A Theoretical Model for the Folding/Unfolding Thermodynamics of Single-Domain Proteins, Based on the Quasi-Gaussian Entropy Theory

Danilo Roccatano,[†] A. Di Nola,[‡] and Andrea Amadei*[,§]

School of Engineering and Science, International University Bremen, Campus Ring 1, D-28725 Bremen, Germany, Dipartimento di Chimica Università di Roma "La Sapienza" P.le A. Moro 5, 00185 Roma, Italy, and Dipartimento di Scienze e Tecnologie Chimiche Università di Roma "Tor Vergata", via della Ricerca Scientifica 1, I-00133 Roma, Italy

Received: November 25, 2003; In Final Form: February 19, 2004

The quasi-Gaussian entropy (QGE) theory was used to formulate a statistical mechanical model describing the thermodynamics of the folding/unfolding process of single-domain proteins. The model was parametrized using experimental data obtained from differential scanning calorimetry (DSC) of a set of proteins. The results showed that the model is able to reproduce the experimental behavior in the usual temperature range, for all the analyzed proteins. Furthermore, a remarkable similarity of some parameters of the model, when normalized per residue and corresponding to well-defined physical properties, was found. Interestingly, at low temperature, the model provides cold denaturation features for all the proteins. Finally, a general description of the folding/unfolding process and stability, based on the physical view provided by the model, is discussed.

Introduction

Protein folding is one of the most fascinating problems of structural biology. The nature of this phenomenon is very elusive since it involves a subtle equilibrium of interactions apparently not simply related to the amino acid sequence.[1] Despite the microscopic complexity of the folding mechanism, the thermodynamics of the process, at least for single-domain proteins, is relatively simple, and it has been experimentally well-known for almost 30 years.[2]

For single-domain proteins, the thermodynamics of the folding/unfolding process is typically described by semiempirical models based on simple thermodynamical assumptions and approximations.[1−3] The simplicity of such models allows a macroscopic phenomenological description of the process and furnishes interesting information on the thermodynamic stability of proteins.[2] However, no connections with the underlying microscopic physical mechanism of protein stability and folding/unfolding equilibrium can be inferred from such models, and their semiempirical character cannot provide a real physical consistency. This implies that their predictions must be always be considered as a good approximation only in a limited temperature range. To overcome these problems and try to bridge the microscopic nature of the phenomenon with the macroscopic thermodynamics of the process, statistical mechanical models have also been proposed. These models are based on simplified molecular Hamiltonians (chain−chain Hamiltonians) used to describe the interactions in the system. More realistic Hamiltonians can be in principle used in atomistic simulations, but the high dimensionality of the configurational space and complexity of the corresponding energy surface for solvated proteins make simulations still unable to provide the complete thermodynamics of such systems. The use of simplified models was first proposed by Zimm and Bragg[4] to describe the helix−coil transition of α-helices forming peptides. Similar models have been used to describe the folding/unfolding kinetics and thermodynamics of α-helices and other peptides forming secondary structures in solution[5] (like *β*-hairpin) and single-domain proteins.[6−12] These kinds of models typically provide only a qualitative agreement with experimental data, as the idealized molecular Hamiltonians used oversimplify the physics of the proteins.

In this paper we use the framework of the quasi-Gaussian entropy (QGE) theory, based on enthalpy fluctuations in the isothermal−isobaric ensemble,[13] to formulate a general model for the description of the thermodynamic properties of the folding/unfolding of single-domain proteins. This statistical mechanical theory, which is basically an extension of the fluctuation theory, has provided in the past few years excellent models for homogeneous fluids,[13−18] solids,[19] and recently for the thermodynamics of flexible organic molecules in vacuo[20] and solutes.[21] Our approach has analogies with those proposed by Rosgen et al. and Linder and Kromhout,[11,12] although based on a completely different theoretical framework. The model we propose was applied to experimental denaturation heat capacity data for different proteins (SH3, cytochrome *c*, barnase, lysozyme) in different denaturating conditions.

Methods

The QGE model proposed in this paper was parametrized on the heat capacity data obtained from DSC experiments of a set of proteins reported in the literature: α-spectrin SH3,[22] cytochrome *c*,[3] barnase,[23] and egg white lysozyme.[3] The experimental calorimetric curves were kindly provided by Professor P. L. Privalov and Dr. F. Conejero-Lara. The curve fitting was performed by a graphical Unix tool (DSC modeler) specifically designed for this purpose. This program is available from the author (D.R.) by request.

* Author to whom the correspondence should be addressed. E-mail: andrea.amadei@uniroma2.it. Fax: +39-6-72594328.
[†] International University Bremen.
[‡] Università di Roma "La Sapienza".
[§] Università di Roma "Tor Vergata".

