Eur. Phys. J. B **6**, 157–161 (1998)

THE EUROPEAN PHYSICAL JOURNAL B

Statistical mechanics of warm and cold unfolding in proteins

A. Hansen[a], M.H. Jensen[b], K. Sneppen, and G. Zocchi

Niels Bohr Institute and NORDITA, Blegdamsvej 17, 2100, Denmark

Received: 19 March 1998 / Revised and Accepted: 25 May 1998

Abstract. We present a statistical mechanics treatment of the stability of globular proteins which takes explicitly into account the coupling between the protein and water degrees of freedom. This allows us to describe both the cold and the warm unfolding, thus qualitatively reproducing the known thermodynamics of proteins.

PACS. 87.10.+e General, theoretical, and mathematical biophysics (including logic of biosystems, quantum biology, and relevant aspects of thermodynamics, information theory, cybernetics, and bionics) – 05.70.Jk Critical point phenomena – 82.20.Db Statistical theories (including transition state)

The folded conformation of globular proteins is a state of matter peculiar in more than one respect. The density is that of a condensed phase (solid or liquid), and the relative position of the atoms is, on average, fixed; these are the characteristics of the solid state. However, solids are either crystalline or amorphous, and proteins are neither: the folded structure, while ordered in the sense that each molecule of a given species is folded in the same way, lacks the translational symmetry of a crystal. Unlike any other known solids, globular proteins are not really rigid, being able to perform large conformational motions while retaining locally the same folded structure. Finally, these are mesoscopic systems, consisting of a few thousand atoms.

Quantitatively, the peculiarities of this state of matter are perhaps best appreciated from the thermodynamics. Delicate calorimetric measurements [1–3] on the folding transition of globular proteins reveals the following picture: firstly the transition is first order, at least in the case of single domain proteins. Secondly, the stability of the folded state, *i.e.* the difference in Gibbs potential ΔG between the unfolded and the folded state is at most a fraction of kT_{room} per aminoacid. Following Privalov [3], we will refer to this property as "cooperativity". The Gibbs potential difference ΔG, as a function of temperature, is non monotonic: it has a maximum around room temperature (where $\Delta G > 0$ and so the folded form is stable), then crosses zero and becomes negative both for higher and lower temperatures. Correspondingly, the protein unfolds not only at high, but also at low temperatures. This phenomenon of "cold unfolding", which is observed experimentally, is most peculiar: solids usually do not melt upon cooling! For temperatures around the cold unfolding transition and below, the enthalpy difference ΔH between the unfolded and the folded state is negative; this means that cold unfolding proceeds with a release of heat (a negative latent heat), as is also observed experimentally; at the higher unfolding transition, on the contrary, $\Delta H > 0$ which corresponds to the usual situation of a positive latent heat. Figure 1 shows Privalovs measurements of the specific heat of myoglobin [3]. There are two peaks in the specific heat, corresponding to the two unfolding transitions, and a large gap ΔC in the specific heat between the unfolded and the folded state. This gap is again peculiar to proteins: usually, for a melting transition $\Delta C \approx 0$ (*e.g.* for ice at 0 °C, $C = 1.01$ cal/g K while for water at 0 °C, $C = 1.00$ cal/g K). The existence of this gap ΔC is related to the phenomenon of cold unfolding [3].

From the microscopic point of view, the main driving force for folding is the hydrophobic effect. In the native state of globular proteins hydrophobic residues are generally found on the inside of the molecule, where they are shielded from the water, while hydrophilic residues are typically on the surface. In the following we refer to the difference in free energy between hydrophobic residues interacting with each other in the core of the folded protein and these same residues interacting with the water in the unfolded structure, including any changes in the microscopic states of the water, as the "hydrophobic interaction". Hydrogen bonds within the regular elements of secondary structure (α helices and β sheets), while necessary for the stability of the native state, can hardly be thought of as providing the positive ΔG of the folded structure, since the unfolded structure would form just as many hydrogen bonds with the water. When the protein unfolds, the hydrophobic residues of the interior are exposed; this accounts for most of the gap in the specific heat ΔC [3], according to the known effect that dissolving hydrophobic

[a] *Permanent address:* Department of Physics, Norwegian University of Science and Technology, NTNU, 7034 Trondheim, Norway.

