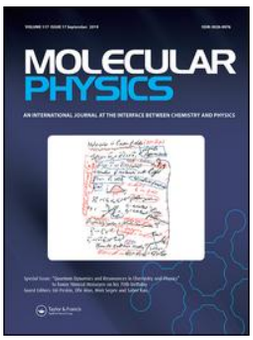

Molecular Physics

ISSN: 0026-8976 (Print) 1362-3028 (Online) Journal homepage: https://www.tandfonline.com/loi/tmph20

# Statistical thermodynamics of hindered rotation from computer simulations

EUGENE DEMCHUK & HARSHINDER SINGH

**To cite this article:** EUGENE DEMCHUK & HARSHINDER SINGH (2001) Statistical thermodynamics of hindered rotation from computer simulations, Molecular Physics, 99:8, 627-636, DOI: 10.1080/00268970010024353

**To link to this article:** https://doi.org/10.1080/00268970010024353

MOLECULAR PHYSICS, 2001, VOL. 99, NO. 8, 627–636

# Statistical thermodynamics of hindered rotation from computer simulations

EUGENE DEMCHUK[1,2*] and HARSHINDER SINGH[3]

[1] Health Effects Laboratory Division, National Institute for Occupational Safety and Health, Morgantown, WV 26505-2888, USA

[2] School of Pharmacy, West Virginia University, Morgantown, WV 26506-9500, USA

[3] Department of Statistics, West Virginia University, Morgantown, WV 26506-6330, USA

(*Received 11 July 2000; revised version accepted 22 November 2000*)

Due to the nonlinearity of internal molecular coordinate space, comprehensive statistical mechanical treatment of polyatomic molecules presents a formidable theoretical problem where traditional statistics does not hold. Here we introduce a new theoretical approach to the hindered internal rotor problem, which relies on simulated statistics of microstates, rather than on first-principles calculations of the partition function. The concept links traditional statistical-mechanical thermodynamics to an authentic treatment of curvilinear descriptive statistics. The results are illustrated by classical molecular dynamics simulations.

## 1. Introduction

Numeric calculations of thermodynamic properties are an increasingly popular approach to molecular modelling. Modern high-performance computing provides sufficient accuracy in calculations of enthalpy and other macroscopic observables which depend on the derivative of a partition function. However, quantities that involve summation over all microstates of the molecular system, e.g. the free energy, are known to converge poorly [1, 2]. These entropy-driven thermodynamic functions remain a computational challenge both in theory and practice [3–8].

Entropy is a measure of the freedom of a molecular system to explore its available phase space. Entropy calculations are important in order to understand the factors involved in the conformational stability, in the change of one conformation to another, and in other physical chemical molecular processes. To calculate the entropy of a statistical thermodynamic system we first consider the configurational probability distribution of the system.

Let $\Theta_1, \Theta_2, \ldots, \Theta_m$ be generalized coordinates of the system with $m$ degrees of freedom, and $f(\theta_1, \theta_2, \ldots, \theta_m)$ be the joint configurational probability density function of $\Theta_1, \Theta_2, \ldots, \Theta_m$. Assume that the kinetic energy part of the Hamiltonian of the system [9] is independent of the generalized coordinates [10]. Then

$$f(\theta_1, \theta_2, \ldots, \theta_m) = \frac{1}{Z(\beta)} \exp\left[-\beta V(\theta_1, \theta_2, \ldots, \theta_m)\right], \quad (1)$$

where $\beta = 1/k_B T$, $k_B$ is the Boltzmann constant, $T$ is the temperature, $V(\theta_1, \theta_2, \ldots, \theta_m)$ is the effective potential energy of the system, when $\Theta_i = \theta_i$, $i = 1, 2, \ldots, m$, and the configurational integral of the system, $Z(\beta)$, is chosen so that $f(\theta_1, \theta_2, \ldots, \theta_m)$ is the probability density function, i.e.

$$Z(\beta) = \int\int \ldots \int \exp\left[-\beta V(\theta_1, \theta_2, \ldots, \theta_m)\right] d\theta_1\, d\theta_2 \ldots d\theta_m.$$

The configurational part of entropy of the system, $S_c$, is given by the Boltzmann expression

$$\begin{aligned} S_c &= -k_B \langle \ln\left[f(\Theta_1, \Theta_2, \ldots, \Theta_m)\right]\rangle \\ &= -k_B \int\int \ldots \int \ln\left[f(\theta_1, \theta_2, \ldots, \theta_m)\right] \\ &\quad \times f(\theta_1, \theta_2, \ldots, \theta_m)\, d\theta_1\, d\theta_2 \ldots d\theta_m \\ &= k_B \ln\left[Z(\beta)\right] + k_B \beta \langle V(\Theta_1, \Theta_2, \ldots, \Theta_m)\rangle, \end{aligned} \quad (2)$$

where $\langle \ldots \rangle$ denotes the expectation (mean). For additional reading refer to Landau and Lifshitz [11] and to an article by Meirovitch in *Reviews in Computational Chemistry* [12] for recent advances in applications to molecular physics.

* Author for correspondence. e-mail: eed5@cdc.gov

*Molecular Physics* ISSN 0026–8976 print/ISSN 1362–3028 online © 2001 US Government
http://www.tandf.co.uk/journals
DOI: 10.1080/00268970010024353

Note that (1) gives the probability distribution of the system being in a particular configuration with the potential energy $V(\theta_1, \theta_2, \ldots, \theta_m)$ specified by the generalized coordinates $\Theta_1, \Theta_2, \ldots, \Theta_m$. Equation (2) is the fundamental equation for entropy. Its generalization is valid in both the classical and quantum mechanical formulations. It indicates the importance of realistic modelling of the probability distribution of coordinates and consequent modelling of the potential energy of the system. The present work, however, is restricted to only classical distributions.

Reliance on probability distributions that do not capture the curvilinear nature of the internal configurational space of polyatomic molecules is a major drawback of current numerical simulation methods. For instance, the coordinates $\Theta_1, \Theta_2, \ldots, \Theta_m$ have been modelled in the literature by assuming an independent Gaussian distribution for each $\theta_i$ [13]. Assume $\Theta_i$ has a normal distribution $N(\mu_i, \sigma_i^2)$, and the $\Theta_i$'s are independent, then from equation (124) in [13] the configurational entropy of the system is given by

$$S_c = k_B \left[ \frac{m}{2} \ln (2\pi) + \sum_{i=1}^{m} \ln (\sigma_i) + \frac{m}{2} \right]. \tag{3}$$

By analogy with (1), the potential energy of the system is then given by

$$V = \frac{1}{2\beta} \sum_{i=1}^{m} Z_i^2, \tag{4}$$

where $Z_i$'s are independent standard Gaussian variables.

