THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

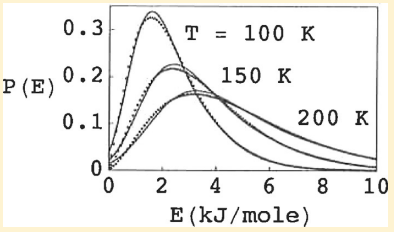
pubs.acs.org/JPCB

Empirical Protein Partition Functions

Douglas Poland

Department of Chemistry, The Johns Hopkins University, Baltimore, Maryland 21218, United States

ABSTRACT: In the present paper, we outline how to construct the partition function for a protein using empirical heat capacity data. The procedure is based on the calculation of a set of energy moments from the temperature dependence of the heat capacity. Given a set of energy moments, one can then use the maximum-entropy method to calculate an approximate energy distribution for the protein; the more energy moments one has, the better the approximation. The energy distribution can then be used to calculate the probability that the molecule is in a given energy level, which, using standard statistical mechanics, gives the degeneracy of the particular energy level. The degeneracy as a function of energy is the central ingredient in the construction of the partition function. Given the partition function, one can calculate all of the thermodynamic functions of the protein (free energy, energy, entropy, heat capacity, and energy probability distribution) as a function of temperature. The three-dimensional plot of the probability that the protein has a given energy at a given temperature tells one graphically (without imposing the assumption) whether or not it is a good approximation to divide the terms in the partition function into two or more groups, reflecting, for example, the presence of distinct native and denatured species.



INTRODUCTION

The statistical mechanics of helix−coil transitions in polypeptides was reviewed several years ago.[1] On the experimental side, the determination of the Zimm−Bragg σ and s parameters for polymers of all of the naturally occurring amino acids has been accomplished over a 20 year period.[2,3] In the present paper, we address the statistical mechanics of proteins, in particular the calculation of protein partition functions based on empirical data. In previous papers,[4−14] we have shown how empirical heat capacity data can be used to calculate energy probability distributions for molecules. One first converts the heat capacity data into a set of energy moments that, in turn, employing the maximum-entropy method, are used to calculate energy probability distributions for molecular species. Here we use the same approach to calculate the partition function for any molecule. While the method is general, we will apply it to proteins in the present paper.

Since we will be treating the internal energy states of a molecule, we will consider the canonical partition function of the molecule[15]

$$Q = \sum_E W(E) \exp[-E/RT] \tag{1}$$

where $W(E)$ is the degeneracy of level-E and the sum is over the internal energy states of the molecule. We note that $W(E)$ is independent of temperature and hence all of the temperature dependence of Q comes from the Boltzmann factor $\exp(-E/RT)$. The temperature dependence of Q is thus the same for all species, and it is $W(E)$ that varies from one species to the next. Thus, if we know $W(E)$, we know Q (and hence all of the thermodynamics).

The probability that the molecule is in level-E is obtained in the standard way

$$P(E) = W(E) \exp[-E/RT]/Q \tag{2}$$

which we can solve for $W(E)$

$$W(E) = P(E) \exp[E/RT]Q \tag{3}$$

As we have shown,[6,7] $P(E)$ can be obtained from empirical heat capacity data via the maximum-entropy method. Likewise, Q can be obtained from the heat capacity using standard thermodynamics (see eq 24). From eq 3, we see that $P(E)$ is the main ingredient in the calculation of $W(E)$. In the next section, we review the calculation of $P(E)$.

Energy Distributions. The energy distribution $P(E)$ is obtained from energy moments which are defined as follows

$$E_n = \langle E^n \rangle = \sum_E E^n P(E) \tag{4}$$

The moments can be obtained from derivatives of Q with respect to $\beta = 1/RT$

$$E_n = [\partial^n Q/\partial(-\beta)^n]/Q \tag{5}$$

For example, the first two moments are

$$E_1 = \langle E \rangle = U$$
$$E_2 = \langle E^2 \rangle = {E_1}^2 + RT^2 C(T) \tag{6}$$

where U is the internal energy and $C(T)$ is the heat capacity.

Special Issue: Harold A. Scheraga Festschrift

Received: December 7, 2011
Revised: February 21, 2012
Published: February 28, 2012

The energy moments can then be obtained from the temperature dependence of the heat capacity which can be represented by the standard Taylor series expansion

$$C(T) = C_0 + C_1\Delta T + C_2\Delta T^2 + ... \tag{7}$$

where

$$C_j = \frac{1}{j!}[\partial^j C/\partial T^j]_{T_m} \tag{8}$$

and

$$\Delta T = T - T_m \tag{9}$$

where T_m is a reference temperature, for example, the temperature corresponding to the maximum in the heat capacity.

To calculate M moments requires the values of C_j through $j = M - 2$. Thus, to obtain four moments requires the C_j coefficients through C_2, while to obtain six moments requires the coefficients through C_4. The C_j coefficients in eq 7 thus are our sole sources of empirical data from which we construct the partition function Q.

The next step is to use the energy moments obtained from the coefficients C_j in eq 7 to calculate the energy distribution $P(E)$. Using the maximum-entropy method,[16,17] we have

$$P(E) = \exp[-\Gamma(E)] \tag{10}$$

where

$$\Gamma(E) = \sum_{m=0}^{M} \lambda_m E^m \tag{11}$$

with M being the number of energy moments one has. Thus, given E_n through $n = M$, one calculates (using a simple nonlinear iteration procedure) M values of the coefficients λ_m. One thus trades M values of E_n for M values of λ_m.

For the case of $M = 2$, the quantity $\Gamma(E)$ of eq 11 will be a quadratic in E and hence will give a unimodal distribution. For the case of $M = 4$, the quantity $\Gamma(E)$ will be a quartic in E and the distribution can then be bimodal.

Given two moments, the maximum-entropy method gives exactly the standard Gaussian distribution

$$P(E) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-(E - E_1)^2/2\sigma^2] \tag{12}$$

where

$$\sigma = \sqrt{E_2 - E_1^2} = \sqrt{RT^2C(T)} \tag{13}$$

We now give an example of the use of $P(E)$ to calculate $W(E)$ and hence the partition function for a simple system.

3-d Harmonic Oscillator. As an example of the use of $P(E)$ to obtain $W(E)$, we treat a model we have used before, namely, the isotropic harmonic oscillator in three dimensions.[6] Taking the lowest energy level as zero, the partition function for this species is given by

$$Q = (\sum_{n=0} X^n)^3 = \left(\frac{1}{1 - X}\right)^3 = \sum_{n=0} W_n X^n \tag{14}$$

where

$$X = \exp[-\varepsilon/RT] \tag{15}$$

and

$$W_n = \frac{1}{2}(n + 1)(n + 2) \tag{16}$$

Epsilon is the standard energy spacing between neighboring levels. The probability of the species being in level-n is given by

$$P(n) = W_n X^n/Q \tag{17}$$

The sums in eq 14 and the distribution in eq 17 are given in terms of the levels n. These can be converted to energy functions using the relation

$$E = n\varepsilon \tag{18}$$

and

$$\Delta n = \Delta E/\varepsilon \tag{19}$$

The energy distribution is then given by

$$P(E) = P(n)/\varepsilon \tag{20}$$

We now want to compare the exact distribution given by eqs 17 and 20 with that obtained using the maximum-entropy method. The required moments are calculated from eq 5 using the closed form for Q given in eq 14. The results are given in Figure 1 where the distributions are shown for T = 100, 150, and 200 K. We take $X = 0.9$ at $T = 100$ K which gives ε = 0.0876 kJ/mol. The solid dots in Figure 1 give the exact distributions, while the solid lines are the maximum-entropy distributions. One sees that for this model the maximum-entropy method very accurately reproduces the energy distributions.

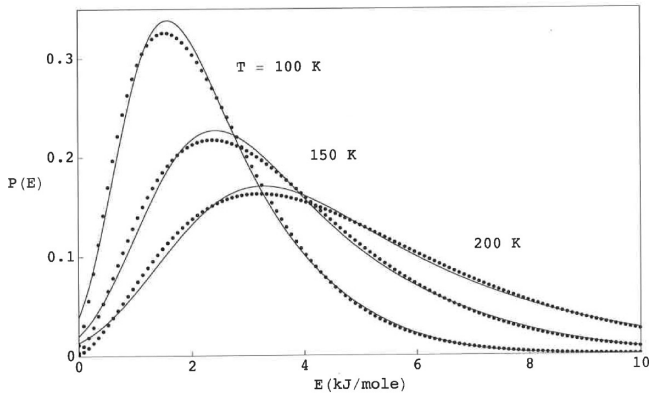


Figure 1. The probability, $P(E)$, that a 3-d harmonic oscillator will be in energy level E at T = 100, 150, and 200 K. For ε = 0.0876, the solid dots give the exact distributions given by eq 20, while the solid lines are the maximum-entropy distributions.

