Determinant of the Inertial Tensor and Rotational Entropy of Random Polymers

Y. Bruce Yu*

The Protein Engineering Network of Centres of Excellence, University of Alberta, Edmonton, AB T6G 2S2, Canada

Weizhen Wang

Department of Mathematics and Statistics, Wright State University, Dayton, Ohio 45435

Received: December 10, 1998; In Final Form: March 19, 1999

The determinant of the inertial tensor of the freely jointed chains is investigated using statistical simulation. It is found that for the straight chain, equivalent to a Markov chain, the determinant of the inertial tensor follows log-normal distribution at large chain length ($N \geq 20$). The ensemble average of this determinant scales with chain length N as $N^{6.0}$. For the symmetrically cross-linked chain (two single chains of identical length with a link in between located at identical positions), it is shown that as the link position moves to the center of the chain, the average of the determinant of the inertial tensor decreases. This result is consistent with the previous experimental observation that the stabilizing effect of inerchain disulfide bond on a dimeric α-helical coiled-coil becomes greater as it moves toward the center of the chain. The average of the determinant of the inertial tensor are then used to calculate the reduction in rotational entropy caused by cross-linking two random polymers, using ideal gas statistics. The values obtained are then compared with experimental data on entropy reduction caused by cross-linking unfolded polypeptide chains in aqueous solutions. The results support the conclusion drawn previously from experimental work that ideal gas statistics grossly overestimate the translational and rotational entropy of macromolecules in aqueous solutions.

Introduction

It is of tremendous interest to give experimental data on protein stability and protein−protein/DNA interaction a molecular interpretation based on the structures.[1−3] This requires that contributions from the overall translational and rotational motions be subtracted from the experimental data so that the resultant thermodynamic quantities correspond to reactions between molecules with fixed center of mass and orientations. However, the estimation of the overall translational and rotational entropy of macromolecules in aqueous solutions has been quite controversial.[2−13] At the center of the controversy is whether ideal gas statistics is applicable to macromolecules in aqueous solutions. Recent experimental data obtained by comparing the unfolding entropy of dimeric proteins with that of their cross-linked monomeric counterparts indicate that ideal gas statistics grossly overestimate the entropy of overall motions of macromolecules in aqueous solutions.[14,15] However, these experimental data are on the translational and rotational entropy of unfolded polypeptide chains. The equation used by ideal gas statistics to calculate the translational entropy (the Sackur−Tetrode equation)[16] contains only molecular weight (MW) and therefore is indifferent to the structure of the molecule. In contrast, the equation for the rotational entropy contains the product of the principal moments of inertia of the molecule, which depends on the structure of the molecule.[16] Therefore, to make a direct comparison between experimental data and results from ideal gas statistics, it is necessary to estimate the product of the principal moments of inertia of the unfolded polypeptide chain. Of particular interest is how the rotational entropy is affected by cross-linking two shorter chains to form a longer chain. This is the objective of the current work.

Models and Methods

The model employed here to mimic the random polymer is the so-called freely jointed chain.[17] In this model, the random polymer is made of N mass points of identical mass m, linked by massless rigid bonds all of identical length l. The orientation of each bond is independent of the other bonds and is not in favor of any particular direction in $\mathbf{R}^3$. Therefore, each bond follows uniform distribution on a unit sphere. Its inertial tensor is given by the following 3×3 matrix:[18]

$$\hat{A} = \begin{bmatrix} \sum_{i=1}^{N} m(Y_i^2 + Z_i^2) & -\sum_{i=1}^{N} mX_iY_i & -\sum_{i=1}^{N} mZ_iX_i \\ -\sum_{i=1}^{N} mX_iY_i & \sum_{i=1}^{N} m(Z_i^2 + X_i^2) & -\sum_{i=1}^{N} mY_iZ_i \\ -\sum_{i=1}^{N} mZ_iX_i & -\sum_{i=1}^{N} mY_iZ_i & \sum_{i=1}^{N} m(Y_i^2 + Y_i^2) \end{bmatrix} \tag{1}$$

where (X_i, Y_i, Z_i) is the coordinate of the ith unit, with the center of mass of the chain being the origin of the coordinate system. The three principal moments of inertia, λ_1, λ_2, λ_3, are eigen values of $\hat{A}$, and their product is just the determinant of $\hat{A}$, i.e.:

$$\lambda_1 \cdot \lambda_2 \cdot \lambda_3 = \det(\hat{A}) \tag{2}$$

It is obvious that the mass m and the bond length l can be factored out of $\det(\hat{A})$ as $m^3 l^6 \det(A)$, with A being the inertial

* Corresponding author. E-mail: bruce.yu@ualberta.ca. Phone: 780-492-6540. Fax: 780-492-1473.