In fact Berendsen and co-workers [13] have pursued a broader approach in which they considered a general unimodal probability density function. The latter assumption results in a Gaussian distribution for the potential energy of a large (macroscopic) system and a gamma distribution in the case of a classical Einstein solid. For liquid methanol and liquid water, an effective gamma approximation has been used for modelling the potential energy.

In another approach, Karplus and Kushick [14] and co-authors [15] have modelled internal torsional coordinates of macromolecules using a multivariate normal distribution. Assuming that the $\Theta_i$'s have a multivariate normal distribution with the variance-covariance matrix $\Sigma$, the configurational torsional entropy is given by

$$S_c = \frac{mk_B}{2} + \frac{k_B}{2} \ln [(2\pi)^m |\Sigma|]. \tag{5}$$

In this case, the potential energy of the system can be represented as a general quadratic function of the coordinates. Amended but less sophisticated clones of this and related approaches are currently popular [16–21].

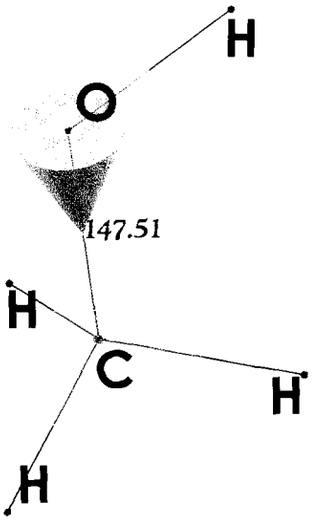


Figure 1. Methanol molecule. The cone denotes a rotatable bond.

In general, however, the assumption of a Gaussian distribution for the coordinates does not hold. For instance, a distribution of the internal Euler angle in the methanol molecule (figure 1) has three modes, and therefore, its modelling as a Gaussian is irrelevant. Generally, a Gaussian distribution is inappropriate for any molecule with large fluctuations around rotatable bonds.

A proper distribution for modelling curvilinear angular coordinates in such situations would be the von Mises distribution. Hitherto, the von Mises distribution has not been widely used although its three-dimensional analogue, the von Mises–Fisher distribution, and closely related Legendre polynomials have been applied in the order parameter theories to describe statistics in spherical polar coordinates [22–24].

In the present work we consider a representation of the angular coordinates of molecular conformations by independent von Mises distributions, and obtain associated expressions for the internal configurational entropy and the probability distribution of the potential energy. As a case study, we model the torsional angle of the methanol molecule by a three-mode von Mises distribution and theoretically derive a bathtub-shaped probability distribution for the torsional energy of the molecule. Our assumption about the three-mode von Mises distribution for the torsional angle and the theoretical derivation of the torsional energy bathtub-shaped distribution is confirmed by molecular dynamics simulations.

In section 2, we briefly discuss the shapes, properties and parametric inferences of unimodal and multimodal von Mises distributions. In section 3, we consider mod-

elling of angular coordinates of molecular conformations by independent von Mises distributions and obtain an expression for the conformational entropy. We also derive the probability density function of the potential energy. In section 4, we consider modelling of the torsional angle of the methanol molecule by a three-mode von Mises distribution and apply the results of section 3 to obtain a bathtub-shaped probability density function of the torsional energy of the molecule. In this section we also report results of a molecular dynamics simulation study. A histogram of observed torsional energies has a bathtub shape as predicted by our theoretical formulations. We fit the von Mises distribution to observations of the torsional angle and obtain an excellent fit. We fit our theoretically derived bathtub distribution to the histogram of torsional energy observations, and the resulting fit is also excellent.

## 2. The von Mises distribution

A circular random variable $\Theta$ is said to have a von Mises distribution, if its probability density function is given by

$$f(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(\theta - \theta_0)],$$

$$-\pi \leqslant \theta \leqslant \pi, \quad \kappa > 0, \quad (6)$$

where $\kappa$ is the concentration parameter, $\theta_0$ is the mean angle and $I_0$ is the modified Bessel function of zero order [25].

This distribution was introduced by von Mises [26] to study deviations of measured atomic weights from integral values. The von Mises distribution is important in statistical inferences on the circle, where its importance is the same as that of the Gaussian distribution on the line. The distribution is unimodal and symmetrical around the mode at $\theta_0$. The larger the value of $\kappa$, the greater the clustering around the mode. For $\kappa = 0$, the von Mises distribution reduces to the uniform distribution on $[-\pi, \pi]$. For large $\kappa$, the von Mises distribution is approximately Gaussian with the mean $\theta_0$ and variance $1/\kappa$; as $\kappa$ tends to $\infty$, the distribution becomes concentrated at the point $\theta_0$. In figure 2, we give plots of the probability density function of the von Mises distribution with $\theta_0 = 0$ and different values of $\kappa$.

Generalization of the unimodal von Mises distribution to a multimodal case is straightforward. A circular random variable $\Theta$ is said to have an $l$-mode von Mises distribution, if its probability density function is given by

$$f(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(l(\theta - \theta_0))],$$

$$-\pi \leqslant \theta \leqslant \pi, \quad l \in \mathbf{N}. \quad (7)$$

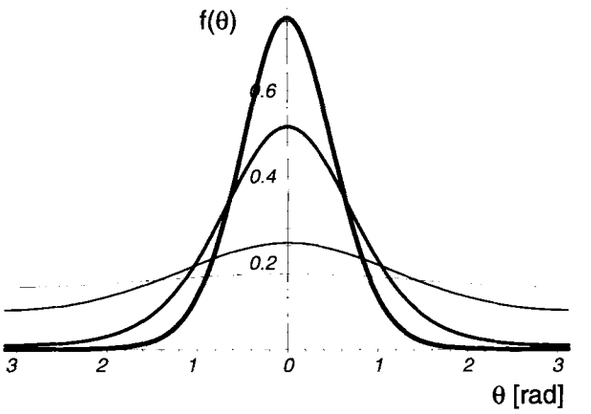


Figure 2. Probability density functions of the von Mises distribution with $\kappa = 0.10$, 0.50, 2.00, 4.00, and $\theta_0 = 0$. Increased thickness corresponds to increasing values of $\kappa$.