Finally, we calculate $W(E)$. We take eq 17, and, using eq 18 to convert to the energy variable, solve for $W(E)$

$$\ln W(E) = \ln P(E) + \beta E + \ln Q \tag{21}$$

where $\beta = 1/RT$. We use eq 14 to calculate Q and the maximum-entropy method to calculate $P(E)$. We compare this result with the exact relation obtained from eqs 16 and 18

$$W(E) = \varepsilon^{-1}(E/\varepsilon + 1)(E/\varepsilon + 2) \tag{22}$$

The results are shown in Figure 2 where the solid dots give the exact result of eq 22 and the solid line gives the maximum-entropy result of eq 21. One sees that the maximum-entropy method based on energy moments very accurately reproduces

dx.doi.org/10.1021/jp211794u | J. Phys. Chem. B 2012, 116, 6683–6693

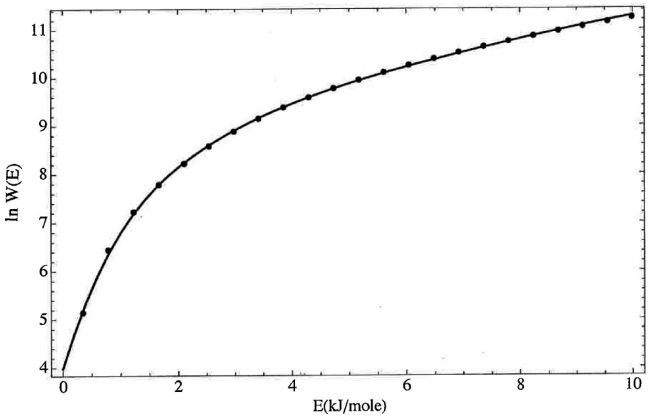


Figure 2. The quantity ln $W(E)$ for the 3-d harmonic oscillator as given by the maximum-entropy method using eq 21 in the solid line, while the solid dots give the exact result of eq 22.

$W(E)$ and hence the partition function Q and all of the thermodynamics for this model. We now apply this method to proteins.

Protein Partition Functions. As an example of the calculation of the partition function for a protein, we take the case of barnase which has been studied in depth in the laboratory of Prof. P. L. Privalov at Hopkins.[18,19] Barnase is a small, stable protein consisting of 110 amino acids with no disulfide cross-links that unfolds reversibly. We have also previously calculated the energy distributions of this protein.[4,7] Our starting point in the calculation of the partition function for barnase is the empirical heat capacity of this protein as a function of temperature as determined by Privalov et al.[18] This is shown in the upper graph in Figure 3 where T_0 is the lowest temperature for which one has heat capacity data. For barnase, T_0 = 325 K (the maximum temperature is 350 K). The internal energy and the entropy relative to the values at T_0 are given by the standard relations

$$\Delta U = \int_{T_0}^{T} C(T)\,\mathrm{d}T = U(T) - U_0(T_0)$$

$$\Delta S = \int_{T_0}^{T} (C(T)/T)\,\mathrm{d}T = S(T) - S_0(T_0) \tag{23}$$

For simplicity, we will take $U_0 = S_0 = 0$. Thus, we measure all of the thermodynamic quantities relative to their values at T_0. This choice of reference does not influence the energy distribution $P(E)$. Using these reference points, we have

$$\Delta A = \Delta U - T\Delta S = -RT \ln Q \tag{24}$$

where A is the Helmholtz free energy (recall from eq 1 that Q is a sum over internal energy states). The quantities $\Delta Y(T)$, $\Delta S(T)$, and $\Delta A(T)$ are given by the solid lines in Figures 3 and 4. The solid dots in Figures 3 and 4 will be discussed in connection with eqs 33 and 34.

To calculate $W(E)$ for barnase, we first use the heat capacity to obtain $P(E)$. We then use eq 3 to give $W(E)$. Our central temperature is determined by the T that corresponds to the maximum in $C(T)$ which is

$$T_m = 337.1 \text{ K} \tag{25}$$

In Figure 5, we show $P(E)$ calculated at three temperatures (T = 325, T_m, and 350 K) and using two different sets of moments (M = 2 and M = 4). The two-moment data are given

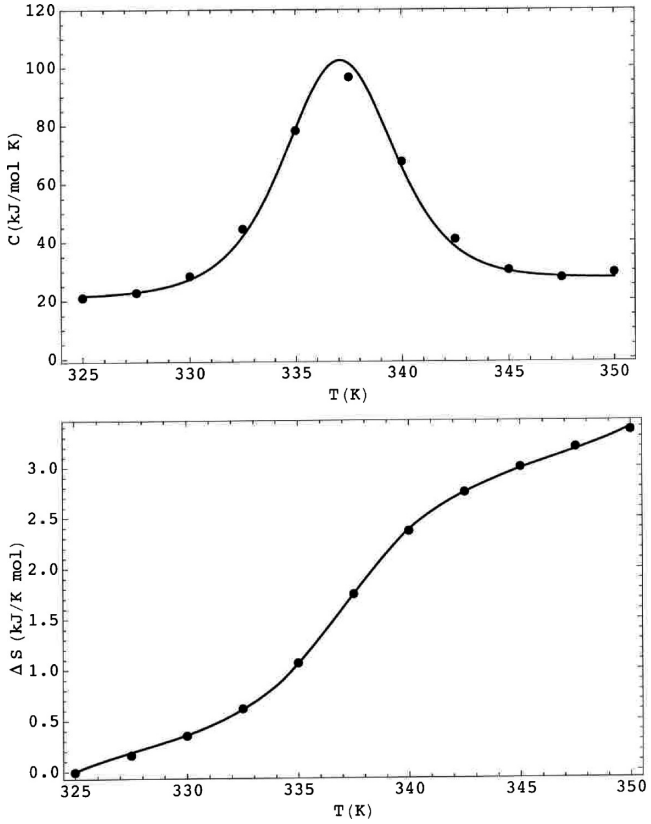


Figure 3. The solid line in the upper graph gives the experimental values of the heat capacity of barnase as a function of temperature. The solid line in the lower graph gives ΔS of barnase obtained from the heat capacity using eq 23. We will return to the solid dots in connection with eqs 33 and 34.

by the solid dots, while the four-moment data are given by the solid lines. One sees that at T = 325 and 350 K, the two-moment and four-moment distributions are essentially identical. On the other hand, at T_m, there is a qualitative difference in the two distributions with the two-moment distribution being unimodal and the four-moment distribution being bimodal.

We then use $P(E)$ as shown in Figure 5 and Q from eq 24 in eq 3 to obtain $W(E)$. Since $W(E)$ is independent of temperature, we should obtain the same function $W(E)$ from the data at different temperatures and this is so to a very good approximation. One can, however, obtain an even better determination of $W(E)$ by recognizing that a given $P(E)$ distribution function has a range of E where the function has a maximum that depends on the value of the temperature, as shown in Figure 5. We thus can use a set of temperatures that give a set of $P(E)$ functions that peak at different E values. At a given temperature T_j, we have

