Energy and Enthalpy Distribution Functions for a Few Physical Systems

K. L. Wu,[†] J. H. Wei,[†] S. K. Lai,*[,†] and Y. Okabe[‡]

Complex Liquids Laboratory, Department of Physics, National Central University, Chungli 320, Taiwan, and Department of Physics, Tokyo Metropolitan University, 1-1 Minami-Osawa, Hachioji, Tokyo 192-0397, Japan

Received: February 27, 2007; In Final Form: April 20, 2007

The present work is devoted to extracting the energy or enthalpy distribution function of a physical system from the moments of the distribution using the maximum entropy method. This distribution theory has the salient traits that it utilizes only the experimental thermodynamic data. The calculated distribution functions provide invaluable insight into the state or phase behavior of the physical systems under study. As concrete evidence, we demonstrate the elegance of the distribution theory by studying first a test case of a two-dimensional six-state Potts model for which simulation results are available for comparison, then the biphasic behavior of the binary alloy Na−K whose excess heat capacity, experimentally observed to fall in a narrow temperature range, has yet to be clarified theoretically, and finally, the thermally induced state behavior of a collection of 16 proteins.

1. Introduction

In statistical mechanics the canonical ensemble theory has played a vital role in its applications to a variety of physical problems. Underlying this ensemble theory is the canonical partition function, which is expressed as[1,2]

$$Q_N = \int \mathrm{d}E\, g(E) \exp(-\beta E)$$

If we write $P(E) = g(E) \exp(-\beta E)/Q_N$, then $P(E)$ is the normalized energy probability distribution function. It can be shown that $P(E)$ displays a Gaussian form, implying that the most probable energy value E^* is $E^* = \langle E \rangle \equiv U$. A well-known example is the ideal gas consisting of N non-interacting particles where $g(E) \propto E^{(3N/2)-1}$ (Figure 3.3 in ref 1). For a real physical system with N interacting particles, one would wonder if the description with the Gaussian distribution remains useful, for under favorable thermodynamic conditions a physical system may undergo phase transition and lead to the phenomenon of phases in coexistence. Indeed, it has been demonstrated in recent computer simulations for the atomic clusters[3,4] and small proteins[5] that $P(E)$ shows a bimodal structure in marked deviation from the Gaussian behavior. The same scenario can be gleaned in an older heat capacity measurement reported by Krier et al.[6] for the binary alloy of composition Na_2K. From their analysis of the excess heat capacity of the Na−K system, the authors inferred that in the temperature range $280 \lesssim T \lesssim 295$ K there exist two coexisting phases, namely, the solid phase sodium and the liquid alloy phase Na_2K. Also, ample experimentally reported heat capacities for a variety of proteins[7] have attributed the cause of denaturation of a protein to the system assuming two states, a native state and an unfolded state coexisting with it. It is thus likely that the non-Gaussian distribution $P(E)$ is a ubiquitous characteristic and may show up in an equilibrium physical system. Unfortunately, the usual canonical ensemble theory does not a priori provide an easy recipe for obtaining a non-Gaussian $P(E)$. It is in fact not clear if the multipeak behavior of $P(E)$ has any relevance to the density of states $g(E)$. Moreover, since $g(E)$ and in most cases the internal energy U are unknown, the multipeak $P(E)$, if it ever exists, has to be determined by some other means. It is the purpose of this work to delve into the functional form of $P(E)$ for a few physical systems that may have undergone phase transition, and we shall proceed by appealing to their experimental heat capacity data.

The present work has been motivated by a series of theoretical studies of the energy or enthalpy distribution functions of the bulk and finite-sized systems by Poland,[8−17] who advanced using the maximum entropy method to find the distribution functions from the moments of the distribution. The main idea is as follows. Supposing that a generalized energy distribution function $P(E)$ exists and that we are given the various moments E_m $(= \int \mathrm{d}E\ E^m P(E))$ by some (independent) means, the maximum entropy method relates the moments $\{E_0, E_1, ..., E_m\}$ to a set of parameters $\{\lambda_0, \lambda_1, ..., \lambda_m\}$ that appear in $P(E) = \exp(\sum_{k=0}^{m} \lambda_k E^k)$. This is a set of coupled nonlinear equations whose solutions, if solvable, would lead to the energy distribution function $P(E) = \exp(\sum_{k=0}^{m} \lambda_k E^k)$. Since $U \equiv E_1$ is generally unknown (but $C_V = (\partial U/\partial T)_V$ is known), one may apply the technique of central moments to redefine the first energy moment $\epsilon_1 = \langle (E - E_1)^1 \rangle$, second moment $\epsilon_2 = \langle (E - E_1)^2 \rangle$, ..., mth moment $\epsilon_m = \langle (E - E_1)^m \rangle$ in which $\langle (E - E_1)^n \rangle = \int \mathrm{d}\epsilon\ \epsilon^n P(\epsilon)$ and arrive at a distribution function $P(\epsilon) = \exp(\sum_{k=0}^{m} \lambda'_k \epsilon^k)$. The fact that E_m is the mth derivative of U with respect to T and that it is generally possible to measure experimentally the heat capacity C_V at various temperatures, the thermal information contained in C_V can be related quite naturally to the moments E_m and hence ϵ_m. Thus, within the context of the maximum entropy method, it is then probable to obtain $P(\epsilon)$ even if U is unknown. In this work, we first test the methodology by applying it to calculate $P(E)$ of the q-state Potts model, then turn to study the enthalphy distribution function $P(\bar{H})$ of the binary alloy Na−K and finally proceed to deduce $P(\bar{H})$ of a class of proteins. Technically, the present work differs from those calculations of Poland[8−17] in that we aim at

* Author to whom correspondance should be addressed. E-mail: sklai@coll.phy.ncu.edu.tw.
[†] National Central University.
[‡] Tokyo Metropolitan University.

10.1021/jp0716280 CCC: $37.00
Published on Web 06/29/2007

deriving more general formulas and working out an iterative numerical procedure that can be used to obtain the energy or enthalpy distribution function at any order of moment.

The organization of the paper is as follows. In the next section, we describe briefly the relationship between the energy (enthalpy) moments and the constant volume (pressure) specific heat whose functional form $C_V(T)$ ($C_P(T)$) is supposedly available from either laboratory or computer experiments. Since considerable mathematical details are available in the literature, we give here essential formulas that relate E_m or ϵ_m to relevant quantities pertaining to C_V. General formulas that are applicable to any order of moments will, however, be given in greater detail. In section 2.2, we introduce the idea of the maximum entropy method,[18] but the content follows mainly the work of Mead and Papanicolaou.[19] We draw attention to the Hausdorff relation,[20] which is here applied numerically to $\{E_0, E_1, ..., E_m\}$. (Readers interested in the numerical aspects may also consult the recent work of Bandyopadhyay et al.[21]) Again a fair amount of technical details have been well documented in refs 19 and 20. Our presentation will thus be in brief. Then, in section 3, we shall apply the distribution theory to three cases of interest. First, we pick up as a test case the two-dimensional six-state Potts model that exhibits the first-order phase transition. Next, we apply the theory to study the binary alloy Na−K and show that our prediction of two coexisting phases is in accordance with the thermodynamic measurement.[6] For the third case, we study the problem of protein denaturation for a total of 16 proteins. It turns out that from merely analyzing the calculated structures of the distribution functions a classification of these proteins into three broad categories can be made. Finally, we summarize, in section 4, our main findings of $P(E)$ or $P(\bar{H})$.

2. Distribution Theory

In this section, we describe as in a recent work of Poland[8] the energy (enthalpy) distribution function in terms of the energy (enthalpy) moments. The connection between the heat capacity and relevant quantities such as E_m or ϵ_m will be given. The iterative scheme used to extract $P(\epsilon)$ or $P(\mathrm{E})$ will be presented also.

2.1. Energy and Enthalpy Distribution Functions. As described in section 1, the energy probability function $P(E)$ for a system lying in the energy range E and $E + \mathrm{d}E$ reads

$$P(E)\,\mathrm{d}E = \frac{g(E)\exp(-\beta E)}{Q_N}\,\mathrm{d}E \tag{1}$$

Making use of $P(E)$, the nth energy moments E_n are

$$E_n = \langle E^n\rangle = \int \mathrm{d}E\, E^n P(E) = \frac{(-1)^n}{Q_N}\,\mathrm{Q}_N^{(n)} \tag{2}$$

where $\mathrm{Q}_N^{(n)} = \partial^n Q_N/\partial\beta^n$. If we carry out a series of derivatives of the internal energy $U \equiv E_1$ with respect to β, then we may express $\partial U/\partial\beta$, $\partial^2 U/\partial\beta^2$, ... in terms of $Q_N^{(n)}$ and Q_N. An iterative procedure that relates $Q_N^{(n)}$ (and hence $\partial^n U/\partial\beta^n \equiv U_\beta^{(n)}$) to E_n is given in Appendix A. Following the procedure set out there, it can be shown easily that

$$E_2 = -U_\beta^{(1)} + {E_1}^2, \quad E_3 = U_\beta^{(2)} - 2{E_1}^3 + 3E_1E_2, \quad ... \tag{3}$$

If $E_1 = U$ is known, then all of the energy moments, in principle, can be calculated using the above formulas. But as mentioned in section 1, U is generally unknown, and we therefore resort to the technique of central moments. In analogy to eq 2 the central moment $\epsilon_m = \langle(E - E_1)^m\rangle$ is defined by

$$\epsilon_m = \langle\epsilon^m\rangle = \int \mathrm{d}\epsilon\, \epsilon^m P(\epsilon) \tag{4}$$

All of the moments ϵ_1, ϵ_2, ... are readily written in terms of $U_\beta^{(n)}$. Knowing (ϵ_1, ϵ_2, ...), it is a simple matter to calculate (E_1, E_2, ...) by employing the recursive relations $E_2 = \epsilon_2 + {E_1}^2$, $E_3 = \epsilon_3 + 3E_2E_1 - 2{E_1}^3$, ... if the first moment $E_1 = U$ is known. Thus, if one has an independent means to obtain ϵ_n, then one can calculate the energy distribution function $P(\epsilon)$ using eq 4 without knowing a priori the internal energy U. A possible way to obtain the energy moments is from the heat capacity data. To this end, let us consider the constant volume heat capacity C_V and suppose that it can be expressed in the form

$$C_V(T) = c_0 + c_1(T - T_\mathrm{r}) + c_2(T - T_\mathrm{r})^2 + c_3(T - T_\mathrm{r})^3 + c_4(T - T_\mathrm{r})^4 + \cdots \tag{5}$$

where T_r is a reference temperature. Since U is a function of T, we may effect a Taylor series expansion on U at T_r

$$U = \sum_n \frac{U_T^{(n)}}{n!}(T - T_\mathrm{r})^n \tag{6}$$

in which $U_T^{(n)} = (\partial^n U/\partial T^n)_{T=T_\mathrm{r}}$. Applying the formula $C_V = (\partial \mathrm{U}/\partial T)_V$ to eq 6 and comparing it with eq 5, a bridge between $U_T^{(n)}$ and the expansion coefficients $\{c_0, c_1, ...\}$ of C_V can be established. It remains to find the explicit formula between T and β to make a direct connection between $U_\beta^{(n)}$ in eq 2 and $U_T^{(n)}$ in eq 6. Appendix A describes the technical details.