Inertia and Entropy of Random Polymers

J. Phys. Chem. B, Vol. 103, No. 36, 1999 **7677**

tensor of the simplified freely jointed chain where each point is of unit mass and each bond is of unit length. For a random polymer, the average of the product of principal moments of inertia is given by

$$\langle \lambda_1 \cdot \lambda_2 \cdot \lambda_3 \rangle = m^3 l^6 \langle \det(A) \rangle \tag{3}$$

The simplest case is the straight chain in which the N units of the freely jointed chain are linked consecutively by $N - 1$ bonds. Mathematically, this is equivalent to a Markov chain. In such a case, both the average and fluctuation (standard deviation) of $\det(A)$ are functions of N only. We denote these two functions as $f(N)$ and $g(N)$, respectively.

The following statistical simulation, coded in the statistics software *Gauss*,[19] was performed to obtain $f(N)$ and $g(N)$ for $N = 3-250$.

1. Generate three independent standard normal variables, U, V, W. Then, $(U', V', W') = (U, V, W)/(U^2 + V^2 + W^2)^{1/2}$, the projection onto the unit sphere, is uniformly distributed.

2. Generate N-triples $(U_1', V_1', W_1'), ..., (U_N', V_N', W_N')$ as in step 1. They are points on a unit sphere with the origin of the coordinate system being the center of the sphere.

3. Use the following transformation to obtain (X_i', Y_i', Z_i'), with $i = 1, 2, ..., N$:

$$X_i' = \sum_{j=1}^{i} U_j' \qquad Y_i' = \sum_{j=1}^{i} V_j' \qquad Z_i' = \sum_{j=1}^{i} W_j' \tag{S-1}$$

then (X_i', Y_i', Z_i') is the coordinate of the ith unit. The center of mass of the entire random chain, $(\langle X' \rangle, \langle Y' \rangle, \langle Z' \rangle)$, is given by

$$\langle X' \rangle = \frac{1}{N} \sum_{i=1}^{N} X_i' \qquad \langle Y' \rangle = \frac{1}{N} \sum_{i=1}^{N} Y_i' \qquad \langle Z' \rangle = \frac{1}{N} \sum_{i=1}^{N} Z_i' \tag{S-2}$$

4. Transform (X_i', Y_i', Z_i') to (X_i, Y_i, Z_i), so that (X_i, Y_i, Z_i) is the coordinate of the ith unit with the center of mass being the origin of the coordinate system

$$X_i = X_i' - \langle X' \rangle \qquad Y_i = Y_i' - \langle Y' \rangle \qquad Z_i = Z_i' - \langle Z' \rangle \tag{S-3}$$

5. Calculate $\det(A)$ in eq 1 using the above generated (X_i, Y_i, Z_i)'s, with $m = 1$.

6. Repeat steps 1–5 10^5 times and calculate $f(N)$, the average of $\det(A)$, as

$$f(N) = \frac{1}{10^5} \sum_{p=1}^{10^5} \det(A)_p \tag{S-4}$$

7. The standard deviation, $g(N)$, is calculated as

$$g(N) = \left[\frac{1}{10^5 - 1} \sum_{p=1}^{10^5} (\det(A)_p - f(N))^2 \right]^{1/2} \tag{S-5}$$

The next level in terms of complexity is the symmetrically cross-linked chain, in which two chains each of N units are linked at the kth unit of both chains (the link is itself a bond). The average of the determinant of the inertial tensor in such a case is a function of both N and k. Denote this function as $h(N, k)$. Obviously, $h(N, k) = h(N, N + 1 - k)$ and $h(N, 1) = h(N, N) = f(2N)$. The following modifications are made to steps 2 and 3 of the above procedure to obtain $h(N, k)$.

