Statistical Thermodynamics of Proteins

Audun Bakk

Preface

This thesis is submitted to the Norwegian University of Science and Technology (NTNU) in partial fulfillment of the requirements for the doctoral degree *Doktor Ingeniør*. The work was carried out in the Theoretical Physics Group at the Department of Physics, NTNU.

I wish to express my gratitude to my supervisor Professor Alex Hansen for bringing me into this fascinating project and giving me invaluable support during the study. I thank Professor Johan S. Høye for hours with very interesting and very helpful discussions. The other members of the Theoretical Physics Group have also been very kind, helpful, and inspiring to me. I thank Mona H. Windseth for her encouragements. I acknowledge the Research Council of Norway for financial support (Contract No. 129619/410).

Audun Bakk
February 26, 2002

Contents

1 List of papers

1. A. Bakk, J.S. Høye, A. Hansen, K. Sneppen, and M.H. Jensen. Pathways in two-state protein folding. *Biophys. J.* **79**, 2722-2727 (2000).

2. A. Bakk, J.S. Høye, A. Hansen, and K. Sneppen. Thermodynamical implications of a protein model with water interactions. *J. Theor. Biol.* **210**, 367-373 (2001).

3. A. Bakk, A. Hansen, and K. Sneppen. Protein model exhibiting three folding transitions. *Physica A* **291**, 60-70 (2001).

4. A. Bakk, J.S. Høye, and A. Hansen. Heat capacity of protein folding. *Biophys. J.* **81**, 710-714 (2001).

5. A. Bakk and J.S. Høye. Microscopic argument for the anomalous hydration heat capacity increment upon solvation of apolar substances. *Physica A* **303**, 286-294 (2002).

6. A. Bakk. Two-state protein model with water interactions: Influence of temperature on the intrinsic viscosity of myoglobin. *Phys. Rev. E* **63**, 061906 (2001).

7. A. Bakk, J.S. Høye, and A. Hansen. Specific heat upon aqueous unfolding of the protein interior: a theoretical approach. *Physica A* **304**, 355-361 (2002).

8. A. Bakk and A. Hansen. Mapping the non-directed polymer model to a non-linear growth equation of Burgers type. *Physica A*. Accepted.

9. A. Bakk, J.O. Fossum, G.J. da Silva, H.M. Adland, A. Mikkelsen, and A. Elgsaeter. Viscosity and transient electric birefringence study of clay colloidal aggregation. *Phys. Rev. E* **65**, 021407 (2002).

10. A. Bakk, J.S. Høye, and A. Hansen. Apolar and polar solvation thermodynamics related to the protein unfolding process. *Biophys. J.* **82**, 713-719 (2002).

11. A. Bakk, P.G. Dommersnes, A. Hansen, J.S. Høye, K. Sneppen, and M.H. Jensen. Thermodynamics of proteins: Fast folders and sharp transitions. *Comput. Phys. Commun.* Accepted.

12. A. Bakk. Heat capacities of solid state proteins: implications for protein stability in solution. Submitted.

2 Summary

The subject of this thesis is to formulate effective energy expressions (Hamiltonians) of proteins and protein related systems. By use of equilibrium statistical mechanics we calculate thermodynamical functions, whereupon we compare the results from theory with experimental data. Papers 1-7 and 10-12 concern this problem. In addition, Paper 8 (P8) and Paper 9 (P9) are attached. Both these papers were finalized during the Ph.D. study. However, they are not related to proteins.

2.1 Introduction to protein thermodynamics

The seminal works of Anfinsen, that led to the Nobel prize in chemistry in 1972, established the "thermodynamic hypothesis" of proteins [1]. A consequence of this is that the folded (native) conformation of a protein is the state of lowest Gibbs free energy. This hypothesis is the foundation of this Ph.D. project. Thus, our first task is to write down the Hamiltonian for a given protein (or a protein related system) [2].

Let $\mathcal{H}_i$ be the value of the Hamiltonian associated with microstate i, thus, the canonical partition function Z for a system of N microstates, yields [3]

$$Z = \sum_{i=1}^{N} \exp\left(-\beta \mathcal{H}_i\right), \tag{1}$$

where $\beta = (k_{\mathrm{B}}T)^{-1}$. The k_{B} is the Boltzmann's constant and T is the absolute temperature. Given the canonical partition function, one may derive other thermodynamical functions as, e.g., the internal energy

$$U = -\frac{\partial}{\partial \beta} \ln Z \tag{2}$$

and the heat capacity

$$C = \frac{\partial U}{\partial T} = \frac{1}{k_{\mathrm{B}}T^2} \frac{\partial^2}{\partial \beta^2} \ln Z. \tag{3}$$

The heat capacity is a thermodynamic quantity that can be directly measured experimentally. Loosely speaking, the heat capacity is the amount of heat necessary to raise the temperature one Kelvin [4].

A major difficulty with respect to proteins is that the Hamiltonian is not obvious, thus, the canonical partition function is not easily attainable for proteins. The reason for this is that a protein is a mesoscopic system that typically consists of thousands of atoms. In addition, proteins can fold and unfold. Furthermore, proteins are normally soluted in water, which makes the protein folding problem (given an amino acid sequence – what is the three-dimensional conformation?) even more complicated [5].

At physiological temperatures (30°C – 40°C) proteins exist in a unique state called the native (folded) state. Upon heating, proteins will denature (unfold). Surprisingly, at sufficient low temperatures (< 10°C) several small globular proteins also tend to unfold. The latter phenomenon is called cold unfolding [6].

Hansen et al. [7] proposed an analytical model that exhibits both cold and warm unfolding of proteins. This model established the starting point of this thesis.

It should be noted that our approach to protein folding is an effective one. An effective approach means that one grasps, and tries to describe, the main characteristics of the system, thus, throwing away details of the system. The gain with such an approach is that fewer free parameters have to be determined compared to an analysis "from scratch". Furthermore, proteins are obviously too complicated to allow more exact modeling so far [8].

To visualize what is meant by an effective description, we will briefly explain the model for a hydrogen bond between two water molecules that is applied in P5, P6, and P10.

It is known upon solvation of apolar (no permanent dipoles or charges) molecules in water, that both enthalpy and entropy are reduced compared to bulk water (see Refs. 11-16 in P7). This is usually attributed to an "iceberg" formation around the apolar molecules (solute) [9]. In order to model this "iceberg" we simply

say that apolar solvation is equivalent to an increased number of hydrogen bonds between water molecules in the vicinity of an apolar molecule compared with bulk water. Furthermore, we assume that the hydrogen bonds are able to bend, as illustrated in Figure 1. The energy associated with the bending of an individual hydrogen

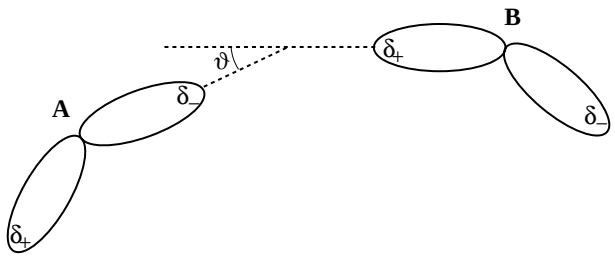

Figure 1: Schematic illustration of two water molecules (A and B) where two of the four orbitals are shown, respectively. The hydrogen bond between the two water molecules is bent an angle ϑ. The figure is taken from P6.

bond (HB) between two water molecules is modeled in analogy with an electric dipole in an external electric field, thus

$$\mathcal{H}_{\mathrm{HB}} = -\varepsilon_{\mathrm{HB}} \cos\vartheta, \qquad \vartheta \in [0, \pi]. \tag{4}$$

The $\varepsilon_{\mathrm{HB}}$ may be interpreted as the energy for breaking one hydrogen bond. The effective Hamiltonian in Eq. (4) is then inserted into the canonical partition function [Eq. (1)]. After integration over the ϑ-angles, thermodynamical functions can be derived and compared with experimental data.

2.2 Protein related papers

In P1 we expand the framework of Hansen et al. [7]. We show, in spite of an hierarchical model where the folding follows (thermodynamically) a specific sequence of folding steps, that the folding and the unfolding transitions may still follow the behavior of a two-state system, respectively. Furthermore, in P11 we discuss the possibility of a two-state unfolding transition for both a single folding pathway, as in P1, and by multiple (parallel) folding pathways, as investigated by Dommersnes et al. [10].

In P3 we study a model exhibiting three folding transitions that is a result of a combination of the Hamiltonians proposed in P1. In P2 we introduce independent water molecules compared with the model of Hansen et al. [7] (in Ref. [7] the water associated with a protein is treated as a unity). Furthermore, we reduce the number of free parameters compared with the model Hansen et al. [7].

The papers P1, P2, and P3 include only a qualitative comparison with experimental data on the heat capacity. In P4 we compare also quantitatively theoretical results with experimental heat capacity data on myoglobin from Privalov et al. [6]. The water interactions are now modeled in analogy with interacting electric dipoles in an external electric field. We also add a vibration term to the energy. This model results in a fairly good fit to the experimental data. However, we now realize that the fit can be done even better by using a (close to) two-state partition function that yields a sharper transition.

After the work that led to P4 we realized, instead of investigating complete protein models, that it might be beneficial to split up protein energetics in two sub-problems: (1) water interactions and (2) internal interactions, and study these contributions separately. Water interactions include both water-water interactions and water-protein interactions.

P5 is a study of the change in the heat capacity upon solvation of small apolar molecules. This work serves as a preceding study of hydration effects related to proteins. In P5 we use the model sketched in Section 2.1, which is the continuum limit of the water model used by Hansen et al. [7]. The model studied in P5 fits well to experimental solvation data on small apolar molecules. This water model is also applied in P6 where we study experimental intrinsic viscosity data. Intrinsic viscosity is a measure of the effective hydrodynamic radius of a macromolecule. In P6 we assume a linear relation between the degree of folding and the effective hydrodynamic radius of the protein. This makes it possible to explain the intrinsic viscosity data on myoglobin.

One problem associated with protein experiments in solution, is that there is no direct way to separate water interactions from internal protein interactions. However, Privalov and Makhatadze show by means of a model compound evaluation that it is possible to quantify the hydration effect of proteins [11]. In P7 we study the latter data by the same water model as in P4, i.e., a model in analogy with interacting electric dipoles in an external electric field. We fit the model, in a mean field solution, to model compound heat capacity data from Privalov and Makhatadze [11]. As our model is an effective description of the solvation process, we found it, and still find it difficult to make a microscopic interpretation of the parameters corresponding to the external field and the intermolecular coupling, respectively.

P10 takes into account the fact that the change in the heat capacity upon solvation ($\Delta C_p^{(\mathrm{solv})}$) is positive for apolar surfaces, whereas $\Delta C_p^{(\mathrm{solv})}$ is (surprisingly) negative for polar (permanent dipoles and charges) surfaces [12]. As a model for the unfolding hydration effect of the apolar parts of the protein we use the same model applied in P5. On the other hand, the polar solvation process is much less understood than the apolar solvation process. We model polar solvation as breaking hydrogen bonds due to bulk water, i.e., the apolar model with a minus sign associated with the heat capacity function. To take into account the presumably strong forces between the polar surface and the nearest water molecules we apply a two-state model. Both the apolar model and the polar model fit well to the model compound data of Makhatadze and Privalov [13]. Furthermore, upon summation of the apolar and the polar contributions from our fittings to the model compound data, we obtain the characteristic curvature of the hydration heat capacity function (versus temperature) as expressed in Figure 3 of P10.

In P12 we show, to first approximation, that heat capacity data of solid state proteins correspond to experimentally measured heat capacities of proteins in solution, where the latter data are corrected for the hydration effect by means of a model compound evaluation

[11]. We also show in P12 that the solid state heat capacity can be represented by an analytical physical model for temperatures from 260 K to 420 K, i.e., the relevant temperatures for water soluted proteins.

2.3 Other works

In P8 we map the non-directed polymer model onto a Kardar-Parisi-Zhang like equation by means of path integrals. We relate the scaling exponents in the two models and discuss extreme behavior.

In P9 we investigate clay suspensions experimentally by two techniques: (1) stationary shear viscometry and (2) transient electric induced birefringence. The data seem to confirm that the clay aggregates increases in size versus increasing salt content.

3 The future

A natural expansion of this protein project is to investigate the hydration heat capacity of native proteins, whose model may show not to be too different from the model of the unfolding hydration heat capacity change as inquired for instance in P10. Furthermore, it would also be interesting to study protein thermodynamics on the level of individual amino acids. This might result in a more predictive protein thermodynamics, compared to the models presented here.

Protein thermodynamics *was* a challenge in the beginning of this Ph.D. project, but is now *even more* challenging.

References

[1] C.B. Anfinsen. Principles that govern the folding of protein chains. *Science*, **181**, 223-230 (1973).

[2] H. Goldstein. *Classical Mechanics*. 2nd. ed., Addison-Wesley, Wokingham, UK (1980).

[3] L.D. Landau and E.M. Lifshitz. *Statistical Physics*. Part 1, 3rd ed. (revised and enlarged by E.M. Lifshitz and L.P. Pitaevskii), Pergamon, Oxford, UK (1980).

[4] B. Wunderlich. *Thermal Analysis*. Academic, London, UK (1990).

[5] C. Branden and J. Tooze. *Introduction to Protein Structure*. 2nd. ed., Garland, New York, USA (1999).

[6] P.L. Privalov, Yu.V. Griko, S.Yu. Venyaminov, and V.P. Kutyshenko. Cold denaturation of myoglobin. *J. Mol. Biol.* **190**, 487-498 (1986).

[7] A. Hansen, M.H. Jensen, K. Sneppen, and G. Zocchi. Statistical mechanics of warm and cold unfolding in proteins. *Eur. Phys. J. B* **6**, 157-161 (1998).

[8] E.I. Shakhnovich. Modeling protein folding: the beauty and power of simplicity. *Fold. Des.* **1**, R50-R54 (1996).

[9] H.S. Frank and M.W. Evans. Free volume and entropy in condensed systems. III. Entropy in binary liquid mixtures; partial molal entropy in dilute solutions; structure and thermodynamics in aqueous electrolytes. *J. Chem. Phys.* **13**, 507-532 (1945).

[10] P.G. Dommersnes, A. Hansen, M.H. Jensen, and K. Sneppen, Parametrization of multiple pathways in proteins: Fast folding versus tight transitions, cond-mat/0006304.

