chem. Phys.* **1977**, *67*, 4608.
(24) Meirovitch, H. *J. Phys. A* **1982**, *15*, L735.
(25) Meirovitch, H. *J. Chem. Phys.* **1983**, *79*, 502.
(26) Yahata, H.; Suzuki, M. *J. Phys. Soc. Jpn.* **1969**, *30*, 657.
(27) Binder, K., Ed. In "Monte Carlo Methods in Statistical Physics"; Springer-Verlag: Berlin, 1979.
(28) Domb, C. *J. Chem. Phys.* **1963**, *38*, 2957.
(29) de Gennes, P.-G. "Scaling Concepts in Polymer Physics"; Cornell University Press: Ithaca, NY, 1979.

Free Energy Determination of Polypeptide Conformations Generated by Molecular Dynamics

Alfredo Di Nola,† Herman J. C. Berendsen,* and Olle Edholm‡

Laboratory of Physical Chemistry, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands. *Received June 30, 1983*

ABSTRACT: Molecular dynamics (MD) simulations of macromolecules are normally restricted to tens of picoseconds and do not produce statistical averages over all relevant conformational states. It is shown for the tetradecapeptide somatostatin as an example that MD at elevated temperatures can be used to generate new conformations. The statistical weight of each conformation depends on its entropy. The entropy for two conformations has been determined by analysis of the distribution functions of generalized coordinates (bond angles and dihedral angles) and by their correlations. The results show that the significant difference in internal energy of the two conformations is compensated by entropy differences, yielding comparable free energies. Effects of constraints, coordinate transformations, and quantum corrections are discussed.

1. Introduction

The method of molecular dynamics (MD), simulating the dynamical behavior of a molecular system by integrating Newton's equations of motion, has so far been successfully applied to a large variety of systems. These include small peptides in solution,[1] proteins in vacuo,[2-5] and a protein including surrounding water in a crystalline environment.[6] Up to now MD has not been applied to flexible polypeptides with a large number of conformations that may have rather different properties. Many biologically important molecules as hormones belong to this class. Although smaller, such molecules are more difficult to study with MD techniques than proteins that cover a more confined region of conformational space. The passage between different conformational states is too rapid to isolate and study the separate conformational states. Thus, the experimentally known properties of such molecules are averages over many conformations. To interpret experimental properties, a range of individual conformations with their relative probabilities has to be studied by theoretical methods. MD is generally limited to time spans in the range 10–100 ps, which is insufficient to obtain appreciable averaging over different conformations.

The purpose of this paper is twofold. First, we show that it is possible to generate new conformations by heating the system, i.e., by performing MD simulations at a temperature of 600–1200 K and subsequently cooling to room temperature. Secondly, we try to estimate thermodynamic properties of the conformational substates, with the aim of determining their statistical weight. The latter is proportional to $\exp(-\Delta G/k_B T)$, where $\Delta G = \Delta H - T\Delta S$ is the

† Permanent address: Istituto di Chimica-Fisica, University of Rome, P. le A. Moro 5, 00185 Rome, Italy.
‡ Permanent address: Department of Theoretical Physics, Royal Institute of Technology, S-100 44 Stockholm 70, Sweden.

0024-9297/84/2217-2044$01.50/0

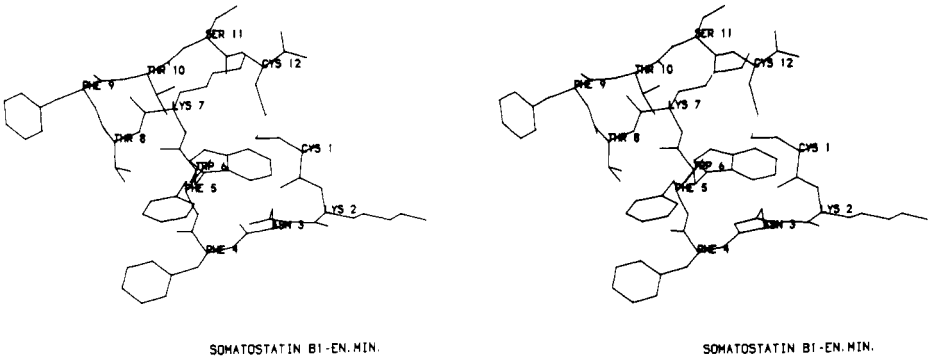

Figure 1. Stereopicture of the conformation B_1 after energy minimization.

Table I
Stereochemical Code[a]

code	φ[b]	ψ[b]	code	φ	ψ
a	−60	−40			
b	−160	160	e	−80	80
c	−140	80	a*	60	40
d	−80	160	e*	80	−80

[a] Reference 20. [b] Reference 21.

difference in Gibbs free energy. While the enthalpy difference ΔH can be obtained directly from the simulation, the entropy difference ΔS is not so easily obtained. We will evaluate methods that have been proposed earlier[7,8] to estimate the configurational entropy from distribution functions of molecular coordinates.

For this purpose we have chosen the tetradecapeptide somatostatin (somatropin release inhibiting factor, SRIF) which has been extensively studied with several different techniques: circular dichroism,[9,10] infrared,[11] NMR,[11-18] and conformational analysis using energy minimization.[11-14,17] These studies suggest the coexistence of a large number of conformations. In another paper[19] we report the comparison of various theoretical conformations with NMR data.