The enthalpy distribution function can be derived in the same manner. Here the constant pressure heat capacity C_P data are used instead. Recalling that $H = E + pV$ and in analogy to eq 1, we write the enthalpy distribution function as

$$P(H) = \frac{g(E)\exp(-\beta H)}{\int_V \mathrm{d}V \int \mathrm{d}E\, g(E)\exp(-\beta H)} \tag{7}$$

Note that the partition function is now isobaric. Similarly, we write for the central moments $\bar{H}_m = \langle(H - H_1)^m\rangle$, which is similar in form to eq 4 above; the rest of the derivation parallels those in obtaining the energy distribution function.

2.2. Maximum Entropy Theory. We give a brief account of the maximum entropy method summarizing some essential points. Following Mead and Papanicolaou,[19] we consider moments $\{\mu_0, \mu_1, ..., \mu_n, ...\}$ and define a positive definite distribution function $h(x)$ on the interval (0, 1) by

$$\mu_n = \langle x^n\rangle = \int_0^1 \mathrm{d}x\, x^n h(x) \tag{8}$$

If n runs through infinite values so that $\{\mu_0, \mu_1, ..., \mu_n, ..., \mu_\infty\}$ are all defined, then an exact $h(x)$, in principle, can be determined. In reality only a finite number of moments are available, say, $\{\mu_0, \mu_1, ..., \mu_m\}$. Under this constraint, we can only expect an approximate $h(x)$. It should be stressed that for a given set of moments $\{\mu_1, ..., \mu_m\}$ eq 8 does not necessarily yield a positive and well-behaved function $h(x)$ unless the moments $\{\mu_1, ..., \mu_m\}$ satisfy the Hausdorff relation[20]

$$\sum_{j=0}^{k} (-1)^j \frac{k!}{j!(k-j)!}\mu_{n+j} > 0 \tag{9}$$

Here, for a given moment m, the indices (n, k), $n = 1, ..., m$, which serve to impose a set of relations on μ_n, run from 0 to an integer $(n + k) \leq m$. The solution for $h(x)$ is guaranteed provided that all sets of (n, k) (each of which gives an inequality relation for μ_n) are obeyed. Having checked on the Hausdorff relation, we next consider the entropy functional S expressed in terms of $h(x)$ as

$$S = -\int_0^1 [h(x) \ln h(x) - h(x)] + \sum_{n=0}^{m} \lambda_n \left(\int_0^1 dx\, x^n h(x) - \mu_n\right) \quad (10)$$

where the second term on the right-hand side is the constraint $h(x)$ imposed by eq 8 and is included by the method of Lagrangian multipliers λ.[22] Maximizing S with respect to $h(x)$ yields

$$h(x) = \exp\left[-\left(\sum_{k=0}^{m} \lambda_k x^k\right)\right] \quad (11)$$

Equations 8 and 11, upon normalizing $h(x)$, can now be written

$$\mu_n = \langle x^n \rangle = \frac{\int_0^1 dx\, x^n \exp\left[-\left(\sum_{k=1}^{m} \lambda_k x^k\right)\right]}{\int_0^1 dx \exp\left[-\left(\sum_{k=1}^{m} \lambda_k x^k\right)\right]} \quad n = 1, ..., m \quad (12)$$

which are m nonlinear coupled equations for m unknown λ's. If we now define

$$M(\lambda_1, ..., \lambda_n) = \ln\left\{\int_0^1 \exp\left[-\sum_{j=1}^{n} \lambda_j (x^j - \mu_j)\right] dx\right\} \quad (13)$$

it can be shown that $M(\lambda_1, ..., \lambda_n)$ has an absolute minimum at a unique set of values $\{\lambda_1, ..., \lambda_n\}$, which is equivalent to the requirement that $(\partial M/\partial \lambda_j)_{\lambda_1,...,\lambda_{j-1},\lambda_{j+1},...,\lambda_n} = 0$. Given a finite number of moments $\{\mu_1, ..., \mu_m\}$ that satisfy eq 12, we may proceed by finding an optimized $\{\lambda_1, ..., \lambda_m\}$ at which values M is a minimum. We defer to the following subsection the numerical procedure for solving the optimized $\{\lambda_1, ..., \lambda_m\}$. We checked that the approach to minimize M is more efficient and stable than the iterative procedure of the Hessian matrix[8] especially in dealing with the higher-order moments. For the present case, the energy ϵ ($\bar{H}$) in $P(\epsilon)$ ($P(\bar{H})$) generally does not lie in the range (0, 1). A change in variable is therefore necessary. Poland[8] has suggested a numerical algorithm to handle this kind of variable transformation. We refer interested readers to his paper[8] for further details.

2.3. Numerical Procedure for Finding $M(\lambda_1, ..., \lambda_n)$. As shown in the preceding subsection, the maximum entropy requirement is equivalent to finding the minimum value of $M(\lambda_1, ..., \lambda_n)$ (eq 13 above) given the set of moments $\{\mu_1, ..., \mu_n\}$ that can be determined via eqs 3, 5, and 6 from the coefficients $\{c_0, c_1, ..., c_n\}$ of $C_V(T)$ or $C_P(T)$. These $\{c_0, c_1, ..., c_n\}$ are, in turn, obtainable by a least-squared fitting of eq 5 to the experimental heat capacity data carried out around a chosen reference temperature $T = T_r$. Since we need to decide on the moment n that best represents the distribution function $P(x)$, an expedient sampling of measured data is necessary. To this end, we suggest checking a deviation parameter ξ, which gives the difference between the peak position of the polynomial fitting curve (eq 5 for the chosen n) and that of the experimental C_V or C_P. This is done numerically as follows. For a given moment n, we first located T_r of the observed $C_V(T)$ or $C_P(T)$, next substituted this T_r into eq 5, and then at the moment n desired truncated c_i. Using all of the experimental data, we did a least-squared fitting. Generally it was found that ξ is large in magnitude implying that the agreement is poor especially in proximity to T_r. To rectify ξ, we reduced, with respect to T_r, the experimental data evenly from both sides and performed a least-squared fitting again. This reduction in the experimentally measured data from both sides, checking in each step for ξ, was repeated until $\xi < 0.01$. The moments μ_n (corresponding up to the power of n) will therefore cover a certain range of C_V or C_P data and were thus determined subject to the criterion $\xi < 0.01$. As an illustration, we show in Figure 1 one case study of the protein barnase. Notice that the number of measured data used increases with higher moments.

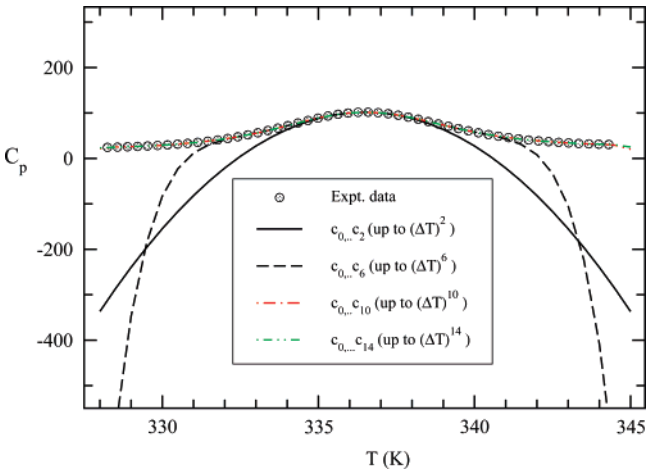

Figure 1. Least-squared fitting of eq 5 at given m moments (corresponding to coefficients up to $c_0, ..., c_m$) to measured data of barnase taken from Makhatadze and Privalov.[7] Notice that the fittings for $m = 10$ (dot−dashed red line) and 14 (dot−dot−dashed green line) are already indistinguishable from experimental data.

We next turn to find the minimum value of $M(\lambda_1, ..., \lambda_n)$. In this work we applied the genetic algorithm[23] within the context of the simplex local optimization technique.[24] We checked that this method works well for the present problem. Assuming that a minimized $M(\lambda_1, ..., \lambda_n)$ is located, we may reconstruct the distribution function $P(x) = \exp(\sum_{j=0}^{n} -\lambda_j x^j)$ and apply the relation $\int_0^1 x^n P(x)\, dx = \mu'_n$ to calculate the approximate moments $\{\mu'_n\}$. These $\{\mu'_n\}$ generally differ slightly from $\{\mu_n\}$ determined iteratively from the experimentally measured data $\{c_0, c_1, ..., c_n\}$. The deviation between $\{\mu'_n\}$ and $\{\mu_n\}$ can be traced to the number of measured data used in fitting eq 5. To be consistent, we now introduce a useful parameter

$$\chi = \left(\sum_{i=1}^{n} |\mu'_i - \mu_i|/\mu_i\right) \times 100\% \quad (14)$$

which was consulted as a criterion for the best representation of $P(x)$. For concreteness, let us analyze a few cases in detail. The numerical data were directed to the protein barnase. Consider first all lower moments $n \leq 4$. The values of χ were calculated and usually found to be less than 0.01%. There occurred, however, numerical difficulties when we proceeded to higher moments $n > 4$. The difficulties lie in the necessity of dealing with large numerical values of $\exp(\lambda_j)$. To circumvent this numerical drawback, we find it is necessary to place a constraint on λ_j, $j = 1, ..., n$, by confining them to fall within the range $-\alpha \leq \lambda_j \leq \alpha$. The question now is: How do we decide on the value of $\alpha > 0$? As it turns out, there is an interesting correlation between χ and α that we may use to fix