[11] P.L. Privalov and G.I. Makhatadze. Contribution of hydration and non-covalent interactions to the heat capacity effect on protein unfolding. *J. Mol. Biol.* **224**, 715-723 (1992).

[12] M.H. Abraham and Y. Marcus. The thermodynamics of solvation of ions. *J. Chem. Soc. Faraday T.* **82**, 3255-3274 (1986).

[13] G.I. Makhatadze and P.L. Privalov. Energetics of protein structure. *Adv. Prot. Chem.* **47**, 307-425 (1995).

Paper 1

Bakk, A ; Høye, JS ; Hansen, A ; Sneppen, K ; Jensen, MH: Pathways in two-state protein folding . *Biophys.* , 79(2000), 2722-2727

Paper 2

J. theor. Biol. (2001) **210**, 367–373
doi:10.1006/jtbi.2001.2311, available online at http://www.idealibrary.com on IDEAL®

Thermodynamical Implications of a Protein Model with Water Interactions

AUDUN BAKK*, JOHAN S. HØYE, ALEX HANSEN AND KIM SNEPPEN†

Department of Physics, Norwegian University of Science and Technology, NTNU, NO-7491 Trondheim, Norway

(*Received on* 14 *July* 2000, *Accepted in revised form on* 7 *March* 2001)

We refine a protein model that reproduces fundamental aspects of protein thermodynamics. The model exhibits two transitions, hot and cold unfolding. The number of relevant parameters is reduced to three: (1) binding energy of folding relative to the orientational energy of bound water, (2) ratio of degrees of freedom between the folded and unfolded protein chain, and (3) the number of water molecules that can access the hydrophobic parts of the protein interior upon unfolding. By increasing the number of water molecules in the model, the separation between the two peaks in the heat capacity curve comes closer, which is more consistent with experimental data. In the end we show that if we, as a speculative assumption, assign only two distinct energy levels for the bound water molecules, better correspondence with experiments can be obtained.

1. Introduction

Proteins are crucial components in all living organisms. In order to have biological functionality at physiological temperatures it is important that they have an exclusively ordered state, termed the native state. Anfinsen (1973) showed that the native state is genetically determined, which means that each protein, with its specific amino acid sequence, folds into a unique conformation. The experiment by Anfinsen also proved that the native state is thermodynamically determined, i.e. the state in which Gibbs free energy of the whole system is lowest. It is now commonly accepted that folding of the polypeptide chain is thermodynamically driven (Makhatadze & Privalov, 1995).

A peculiar feature of proteins is that they fold on time-scales from 10^{-3} to 1 s. If one calculates the folding time of this process simply by taking the folding as stochastic, one finds astronomical time-scales (Levinthal, 1968). This is called the "Levinthal paradox". A resolution of this apparent paradox is outlined by Shakhnovich (1997), where he describes how the protein forms at first a "nucleation-condensate" (Fersht, 1995; Itzhaki *et al.*, 1995) via thermal fluctuations of the polypeptide chain, whereupon a transition state (TS) occurs, carrying common features to the native state, in which the protein descends downhill in the Gibbs free energy landscape to the native state. A recent point of view is that the "TS-pathway" is not a concrete mechanistical pathway on which every position corresponds to a unique conformation. Instead a "statistical pathway" is introduced, where a new step forward on the pathway means reaching a more favorable statistical ensemble of conformations

* Author to whom correspondence should be addressed. E-mail: audun.bakk@phys.ntnu.no
† Present address: Institute for Theoretical Physics, Kohn Hall, UC Santa Barbara, CA 93106, USA.

0022–5193/01/110567 + 07 $35.00/0

with regard to Gibbs free energy. However, every step along the path, each describing an ensemble of conformations, should have some common structural features which acts like checkpoints for the folding. Further, these checkpoints of increasing order is likely to follow a *folding pathway* (Baldvin & Rose, 1999; Hansen *et al.*, 1998a, b, 1999; Bakk *et al.*, 2000, 2001), where one particular point on the pathway depends on the assumption that the main structures of the earlier steps are conserved.

Unfolding of the polypeptide chain by increasing the temperature is somewhat intuitive, but what is rather surprising is that proteins unfold at low temperatures, i.e. they become denaturated and not biologically functional. *Cold* denaturation seems to be a general property of small globular proteins (Privalov, 1990; Chen & Schellman, 1989; Griko *et al.*, 1988). Experiments on low-temperature unfolding are difficult, because most of the proteins seem to unfold around and below the freezing point of water. Unfolding at cold temperatures is thought to result from increased solvation of apolar surface areas of the protein on unfolding, as water becomes more ordered (Privalov, 1992).

The paper is organized as follows. In Section 2, we present the model and calculate the partition function. In Section 3, we discuss the thermodynamics of the model, and present the results for folding and unfolding transitions.

2. The Physical Model

2.1. POLYPEPTIDE CHAIN

We refine a physical model for a small globular protein, which builds on earlier models by Hansen *et al.* (1998a, 1999) and Bakk *et al.* (2000). The protein is viewed as a zipper (Fig. 1), in analogy to the model of Dill *et al.* (1993) which is a one-dimensional model of a folding pathway. The complex three-dimensional protein is equipped with N *contacts*. Each individual contact is assigned an energy of $-\varepsilon_0 < 0$ if it is folded (native), zero otherwise (Bryngelson & Wolynes, 1987, 1990). This means that a folded contact is energetically favorable. Requiring that if contact i is folded, all contacts $j < i$ are also folded, is an implication of the pathway. The point of view

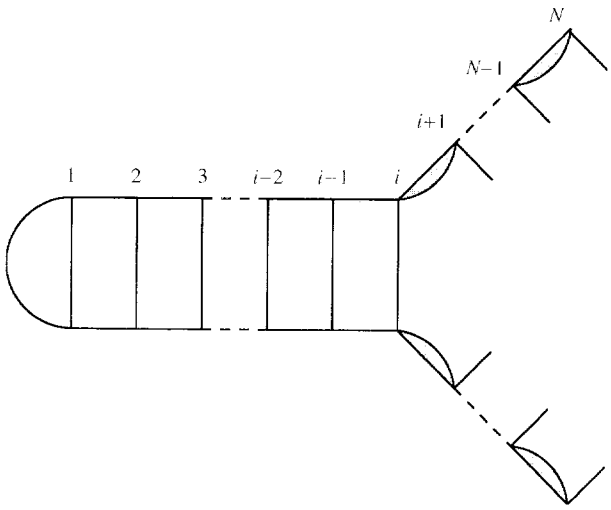

FIG. 1. Schematic illustration of a partly folded protein containing i folded contacts and $N - i$ unfolded contacts associated with water (shadowed).

that the contacts are distinct contact points in space is a simplification. Folding contact 1 means finding the "nucleation-condensate", which is reached through a condensation down to a structure which marks the beginning of the folding pathway, and guides the protein into the native state. Each individual contact is regarded as a statistical ensemble due to the previous discussion in Section 1, and they are likely to form non-local contacts which may be important for the cooperativity (Shakhnovich, 1997; Abkevich *et al.*, 1995; Privalov, 1996). However, the specific contacts do have some common structural motifs. The specific mechanism forming this "nucleation-condensate" is not considered in this paper, but we assume that the condensate exists and restrict the study to the TS-pathway, that eventually folds the protein into its native conformation.

We introduce binary contact variables $\phi_i \in \{0, 1\}$. $\phi_i = 0$ means that contact i is open (unfolded), and $\phi_i = 1$ means that contact i is in contact (folded). Assuming N contacts, a Hamiltonian (H_1) for the energies associated with the polypeptide chain is written in a compact way as (Hansen *et al.*, 1998a, b, 1999; Bakk *et al.*, 2000)

$$H_1 = -\varepsilon_0(\phi_1 + \phi_1\phi_2 + \cdots + \phi_1\phi_2 \cdots \phi_N). \tag{1}$$

Product terms $(\phi_1 \cdots \phi_i)$ meet the assumption about a folding pathway, if $\phi_i = 0$ then all terms containing $j \geq i$ will vanish.

The unfolded protein will access some more degrees of freedom relative to the native protein, because an unfolded polypeptide backbone will have rotational freedom represented by the dihedral angles (Creighton, 1993). This can be further simplified to one "pseudodihedral" angle (Peticolas & Kurtz, 1980), and is incorporated by assigning each single unfolded contact f degrees of freedom. The parameter f is interpreted as the relative increase in the degrees of freedom for an unfolded contact compared to a folded contact.

2.2. WATER INTERACTIONS

Introduction of water is important for several reasons. First, proteins are *in vivo* exposed to water, and second, water has several peculiar properties due to the polarity of, and the hydrogen bonds between water molecules. Makhatadze & Privalov (1995) state that the sum of the hydration effects destabilize the native state, and decreasing temperature implies increasing destabilizing action. This is termed the *hydrophobic force*, and the water–protein interaction is incorporated by an "energy ladder" representing each individual water molecule associated with the unfolded parts of the protein (i.e. all contacts where $\phi_i = 0$) (Hansen *et al.*, 1998a, 1999; Bakk *et al.*, 2000):

$$\omega_{ij} = \begin{cases} -\varepsilon_w + (g-1)\delta \\ \quad \vdots \\ -\varepsilon_w + 2\delta, \\ -\varepsilon_w + \delta, \\ -\varepsilon_w. \end{cases} \tag{2}$$

Here ω_{ij} is the energy for water molecule j at contact i. $\varepsilon_w > 0$, is the energy minimum, and δ is the spacing between the energy levels. Interactions between the water molecules are not considered in this paper. Equation (2) is interpreted as all available energies for *one* water molecule associated with the unfolded contact i. Here we will let M water molecules be associated with each unfolded contact, whereas Hansen *et al.* (1998a, 1999) and Bakk *et al.* (2000) restricted this number to one. No water is supposed to access a folded contact, i.e. the protein interior.

The "ladder" contains g equidistant energies which give an entropy contribution while contact i is folded, because then the water is unbounded. Hence, a folded contact implies an entropy contribution from g^M degrees of freedom. The water energy is, of course, a simplification, and is connected to the need of some sort of energy levels to make it energetically favorable to unfold at low temperatures. Thus, the "energy ladder" in eqn (2) is introduced for computational convenience. We note that the proposed water energy in fact is nothing but the quantized energy levels of a magnetic dipole in an external magnetic field. However, in the limit $g \to \infty$ (with $g\delta$ finite), the classical limit for a magnetic moment of a fixed length is obtained. This in turn is equivalent to an electric dipole in an electric field. The latter can be interpreted as a direct physical model of dipolar water molecules that feel an effective electric field from the protein. In a protein, an electrical field arises from the permanent and induced charges on the protein surface that becomes exposed after unfolding of a contact. This field will interact with the nearest water molecules (dipoles) and structure them. The quantitative aspects of the folding problem will probably need a discussion of additional interactions, but this will not be considered here. Figure 1 is a schematic illustration of a partly folded protein containing some water associated with the hydrophobic parts that are exposed upon unfolding of the contacts.

By using the same notation as in eqn (1), the energy associated with water–protein interactions H_2, becomes

$$\begin{aligned} H_2 = {} & (1-\phi_1)(\omega_{11} + \omega_{12} + \cdots + \omega_{1M}) \\ & + (1-\phi_1\phi_2)(\omega_{21} + \omega_{22} + \cdots + \omega_{2M}) \\ & + \cdots + (1-\phi_1\phi_2\cdots\phi_N) \\ & \times (\omega_{N1} + \omega_{N2} + \cdots + \omega_{NM}). \end{aligned} \tag{3}$$

2.3. THE PARTITION FUNCTION

The Hamiltonian $H = H_1 + H_2$ describing the entire system is then

$$H = -\varepsilon_0(\phi_1 + \phi_1\phi_2 + \cdots + \phi_1\phi_2\cdots\phi_N)$$

$$+ (1 - \phi_1)(\omega_{11} + \omega_{12} + \cdots + \omega_{1M})$$

$$+ (1 - \phi_1\phi_2)(\omega_{21} + \omega_{22} + \cdots + \omega_{2M})$$

$$+ \cdots + (1 - \phi_1\phi_2 \cdots \phi_N)$$

$$\times (\omega_{N1} + \omega_{N2} + \cdots + \omega_{NM}). \quad (4)$$

The partition function $Z = \sum_{i=0}^{N} Z_i$, where the term Z_i corresponds to folding of all contacts $\leqslant i$ (pathway assumption), becomes

$$Z_i = f^{N-i} g^{iM} e^{i\varepsilon_0\beta} \left(e^{\varepsilon_w\beta} \frac{1 - e^{-g\delta\beta}}{1 - e^{-\delta\beta}} \right)^{(N-i)M}. \quad (5)$$

$\beta \equiv 1/T$ is a rescaled inverse absolute temperature where the Boltzmann constant is absorbed in T. Z_0 means that all contacts are open, i.e. a complete unfolded protein. The factor f^{N-i} in eqn (5) arises from the degrees of freedom in the polypeptide chain that are available in the $N - i$ unfolded contacts. Further, the product term g^{iM} is the entropy contribution from M free non-interacting water molecules associated with i folded contacts. $e^{i\varepsilon_0\beta}$ is the Boltzmann factor from i contact energies $-\varepsilon_0$ in the polypeptide chain. The last term in brackets is simply the sum over all distinct levels in one "water ladder" raised to the power of the number of water molecules $(N - i)M$, bounded to the unfolded hydrophobic parts of the protein. A rearrangement of eqn (5) gives

$$Z_i = (g^M e^{\varepsilon_0\beta})^N r^{i-N}, \quad (6)$$

where we have defined

$$r \equiv \left[\frac{g}{f^{1/M}} e^{(\varepsilon_0/M - \varepsilon_w)\beta} \frac{1 - e^{-\delta\beta}}{1 - e^{-g\delta\beta}} \right]^M. \quad (7)$$

We put or assume that $\delta\beta \ll 1$ (i.e. $g \to \infty$), which means an infinite small level spacing between the water energies. Hence, a Taylor expansion yields $1 - e^{-\delta\beta} \approx \delta\beta$, and thus eqn (7) can be rewritten as

$$r = \left[a e^{\mu\beta} \frac{\beta}{\sinh \beta} \right]^M, \quad (8)$$

where $a \equiv 1/f^{1/M}$, and the inverse temperature is rescaled by $g\delta\beta/2 \to \beta$. The parameter a reflects the ratio of the degrees of freedom between the folded and the unfolded units of protein chain. The new energy parameter $\mu \equiv (\varepsilon_0/M - \varepsilon_w + g\delta/2)/(g\delta/2)$ is proportional to the binding energy of each contact, and may be interpreted as an *effective chemical potential* for each single protein. Changing the environments of the protein, i.e. adding denaturants or changing pH, changes this chemical potential.