10.1021/jp037591a CCC: $27.50
Published on Web 04/09/2004

Theory

For a fluid state system of N solute molecules at high dilution, the canonical partition function can be expressed as[24]

$$Q = \frac{(8\pi^2 V)^N}{N!}(\Theta \sum_l \int^* \mathrm{e}^{-\beta \mathscr{U}_l} \mathrm{d}\boldsymbol{\xi}\, \mathrm{d}\mathbf{x}\, \mathrm{d}\boldsymbol{\pi}\, \mathrm{d}\mathbf{p})^N \quad (1)$$

$$\Theta^{-1} = n_s!(1+\gamma)(1+\gamma_s)^{n_s} h^{(d+d_s)} \quad (2)$$

where the summation runs over the quantum vibronic states, $\mathscr{U}_l$ is the total energy, including the quantum vibronic energy, of the subsystem defined by a single solute molecule and n_s solvent molecules, V is the overall volume of the system, $\boldsymbol{\xi}$ are the generalized internal coordinates of a single solute molecule with fixed rototranslational coordinates, $\boldsymbol{\pi}$ are the conjugated momenta of the solute molecule, and $\mathbf{x}$, $\mathbf{p}$ are the coordinates and conjugated momenta of the n_s solvent molecules within the solute molecular volume. Moreover, $1+\gamma$ and $1+\gamma_s$ are the symmetry coefficients for the solute and the solvent respectively, d and d_s are the number of classical degrees of freedom in the solute and solvent molecules of the subsystem, and h is the Planck's constant. Finally, the integral is taken within the solute molecular volume $V_m = V/N$, and the asterisk denotes an integration only over the accessible configurational space. From the previous equation it follows, using the approximation $N! \cong N^N \mathrm{e}^{-N}$, that the whole partition function can be obtained from the solute molecular partition function by $Q = Q_m^N$

$$Q_m = \frac{8\pi^2 V_m \Theta}{\mathrm{e}^{-1}} \sum_l \int^* \mathrm{e}^{-\beta \mathscr{U}_l} \mathrm{d}\boldsymbol{\xi}\, \mathrm{d}\mathbf{x}\, \mathrm{d}\boldsymbol{\pi}\, \mathrm{d}\mathbf{p} \quad (3)$$

Hence, the whole thermodynamics is defined by Q_m as $A = -NkT \ln Q_m$. This clearly means that if we want to describe the thermodynamics of the same system using the isobaric ensemble we must use a solute molecular isobaric partition function defined as

$$\Xi_m = \int \mathrm{e}^{-\beta p V_m} Q_m(\beta, V_m) \frac{\mathrm{d}V_m}{v} \quad (4)$$

providing $G = -NkT \ln \Xi_m$ (note that v is an arbitrary volume constant necessary to make a dimensional Ξ_m). Following the theoretical scheme described in a previous article,[13] we then have

$$\beta G(\beta) - \beta_0 G(\beta_0) = -N \ln\left\{\frac{\Xi_m(\beta)}{\Xi_m(\beta_0)}\right\} = -N \ln \langle \mathrm{e}^{-\Delta\beta \mathscr{H}} \rangle_{\beta 0} \quad (5)$$

with $\Delta\beta = \beta - \beta_0$, the subscript β_0 indicating an average in the β_0 ensemble and $\mathscr{H} = \mathscr{U}_l + pV_m$. Note that $G/N = \mu + n_s\mu_s = -kT \ln \Xi_m$ with μ and μ_s the solute and solvent chemical potentials. The ensemble average in eq 5 can be expressed as

$$\langle \mathrm{e}^{-\Delta\beta \mathscr{H}} \rangle_{\beta 0} = \int \rho(\mathscr{H})\, \mathrm{e}^{-\Delta\beta \mathscr{H}} \mathrm{d}\mathscr{H} \quad (6)$$

where $\rho(\mathscr{H})$ is the enthalpy probability distribution function in the β_0 ensemble. Instead of using a perturbation expansion, in the QGE theory the free energy is obtained by modeling the distribution function and hence its moment generating function[25,26] or Laplace transform, eq 6. For homogeneous fluid state systems it was shown that a rather good model in the isobaric ensemble is the diverging gamma state model[13] for enthalpy fluctuations. In the present paper where we deal with a very complex system involving a macromolecule, it is likely that we need more sophisticated models. In recent articles[20,27] we showed, for the canonical ensemble, that the use of mixing distributions for gamma state models provides a very powerful method to obtain more sophisticated and accurate models for fluid state systems. We can use a similar approach in the present case, assuming that the solute molecular configurational space of the internal coordinates can be partitioned into a set of L subspaces, each one defining a solute−solvent system exactly described by a "local" diverging gamma state (note that pure water thermodynamics, along an isobar, is well described over a wide temperature range by a single diverging gamma state). We can rewrite the total free energy change as

$$\beta G(\beta) - \beta_0 G(\beta_0) = -N \ln \sum_{i=1}^{L} \frac{\Xi_{m,i}(\beta)}{\Xi_{m,i}(\beta_0)} \epsilon_i = -N \ln\{\sum_{i=1}^{L} \epsilon_i\, \mathrm{e}^{-[n_s \Delta(\beta\mu_s) + \Delta(\beta\mu_i)]}\} \quad (7)$$

$$\Delta(\beta\mu_s) = \beta\mu_s(\beta) - \beta_0\mu_s(\beta_0) \quad (8)$$

$$\Delta(\beta\mu_i) = \beta\mu_i(\beta) - \beta_0\mu_i(\beta_0) \quad (9)$$

where $\Xi_{m,i}$ is the partition function corresponding to the ith conformation, μ_i is the chemical potential of the solute in the ith conformation and