$$W_j(E) = P_j \exp[E/RT_j]Q \tag{26}$$

At a given value of E, we calculate the relative values of the P_j functions

$$f_j = P_j / \sum_i P_i \tag{27}$$

We then construct the weighted average

$$\langle \ln W(E) \rangle = \sum_j f_j \ln W_j \tag{28}$$

dx.doi.org/10.1021/jp211794u | J. Phys. Chem. B 2012, 116, 6683–6693

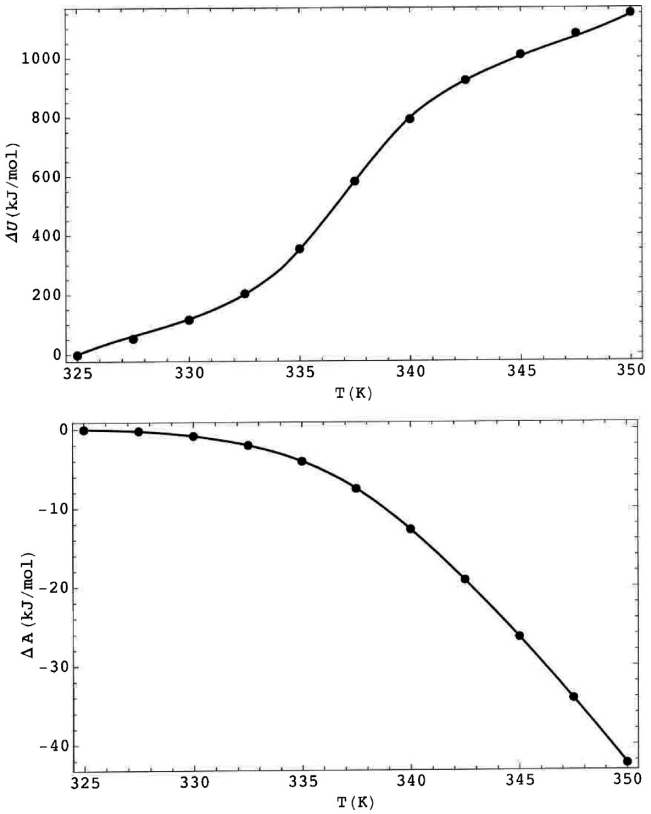


Figure 4. The solid line in the upper graph gives ΔU of barnase obtained from the heat capacity using eq 23. The solid line in the lower graph gives ΔA of barnase obtained from the heat capacity using eq 23.

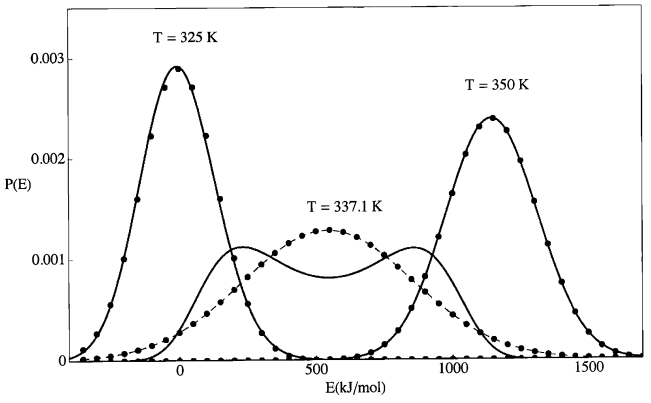


Figure 5. $P(E)$ for barnase calculated at three temperatures (T = 325, T_m, and T = 350 K) using two sets of moments (with M = 2 and M = 4). The solid dots give the two-moment data, while the solid lines give the four-moment data.

This function now determines $W(E)$ using the temperature where $P(E)$ is a maximum for a given value of E.

The ln $W(E)$ given by eq 28 is shown in Figure 6 by the solid line. It is somewhat unexpected that the function is almost a straight line. Referring to eq 1, we can define the quantity

$$\phi(E, T) = -\ln W(E) + E/RT \tag{29}$$

which is the free energy divided by RT for energy level-E. For a protein to denature at T_m, one requires that ϕ is approximately

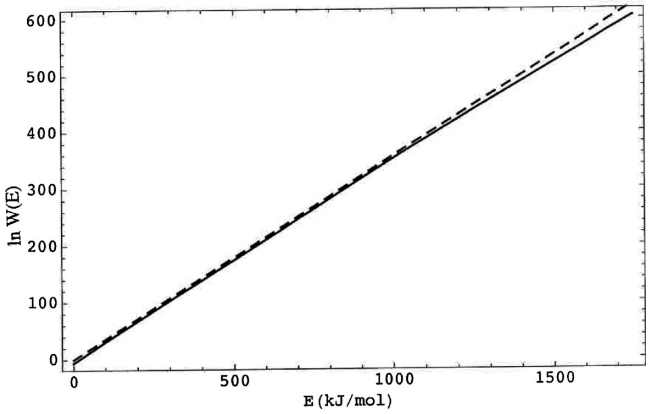


Figure 6. The solid line gives ln $W(E)$ for barnase given by eq 28, while the dashed curve gives the quantity E/RT_m.

zero or that ln W is approximately E/RT_m. We then define the reference ϕ

$$\phi_m = -\ln W(E) + E/RT_m \tag{30}$$

The quantity E/RT_m is also shown in Figure 6 (dashed curve), and one sees that it can almost be superimposed on ln W. Thus, the quantity ϕ_m in eq 30 is the small difference between two large quantities and it will be more convenient to use the quantity ϕ_m rather than ln $W(E)$ by itself.

At other temperatures, we have (with no approximation)

$$\phi(E, T) = \phi_m + \frac{E}{R}\left[\frac{1}{T} - \frac{1}{T_m}\right] \tag{31}$$

Thus, the function $\phi(E,T)$ is given at general E and T in terms of ϕ_m and therefore ϕ_m gives all of the thermodynamics of the protein. The function $\phi_m(E)$ of eq 30 is plotted in Figure 7. We

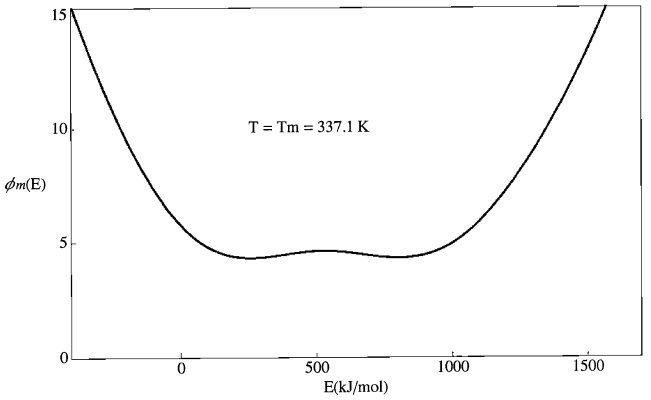


Figure 7. The function $\phi_m(E)$ for barnase as given by eq 30.

illustrate $\phi(E,T)$ for T = 325, T_m, and 350 K in Figure 8. The curves for T = 325 and 350 K are obtained from ϕ_m given in Figure 7 by utilizing eq 31.

Given ϕ_m as shown in Figure 7, we need to check that this function indeed generates all of the thermodynamics for barnase. To this end, we define a set of integrals (treating E as a continuous variable)

$$I_n(T) = \int_E E^n \exp[-\phi(E, T)]\, dE \tag{32}$$

dx.doi.org/10.1021/jp211794u | J. Phys. Chem. B 2012, 116, 6683–6693

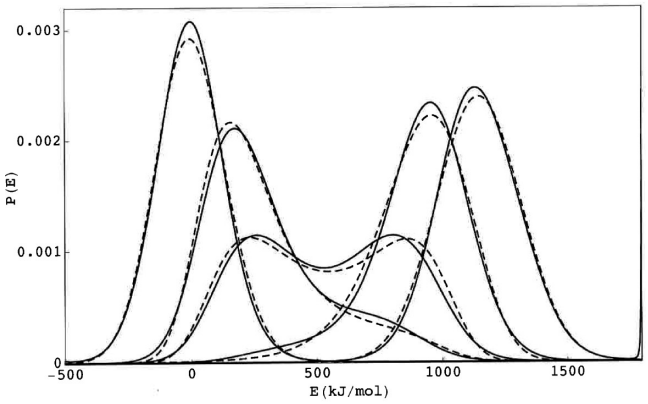


Figure 8. The energy distribution functions for barnase obtained from the maximum-entropy procedure (solid lines) and from eq 35 using eqs 30 and 31 (dashed lines). The temperatures represented are (left to right): T = 325, 334.25, 337.1, 342, and 350 K.