We calculate the partition function by simply summing up the Z_i terms in eqn (6):

$$Z = \sum_{i=0}^{N} Z_i = g^{NM} e^{c\beta} \frac{1}{r^N} \sum_{i=0}^{N} r^i$$

$$= g^{NM} e^{c\beta} \frac{1 - r^{-(N+1)}}{1 - r^{-1}}, \quad (9)$$

where $c \equiv 2N\varepsilon_0/g\delta$.

The *order parameter* ("reaction coordinate") in this system is n, which is the degree of folding, i.e. the mean of the number of folded contacts divided by N, is

$$n = \frac{1}{N} \frac{\sum_{i=0}^{N} i Z_i}{Z} = \frac{1}{M} \frac{\partial}{\partial(\mu\beta)} (\ln Z)$$

$$= \frac{r}{N} \frac{Nr^{N-1} - (N+1)r^N + 1}{(1 - r^{N+1})(1 - r)}. \quad (10)$$

3. Thermodynamical Calculations and Discussion

3.1. CONTINUUM LIMIT OF THE WATER ENERGY LEVELS

The heat capacity is $C = \beta^2 \partial^2(\ln Z)/\partial\beta^2$. This function is independent of the prefactor $g^{MN} e^{c\beta}$ in Z. Furthermore, Z contains the function r, which has only three parameters: the amplitude factor a, the effective chemical potential μ and the number of water molecules per unfolded contact M. We assume that the number of contacts is a constant, let us say $N = 100$, reflecting a typical number of residues in a small protein. The number of *relevant* parameters in our physical model is now reduced from the initial six: $f, g, \varepsilon_0, \varepsilon_w, \delta$ and M, to only three parameters: a, μ and M.

The partition function in eqn (9), and thus the heat capacity C, is apparently most sensitive to changes in r for values $r \approx 1$. The function r

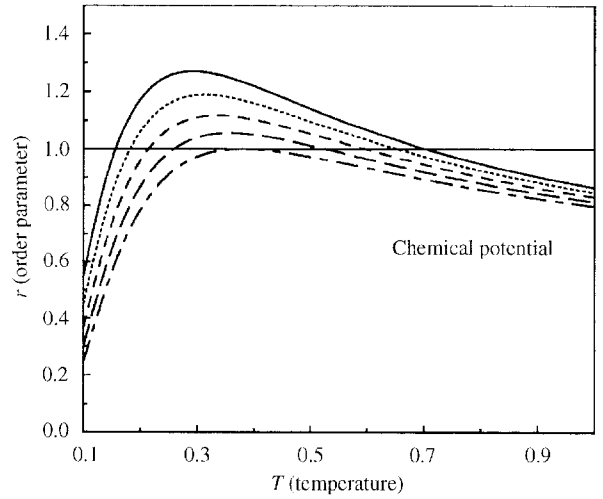

FIG. 2. The function $r(T)$ in eqn (8) for a variable effective chemical potential μ, $a = 0.5$ and $M = 1$. The absolute temperature T is rescaled. (——) 0.71; (·····) 0.69; (- - -) 0.67; (——) 0.65; (— — —) 0.63.

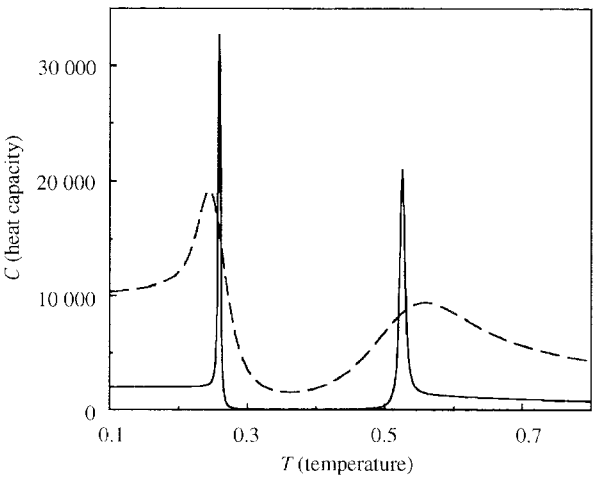

FIG. 3. Heat capacity $C(T)$ for $M = 1$ (scaled by a factor 50) and $M = 10$ showing two characteristic peaks for cold and hot unfolding, where $a = 0.5$ and $\mu = 0.65$. Both axes are rescaled (——) $M = 1$; (——) $M = 20$.

is plotted in Fig. 2 for $a = 0.5$ and $M = 1$. We see the effect of a decreasing effective chemical potential, by the decreasing separation of the two intersections for $r = 1$. Larger M implies only a smaller and higher function r, while the intersections for $r = 1$ is independent of the specific value of M. $\mu_c = 0.63$ is a critical effective chemical potential, and $\mu < \mu_c$ makes the protein denaturated at all temperatures. This critical point was studied for $M = 1$ by Hansen *et al.* (1999).

The heat capacity $C(T)$ in Fig. 3 shows two characteristic peaks. Calculating the order

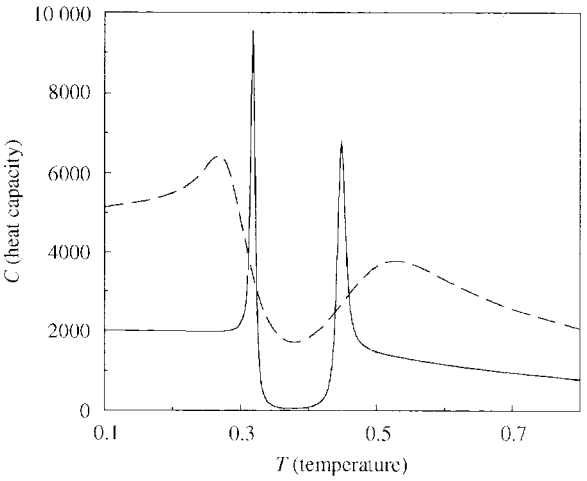

FIG. 4. Heat capacity $C(T)$ for $M = 1$ (scaled by a factor 50) and $M = 20$. $a = 0.5$ and $\mu = 0.635$. Both axes are rescaled (——) $M = 1$; (——) $M = 20$.

parameter n in eqn (10), reveals that the protein is essentially unfolded in the hot and cold temperature regions. This is notable, because as mentioned earlier hot and cold unfolding is a common feature of small globular proteins. It makes sense that the protein is unfolded at low temperatures because this is a question of energy minimizing. Increasing temperature implies folding, regarded as a compromise between entropy and energy. Further increase in temperature shakes the protein, whereupon it eventually unfolds, i.e. the residual entropy of the polypeptide chain dominates in the Gibbs free energy. It is interesting to note that the temperatures for the intersection $r = 1$ for $\mu = 0.65$ in Fig. 2, corresponds to the transition temperatures for the heat capacity in Fig. 3 for $M = 20$. The heat capacity for $M = 1$ is somewhat smeared out, implying a slightly broader separation between the cold and hot unfolding peaks.

Although the temperature in our model is rescaled it may be important that the relative difference between the two temperatures (T_1 and T_2, respectively) $(T_2 - T_1)/T_2$, associated with the peaks (1 and 2) of the heat capacity, corresponds to experimental data, where a typical value is 0.1–0.3 depending on the chemical potential (Privalov, 1990). In order to make the separation between the peaks smaller in our model, we can either decrease μ or a, or decrease both μ and a. In Fig. 4, the value of $\mu = 0.635$ is slightly decreased compared to Fig. 3 where $\mu = 0.65$.

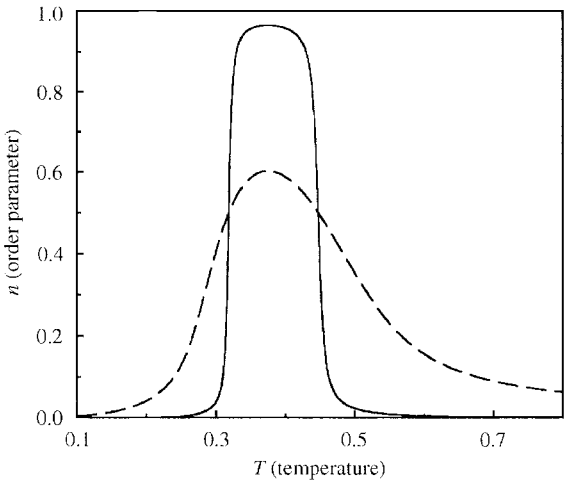

FIG. 5. Order parameter $n(T)$ for $a = 0.5$ and $\mu = 0.635$, where T is rescaled (——) $M = 1$; (— — —) $M = 20$.

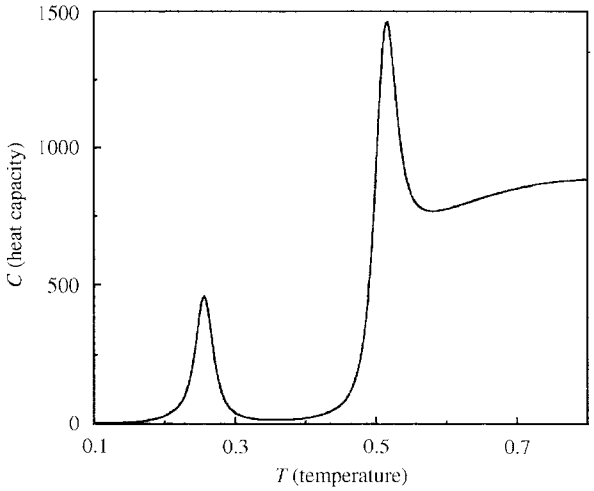

FIG. 6. Heat capacity $C(T)$ for $\tilde{a} = 0.52$, $\tilde{\mu} = 0.99$ and $M = 20$. This plot is based on the function r in eqn (11). Both axes are rescaled.

Obviously this results in a smaller peak separation. The order parameter n in Fig. 5 shows that for $M = 1$ the protein is only partly folded between the two transition temperatures, while for $M = 20$ the protein is nearly completely folded. This fact suggests that for a fixed system size N, several water molecules per unfolded contact ($M \gg 1$) is important in order to get a more realistic separation between the two peaks in the heat capacity.

3.2. TWO LEVEL WATER INTERACTION ENERGY

Finally, in this paper we will discuss the case $g = 2$ for the function r in eqn (7). This corresponds to an Ising spin model (Ising, 1925) with only two energy states per water molecule. A rearranged version of r then becomes

$$r = \left[\frac{\tilde{a}e^{\tilde{\mu}\beta}}{\cosh\beta}\right]^M, \tag{11}$$

where $\tilde{a} \equiv g/f^{1/M}$, and $\tilde{\mu} \equiv (\varepsilon_0/M - \varepsilon_w + \delta/2)/(\delta/2)$. In Fig. 6, based on r in eqn (11), one sees that the warm peak is higher than the cold peak, which is the opposite of the situation in Figs 3 and 4. This first feature corresponds better to experimental results from Privalov *et al.* (1986), Privalov (1990) and Griko *et al.* (1988). Experiments show that, for the warm unfolding transition, the heat capacity of the unfolded state is higher than for the folded state, and it has an upward slope that decreases with increasing temperature (Makhatadze & Privalov, 1995; Privalov, 1997, 1992), with which Fig. 6 is consistent in a qualitative way.

Although this two-level representation of water molecules gives results with interesting features, it is not a proper representation of water. But it can give a *clue* to a better physical model of the system, leading to the same features of interest.

4. Conclusion

We have in this paper refined a protein model (Hansen *et al.*, 1998a, 1999; Bakk *et al.*, 2000) by increasing the number of water molecules that can access the hydrophobic interior of the protein. The refined model exhibits both the hot and cold unfolding transitions. By increasing the number of water molecules that can access the hydrophobic parts of the protein interior, we have shown that the separation between the hot and cold unfolding transition peaks in the heat capacity curve can be made more sharply peaked, and can thus come closer in comparison to the earlier protein models. This is more consistent with the experimental data. By assuming the water–protein interactions to be two level, which is a speculative assumption, the heat capacity peak corresponding to the cold transition becomes smaller than the peak at the hot transition.

This is in agreement with experimental data, and opposite to the situation found in the earlier protein models of Hansen *et al.* (1998a, 1999) and Bakk *et al.* (2000).

A. B. thanks the Research Council of Norway for financial support (Contract No. 129619/410). A. H. thanks NORDITA and Niels Bohr Institute for warm hospitality and support.

REFERENCES

ABKEVICH, V. I., GUTIN, A. M. & SHAKHNOVICH, E. I. (1995). Impact of local and nonlocal interactions on thermodynamics and kinetics of protein-folding. *J. Mol. Biol.* **252,** 460–471.

ANFINSEN, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181,** 223–230.

BAKK, A., HØYE, J. S., HANSEN, A., SNEPPEN, K. & JENSEN, M. H. (2000). Pathways in two-state protein folding. *Biophys. J.* **79,** 2722–2727.

BAKK, A., HANSEN, A. & SNEPPEN, K. (2001). Protein model exhibiting three folding transitions. *Physica A* **291,** 60–70.

BALDWIN, R. L. & ROSE, G. D. (1999). Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem. Sci.* **24,** 77–83.

BRYNGELSON, J. D. & WOLYNES, P. G. (1987). Spin-glasses and the statistical-mechanics of protein folding. *Proc. Natl Acad. Sci. U.S.A.* **84,** 7524–7528.

BRYNGELSON, J. D. & WOLYNES, P. G. (1990). A simple statistical field-theory of heteropolymer collapse with application to protein folding. *Biopolymers* **30,** 177–188.

CHEN, B.-LU & SCHELLMAN, J. A. (1989). Low-temperature unfolding of a mutant of phage T4 lysozyme. 1. Equilibrium studies. *Biochemistry U.S.A.* **28,** 685–691.