It has $l$ modes situated $2\pi/l$ rad apart with the first mode at $\theta = \theta_0$. This distribution approximates a mixture of $l$ equally spaced unimodal von Mises distributions. In figure 3, we give plots of the probability density function of a three-mode von Mises distribution with $\kappa = 0.10, 4.00$ and $\theta_0 = -2\pi/3$. Below we give the maximum likelihood estimators of parameters $\kappa$ and $\theta_0$ of the von Mises distribution [25].

Let $\theta^{[1]}, \theta^{[2]}, \ldots, \theta^{[n]}$ be a random sample from an $l$-mode von Mises distribution, where $l$ is known. Let

$$A = (1/n) \sum_{i=1}^{n} \cos(l\theta^{[i]}),$$

$$B = (1/n) \sum_{i=1}^{n} \sin(l\theta^{[i]}),$$

$$R = (A^2 + B^2)^{1/2}.$$

The maximum likelihood estimate of $\kappa$, $\hat{\kappa}$, is given by the solution of the equation

$$A(\hat{\kappa}) = R, \quad (8)$$

where $A(\kappa) = I_1(\kappa)/I_0(\kappa)$, and $I_0(\kappa)$ and $I_1(\kappa)$ are the modified Bessel functions of the order 0 and 1, respectively. Precalculated tables are available to determine the value of $\hat{\kappa}$ from the value of $R$ [25]. The maximum likelihood estimate of $\theta_0$ is $\hat{\theta}_0 = m_0/l$, where $m_0$ is given by the solution of equations

$$\cos(m_0) = \frac{A}{R}, \quad (9)$$

$$\sin(m_0) = \frac{B}{R}. \quad (10)$$

A large sample 95% confidence interval for $\kappa$ is given by

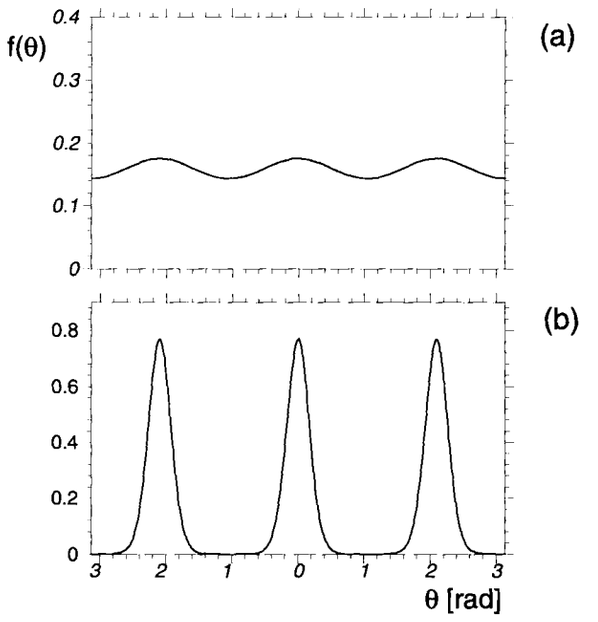


Figure 3. Probability density functions of the three-mode von Mises distribution with (*a*) $\kappa = 0.10$ and (*b*) $\kappa = 4.00$, while $\theta_0 = -2\pi/3$.

$$\hat{\kappa} \pm 1.96\frac{\hat{\sigma}(\hat{\kappa})}{n^{1/2}}, \tag{11}$$

where

$$\hat{\sigma}^2(\hat{\kappa}) = \frac{\hat{\kappa}}{\hat{\kappa} - \hat{\kappa}A^2(\hat{\kappa}) - A(\hat{\kappa})}. \tag{12}$$

## 3. Calculation of entropy assuming independent von Mises distributions for the coordinates

Consider a molecular system with the internal curvilinear torsional coordinates $\Theta_1, \Theta_2, \ldots, \Theta_m$. Let the $\Theta_i$ angles be independent and let each $\Theta_i$ have an $l_i$-mode von Mises distribution with the probability density function

$$f_i(\theta_i) = \frac{1}{2\pi I_0(\kappa_i)} \exp\left[\kappa_i \cos\left(l_i(\theta_i - \theta_{0i})\right)\right], \quad -\pi \leqslant \theta_i \leqslant \pi, \tag{13}$$

where $\kappa_i > 0$, $i = 1, 2, \ldots, m$.

The joint probability density function of $\Theta_1, \Theta_2, \ldots, \Theta_m$ is given by

$$f(\theta_1, \theta_2, \ldots, \theta_m) = \frac{1}{(2\pi)^m} \exp\left(\sum_{i=1}^{m} \kappa_i\right) \times \left(\prod_{i=1}^{m} \frac{1}{I_0(\kappa_i)}\right) \exp(-\beta V), \tag{14}$$

where

$$V = \frac{1}{\beta}\sum_{i=1}^{m} \kappa_i\{1 - \cos[l_i(\Theta_i - \theta_{0i})]\}. \tag{15}$$

The configurational torsional entropy of the system is given by

$$S_c = -k_B\langle \ln\left(f(\Theta_1, \Theta_2, \ldots, \Theta_m)\right)\rangle = k_B\left\{m \ln(2\pi) + \sum_{i=1}^{m} \ln[I_0(\kappa_i)] - \sum_{i=1}^{m} \kappa_i \langle \cos[l_i(\Theta_i - \theta_{0i})]\rangle\right\}.$$

Taking into account that $\langle \cos[l_i(\Theta_i - \theta_{0i})]\rangle = I_1(\kappa_i)/I_0(\kappa_i)$, the configurational torsional entropy of the system reduces to

$$S_c = k_B\left[m \ln(2\pi) + \sum_{i=1}^{m} \ln[I_0(\kappa_i)] - \sum_{i=1}^{m} \kappa_i \frac{I_1(\kappa_i)}{I_0(\kappa_i)}\right], \tag{16}$$

i.e. given the values of $\kappa_i$'s (or their estimates), the classical configurational torsional entropy (or its estimate) can be found using the tables of modified Bessel functions.