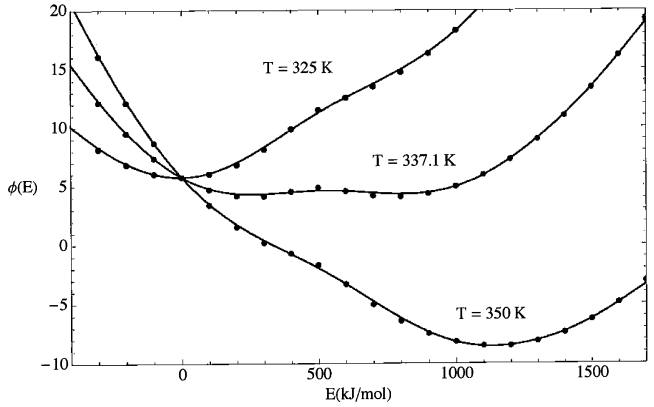


Figure 9. The solid lines give the quantity $\phi(E)$ for barnase at three different temperatures (T = 325, T_m, and T = 350 K) from eq 31. The solid dots give ϕ^* constructed from the two-species relations of eq 40.

where $\phi(E,T)$ is given by eq 31 in terms of $\phi_m(E)$. The thermodynamic quantities are then given by

$$\Delta U = \langle E \rangle = I_1/I_0$$

$$\Delta A = -RT \ln I_0$$

$$\Delta S = (\Delta U - \Delta A)/T = R \ln I_0 + (I_1/I_0)/T \quad (33)$$

and

$$C = \partial \Delta U/\partial T = [(I_2/I_0) - (I_1/I_0)^2]/RT^2 \quad (34)$$

The quantities given above are indicated by the solid dots in Figures 3 and 4, and one sees that the single function $\phi_m(E)$ given in Figure 7 gives all of the standard thermodynamic functions with great accuracy. Finally, we can also obtain the $P(E)$ functions as follows

$$P(E) = W(E) \exp[-E/RT]/Q$$

$$= \exp[-\phi(E, T)]/\int_E \exp[-\phi(E, T)]\, dE \quad (35)$$

where again we calculate ϕ from $\phi_m(E)$ as in eq 31.

The comparison of the $P(E)$ obtained from the maximum-entropy procedure (solid lines) and those obtained from ϕ_m using eq 35 (dashed lines) are shown in Figure 8. Again, the results of the two procedures are in good agreement.

Figure 9 shows three isotherms of the function $\phi(E,T)$ calculated using eq 31 (solid lines). The solid dots give the quantity ϕ^* constructed from the two-species relations of eq 41 to be discussed in the next section. Figure 10 shows a three-dimensional plot of the function $\phi(E,T)$. This figure is a graphic display of the empirical partition function.

In Figure 11, we show isotherms of the function $P(E)$ in a continuous three-dimensional graph. The function is given by eq 35. It is clear in Figure 11 that there are two distinct branches of the $P(E)$ function, one at low temperature and one at high temperature. It is natural to associate the low-temperature branch with the native form of the protein (designated by the N in Figure 11) and the high-temperature branch with the denatured form (designated by the D in Figure 11). We note that this interpretation is an assumption that we will explore further in the next section.

In Figure 12, we plot the loci (solid lines) of the ridge points for the two branches (labeled N and D, respectively). This

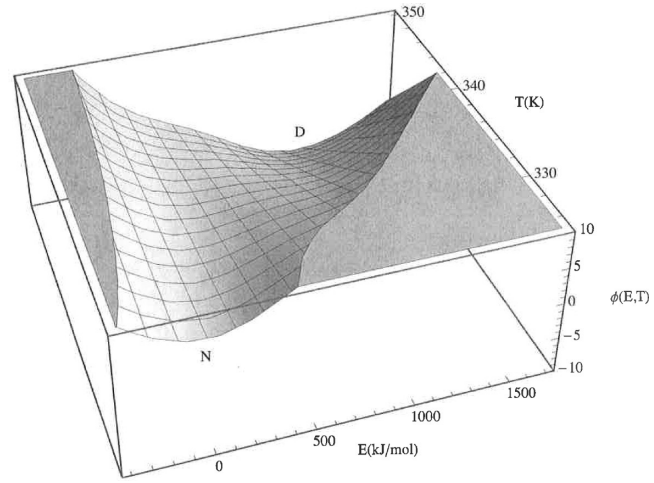


Figure 10. Three-dimensional plot of $\phi(E,T)$ for barnase obtained from eqs 30 and 31. The symbols N and D represent the loci of the native and denatured forms, respectively.

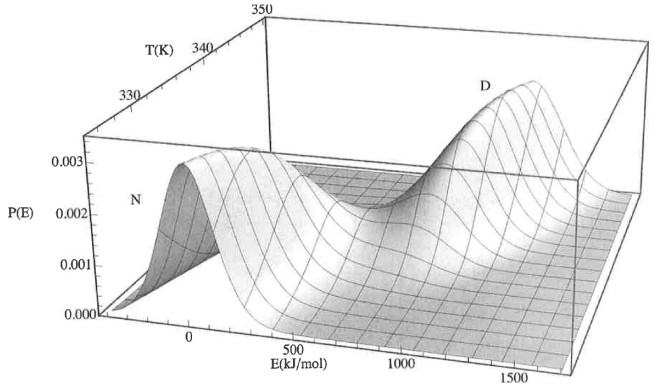


Figure 11. Three-dimensional plot of $P(E,T)$ for barnase obtained from eq 35. The symbols N and D represent the loci of the native and denatured forms, respectively.

clearly shows that the maximum in $P(E)$ goes from the low-temperature branch over to the high-temperature branch as the temperature is increased. The solid dots in Figure 12 indicate the loci of the minima between the two ridge lines.

Two-Species Approximation. The partition function Q is in general a sum over states. The structure of the $P(E)$ function

dx.doi.org/10.1021/jp211794u | J. Phys. Chem. B 2012, 116, 6683−6693

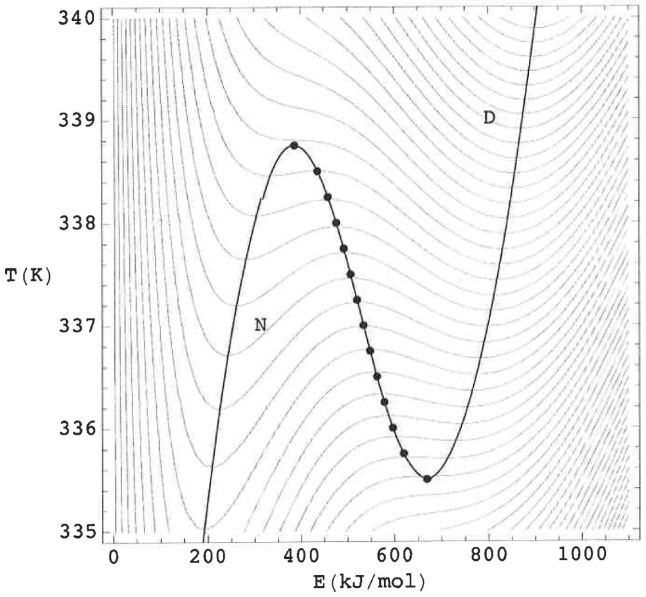


Figure 12. The loci of the ridgelines in the $E-T$ plane shown in Figure 11 labeled N and D. The solid dots represent the loci of the minima between the two ridgelines.

shown in Figure 11 suggests that there are two distinctly different groups of states which it is reasonable to identify as the native (N), low-temperature states and the denatured (D), high-temperature states. This has been done before, approximating Q by a Gaussian function (constructed from two energy moments) but not from the point of view of the $\phi(E,T)$ function obtained from the $P(E,T)$ function.[20]

We thus make the approximation that we can write Q as follows

$$Q^* = Q_N + Q_D \tag{36}$$

where

$$Q^* = \sum_E \exp[-\phi(E, T)] \tag{37}$$

with

$$Q_N = \sum_E \exp[-\phi_N(E)]$$
$$Q_D = \sum_E \exp[-\phi_D(E)] \tag{38}$$

Using Q_N and Q_D of eq 38 in eq 36, one has

$$\exp[-\phi^*(E)] = \exp[-\phi_N(E)] + \exp[-\phi_D(E)] \tag{39}$$

or equivalently

$$\phi^*(E) = -\ln[\exp[-\phi_N(E)] + \exp[-\phi_D(E)]] \tag{40}$$

Now the quantities ϕ_N and ϕ_D in eqs 39 and 40 have the same structure as ϕ in eq 31 that we write as

$$\phi_N(E) = \phi_N(T_N) - \frac{E}{R}\left[\frac{1}{T_N} - \frac{1}{T}\right]$$
$$\phi_D(E) = \phi_D(T_D) - \frac{E}{R}\left[\frac{1}{T_D} - \frac{1}{T}\right] \tag{41}$$

where we take

$$T_N = 325 \text{ K}$$
$$T_D = 350 \text{ K} \tag{42}$$

We refer to Figure 11 and give the interpretation that the low-temperature ridge represents N while the high-temperature ridge represents D.