The molecule actually simulated in this study is the *N*-acetyl (des Ala-Gly) analogue of somatostatin: CH_3-CO-Cys-Lys-Asn-Phe-Phe-Trp-Lys-Thr-Phe-Thr-Ser-Cys. In section 2 the interaction model and the MD methods used are described. Section 3 describes the theoretical background for the analysis of the thermodynamic properties of simulated conformations. Two conformations are compared in section 4, and the relative importance of various contributions to the entropy is discussed. The general applicability and limitations of the proposed methods are discussed in section 5.

2. Generation of Conformations

Recently, Knappenberg et al.[15] have performed a conformational analysis of the somatostatin analogue. We chose one of the low-energy conformations compatible with NMR results, named B_1 (Figure 1), which is characterized by the stereochemical code[20] (see Table I) eebaeea*ea*bdb, as a basis for MD simulations. Dynamic simulations were performed by solving the equations of motion using the "leap frog" algorithm[22] and incorporating a weak coupling to a constant-temperature bath.[23] The results of simulations in water will be published elsewhere;[19] in the present paper no solvent interactions were taken into account, but the charges on the amino group of the two Lys were omitted and the charge on the COO^- terminal was replaced by a dipole. The interaction functions used are the same as used in a full MD of trypsin inhibitor,[6] the details of which will be published elsewhere. The potential function comprises terms for bond angles, dihedral angles, and

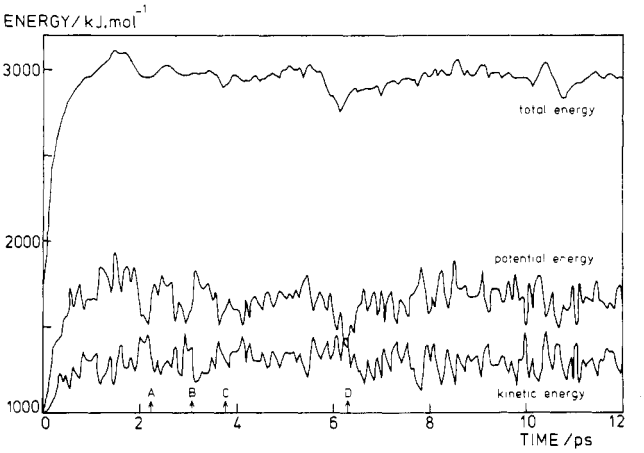

Figure 2. Total, kinetic, and potential energies as a function of time for the MD run at 1200 K. The times for which the A, B, C, and D conformations are obtained after cooling and energy minimization are marked.

electrostatic and van der Waals energies. Hydrogen atoms attached to carbon are incorporated into the latter, but other hydrogen atoms are individually treated. No explicit hydrogen-bonding interaction is applied since this interaction is adequately modeled by electrostatic and van der Waals forces. A dielectric constant of 1 is used and interactions are truncated at 8 Å. All bond lengths are constrained in the simulation using the SHAKE method.[24] The time step was taken as 0.002 ps; 1 ps of simulation required 100 s of CP time on a CDC Cyber 170/760.

The original conformation was first energy refined by 98 cycles of steepest descent minimization. Subsequently an equilibration of 10 ps at 300 K was allowed; several structural transitions occurred in the first 5 ps. For the analysis a further run of 20 ps was carried out at 300 K, during which time no changes in backbone configuration occurred.

In order to generate new conformations, the temperature was increased first during 10 ps to 600 K by coupling to a 600 K bath with a time constant of 0.1 ps. The starting point was the conformation obtained after 16 ps of simulation at 300 K. Subsequently the temperature was raised to 1200 K, and a further 14 ps of simulation was performed, using a time step of 0.0014 ps. The time course of kinetic, potential, and total energy of this simulation in the first 12 ps is given in Figure 2. On a short time scale fluctuations in kinetic and potential energy cancel approximately; on a longer time scale the heat bath coupling allows fluctuation of the total energy as well. A conspicuous fluctuation in the total energy (accompanied by a potential energy fluctuation) signals a possibly significant conformational change. At four points in the simulation (denoted by A–D in Figure 2), where minima in the total

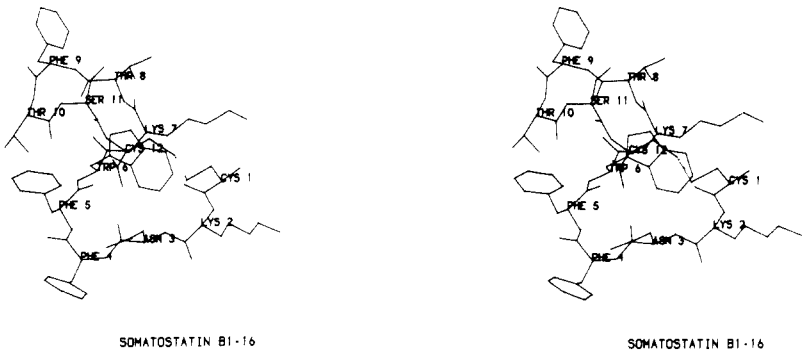

Figure 3. Stereopicture of conformation B_1 16 ps after energy minimization.