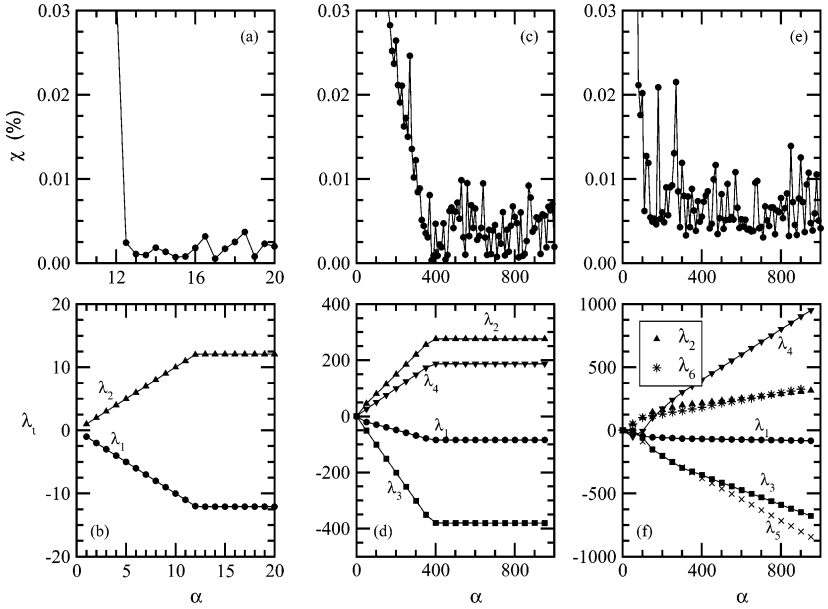

Figure 2. (a) Parameter χ (%) vs integration range α and (b) parameter λ_i appearing in P_m vs α of integration range for the protein barnase. On the left are two moments, and four and six moments are in the middle and on the right, respectively.

α. Take the simple case of the lowest moment $n = 2$. Figure 2a depicts the variation of χ versus α. It can be seen that χ is less than 0.01% if $\alpha > 12$. This indicates that the approximate $\{\mu_n'\}$ converge to "exact" $\{\mu_n\}$ for $\alpha > 12$. Interestingly, the corresponding $\{\lambda_j\}$ also converge (Figure 2b). These results for χ versus α and λ_j versus α allude to a good criterion for choosing α. The same features apply to four moments as can be readily identified in Figures 2c and 2d for the curves χ versus α and λ_j versus α. Obviously, the fluctuation in χ versus α (Figure 2c) for moment $n = 4$ is relatively larger than that for the two moments, but $\chi < 0.01\%$ may still be prescribed unambiguously for $\alpha > 400$. Nonetheless, when comparing λ_j as a function of $\alpha \gtrsim 400$ in Figure 2d, we noticed immediately that $\{\lambda_j\}$ converge to the same values after $\alpha = 400$. Moving on further to six moments, the curve χ versus α (Figure 2e) has a sudden descent to $\chi < 0.01\%$ around $\alpha = 100$. It is noteworthy to point out that $\{\lambda_j\}$ seem not converging (in the sense of two and four moments) after $\alpha = 100$ (Figure 2f). These results for λ_j versus α differ somewhat from those of two and four moments. It would thus be worthwhile to look further into the relation between χ and the distribution function $P(x)$.

Let us first consider four moments for the distribution function $P(x)$ varying for different α (Figure 3a). Here we calculated five different cases in increasing magnitude of α, and for each α, we gave the corresponding value of χ. It is readily seen that as χ becomes less than 0.01% $P(x)$ converges. For six moments, we observed in Figure 3b the same trend as the moment $n = 4$ for $P(x)$. The function $P(x)$ converges after $\alpha \gtrsim 100$, although $\{\lambda_n\}$ (Figure 2f) do not appear to do so. The explanation is that the set of values $\{\lambda_n\}$ just plays the role of variational parameters yielding $M(\lambda_1, ..., \lambda_n)$, a minimum. So long as moments $\{\mu_n'\}$ are close to the "exact" moments $\{\mu_n\}$, a reasonably converged $P(x)$ should be anticipated. For the three physical systems studied here, we found that the energy or enthalpy distribution function calculated up to 12 moments is sufficient, and we have therefore terminated below our calculations up to this order of moment.

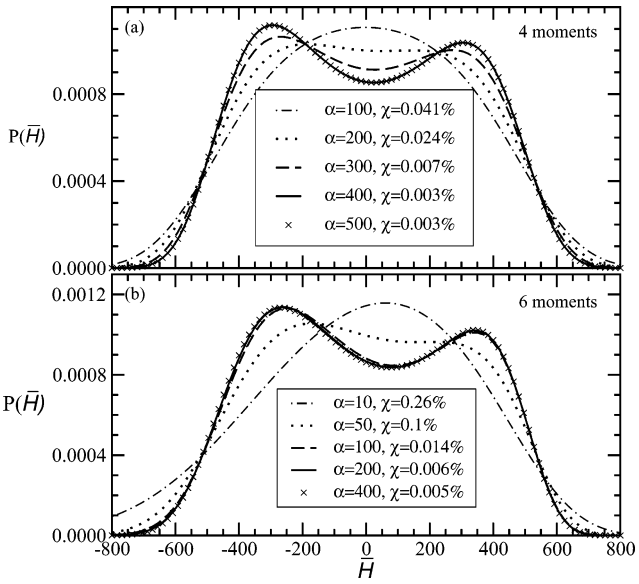

Figure 3. Enthalpy distribution function $P(\bar{H})$ vs enthalpy $\bar{H}$ (in units of kJ/mol) for the protein barnase calculated using the minimization procedure described in the text. Notations for each of the ranges α of λ_i and the corresponding parameter χ are given in the inset. The top figure (a) is for four moments, and the bottom figure (b) for six moments.

3. Numerical Results and Discussion

We consider three cases. First, as a test case, we apply the theory to a statistical mechanical phase transition problem focusing on the q-state Potts model for which both computer-simulated C_V and $P(E)$ are available for comparison, then the binary alloy Na−K for which measured C_P values were reported, and finally 16 proteins where the thermal data for these complex systems have been experimentally reported also.

3.1. *q*-State Potts Model: *P*(*E*). The q-state Potts model is one of the basic models for studying phase transitions.[25] The model is an extension of the Ising model, and its Hamiltonian is given by

$$\underline{H} = -J \sum_{\langle i,j \rangle} \delta_{\sigma_i,\sigma_j} \tag{15}$$

where the Potts variable σ_i takes on the values 1, 2, ..., q, and the summation is taken over the nearest-neighbor pairs of sites

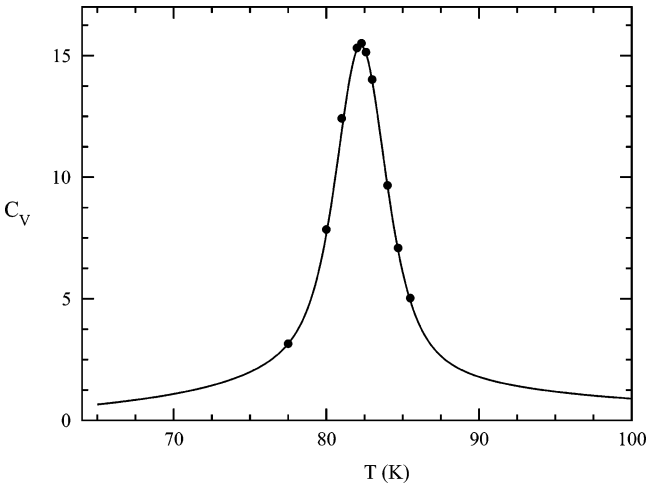

Figure 4. Simulation data (full curve) for the specific heat per spin C_V (in units of J/K) vs T (K) for the statistical Potts model of size 12 × 12. In plotting C_V, the exchange coupling J in eq 15 is set at J/k_B = 100 K. The C_V calculated backward from energy moment data and $P(E)$ (full circles) at several selected temperatures are included for comparison (see text).

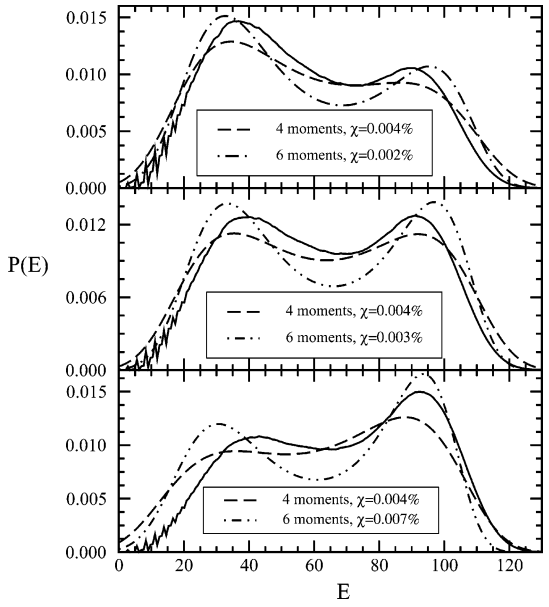

Figure 5. Energy distribution function $P(E)$ vs E calculated using the maximum entropy theory (dashed curve for four moments and dash−dotted curve for six moments) at temperature T = 82.0 (top), 82.3 (middle), and 82.6 K (bottom) compared with simulation results (full curve).

on a lattice. The $q = 2$ Potts model is nothing but the Ising model. The two-dimensional ferromagnetic ($J > 0$) Potts models with $q \leq 4$ possess the second-order phase transition, whereas the first-order phase transition occurs for $q \geq 5$. The system that exhibits the first-order phase transition has a double-peak structure in the energy probability function at the first-order phase transition point.