$$\epsilon_i = \frac{\Xi_{m,i}(\beta_0)}{\Xi_m(\beta_0)} = \mathrm{e}^{-\beta_0(\mu_{0,i} - \mu_0)} \quad (10)$$

with $\mu_0 = \mu(\beta_0) = -kT_0 \ln\Xi(\beta_0) - n_s\mu_s(\beta_0)$ the solute chemical potential at the reference temperature $T_0 = 1/(k\beta_0)$. Note that in eq 10 we use the fact that at high dilution the solvent molecular partial properties are identical to the pure solvent ones (hence independent of the solute). Within the assumption that each $n_s\mu_s + \mu_i$ can be well modeled by a "local" diverging gamma state, we have[13,21]

$$\mu_i = h_{0,i} - T_0 c_{p0,i} + T(c_{p0,i} - s_{0,i}) + T c_{p0,i} \ln \frac{T_0}{T} \quad (11)$$

with $h_{0,i}$, $c_{p0,i}$, and $s_{0,i}$ the partial molecular enthalpy, heat capacity, and entropy of the solute in the ith conformation at the reference temperature T_0. From the other general equations of the diverging gamma state properties we can also obtain all the other partial molecular properties of the solute, for example, the enthalpy h_i and heat capacity $c_{p,i}$

$$h_i = \left(\frac{\partial \beta\mu_i}{\partial \beta}\right)_{p,n_s} = h_{0,i} + (T - T_0) c_{p0,i} \quad (12)$$

$$c_{p,i} = \left(\frac{\partial h_i}{\partial T}\right)_{p,n_s} = c_{p0,i} \quad (13)$$

(note that in this paper, for sake of simplicity, we always omit in the partial derivatives the notation for the fixed number of solute molecules). In the present case where we deal with complex macromolecules such as proteins at fixed pressure, the summation in eq 7 is likely to involve a very large number of gamma states corresponding to different protein configurational subspaces (conformations). Hence, in order to keep the mathematical derivations and especially the model application

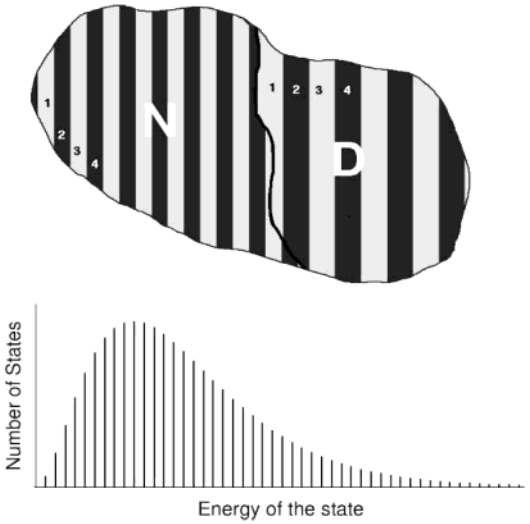

Figure 1. Simple picture of the configurational space partitioned into subspaces corresponding to the gamma states. In the figure is also shown the distribution of the enthalpy gaps of the subspaces.

manageable, we must use drastic simplifications. We will first assume that we can decompose the huge number of gamma states into two subgroups: one associated with the folded state of the protein and one with the unfolded state. Moreover, we will also assume that within each subgroup the partial molecular heat capacity is the same for the different gamma states and the partial molecular enthalpies are given by $h_{0,i} \cong h_0^0 + j\delta$ ($j = 0, 1, 2, 3, ...,$) with h_0^0 and δ the overall "ground state" enthalpy (at T_0) and enthalpy gap of the subgroup. We may define such kind of subgroup a gamma states family. In Figure 1 we give a simple picture of the gamma states partition used, where we indicate the native (N) and denatured (D) gamma states families. Hence, using the subscripts f and u to define the folded and unfolded state properties, respectively, we obtain from eqs 7−10

$$\Delta(\beta\mu) = \Delta\left(\beta\frac{G(T)}{N}\right) - n_s\Delta(\beta\mu_s) = -\ln\{e^{-(\beta\mu_f-\beta_0\mu_0)} + e^{-(\beta\mu_u-\beta_0\mu_0)}\}$$

$$e^{-(\beta\mu_f-\beta_0\mu_0)} = e^{-\Delta(\beta\mu^0_f)}\epsilon_f\langle e^{-\Delta\beta\delta_{fj}}\rangle_f \quad (14)$$

$$e^{-(\beta\mu_u-\beta_0\mu_0)} = e^{-\Delta(\beta\mu^0_u)}\epsilon_u\langle e^{-\Delta\beta\delta_{uj}}\rangle_f \quad (15)$$

$$\langle e^{-\Delta\beta\delta_{fj}}\rangle_f = \sum_{j=0}^{L_f-1} e^{-\Delta\beta\delta_{fj}}w_f(j) \quad (16)$$

$$\langle e^{-\Delta\beta\delta_{uj}}\rangle_u = \sum_{j=0}^{L_u-1} e^{-\Delta\beta\delta_{uj}}w_u(j) \quad (17)$$