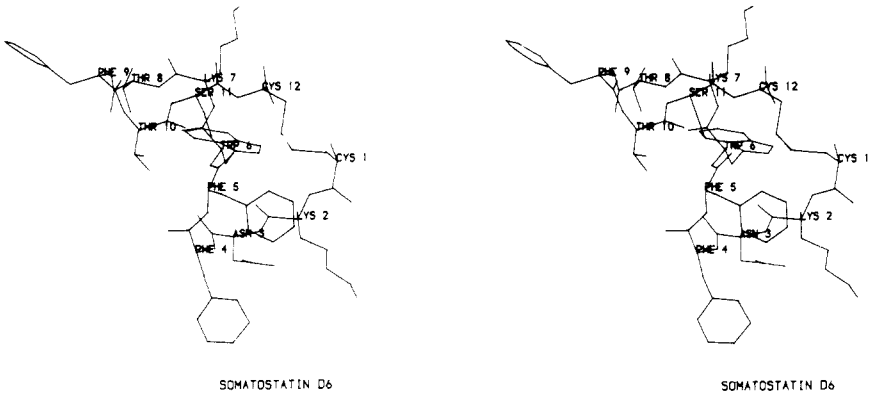

Figure 4. Stereopicture of conformation D 6 ps after energy minimization.

Table II
Conformations at Different Times of the MD Simulation

time after minimization, ps	conf	stereochem code	$\langle E_{pot} \rangle$, kJ/mol	$\langle \Delta E_{pot} \rangle^{1/2}$, kJ/mol
3–5	B_1	eebaeea*ea*bdb	610	15
11–21	B_1'	eebaedaee*eee*	574	13
21–31	B_1'	eebaedaee*eee*	578	10
3–5	A	ee*daeea*ea*a*ec	613	20
3–5	B	ba*e*aede*ae*eee*	657	17
3–6	C	bde*aeee*ee*ebd	617	17
3–6	D	abaacde*ee*ee*e	635	23
6–16	D′	ace*aebaee*ee*e	601	28
16–26	D″	ace*aedaee*ee*a	604	10

energy occurred, the configurations were taken as starting points for 300 K simulations, after energy minimization and 5–6 ps of equilibration at 300 K. In Table II the stereochemical code for these conformations at the end of each equilibration are given, including the average potential energy and the fluctuation in the potential energy during the given time interval. Table II also gives the configuration of the conformer B_1 after 5, 21, and 31 ps, respectively.

Figure 3 gives the conformation of B_1 16 ps after minimization (denoted by B_1' in Table II). The high-temperature run rapidly produces appreciable conformational changes. The conformation D was selected for further analysis; its structure 6 ps after minimization is shown in Figure 4. After minimization and equilibration conformation D has a higher potential energy than B_1; moreover, it shows several conformational changes during the simulation used for analysis (6–26 ps). The average energy of the D′/D″ conformations is significantly higher than the average energy of the B_1' conformation. The accuracy in the average potential energy is estimated as a few kJ/mol.

3. Analyses of Entropy and Free Energy

The entropy of a macromolecule is determined classically by the distribution function over a sufficiently large ensemble in phase space $P(p,q)$. In Cartesian coordinates the contribution due to the distribution in impulse space can be integrated,[8] yielding a term that is the same for all conformations. A configurational entropy S_C remains:

$$S_C = -k_B \int d^{3N}\mathbf{x}\, P(\mathbf{x}) \ln P(\mathbf{x}) \qquad (1)$$

A complication arises when constraints are incorporated into the simulation, yielding a distribution on a hypersurface in configurational space. If constrained degrees of freedom are eliminated by proper choice of generalized coordinates and impulses, the contribution of the impulse part can no longer be separated. Integration over impulse space yields a metric determinant in the configurational integrals. If only bond lengths (no bond angles) are constrained, this metric determinant will only depend upon the fixed bond lengths and the bond angles.[25,26] Since bond angles are confined to a very narrow distribution, with equal mean value for different conformations, this metric determinant will only have a slight influence on the entropy or free energy differences between conformations. It is important, however, in order to obtain correct absolute values of the entropy. If bond angles are constrained two difficulties arise: first, the metric tensor now becomes a function of the dihedral angles as well, which are not confined to a narrow conformation-independent distribution. Secondly, the simulation should take the metric tensor effect into account, e.g., by using a corrected potential.[26-28] Moreover, the statistical and dynamical behavior of macromolecules is significantly distorted by constraining bond angles.[27-29] For these reasons we have only constrained bond lengths in our simulations.