2'. Generate $2N - 2$ triples (U_i', V_i', W_i') as in step 1 for $i = 1$ to $2N$ except k and $N + k$ ($1 \le k \le N$).

3'. Let $(X_k', Y_k', Z_k') = (0, 0, 0)$ and $(X_{N+k}', Y_{N+k}', Z_{N+k}') = (1, 0, 0)$.

For $i = k - 1, k - 2, ..., 2, 1$, let

$$X_i' = X_k' - \sum_{j=i}^{k-1} U_j' \qquad Y_i' = Y_k' - \sum_{j=i}^{k-1} V_j'$$

$$Z_i' = Z_k' - \sum_{j=i}^{k-1} W_j' \tag{S-6}$$

For $i = k + 1, k + 2, ..., N$, let

$$X_i' = X_k' + \sum_{j=k+1}^{i} U_j' \qquad Y_i' = Y_k' + \sum_{j=k+1}^{i} V_j'$$

$$Z_i' = Z_k' + \sum_{j=k+1}^{i} W_j' \tag{S-7}$$

Then (X_i', Y_i', Z_i') is the coordinate of the ith unit in the first chain.

For $i = N + k - 1, N + k - 2, ..., N + 2, N + 1$, let

$$X_i' = X_{N+k}' - \sum_{j=i}^{N+k-1} U_j' \qquad Y_i' = Y_{N+k}' - \sum_{j=i}^{N+k-1} V_j'$$

$$Z_i' = Z_{N+k}' - \sum_{j=i}^{N+k-1} W_j' \tag{S-8}$$

For $i = N + k + 1, N + k + 2, ..., 2N$, let

$$X_i' = X_{N+k}' + \sum_{j=N+k+1}^{i} U_j' \qquad Y_i' = Y_{N+k}' + \sum_{j=N+k+1}^{i} V_j'$$

$$Z_i' = Z_{N+k}' + \sum_{j=N+k+1}^{i} W_j' \tag{S-9}$$

Then (X_i', Y_i', Z_i') is the coordinate of the ith unit in the second chain.

The center of mass of the entire random chain, $(\langle X' \rangle, \langle Y' \rangle, \langle Z' \rangle)$, is given by

$$\langle X' \rangle = \frac{1}{2N} \sum_{i=1}^{2N} X_i' \qquad \langle Y' \rangle = \frac{1}{2N} \sum_{i=1}^{2N} Y_i' \qquad \langle Z' \rangle = \frac{1}{2N} \sum_{i=1}^{2N} Z_i' \tag{S-10}$$

The same procedure for calculating $f(N)$ is used to calculate $h(N, k)$.

Results and Discussion

Figure 1a plots $\ln f(N)$ against $\ln N$. Clearly, a linear pattern can be seen. All the points fall on a straight line almost perfectly for $N \ge 20$. If one uses the least squares method to fit the points by a linear function, the coefficient of determination, r^2, is 100.0% and the sum of residuals is less than 0.01, which indicate an excellent fit.[20] The fitted line satisfies the following equation:

$$\ln f(N) = -6.8 + 6.0 \ln N \tag{4}$$

which is equivalent to

$$f(N) = 1.11 \times 10^{-3} N^{6.0} \tag{5}$$

This relationship means that the determinant of the inertial tensor

of a freely jointed chain satisfies the scaling law,[21,22] with the critical exponent $\nu = 6.0$. In addition to scale invariance, another feature of systems satisfying the scaling law is large fluctuation. It can be seen from Figure 1b that the ratio of the relative fluctuation, i.e., the coefficient of variation, $g(N)/f(N)$, increases at first as N increases and then stabilizes to a constant value of about 1.6 at large N ($N \geq 80$). Thus, the fluctuation is of the same order as the average as N goes to infinity. The probability density function of the random variable $\ln[\det(A)]$ is well approximated by a Gaussian function at large N, as indicated by the Wilk–Shapiro test[23] (Figure 2). For example, at $N = 40$, a Wilk–Shapiro test was conducted against a Gaussian distribution with mean 14.72 and standard deviation 1.21. The resultant p value is 0.13. Consequently, $\det(A)$ itself follows a log-normal distribution, one of the most commonly encountered distributions for scale invariant processes.[24] More specifically, the probability density function of $\det(A)$ is