$$w_f(j) = \frac{\epsilon_{f,j}}{\epsilon_f} \quad (18)$$

$$w_u(j) = \frac{\epsilon_{u,j}}{\epsilon_u} \quad (19)$$

(where ϵ_f, ϵ_u, μ_f, and μ_u are the total fractions and chemical potentials of the folded and unfolded subgroups and

$$\Delta(\beta\mu^0_f) = h^0_{0,f}\Delta\beta - c_{p0,f}T_0\Delta\beta - \frac{c_{p0,f}}{k}\ln\frac{T}{T_0} \quad (20)$$

$$\Delta(\beta\mu^0_u) = h^0_{0,u}\Delta\beta - c_{p0,u}T_0\Delta\beta - \frac{c_{p0,u}}{k}\ln\frac{T}{T_0} \quad (21)$$

From the previous equations we readily obtain the other partial molecular properties, that is, enthalpy, entropy, and heat capacity

$$h = \left(\frac{\partial\beta\mu}{\partial\beta}\right)_{p,n_s} = h_f + \chi(h_u - h_f) \quad (22)$$

$$\chi = \frac{e^{-\beta(\mu_u-\mu_f)}}{1 + e^{-\beta(\mu_u-\mu_f)}} \quad (23)$$

$$s = \frac{h-\mu}{T} \quad (24)$$

$$c_p = \left(\frac{\partial h}{\partial T}\right)_{p,n_s} = c_{p,u} - (1-\chi)(c_{p,u} - c_{p,f}) + (1-\chi)\chi\frac{(h_u - h_f)^2}{kT^2} \quad (25)$$

where obviously

$$h_f = \left(\frac{\partial\beta\mu_f}{\partial\beta}\right)_{p,n_s} \quad (26)$$

$$h_u = \left(\frac{\partial\beta\mu_u}{\partial\beta}\right)_{p,n_s} \quad (27)$$

$$c_{p,f} = \left(\frac{\partial h_f}{\partial T}\right)_{p,n_s} \quad (28)$$

$$c_{p,u} = \left(\frac{\partial h_u}{\partial T}\right)_{p,n_s} \quad (29)$$

are the corresponding partial molecular properties in the folded or unfolded state. Note that in the limit of a differential δ_f and δ_u a continuous gamma states partition of the solute intramolecular phase space is involved, as previously described for the canonical ensemble.[27] However, in this paper we will consider only the discrete-like diverging gamma states partition as it seems to provide a better general description of protein behavior. The use of a discrete-like diverging gamma states partition implies that the thermodynamics of a solvated protein should be a complex mixture between a typical fluid state behavior and a discrete-like "energy" fluctuation. In order to proceed further we must model the discrete probability distribution $w(j)$. A simple discrete distribution, which is physically acceptable and proved to be successful to model the quantum solid state, is the negative binomial distribution providing

$$\langle e^{-\Delta\beta\delta_{fj}}\rangle_f = \left\{\frac{1-q_f}{1-q_f e^{-\Delta\beta\delta_f}}\right\}^{Z_f} \quad (30)$$

$$\langle e^{-\Delta\beta\delta_{uj}}\rangle_u = \left\{\frac{1-q_u}{1-q_u e^{-\Delta\beta\delta_u}}\right\}^{Z_u} \quad (31)$$

where q and Z are two pure numbers characteristic of the negative binomial distribution. With these last equations we can express the partial molecular properties of the folded and unfolded states as

$$\beta\mu_f - \beta_0\mu_0 = h^0_{0,f}\Delta\beta - c_{p0,f}T_0\Delta\beta - \frac{c_{p0,f}}{k}\ln\frac{T}{T_0} - \ln\epsilon_f - Z_f\ln\left(\frac{1-q_f}{1-q_f\,e^{-\Delta\beta\delta_f}}\right)$$

$$= Z_f\left\{\frac{h^0_{0,f}}{Z_f}\Delta\beta - \frac{c_{p0,f}}{Z_f}T_0\Delta\beta - \frac{c_{p0,f}}{Z_f k}\ln\frac{T}{T_0} - \frac{1}{Z_f}\ln\epsilon_f - \ln\left(\frac{1-q_f}{1-q_f\,e^{-\Delta\beta\delta_f}}\right)\right\}$$

$$= -Z_f\ln\Omega_f$$

$$\beta\mu_u - \beta_0\mu_0 = h^0_{0,u}\Delta\beta - c_{p0,u}T_0\Delta\beta - \frac{c_{p0,u}}{k}\ln\frac{T}{T_0} - \ln\epsilon_u - Z_u\ln\left[\frac{1-q_u}{1-q_u\,e^{-\Delta\beta\delta_u}}\right]$$

$$= Z_u\left\{\frac{h^0_{0,u}}{Z_u}\Delta\beta - \frac{c_{p0,u}}{Z_u}T_0\Delta\beta - \frac{c_{p0,u}}{Z_u k}\ln\frac{T}{T_0} - \frac{1}{Z_u}\ln\epsilon_u - \ln\left(\frac{1-q_u}{1-q_u\,e^{-\Delta\beta\delta_u}}\right)\right\}$$