[b] e-mail: `mhjensen@nbi.dk`

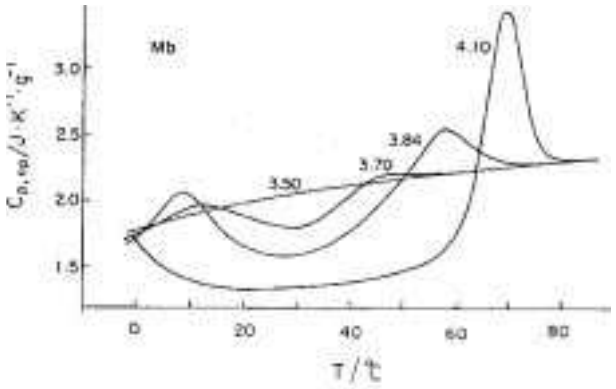


Fig. 1. Calorimetric measurements of the specific heat of Myoglobin at four different values of pH, as presented by Privalov in reference [3]. At sufficiently low pH the native structure of the protein never becomes stable, thus the protein remains in its unfolded structure with approximately constant heat capacity over the measured temperature range. By increasing pH the native structure becomes stabilized for intermediate temperatures, defining a transition to an unfolded state at both low and high temperatures, denoted respectively cold and warm denaturation. There is a gap in the specific heat between the folded and the unfolded states.

substances in water raises the heat capacity of the solution [4].

As in other branches of physics, once the thermodynamics of a system is known it is desirable to develop a corresponding statistical mechanics picture. Several models have been proposed which address some aspects of the folding transition. In the "zipper model" [5] , which was introduced to describe the helix–coil transition, the relevant degrees of freedom (conformational angles) are treated as a set of variables which can take two values: one corresponding to matching the ordered structure (helix), and the other corresponding to the "coil" state. The problem is then equivalent to the 1D Ising model. A related parametrization for the 3-d folding transition has been proposed by Zwanzig [6], describing the folding transition in terms of variables each of which awards match with the correct ground state. A zipper model that deals with the initial pathway of protein folding has been proposed by Dill, Fiebig and Chan [7]. For a review see [8]. A recent discussion of hydrophobicity in protein folding is in reference [9].

However, to our knowledge no model exists which reproduces all the thermodynamic features surveyed above. With the present work, we address this question.

We start with a Hamiltonian which we have recently introduced [10] to describe self-assembly of a cooperatively stabilized (in the sense defined above) structure:

$$H \;=\; -\,\mathcal{E}_0\,(\varphi_1 + \varphi_1\varphi_2 + \varphi_1\varphi_2\varphi_3 + \cdots + \varphi_1\varphi_2\cdots\varphi_N), \tag{1}$$

here the φ's are variables which take on the values 0 and 1, and, in the spirit of the zipper model, we define the ground state ($\varphi_i = 1\,\forall\, i$) as the template corresponding to the ordered, aperiodic structure, *i.e.* the folded state. One can think of the φ's as appropriately coarse-grained angle variables which define the conformation of the polypeptide chain. The above Hamiltonian is then a description of a system that has a specific folding pathway; a property that is well documented for proteins [11–15]. In particular, in terms of folding the first variable, φ_1, may be vieved as a nucleation center, whereas subsequent variables represent subsequent addition to structures onto it, much like what is suggested in the experimental analysis of folding of the protein Barnase from microseconds to seconds in reference [16]. The total number of steps in the subsequent folding process is presumably less than the number of amino acids, but a priori unknown. It is important to realize that if one parametrizes the folding with fewer steps N, each unit will be larger and energies and entropies appropriately increased (inversely proportional to N). Thus if one uses fewer steps N to parametrize the folding, the transition temperature for the system remains unchanged, only the sharpness of the transition will decrease.