$$
p_N(x) = \frac{1}{\left(2\pi \ln\left[\frac{g^2(N)}{f^2(N)} + 1\right]\right)^{1/2}} \frac{1}{x} \times
$$

$$
\exp\left(- \frac{\left\{\ln x - \ln f(N) - 0.5 \ln\left[\frac{g^2(N)}{f^2(N)} + 1\right]\right\}^2}{2 \ln\left[\frac{g^2(N)}{f^2(N)} + 1\right]}\right)
$$

$$
\approx \frac{1}{2.82} \frac{1}{x} \exp\left(- \frac{\{\ln x - 6.0 \ln N + 6.17\}^2}{2.54}\right) \tag{6}
$$

The second equality is obtained by using eq 3 and the result that $g(N)/f(N)$ is about 1.6. This is approximately valid at $N \geq 80$.

Figure 3 plots $\ln h(35, k)$ as a function of k. As can be seen, as the link position moves from the terminal to the center of the chain, the average of the determinant of the inertial tensor decreases, indicating that the cross-linked chain becomes more compact and less random. This leads to greater reduction in rotational entropy of the unfolded state (see below). Consequently, stabilization of the folded conformation by a cross-link will become greater as the link moves to the center of the chain. Experimentally, it has been observed that as an interchain disulfide bond moves from the termini to the center of a two-stranded α-helical coiled-coil with each chain of 35 residues, the coiled-coil becomes more stable provided there is no steric strain.[25] The previous explanation of this positional effect was based on the local stereochemical environment of the disulfide bond in the folded state. The results obtained here indicate that part of the explanation lies in the reduction of the rotational entropy of the unfolded state.

Since no excluded volume effect is considered, the freely jointed chain should be more compact than the actual random polymer.[17] Therefore, the average value calculated here for the determinant of the inertial tensor should be lower than the actual value. On the other hand, this quantity can be calculated for the folded structure without involving any model. Since the folded globular state is more compact than the unfolded state and therefore possesses smaller moments of inertia, the values for the folded structure with the same N offers an objective criterion for the lower limit of such calculations. In other words, the freely jointed chain should not be more compact than the actual folded structure. Here is one such test. Using the m (113 Da) and N values for the insulin monomer ($N = 51$) and dimer
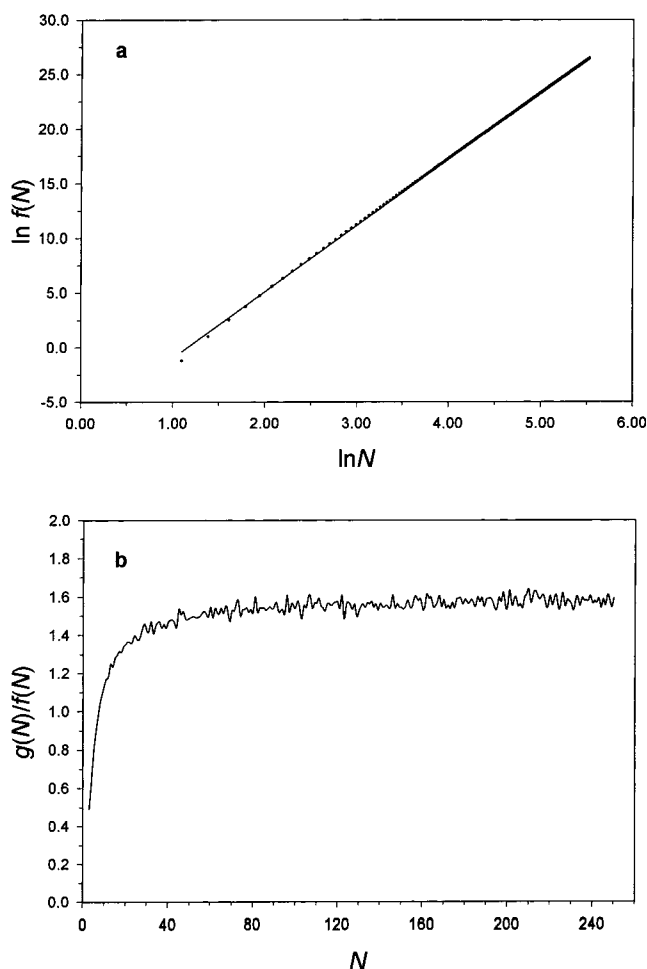