$$= -Z_u\ln\Omega_u$$

$$h_f = -Z_f\left(\frac{\partial\ln\Omega_f}{\partial\beta}\right)_{p,n_s} = h^0_{0,f} + (T-T_0)c_{p0,f} + \frac{Z_f q_f\delta_f}{e^{\Delta\beta\delta_f} - q_f}$$

$$h_u = -Z_u\left(\frac{\partial\ln\Omega_u}{\partial\beta}\right)_{p,n_s} = h^0_{0,u} + (T-T_0)c_{p0,u} + \frac{Z_u q_u\delta_u}{e^{\Delta\beta\delta_u} - q_u}$$

and so

$$c_{p,f} = c_{p0,f} + \frac{Z_f q_f k(\delta_f\beta)^2\,e^{-\Delta\beta\delta_f}}{(1-q_f\,e^{-\Delta\beta\delta_f})^2}$$

$$c_{p,u} = c_{p0,u} + \frac{Z_u q_u k(\delta_u\beta)^2\,e^{-\Delta\beta\delta_u}}{(1-q_u\,e^{-\Delta\beta\delta_u})^2}$$

$$s_f = \frac{h_f - \mu_f}{T}$$

$$s_u = \frac{h_u - \mu_u}{T}$$

Note that from the last equations we can interpret the folding and unfolding states thermodynamics as due to a set (Z_f and Z_u) of independent "modes" each defined by the "partition function" Ω_f or Ω_u. Each folded or unfolded "mode" corresponds to an independent type of enthalpy fluctuation and can be excited providing a thermodynamic transition of the solvent−solute system. Interestingly, such independent thermodynamic "modes" should correspond to basic sets of conformational subspaces which, via combination, provide all the possible gamma state conformational subspaces of the system.

We can simplify further the model assuming that

$$q_f = q_u = q \tag{32}$$

$$\delta_f = \delta_u = \delta \tag{33}$$

and taking the reference temperature T_0 as the equilibrium temperature, that is, $\epsilon_f/\epsilon_u = 1$. With these simplifications we obtain

$$\beta(\mu_u - \mu_f) = (h^0_{0,u} - h^0_{0,f})\Delta\beta - (c_{p0,u} - c_{p0,f})T_0\Delta\beta - \frac{(c_{p0,u} - c_{p0,f})}{k}\ln\frac{T}{T_0} - (Z_u - Z_f)\ln\left\{\frac{1-q}{1-q\,e^{-\Delta\beta\delta}}\right\} \tag{34}$$

$$h_u - h_f = h^0_{0,u} - h^0_{0,f} + (c_{p0,u} - c_{p0,f})(T-T_0) + \frac{(Z_u - Z_f)q\delta}{e^{\Delta\beta\delta}} \tag{35}$$

$$c_{p,u} - c_{p,f} = c_{p0,u} - c_{p0,f} + \frac{(Z_u - Z_f)qk(\delta\beta)^2\,e^{-\Delta\beta\delta}}{(1-q\,e^{-\Delta\beta\delta})^2} \tag{36}$$

which can be used to obtain the solute partial molecular properties, for example, via eq 25 the partial molecular heat capacity. Finally, it must be remarked that we derived a multistate model and hence the use of two gamma states families, corresponding to the folded and unfolded conditions, is not equivalent to the usual "two states" approximation where two simple thermodynamic states describe completely the folding thermodynamics. However, the approximations introduced in the general theory to simplify the gamma states properties in each subgroup reduce the ability of the model to treat more complex systems where more than two gamma states families could be necessary. More general multistate models, although possible, are in practice very difficult to use as the number of physical parameters defining the model becomes very large.

TABLE 1: Number of Residues, pH Value Used for DSC, Equilibrium Temperature Used in the QGE Model, and the rmsd between Experimental and Theoretical Heat Capacity for Each Protein

protein	n	pH	T_0 (K)	rmsd (J K^{-1} mol^{-1})
SH_3	62	2.0	307.93	183
		2.5	318.73	170
		3.0	330.63	153
		3.5	336.23	174
		4.0	339.73	113
cytochrome *c*	104	4.5	345.58	552
barnase	108	5.5	328.00	751
lysozyme	129	2.5	341.28	983

Results

The theoretical model described in the theory section was applied to four different single-domain proteins: SH3, cytochrome *c*, barnase, and lysozyme (see Table 1). The model parameters were obtained by fitting experimental heat capacity data (DSC experiments) with the corresponding theoretical expression, via a multistep procedure where the folded and unfolded data were used separately. The simplified QGE model used in this paper utilizes the same number of parameters as the more sophisticated semiempirical models based on thermodynamic expansions (7 to 8 parameters). In Figure 2 we compare the experimental partial molar heat capacities of SH3, at three pH values, with the corresponding curves due to our theoretical model. In Figure 3 we also show the folding free energies, entropies, and enthalpies of the QGE model at the three pH values. Note that each folding free energy crosses the abscissa twice showing clearly that two equilibrium temperatures are present. The higher is the experimental equilibrium (folding) temperature we used in the model, the other, lower temperature

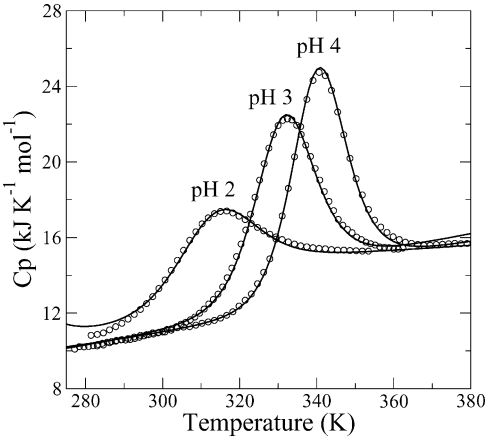

Figure 2. Experimental folding heat capacity (circles) and corresponding QGE model values (solid line) for SH3.