In the following we use a classical treatment for the calculation of entropy. This procedure is not correct for modes with frequencies larger than kT/h. Such modes include vibrations involving bond angles and dihedral angles. Although quantum corrections can be made, we have not attempted to include them. When the high-frequency range of the spectral density is the same for

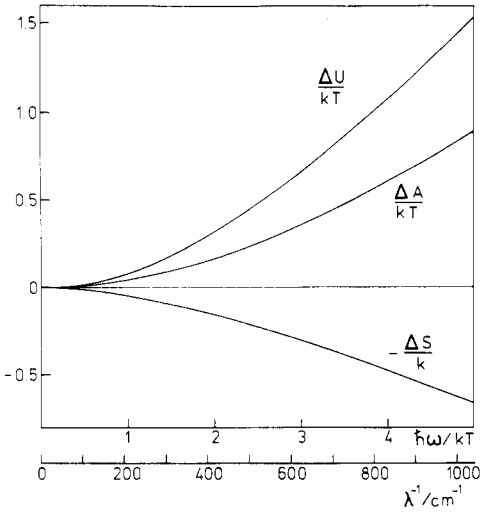

Figure 5. Quantum corrections to entropy, internal energy, and free energy for a harmonic degree of freedom as a function of frequency. The wavenumber scale applies to a temperature of 300 K.

different conformations, the quantum corrections will cancel when differences are compared. Only wavenumbers higher than 500 cm^{-1} involve appreciable corrections to the free energy. In Figure 5 quantum corrections to both entropy and free energy are shown; it is interesting to note that in the free energy corrections the energy term (mainly resulting from zero-point energy) more than compensates the entropy term. It would thus not be justified to apply a quantum correction to the entropy alone. We may safely assume that the higher frequency modes are not strongly conformation-dependent. Although corrections are important for absolute values of the entropy, they can be disregarded for differences between conformations. A similar conclusion was reached by Karplus and Kushick.[7]

Transforming to internal coordinates **q**, such as dihedral and bond angles, the total entropy becomes

$$S = S_q + S_p \tag{2}$$

where

$$S_q = -k_B \int P(\mathbf{q}) \ln P(\mathbf{q}) \, d\mathbf{q} \tag{3}$$

and

$$S_p = k_B \left[\frac{N_\alpha}{2} + \frac{N_\alpha}{2} \ln \left(\frac{2\pi k_B T}{h^2} \right) + \frac{1}{2} \int d\mathbf{q} \, P(\mathbf{q}) \ln (\det \mathbf{G}^\alpha(\mathbf{q})) \right] \tag{4}$$

N_α is the number of internal coordinates and the metric tensor $\mathbf{G}^\alpha$ is

$$G_{ij}^{\alpha} = \sum_l m_l \frac{\partial x_l}{\partial q_i} \frac{\partial x_l}{\partial q_j} \tag{5}$$

We note that S_q and S_p cannot be identified as a configurational part and an impulse part.

Equation 4 is given for the general case when det $\mathbf{G}^\alpha$ is conformation-dependent. If only bond lengths are constrained, the term S_p is almost conformation-independent because the distribution of bond angles is narrow and does not strongly depend on conformation. Differences in entropy will then result mainly from S_q. The absolute value of S_q depends on the choice of coordinates and units and has physical significance only in conjunction with S_p.[8]

The evaluation of S_q is carried out along the principles outlined before.[8] If the distribution function $P(\mathbf{q})$ is a multivariate Gaussian, S_q follows from the treatment of Karplus and Kushick:[7]

$$S_q(\text{Gaussian}) = \tfrac{1}{2} k_B [N_\alpha (1 + \ln 2\pi) + \ln \det \sigma] \tag{6}$$

where σ is the covariance matrix

$$\sigma_{ij} = \langle (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \rangle \tag{7}$$

The Gaussian entropy of (6) can be considered to consist of an uncorrelated or diagonal contribution

$$S_d(\text{Gaussian}) = \tfrac{1}{2} k_B [N_\alpha (1 + \ln 2\pi) + \sum_i \ln \sigma_{ii}] \tag{8}$$

and a term due to correlations between these degrees of freedom. The real multidimensional distribution $P(\mathbf{q})$ deviates from Gaussian but cannot be evaluated in a statistically accurate way from the simulation. It is however possible to determine the one-dimensional distributions $P_i(q_i)$ with sufficient accuracy to determine the "exact" diagonal contribution:

$$S_d = -k_B \sum_i \int dq_i \, P(q_i) \ln P(q_i) \tag{9}$$

The best approximation to the total configurational entropy is obtained by using (9) for the diagonal contribution and the Gaussian approximation for the contribution due to correlation. In this way the best entropy estimate is obtained by using the knowledge of σ as well as $P(q_i)$. Thus we obtain

$$S_q = -k_B \sum_i \int dq_i \, P(q_i) \ln P(q_i) + \frac{k_B}{2} \ln \left(\frac{\det \sigma}{\prod_i \sigma_{ii}} \right) \tag{10}$$

The evaluation of the first term in (10), S_d, is based on histograms of P_i obtained from the simulation. Depending on the interval sizes used, systematic errors arise due to the nonlinear form of the functional. The systematic error is eliminated by evaluating $S_d(n)$ as a function of the number n of intervals. $S_d(n)$ can be shown[8] to relate to n as

$$S_d(n) = S_d + nA - B/n^2 \tag{11}$$

The coefficients A and B and the final estimate S_d are obtained from a least-squares fit to eq 11 of the $S_d(n)$ obtained from the simulation.