CREIGHTON, T. E. (1993). *Proteins - Structures and Molecular Properties*, 2nd edn. Chapter 5. New York: W. H. Freeman and Company.

DILL, K. A., FIEBIG, K. M. & Chan, H. S. (1993). Cooperativity in protein-folding kinetics. *Proc. Natl Acad. Sci. U.S.A.* **90,** 1942–1946.

FERSHT, A. R. (1995). Optimization of rates of protein-folding: The nucleation-condensation mechanism and its implications. *Proc. Natl Acad. Sci. U.S.A.* **92,** 10869–10873.

GRIKO, YU. V., PRIVALOV, P. L. & VENYAMINOV, S. YU & KUTYSHENKO, V. P. (1988). Thermodynamic study of the apomyoglobin structure. *J. Mol. Biol.* **202,** 127–138.

HANSEN, A., JENSEN, M. H., SNEPPEN, K. & ZOCCHI, G. (1998a). Statistical mechanics of warm and cold unfolding in proteins. *Eur. Phys. J. Ser. B* **6,** 157–161.

HANSEN, A., JENSEN, M. H., SNEPPEN, K. & ZOCCHI, G. (1998b). A hierarchical scheme for cooperativity and folding in proteins. *Physica A* **250,** 355–361.

HANSEN, A., JENSEN, M. H., SNEPPEN, K. & ZOCCHI, G. (1999). Hot and cold denaturation of proteins: critical aspects. *Eur. Phys. J. Ser. B* **10,** 193–196.

ISING, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeit. Phys.* **31,** 253–258.

ITZHAKI, L. S., OTZEN, D. E. & FERSHT, A. R. (1995). The structure of the transition-state for folding of chymotrypsin inhibitor-2 analyzed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein-folding. *J. Mol. Biol.* **254,** 260–288.

LEVINTHAL, C. (1968). Are there pathways for protein folding? *J. Chim. Phys. Phys.-Chim. Biol.* **65,** 44–45.

MAKHATADZE, G. I. & PRIVALOV, P. L. (1995). Energetics of protein structure. *Adv. Prot. Chem.* **47,** 307–425.

PETICOLAS, W. L. & KURTZ, B. (1980). Transformation of the ϕ–ψ plot for proteins to a new representation with local helicity and peptide torsional angles as variables. *Biopolymers* **19,** 1153–1166.

PRIVALOV, P. L., GRIKO, YU. V. & VENYAMINOV, S. YU. (1986). Cold denaturation of myoglobin. *J. Mol. Biol.* **190,** 487–498.

PRIVALOV, P. L. (1990). Cold denaturation of proteins. *Biochem. Mol. Biol.* **25,** 281–305.

PRIVALOV, P. L. (1992). Physical basis of the stability of the folded conformations of proteins. In: *Protein Folding* (Creighton, T. E., ed.), pp. 83. New York: W. H. Freeman and Company.

PRIVALOV, P. L. (1996). Intermediate states in protein folding. *J. Mol. Biol.* **258,** 707–725.

PRIVALOV, P. L. (1997). Thermodynamics of protein folding. *J. Chem. Thermodyn.* **29,** 447–474.

SHAKHNOVICH, E. I. (1997). Theoretical studies of protein-folding thermodynamics and kinetics. *Curr. Opin. Struct. Biol.* **7,** 29–40.

Paper 3

ELSEVIER

Physica A 291 (2001) 60–70

PHYSICA A

www.elsevier.com/locate/physa

Protein model exhibiting three folding transitions

Audun Bakk[a,*], Alex Hansen[a], Kim Sneppen[b]

[a] *Department of Physics, Norwegian University of Science and Technology, NTNU, NO-7491 Trondheim, Norway*

[b] *NORDITA, Blegdamsvej 17, DK-2100 Copenhagen, Denmark*

Received 31 July 2000

Abstract

We explain the physical basis of a very simple hierarchical model for small globular proteins with water interactions. The water is supposed to accesses the protein interior in an "all-or-none" manner during the unfolding of the protein chain. As a consequence of this mechanism (somewhat speculative), the model exhibits fundamental aspects of protein thermodynamics, as cold and warm unfolding of the polypeptide chain. Decreasing the temperature below the cold unfolding the protein folds again. Accordingly, the heat capacity has three characteristic peaks. The cold and warm unfolding has a sharpness close to a two-state system, while the cold folding yields a less sharp transition. Interestingly, the entropy of the protein chain drives both the cold folding and the warm unfolding.

PACS: 05.70.Jk; 87.14.Ee; 87.15.Cc; 87.10.+e

Keywords: Protein folding; Hydrophobicity; Phase transition

1. Introduction

In order to have a precise function in the biological "machinery", it is important for proteins to have an unique conformation at physiological temperatures. This is termed the *native* (folded) state. Anfinsen [1] proved in his famous experiment with ribonuclease the important fact that the folding of the polypeptide chain is thermodynamically determined.

One simplified view of protein folding is that the protein is supposed to follow a specific *folding pathway* of conformations in the Gibbs free energy landscape [2–9].

* Corresponding author.
E-mail address: audun.bakk@phys.ntnu.no (A. Bakk).

0378-4371/01/$ - see front matter
PII: S0378-4371(00)00611-7

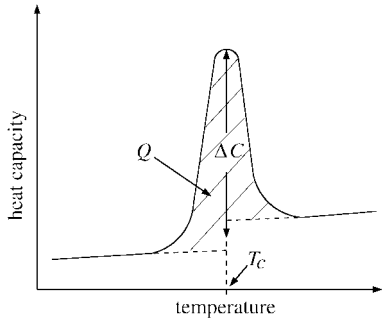

Fig. 1. Schematic illustration of the heat capacity around an unfolding transition showing the parameters in the van't Hoff enthalpy relation (Eq. (1)). T_c is the transition temperature, Q (area of the peak) is the released energy (latent heat) and ΔC is the peak height of the transition.

This is a picture of a folding protein that is forced to follow a specific "path" of successive conformational steps of increasing structural order. We will use this pathway assumption in this paper.

A protein in physiological environments (pH, ionic strength, etc.) and temperatures is packed in a very compact way. An increase of the temperature will eventually denaturate (unfold) the protein. Other ways to unfold the protein are for instance to change the pressure, denaturant concentration or the pH. The fact that proteins also unfold at low temperatures, termed as *cold unfolding* [10,11], makes the system very unusual. A major difficulty in experiments of cold unfolding is that the temperature is around and below the freezing point of water. In a frozen aqueous solution, one cannot observe any conformational transitions [12].

A general feature on small globular proteins is that they *thermodynamically* seems to unfold in an "all-or-none" manner. This means that they unfold cooperatively without noticeable intermediates [13–18], with a deviation from a two-state system not exceeding 5%. The deviation from a single macroscopic system can be explained by the presence of unstable intermediates [8,14,19]. It is worth noting that all these experiments have been done only for the warm unfolding. The occurrence of intermediate states in larger proteins [2,3,20] is not a contradiction to the two-state behavior in the experiments in Refs. [13–18], because the latter only consider small globular proteins.

The van't Hoff enthalpy relation (for heat of reaction) [14,21]

$$\Delta H = \alpha k_B T_c^2 \frac{\Delta C}{Q} \tag{1}$$

is a powerful way to quantify the *sharpness* of a smoothed out first-order phase transition. As shown in Fig. 1, T_c is the transition temperature (at the middle of the peak), Q which is the same as ΔH (no pressure) is the released energy (latent heat), and ΔC is the peak height of the transition. α is a dimensionless proportionality factor, and k_B is the Boltzmann constant. For a given ΔH and Q, then the value of α is inversely proportional to ΔC. In this respect a smaller α corresponds to a sharper transition.

In this article we will explain the physical basis of a protein model that reformulates the water interactions proposed in earlier models by Hansen et al. [5,6] and Bakk et al. [8,9]. We will compare thermodynamical quantities, as the heat capacity, to experiments. The protein is also investigated in a temperature region below accessible experimental data.

2. Modeling the protein

2.1. The polypeptide chain

The polypeptide chain is modeled as in earlier articles by Hansen et al. [5–7] and Bakk et al. [8,9], where the protein is supposed to follow a pathway as described in Section 1. The protein is equipped with N contacts. As a very simple assumption we say that all contacts contribute equally with respect to energy and entropy. Due to the fact that a protein is a complex three-dimensional system, a folded contact likely has non-local interactions with respect to the amino acid sequence in the polypeptide chain. A weakness in the model is that it has no distinct conformational information for specific proteins within the contacts. However, in this work we want to investigate aspects of the *mechanism* for protein folding. In this respect the generality of the model is an advantage.

We parameterize the protein in a way analogous to Zwanzig [22], where we assign binary variables $\Psi_i \in \{0,1\}$ describing an open and closed contact i, respectively. The pathway implies the constraint

$$\Psi_i \geqslant \Psi_{i+1}\,, \tag{2}$$

because a folded contact i is not assumed to unfold when a contact $j > i$ is folded. This means that it is difficult to unfold a part of the protein within an already larger folded structure. In order to implement this into a Hamiltonian we introduce a second set of binary variables $\xi_i \in \{1, -B\}$, and $B \gg 1$. The state B is assumed to have a degeneracy $f-1$. Thus, for a system of N contacts the chain–chain Hamiltonian becomes

$$\mathscr{H}_c = -\varepsilon_0 \sum_{i=1}^{N} \Psi_i \xi_i\,, \tag{3}$$

and ε_0 is the energy gain to fold one contact [23,24]. From the constrains in Eq. (2) and requiring for simplicity $B \to \infty$, the Hamiltonian in Eq. (3) can be reformulated by the transformation $\Psi_i = \varphi_1 \varphi_2 \cdots \varphi_i$, where $\varphi_j \in \{0,1\}$ are binary variables. In particular $\Psi_1 = \varphi_1$. The value $\phi_i = 0$ means an unfolded contact, while $\phi_i = 1$ is equivalent to a folded contact. The product term meets the assumption about a folding pathway. Thus, Eq. (3) becomes

$$\mathscr{H}_c = -\varepsilon_0(\phi_1 + \phi_1\phi_2 + \cdots + \phi_1\phi_2\cdots\phi_N)\,. \tag{4}$$

All open contacts will "flap" freely due to zero energy cost (see Eq. (4)). Having in mind the degeneracy $f-1$ of the state B of the ξ_i-variable, each open contact has a

degeneracy f. The parameter f is interpreted as the relative increase in the degrees of freedom for an unfolded contact compared to a folded contact.

We note that the chain–chain Hamiltonian emerges from a very simple model, even simpler than the already simple pathway models by Zwanzig [22], Micheletti et al. [25], Galzitskaya and Finkelstein [26], and Muñoz and Eaton [27] applied to study different aspects of protein folding. However, none of these consider the heat capacity aspect, and in particular the phenomena of cold and warm unfolding. The latter seems to be connected closely to the water interactions [10], as discussed below.

2.2. Water interactions

Interactions between water and protein surface are important. Proteins are during the evolution "designed" to interact with water, simply because they are exposed to water in vivo. Makhatadze and Privalov [28] state that in sum, hydration effects destabilize the native state, and decreasing temperature implies increasing destabilizing action. The water that accesses the exposed hydrophobic protein interior during unfolding is supposed to obtain an "ice-like" structure around the apolar surfaces [29]. Hence, this structured water will both decrease the entropy and the energy compared to "free" water [30], and thus impacts the thermodynamics of the system.

Hansen et al. [5,6] proposed a simple model where each water molecule associated to a contact was represented by a "ladder" of g equidistant energies,

$$\omega_i = \begin{cases} -\varepsilon_w + (g-1)\delta\,, \\ \vdots \\ -\varepsilon_w + 2\delta\,, \\ -\varepsilon_w + \delta\,, \\ -\varepsilon_w\,, \end{cases} \tag{5}$$

which we will also apply in the model considered in this text. The interpretation of ω_i is the energy difference between a "frozen" water molecule, associated to the unfolded parts of the protein, and a "free" water molecule in the bulk.

The observable states in a small globular protein is either the unfolded ($\phi_1 = 0$), with water bound to the surface that uncovers during unfolding of the protein, or the folded state ($\phi_1 \cdots \phi_N = 1$) with no water in the protein interior. No intermediate states are thermodynamically detected for small globular proteins [31]. Hence, one cannot know for sure how the water enters the protein interior during the unfolding. Hansen et al. [5,6] and Bakk et al. [8,9] have earlier only considered that the amount of water interactions increase in proportion to the number of unfolded contacts, and with that the contact energy of the chain. In this paper we study, as a more speculative assumption, the case when a macroscopic contribution of water enters the protein surface when the last contact is unfolded.

We note that Eq. (5) is the quantized energy levels of a magnetic dipole in an external field. In the continuum limit where $g \to \infty$ (with $g\delta$ finite), a classical magnetic dipole in an external field is obtained, and this again is analogous to an electrical

dipole in an external electrical field. The dipolar water molecules are, in addition to water–water interactions, exposed to an electrical field from the permanent, and induced charges on the protein surface. Thus, Eq. (5) is a effective representation of that.

By using the same notation as in Eq. (4), we propose the Hamiltonian that corresponds with the water–protein interactions

$$\mathscr{H}_w = (1 - \phi_1\phi_2\cdots\phi_N)(\omega_1 + \omega_2 + \cdots + \omega_M)\,, \tag{6}$$

where M is the number of water molecules. The folded protein is a highly ordered and dense packed structure where no water can access the interior. Due to Eq. (4), unfolding of the last contact ($\phi_N = 0$) implies a less dense packing of the protein, and the cavities are now supposed to be big enough to let water access the interior of the protein. The next step, unfolding of contact $N-1$, implies likely an even lesser dense packing, and allows more water in the protein interior. We assume in this work that the water entering upon unfolding of contact $N-1$ will not interact with the protein surface, because it is then regarded as a second layer containing "free water". Cohn and Edsall [32] state that roughly a monolayer of water is bounded to the protein, implying that the protein is only interacting with the first monolayer, thus the second, and third, etc., water layers, successively entering the protein during unfolding, are regarded as bulk water. Thus, according to the latter possible (but somewhat speculative) explanation of how the water accesses the apolar interior of the protein, unfolding of contacts $i < N$ does not contribute energy to the water Hamiltonian ($\mathscr{H}_w$) and consequently not to the thermodynamics.