In Figure 13, we show isotherms (solid lines) of $\phi(E,T)$ for $T = T_N$ and $T = T_D$ based on $P(E)$ calculated using four energy

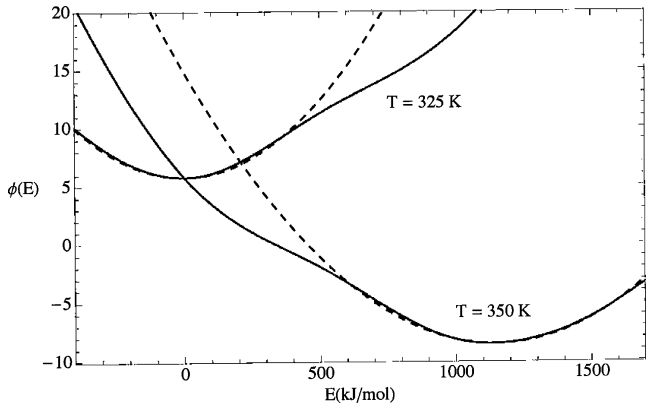


Figure 13. Isotherms (solid lines) of $\phi(E)$ for $T = T_N$ and $T = T_D$ based on $P(E)$ calculated using four moments. The dashed curves represent $\phi_N(E)$ and $\phi_D(E)$.

moments. The dashed curves represent $\phi_N(E)$ and $\phi_D(E)$ based on $P(E)$ calculated using two energy moments. Hence, ϕ is based on a quartic Γ, while ϕ_N and ϕ_D are based on two quadratic Γ's (see eq 11). Thus, two quadratic functions give essentially the same result as a single quartic function, as shown in Figure 13.

We can use eq 41 for ϕ_N and ϕ_D in eq 40 to obtain the two-species approximation of $\phi^*(E,T)$. The solid dots in Figure 9 give the values of ϕ^* based on the two-moment $P(E)$ functions, while the solid lines give the $\phi(E,T)$ based on the four-moment $P(E)$ function. We see that the two-species approximation form $\phi^*(E,T)$ is essentially identical to $\phi(E,T)$.

It is useful to examine the three-dimensional functions $P_N(E)$ and $P_D(E)$ based on ϕ_N and ϕ_D as follows

$$P_N(E, T) = \exp[-\phi_N(E)]/\int_E \exp[-\phi_N(E)]\, dE$$
$$P_D(E, T) = \exp[-\phi_D(E)]/\int_E \exp[-\phi_D(E)]\, dE \tag{43}$$

In Figure 14, we show independent isotherms of these two functions. The two ridges give a ridge (left) for the native species (N) and a ridge (right) for the denatured species (D). If we compare the structure of the ridges shown in Figure 14 with those shown in Figure 11, we see that for the real system (Figure 11) the molecule follows the N-ridge and then, at the transition temperature, it moves over to the D-ridge at high temperature.

Finally, we compare the approximate function ϕ^* given by eq 40 with the functions ϕ_N and ϕ_D for the independent branches. The results are shown in Figures 15, 16, and 17, respectively, for the temperatures $T = 325$, T_m, and 350 K. We recall from the results shown in Figure 9 that the ϕ^* and ϕ functions are essentially identical. Figures 15−17 show that the

dx.doi.org/10.1021/jp211794u | J. Phys. Chem. B 2012, 116, 6683−6693

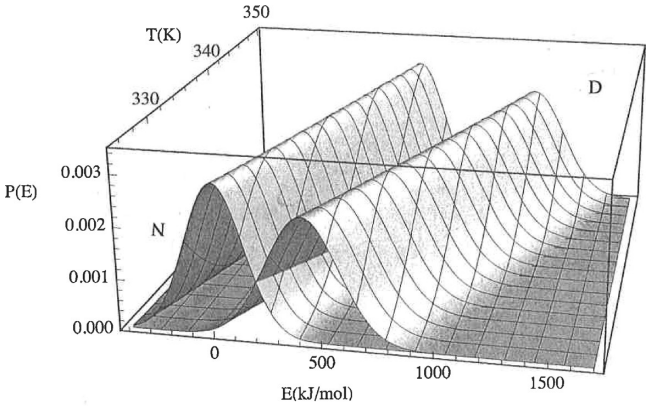


Figure 14. Three-dimensional plots of the independent functions $P_N(E,T)$ and $P_D(E,T)$ given by eq 43. The ridge on the left represents the native species (N), while the ridge on the right represents the denatured species (D).

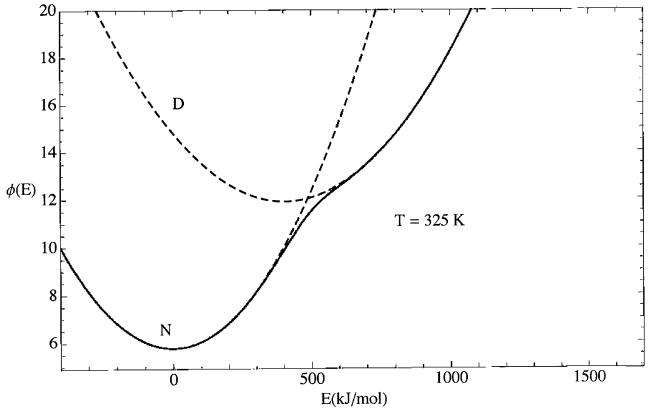


Figure 15. The function $\phi(E)$ (solid line) at T = 325 K as given by ϕ^* of eq 40 compared with $\phi_N(E)$ and $\chi(E)$ as given by eq 41.

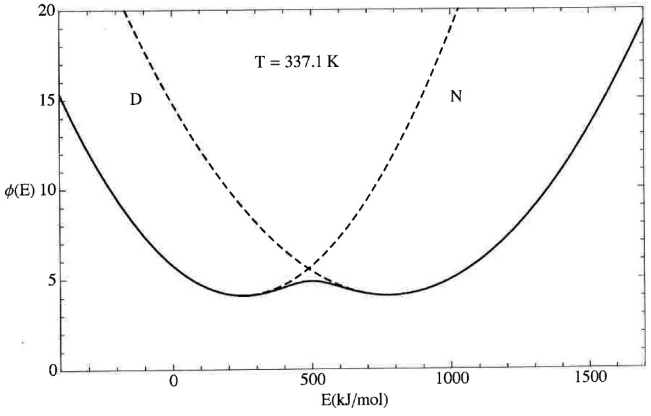


Figure 16. The function $\phi(E)$ (solid line) at T = 337.1 K as given by ϕ^* of eq 40 compared with $\phi_N(E)$ and $\phi_D(E)$ as given by eq 41.

two functions ϕ_N and ϕ_D reproduce the behavior of the ϕ (or ϕ^*) function.

Two-Species Thermodynamics. We have seen in the previous section that it is a good approximation to interpret the denaturation thermodynamics of barnase in terms of two qualitatively different species, the native (N) and denatured (D),

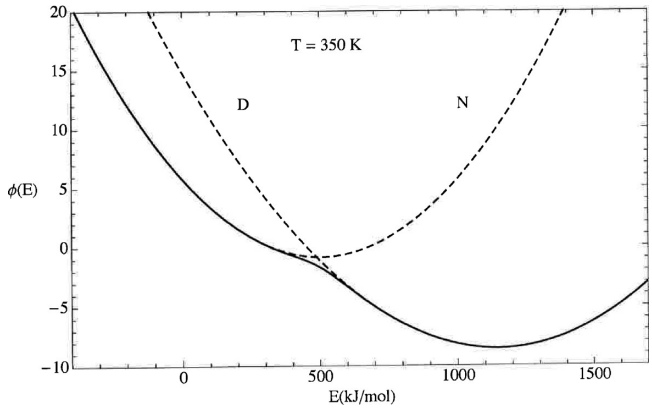


Figure 17. The function $\phi(E)$ (solid line) at T = 350 K as given by ϕ^* of eq 40 compared with $\phi_N(E)$ and $\phi_D(E)$ as given by eq 41.