Figure 1. (a) Chain length (N) dependency of the average of the determinant of the inertial tensor ($f(N)$) of a freely jointed chain. The dots are simulation results. The solid line is the fitted line obtained by linear regression. (b) Relationship between the chain length (N) and the relative fluctuation of the determinant of the inertial tensor ($g(N)/f(N)$).
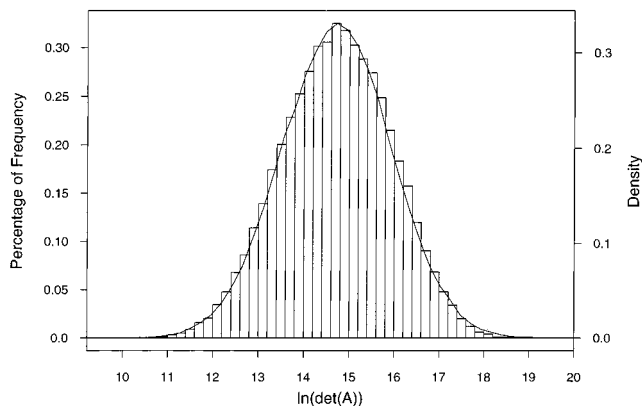

Figure 2. Histogram of the determinant of the inertial tensor after logarithmic transformation for $N = 40$. The solid curve is the density function of a Gaussian distribution with mean 14.72 and standard deviation 1.21.

($N = 102$), the calculated average values of the determinant of the inertial tensor for the unfolded state are 7.00×10^{16} and 4.57×10^{18} $(\text{g·mol}^{-1})^3$ Å^6 for $N = 51$ and $N = 102$, respectively. The values for the folded insulin monomer and dimer are 4.73×10^{16} and 2.15×10^{18} $(\text{g·mol}^{-1})^3$ Å^6, respectively.[26] In both cases, the values calculated here using the freely jointed chain model for the unfolded state are larger

Inertia and Entropy of Random Polymers

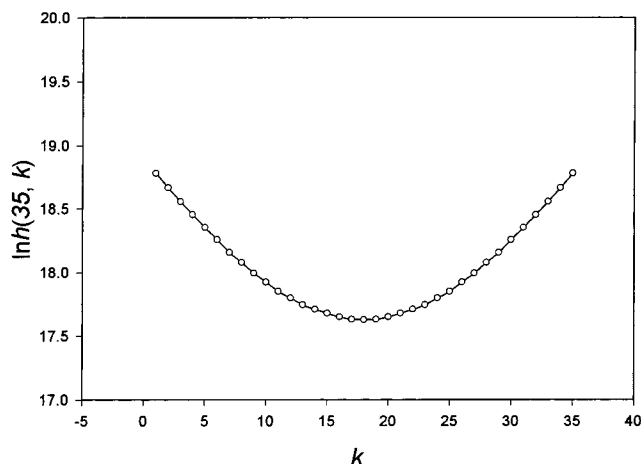*J. Phys. Chem. B, Vol. 103, No. 36, 1999* **7679**

Figure 3. Dependency of the average of the determinant of the inertial tensor of a symmetrically cross-linked chain $h(35, k)$ on the link position (k). 35 is the length of each individual chain.

than the values for actual folded structures with the same N. Satisfying this criterion means that as simplified as it is, the model employed here nonetheless catches the essence of the problem and the values it gives are in the right range.