As discussed in [10], this system has a first order phase transition from an ordered to a disordered state at temperature $T_m = 1/\ln 2$. The Hamiltonian (1) exhibits a hierarchical structure: if one of the variables $\varphi_1, \varphi_2, \ldots, \varphi_i$ equals zero, it does not matter what value the remaining variables $\varphi_{i+1}, \varphi_{i+2}, \ldots, \varphi_N$ assume. As a consequence, the system displays cooperativity, in the sense that the binding energy per degree of freedom in the ordered state, for $T \approx T_m$, is only of order kT_m.

In order to proceed further, it is necessary to take also the water into account. The relevant physics here is that dissolving a hydrophobic substance in water causes a large decrease in the entropy of the system [3]. This entropy change is attributed to a partial ordering of the water molecules around the hydrophobic solute. The gradual melting of this additional structure upon heating causes the increase in heat capacity. Consequently, we introduce a second set of variables $\mu_1, \mu_2, \ldots, \mu_N$ which describe the water. These water degrees of freedom couple to the unfolded protein degrees of freedom because these expose hydrophobic amino acids to the water. This is achieved by the Hamiltonian:

$$\begin{aligned} H = &-\mathcal{E}_0\,(\varphi_1 + \varphi_1\varphi_2 + \varphi_1\varphi_2\varphi_3 + \cdots + \varphi_1\varphi_2\cdots\varphi_N) \\ &+ (1-\varphi_1)\mu_1 + (1-\varphi_1\varphi_2)\mu_2 + \cdots \\ &+ (1-\varphi_1\varphi_2\cdots\varphi_N)\mu_N. \end{aligned} \tag{2}$$

The φ's take on the values 0 or 1, as before. Each of the μ_i's can take a value from the set $\{\mathcal{E}_{min} + s\Delta\mathcal{E}$, $s = 0, 1, 2, \ldots, g-1\}$ where $\mathcal{E}_{min} < 0$, $\Delta\mathcal{E} > 0$. If at least one of the variables $\varphi_1, \ldots, \varphi_i$ equals zero, the corresponding contribution of the i'th water variable to the energy is μ_i and zero otherwise. Therefore when $\varphi_1 \cdots \varphi_i = 1$, the states for the corresponding water degree of freedom μ_i are degenerate with zero contribution to the energy and degeneracy g. The equidistant energy levels reflect the experimentally observed approximate constant heat capacity at intermediate temperatures, whereas the finite number of levels g takes into account that protein-water interactions vanish above 120 °C.

To reiterate, the physical meaning of this Hamiltonian is that the water molecules in contact with an unfolded portion of the protein go to a lower entropy state (compared to the water molecules in contact with a folded portion), but also, for low temperatures, to a more tightly bound state. The more specific features of the model (2), *e.g.* the structure of the energy spectrum, the particular coupling of the μ's to the φ's, *etc.* can be varied while maintaining the overall thermodynamic behavior described below. Here we just present the case which is simplest to solve analytically.

The calculation of the partition function is straightforward. We parametrize the states of the system by the number n of consecutive matches $\varphi_1 = 1, \varphi_2 = 1, ..., \ \varphi_n = 1$ and ending with $\varphi_{n+1} = 0$ and the values $\{s_{n+1}, ..., s_N\}$ where each $s_i \in \{0, 1, 2, ..., g-1\}$ for the $(N-n)$ μ variables coupled to the unfolded portion of the protein. The energy of this state is

$$\varepsilon(n, s_{n+1}, ..., s_N) \ = \ -\, n\ \mathcal{E}_0 \ + \ \sum_{i=n+1}^{N} (\mathcal{E}_{min} + \Delta\mathcal{E}\ s_i) \tag{3}$$

where we have introduced the energy scale $\mathcal{E}_0$ for the protein variable in order to make the formulas dimensionally more transparent (up to now we used $\mathcal{E}_0 = 1$). Denoting $\beta = 1/T$ as the reciprocal temperature, the partition function is

$$\begin{aligned} Z = & \sum_{n=0}^{N-1} 2^{N-n-1} g^n \\ & \times \sum_{s_{n+1}=0}^{g-1} \sum_{s_{n+2}=0}^{g-1} \cdots \sum_{s_N=0}^{g-1} \exp(\ -\beta\varepsilon(n, s_1, \cdots, s_N)) \\ & +\ g^N \exp(\ \beta\mathcal{E}_0 N). \end{aligned} \tag{4}$$