as indicated in eq 36. Accepting this approximation, one can treat the system as a reaction between the two species

$$N \rightleftarrows D \tag{44}$$

We can then use the $\phi_N(E)$ and $\phi_D(E)$ functions given in eq 41 to give the partition functions for the respective species

$$Q_N = \int_E \exp[-\phi_N(E)]\, dE$$
$$Q_D = \int_E \exp[-\phi_D(E)]\, dE \tag{45}$$

The probabilities of the two species are then given by the standard relations

$$P_N = \frac{Q_N}{Q_N + Q_D} \qquad P_D = \frac{Q_D}{Q_N + Q_D} \tag{46}$$

These quantities are shown as a function of temperature in Figure 18. Given P_N and P_D, one can calculate the equilibrium constant for the reaction given in eq 44

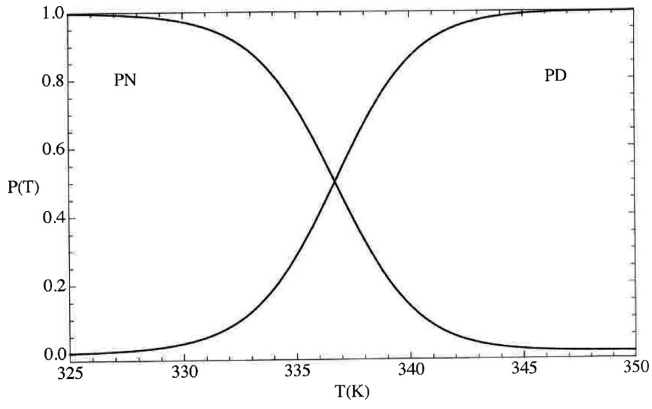


Figure 18. The two-state probabilities of the native (N) and denatured (D) forms of barnase as a function of temperature as determined from eq 46.

$$K = P_D/P_N = Q_D/Q_N \tag{47}$$

This quantity is illustrated as a function of temperature in Figure 19. At the midpoint of the reaction where $P_N = P_D = 1/2$, we of course have $K = 1$.

dx.doi.org/10.1021/jp211794u | J. Phys. Chem. B 2012, 116, 6683−6693

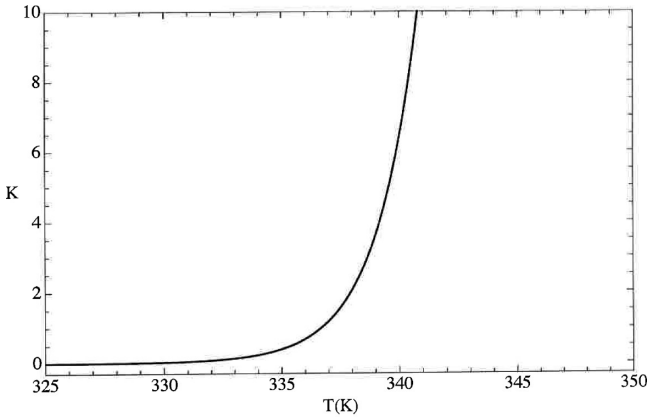


Figure 19. The two-state equilibrium constant K for the reaction N-to-D for barnase as given by eq 47.

The main thermodynamic quantities for the reaction of eq 44 are obtained from K by the standard relations

$$\Delta A = -RT \ln K$$
$$\Delta U = RT^2 \partial \ln K/\partial T$$
$$\Delta S = (\Delta U - \Delta A)/T \quad (48)$$

The quantity ΔA is shown as a function of temperature in Figure 20 where the dashed curve indicates the locus of $\Delta A = 0$

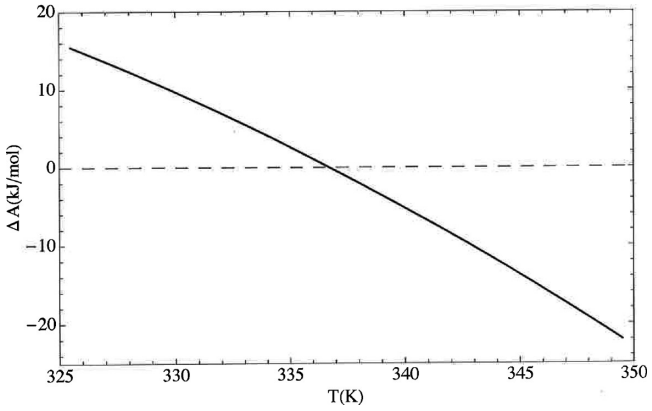


Figure 20. The two-state ΔA as a function of temperature for the reaction N-to-D for barnase as given by eq 48.

which is the half-point of the reaction. Figure 21 shows the temperature dependence of ΔU and ΔS. One notes that all three of these functions, ΔA, ΔH, and ΔS, are almost simple linear functions of temperature.

The functions shown in Figures 20 and 21 are the difference in the respective quantities for the reaction of eq 44. Given Q_N and Q_D, we can also calculate the thermodynamics for each species N and D using the analogues of eq 48, where K is replaced by Q_N or Q_D.

Taking the internal energy as an example, we have

$$U_N = RT^2 \partial \ln Q_N/\partial T$$
$$U_D = RT^2 \partial \ln Q_D/\partial T \quad (49)$$

ΔU for the reaction of eq 44 is then given as the difference in these quantities

$$\Delta U = U_D - U_N \quad (50)$$

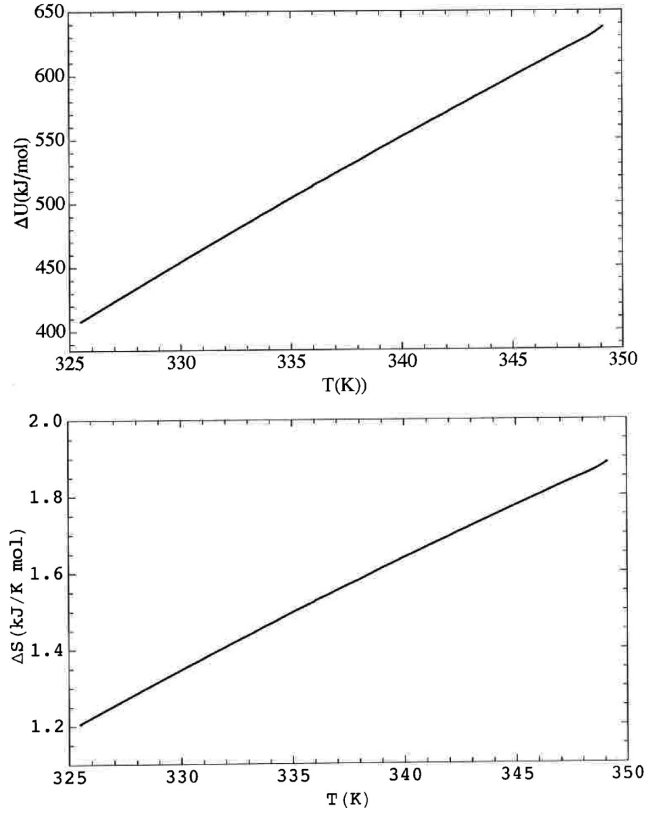


Figure 21. The two-state ΔU and ΔS as functions of temperature for the reaction N-to-D for barnase as given by eq 48.

Finally, the average internal energy is given by

$$\langle U \rangle = P_N U_N + P_D U_D \quad (51)$$

Analogous relations can be given for the free energy and the entropy. The thermodynamics obtained in this manner should be the same as that obtained from $C(T)$ data, as given in eq 23.

In Figure 22, the dashed curves give A_N and A_D, while the solid curve gives $\langle A \rangle$. Figure 23 gives the analogous quantities for the internal energy and the entropy. The solid curves in Figures 22 and 23 are essentially identical with the same quantities obtained from the heat capacity, as shown in Figures 3 and 4.

Two-Species Approximation: Unimodal and Bimodal Distributions. We have seen that for the protein barnase the functions $P(E)$ and $\phi_m(E)$ exhibit bimodal behavior, as seen respectively in Figures 5 and 7. We could then interpret this behavior in terms of a two-species model, as indicated in eqs 36 and 44. We have previously shown that all proteins do not exhibit such bimodal behavior.[7] To illustrate this feature, we will take the example of the DNA binding domain of the interferon regulatory factor-1 (DBD1 for short).