4. Results for Two Conformations

We have performed a 20-ps MD simulation of the B_1' and D′/D″ conformations (see Table II) and calculated the total conformational entropy and different contributions to it by different methods.

Figure 6 shows the total conformational entropy and contributions of different types of degrees of freedom to it. The values were derived from eq 10, using degrees as units for bond and dihedral angles. We emphasize that the absolute values of the entropy scale and also the differences between different curves have no physical meaning, unless the contributions of S_p of the various subspaces (eq 4) are added. The calculation of these contributions is very involved and the results are not relevant for the total entropy of the molecule.

The total configurational entropy is still increasing after 20 ps, although rather slowly. Contributions involving fewer degrees of freedom increase still more slowly. The convergence of the curves seems to be well described by a t^{-1} behavior (Figure 7). This is also what would be expected theoretically, since the number of different conformations is proportional to time and we expect the systematic error due to insufficient statistics in phase space

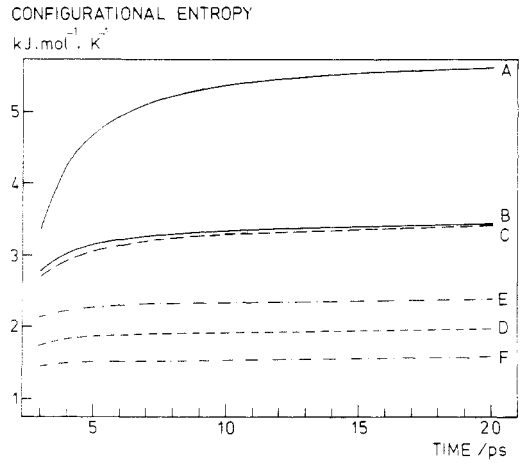

Figure 6. Configurational entropy of the D conformation as obtained with different lengths of simulation. A is total, B total backbone, C total dihedral, D backbone dihedral, E total bond angle, and F backbone bond angle configurational entropy. The curves show the different rates of convergence. To obtain comparable absolute sizes the momentum and Jacobian contributions S_p must be added.

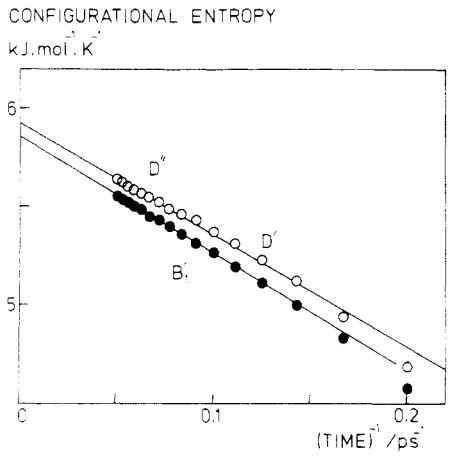

Figure 7. Total configurational entropy of the B_1 and D conformations plotted against the inverse of the simulation length used. The points follow straight lines that are extrapolated to give a final estimate of the entropy.

to be proportional to the inverse of the number of independent configurations.[8] We may thus extrapolate to get a final entropy estimate.

The accuracy of the extrapolation can be increased by a longer simulation, but then the probability increases that the system will pass into other conformations. The total entropy from a simulated trajectory covering more than one conformation can be shown to equal the weighted average entropy of the conformations plus a mixing term $-k_B\sum_i w_i \ln w_i$, w_i being the statistical weight of the ith conformation, if the distributions of the different conformations do not overlap. If the constituent conformations and their entropies are not very different, we may consider them as one conformation. There is room for some arbitrariness in the definition of a conformation. For example, side-chain dihedral changes occur frequently during the simulation; these are not considered to produce new conformations. The D′/D″ trajectory actually contains two slightly different backbone conformations (Table II) and the effect of this is visible in the change of slope in the entropy curve (Figure 7) around $t^{-1} = 0.1$ ps^{-1}.

Figure 6 gives the total entropy, including both the effect of coupling (in the Gaussian approximation) and the effect of the non-Gaussian character of the distribution functions

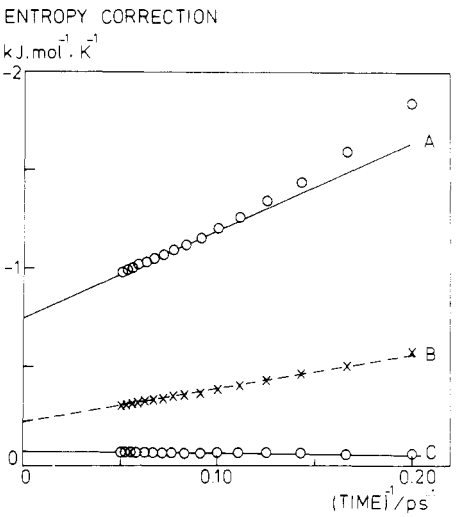

Figure 8. Corrections to the total Gaussian uncoupled entropy due to (A) coupling between all types of degrees of freedom, (B) couplings only in the backbone, and (C) non-Gaussian character of uncoupled distribution functions.