The Monte Carlo simulation is a powerful numerical method to study the many-body problems in statistical physics. Recently, efficient Monte Carlo algorithms to directly calculate the energy density of states $g(E)$ have been developed,[26,27] and with these techniques, one obtains the accurate double-peak energy probability function for the first-order transition. Here, we use the Wang−Landau algorithm[27] to simulate the six-state Potts model on the square lattice with the periodic boundary conditions. The size of the system is chosen to be 12 × 12. The simulations yield the energy density of states $g(E)$ from which we calculate the energy probability function $P(E)$ at temperature $T = 1/(k_B\beta)$ using eq 1 and also the heat capacity $C_V(T)$, which is computed in the standard way as

$$C_V(T) = k_B\beta^2(\langle E^2\rangle - \langle E\rangle^2) \tag{16}$$

where E is the total energy per spin. Figure 4 portrays the heat capacity per spin for the six-state Potts model. As for the exchange coupling J, we have chosen J/k_B = 100 K for convenience. From the simulation data of the temperature variations of total energy $E(T)$ and heat capacity $C_V(T)$ (Figure 4), we apply the present maximum entropy method to calculate $P(E)$ from the moments of the distribution. We compare the original $P(E)$ from simulations and that calculated by appealing to the computer-simulated heat capacity data. Figure 5 details the comparison of $P(E)$ at T = 82.0, 82.3, and 82.6 K for moments 4 and 6. The two sets of double-peak structures are reproduced, albeit qualitatively. Note that since C_V is an integrated quantity one should be able to check furthermore the calculated $P(E)$ as follows. From the definition of the energy moment at a given temperature T_0, which is defined by

$$E_n(T)_{T=T_0} = \frac{\int_0^1 E^n P(E,T)_{T=T_0}\,\mathrm{d}E}{\int_0^1 P(E,T)_{T=T_0}\,\mathrm{d}E}$$

we may calculate the energy moments from the theoretical $P(E, T)$ and ask if the latter recovers the original energy moment γ_m supposedly defined in the energy range $[\alpha_1, \alpha_2]$, i.e.,

$$E_n/L^n = \sum_{m=0}^{n}(-\alpha_1)^{n-m}\binom{m}{n}\gamma_m$$

where $L = \alpha_2 - \alpha_1$ and $\binom{m}{n}$ are the binomial coefficients. The heat capacity at T_0 is then determined by the following equation

$$C_V(T)_{T=T_0} = (\gamma_2 - {\gamma_1}^2)/(k_B{T_0}^2)$$

Figure 4 confirms that C_V is indeed reasonably well-reproduced.

3.2. Na−K System: $P(\bar{H})$ Function. The phase diagram of the Na−K system is given in Figure 4 of a communication by Potter and Rand.[28] In the temperature regime between 280.2 and 300 K, the phase diagram shows that a solid phase sodium coexists with an liquid binary alloy of composition Na_2K. Interestingly, it was experimentally reported by Krier et al.[6] that the heat capacity C_P in the same temperature range exhibits a maximum. According to their analysis this restricted temperature domain is a two-phase region in which they observed a stable solid sodium phase coexisting with a stable liquid alloy phase of composition Na_2K. These authors went further to ascertain carefully in this part of the temperature region their findings of the experimental C_P. They compared their measured enthalpy change ΔH_{expt} (for T increasing through the two-phase temperature region), which is intimately related to C_P with the enthalpy change ΔH_c, which was calculated from the thermal data of Ginnings et al.[29] The experimental ΔH_{expt} = 443 cal/(g atom) agrees excellently with ΔH_c within a maximum uncertainty of ±4 cal/(g atom). In Figure 6a, we display the C_P versus T curve reported by Krier et al.[6] One sees clearly that there is indeed a discernible peak in C_P. Since the available C_P data are limited, our theory above can only be applied to calculate $P_m(\bar{H})$ subject to a larger χ criterion. As illustrated in Figure 7, we found in

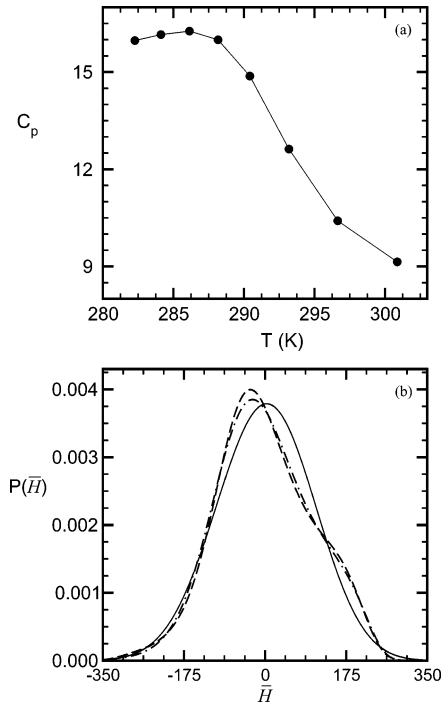

Figure 6. (a) Constant pressure heat capacity C_P (in units of cal/(g atom K) vs T (in K) and (b) enthalpy distribution function $P_m(\bar{H})$ vs $\bar{H}$ (in units of cal/(g atom), calculated for the Na−K system for moments $m = 4$ (full line), 8 (dashed line), and 12 (dot−dashed line). Note that $P_m(\bar{H})$ at $m = 6$ (not shown) remains in a one-peak structure.

fact $\chi \lesssim 0.1\%$ for $m < 12$ and $\chi \approx 0.16\%$ for $m = 12$. The calculated enthalpy distribution functions for moments $m = 4$, 8, and 12 are shown in Figure 6b. Up to six moments (not shown), $P_{m\leq 6}(\bar{H})$ does not point to the coexistence of phases. At moment $m > 6$, we begin to observe bimodal characteristics. The implication of $P_{m>6}(\bar{H})$ is that one of the peaks would correspond to the solid phase sodium and the other is the liquid alloy phase of composition Na_2K. Unfortunately, the authors did not publish their results for H versus T, making the direct conversion of $P(\bar{H})$ to $P(H)$ and hence the attempt to effect a more in-depth analysis of the underlying thermodynamic phase transition not feasible.

3.3. Protein: $P(\bar{H})$ Function. Protein denaturation is a ubiquitous phenomenon. Underlying this equilibrium state transition is the scenario of a protein exhibiting the heterostate fluctuation between two states: the native state (denoted below by N) and the denatured or unfolded state (denoted below by U). Of direct relevance to the present work is the experimental calorimetric data for proteins extensively studied by Makhatadze and Privalov.[7] These authors documented in a review article their measured thermodynamic data for a total of 18 proteins out of which only 16 sets are reliable. Using these thermal data for proteins, the C_P versus T behavior can be determined as follows. We denote by

$$\Delta H(T) = H_U - H_N$$

$$\Delta S(T) = S_U - S_N$$

$$\Delta G(T) = G_U - G_N$$

$$\Delta C_P(T) = C_U - C_N \quad (17)$$

the change of equilibrium states of enthalpy, entropy, Gibbs free energy, and heat capacity of a protein, respectively. In eq 17 the protein is given in the unfolded state with respect to the native one, and the subscript U or N represents a protein distribution of minor variation in their respective basic structure (for instance, N may describe a collection of native conformations, and U the unfolded conformations). On the basis of their extensive analysis of the measured calorimetric data of proteins, Makhatadze and Privalov[7] interpret the denaturation as a simple equilibrium between the N and the U states and thus presented the thermodynamic quantities given in eq 17 as functions of temperature. The ΔG, for example, may be used to account for the probability of occurrence f for the native and unfolded states by

$$f_{N,U} = \frac{1}{1 + \exp[\mp \Delta G/(RT)]} \quad (18)$$

where the − and + correspond to f_N and f_U, respectively, and R is the gas constant. The constant pressure heat capacity is then calculated by the definition $C_P = (\partial H/\partial T)_P$, yielding

$$C_P = (1 - f_N) f_N [(\Delta H)^2/(RT^2)] + C_U + (1 - f_N)\Delta C_P \quad (19)$$

where we have constructed the enthalpy of the protein between the N and the U states by writing

$$H = f_N H_N + f_U H_U \quad (20)$$

which is made use of in deriving C_P. Equation 19 can be used to estimate the measured C_P of a protein if the experimental ΔG (and hence f_N), ΔH, C_U, and ΔC_P as functions of T are known. Equation 19 has been examined by Griko et al.,[30] who made a comparative study between their measured C_P values and those calculated using eq 19. They found that this formula reproduces their measured data with high accuracy. For the present work, we study 16 proteins whose related thermodynamic quantities have been well documented in the literature.[7] Figures 8a−g, 9a−f, and 10a−c delineate C_P versus T for all of these proteins calculated by eqs 17−20. These sets of C_p data are then fitted separately to a polynomial that reads

$$C_P(T) = c_0 + c_1(T - T_r) + c_2(T - T_r)^2 + c_3(T - T_r)^3 + c_4(T - T_r)^4 + \cdots \quad (21)$$

and from each of them we determine the coefficients c_0, c_1, ... of the above equation. The T_r in eq 21 is the reference temperature, which is chosen at the principal peak of C_P. The numerical procedure for determining the enthalpy distribution function $P_m(\bar{H})$ at m moments is now applied to each of the 16 proteins, and the results of computations are illustrated in the same Figures 8a−g, 9a−f, and 10a−c next to C_P curves. In each of the calculated $P(\bar{H})$, we have presented three moments, P_4, P_8, and P_{12}, showing the change of $P_m(\bar{H})$ in the ascending moment.

Before scrutinizing these figures, we should comment on two aspects of our numerical works. First, we have obtained C_P using eq 19 with ΔG (i) taken directly from Table 3 of ref 7 and (ii) calculated from the measured values ΔH and ΔS by the thermodynamic relation $\Delta G = \Delta H - T\Delta S$. The f_N (and hence C_P) defined by eq 18 is therefore obtained at any T by taking the average of i and ii. Second, we have to decide on the largest "stabilized" moment in choosing a "converged" $P_m(\bar{H})$. To this end, we have been guided by the χ criterion ($<0.01\%$, see eq 14 above) in finding out the reasonable range of integration to be used for each moment. This latter numerical aspect has been handled with caution for the calculation is quite sensitive to the range of integration. It turns out in fact that an erroneous

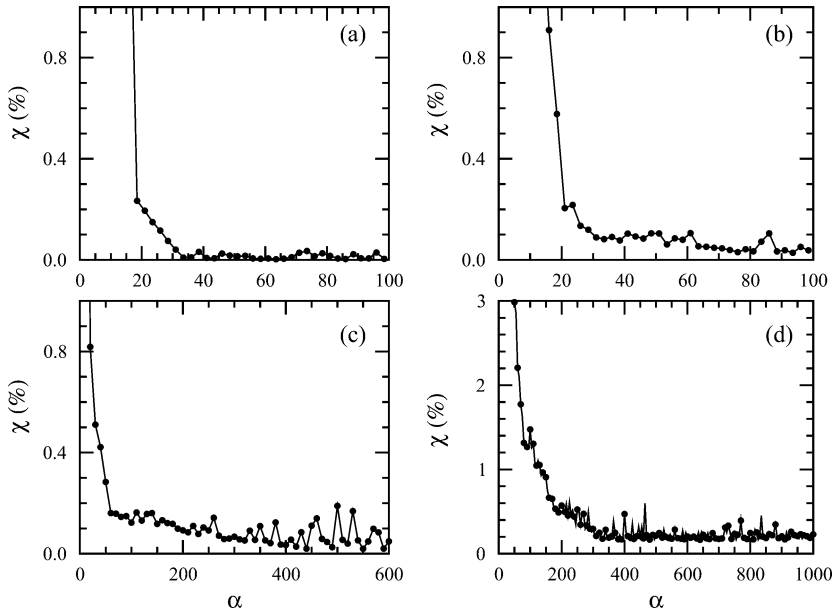

Figure 7. (a) Parameter χ (%) vs integration range α for moments (a) $m = 4$, (b) $m = 6$, (c) $m = 8$, and (d) $m = 12$ calculated for the Na−K binary alloy.

choice of the integration range could lead to prediction of spurious structures.