2.3. The statistical framework

The system Hamiltonian ($\mathscr{H}$) describing both chain-specific energy ($\mathscr{H}_c$) and water interactions ($\mathscr{H}_w$) is

$$\begin{aligned}\mathscr{H} = \mathscr{H}_c + \mathscr{H}_w = &-\varepsilon_0(\phi_1 + \phi_1\phi_2\cdots + \phi_1\phi_2\cdots\phi_N) \\ &+ (1 - \phi_1\phi_2\cdots\phi_N)(\omega_1 + \omega_2 + \cdots + \omega_M)\,.\end{aligned} \tag{7}$$

Let now Z_i be term number i in the partition function which corresponds with folding of all contacts $\leqslant i$ (pathway assumption), thus

$$Z_i = f^{N-i}\mathrm{e}^{\mathrm{i}\varepsilon_0\beta}\left(\mathrm{e}^{\varepsilon_w\beta}\frac{1 - \mathrm{e}^{-g\delta\beta}}{1 - \mathrm{e}^{-\delta\beta}}\right)^M \quad (i < N)\,. \tag{8}$$

$\beta \equiv 1/T$ is a rescaled inverse absolute temperature in which the Boltzmann constant is absorbed. Z_0 is the term where all contacts are zero, i.e., a complete unfolded protein, while Z_N corresponds with a folded protein. The factor f^{N-i} in Eq. (8) are the chain specific degrees of freedom deliberated from the $N-i$ unfolded contacts. $\mathrm{e}^{\mathrm{i}\varepsilon_0\beta}$ is the Boltzmann factor from i contact energies $-\varepsilon_0$. The last term in brackets is simply the sum over all distinct levels in one water molecule raised to the power of the number of independent water molecules M bound to the unfolded parts of the protein. We

assume that $\delta\beta \ll 1$ (i.e., $g \to \infty$), which is equal to an infinite small level spacing in Eq. (5). A first-order Taylor expansion of the denominator in Eq. (8), yields

$$Z_i = f^N \left(\frac{e^{\varepsilon_w \beta}}{\delta\beta}\right)^M e^{i(\varepsilon_0\beta - \ln f)} \quad (i < N), \tag{9}$$

assuming $1 - e^{-g\delta\beta} \approx 1$ when $g \to \infty$. The last term in the partition function (Z_N) corresponds with a complete folded protein, where there are g^M degrees of freedom from M unbound water molecules and N contact energies $-\varepsilon_0$, yielding

$$Z_N = g^M e^{N\varepsilon_0\beta}. \tag{10}$$

By summing up the Z_i terms in Eqs. (9) and (10), we obtain the partition function for the protein

$$Z = \sum_{i=0}^{N} Z_i = f^N g^M \left[\left(\frac{a e^{\mu\beta}}{\beta}\right)^M \frac{1 - r^N}{1 - r} + r^N\right], \tag{11}$$

where $r \equiv e^{\beta - \ln f}$, $a \equiv \varepsilon_0/(g\delta)$ and $\mu \equiv \varepsilon_w/\varepsilon_0$. The inverse temperature is rescaled by $\varepsilon_0\beta \to \beta$. The parameter μ in Eq. (11) measures the strength of the water interactions relative to the chain contact energy. Thus, μ may be interpreted as an *effective chemical potential*. Changing μ may mean to add denaturants, change pH or salt concentration, etc.

The order parameter n in this system measures the degree of folding, i.e., the mean number of folded contacts divided by N

$$n = \frac{\sum_{i=0}^{N} i Z_i}{Z} = \frac{1}{N} \frac{(a e^{\mu\beta}/\beta)^M((N-1)r^{N+1} - Nr^N + r)/(1-r)^2 + Nr^N}{(a e^{\mu\beta}/\beta)^M(1 - r^N)/(1-r) + r^N}. \tag{12}$$

$n = 0$ corresponds with an unfolded protein, while $n = 1$ is a completely folded protein.

3. Calculations and discussion

The heat capacity is $C = \beta^2 \partial^2(\ln Z)/\partial\beta^2$. Fig. 2 shows a typical plot of the heat capacity $C(T)$ with three peaks (numbered 1–3 from left). These characteristic peaks correspond with three critical transition temperatures: T_1, T_2 and T_3, measuring the temperatures at the respective peak maxima. The corresponding order parameter $n(T)$ in Fig. 3, calculated from Eq. (12), indicates that the protein is essentially folded at $T < T_1$ and $T_2 < T < T_3$, while the protein is nearly unfolded at $T_1 < T < T_2$ and $T > T_3$. From this picture it is reasonable to state that the physiological temperature interval is between peaks 2 and 3. Accordingly, with reference to this temperature region, we call peak 1 for *cold folding* and peaks 2 and 3, respectively, for *cold* and *warm unfolding*. Peaks 2 and 3 are both observed in experiments [10,11], and they are also seen in the model of Hansen et al. [5,6] and Bakk et al. [8,9]. The model in this work has, in addition to the cold and warm unfolding, the peculiarity of *cold folding*.

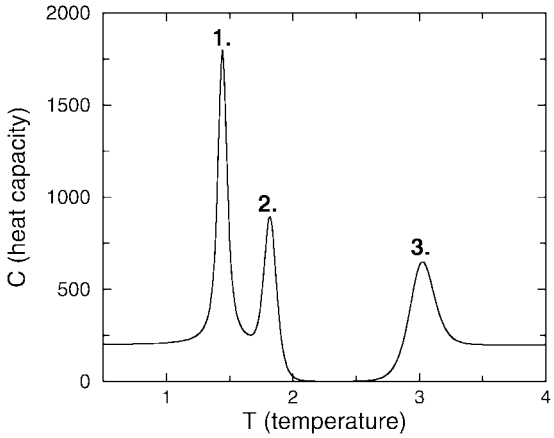

Fig. 2. Heat capacity $C(T)$. Parameters chosen: $a = 0.077$, $\mu = 3.3$, $f = 2$, and $N = M = 200$, exhibiting three peaks. With reference to the temperature region between peaks 2 and 3 (physiological temperatures) we call the transitions: (1) *cold folding*; (2) *cold unfolding*; and (3) *warm unfolding*.

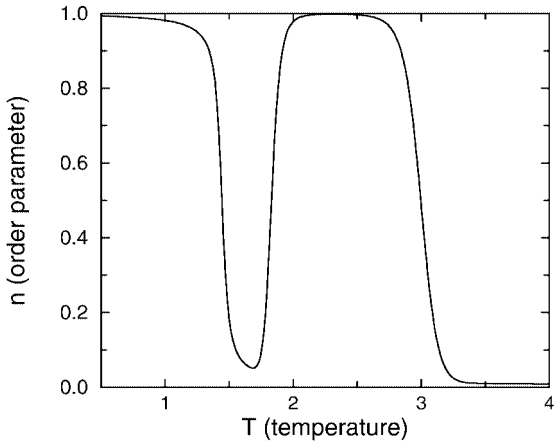

Fig. 3. Order parameter $n(T)$ corresponding with Fig. 2, describing the degree of folding. Parameters are chosen equivalent to Fig. 2. $n = 0$ corresponds with an unfolded protein while $n = 1$ is interpreted as a completely folded protein.

Experiments on cold unfolding are very difficult because most proteins unfold below the freezing point of water. Chen and Chellman [11] and Privalov et al. [33] have all done experiments where the cold unfolding temperature is elevated by denaturants, but denaturants make the interpretation of the data very difficult. However, Privalov [10] did experiments in super-cooled water, which is easier to interpret. Unfortunately, he was not able to detect the sharpness of the cold unfolding, and not at all the heat capacity below the cold unfolding. This means that our model *may* predict a cold folding transition at a temperature below the cold unfolding transition.

At temperatures below the cold folding ($T \to 0$) analysis of the model gives $n = 0.99$, i.e., only the last contact is unfolded, and corresponds to the global energy minimum. From Fig. 2 it is seen that $T_1 = 1.5$. In Eq. (11) the critical $r \equiv e^{1/T - \ln f} = 1$, which implies $T_1 = 1/\ln 2 \approx 1.44$ when $f = 2$. Hence, this is nothing but an initiated transition of the chain entropy. An increase of the temperature from T_1 takes the protein through a nearly unfolded state, whereupon the protein folds again at $T_2 = 1.8$. The temperature is now so high that the energy of water (Eq. (5)) is small (thermal exited) compared to the chain contact energy ε_0, thus the protein prefers to fold again. Further increase of the temperature causes warm unfolding at $T_3 = 3.0$, because then the entropy of the chain again dominates the Gibbs free energy. It is interesting to note that the entropy of the chain causes two transitions, both cold folding and warm unfolding.

We now turn our interest to the sharpness of the transitions, i.e., the value of the parameter α in the van't Hoff enthalpy relation (Eq. (1)). For $M = 200$, $\alpha = 4.0$ both for the cold and warm unfolding. This means that the protein is thermodynamically regarded as a two-state system that folds in an "all-or-none" manner. Privalov [14] measured $\alpha = 4.0$ for the warm unfolding. As far as we know there are no experiments on the sharpness of the cold unfolding, but Privalov [10] indicates a sharpness for the cold unfolding as well, which agrees with our model. The cold folding transition has a sharpness $\alpha = 12.0$. This value is typical for a transition where one has small energy/entropy differences between the folded/unfolded states and the intermediate states. Remember that the "folded" state for $T < T_1$ is actually the first contact unfolded. Thus, the unfolding will essential depend on the polypeptide chain with the Hamiltonian in Eq. (4), which can be shown to correspond with a sharpness $\alpha = 12.0$ [8,19].

The consequence of a decreased μ is increasing separation between the cold and warm unfolding, as seen in Fig. 4. This makes sense. A smaller μ is equivalent to a relatively smaller ε_w compared to ε_0 (see Eq. (11)), i.e., it is less favorable for the protein interior to bind water. The consequence is that the protein prefers to be folded in a larger temperature interval, in which the water is expelled to the bulk. However, the transition temperature T_1 is not changed because this transition is given by the value $T_1 = 1/\ln f$. It is also seen that a smaller μ is qualitatively equivalent to a smaller a.

An increase in M is the same as a decrease in N, because then the water becomes more important relative to the chain, and allows a broader separation between T_2 and T_3. The broader separation is also seen for decreasing f, because this is equivalent to a larger M. Further increase of μ will eventually merge peaks 1 and 2. It is interesting to note that $\alpha = 4$ for the merged peaks, because then the transition of the M water molecules is energetically dominated, which causes the transition at T_2 (Fig. 2). Consequently, the cold folding transition is not very robust against change in the parameters. In order to possibly reveal experimentally such a mechanism, the measurements must be more reliable at/below the cold unfolding transition.

In sum, the qualitative change from Fig. 3 to Fig. 4, by a decreasing μ, is also obtained by a decrease of a, f or N or an increase of M.

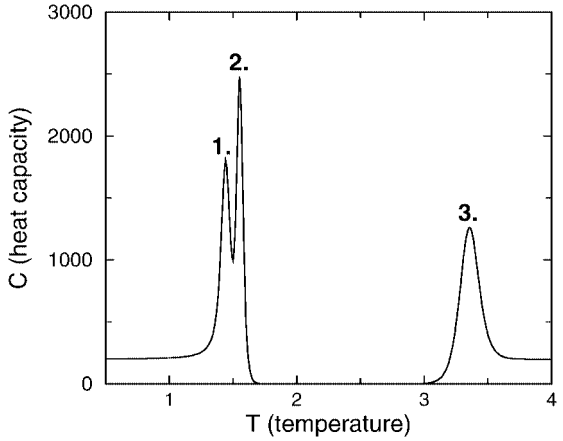

Fig. 4. Heat capacity where the effective chemical potential is $\mu = 3.2$, slightly decreased from the value in Fig. 2 ($\mu = 3.3$). All other parameters are chosen as in Fig. 2. The qualitative picture is a broader separation between T_2 and T_3 compared to Fig. 2.

Finally, we note that a slightly reformulated version of the model seems to be in very good qualitatively correspondence with experiments on heat capacity exhibiting cold and warm unfolding [34].

4. Summary

We have, in this work, studied a hierarchical protein model with water interactions. The model is based on earlier models by Hansen et al. [5,6] and Bakk et al. [8,9]. In contrast to these similar models, where the water amount was supposed to increase linearly to the degree of unfolding of the polypeptide chain, we have, as a more speculative assumption, studied the situation where a macroscopic amount of water access the protein interior during unfolding of the last contact.

With reference to physiological temperatures we find that the protein exhibits cold and warm unfolding transitions, which is an experimental fact to small globular proteins [10,11]. These transitions are associated by a sharpness indicating, from a thermodynamical point of view, a two-state system which is also experimentally established [13–18]. Decreasing the temperature further below the cold unfolding region the protein folds again (cold folding). This folding, caused by the chain entropy, has a less sharp transition. In sum, the model exhibits three unfolding/folding transitions. The cold folding transition seems not to be very robust against a parameter change, in contrast to the cold and warm unfolding.

It is interesting to note that *both* the cold and the warm unfolding is due to the polypeptide chain entropy.

Acknowledgements

We thank J.S. Høye for enthusiastic and enlightening discussions. A.H. thanks NORDITA and Niels Bohr Institute for warm hospitality and support. A.B. thanks the Research Council of Norway for financial support (Contract No. 129619/410).