In the above equation the factor 2^{N-n-1} is the degeneracy of the unfolded protein degrees of freedom and the factor g^n is the degeneracy of water which is not exposed to the inside of the protein. Factorizing the sums over s_i into partition functions Z_w for each water degree of freedom we write:

$$Z \ = \ \frac{1}{2}\ (2Z_w)^N \sum_{n=0}^{N-1} \left(\frac{g \exp(\beta\mathcal{E}_0)}{2Z_w} \right)^n + (g \exp(\ \beta\mathcal{E}_0))^N \tag{5}$$

where the phase space for a water degree of freedom exposed to an unfolded protein degree of freedom is

$$\begin{aligned} Z_w \ &= \ \sum_{s=0}^{g-1} \exp(-\beta(\mathcal{E}_{min} + s\Delta\mathcal{E})) \\ &= \ \frac{(\exp(-\beta\mathcal{E}_{min}) - \exp(-\beta\mathcal{E}_{max})}{(1 - \exp(-\beta\Delta\mathcal{E}))} \end{aligned} \tag{6}$$

where $\mathcal{E}_{max} = \mathcal{E}_{min} + g\Delta\mathcal{E}$. From equation (5) one sees directly that the state of the system is determined by the size of the quantity

$$\frac{g \exp(\beta\mathcal{E}_0)}{2Z_w} \ = \ \exp(\beta\Delta f). \tag{7}$$

If $\Delta f > 0$ then the system will be in the folded state because the sum in equation (5) is dominated by the last term, whereas for $\Delta f < 0$ the system will be unfolded.

The sum in equation (5) can be readily performed and the total partition function is

$$\begin{aligned} Z = & \frac{1}{2}(2Z_w)^N \ \frac{1 - (g \exp(\mathcal{E}_0\beta) \ / \ (2Z_w))^N}{1 - (g \exp(\mathcal{E}_0\beta) \ / \ (2Z_w))} \\ & + (g \exp(\ \beta\mathcal{E}_0))^N . \end{aligned} \tag{8}$$

The free energy is $F = -T \ln(Z)$, the energy $E = -\frac{d \ln(Z)}{d\beta}$ and the heat capacity $C \ = \ dE/dT$. Because there is no pressure in the model, the energy E takes the place of the enthalpy $H = E + pV$ and the free energy $F = E - TS$ takes the place of the Gibbs potential $G = H - TS$. In Figure 2 we show the heat capacity per degree of freedom for four different choices of $\mathcal{E}_{min}$, representing four different values of the chemical potential, which we discuss later. The characteristic feature is that there are two peaks corresponding to warm and cold unfolding, and a gap ΔC in the heat capacity between the unfolded and the folded form. At higher temperatures, *i.e.* $T > g\Delta\mathcal{E}$, the gap goes to zero because the water becomes effectively degenerate again. In Figure 3a we show the order parameter $\langle n \rangle$ as function of temperature. The figure indeed confirms that the protein is folded between the two transitions.

We now calculate explicitly the difference in the thermodynamic functions between the unfolded and the folded state. We consider these quantities per degree of freedom, denoted by small letters, *i.e.* $f = F/N$ *etc.* The thermodynamic functions associated to a folded (f) protein variable

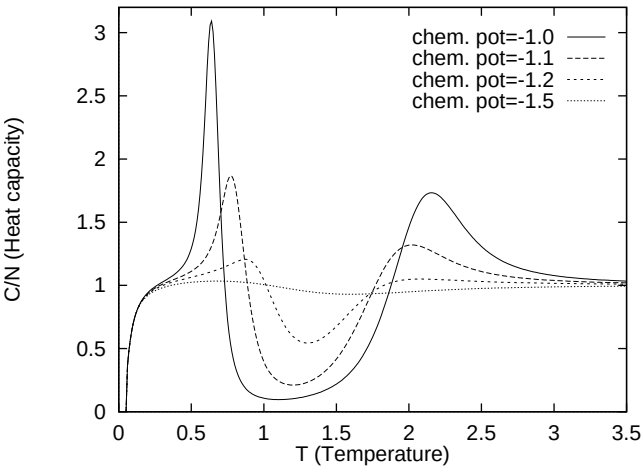


Fig. 2. Specific heat as function of temperature T for the model, with four different values of the chemical potential $\mu = -(\mathcal{E}_0 + \mathcal{E}_{min}) = -1.0, \ -1.1, \ -1.2, \ -1.5$ and with fixed level spacing $\Delta\mathcal{E} = 0.2$, $g = 35$ and system size $N = 60$. As the chemical potential is lowered, it becomes increasingly difficult to fold, and finally for sufficiently low μ the protein stays unfolded.