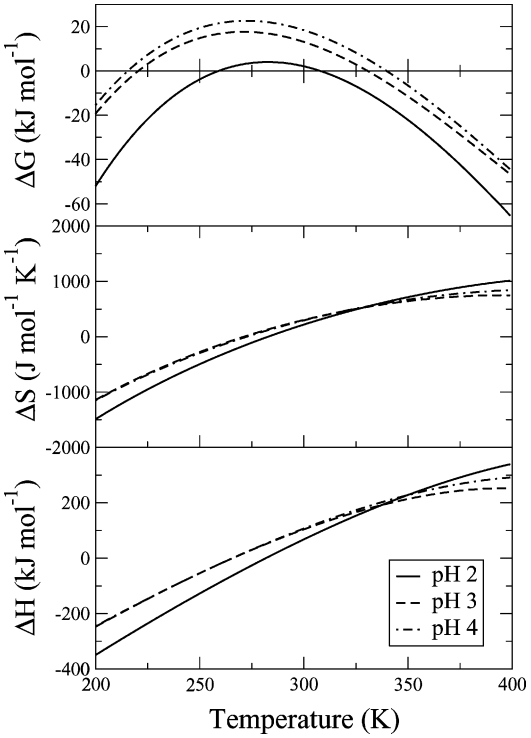

Figure 3. QGE model folding free energy, entropy, and enthalpy of SH3.

corresponds to the experimentally well-known cold denaturation.[28−30] A similar behavior was actually found for all the proteins studied, implying that within our model cold denaturation is a rather general feature, in agreement with the previous observations on folding thermodynamics. The same comparison, between experimental and model heat capacity, is shown for cytochrome *c* (Figure 4), barnase (Figure 5), and lysozyme (Figure 6).

For all the proteins the agreement between the experimental partial molar heat capacity and theoretical prediction is rather good in the whole temperature range. The corresponding root-mean-square deviations (rmsd), see Table 1, are comparable to the ones obtained using the most sophisticated semiempirical models, based on thermodynamic expansion, which utilizes the same number of parameters. These results suggest that the model used, although simplified, captures the essential features of the folding thermodynamics of the four proteins studied, and hence

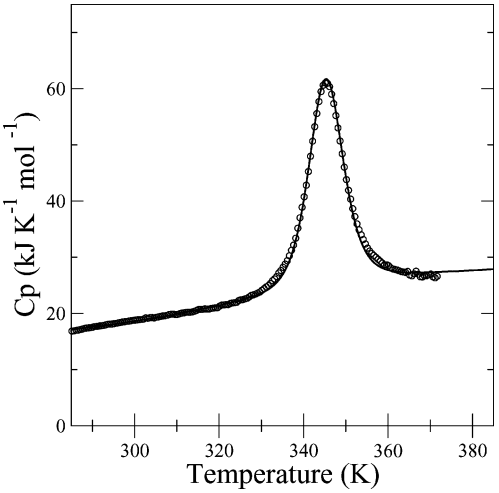

Figure 4. Experimental folding heat capacity (circles) and corresponding QGE model values (solid line) for cytochrome *c*.

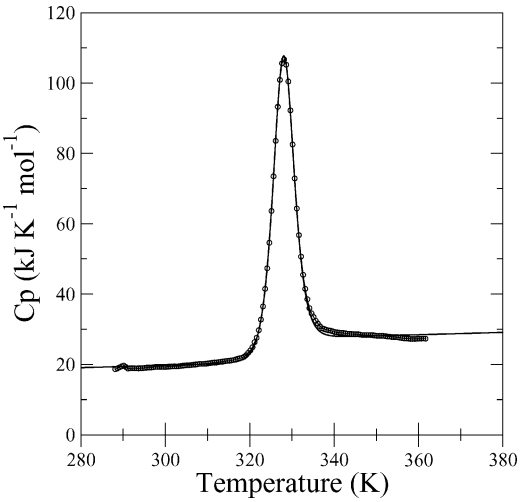

Figure 5. Experimental folding heat capacity (circles) and corresponding QGE model values (solid line) for barnase.