References

[1] C.B. Anfinsen, Principles that govern the folding of a protein chain (Nobel Lecture), Science 191 (1973) 223–230.
[2] R.L. Baldwin, G.D. Rose, Is protein folding hierarchic? I. Local structure and peptide folding, Trends Biochem. Sci. 24 (1999) 26–33.
[3] R.L. Baldwin, G.D. Rose, Is protein folding hierarchic? II. Folding intermediates and transition states, Trends Biochem. Sci. 24 (1999) 77–83.
[4] Yi Wang, D. Shortle, The equilibrium folding pathway of staphylococcal nuclease: identification of the most stable chain–chain interactions by NMR and CD spectroscopy, Biochemistry USA 34 (1995) 15895–15905.
[5] A. Hansen, M.H. Jensen, K. Sneppen, G. Zocchi, Statistical mechanics of warm and cold unfolding in proteins, Eur. Phys. J. Ser. B 6 (1998) 157–161.
[6] A. Hansen, M.H. Jensen, K. Sneppen, G. Zocchi, A model for the thermodynamics of globular proteins, Physica A 270 (1999) 278–287.
[7] A. Hansen, M.H. Jensen, K. Sneppen, G. Zocchi, A hierarchical scheme for cooperativity and folding in proteins, Physica A 250 (1998) 355–361.
[8] A. Bakk, J.S. Høye, A. Hansen, K. Sneppen, M.H. Jensen, Pathways in two-state protein folding, Biophys. J. 79 (2000) 2722–2727.
[9] A. Bakk, J.S. Høye, A. Hansen, K. Sneppen, Thermodynamical implications of a protein model with water interactions, J. Theor. Biol., e-print: cond-mat/0007078, submitted for publication.
[10] P.L. Privalov, Cold denaturation of proteins, Biochem. Mol. Biol. 25 (1990) 281–305.
[11] B. Chen, J.A. Schellman, Low-temperature unfolding of a mutant of phage T4 lysozyme, 1. Equilibrium Studies, Biochemsitry USA 28 (1989) 685–691.
[12] T.E. Creighton, Protein Folding, W.H. Freeman and Company, New York, 1992 (Chapter 3).
[13] P.L. Privalov, N.N. Khechinashvili, A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study, J. Mol. Biol. 86 (1974) 665–684.
[14] P.L. Privalov, Stability of proteins, small globular proteins, Adv. Protein Chem. 33 (1979) 167–241.
[15] G.I. Makhatadze, P.L. Privalov, Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration, J. Mol. Biol. 232 (1993) 639–659.
[16] S.E. Jackson, A.R. Fersht, Folding of chymotrypsin inhibitor 2. Evidence for a two-state transition, Biochemistry USA 30 (1991) 10428–10435.
[17] S.J. Bae, J.M. Sturtevant, Thermodynamics of thermal unfolding of eglin c in the presence and absence of guanidinium chloride, Biophys. Chem. 55 (1995) 247–252.
[18] Y.V. Griko, G.I. Makhatadze, P.L. Privalov, R.W. Hartley, Thermodynamics of barnase unfolding, Protein Sci. 3 (1994) 669–676.
[19] P.G. Dommersnes, A. Hansen, M.H. Jensen, K. Sneppen, Parametrization of multiple pathways in proteins: fast folding versus tight transitions, Proc. Roy. Soc. Ser. A, submitted for publication.
[20] D. Shortle, Yi Wang, J.R. Gillespie, J.O. Wrabl, Protein folding for realists: a timeless phenomenon, Protein Sci. 5 (1996) 991–1000.
[21] H.B. Callen, Thermodynamics and an Introduction to Thermostatics, 2nd Edition, Wiley, New York, 1985 (Chapter 13).
[22] R. Zwanzig, Simple model of protein folding kinetics, Proc. Natl. Acad. Sci. USA 92 (1995) 9801–9804.
[23] J.D. Bryngelson, P.G. Wolynes, Spin-glasses and the statistical-mechanics of protein folding, Proc. Natl. Acad. Sci. USA 84 (1987) 7524–7528.

[24] J.D. Bryngelson, P.G. Wolynes, A simple statistical field-theory of collapse with application to protein folding, Biopolymers 30 (1990) 177–188.
[25] C. Micheletti, J.R. Banavar, A. Maritan, F. Seno, Protein structures and optimal folding from a geometrical variational principle, Phys. Rev. Lett. 82 (1999) 3372–3375.
[26] O.V. Galzitskaya, A.V. Finkelstein, A theoretical search for folding/unfolding nuclei in three-dimensional protein structures, Proc. Natl. Acad. Sci. USA 96 (1999) 11299–11304.
[27] V. Muñoz, W.A. Eaton, A simple model for calculating the kinetics of protein folding from three-dimensional structures, Proc. Natl. Acad. Sci. USA 96 (1999) 11311–11316.
[28] G.I. Makhatadze, P.L. Privalov, Energetics of Protein Structure, Adv. Protein Chem. 47 (1995) 308–424.
[29] H.S. Frank, M.W. Evans, Free volume and entropy in condensed systems. III entropy in binary liquid mixtures; partial molal entropy in dilute solutions; structure and thermodynamics in aqueous electrolytes, J. Chem. Phys. 13 (1945) 507–532.
[30] P.L. Privalov, G.I. Makhatadze, Contribution of hydration and non-covalent interactions to the heat capacity effect on protein unfolding, J. Mol. Biol. 224 (1992) 715–723.
[31] P.L. Privalov, Intermediate states in protein folding, J. Mol. Biol. 258 (1996) 707–725.
[32] E.J. Cohn, J.T. Edsall, Proteins, Amino Acids and Peptides, Reinhold Publishing, New York, 1943.
[33] P.L. Privalov, Yu.V. Griko, S.Yu. Venyaminov, Cold denaturation of myoglobin, J. Mol. Biol. 190 (1986) 487–498.
[34] A. Bakk, J.S. Høye, A. Hansen, Heat capacity of protein folding, Biophysical Journal, submitted for publication.

Paper 4

Bakk, A ; Høye, JS ; Hansen, A: Heat Capacity of Protein Folding. *Biophys. J.*, 81(2001): no. 2, 710-714

Paper 5

ELSEVIER

Physica A 303 (2002) 286–294

PHYSICA A

www.elsevier.com/locate/physa

Microscopic argument for the anomalous hydration heat capacity increment upon solvation of apolar substances

Audun Bakk*, Johan S. Høye

Department of Physics, Norwegian University of Science and Technology, NTNU, NO-7491 Trondheim, Norway

Received 27 June 2001

Abstract

The contribution of the hydration heat capacity upon solvation of apolar molecules is derived by applying equilibrium statistical mechanics on a simple model, which we compare to experimental data on the linear alkanes: methane, ethane, and propane, and to the aromatic compounds: benzene and toluene. The model is based on a microscopic consideration, where we assign an effective bending energy of the hydrogen bonds between the water molecules in the solvation shell around the solute molecule. Thus, we assume that the hydrophobicity is only connected to forces between the water molecules in the solvation shell, and not to forces between the apolar molecule and the surrounding water. We find that the model, by fitting the parameters, corresponds well to the experimental data. The proportionality of the heat capacity versus surface area is also discussed.

Keywords: Apolar molecule; Hydrophobicity; Hydrogen bond; Heat capacity

1. Introduction

The thermodynamics related to aqueous solvation of apolar molecules is of interest in various respects. E.g., in recent years there has been a rapidly growing amount of protein research. The understanding of water interactions upon unfolding of proteins seems to be important in order to understand their stability in general [1–4], besides the more specific feature of cold and warm unfolding of small globular proteins [5–13]. Thus, the understanding of the thermodynamics related to solvation of smaller

* Corresponding author.
E-mail addresses: audun.bakk@phys.ntnu.no (A. Bakk), johan.hoye@phys.ntnu.no (J.S. Høye).

0378-4371/02/$ - see front matter
PII: S0378-4371(01)00494-0

molecules may be beneficial in order to understand and predict the structure and stability of large biomolecules [14].

In this work we want to study of the pure hydration effect upon solvation of small apolar molecules. It is known from experiments that the heat capacity change upon aqueous dissolution of apolar molecules from the gaseous state is positive and proportional to the density of solute molecules [15], and that this change is decreasing with increasing temperature [16]. It is also known that the solution of an apolar substance in water, at room temperature, is associated with a negative entropy change, which decreases in absolute value with increasing temperature [17]. The hydration effect of an apolar molecule has traditionally been regarded as a gradual melting of some kind of ordered "ice-like" shell around these compounds [18].

In order to explain the anomalous heat capacity increment of solvation of apolar substances, we were inspired by the distorted hydrogen bond model introduced by Pople [19,20]. By applying statistical mechanics on a simplified model we can calculate its heat capacity and compare it to experimental data on methane, ethane, and propane from Naghibi et al. [21,22], and on benzene and toluene from Makhatadze and Privalov [23].

2. Theoretical

Solvation of a molecule in water is a complex process that includes formation of a cavity in water, and there are interactions between the water molecules and the solute molecule. Due to these interactions the water molecules are rearranged around the inserted molecule [24,25].

This work is based upon the simple assumption of considering the solvation shell around the solute molecule as an "ice-shell" formation, due to Frank and Evans [18], i.e., an increased number of hydrogen bonds compared to bulk water. We emphasize that this, which is meant to be an effective description of apolar solvation, may be in contrast to recent advances in techniques applied on the structural behavior of apolar solvation. Especially, we note high intensity neutron scattering [26–29] and X-ray absorption spectroscopy investigations [30–32], where there seems to be no evidence for *significant* increase of the internal water interactions around apolar solutes. However, in this work we will nevertheless conceptually use a kind of "ice-shell" analogy that enhance formation of hydrogen bonds compared to bulk water, since the reduction of both enthalphy [33,34] and entropy [35,36] seems to be well established upon apolar solvation.

We define here the *excess energy* (ΔE), as the energy difference between a hydrogen bond in the "ice-shell" around the apolar molecules, and "free" bulk water. The excess energy is assumed to be independent of the intrinsic properties of the solute itself [37]. ΔE is negative and is supposed to be of the size of breaking a hydrogen bond in water (5.5 kJ mol^{-1}) [37,38]. In order to estimate the energy of a hydrogen bond, we use the distorted hydrogen bond model introduced by Pople [19,20]. Water has four localized

orbitals, and to first approximation we may regard the bending of the hydrogen bonds to be independent of each other. So in a simplified model each hydrogen bond can then be considered to act in a way similar to an electric field that directs a dipolar moment. Another way to consider this is that the rectifying effect of the apolar surface, together with the hydrogen bonds acts like an effective electric field upon the dipole moment of a water molecule. With this latter simplification, the breaking of hydrogen bond is modeled as a dipole moment that rotates in an electric field. This model is the same as the classical Heisenberg [39] model for a magnetic moment. With this, the energy for *one* rotator or hydrogen bond becomes

$$E(\vartheta) = -\varepsilon\cos\vartheta\,, \tag{1}$$

where ε is a bending distortion constant. The angle ϑ is the polar angle of orientation.

The partition function that follows from Eq. (1) is

$$\begin{aligned} Z &= \int_0^{2\pi} \mathrm{d}\varphi \int_0^{\pi} \mathrm{d}\vartheta \sin\vartheta \mathrm{e}^{-E(\vartheta)/(k_\mathrm{B}T)} \\ &= \frac{4\pi T}{b}\sinh(b/T)\,, \end{aligned} \tag{2}$$

where $b = \varepsilon/k_\mathrm{B}$ (k_B is the Boltzmann constant) and φ is the azimuthal angle. This yields the internal energy per mole of solute:

$$U = aT^2\frac{\partial}{\partial T}(\ln Z) = a[T - b\coth(b/T)]\,. \tag{3}$$

Here $a = N_\mathrm{H}N_\mathrm{A}k_\mathrm{B} = N_\mathrm{H}R$, where N_H is the number of hydrogen bonds per solute molecule, N_A is Avogadro's number, and $R = 8.314\ \mathrm{J\,K^{-1}\,mol^{-1}}$ is the molar gas constant. From Eq. (3) we obtain the specific heat change upon apolar solvation per mole solute molecules:

$$\Delta C = \frac{\partial U}{\partial T} = a\left[1 - \left(\frac{b}{T\sinh(b/T)}\right)^2\right]. \tag{4}$$

Experimentally one measures the bending distortion constant ε_H per mole which we will identify with ε in the model above. Thus we have $b = \varepsilon_\mathrm{H}/R$.

One can note since this model is a classical one, the specific heat stays non-zero as $T \to 0$. However, we assume that this is appropriate in our case as only temperatures above the melting point of ice are considered. Also trying to take quantum effects into account would complicate solution of our model—which is a simplification of reality anyway. On the other hand, quantum effects are important at low temperatures as is the case with respect to rotations and vibrations of molecules. One should also keep in mind that the hydrogen atom is the lightest element such that one may not rule out noticeable quantum effects, even at room temperature, before more careful investigations are performed. The two-state model may thus act as a very simplified model of such quantization [23,37,40,41,50], which we believe is of less relevance of the temperatures considered here for apolar surfaces.

The parameter b in Eq. (4) contains the bending constant ε_H, which is actually a measure of the energy of breaking 1 mol of hydrogen bonds in ice and transfering them to "unbound" water. Gill et al. [37] points out that this is not the same energy as to fully break 1 mol of hydrogen bonds and "transfer them into vacuum", because an "unbound" water molecule will still interact with its neighbors, but with a weaker coupling compared to ice. Némethy and Scheraga [38] have estimated this value to $\varepsilon_H = 5.5$ kJ mol^{-1}, which is substantially lower than the widely quoted value 18.8 kJ mol^{-1} for breaking a hydrogen bond proposed by Pauling [42].

3. Results and discussion

We now want to compare Eq. (4) with experimental data for the hydration heat capacity of several small apolar molecules. All these data are based upon the definition that the heat capacity change is the difference in the heat capacity of the solute in its ideal gaseous phase and its water soluted phase, i.e., the Ben-Naim definition [43].

In Fig. 1 we plot the experimental data on methane, ethane, and propane from Naghibi et al. [21,22], and in Fig. 2 on benzene and toluene from Makhatadze and Privalov [23]. The continuous lines are based upon best fit of the parameters a and b in Eq. (4). All the parameters used in the plots are listed in Tables 1 and 2. The theoretical lines are clearly within the error-estimates, which is satisfactory in view of the simplicity of the model.

From Eq. (3) one sees that a is proportional to the number of hydrogen bonds. Naghibi et al. [21,22], and Makhatadze and Privalov [23] all conclude in that it is only the first solvation shell that is responsible for the large contribution to the heat capacity upon solvation of apolar substances. In order to check this in our model we simply investigate the fraction

$$\Phi = \frac{a}{A}\,, \tag{5}$$

where A is the accessible surface area for the water molecules which we take from Hermann [44]. Apparently from Table 1 the deviation from $\Phi \approx 2.0$ J K^{-1}mol^{-1} Å^{-2} is quite small for methane, propane, benzene, and toluene, while ethane has a distinct deviation from this value. This trend repeats for the parameter b where methane, propane, benzene, and toluene are all around $b = 600$ K. Again ethane deviates distinct from this value. Note that $b = 600$ K corresponds to $\varepsilon_H = 5.0$ kJ mol^{-1}, close to the estimated value from Némethy and Scheraga for breaking 1 mol of hydrogen bonds in ice and transfer them to bulk water [38].