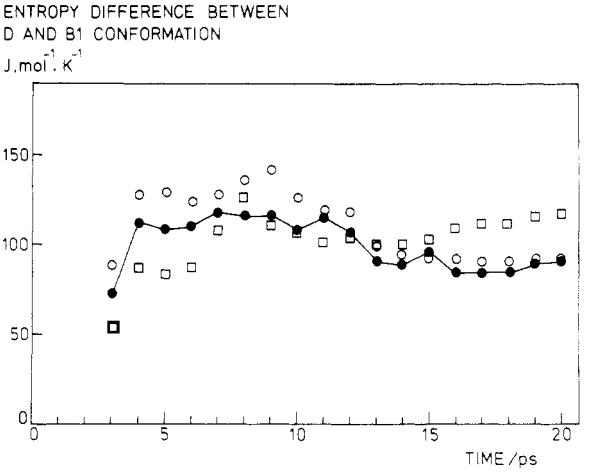

Figure 9. Difference in configurational entropy between D and B_1 conformation as a function of simulation length used for (○) multivariate Gaussian and (□) independent non-Gaussian. With (●) the corrected results according to (10) are shown.

Table III
Entropy Corrections due to Coupling (J mol^{-1} K^{-1})

	total	backbone
dihedrals	−636	−125
bond angles	−43	−34
bond angle − dihedrals	−68	−62
	−747	−224

(in the diagonal approximation). In Figure 8 these two contributions are shown. It is clear that the effect of coupling is quite significant and exceeds the effect of the non-Gaussian character by an order of magnitude. This dominance is not general and depends on the choice of coordinates; an example where the non-Gaussian character predominates is a bilayer membrane of hydrocarbon chains.[8] The coupling correction is for about 2/3 due to the side chains themselves and the side chain–backbone coupling, and for $^1/_3$ to the backbone (Figure 8). The small correction for the non-Gaussian character is almost entirely due to the side chains.

The total coupling correction results from three types: mutual coupling between bond angles, mutual coupling between dihedral angles, and coupling between bond and dihedral angles. As can be seen from Table III, the most important contribution is the mutual coupling between

Table IV
Thermodynamic Results (T = 300 K)

	B_1'	D′/D″	$B_1' - D'/D''$
config entropy (J mol^{-1} K^{-1})	5845	5911	−70 ± 20
abs total entropy (J mol^{-1} K^{-1})	1876	1937	−60 ± 40
pot energy (kJ mol^{-1})	576	603	−27 ± 3
int energy (kJ mol^{-1})	894	921	−27 ± 3
$-TS$ (kJ mol^{-1})	−563	−581	+18 ± 12
free energy $U - TS$ (kJ mol^{-1})	331	340	−9 ± 12

dihedrals. It is interesting to note that the contribution of coupling of bond angles with dihedrals exceeds that of the mutual coupling of bond angles.

In Figure 9 the difference in total configurational entropy between the D′/D″ and B_1' conformations is shown. The various approximations used give very similar results. Already after 5 ps the differences between the two conformations are close to their final values, although the entropy of each conformation is still 1000 J mol^{-1} K^{-1} removed from its extrapolated final value (Figure 7). We estimate the error of the final difference (90 J mol^{-1} K^{-1} at 20 ps (Figure 9), 70 J mol^{-1} K^{-1} extrapolated from Figure 7) to be about 20 J mol^{-1} K^{-1}. The accuracy of a least-squares extrapolation is better, but the uncertainty of the procedure does not warrant a higher precision.

In order to estimate the absolute entropy we also calculated the momentum/Jacobian contributions using eq 4 and 5. To calculate det ($\mathbf{G}^\alpha$) we used

$$\det(\mathbf{G}^\alpha) = \det(\mathbf{H}^\beta)/\det(\mathbf{H}) \tag{12}$$

where

$$H_{ij} = \sum_l \frac{1}{m_l} \frac{\partial q_i}{\partial x_l} \frac{\partial q_j}{\partial x_l} \tag{13}$$

and H_{ij}^β is a similar matrix using only the constrained degrees of freedom q_i^β. This method of computing det ($\mathbf{G}^\alpha$), due to Fixman,[26] makes the computation, which is prohibitive by direct determination of $\mathbf{G}^\alpha$, feasible.

The final result for the total molecule depends only on the absolute entropies using bond constraints only. It turns out that the contribution of S_p is slightly conformation-dependent because the distribution of bond angles is not infinitely sharp. The values for S_p used to compute absolute entropies given in Table IV were obtained by averaging over a number of sample configurations from the simulations. Table IV lists the final thermodynamic quantities for the B_1' and the D′/D″ conformations.

5. Discussion

The results summarized in Table IV clearly show the importance of adding entropic contributions to the energy of simulated conformations. While from potential energy considerations the B_1' conformation would be significantly preferred to the D′/D″ conformation, this is not so if entropic contributions are included. The method used to compute entropy is quite feasible as long as solvent effects are excluded. The accuracy does not allow discrimination of various conformations at the level lower than kT in free energy differences. The obtained accuracy of about $5kT$ is quite satisfactory in view of the uncertainties in the energy due to inadequacies of the presently available potential functions.