Now we derive the probability density function of the potential energy $V$. For simplicity, let $\theta_{0i} = -(l_i - 1)\pi/l_i$. From (15), we have

$$V = \sum_{i=1}^{m} V_i, \tag{17}$$

where

$$V_i = \frac{1}{\beta}\kappa_i\{1 - \cos[l_i(\Theta_i - \theta_{0i})]\} = \frac{V_{0i}}{2}\{1 - \cos[l_i(\Theta_i - \theta_{0i})]\}. \tag{18}$$

$V_i$ is the potential energy contribution from the coordinate $\Theta_i$ and $V_{0i} = 2\kappa_i/\beta$ is the maximum value of $V_i$. Rewriting the probability density function (13) as

$$f_i(\theta_i) = \frac{1}{2\pi I_0(\kappa_i)} \exp(\kappa_i) \exp\left(-\kappa_i\{1 - \cos[l_i(\theta_i - \theta_{0i})]\}\right), \quad -\pi \leqslant \theta_i \leqslant \pi, \tag{19}$$

and transforming

$$v_i = \frac{V_{0i}}{2}\{1 - \cos[l_i(\theta_i - \theta_{0i})]\},$$

we have

$$\left|\frac{dv_i}{d\theta_i}\right| = l_i(v_i)^{1/2}(V_{0i} - v_i)^{1/2}.$$

Thus, the probability density function of $V_i$ is given by

$$g_i(v_i) = (2l_i)\frac{1}{2\pi I_0(\kappa_i)} l_i^{-1} v_i^{-1/2} \times (V_{0i} - v_i)^{-1/2} \exp\left[\kappa_i\left(1 - \frac{2v_i}{V_{0i}}\right)\right] = \frac{1}{\pi I_0(\kappa_i)} \exp\left[\kappa_i\left(1 - \frac{2v_i}{V_{0i}}\right)\right] v_i^{-1/2}(V_{0i} - v_i)^{-1/2}, \quad 0 \leqslant v_i \leqslant V_{0i}. \quad (20)$$

Since $V_1, V_2, \ldots V_m$ are independent, their joint probability density function $g(v_1, v_2, \ldots, v_m)$ is given by the product of the individual probability density functions, i.e.

$$g(v_1, v_2, \ldots, v_m) = \frac{1}{\pi^m}\left[\prod_{i=1}^{m}\frac{1}{I_0(\kappa_i)}\right] \exp\left[\sum_{i=1}^{m}\kappa_i\left(1 - \frac{2v_i}{V_{0i}}\right)\right] \times \left[\prod_{i=1}^{m} v_i^{-1/2}(V_{0i} - v_i)^{-1/2}\right], \quad 0 \leqslant v_i \leqslant V_{0i}, \quad i = 1, 2, \ldots, m. \quad (21)$$

Let

$$W_i = \sum_{j=1}^{i} V_j, \quad i = 1, 2, \ldots, m. \quad (22)$$

We note that $W_m = V$ is the total potential energy of the configuration. The Jacobian of the transformation is 1. Therefore, the joint probability density function of $W_1, W_2, \ldots, W_m$ is given by

$$h(w_1, w_2, \ldots, w_m) = \frac{1}{\pi^m}\left[\prod_{i=1}^{m}\frac{1}{I_0(\kappa_i)}\right] \times \exp\left[\left(\sum_{i=1}^{m}\kappa_i\right) - \beta w_m\right] \times \prod_{i=1}^{m}(w_i - w_{i-1})^{-1/2} \times (V_{0i} - w_i + w_{i-1})^{-1/2}, \quad (23)$$

where $w_{i-1} \leqslant w_i \leqslant \min(w_{i+1}, V_{0i}^*)$, $i = 1, 2, \ldots, m$ with $w_0 = 0$, $w_{m+1} = \infty$ and $V_{0i}^* = \sum_{j=1}^{i} V_{0j}$, $i = 1, 2, \ldots, m$.

Thus, for $m \geqslant 2$ the probability density function of $V$ is given by

$$\check{h}(v) = \frac{1}{\pi^m}\left[\prod_{i=1}^{m}\frac{1}{I_0(\kappa_i)}\right] \exp\left[\left(\sum_{i=1}^{m}\kappa_i\right) - \beta v\right] \times \int_{\max(0, v - V_{0m})}^{\min(v, V_{0m-1}^*)} \mathrm{d}w_{m-1}(v - w_{m-1})^{-1/2} \times (V_{0m} - v + w_{m-1})^{-1/2} \times \left\{\left[\prod_{i=1}^{m-2}\int_{\max(0, w_{m-i} - V_{0m-i})}^{\min(w_{m-i}, V_{0m-1-i}^*)} \mathrm{d}w_{m-1-i} \times (w_{m-i} - w_{m-1-i})^{-1/2} \times (V_{0m-i} - w_{m-i} + w_{m-1-i})^{-1/2}\right] \times w_1^{-1/2}(V_{01} - w_1)^{-1/2}\right\}, \quad (24)$$

where $0 \leqslant v \leqslant V_{0m}^*$. Here $\prod_{i=1}^{l} a_i \equiv 1$, when $l < 1$. The function $\check{h}(v)$ can not be expressed explicitly. However, its values and its plot can be obtained numerically for given values of $m$, $\kappa_i$'s and $V_{0i}$'s.

If $m$ is very large, according to the central limit theorem, the distribution of $V$ will take an approximate Gaussian shape. If $m$ is not large but all $\kappa_i$ are large, then a gamma distribution will approximate $V$. However, if $m$ is not large and all $\kappa_i$ are not large, $V$ will be distributed differently from either a Gaussian or gamma distribution. In the next section we show that for $m = 1$, the probability density function of $V$ has a bathtub shape for small values of the concentration parameter.

## 4. Probability distribution of torsional energy of methanol

Let $\Phi$ denote the internal torsional angle of the methanol molecule. We consider modelling $\Phi$ by a three-mode von Mises distribution. Let $\Phi$ have a three-mode von Mises distribution with the probability density function

$$f(\phi) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos[3(\phi - \phi_0)]\}, \quad -\pi \leqslant \phi \leqslant \pi,$$

where $\kappa > 0$ and $\phi_0 = -2\pi/3$.

From (16), the configurational torsional entropy of methanol is given by

$$S_c = k_B\left(\ln(2\pi) + \ln[I_0(\kappa)] - \kappa\frac{I_1(\kappa)}{I_0(\kappa)}\right). \quad (25)$$

Figure 4 gives a plot of $S_c/k_B$ as a function of $\kappa$. It is analogous to an $XY$ cross-section of a three-dimensional surface produced by Ercolani [8] from the revised data of Pitzer and Gwinn [27, 28]. However, in our curve the

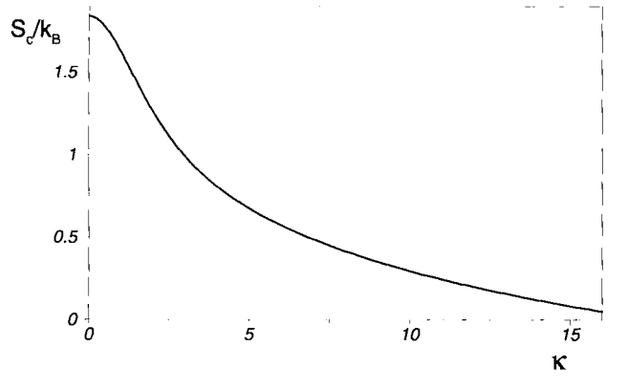


Figure 4. Plot of $S_c/k_B$ as a function of $\kappa$.

kinetic energy component is not counted, i.e. the curve represents only the configurational entropy of the molecule. Equation (25) can also be verified from the first principles, but generally that would be a less productive approach.