is the energy $e_f = -\mathcal{E}_0$, the entropy $s_f = \ln(g)$ and the free energy $f_f = -\mathcal{E}_0 - T\ln(g)$. The free energy associated to an unfolded (u) protein variable is given by the corresponding partition function of water multiplied by the degeneracy factor of an unfolded part of the protein: $f_u = -T\ln(Z_w\ 2)$. The difference in free energy between folded and unfolded state is accordingly

$$\Delta f \ = \ f_u \ - \ f_f \ = \ T \ \ln\left(\frac{g\ \exp(\beta\mathcal{E}_0)}{2\ Z_w}\right) \qquad (9)$$

which is the quantity we earlier identified as the one which decides whether the system cooperatively selects the folded or the unfolded state. To clarify the contents of this formula, we rewrite it for small energy level spacings $\Delta\mathcal{E} \ll T$:

$$\Delta f = \ \mathcal{E}_0 \ + \ \mathcal{E}_{min} \ + \ T\ln(\frac{g\Delta\mathcal{E}}{2\ T}) \\ - \ T\ln\left(1 - \exp(-(\mathcal{E}_{max} - \mathcal{E}_{min})/T)\right). \qquad (10)$$

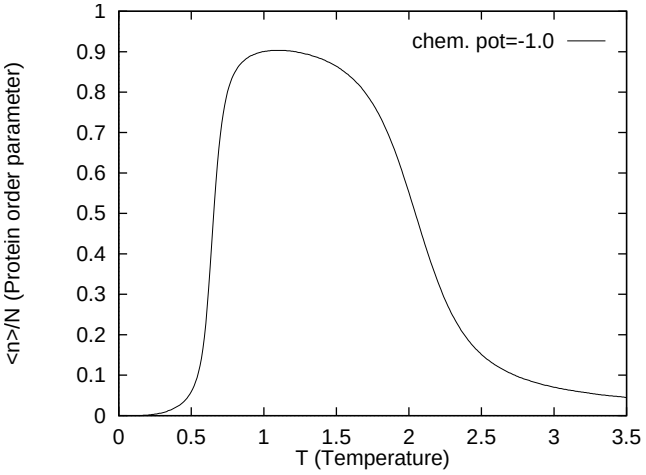
From this expression for the difference in free energy one easily obtains the corresponding differences in energy, entropy and specific heat. In particular, we obtain a gap in the specific heat between the folded and unfolded state $\Delta c = (\Delta\mathcal{E}/T)^2/(e^{\Delta\mathcal{E}/T} - 1)^2\ e^{\Delta\mathcal{E}/T} \ \sim \ \exp(\Delta\mathcal{E}/T) \sim 1$ for temperatures $T \in [\Delta\mathcal{E}, \mathcal{E}_{min} + \mathcal{E}_{max}]$, see Figure 2.

To simplify the discussion let us consider the limit of large $\mathcal{E}_{max}$ in (10). It is easily seen that Δf has a maximum at the temperature $T_m \approx g\Delta\mathcal{E}/2e$. The corresponding value of Δf is $\Delta f(T_m) \approx \ (\mathcal{E}_{min} + \mathcal{E}_0) \ + \ g\Delta\mathcal{E}/2e$, so the condition for the existence of a region of stability of the ordered structure ($\Delta f > 0$) is:

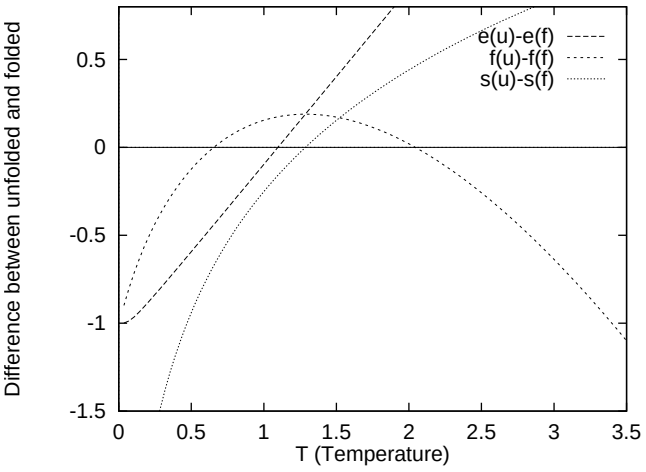
$$\frac{g\Delta\mathcal{E}}{2e} > -(\mathcal{E}_{min} + \mathcal{E}_0). \qquad (11)$$

This is of course always satisfied if $(\mathcal{E}_{min} + \mathcal{E}_0) > 0$. However, the more interesting situation is $(\mathcal{E}_{min} + \mathcal{E}_0) < 0$, since then $\Delta f < 0$ at sufficiently low temperature, *i.e.* the phenomenon of cold unfolding appears. Under these conditions ΔE is also negative at sufficiently low temperature which means that we have a negative latent heat for cold unfolding. Figure 3b shows these thermodynamic functions. They qualitatively reproduce the known thermodynamic behavior of globular proteins as described in the introduction [3].

In our description the mechanism for the transitions is the following. The excistence of an unfolded state at low temperature is associated to the energy gain of water ordering around the hydrophobic residues. At zero T this contribution is $\mathcal{E}_{min}$ per degree of freedom, and in order for cold unfolding to take place, this contribution has to dominate the chain energy $\mathcal{E}_0$. With increasing temperature this is in competition with the folding entropy gain of water due to a shielding of folded hydrophobic residues from the water. This difference in the water degrees of freedom between the folded and unfolded states deminishes with increasing temperature because an increasing fraction of water energy levels are accesible thereby finally opening for the warm unfolding.



(a)



(b)

Fig. 3. (a) Average fraction of folded protein variables as function of temperature for $\mu = -(\mathcal{E}_0 + \mathcal{E}_{min}) = -1.0$ and the other parameters as in Figure 1. The figure shows that in between the transitions the protein is folded. (b) Difference of thermodynamic functions between folded and unfolded configurations for the same chemical potential as in (a). The difference in free energy Δf has a maximum and becomes negative for both high and low temperature (cold unfolding). $\Delta S/N$ and $\Delta E/N$ increase with temperature. $\Delta E/N$ is negative at the cold unfolding transition corresponding to a negative latent heat.

At the high temperature transition the unfolding is due to the entropy gain of the variables φ, which here is log(2) per degree of freedom. This competes with a high temperature refolding favored partly by the energy gain of the chain, which is $\mathcal{E}_0$ per degree of freedom, and partly by the already mentioned entropy gain of the water when the hydrophobic residues are shielded. Therefore the transition temperature will be higher if the entropy contribution of the water is increased (making g larger at fixed $\Delta\mathcal{E}$). This can be read off from equation (7), where the influence of water is expressed by the ratio g/Z_w.

For both warm and cold unfolding the phase transition is parametrized through the same factor $g\exp(\beta\mathcal{E}_0)/(2Z_w)$

and the mathematical interplay between the degrees of freedom is similar in the two cases, as stated in the summation of equation (5).