The main inadequacy of the potential functions used is the omission of solvent interactions. Solvent can be included either by introducing solvent molecules[6] or by using appropriate potentials of mean force.[30] Only in the latter case is it still possible to compute the entropy in the same manner as described here: entropic contributions due to macromolecule–solvent interactions appear in the potential energy of the simulations because the potentials of mean force represent free energies with respect to degrees of freedom of the solvent.

It appeared that correlations between various degrees of freedom are of considerable influence on the total values of the configurational entropies. However, for the entropy differences of the two conformations considered here, they are not important. The non-Gaussian character of the diagonal distribution functions is less important than the correlations are. It is not possible to assess the importance of the non-Gaussian character of off-diagonal correlations. Their magnitude depends on the choice of generalized coordinates.[8]

No general guiding principle exists to derive the best generalized coordinates. There is a unique criterion, however, by which the preference of one choice among others can be established: the lowest entropy obtained will be the best estimate. This is a consequence of the fact that the method described here is an application of the maximum entropy formalism of Jaynes:[31] the derived entropy is the maximum estimate consistent with the "knowledge" of the configurational distribution embodied in the measured quantities $\langle q_i q_j \rangle$ and $P_i(q_i)$. Thus the entropy obtained is an upper bound of the actual entropy of the system. The choice of coordinates that gives the lowest entropy abstracts the most information of the configurational distribution and hence gives the best entropy estimate.

Acknowledgment. We are indebted to Dr. K. Hallenga for providing initial conformations of somatostatin and to Dr. W. F. van Gunsteren for providing molecular dynamics programs. A.D.N. acknowledges support by a fellowship of the Italian National Research Council (CNR) through the Netherlands Organization of Pure Research (ZWO). O.E. acknowledges support by a fellowship of the Swedish Natural Scientific Research Council (NFR).

Registry No. CH_3-CO-Cys-Lys-Asn-Phe-Phe-Trp-Lys-Thr-Phe-Thr-Ser-Cys, 54518-52-4; somatostatin, 38916-34-6.

References and Notes

(1) Rossky, P. J.; Karplus, M. *J. Am. Chem. Soc.* **1979**, *101*, 1913.
(2) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature (London)* **1977**, *267*, 585.
(3) McCammon, J. A.; Karplus, M. *Annu. Rev. Phys. Chem.* **1980**, *31*, 29.
(4) Karplus, M.; McCammon, J. A. *CRC Crit. Rev. Biochem.* **1981**, *9*, 293.
(5) Levitt, M. *Nature (London)* **1981**, *294*, 379.
(6) van Gunsteren, W. F.; Berendsen, H. J. C.; Hermans, J.; Hol, W. G. J.; Postma, J. P. M. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 4315.
(7) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325.
(8) Edholm, O.; Berendsen, H. J. C. *Mol. Phys.* **1984**, *51*, 1011.
(9) Holladay, L. A.; Puett, D. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 1193.
(10) Holladay, L. A.; Rivier, J.; Puett, D. *Biochemistry* **1977**, *16*, 4895.
(11) Shaw-hin, Han; Rivier, J. E.; Scheraga, H. A. *Int. J. Pep. Protein Res.* **1980**, *15*, 355.
(12) Momany, F. A.; Drake, L. G.; Aubuchon, J. R. *Int. J. Quant. Chem.* **1978**, *5*, 381.
(13) Hallenga, K.; Van Binst, G.; Knappenberg, M.; Brison, J.; Michel, A.; Dirkx, J. *Biochim. Biophys. Acta* **1979**, *577*, 82.
(14) Knappenberg, M.; Brison, J.; Dirkx, J.; Hallenga, K.; Deschrijver, P.; Van Binst, G. *Biochim. Biophys. Acta* **1979**, *580*, 266.
(15) Knappenberg, M.; Michel, A.; Scarso, J.; Brison, J.; Zanen, J.; Hallenga, K.; Deschrijver, P.; Van Binst, G. *Biochim. Biophys. Acta* **1982**, *780*, 229.
(16) Buffington, K.; Garsky, V.; Massiot, G.; Rivier, J.; Gibbons, W. A. *Biochim. Biophys. Res. Commun.* **1980**, *93*, 376.
(17) Hallenga, K.; Van Binst, G.; Scarso, A.; Michael, A.; Knappenberg, M.; Dremier, C.; Brison, J.; Dirkx, J. *FEBS Lett.* **1980**, *119*, 47.
(18) Kaptein, R.; Dijkstra, K. *J. Magn. Reson.* **1978**, *31*, 171.

(19) Di Nola, A.; Berendsen, H. J. C.; Hallenga, K., to be published.
(20) Liquori, A. M. *Q. Rev. Biophys.* **1969**, *2*, 65.
(21) IUPAC–IUB Commission on Biochemical Nomenclature *Biochemistry* **1970**, *9*, 3471.
(22) Hockney, R. W.; Eastwood, J. W. "Computer Simulations Using Particles"; McGraw-Hill, New York, 1981.
(23) Berendsen, H. J. C.; van Gunsteren, W. F.; Postma, J. P. M.; Haak, J. R.; Di Nola, A., *J. Chem. Phys.*, in press.
(24) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(25) Gō, N.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 535.
(26) Fixman, M. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 3050.
(27) Berendsen, H. J. C.; van Gunsteren, W. F. In "Proceedings, NATO Advanced Study Institute on Physics of Superionic Conductors and Electrode Materials"; Perram, J. W., Ed.; Plenum Press: New York, 1983.
(28) van Gunsteren, W. F. *Mol. Phys.* **1980**, *40*, 1015.
(29) van Gunsteren, W. F.; Karplus, M. *Macromolecules* **1982**, *15*, 1528.
(30) Paterson, Y.; Némethy, G.; Scheraga, H. A. *J. Solution Chem.* **1982**, *11*, 831.
(31) Jaynes, E. T. *Phys. Rev.* **1957**, *106*, 620.