The probability density function of $\Phi$ can be written as

$$f(\phi) = \frac{\exp(\kappa)}{2\pi I_0(\kappa)} \exp(-\kappa\{1 - \cos[3(\phi - \phi_0)]\}) = \frac{\exp(\kappa)}{2\pi I_0(\kappa)} \exp(-\beta v), \quad (26)$$

where

$$v = \frac{\kappa}{\beta}\{1 - \cos[3(\phi - \phi_0)]\}.$$

Thus, by analogy with (1),

$$V = \frac{\kappa}{\beta}\{1 - \cos[3(\Phi - \phi_0)]\} \quad (27)$$

is the torsional energy of the molecule corresponding to the torsional angle $\Phi$. Let $V_0$ be the maximum torsional energy, then from (27) we have

$$V_0 = 2\kappa/\beta, \quad (28)$$

and thus

$$V = (V_0/2)\{1 - \cos[3(\Phi - \phi_0)]\} = (V_0/2)[1 + \cos(3\Phi - \phi_1)], \quad (29)$$

where $\phi_1 = 3\phi_0 - \pi$. We note that the above expression for the torsional energy agrees with the expression for the torsional potential energy of a typical classical molecular mechanical force field [29].

Using (20), the probability density function of the torsional energy $V$ is given by

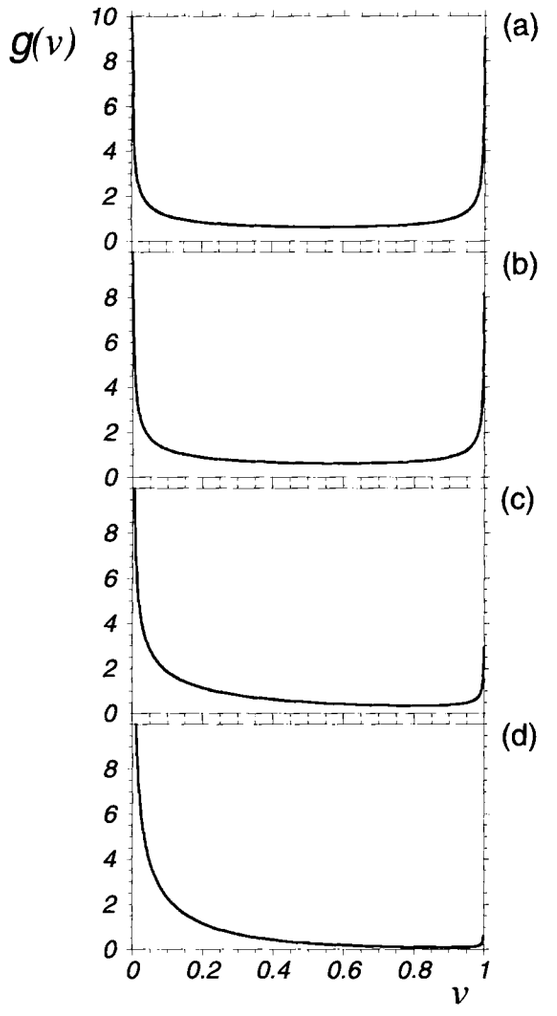


Figure 5. Plots of probability density function $g(v)$ of the torsional energy for $\kappa$ (*a*) 0.10, (*b*) 0.20, (*c*) 1.0 and (*d*) 2.0, when $V_0 = 1$.

$$g(v) = \frac{1}{\pi I_0(\kappa)} \exp\left[\kappa\left(1 - \frac{2v}{V_0}\right)\right] v^{-1/2}(V_0 - v)^{-1/2}, \quad 0 \leqslant v \leqslant V_0. \quad (30)$$

In figure 5(*a*)–(*d*), we give plots of $g(v)$ for $\kappa = 0.10$, 0.20, 1.0 and 2.0 with $V_0 = 1$. We observe a bathtub shape for the probability density function of $V$ for the values of $\kappa$ that are not too large. Since $2\kappa/\beta = V_0$ is fixed, the value of $\kappa$ is small at high temperatures.

Consider the relative potential energy $W = V/V_0$. Its probability density function is given by

$$b(w) = \frac{1}{\pi I_0(\kappa)} \exp[\kappa(1 - 2w)] w^{-1/2}(1 - w)^{-1/2}, \quad 0 \leqslant w \leqslant 1. \quad (31)$$

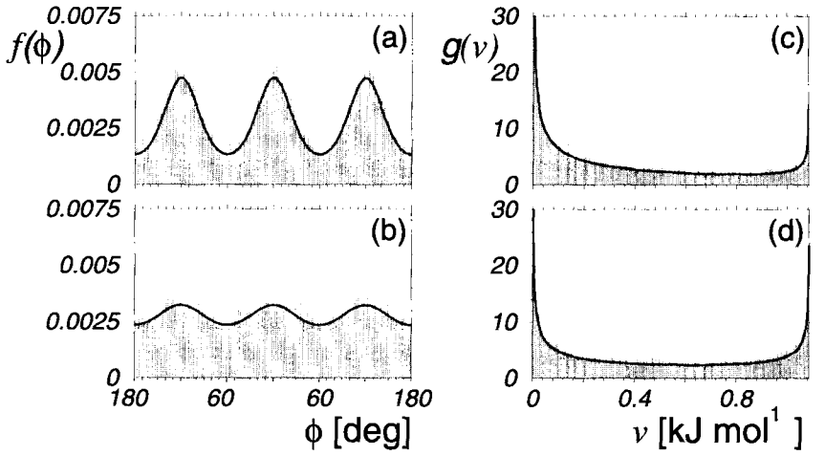


Figure 6. Histograms of torsion angles and energy of methanol at two temperatures: (*a*) angles at 298 K and (*b*) 1000 K and (*c*) energy at 298 K and (*d*) 1000 K. Fitted lines are bold.

For $\kappa = 0$,

$$b(w) = \frac{1}{\pi} w^{1/2-1}(1-w)^{1/2-1}, \quad 0 \leqslant w \leqslant 1,$$

which is the probability density function of a beta distribution with parameters 1/2 and 1/2.