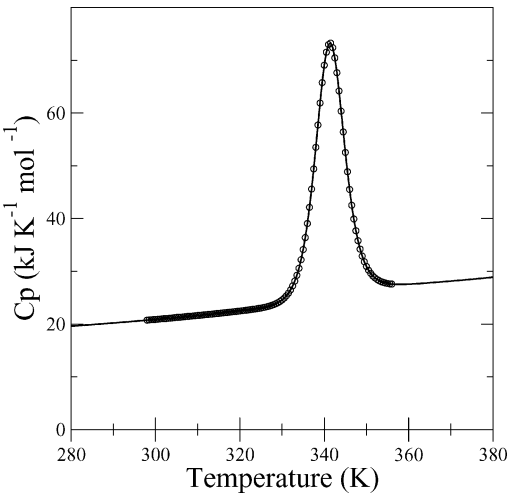

Figure 6. Experimental folding heat capacity (circles) and corresponding QGE model values (solid line) for lysozyme.

the folded or unfolded condition of these proteins can be well described by a single family of gamma states (more complex proteins could require the use of more gamma states families). The correspondence of the folded and unfolded conditions with

TABLE 2: Parameters Used in the QGE Models of the Different Proteins[a]'

protein	pH	ΔZ	Z_f	δ/n	$\Delta c_{p0}/n$	$c_{p0,f}/n$	$\Delta h_0^0/n$
SH_3	2.0	−165.09	265.13	9.74	114.21	88.98	11.81
	2.5	−165.09	351.73	10.56	118.72	49.90	13.42
	3.0	−165.09	351.73	11.92	108.94	40.55	15.43
	3.5	−165.09	351.73	11.92	105.90	47.00	15.67
	4.0	−165.09	351.73	11.92	105.90	50.23	15.90
cytochrome *c*	4.5	−162.36	172.31	14.52	166.58	80.60	18.55
barnase	5.5	−163.29	402.35	6.31	98.74	102.69	11.62
lysozyme	2.5	−161.26	270.31	7.87	76.99	86.52	11.38

[a]When indicated the property is per residue, i.e., the value is divided by the number of residues *n*. In the table we also give for each protein its pH value used for DSC. Note that δ (J/mol), $\Delta c_{p0} = c_{p0,u} - c_{p0,f}$ (J mol^{-1} K^{-1}), $\Delta h_0^0 = h_{0,u}^0 - h_{0,f}^0$ (kJ/mol), and $\Delta Z = Z_u - Z_f$. Finally, for all the proteins $q = 0.864$.

two large ensembles of simple thermodynamic states is actually in agreement with simulation results showing that the native protein configurational space can be subdivided into a low-dimensional subspace, characterized by large conformational fluctuations,[31] and a high-dimensional one, defined by quasi-constrained degrees of freedom, where the folding or unfolding paths occur.[32] In Table 2 we provide the physical parameters defining the QGE models used. Interestingly, all the properties of the (negative binomial) distribution of the gamma states reference enthalpies ($h_{0,u}$, $h_{0,f}$), are quite size independent being either identical (as for q) or similar (as for Z_f and Z_u). On the other hand, all the parameters describing the thermodynamics of the gamma states ($h_{0,f}^0$, $h_{0,u}^0$, δ, $c_{p0,f}$, $c_{p0,u}$) are rather clearly correlated to the protein size, their values per residue (reported in Table 2) being rather similar for the different proteins. From the physical parameters defining the QGE models it is evident that for the single gamma state the unfolded condition is characterized by a higher enthalpy and entropy, the latter due to an increase of the heat capacity, while the contribution of the reference enthalpies distribution provides a lower enthalpy and entropy for the unfolded ensemble ($Z_u - Z_f < 0$). Such a conflicting behavior, also providing cold denaturation, is due to the fact that the high enthalpy−entropy unfolded gamma state subspaces are defined as the combination of fewer "modes" than the low enthalpy−entropy folded gamma state subspaces. The presence of these conformational "modes" suggests a hierarchical organization of the conformational−thermodynamical protein behavior: a set of combined "modes" generates the gamma state subspaces and these are grouped into two families corresponding to the folded and unfolded conditions. Hence, within this model, folding thermodynamics essentially emerges from the combination of the independent "modes" providing excitations from one gamma state to another. It is also worth noting that the enthalpy gap for the gamma states is only about 10 J/mol in all the studied proteins, implying that at physiological temperature such excitations are largely accessible.

Conclusions

In this paper we developed a new theoretical model for describing protein folding thermodynamics, based on the coupling of relatively simple QGE models, that is, the diverging gamma states for the enthalpy fluctuation in the isothermal−isobaric ensemble. Such a coupling is based on the assumption that protein configurational space can be partitioned into a very large set of subspaces, each well described by a single gamma state. Gamma states themselves are assumed to be grouped into two families (i.e., subgroups where the gamma states differ only for the reference enthalpy) each describing either the folded or the unfolded condition. The model distribution of the reference enthalpies we use implies that each gamma states family is determined by combinations of independent conformational "modes" describing all the possible thermodynamic transitions for the protein−solvent system. The results obtained for the set of proteins studied show that such a model is rather accurate in reproducing the experimental behavior and hence captures some of the essential features of protein folding. The corresponding thermodynamics essentially emerges from the hierarchical organization of the gamma states subspaces, defined by a specific type of discrete distribution (negative binomial) for the reference enthalpies. Such a picture seems to point out that proteins could be a peculiar mix of fluid and solid state behavior where the complexity of the conformational space results in a relatively simple thermodynamics. Finally, comparison of the QGE models for the different proteins used shows a rather homogeneous folding behavior: the gamma states properties are essentially defined by the number of residues involved while the properties of the distribution of the reference enthalpies seems to be a molecular property which cannot be reduced to the single residue behavior.