The temperature dependence of the heat capacity of DBD1 as determined in the laboratory of Prof. Privalov[21] is shown in Figure 24 and is qualitatively the same as that for barnase shown in Figure 3 in that $C(T)$ for both proteins exhibit a strong maximum as a function of temperature. However, when we construct the isotherm ϕ_m at T_m, for this protein, as shown in Figure 25 (where T_m is the temperature of the $C(T)$ maximum, T_m = 320.2 K), it exhibits unimodal behavior. This is in contrast to the same function for barnase shown in Figure 7. This indicates that at T_m this species is not divided into two

dx.doi.org/10.1021/jp211794u | J. Phys. Chem. B 2012, 116, 6683–6693

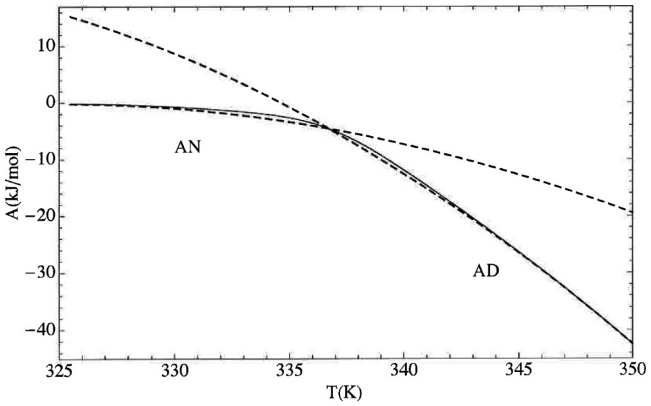


Figure 22. The quantities A_N and A_D (dashed lines) calculated from the two-state Q_N and Q_D. The solid curve gives the average A (analogue of eq 51).

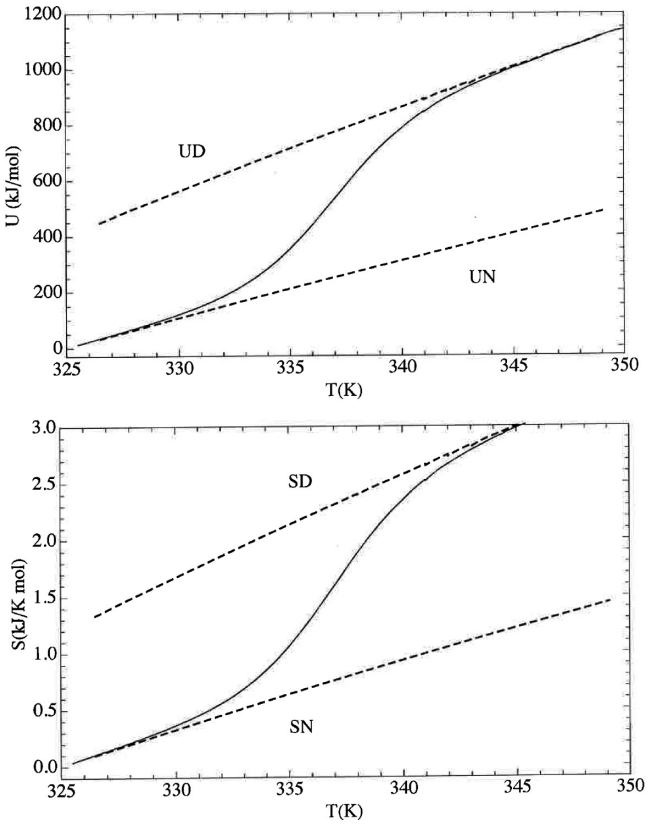


Figure 23. The upper graph gives the quantities U_N and U_D (dashed lines) as given by eq 49. The solid curve gives the average U as given by eq 51. The lower graph gives the quantities S_N and S_D (dashed lines) as given by the analogue of eq 49. The solid curve gives the average S (analogue of eq 51).

qualitatively different species. We obtain more detail by plotting $\phi(T,E)$ where the three-dimensional graph is shown in Figure 26, where $\phi(T,E)$ is given by eqs 30 and 31. We still refer to the low and high limits of the function as native (N) and denatured (D), but here there is no distinct barrier between the limits, as shown in Figure 12 for barnase. Rather, the function $\phi(T,E)$ is a broad open surface that evolves smoothly from N to D.

In Figure 27, we show isotherms of the function $P(E,T)$ for DBD1. This function exhibits one continuous ridge in going

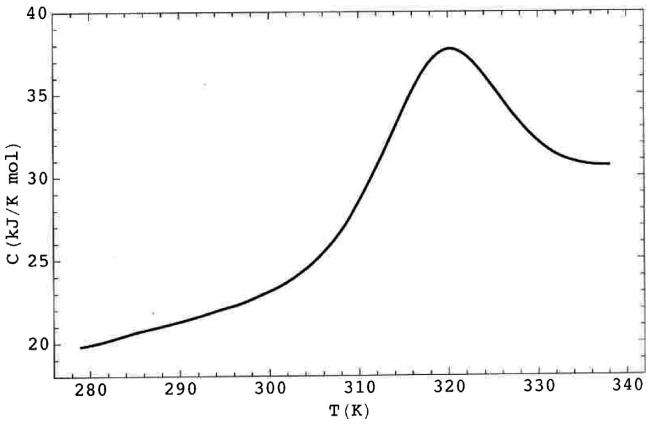


Figure 24. The heat capacity of DBD1 as a function of temperature.

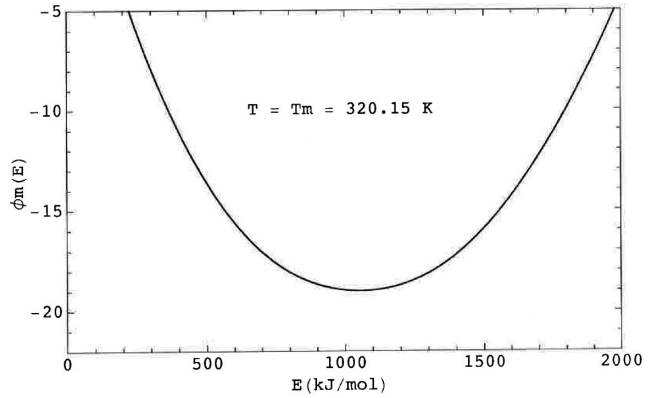


Figure 25. The function $\phi_m(E)$ at T_m = 320.2 K for DBD1.

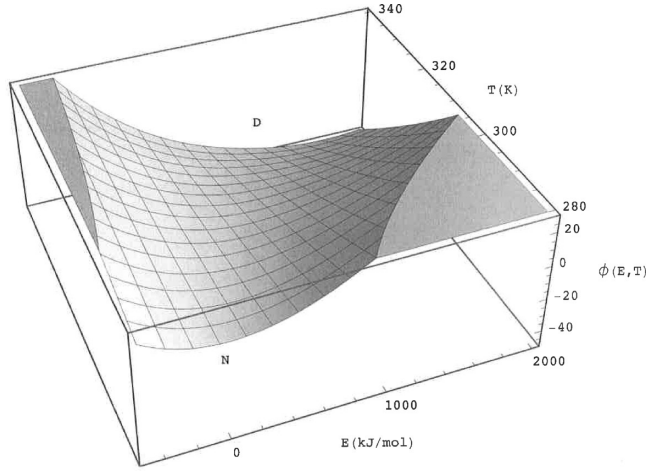


Figure 26. The function $\phi(E,T)$ for DBD1.

from N to D, which is quite different from the behavior of the corresponding function for barnase shown in Figure 11.

Clearly, a maximum in the heat capacity indicates that major changes are occurring in the internal energy (and hence also in the molecular conformation). The question then is why does one protein (barnase) exhibit bimodal behavior while another (DBD1) does not.