We now turn to examine the $P(\bar{H})$ of proteins. Quite generally the calculated $P(\bar{H})$ can be classified into three broad categories. In the first category (Figures 8a−g), which includes the proteins BPTI, SH3 domain, G protein, tendamistat, eglin c, CI-2, and interleukin 1β, the calculated enthalpy distribution functions exhibit only one principal peak. It would appear that proteins in this category occupy only the native and denatured states on their folding pathway, but the unfolding and refolding are a monophasic process kinetically. If one knows the $P(H)$ near $T = T_r$, then one would be able to tell in which state the protein is in. Coming to the second category (Figures 9a−f) to which the proteins ubiquitin, RNase T1, barnase, lysozyme, T4 lysozyme, and chymotrypsin belong, we see distinct bimodal structures. These proteins manifest heterostate fluctuations near $T = T_r$ with the protein in state N coexisting with the protein in the U state. For these proteins, well-resolved double-peak characteristics are observed even at a lower moment $P_4(\bar{H})$, and the bimodal structures develop more clearly at higher moments. Finally, the third category (Figures 10a−c) is characterized by the $P(\bar{H})$ revealing a weakly discernible two-peak structure, and proteins cytochrome c, RNase A, and myoglobin are predicted to fall into this category. Apparently, these less well-developed $P(\bar{H})$ structures lie in between the preceding two categories. Intuitively, one would reasonably conjecture that proteins in this category have equal likelihood to take on the N and U states. Before proceeding further, we should mention at this point that Poland[31] has recently reviewed the methodology designed to study the distribution functions from (given) moments and the strategy of maximum entropy theory. He reported results for the $P_m(\bar{H})$ of tendamistat,[10,11] barnase,[8,10] ubiquitin,[11] lysozyme,[11] chymotrypsin,[11] cytochrome c,[11] RNase A[11], and myoglobin,[11−13] but his calculations were all done for the moments $m \leq 6$. Their enthalpy distribution functions are similar to the corresponding $P(\bar{H})$ given in Figures 8−10. There are, however, discernible discrepancies in some of these $P(\bar{H})$ since we have considered $P_m(\bar{H})$ for $m > 6$. Care must thus be exercised in making inference from $P(\bar{H})$ since for some proteins (see, for example, ubiquitin in Figure 9a and also the remark made regarding Figure 6b in subsection 3.2) their $P_m(\bar{H})$ change irregularly with increasing moment m and do not appear stabilized even up to $m = 12$.

Now that we have calculated the $P(\bar{H})$ for 16 proteins and on the basis of their structural profiles, classified them roughly into three kinds, two immediate questions come to mind. Can we interpret these $P(\bar{H})$ data generally? If so, can we explain further the mechanism behind the classification of $P(\bar{H})$ into three categories? To answer the first question, we have followed a scheme most often employed by experimentalists in thermodynamic studies of proteins, i.e., the approximation of a two-state model.[7,8] The scheme is as follows. First, we construct an enthalpy distribution function $Q(H)$ within a two-state model by

$$Q(H) = f_N \rho_N(H) + f_U \rho_U(H) \tag{22}$$

where $\rho_j(H) = \exp[-(H - H_1^j)^2/(2k_B T^2 C_j)]/\sqrt{2\pi k_B T^2 C_j}$ in which H_1^j is the first moment enthalpy for the protein in the native (j = N) or unfolded (j = U) state and C_j is the specific heat in the native (j = N) or unfolded (j = U) state. We then apply eq 22 to calculate the two-state model $Q(H)$, making use of experimental data C_N, C_U, and ΔH at the temperature of the main peak of C_P (equivalent to fixing the origin of $\rho_j(H)$ at $H = H_1^j$). Two relevant remarks are in order. First, $Q(H)$ is actually a measured enthalpy distribution function (within the two-state model) since all input data used in constructing $Q(H)$ are taken from experiments. In fact Poland[8] has compared his calculated $P_6(H)$ of protein barnase with $Q(H)$ and found the two functions strikingly similar. Second, we note that, in solving for $P(\bar{H})$, the $Q(H)$, if it can be constructed, provides a good guide for choosing the integration range, which is a priori unknown. The answer to the first question can therefore be caught sight of from calculating $Q(H)$. We portray in Figures 11−13 the $Q(H)$ of all 16 proteins of interest here. These figures may be compared with the $P_{12}(\bar{H})$ depicted in Figures 8−10. A glance at these enthalpy distribution functions indicates that the calculated $P_{12}(\bar{H})$ for the 16 proteins and their respective $Q(H)$ run from qualitative to semiquantitative agreement. There is, thus, a fair agreement between our calculated

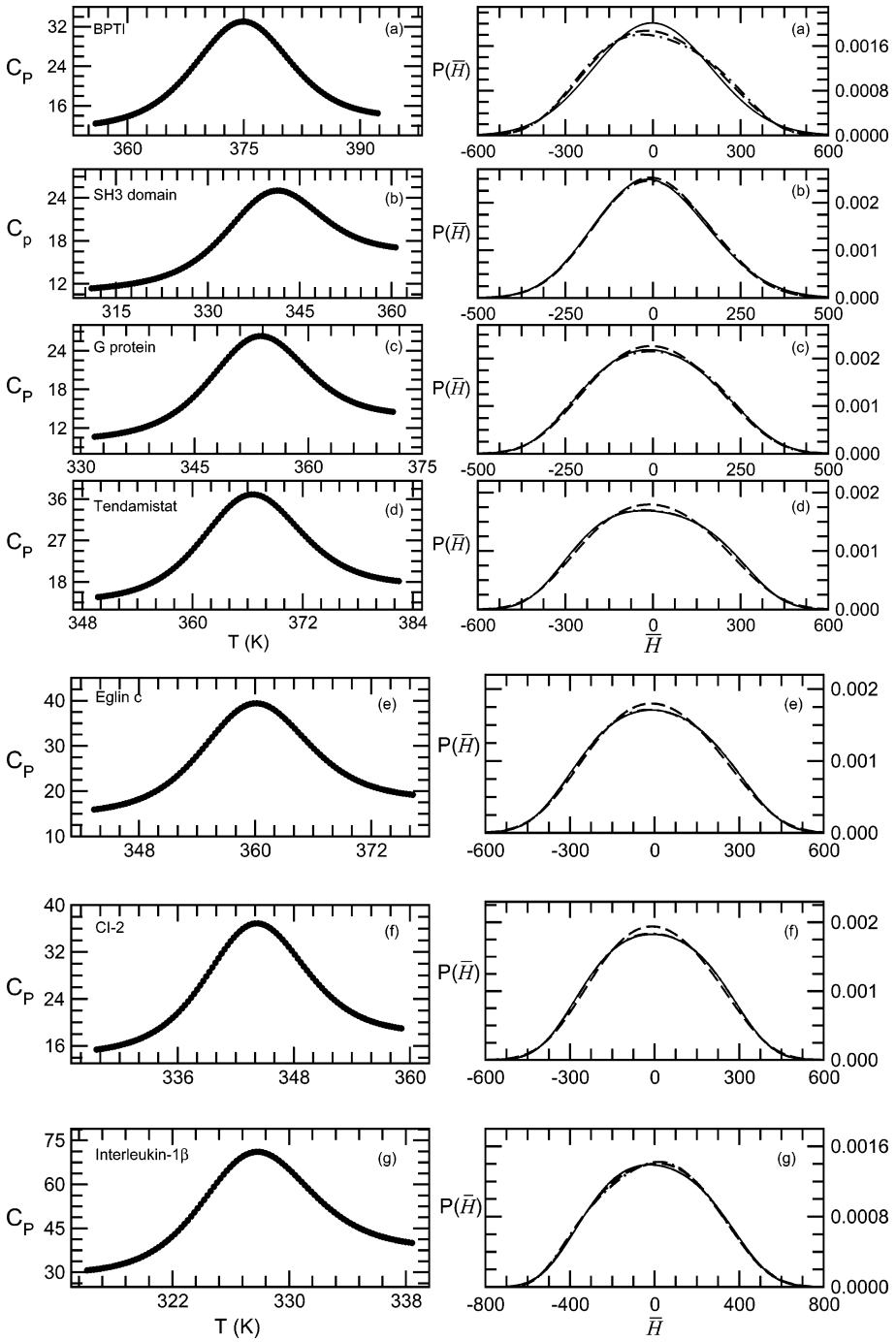

Figure 8. Constant pressure heat capacity C_P (in units of kJ/mol) vs T (in K) and enthalpy distribution function $P_m(\bar{H})$ vs $\bar{H}$ (in units of kJ/mol) calculated for protein (a) BPTI, (b) SH3 domain, (c) G protein, (d) tendamistat, (e) eglin c, (f) CI-2, and (g) interleukin 1β. The calculations were carried out at the reference temperature T_r, which is the main peak of C_P. Symbols used are: full curve, four moments; dashed curve, eight moments; dot−dashed curve, twelve moments.

results and experimentally observed data, and within the approximation of two-state model, the present theory interprets the $P(\bar{H})$ of proteins generally. Quantitatively, however, the proteins in the second category, in contrast to those in the first and third categories, exhibit larger deviation from the two-state model since the $P_m(\bar{H})$ of these proteins change with moment m even at the highest moment $P_{12}(\bar{H})$ considered here. This brings us to the second question raised above.