In Fig. 1 we also present a plot based upon choosing a fixed value $b = 600$ K whereupon a is adjusted accordingly (dashed line). The corresponding values for a and Φ are listed in Table 2. The best fit b-values for benzene and toluene are both very close to $b = 600$ K, as seen when comparing Tables 1 and 2, and are accordingly not drawn. Methane and propane are both within the error-estimates, which is not the case for ethane. Thus, we cannot from our model conclude for these apolar molecules that

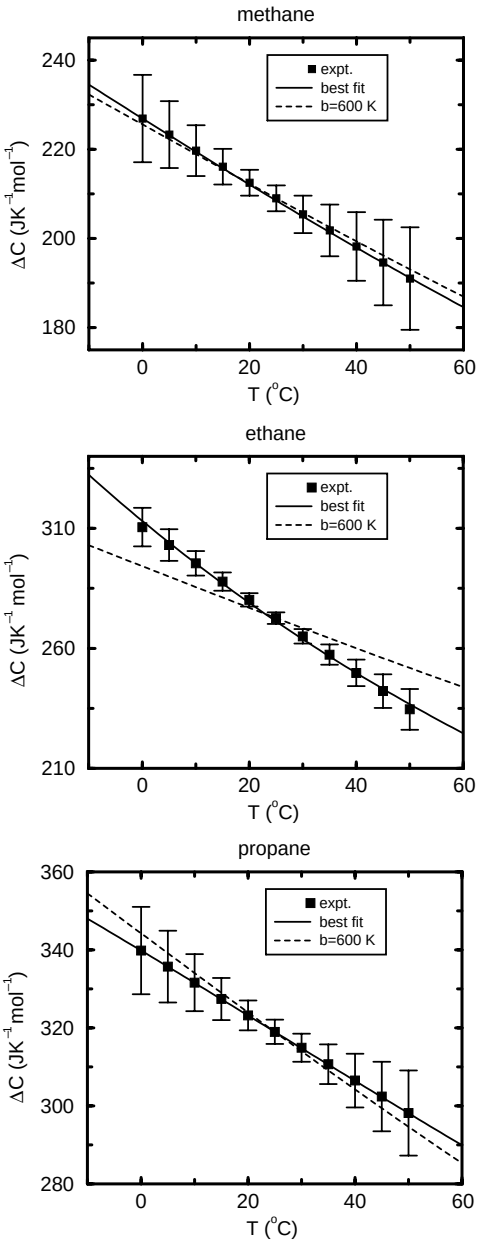

Fig. 1. Experimental data from Naghibi et al. [21,22] on the aqueous solvation contribution to the heat capacity for the linear alkanes ethane, methane, and propane as function of the absolute temperature. The continuous lines are best fit of the parameters a and b in Eq. 4, while the dashed lines correspond to best fit of parameters a with parameter b fixed to 600 K. The parameter values are listed in Tables 1 and 2.

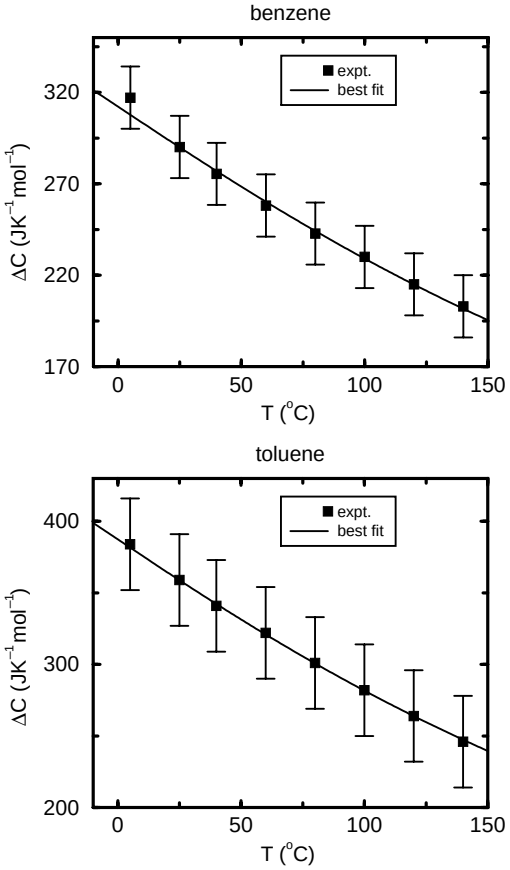

Fig. 2. Experimental data from Makhatadze and Privalov [23] on the aqueous solvation contribution to the heat capacity for the aromatic hydrocarbons benzene and toluene as function of the absolute temperature. The continuous lines are the best fit of the parameters a and b in Eq. (4). Values of the parameters are listed in Table 1.

Table 1
Best fit of the parameters a and b in Eq. (4) that correspond to the continuous lines in Figs. 1 and 2[a]

Solute	a (J K^{-1} mol^{-1})	b (K)	A (Å^2)	Φ (J K^{-1} mol^{-1} Å^{-2})
Methane	319.2	559.1	152.4	2.10
Ethane	1047.5	286.9	191.5	5.47
Propane	415.8	667.7	223.4	1.86
Benzene	403.8	612.9	240.7	1.68
Toluene	513.5	599.0	273.9	1.87

[a] A is the accessible surface area [44] and Φ is defined in Eq. (5).

Table 2
Parameter b is fixed to 600 K and parameter a is obtained by a corresponding best fit, due to Eq. (4)[a]

Solute	a ($J\,K^{-1}\,mol^{-1}$)	Φ ($J\,K^{-1}\,mol^{-1}\,Å^{-2}$)
Methane	298.6	1.96
Ethane	389.5	2.03
Propane	455.7	2.04
Benzene	417.2	1.73
Toluene	512.8	1.87

[a]The value of b and the ones for a correspond to the dashed lines in Fig. 1.

there exists a "universal" coupling constant ε_H in water. The reason for this is not clear to us, but apolar solvation can for instance be sensible to surface geometry. This may possibly explain the irregular values of the parameters for ethane compared to the other four molecules examined. In this context one can look at, e.g., cyclohexane which has a cost in free energy 5.2 kJ mol^{-1} when put into water from vapor at 25°C, while hexane has a cost of 10.7 kJ mol^{-1} [45]. This amounts to a free energy per unit surface area accessible to water of 20 J mol^{-1} $Å^{-2}$ for cyclohexane versus 36 J mol^{-1} $Å^{-2}$ for hexane. This particular example indicates that the hydrophobic effect is sensitive to surface geometry. Thus, the conclusion that thermodynamical quantities upon solvation of apolar compounds are proportional to water accessible surface area [46–49], regardless of the surface geometry, is at best approximate.

Gill et al. [37], and Makhatadze and Privalov [23] consider a two-state model which they fit to experimental data of apolar solvation. They conclude that the relatively good correspondence between experiments and their theory seems to support the view that water molecules in the solvation shell behave almost independently. This latter feature is common to our viewpoint. Although a two-state model has a more correct behavior in the $T \to 0$ limit, as discussed in Section 2, we cannot see that a two-state model has a more direct connection to the molecular basis of the problem at the temperatures considered here. In this respect we find our model more satisfactory, which is also supported by a better fit to the experimental data including the slight curvature as seen in Figs. 1 and 2.

Nevertheless, Sharp and Madan [40] conclude that various qualitative effects of apolar (also polar) solution can be satisfactorily reproduced by a two-state model. On the other hand their Monte Carlo simulations on a water network model reveal a more complex behavior than provided by the two-state picture.

4. Conclusion

We have investigated a new model on the microscopic level for the excess heat capacity of aqueous solvation of the apolar substances methane, ethane, propane, benzene,

and toluene. In the model we apply the energy of a hydrogen bond in the "ice-shell" around the solvated apolar molecule, which models the excess energy of solvation compared to bulk water, by using a hydrogen bond distortion model introduced by Pople [19].

The heat capacity is calculated by means of equilibrium statistical mechanics and the model fits experimental data well for all the proteins considered. One notes from Table 1 that the energy parameter b for bending hydrogen bonds is nearly the same for all substances except for ethane. As discussed in Section 3 this may be related to surface geometry, but otherwise the reason for the latter deviation is not clear.

Acknowledgements

A.B. thanks the Research Council of Norway (Contract No. 129619/410) for financial support.

References

[1] P.L. Privalov, Adv. Protein Chem. 33 (1979) 167.
[2] G.N. Phillips Jr., B.M. Pettitt, Protein Sci. 4 (1995) 149.
[3] T. Lazaridis, M. Karplus, Proteins 35 (1999) 133.
[4] G.W. Robinson, C.H. Cho, Biophys. J. 77 (1999) 3311.
[5] B.-lu Chen, J.A. Schellman, Biochemistry USA 28 (1989) 685.
[6] P.L. Privalov, Crit. Rev. Biochem. Mol. 25 (1990) 281.
[7] G. Graziano, F. Catanzano, A. Riccio, G. Barone, J. Biochem. 122 (1997) 395.
[8] F. Franks, Adv. Protein Chem. 46 (1995) 105.
[9] P. Bruscolini, L. Casetti, Phys. Rev. E 61 (2000) R2208.
[10] P. De Los Rios, G. Caldarelli, Phys. Rev. E 62 (2000) 8449.
[11] A. Hansen, M.H. Jensen, K. Sneppen, G. Zocchi, Physica A 288 (2000) 21.
[12] A. Bakk, J.S. Høye, A. Hansen, K. Sneppen, M.H. Jensen, Biophys. J. 79 (2000) 2722.
[13] A. Bakk, A. Hansen, K. Sneppen, Physica A 291 (2001) 60.
[14] T. Simonson, A. Brünger, J. Phys. Chem. 98 (1994) 4683.
[15] J.T. Edsall, J. Am. Chem. Soc. 57 (1935) 1506.
[16] P.L. Privalov, S.J. Gill, Adv. Prot. Chem. 39 (1988) 191.
[17] P.L. Privalov, S.J. Gill, Pure Appl. Chem. 61 (1989) 1097.
[18] H.S. Frank, M.W. Evans, J. Chem. Phys. 13 (1945) 507.
[19] J.A. Pople, Proc. Roy. Soc. Ser. A 205 (1951) 163.
[20] D. Eisenberg, W. Kauzmann, The Structure and Properties of Water, Oxford University Press, London, 1969, p. 167.
[21] H. Naghibi, S.F. Dec, S.J. Gill, J. Phys. Chem. 90 (1986) 4621.
[22] H. Naghibi, S.F. Dec, S.J. Gill, J. Phys. Chem. 91 (1987) 245.
[23] G.I. Makhatadze, P.L. Privalov, J. Chem. Thermodyn. 20 (1988) 405.
[24] B. Lee, Biopolymers 24 (1985) 813.
[25] B. Lee, Biopolymers 31 (1991) 993.
[26] J.L. Finney, A.K. Soper, J.Z. Turner, Pure Appl. Chem. 65 (1993) 2521.
[27] J.L. Finney, A.K. Soper, Chem. Soc. Rev. 23 (1994) 1.
[28] A.K. Soper, J.L. Finney, Phys. Rev. Lett. 71 (1993) 4346.
[29] J.Z. Turner, A.K. Soper, J. Chem. Phys. 101 (1994) 6116.
[30] A. Filipponi, D.T. Bowron, C. Lobban, J.L. Finney, Phys. Rev. Lett. 79 (1997) 1293.
[31] D.T. Bowron, A. Filipponi, M.A. Roberts, J.L. Finney, Phys. Rev. Lett. 81 (1998) 4164.

[32] D.T. Bowron, A. Filipponi, C. Lobban, J.L. Finney, Chem. Phys. Lett. 293 (1998) 33.
[33] G. Olofsson, A.A. Oshodj, E. Qvarnström, I. Wadsö, J. Chem. Thermodyn. 16 (1984) 1041.
[34] B. Madan, K. Sharp, J. Phys. Chem. Ser. B 101 (1997) 11 237.
[35] E. Wilhelm, R. Battino, R.J. Wilcock, Chem. Rev. 77 (1977) 219.
[36] S.F. Dec, S.J. Gill, J. Solution Chem. 13 (1984) 27.
[37] S.J. Gill, S.F. Dec, G. Olofsson, I. Wadsö, J. Phys. Chem. 89 (1985) 3758.
[38] G. Némethy, H.A. Scheraga, J. Chem. Phys. 36 (1962) 3382.
[39] W. Heisenberg, Z. Phys. 49 (1928) 619.
[40] K.A. Sharp, B. Madan, J. Phys. Chem. Ser. B 101 (1997) 4343.
[41] G. Graziano, Phys. Chem. Chem. Phys. 1 (1999) 3567.
[42] L. Pauling, The Nature of the Chemical Bond: and the Structure of Molecules and Crystals, 2nd Edition, Cornell University Press, Ithaca, New York, 1940, p. 304.
[43] A. Ben-Naim, Solvation Thermodynamics, Plenum Press, New York, 1987.
[44] R.B. Hermann, J. Phys. Chem. 76 (1972) 2754.
[45] A. Ben-Naim, Y. Marcus, J. Chem. Phys. 81 (1984) 2016.
[46] S.J. Gill, N.F. Nichols, I. Wadsö, J. Chem. Thermodyn. 8 (1976) 445.
[47] R.S. Spolar, J.-H. Ha, M.T. Record Jr., Proc. Natl. Acad. Sci. USA 86 (1989) 8382.
[48] J.R. Livingstone, R.S. Spolar, M.T. Record Jr., Biochemistry USA 30 (1991) 4237.
[49] G. Graziano, G. Barone, J. Am. Chem. Soc. 118 (1996) 1831.
[50] A. Bakk, J.S. Høye, A. Hansen, K. Sneppen, J. Theor. Biol. 210 (2001) 367.

Paper 6

Two-state protein model with water interactions: Influence of temperature on the intrinsic viscosity of myoglobin

Audun Bakk*
Department of Physics, Norwegian University of Science and Technology, NTNU, NO-7491 Trondheim, Norway
(Received 31 October 2000; revised manuscript received 10 January 2001; published 24 May 2001)

We describe a single-domain protein as a two-state system with water interactions. Around the unfolded apolar parts of the protein we incorporate the hydration effect by introducing hydrogen bonds between the water molecules in order to mimic the ''icelike'' shell structure. Intrinsic viscosity, proportional to the effective hydrodynamic volume, for sperm whale metmyoglobin is assigned from experimental data in the folded and in the denaturated state. By weighing statistically the two states against the degree of folding, we express the total intrinsic viscosity. The temperature dependence of the intrinsic viscosity, for different chemical potentials, is in good correspondence with experimental data [P. L. Privalov *et al.*, J. Mol. Biol. **190**, 487 (1986)]. Cold and warm unfolding, common to small globular proteins, is also a result of the model.