Distribution (31) is bathtub shaped with the modes at 0 and 1. The reason for the singularities at $w = 0$ and at $w = 1$ is that for these values of $w$, the potential energy is at extremum. Thus the peculiar shape of the potential energy distribution is due to the fact that the potential energy has several extrema, and the potential energy is a function of only one variable. Effects of this kind, so called van Hove singularities, are well known from solid-state physics [30]. In our case the peaks correspond to similar singularities but for a one-dimensional density of states.

Now we derive the maximum likelihood estimator for the parameter $\kappa$ of the probability density function of $V$ from independent observations on $V$. Let $v^{[1]}, v^{[2]}, \ldots, v^{[n]}$ be a random sample of the torsional energy terms. Then the likelihood function is given by

$$L(\kappa) = \pi^{-n}(I_0(\kappa))^{-n} \exp\left(n\kappa - \frac{2\kappa}{V_0}\sum_{i=1}^{n} v^{[i]}\right)$$

$$\times \prod_{i=1}^{n} (v^{[i]})^{-1/2}(V_0 - v^{[i]})^{-1/2}.$$

Thus, the log-likelihood is given by

$$\ln L(\kappa) = -n \ln(\pi) - n \ln[I_0(\kappa)] + n\kappa - \frac{2\kappa}{V_0}\sum_{i=1}^{n} v^{[i]}$$

$$-\frac{1}{2}\sum_{i=1}^{n} \ln[v^{[i]}(V_0 - v^{[i]})].$$

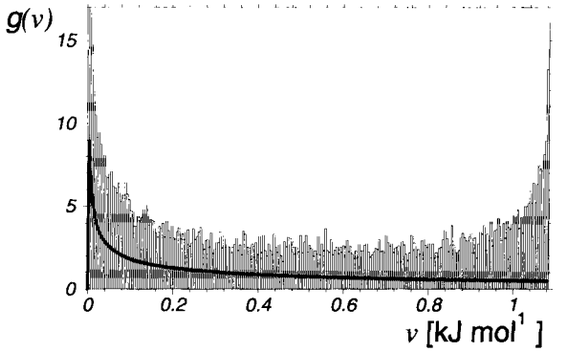


Figure 7. Fit (bold line) of the torsion energy data for methanol using the probability density function of $V$ based on the traditional quasi-harmonic approach, refer to equation (5).

The derivative of the above equation equal to 0 gives

$$-n\frac{I_1(\kappa)}{I_0(\kappa)} + n - \frac{2}{V_0}\sum_{i=1}^{n} v^{[i]} = 0.$$

Thus, the maximum likelihood estimate of $\kappa$ is given by the solution of the equation

$$A(\kappa) = 1 - 2\bar{v}/V_0, \tag{32}$$

where $\bar{v} = \sum_{i=1}^{n} v^{[i]}/n$ is the mean observed potential energy and $A(\kappa) = I_1(\kappa)/I_0(\kappa)$.

We carried out computational experiments in which we collected the torsional angle and corresponding torsional energy term of methanol. In each experiment 15 000 configurations of methanol molecules were sampled in an *NVT* ensemble. In figure 6 we report results of a typical simulation performed at 298 and 1000 K using AMBER-style potentials. Molecular dynamics simulations were carried out using

CDISCOVER from Molecular Simulations, Inc. The CVFF/NOMORSE/NOCROSS molecular mechanical force field [31] was used.

In order to fit a three-mode von Mises distribution to observations of the angle, we obtained maximum likelihood estimates of $\phi_0$ and $\kappa$ from the angular data, using (8)–(10). At the two temperatures, $\hat{\phi}_0$ is $-119.5°$ and $-119.8°$, respectively, i.e. the values of $\hat{\phi}_0$ are close to $-2\pi/3$ as assumed in the model. Corresponding estimates of $\hat{\kappa}$ are 0.633 at 298 K and 0.159 at 1000 K, respectively. We also obtained maximum likelihood estimates of $\kappa$ from the torsional energy data using (32), and they turned out to be identical to the fourth digit to those from the angular data. The three-mode von Mises distributions fitted to the torsional angle data and the bathtub distributions fitted to the energy data are shown in figure 6. They demonstrate that our approach results in excellent fits confirmed by the $\chi^2$ goodness-of-fit test. To appreciate the difference, in figure 7 we show the same histogram of the torsional energy data at 1000 K but with the best possible fit achieved by the traditional quasi-harmonic approximation.

Using (11) the 95% confidence interval for $\kappa$ at 1000 K is given by $0.159 \pm 0.022$. Thus, from (25) the 95% confidence interval for the configurational component of the classical torsional entropy of the methanol molecule at 1000 K is given by $1.831 \pm 0.002 k_B$. Torsional vibrations at very high temperatures are reasonably approximated by the free rotator model which puts the upper limit on our estimates. For methanol, an ideal asymptotic free-rigid-rotator estimate, $1.837 k_B$ is close to, but slightly above, the derived number. The classical configurational torsional entropy of methanol at room temperatures calculated from the simulated data is $1.744 \pm 0.004 k_B$.

These results suggest a high efficacy of the new theoretical approach. They open a way to the comprehensive statistical mechanical treatment of complex polyatomic molecular systems susceptible to conformational flexibility.

## 5. Discussion

The objective of this paper is to introduce a new theoretical approach to the statistical mechanics of polyatomic molecules. Essentially, it formulates the statistical mechanics of polyatomic molecules in terms of one-dimensional Euler angles, since the other degrees of freedom usually could be rationally factored out. Therefore, the proposed approach relies on circular, rather than linear statistics [25]. Unlike other authors, who start from the first-principles ideas of Pitzer [27, 32] and his contemporaries, we formulate the problem in terms of statistics obtained by means of numeric simulations.