Coming back to the partition function (3, 4), we may write:

$$\mathcal{E} = -N\mathcal{E}_0 \; + \; (N-n)(\mathcal{E}_0 + \mathcal{E}_{min}) + \sum_{i=n+1}^{N} \Delta\mathcal{E} \; s_i$$

$$= \; -N\mathcal{E}_0 \; + \; \sum_{i=n+1}^{N} \; [\Delta\mathcal{E} \; s_i + \mathcal{E}_0 + \mathcal{E}_{min}] \qquad (12)$$

and

$$Z = e^{\beta N \mathcal{E}_0} \sum_{n=0}^{N-1} \; 2^{N-n-1} g^n \; \sum_{\{s_i\}} \; e^{-\beta \sum_{i=n+1}^{N} \; (\mathcal{E}_i - \mu)}$$

$$+ \; g^N \exp(\; \beta \mathcal{E}_0 N) \qquad (13)$$

where we have set $\mathcal{E}_i = \Delta\mathcal{E} \; s_i \; , \; \mu = -(\mathcal{E}_0 + \mathcal{E}_{min})$. From this expression for Z we can identify μ with the chemical potential *of the water*, or, to be more precise, the difference in chemical potential of the water when it is in contact with the hydrophobic interior of the protein and when it is not. Therefore, $\mu > 0$ is the physically relevant situation. Experimentally, μ can be changed by adding denaturants, changing pH, *etc.*, which indeed alters the stability of the ordered structure. The four curves in Figure 2, which are to be compared with the experimental data in Figure 1, are the results of the model for different values of the chemical potential μ.

In conclusion, this paper introduces a new model for the stability of proteins which reproduces their known thermodynamics. We obtain: 1) first order unfolding transitions; 2) both warm and cold unfolding; 3) cooperativity in the sense that the free energy difference stabilizing the folded state is only a fraction of kT_{room} per degree of freedom; 4) a qualitatively correct behavior of the specific heat both as a function of temperature and chemical potential; 5) a gap in the specific heat between the unfolded and folded state; 6) a negative latent heat for the cold unfolding. A deficiency of the model is that our description of the water-protein coupling is simplified. As a result, the two transitions are too far apart in absolute temperature and in the model the cold unfolding appears sharper than the warm unfolding, which is not seen in experiment. This deficiency calls for some modifications, in particular by introducing both hydrophobic and hydrophilic φ_i's one can influence the relative strength of the two transitions.

References

1. P.L. Privalov, N.N. Khechinashvili, J. Mol. Biol. **86**, 665-684 (1974).
2. P.L. Privalov, E.I. Tiktopulo, S.Yu Venyaminov, Yu.V. Griko, G.I. Makhatadze, N.N. Khechinashvili, J. Mol. Biol. **205**, 737-750 (1989).
3. P.L. Privalov, *Physical basis of the stability of the folded conformations of proteins*, in *Protein Folding,* edited by T.E. Creighton (W.H. Freeman, San Fransisco, 1992).
4. J.T. Edsall, J. Am. Soc. **57**, 1506 (1935).
5. J.A. Schellman, J. Phys. Chem. **62**, 1485 (1958); B.H. Zimm, J.K. Bragg, J. Chem. Phys. **31**, 526 (1959); C.R. Cantor, P.R. Schimmel, *Biophysical Chemistry,* (W.H. Freeman, San Fransisco, 1980), Chap. 20.
6. R. Zwanzig, Proc. Natl. Acad. Sci. **92**, 9801-9804 (1995).
7. K.A. Dill, K.M. Fiebig, H.S. Chan, Proc. Natl. Acad. Sci. **90**, 1942-1946 (1993).
8. K.A. Dill, S. Bromberg, K. Yue, K.M. Fiebig, D.P. Yee, P.D. Thomas, H.S. Chan, Protein Sci. **4**, 561 (1995).
9. H. Li, C. Tang, N.S. Wingren, NEC preprint (1997).
10. A. Hansen, M.H. Jensen, K. Sneppen, G. Zocchi, Physica A **250**, 355 (1998).
11. C. Levinthal, J. Chim. Phys. Phys.-Chim. Biol. **65**, 44-45 (1968).
12. A.R. Fersht, FEBS **325**, 5-16 (1993).
13. T.E. Creighton, J. Mol. Biol. **87**, 603-624 (1974).
14. T.E. Creighton, *Proteins* (W.H. Freeman, San Fransisco, 1993).
15. G. Zocchi, Proc. Natl. Acad. Sci. **94**, 10647-10651 (1997).
16. B. Nolting, R. Golbik, J.L. Neira, A.S. Soler-Gonzales, G. Schreiber, A.R. Fersht, Proc. Natl. Acad. Sci. **94**, 826 (1997).