Why Does the Generalized Stokes–Einstein Equation Work?†

George D. J. Phillies

Department of Chemistry, The University of Michigan, Ann Arbor, Michigan 48109.
Received February 2, 1984

ABSTRACT: The generalized Stokes–Einstein equation (GSE) $D_m = D_s\beta(\partial\pi/\partial c)_{P,T}(1 - \phi)$ and various microscopic theories for the mutual diffusion coefficient D_m are compared with the available experimental data on solutions of rigid macromolecules. The GSE and microscopic theories do not agree; the bulk of the experimental data favors the GSE. Methods of modifying microscopic theories in order to obtain agreement with the GSE are considered. If one neglects the Oseen $(a/r)^1$ term in the cross-diffusion tensor, satisfactory results are obtained. A rationale for this neglect, based on the nonzero frequencies of molecular motions, is noted.

I. Introduction

Our objective in this paper is to compare the available experimental data on the diffusion of macroparticulate species at high concentration with some of the theories which have been proposed to explain this behavior. Two major classes of theory are noted: semimicroscopic theories and microscopic theories. In the microscopic theories, diffusion coefficients are calculated from ensemble averages over products of intermolecular potentials and hydrodynamic interaction tensors. In semimicroscopic theories, calculations are made in terms of concentration gradients, associated fluxes, and cross-diffusion tensors, but hydrodynamic and direct interactions between individual particles are not given a detailed representation. It is often not sufficiently emphasized that these two classes of theory do not make the same predictions for D at high concentration. Furthermore, within the class of microscopic theories, there are significant disagreements as to what quantities are to be averaged to obtain D.

Our major emphasis is on the behavior of rigid spheroidal macroparticles, primarily protein molecules and silica spheres. One could also consider the dynamics of interacting, flexible polymer chains. However, this latter problem is far more complex and has substantially more free parameters. It is here preferred not to treat the chain problem before the dynamics of points is demonstrably understood. The virtues of this choice are suggested in section IV.

Section II of this paper summarizes semimicroscopic and microscopic theories for D. Section III notes useful experimental data on this problem and makes a numerical comparison of various theories with experiment. In most cases, experiment is in good agreement with the semimicroscopic theory. A discussion is found in section IV.

II. Theories of the Diffusion Coefficients

As is well-known, a characterization of macroparticle diffusion requires two translational diffusion coefficients. One of these, the mutual diffusion coefficient D_m, describes the relaxation of a real concentration gradient by the Brownian motion of the individual macromolecules and by the intermacromolecular forces. The intermacromolecular forces are reflected in the excess chemical potential μ^e of the solute. The other diffusion coefficient, the self (or tracer) diffusion coefficient D_s, describes the motion of an individual solute molecule through a uniform solution. D_s may be obtained by labeling some of the solute molecules, establishing countervailing gradients in the concentrations of labeled and unlabeled molecules (so that the total solute concentration is everywhere the same), and measuring the relaxation of the labeled solute's concentration gradient.

Treatments based on irreversible thermodynamics, i.e., semimicroscopic treatments, obtain[1,2]

$$D_s = k_B T/f_s \tag{2.1a}$$

$$D_m = \left[\left(\frac{\partial\pi}{\partial c}\right)_{P,T}(1 - \phi)\right] \Big/ f_m \tag{2.1b}$$

Here, k_B is Boltzmann's constant, T is the absolute temperature, $(\partial\pi/\partial c)_{P,T}$ is the isothermal osmotic compressibility, f_s and f_m are the drag coefficients for self and mutual diffusion, $\beta = (k_B T)^{-1}$, and ϕ is the volume fraction of solute ($\phi = c\bar{v}$, $\bar{v}$ being the thermodynamic partial volume of the solute). From the Onsager reciprocal relations and the physical quasi-identity of labeled and unlabeled molecules, it is possible to construct a self-consistent treatment of D_s and D_m which relates f_s and f_m, indicating[3]

$$f_s = f_m \tag{2.2}$$

from which follows the generalized Stokes–Einstein equation[3] (GSE)

$$D_m = D_s\beta\left(\frac{\partial\pi}{\partial c}\right)_{P,T}(1 - \phi) \tag{2.3a}$$

or[4]

$$D_m = D_s[S(k)]^{-1}(1 - \phi) \tag{2.3b}$$

where $S(k)$ is the static structure factor at wave vector k.

†Support of this work by the National Science Foundation under Grant CHE82-13941 is gratefully acknowledged.

0024-9297/84/2217-2050$01.50/0