The classical partition function for the overall rotation of a polyatomic molecule in the ideal gas phase is given by[16]

$$q_r = \pi^{1/2}\left(\frac{8\pi^2 kT}{h^2}\right)^{3/2}(\lambda_1 \cdot \lambda_2 \cdot \lambda_3)^{1/2} = \pi^{1/2}\left(\frac{8\pi^2 kT}{h^2}\right)^{3/2}(\det(\hat{A}))^{1/2} \quad (7)$$

with k being the Boltzmann constant, T the absolute temperature, and h Planck's constant. For a random polymer represented by a freely jointed chain, the vibrational modes are unexcited and q_r is given by

$$q_r = \pi^{1/2}\left(\frac{8\pi^2 kT}{h^2}\right)^{3/2}\langle\lambda_1 \cdot \lambda_2 \cdot \lambda_3\rangle^{1/2} =$$

$$\pi^{1/2}\left(\frac{8\pi^2 ml^2 kT}{h^2}\right)^{3/2}\langle\det(A)\rangle^{1/2} \quad (8)$$

The molar entropy of overall rotation is

$$S_r(N) = R\frac{d}{dT}(T\ln q_r) = \frac{R}{2}\left[3 + \ln\pi + 3\ln\frac{8\pi^2 ml^2 kT}{h^2} + \ln\langle\det(A)\rangle\right] \quad (9)$$

R is the ideal gas constant. For proteins, $m \approx 115$ Da and $l \approx 3.63$ Å.[27] At 300 K, one obtains

$$S_r(N) = 13.44R + 3.0R\ln N \quad (10)$$

for straight chains (Markov chains) and

$$S_r(N, k) = 16.84R + 0.5R\ln h(N, k) \quad (11)$$

for symmetrically cross-linked chains.

From eqs 10 and 11, the entropy reduction due to cross-linking two chains each of length N at identical position k is given as

$$\delta S_r = 2S_r(N) - S_r(N, k) = R(10.0 + 6.0\ln N - 0.5\ln h(N, k)) \quad (12)$$

In the study of Tamura and Privalov,[14] $N = 113$, $k = 83$. The

simulation gives $\ln h(113, 83) = 24.92$. The corresponding rotational entropy reduction is $25.9R$, or about 51 eu (1 eu = 1 cal·K^{-1}·mol^{-1} = 4.184 J·K^{-1}·mol^{-1}). The reduction in translational entropy at 1 M standard state calculated from ideal gas statistics (the Sackur−Tetrode equation) is 45 eu (MW of each single chain is 11 483). In the study of Yu et al.,[15] $N = 36$, $k = 3$, $\ln h(36, 3) = 18.73$, and the rotational entropy reduction is $22.1R$, or 44 eu. The translational entropy reduction is 42 eu (MW of each single chain is 4070). In both cases, the reductions in translational and rotational entropy are about equal and the combined reduction in translational and rotational entropy caused by the cross-link is about 90−100 eu. If the excluded volume effect is considered, the calculated value of reduction will be even larger. On the other hand, the experimentally measured reduction in translational and rotational entropy in both cases is about 5 eu. The calculation here on the rotational entropy of random polymers upholds the conclusion drawn from the experimental work that ideal gas statistics grossly overestimate overall translational and rotational entropy of macromolecules in aqueous solutions.

Note Added in Proof

The quantity being investigated here is the average of the determinant of the inertial tensor of random chains, $\langle\det A\rangle$. It is different from the determinant of the average inertial tensor of random chains, which is $\det\langle A\rangle$. $\langle\det A\rangle = \det\langle A\rangle$ if and only if the two operators, $\langle\,\rangle$ and det, commute. For unbranched freely jointed chains (Markov chains), the average conformation is spherically symmetric and therefore $\langle A\rangle$ can be easily estimated because its off-diagonal terms are all zeros while its diagonal terms all equal $2N\langle s^2\rangle_0/3$, where $\langle s^2\rangle_0$ is the mean-square radius of gyration. Since $\langle s^2\rangle_0 = (N^2 - 1)/6N$ for any N,[17] $\det\langle A\rangle = (N^2 - 1)^3/3^6$ and $\ln\det\langle A\rangle = -6\ln 3 + 3\ln(N^2 - 1)$. Therefore, $\ln\det\langle A\rangle$ is a linear function of $\ln(N^2 - 1)$. On the other hand, the linear relationship between $\ln\langle\det A\rangle$ and $\ln(N^2 - 1)$ is only asymptotic (not valid for small N) according to simulation results. Furthermore, the constant term of $\ln\langle\det A\rangle$ (-6.8) differs from that of $\ln\det\langle A\rangle$ ($-6\ln 3$) by 0.2, which cannot be accounted for by simulation error. This shows that $\langle\det A\rangle \neq \det\langle A\rangle$, and therefore, the two operators, $\langle\,\rangle$ and det, do not commute. This is hardly surprising considering that det is nonlinear. In contrast, the trace operator, Tr, is linear and commutes with $\langle\,\rangle$. Consequently, $\langle\text{Tr } A\rangle = \text{Tr}\langle A\rangle = (N^2 - 1)/3$.