In a previous paper,[7] we have described the formal cause for this difference in behavior and we review that cause here. Our construction of the partition function for a protein begins with the temperature dependence of the heat capacity. Using a

dx.doi.org/10.1021/jp211794u | J. Phys. Chem. B 2012, 116, 6683–6693

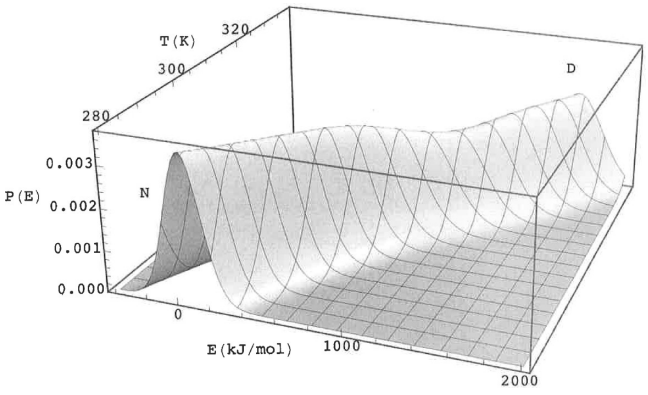


Figure 27. The function $P(E,T)$ for DBD1.

Taylor-series expansion about T_m, the temperature maximum of C, we have

$$C = C_0 + C_2\Delta T^2 \tag{52}$$

Since we are expanding about the maximum in the heat capacity, we have the condition $C_1 = 0$. Using the coefficients C_0 and C_2 in eq 5 gives four energy moments. We then use the maximum-entropy method to convert the energy moments into the coefficients λ_m of the function $\Gamma(E)$ given in eq 11. The condition for an extremum in $P(E)$ is

$$\partial P/\partial E = -(\partial\Gamma/\partial E)P = 0 \tag{53}$$

Since $P > 0$, the condition in eq 7.2 is also

$$\partial\Gamma/\partial E = 0 \tag{54}$$

Using four energy moments gives a function Γ that is quartic in E and a function $\partial\Gamma/\partial E$ that is cubic in E. If $\partial\Gamma/\partial E$ is cubic, then there will be three roots that satisfy eq 54. If all three roots are real, then $P(E)$ will exhibit two maxima separated by a single minimum, as seen in Figure 5 for barnase for the four-moment function calculated at T_m. This is the case of bimodal behavior. The point where one goes from the state of having three real roots to the state of having one real root with two imaginary roots is the point where one goes from bimodal to unimodal behavior. This is also the point where one has

$$\partial^2 P/\partial E^2 = 0 \tag{55}$$

The overall procedure then is as follows. One picks a value of C_0 at a fixed value of T and then varies C_2 (going through the maximum-entropy procedure) and finds the critical value C^*_2 where one goes from having three real roots to one real root. We introduce the variable (in terms of the coefficients in eq 52)

$$r = -C_2/C_0 \tag{56}$$

which is positive, since C_2 will be negative at T_m. We designate the critical value of r by $r_c = -C^*_2/C_0$. We find a very accurate universal empirical relation for r_c

$$r_c = k(T)C_0 \tag{57}$$

where

$$k(T) = [2.681 - (1.049)10^{-2}T + (1.151)10^{-5}T^2]/1000 \tag{58}$$

The condition that determines whether one has unimodal or bimodal behavior then is

$$r > r_c \quad \text{bimodal}$$
$$r < r_c \quad \text{unimodal} \tag{59}$$

The conditions of eq 59 determine the structure of a phase diagram for the occurrence of bimodal and unimodal behavior.

In Figure 28, we illustrate the conditions shown in eq 59 using the two proteins we have treated. The solid dots indicate

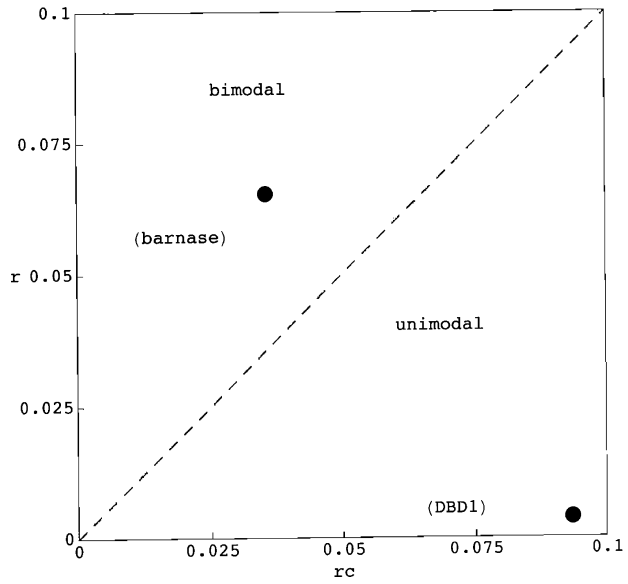


Figure 28. The phase diagram for bi- and unimodal behavior in $P(E)$.

the loci of the points (r_c, r) with the point for barnase clearly in the area for bimodal behavior and the point for DBD1 in the area for unimodal behavior. What we see clearly is that whether or not a molecule exhibits bimodal behavior is not a qualitative property of $C(T)$ but is a quantitative condition determined by the values of r and r_c which in turn are determined by the heat capacity coefficients C_0 and C_2 given in eq 52.

CONCLUSION

In the present paper, we have extended our previous work on using energy moments obtained from empirical heat capacity data to construct, using the maximum-entropy method, energy probability distributions for proteins. The extension involves using the distribution functions to calculate the energy degeneracy, $W(E)$, and hence the partition function as given in eq 1. Our model calculation of $W(E)$ for the 3-d harmonic oscillator, using the maximum-entropy method, as shown in Figures 1 and 2, shows that the method can be used to give an accurate representation of the degeneracy $W(E)$ and hence of the partition function $Q(E)$.

In this paper, we apply the same approach to the protein barnase. Using the quantity $\phi(E,T)$ defined in eqs 29 and 31, we can construct the partition function for barnase that is illustrated in Figure 10. In Figure 11, we show isotherms of the function $P(E)$, the probability of a molecule having energy E. In this figure, it is clear that there are two distinct regions of high probability that we naturally associate with the native and denatured states. This feature arises naturally from the function $\phi(E,T)$ which in turn is determined by the quartic nature of the energy distribution which is obtained from the temperature

dx.doi.org/10.1021/jp211794u | J. Phys. Chem. B 2012, 116, 6683–6693

dependence of the heat capacity. Thus, we are able to get an overall picture of the energy distribution of a protein, as illustrated in Figures 10 and 11.

AUTHOR INFORMATION

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Poland, D.; Scheraga, H. A. *Theory of Helix-Coil Transitions in Biopolymers*; Academic Press: New York, 1970.
(2) von Dreele, P. H.; Poland, D.; Scheraga, H. A. *Macromolecules* **1971**, *4*, 396−407.
(3) Wojcik, J.; Altmann, K.-H.; Scheraga, H. A. *Biopolymers* **1990**, *30*, 121−134.
(4) Poland, D. *J. Chem. Phys.* **2000**, *112*, 6554−6562.
(5) Poland, D *J. Chem. Phys.* **2000**, *113*, 4774−4784.
(6) Poland, D *J. Chem. Phys.* **2000**, *113*, 9930−9941.
(7) Poland, D. *Biopolymers* **2001**, *58*, 89−105.
(8) Poland, D. *Proteins: Struct., Funct., Genet.* **2001**, *45*, 325−336.
(9) Poland, D. *Biopolymers* **2002**, *63*, 59−65.
(10) Poland, D. *J. Protein Chem.* **2001**, *21*, 187−194.
(11) Poland, D. *Biophys. Chem.* **2002**, *101*, 485−495.
(12) Poland, D. *Methods Enzymol.* **2004**, *383*, 233−465.
(13) Poland, D. *Biopolymers* **2006**, *81*, 127−135.
(14) Poland, D. *Biophys. Chem.* **2007**, *125*, 497−507.
(15) Hill, T. L. *An Introduction to Statistical Thermodynamics*; Dover: New York, 1986.
(16) Tagliani, A. *J. Math. Phys.* **1983**, *34*, 326−334.
(17) Mead, L. R.; Papanicolaou, N. *J. Math. Phys.* **1984**, *25*, 2404−2413.
(18) Griko, Y. V.; Makhatadze, G. I.; Privalov, P. L.; Hartley, R. W. *Protein Sci.* **1994**, *3*, 669−676.
(19) Makhatadze, G. J.; Privalov, P. L. *Adv. Protein Chem.* **1995**, *47*, 307.
(20) Freire, E.; Biltonen, R. L. *Biopolymers* **1978**, *17*, 463−479.
(21) Hargreaves, V. V.; Makeyeva, E. N.; Dragan, A. I.; Privalov, P. L. *Biochemistry* **2005**, *44*, 14202−14211.