A hint to the answer to the second question may be sought if we explore further the $P(\bar{H})$ in the second category. Referring to Figures 9a−f, the $P_m(\bar{H})$ in this group of proteins vary with increasing moment m, and their structures reveal no sign of convergence although their bimodal structures mimic those of $Q(H)$ (Figures 12a−f). This is in contrast to the proteins in the first category (Figures 8a−g) where the $P_8(\bar{H})$ of many proteins are already remarkably close to $P_{12}(\bar{H})$. Since the agreement between $P_{12}(\bar{H})$ (Figures 8a−g) and $Q(H)$ (Figures 11a−g) for proteins in the first category are generally good (both manifesting broad single peaks) and these small proteins are characterized mainly by fast kinetics in folding−unfolding,[32] the $P(\bar{H})$ structures of proteins in the second category imply that the two-state behavior may not be an adequate description. Indeed, it has been pointed out in the literature that the non-two-state folding−unfolding is perhaps a more appropriate description for proteins such as ubiquitin,[32,33] barnase,[34] and T4 lysozome[35] in the second category; these proteins denature via intermediates

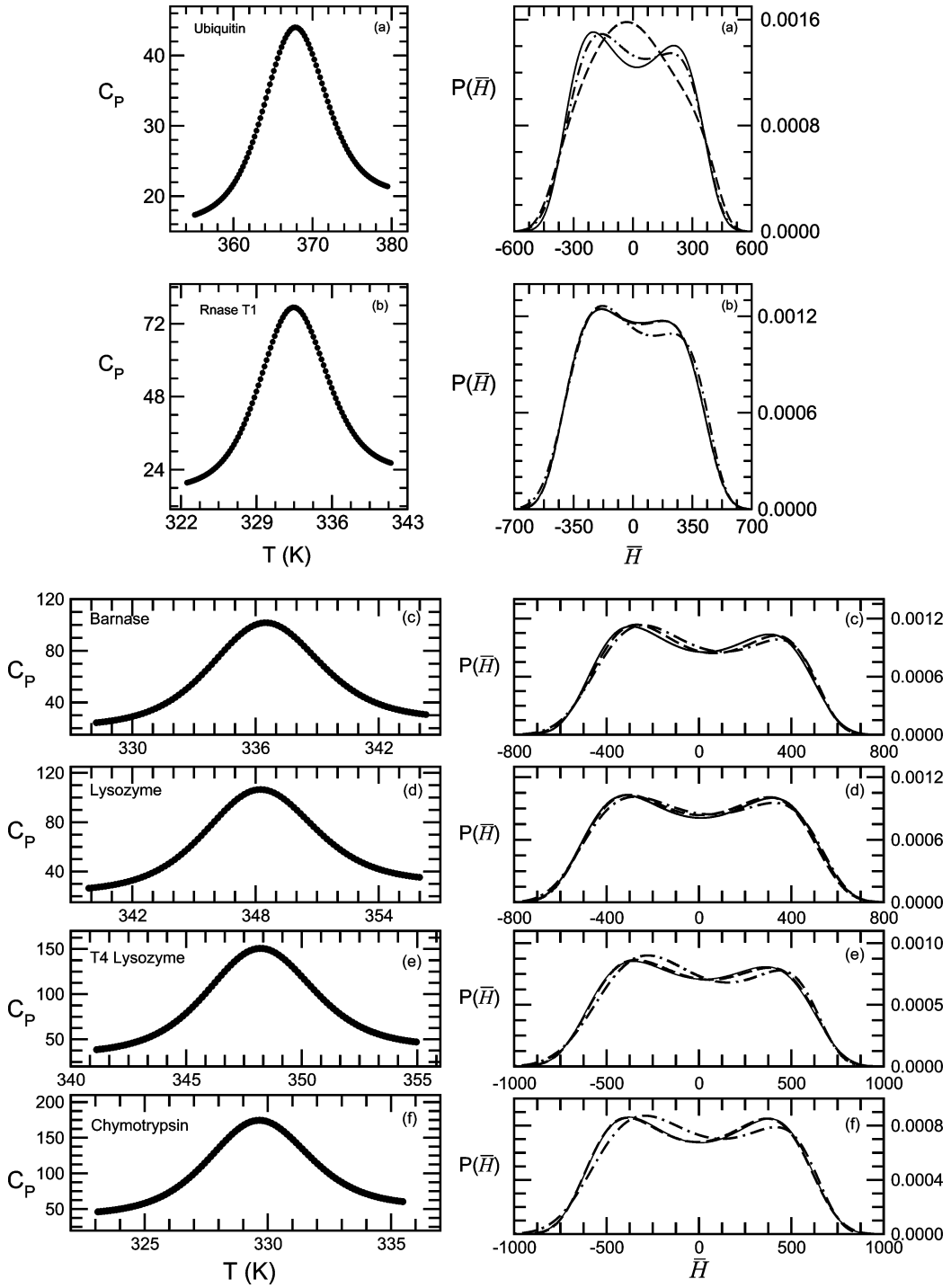

Figure 9. Same as Figure 8 but for (a) ubiquitin, (b) RNase T1, (c) barnase, (d) lysozome, (e) T4 lysozome, and (f) chymotrypsin.

that were observed to slow down the folding−unfolding kinetics. Their behaviors are thus in marked contrast to the smaller proteins SH3 domain[32,36,37] and CI-2[32,38] in the first category where recent thermal and NMR experiments have found them to obey the two-state model quite well. In particular, Korzhnev et al.[37] have demonstrated clearly in their most recent NMR experiment on the SH3 domain that this protein undergoes a two-state transition, and the process proceeds with low-populated folding intermediates and hence signifies fast kinetics. One qualitative means for glimpsing at such variances in the folding−unfolding kinetics is to look more closely at the changes of f_U and f_N as functions of temperature. These probability functions describe the occurrence of U and N states and they vary with temperature. For simplicity, we have chosen three representative proteins, namely, SH3 domain, barnase, and RNase A, one from each category. These functions are displayed in Figures 14a and 14b. The f_N of SH3 domain decays with increasing T in a manner slower than proteins barnase (Figure 14a) and RNase A (Figure 14b). This is due to the fact that in the absence or paucity of intermediate states the transition between native and unfolded states will be unresponsive to temperature until the protein is in the vicinity of $T \approx T_r$. In marked contrast, the f_N of proteins in the second (Figure 14a) and third (Figure 14b) categories, due to the probable occurrence of the stable intermediates, are seen to decline as soon as they are driven thermally. At this point, we should perhaps examine in greater detail the proteins in the third category. For these proteins, we find their $Q(H)$ (Figures 13a−c) exhibiting well-

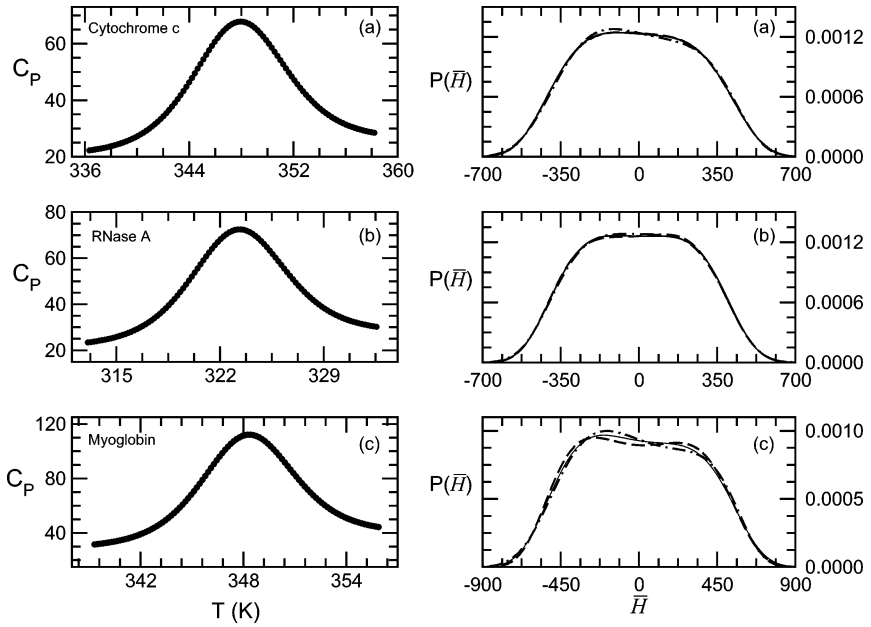

Figure 10. Same as Figure 8 but for (a) cytochrome c, (b) RNase A, and (c) myoglobin.

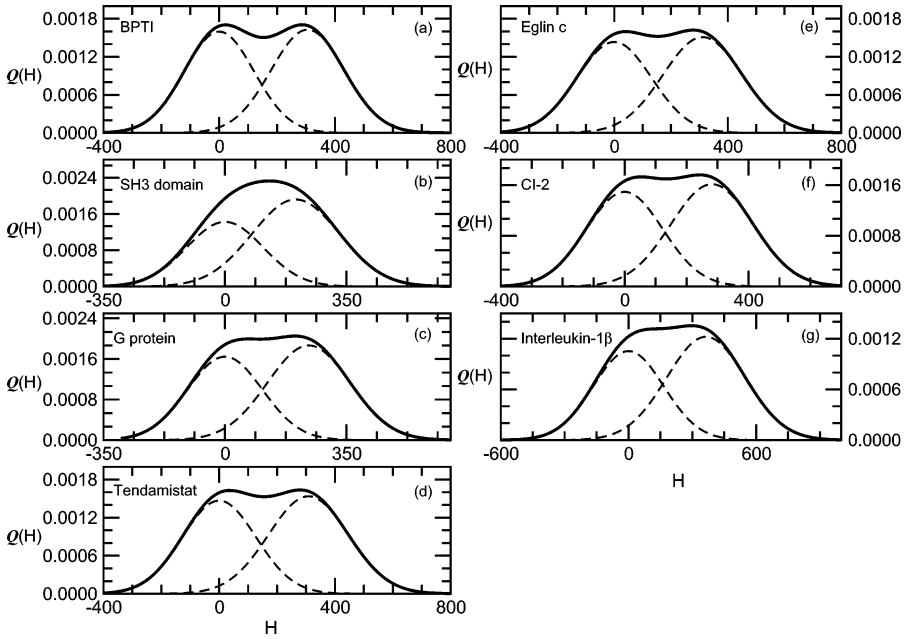

Figure 11. Enthalpy distribution function $Q(H)$ vs H (in units of kJ/mol) calculated by the two-state model given by eq 22 for (a) BPTI, (b) SH3 domain, (c) G protein, (d) tendamistat, (e) eglin c, (f) CI-2, and (g) interleukin 1β. These $Q(H)$ may be compared with the $P_{12}(\bar{H})$ curve presented in Figure 8. Note that, in calculating $Q(H)$, we have set the origin at $H = H_1^N$, and accordingly, $\rho_N(H) = \exp[-H^2/(2k_BT^2C_N)/(\sqrt{2\pi k_BT^2C_N})]$ and $\rho_U(H) = \exp[-(H - (H_1^U - H_1^N))^2/(2k_BT^2C_U)/(\sqrt{2\pi k_BT^2C_U})]$.