The laws of circular statistics, applied to angular molecular data, allowed us to formulate classical-approximation relations among fundamental statistical mechanical quantities, so that they resemble and asymptotically converge in the limit of small fluctuations to corresponding expressions for conventional linear-space statistical thermodynamics. For instance, equation (16) is a generalization of equation (3). If fluctuations are small, $I_0(\kappa_i) \approx \exp(\kappa_i)/(2\pi\kappa_i)^{1/2}$ and $A(\kappa_i) = I_1(\kappa_i)/I_0(\kappa_i) \approx 1 - 1/2\kappa_i$, (16) reduces to (3) with $\sigma_i = 1/\kappa_i^{1/2}$. In the case of small fluctuations in methanol, where there is only one rotational degree of freedom, both (3) and (5) would approximate (25). However, they inevitably fail when fluctuations are large. For instance, applied to our data simulated at 298 and 1000 K, the Karplus–Kushick equation results in entropy estimates of $2.034 \pm 0.008 k_B$ and $2.018 \pm 0.008 k_B$, respectively. Both are clearly incongruous numbers exceeding the classical free-rigid-rotator limit, $k_B \ln(2\pi)$. For compact conformations of macromolecules, in which motions of buried groups are presumably confined, an average per-bond error of the Karplus–Kushick equation is believed to decrease [14, 15]. Nevertheless, for the surface groups, which are most important, utility of the Karplus–Kushick equation remains uncertain.

For simplicity, the kinetic energy part of the Hamiltonian was omitted in our derivations, because in the classical rigid rotator approximation it is independent of coordinates [33]. Therefore, equations (16) and (25) count for the *configurational part* of internal entropy only, i.e. they are applicable for assessment of only relative entropies. However, even in the general definition the absolute classical entropy is undefined [11].

The absolute entropy of hindered rotation is obtained quantum mechanically. For instance, using the rescaling approximation of Pitzer and Gwinn [27], equation (16) transforms to

$$S = 0.42 k_B m + k_B \sum_{i=1}^{m} \left\{ \ln [\kappa_i^{1/2} I_0(\kappa_i)] - \kappa_i \frac{I_1(\kappa_i)}{I_0(\kappa_i)} \right\} + \sum_{i=1}^{m} S_i^{HO}, \quad (33)$$

where $S_i^{HO}$ is the entropy of a harmonic oscillator conjugated to the $i$th vibrational mode of rotation. In particular, equation (33) should be used in simulations of grand canonical ensembles and in applications to adiabatic processes.

The Pitzer–Gwinn approximation has been tested by a number of authors [4–6, 27] and appeared to be very accurate over the entire domain of chemically meaningful parameters. In other words, equation (33)

is appropriate in a wide range of physical chemical problems. It signifies the difference between the quasi-harmonic analysis and our approach. Indeed, the last term in the equation may be interpreted as a contribution to the entropy of hindered rotation from the statistically fitted imaginary 'harmonic modes', whereas the second term takes care of 'anharmonicity'. The last term alone obviously does not provide a full description for large-scale internal rotational motions. Therefore, the quasi-harmonic approximation bears a major caveat when applied outside of the scopes of solid state physics. Remarkably, the quasi-harmonic and normal mode analysis [34] has often been engaged in calculations on flexible molecules, e.g. in [14–21] and others, perhaps literally depicting Schrödinger's definition of a protein as an 'aperiodic crystal' [35]. However, his perception is not quite true. Proteins undergo large-scale conformational fluctuations [36–41], and their surface is conformationally labile [7]. Further theoretical advances hopefully will properly capture the physics of these processes.

Estimators of parameters in the new relations were obtained using the maximum likelihood approach. We also derived a theoretical probability distribution of the potential energy of the system and tested the results by using molecular dynamics simulations. Molecular dynamics tend to support our formulations. We chose the methanol molecule as a case study, because of the low internal hindered rotation barrier [42]. The bathtub-shaped probability density function for the torsional potential energy of methanol, which we obtained in section 4, is a new addendum to probability distributions. It signifies an unusual contra-intuitive property of the potential energy of hindered rotation when high energy states are more populated than the lower ones. The paradox can be understood in terms of the confined configurational space of the molecule, expressed by the internal Euler angle, and the specific shape of the potential energy function with 'saddle points' [30]. Overall, our formulations and numerical estimates agree with the classical-approximation results published in the literature [4, 6, 8, 27, 28].

Our approach is most pertinent to liquid simulations. In liquid, as compared to the gas phase, the intramolecular potential of a solvated molecule is significantly biased by non-bonded intermolecular interactions so that comprehensive first-principles calculations often are impractical. Even a computationally affordable empirical force-field approach, e.g. à la one recently exercised by Harvard researchers [7] (and also [8]), combined with an external potential of mean force [3] is unlikely to provide an unambiguous solution because of the complexity of the dielectric response of solvated molecules [43–45]. Therefore, extensive computer simulations with explicitly represented particles of the solvent, perhaps will continue to be the only reliable approach to quantitative statistical thermodynamics in the liquid phase. The net effect of the liquid matrix on intramolecular potentials in the imbedded molecule is usually associated with smoothing of the adiabatic gas-phase potential energy surface of the molecule [46]. Our equations provide an elegant way to quantify effects of the environment on the hindered rotation of a solvated molecule. This is achieved by using a sampled configurational ensemble alone. Depending on the geometry optimization procedure, the full gas-phase rotational barrier of methanol, as given by CDISCOVER, is 3.26–4.10 $kJ\,K^{-1}\,mol^{-1}$. From the angular data of liquid simulations performed at 298 and 1000 K, using (28), we obtain 3.16 and 2.64 $kJ\,K^{-1}\,mol^{-1}$, respectively. When the supercritical fluid was allowed to evaporate, the rotational barrier increased as described in the literature.

Classical computer simulation techniques are an indispensable tool in modern molecular science. They are especially useful in applications to the aggregated state and large macromolecular systems [2, 12, 18–20]. For them even simplified quantum-mechanical calculations are unfeasible. Not surprisingly, these complex statistical-mechanical systems also pose a formidable challenge to our methodology. So far, we approached the case of complex polyatomic molecules following a recipe of Pitzer and Gwinn [27], i.e. by modelling the angular probability density function as a collection of independent von Mises distributions. In other words, our expressions should be valid, where the internal rotors are far apart and coupling interactions between them are negligible. For instance, this approximation should reasonably represent properties of hydroquinone and *para*-xylene but probably not of *o*-cresol and *ortho*-xylene, or butane. However, characteristic coordinates of complex polyatomic molecules are not necessarily independent. In fact, the opposite is fairly common, especially in macromolecules. Therefore, a realistic computational approach to entropy in curvilinear coordinates would require theoretical development of the multivariate von Mises distribution.

We are grateful to Dan Sharp, James Harner and Sidney Soderholm for continuous support of the project; to Vladimir Hnizdo, Doug Landsittel and Molly Pickett-Harner for a thorough review of the manuscript; to Lingyi Zheng for help with preparation of the data; to Tina Yeh and Jodi Shaulsky for help with CDISCOVER; and to the anonymous reviewer of the manuscript for the insightful comments.

HS was partially supported by the Biostatistics Branch of NIOSH Health Effects Laboratory Division, #HELD99B0030.