Acknowledgment. We thank Professor P. Privalov who kindly provided us with the experimental calorimetric data of the lysozyme, barnase, and cytochrome *c* proteins and Dr. F. Conejero-Lara for the SH3 data.

References and Notes

(1) Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*; W. H. Freeman and Company: New York, 2000.
(2) Creighton, T. E. *Protein Folding;* Freeman: San Francisco, CA, 1992.
(3) Privalov, P. L.; Khechinashvili, N. N. *J. Mol. Biol.* **1974**, *86*, 665−684.
(4) Parthasarathy, R.; Chaturvedi, S.; Go, K. *Prog. Biophys. Mol. Biol.* **1995**, *64*, 1−54.
(5) Muñoz, V. *Curr. Opin. Struct. Biol.* **2001**, *11*, 212−216.
(6) Zwanzig, R. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9801−9804.
(7) Micheletti, C.; Banavar, J. R.; Maritan, A.; Seno, F. *Phys. Rev. Lett.* **1999**, *82*, 3372−3375.
(8) Galzitskaya, O. V. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *96*, 11311−11316.
(9) Bakk, A.; Høye, J. S.; Hansen, A.; Sneppen, K.; Jensen, M. H. *Biophys. J.* **2000**, *79*, 2722−2727.
(10) Bakk, A.; Hansen, A.; Sneppen, K. *Physica A* **2001**, *291*, 60−70.
(11) Linder, B.; Kromhout, R. A. *J. Phys. Chem. B* **1999**, *103*, 10325−10330.
(12) Linder, B.; Kromhout, R. A. *J. Phys. Chem. B* **2001**, *105*, 6387−6395.
(13) Amadei, A.; Apol, M. E. F.; Berendsen, H. J. C. *J. Chem. Phys.* **1998**, *109*, 3004−3016.
(14) Amadei, A.; Apol, M. E. F.; Berendsen, H. J. C. *J. Chem. Phys.* **1997**, *106*, 1893−1912.
(15) Apol, M. E. F.; Amadei, A.; Berendsen, H. J. C. *Chem. Phys. Lett.* **1996**, *256*, 172−178.
(16) Roccatano, D.; Amadei, A.; Apol, M. E. F.; Di Nola, A.; Berendsen, H. J. C. *J. Chem. Phys.* **1998**, *109*, 6358−6363.
(17) Amadei, A.; Apol, M. E. F.; Chillemi, G.; Berendsen, H. J. C.; Di Nola, A. *Mol. Phys.* **1999**, *96*, 1469−1490.
(18) Apol, M. E. F.; Amadei, A.; Berendsen, H. J. C. *J. Chem. Phys.* **1998**, *109*, 3017−3027.
(19) Apol, M. E. F.; Amadei, A.; Berendsen, H. J. C.; Di Nola, A. *J. Chem. Phys.* **1999**, *111*, 4431−4441.
(20) Iacono, A. A. B.; Grego, S.; Chillemi, G.; Apol, M. E. F.; Paci, E.; Delfini, M.; Di Nola, A. *J. Phys. Chem. B* **2001**, *105*, 1834−1844.
(21) Amadei, A.; Apol, M. E. F.; Di Nola, G. B. A. *J. Chem. Phys.* **2002**, *116*.
(22) Viguera, A. R.; Martínex, J. C.; Filimonov, V. V.; Mateo, P. L.; Serrano, L. *Biochemistry* **1994**, *33*, 2142−2150.
(23) Griko, Y.; Makhatadze, G.; Privalov, P. L.; Hartley, R. W. *Protein Sci.* **1994**, *3*, 669−676.

(24) Amadei, A.; Chillemi, G.; Ceruso, M. A.; Grottesi, A.; Di Nola, A. *J. Chem. Phys.* **2000**, *112*, 9−23.
(25) Patel, J. K.; Kapadia, C. H.; Owen, D. B. *Handbook of Statistical Distributions*; Marcel Dekker: New York, 1976.
(26) Stuart, A.; Ord, J. K. *Kendall's Advanced Theory of Statistics*, 5th ed.; Griffin: London, 1987; Vol. 1.
(27) Apol, M. E. F.; Amadei, A. *J. Phys. Chem. B* **2003**, *107*, 1410−1422.
(28) Privalov, P. Protein Folding. Creighton, T., Ed.; W. H. Freeman: New York, 1992; Chapter 3.
(29) Privalov, P. L.; Privalov, G. *Methods Enzymol.* **2000**, *323*, 31−62.
(30) Kunugi, S.; Tanaka, N. *Biochim. Biophys. Acta* **2002**, *1529*, 329−344.
(31) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct., Funct., Gen.* **1993**, *17*, 412−425.
(32) Daidone, I.; Amadei, A.; Roccatano, D.; Nola, A. D. *Biophys. J.* **2003**, *85*, 2865−2871.