resolved two-peak characteristics similar to those proteins in the second category (Figures 12a−f), whereas all of our calculated $P_{12}(\bar{H})$ in the third category show only the weakly visible bimodal structures evident in strike disparity. This discrepancy appears to be the inadequacy of the two-state model for it was reported[39] that cytochrome c (the third category protein, see Figure 13a), though a small, simple, and single domain protein, manifests slow down kinetics,[40] which thus implies non-two-state folding. On the other hand, Chan et al.[41] have demonstrated in their ultrarapid mixing studies on this same protein that it folds with two-state kinetics. Apparently, further investigation, both theoretically and experimentally, is in need for clarifying the kinetic aspects of this protein. The case of the protein chymotrypsin is perhaps even more probable to exhibit the folding−unfolding process via stable intermediates since it is the largest protein considered here and its structure has multiple domains.

4. Conclusion

In statistical mechanics the canonical ensemble theory has been widely used to study the thermodynamic properties of different kinds of physical systems. Central to this ensemble theory is the canonical partition function. Experience shows that an analytical evaluation of the partition function is by no means easy especially for complex systems. The energy or enthalpy

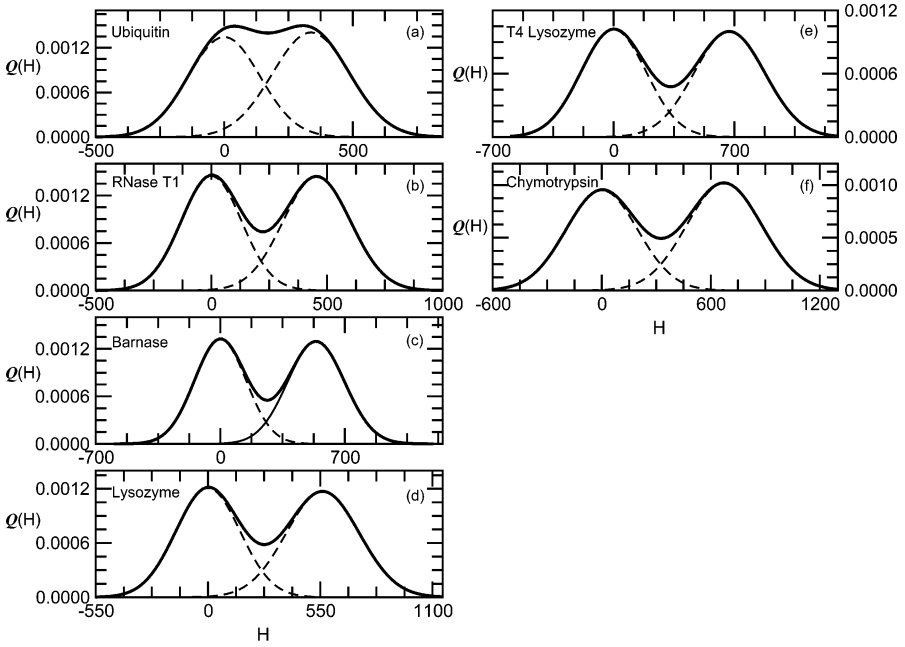

Figure 12. Same as Figure 11 but for (a) ubiquitin, (b) RNase T1, (c) barnase, (d) lysozome, (e) T4 lysozome, and (f) chymotrypsin. These $Q(H)$ may be compared with the $P_{12}(\bar{H})$ curve presented in Figure 9.

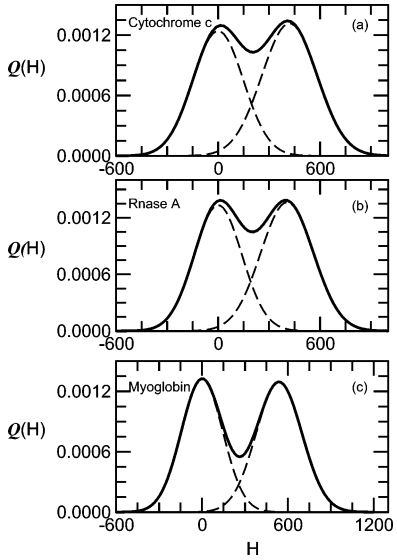

Figure 13. Same as Figure 11 but for (a) cytochrome c, (b) RNase A, and (c) myoglobin. These $Q(H)$ may be compared with the $P_{12}(\bar{H})$ curve presented in Figure 10.

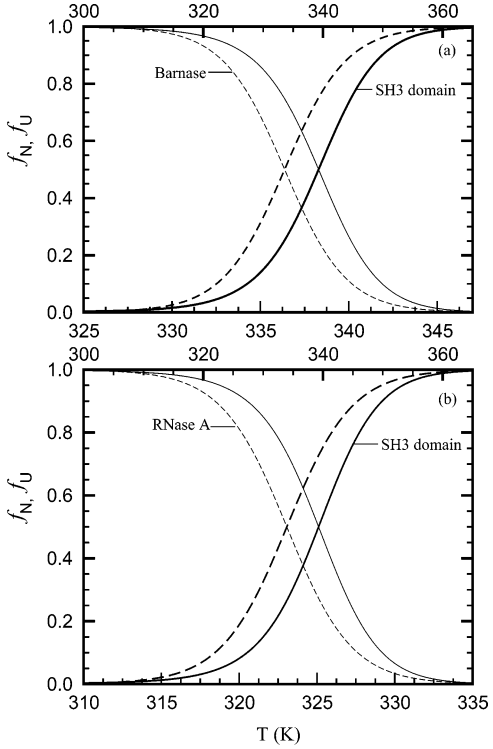

Figure 14. (a) Probability of occurrence f_j vs T (in K), j = N, for the N state (thin full line, SH3 domain; thin dashed line, barnase) and j = U for the U state (thick full line, SH3 domain; thick dashed line, barnase). For barnase read the lower panel beginning at 325 K, and for SH3 domain the top panel beginning at 300 K. (b) Same as part a except the N state (thin dashed line) and the U state (thick dashed line) are for RNase A, and its x-axis is the lower panel beginning at 310 K.

distribution function from which the partition function can be calculated has thus played a vital role in statistical thermodynamic studies. In this work, we apply the maximum entropy method to calculate the energy or enthalpy distribution function from the moments of the distribution. The method has the attractive feature that it requires only the experimental heat capacity data. To evaluate the distribution theory, we calculate, as a test example, the energy distribution function for the q-state Potts model and extract the enthalpy distribution functions for the binary alloy Na−K and for 16 proteins. For the former, our calculated $P(E)$ is in fair agreement with computer simulation results, and for the latter, our analysis of the calculated $P(\bar{H})$ sheds some light on their thermal behavior even without a priori knowledge of the temperature dependence of enthalpy. Specifically, we found that the Na−K system exhibits the heterostate fluctuations where, out of the three intepretations conjectured by Krier et al.,[6] the coexistence of a solid phase sodium and a liquid alloy phase of composition Na_2K is the more probable interpretation of the anomaly in excess C_P. For the 16 proteins, the temperature-driven denaturation may be gleaned from scrutinizing $P(\bar{H})$, and our detailed analysis reveals two broad

qualitative features. For proteins BPTI, SH3 domain, G protein, tendamistat, eglin c, CI-2, interleukin 1β, cytochrome c, and RNase A, in the temperature regime studied here, they populate only the folded and unfolded states on the folding−unfolding pathway, and unfolding and refolding are monophasic processes signaling two-state kinetics, whereas for proteins ubiquitin, RNase T1, barnase, lysozyme, myoglobin, T4 lysozyme, and chymotrypsin, the folding−unfolding transition probably occurs in coexisting states with the native state coexisting with the unfolded state. In the course of folding−unfolding transition process, this second group of proteins is more likely proceeded with a slow kinetics via stable intermediate states.

Acknowledgment. This work is supported by the National Science Council (Grant No. NSC94-2112-M-008-036). S.K.L. thanks Professor Y. Okabe for hosting his visit to the Tokyo Metropolitan University where the final part of this work was written.