## References

[1] ALLEN, M. P., and TILDESLEY, D. J., 1989, *Computer Simulation of Liquids* (Oxford: Clarendon Press).
[2] BEVERIDGE, D. L., and DICAPUA, F. M., 1989, *Computer Simulation of Biomolecular Systems*, edited by W. F. van Gunsteren and P. K. Weiner (Leiden: ESCOM), pp. 1–25.
[3] DEMCHUK, E., BASHFORD, D., GIPPERT, G. P., and CASE, D. A., 1997, *J. molec. Biol.*, **270,** 305.
[4] EAST, A. L. L., and RADOM, L., 1997, *J. chem. Phys.*, **106,** 6655.
[5] KNYAZEV, V. D., 1998, *J. phys. Chem.* A, **102,** 3916.
[6] AYALA, P. Y., and SCHLEGEL, H. B., 1998, *J. chem. Phys.*, **108,** 2314.
[7] MAMMEN, M., SHAKHNOVICH, E. I., and WHITESIDES, G. M., 1998, *J. org. Chem.*, **63,** 3168.
[8] ERCOLANI, G., 1999, *J. org. Chem.*, **64,** 3350.
[9] ECKART, C., 1935, *Phys. Rev.*, **47,** 552.
[10] CICCOTTI, G., and RYCKAERT, J. P., 1986, *Comp. Phys. Rep.*, **4,** 345.
[11] LANDAU, L. D., and LIFSHITZ, E. M., 1980, *Statistical Physics*, Part 1, 3rd Edn (Oxford: Pergamon).
[12] MEIROVITCH, H., 1998, *Reviews in Computational Chemistry*, Vol. 12, edited by K. B. Lipkowitz and D. B. Boyd (New York: Wiley), pp. 1–74.
[13] AMADEI, A., APOL, M. E. F., NOLA, A. D., and BERENDSEN, H. J. C., 1996, *J. chem. Phys.*, **104,** 1560.
[14] KARPLUS, M., and KUSHICK, J. N., 1981, *Macromolecules*, **14,** 325.
[15] LEVY, R. M., KARPLUS, M., KUSHICK, J., and PERAHIA, D., 1984, *Macromolecules*, **17,** 1370.
[16] TIDOR, B., and KARPLUS, M., 1993, *Proteins*, **15,** 71.
[17] BRADY, G. P., and SHARP, K. A., 1997, *Curr. Opin. Struct. Biol.*, **7,** 215.
[18] CHEATHAM, T. E., SRINIVASAN, J., CASE, D. A., and KOLLMAN, P. A., 1998, *J. biomolec. Struct. Dynam.*, **16,** 265.
[19] SRINIVASAN, J., MILLER, J., KOLLMAN, P. A., and CASE, D. A., 1998, *J. biomolec. Struct. Dynam.*, **16,** 671.
[20] VOROBJEV, Y. N., and HERMANS, J., 1999, *Biophys. Chem.*, **78,** 195.
[21] WRABL, J. O., SHORTLE, D., and WOOLF, T. B., 2000, *Proteins*, **38,** 123.
[22] MAIER, W., and SAUPE, A., 1959, *Z. Naturforsch.*, **14A,** 882.
[23] LIPARI, G., and SZABO, A., 1982, *J. Am. chem. Soc.*, **104,** 4546.
[24] YANG, D., and KAY, L. E., 1996, *J. molec. Biol.*, **263,** 369.
[25] MARDIA, K. V., and JUPP, P. E., 1999, *Directional Statistics* (Chichester: Wiley).
[26] VON MISES, R., 1918, *Physikal. Z.*, **19,** 490.
[27] PITZER, K. S., and GWINN, W. D., 1942, *J. chem. Phys.*, **10,** 428.
[28] LEWIS, N., and RANDALL, M., 1961, *Thermodynamics*, 2nd Edn, revised by K. S. Pitzer and L. Brewer (New York: McGraw-Hill).
[29] MACKERELL JR, A. D., 1998, *Encyclopedia of Computational Chemistry*, edited by P. V. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III and P. R. Schreiner (Chichester: Wiley), pp. 2191–2200.
[30] VAN HOVE, L., 1953, *Phys. Rev.*, **89,** 1189.
[31] DAUBER-OSGUTHORPE, P., ROBERTS, V. A., OSGUTHORPE, D. J., WOLFF, J., GENEST, M., and HAGLER, A. T., 1988, *Proteins*, **4,** 31.
[32] KEMP, J. D., and PITZER, K. S., 1937, *J. Am. chem. Soc.*, **59,** 276.
[33] LANDAU, L. D., and LIFSHITZ, E. M., 1976, *Mechanics*, 3rd Edn (Oxford: Pergamon).
[34] BROOKS, B. R., JANEŽIČ, D., and KARPLUS, M., 1995, *J. comp. Chem.*, **16,** 1522.
[35] SCHRÖDINGER, E., 1944, *What is Life? The Physical Aspect of the Living Cell* (Cambridge: Cambridge University Press).
[36] ERIKKSON, A. E., BAASE, W., WOZNIAK, J. A., and MATTHEWS, B. W., 1992, *Nature*, **355,** 371.
[37] BAI, Y., SOSNICK, T. R., MAYNE, L., and ENGLANDER, S. W., 1995, *Science*, **269,** 192.
[38] HOLLIEN, J., and MARQUSEE, S., 1999, *Proc. Nat. Acad. Sci.* USA, **96,** 13 674.
[39] HAYWARD, S., KITAO, A., and GO, N., 1995, *Proteins*, **23,** 177.
[40] KITAO, A., HAYWARD, S., and GO, N., 1998, *Proteins*, **33,** 496.
[41] STILL, W. C., GUARNIERI, F., and SENDEROWITZ, H., 1998, Rapidly convergent method for Boltzmann-weighted ensemble generation in free energy simulations, US Patent #574 007 2.
[42] MCQUARRIE, D. A., 1976, *Statistical Mechanics* (New York: Harper Collins).
[43] SIMONSON, T., and PERAHIA, D., 1995, *Proc. Nat. Acad. Sci.* USA, **92,** 1082.
[44] DEMCHUK, E., and WADE, R. C., 1996, *J. phys. Chem.*, **100,** 17 373
[45] JUFFER, A. H., ARGOS, P., and VOGEL, H. J., 1997, *J. phys. Chem.* B, **101,** 7664.
[46] LEACH, A. R., 1996, *Molecular Modeling: Principles and Applications* (Singapore: Addison Wesley).