References and Notes

(1) Makhatadze, G. I.; Privalov, P. L. *Adv. Protein Chem.* **1995**, *47*, 307.

(2) Luque, I.; Freire, E. *Methods Enzymol.* **1998**, *295*, 100.

(3) Baker, B. M.; Murphy, K. P. *Methods Enzymol.* **1998**, *295*, 294.

(4) Steiberg, I. Z.; Scheraga, H. A. *J. Biol. Chem.* **1963**, *238*, 172.

(5) Finkelstein, A. V.; Janin, J. *Protein. Eng.* **1989**, *3*, 1.

(6) Janin, J. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 30.

(7) Brady, G. P.; Sharp, K. A. *Curr. Opin. Struct. Biol.* **1997**, *7*, 215.

(8) Murphy, K. P.; Xie, D.; Thompson, K. S.; Amzel, L. M.; Freire, E. *Proteins: Struc. Funct. Genet.* **1994**, *18*, 63.

(9) Holtzer, A. *Biopolymers* **1995**, *35*, 595.

(10) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047.

(11) Hermans, J.; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2707.

(12) Bohm, H.-J. *J. Compu.-Aided Mol. Des.* **1994**, *8*, 243.

(13) Amzel, L. M. *Proteins: Struct., Funct., Genet.* **1997**, *28*, 144.

(14) Tamura, A.; Privalov, P. L. *J. Mol. Biol.* **1997**, *273*, 1048.

(15) Yu, Y. B.; Lavigne, P.; Kay, C. M.; Hodges, R. S.; Privalov, P. L. *J. Phys. Chem. B* **1999**, *103*, 2270.

(16) Mayer, J. E.; Mayer, M. G. *Statistical Mechanics,* 2nd ed; John Wiley & Sons: New York, 1977.

(17) Flory, P. J. *Statistical Mechanics of Chain Molecules*, 2nd ed; Hanser: New York, 1988; pp 1−29.

(18) Goldstein, H. *Classical Mechanics*, 2nd ed.; Addison-Wesley: Reading, MA, 1980; pp 195−203.

(19) *Gauss*, version 3.2.31; Aptech Systems, Inc.: Maple Valley, WA, 1996.

(20) Neter, J.; Wasserman, W.; Kunter, M. H. *Applied Statistical Models*, 3rd ed.; Irwin: Chicago, 1990; pp 47 and 102.

(21) De Gennes, P.-G. *Scaling Concepts in Polymer Physics*; Cornell University Press: Ithaca, NY, 1979.

(22) Grosberg, A. Yu.; Khokhlov, A. R. *Statistical Physics of Macromolecules*; AIP Press: New York, 1994.

(23) Shapiro, S. S.; Wilk, M. B. *Biometrika* **1965**, *52*, 591.

(24) Dubrulle, B.; Breon, F.-M.; Graner, F.; Pocheau, A. *Eur. Phys. J. B* **1998**, *4*, 89.

(25) Zhou, N. E.; Kay, C. M.; Hodges, R. S. *Biochemistry* **1993**, *32*, 3178.

(26) Tidor, B.; Karplus, M. *J. Mol. Biol.* **1994**, *238*, 405.

(27) Creighton, T. E. *Proteins*; Freeman: New York, 1984; p 4.