Appendix A

In this appendix, we describe the iterative procedure used to calculate the distribution function up to any order of moments. Recall from eq 5 that C_V is expressed as a power series and is defined by $C_V = (\partial U/\partial T)_V$. Accordingly, the expansion coefficients $\{c_0, c_1, ...,c_n\}$ of C_V, supposedly known, are associated (via eq 6) with the quantity $U_T^{(n)} = \partial^n U/\partial T^n$ upon which the moments depend. The calculation of moments, however, requires the derivative of U with respect to β (see eq 4). This can be obtained as follows. Let us denote $M_\beta^{(i)}$ ($M_T^{(i)}$) to be the ith derivative of a quantity M ($E_1 = U$ or H_1) with respect to β (T). We introduce further $M_\beta^{(i,j)}$ ($M_T^{(i,j)}$), which represents the contributions of $M_\beta^{(i)}$ with $j \le i$ in descending order, taking i, $i - 1$, $i - 2$, ..., 1. Each separate contribution i contains the product of $M_T^{(m)}$ and $T_\beta^{(n)} = \partial^n T/\partial\beta^n$. The first four $M_\beta^{(4)}$ can be calculated in terms of $M_\beta^{(i,j)}$ or written as the product of $M_T^{(m)}$ and $T_\beta^{(n)}$ as

$$M_\beta^{(1)} = M_\beta^{(1,1)} = M_T^{(1)} T_\beta^{(1)} \tag{A.1}$$

$$M_\beta^{(2)} = M_\beta^{(2,2)} + M_\beta^{(2,1)}$$

$$= \frac{2!}{(1!)^2 2!} M_T^{(2)} (T_\beta^{(1)})^2 + \frac{2!}{2!} M_T^{(1)} T_\beta^{(2)} \tag{A.2}$$

$$M_\beta^{(3)} = M_\beta^{(3,3)} + M_\beta^{(3,2)} + M_\beta^{(3,1)}$$

$$= \frac{3!}{(1!)^3 3!} M_T^{(3)} (T_\beta^{(1)})^3 + \frac{3!}{2!1!} M_T^{(2)} T_\beta^{(2)} T_\beta^{(1)} + \frac{3!}{3!} M_T^{(1)} T_\beta^{(3)} \tag{A.3}$$

$$M_\beta^{(4)} = M_\beta^{(4,4)} + M_\beta^{(4,3)} + M_\beta^{(4,2)} + \mathrm{M}_\beta^{(4,1)}$$

$$= \frac{4!}{(1!)^4 4!} M_T^{(4)} (T_\beta^{(1)})^4 + \frac{4!}{2!(1!)^2 2!} M_T^{(3)} T_\beta^{(2)} T_\beta^{(1)} T_\beta^{(1)} + \frac{4!}{3!1!} M_T^{(2)} T_\beta^{(3)} T_\beta^{(1)} + \frac{4!}{(2!)^2 2!} M_T^{(2)} (T_\beta^{(2)})^2 + \frac{4!}{4!} M_T^{(1)} T_\beta^{(4)} \cdots \tag{A.4}$$

$$M_\beta^{(n)} = M_\beta^{(n,n)} + M_\beta^{(n,n-1)} + \cdots M_\beta^{(n,2)} + M_\beta^{(n,1)} \tag{A.5}$$

Equations A.1−A.5 can now be employed to iterate the moments. For definiteness, we first define $B_j = (j - 1)!(-1)^j$ and write the nth moment q as q_n. The first four moments can be calculated as the following

$$M_\beta^{(0)} = -\left(\frac{Q_N^{(1)}}{Q_N}\right) = \frac{1!}{1!} B_1 q_1$$

$$q_1 = -M_\beta^{(0)} \tag{A.6}$$

$$M_\beta^{(1)} = \left(\frac{Q_N^{(1)}}{Q_N}\right)^2 - \left(\frac{Q_N^{(2)}}{Q_N}\right)$$

$$= \frac{2!}{(1!)^2 2!} B_2 (q_1)^2 + \frac{2!}{2!} B_1 q_2 = A_{2,2} + A_{2,1}$$

$$q_2 = A_{2,2} - M_\beta^{(1)} \tag{A.7}$$

$$M_\beta^{(2)} = -2\left(\frac{Q_N^{(1)}}{Q_N}\right)^3 + 3\left(\frac{Q_N^{(1)}}{Q_N}\right)\left(\frac{Q_N^{(2)}}{Q_N}\right) - \left(\frac{Q_N^{(3)}}{Q_N}\right)$$

$$= \frac{3!}{(1!)^3 3!} B_3 (q_1)^3 + \frac{3!}{1!2!} B_2 q_1 q_2 + \frac{3!}{3!} B_1 q_3$$

$$= A_{3,3} + A_{3,2} + A_{3,1}$$

$$q_3 = A_{3,3} + A_{3,2} - M_\beta^{(2)} \tag{A.8}$$

$$M_\beta^{(3)} = 6\left(\frac{Q_N^{(1)}}{Q_N}\right)^3 - 2\left(\frac{Q_N^{(1)}}{Q_N}\right)^2\left(\frac{Q_N^{(2)}}{Q_N}\right) + 4\left(\frac{Q_N^{(3)}}{Q_N}\right)\left(\frac{Q_N^{(1)}}{Q_N}\right) + \frac{1}{2}\left(\frac{Q_N^{(2)}}{Q_N}\right)^2 - \left(\frac{Q_N^{(4)}}{Q_N}\right)$$

$$= \frac{4!}{(1!)^4 4!} B_4 (q_1)^4 + \frac{4!}{2!(1!)^2 2!} B_3 q_2 (q_1)^2 + \frac{4!}{3!1!} B_2 q_3 q_1 + \frac{4!}{(2!)^2 2!} B_2 (q_2)^2 + \frac{4!}{4!} B_1 q_4$$

$$= A_{4,4} + A_{4,3} + A_{4,2} + A_{4,1}$$

$$q_4 = A_{4,4} + A_{4,3} + A_{4,2} - M_\beta^{(3)} \cdots \tag{A.9}$$

$$q_n = A_{n,n} + A_{n,n-1} + ...A_{n,2} - M_\beta^{(n-1)} \tag{A.10}$$

Notice that the sets of eqs A.1−A.5 and A.6−A.10 have the same mathematical structures. In this way we obtain the set of moments (q_1, q_2, .., q_n).

References and Notes

(1) Pathria, R. K. *Statistical Mechanics*, 1st ed.; Pergamon Press: Oxford, U. K., 1972; p 65.
(2) Hill, T. L. *An Introduction to Statistical Thermodynamics*; Dover: New York, 1986.
(3) Labastie, P.; Whetten, R. L. *Phys. Rev. Lett.* **1990**, *65*, 1567.
(4) (a) Bulgac, A.; Kusnezov, D. *Phys. Rev. Lett.* **1992**, *68*, 1335. (b) Ju, N.; Bulgac, A. *Phys. Rev. B* **1993**, *48*, 2721.
(5) Ding, F.; Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. *Biophys. J.* **2002**, *84*, 3525.
(6) Krier, C. A.; Craig, R. S.; Wallace, W. E. *J. Phys. Chem.* **1957**, *61*, 522.
(7) Makhatadze, G. I.; Privalov, P. L. *Adv. Protein Chem.* **1995**, *47*, 307.
(8) Poland, D. *J. Chem. Phys.* **2000**, *112*, 6554.
(9) Poland, D. *J. Chem. Phys.* **2000**, *113*, 4774.
(10) Poland, D. *Proteins: Struct., Funct., Genet.* **2001**, *45*, 325.
(11) Poland, D. *Biopolymers* **2001**, *58*, 89.
(12) Poland, D. *J. Protein Chem.* **2002**, *21*, 187.
(13) Poland, D. *Biopolymers* **2002**, *63*, 59.

(14) Poland, D. *Biophys. Chem.* **2002**, *101*−*102*, 485.
(15) Poland, D. *J. Chem. Phys.* **2005**, *123*, 024707.
(16) Poland, D. *Biopolymers* **2006**, *81*, 127.
(17) Poland, D. *J. Chem. Phys.* **2007**, *126*, 054507.
(18) Jaynes, E. T. *Phys. Rev.* **1957**, *106*, 620.
(19) Mead, L. R.; Papanicolaou, N. *J. Math. Phys.* **1993**, *25*, 2404.
(20) Poland, D. *J. Chem. Phys.* **1995**, *102*, 2604.
(21) Bandyopadhyay, K.; Bhattacharya, A. K.; Biswas, P.; Drabold, D. A. *Phys. Rev. E* **2005**, 71, 057701 and references therein.
(22) Ter Haar, D.; Wergeland, H. *Elements of Thermodynamics*; Addison-Wesley: Reading, MA, 1966.
(23) Besley, N. A.; Johnson, R. L.; Stace, A. J.; Uppenbrink, J. *J. Mol. Struct. (THEOCHEM)* **1995**, *341*, 75.
(24) Nelder, J. A.; Mean, R. *Comput. J.* **1965**, *7*, 308.
(25) (a) Potts, R. B. *Proc. Cambridge. Philos. Soc.* **1952**, *48*, 106. (b) Wu, F. Y. *Rev. Mod. Phys.* **1982**, *54*, 235.
(26) (a) Berg, B. A.; Neuhaus, T. *Phys. Lett. B* **1991**, *267*, 249. (b) Berg, B. A.; Neuhaus, T. *Phys. Rev. Lett.* **1992**, *68*, 9. (c) Lee, J. *Phys. Rev. Lett.* **1993**, *71*, 211.
(27) Wang, F.; Landau, D. P. *Phys. Rev. Lett.* **2001**, *86*, 2050.
(28) Potter, P. E.; Rand, M. H. In *Handbook of Thermodynamic and Transport Properties of Alkali Metals*; Ohse, R. W., Ed.; Blackwell Scientific Publications: Oxford, U. K., 1985; Chapter 9.1.
(29) Ginnings, D. C.; Douglas, T. B.; Ball, A. F. *J. Res. Natl. Bur. Stand.* **1950**, *45*, 23.
(30) Griko, Y.; Makhatadze, G. I.; Privalov, P. L.; Hartley, R. W. *Protein Sci.* **1994**, *3*, 669.
(31) Poland, D. *Methods Enzymol.* **2004**, *383*, 427.
(32) Jackson, S. E. *Folding Des.* **1998**, *3*, R81.
(33) Khorasanizadeh, S.; Peters, I. D.; Butt, T. R.; Roder, H. *Biochemistry* **1993**, *32*, 7054.
(34) Dalby, P. A.; Oliveberg, M.; Fersht, A. R. *J. Mol. Biol.* **1998**, *276*, 625.
(35) Llinás, M.; Gillespie, B.; Dahlquist, F. W.; Marquesee, S. *Nat. Struct. Biol.* **1999**, *6*, 1072.
(36) Viguera, A. R.; Martinez, J. C.; Filimonov, V. V.; Mateo, P. L.; Serrano, L. *Biochemistry* **1994**, *33*, 2142.
(37) Korzhnev, D. M.; Salvatella, X.; Vendruscolo, M.; Di Nardo, A. A.; Davidson, A. R.; Donson, C. M.; Kay, L. E. *Nature* **2004**, *430*, 586.
(38) Jackson, S. E.; Fersht, A. R. *Biochemistry* **1991**, *30*, 10428.
(39) Plaxco, K. W.; Simons, K. T.; Ruczinski, I.; Baker, D. *Biochemistry* **2000**, *39*, 11177.
(40) Mines, G. A.; Pascher, T.; Lee, S. C.; Winkler, J. R.; Gray, H. B. *Chem. Biol.* **1996**, *3*, 491.
(41) Chan, C. K.; Hu, Y.; Takahashi, S.; Rousseau, D. L.; Eaton, W. A.; Hofrichter, J. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 1779.