DOI: 10.1103/PhysRevE.63.061906 PACS number(s): 87.14.Ee, 87.15.Cc, 66.20.+d, 05.70.Ce

I. INTRODUCTION

Proteins are macromolecules consisting of thousands of atoms. Despite their complexity, Privalov and Khechinashvili [1] showed by a van't Hoff analysis that several small globular proteins (<200 residues) are nearly a two-state system, i.e., either the protein is thermodynamically stable in the folded (native) state, or it is stable in the unfolded (denaturated) conformation.

Proteins are in a compact native state around physiological temperatures and natural chemical environments. An increase of the temperature denaturates the protein, which is quite intuitive from a physical point of view, e.g., thinking about thermal expansion of materials. But, what is rather surprising is that some proteins lose their stability at subphysiological temperatures [2–4]. This is called cold denaturation.

In this work we apply a simple two-state description for a protein, which is a reformulated version of a model proposed by Hansen *et al.* [5] and Bakk *et al.* [6,7]. In the denaturated state water is allowed to access the unfolded regions of the protein. The water molecules in this hydration shell are assigned a bending energy in order to mimic the ''frozen'' structure around an apolar surface [8]. By means of statistical mechanics we calculate an order parameter, which we apply in an expression for the *intrinsic viscosity* (IV). The IV is proportional to the effective hydrodynamic volume of a macromolecule [9], and is *not* equivalent to the *internal viscosity*, where the latter describes a resistance to extension or compression of a macromolecule [10]. Finally we compare the model with experimental data from Privalov *et al.* on sperm whale myoglobin [3].

II. PROTEIN MODEL

To first approximation, a small single-domain globular protein may be regarded as a two-state macroscopic system [1,11]. However, as shown in Ref. [6] the folding of such a protein can be regarded as a multiple process, i.e., a hierarchical folding of M contacts [12], and still be a two-state system from a thermodynamical point of view. Analogous to Zwanzig [13] we assign binary variables $\Psi_j \in \{0,1\}$ corresponding to an open (unfolded) and closed (folded) contact j, respectively. The hierarchical folding implies the constraint

$$\Psi_j \geqslant \Psi_k, \quad k \geqslant j \tag{1}$$

simply because contact $j \leqslant k$ cannot unfold while k is folded. This can further be parametrized by a second set of binary variables $\xi_j \in \{1,-B\}$. The ξ_j variables may be interpreted as a simplified representation of the dihedral angles [9] with only two ''angles'' accessible at each contact. Let ϵ be the energy gain to fold one contact [14], and let the binary variable $\chi_j \in \{1,-C\}$ distinguish between the fully folded state ($\chi_M = 1$) and the intermediate states ($\chi_{j<M} = -C$), respectively. The enthalpy for contact thus becomes

$$E^c = -i\epsilon \Psi_i \xi_i \chi_i, \tag{2}$$

when the unfolded enthalpy is set to zero.

In the two-state limit the intermediate states are unstable, i.e., $C \to \infty$. For simplicity we assume $B \to \infty$, thus the chain-chain enthalpy in Eq. (2) effectively becomes

$$E_i^c = -i\epsilon_c, \quad i \in \{0,1\}, \tag{3}$$

which corresponds to the native state ($i=1$) and the denaturated state ($i=0$) for the complete protein. The protein contact energy ϵ_c is simply the sum of M contact energies ϵ, i.e., $\epsilon_c = M\epsilon$.

For simplicity we assume that the denaturated state has g_c chain-related degrees of freedom compared to the thermodynamically unique native state of zero entropy. The present two-state model fulfills the van't Hoff enthalpy relation as shown in Ref. [6], which is also experimentally established for several globular proteins, myoglobin included [1]. As for ϵ_c in Eq. (3) g_c varies little with respect to the temperature [15], thus we assume g_c is independent of the temperature.

Solvation of a molecule in water, in analogy to protein unfolding, is a complex affair. It includes a cavity formation

*Electronic address: Audun.Bakk@phys.ntnu.no

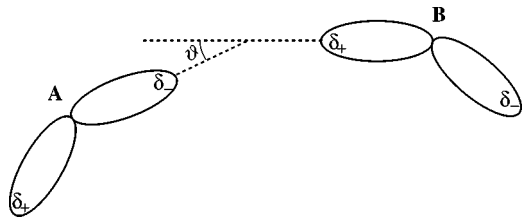

FIG. 1. Schematic illustration of two water molecules (A and B) between which a hydrogen bond is bent an angle ϑ. Only two of the four orbitals, where the polarity is indicated by δ_+ and δ_-, are shown for each molecule. The hydrogen bonds are meant to mimic the ''icelike'' structure of water around the unfolded apolar regions of the protein.

in water, interactions between water molecules and the surface of the solute molecule, and finally a rearrangement of the water around the solute molecule [16]. In this work we will only consider the latter effect.

Proteins consist of apolar[1] as well as polar surfaces [17]. As a simplification, we will in this work only consider the hydration effect upon unfolding around the exposed apolar parts of the protein. It is known from solvation of apolar substances in water that the hydration contribution to the entropy is negative, moreover, it decreases in absolute value for increasing temperature [18]. Frank and Evans [8] attributed this to a gradual melting of an ''ice shell'' around the apolar molecules. In analogy to this, we regard the water in the solvation shell around the unfolded apolar parts of the protein as hydrogen bonded (HB), while upon folding this water is expelled to the bulk, and is there regarded as a ''non-hydrogen-bonded liquid'' [19].

Inspired by Pople [20,21] we define an effective bending energy of one individual hydrogen bond in the solvation shell

$$E_i^{\mathrm{HB}}(\vartheta) = -(1-i)\epsilon_{\mathrm{HB}}\cos\vartheta, \quad \vartheta \in [0,\pi]. \tag{4}$$

The polar angle ϑ is the bending or distortion of a hydrogen bond as illustrated in Fig. 1. One sees from Eq. (4) that it is enthalpically favorable to let water access the unfolded apolar protein surfaces (i.e., $i=0$), otherwise, if water is expelled to the bulk we put this enthalpy to zero. ϵ_{HB} is a bending distortion constant and is supposed to be of the size of breaking one mole of hydrogen bonds and transferring them to bulk water. Némethy and Scheraga [19] estimated 5.5 kJ/mol for this constant, which we will apply in this work. Water molecules in the bulk will also have internal interactions, but with a weaker coupling compared to ice. Thus, the value from Némethy and Scheraga is substantially lower than the widely quoted value 18.8 kJ/mol for breaking one mole of hydrogen bonds and transferring them to vacuum as proposed by Pauling [22]. Each individual water molecule expelled to the bulk is assigned a degeneracy g_{w} in order to take into account the entropy loss of solvated water [18]. Let N be the effective number of hydrogen bonds in the solvation shell around the apolar surfaces of the unfolded protein. Thus, the total degeneracy of the protein is $g_1 = g_w^N$ in the folded state due to the water degrees of freedom, while the degeneracy of the unfolded state is $g_0 = g_c$ due to the chain flexibility. This yields a degeneracy corresponding to state i

$$g_i = g_c^{1-i} g_w^{Ni}. \tag{5}$$

The Hamiltonian for the protein is simply the sum of chain-chain enthalpies [see Eq. (3)] and protein-water interactions [see Eq. (4)]

$$\mathcal{H}_i(\vartheta) = E_i^c + E_i^{\mathrm{HB}}(\vartheta), \tag{6}$$

whereupon the canonical partition function becomes

$$\begin{aligned} Z &= \sum_{i=0}^{1} g_i e^{-E_i^c/(RT)} \left(\int_0^{\pi} d\vartheta \sin\vartheta e^{-E_i^{\mathrm{HB}}(\vartheta)/(RT)} \right)^N \\ &= g_c \left[\frac{2RT}{\epsilon_{\mathrm{HB}}} \sinh\{\epsilon_{\mathrm{HB}}/(RT)\} \right]^N + g_w^N e^{\epsilon_c/(RT)} 2^N \\ &= 2^N g_w^N e^{\epsilon_c/(RT)} (r+1) \equiv \sum_{i=0}^{1} Z_i. \end{aligned} \tag{7}$$

$R = 8.31$ J/(K mol) is the molar gas constant, T is the absolute temperature, and the function r is defined as

$$r \equiv [aTe^{-\mu/T}\sinh(b/T)]^N, \tag{8}$$

where $a \equiv Rg_c^{1/N}/(\epsilon_{\mathrm{HB}} g_w)$, $\mu \equiv \epsilon_c/(NR)$, and $b \equiv \epsilon_{\mathrm{HB}}/R$. The power of N in Eq. (7) is due to the N hydrogen bonds that are supposed to act individually. This is a coarse simplification because ice is supposed to have long-range order [21].

In Sec. III below, which concerns the intrinsic viscosity, we will need a quantity or a measure of the degree of folding. Thus, we define an order parameter[2] n for the system. According to the previous notation where $i=0$ and $i=1$ corresponds to a denaturated and a native protein, respectively, we weigh the two states by the corresponding Boltzmann weights Z_i defined in Eq. (7). The order parameter becomes

$$n \equiv \frac{\sum_{i=0}^{1} iZ_i}{\sum_{i=0}^{1} Z_i} = \frac{1}{r+1}. \tag{9}$$

Up to this point the model is general, only restricted to single-domain (small) proteins exhibiting two states. For positive values of a and b, which is valid throughout this work, $r>0$ is a consequence for all temperatures. Moreover,

[1] *Apolar* means that the molecule exhibits no permanent dipole moment, as opposed to *polar*.

[2] *Order parameter* in physics is equivalent to *reaction coordinate*, commonly used in chemistry and protein literature.

the order parameter has the following limits: $\lim_{r\to 0} n=1$ and $\lim_{r\to\infty} n=0$. Thus, we have constructed an order parameter confined to the interval $n\in[0,1]$, which statistically describes the degree of folding.

III. INTRINSIC VISCOSITY

We now have the ''machinery'' to describe the IV. First we will calculate IV separately in the native and denaturated state. By quantifying the population of the two states by the order parameter [see Eq. (9)], we are finally able to express the total IV for myoglobin as a sum of native and denaturational IV weighed against the degree of folding.

IV is in general defined as [24]

$$[\eta]\equiv\lim_{c\to 0}\frac{\eta'-\eta}{c\,\eta}, \tag{10}$$

which is the limit of zero concentration c of the *reduced viscosity* [25]. η' is the macroscopic viscosity (water + protein), η is the viscosity of pure solvent (water), and c is the protein concentration. For a fixed conformation IV is independent of the solution. However, the conformation will strongly depend on the solution, e.g., pH. Thus, in this respect IV will implicitly depend upon the solution, as discussed further in Sec. IV.

It can be shown that IV for a compact macromolecule of arbitrary shape can be written by heuristic means as [24]

$$[\eta]_1=\nu(\overline{V}_p+\delta\overline{V}_w), \tag{11}$$

where ν is the Simha factor containing all the shape dependence, $\overline{V}_p$ and $\overline{V}_w=1.0$ cm^3/g are the partial specific volumes of protein and pure water, respectively, and δ is the hydration ratio. From Eq. (11) one sees that IV can be regarded as an effective measure of the size of a macromolecule. In this work we study sperm whale metmyoglobin that has the following data: $\overline{V}_p=0.75$ cm^3/g [9], $\nu=2.8$ and $\delta=0.35$ [24]. Thus, according to Eq. (11), IV for myoglobin in the *native* state (1) is

$$[\eta]_1=3.1\ \text{cm}^3/\text{g}, \tag{12}$$

i.e., independent of temperature. The data, leading to Eq. (12), is measured at 20 °C. However, to first approximation we assume that Eq. (12) is valid at all temperatures.

IV in the denaturated state is a bit more complicated, where we *may* regard the protein as a random coil [26,27]. Flory [28] proposed

$$[\eta]_{\text{Flory}}=\frac{\Phi\langle r^2\rangle^{3/2}}{M} \tag{13}$$

for the intrinsic viscosity of a non-free-draining coil. $\Phi=3.62\times 10^{21}$ is a universal constant, $\langle r^2\rangle$ is the mean-square end-to-end distance, and M is the molecular weight. According to the data from Privalov *et al.* [3] there seems to be a pronounced temperature dependence of IV in the denaturated state (see upper curve in Fig. 2). Moreover, it is known that

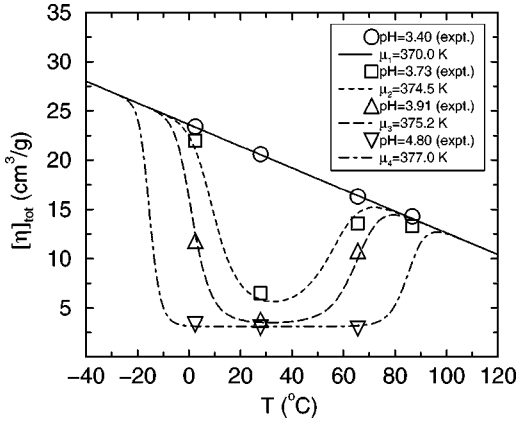

FIG. 2. Temperature dependence of myoglobin intrinsic viscosity at different chemical potentials μ. The curves are based upon Eq. (15) where $a=2.59\times 10^{-3}$ K^{-1} and $b=662$ K [19]. μ_1 corresponds to a denaturated protein, while μ_4 corresponds to the native state in the horizontal region between -10 °C and 80 °C. Experimental (expt.) data at various pH′s from Privalov *et al.* [3].

facilitation of rotational degrees of freedom in the backbone will cause a decrease of the dimension [27,29], hence, $\langle r^2\rangle$ decreases with increase in temperature, as well. Thus, according to Eq. (13), the IV also decreases. In order to incorporate the latter effect, we do a linear regression of data on unfolded myoglobin.[3] This implies the following temperature[4] dependent expression on IV of *denaturated* myoglobin

$$[\eta]_0=23.6-0.11T\ \text{cm}^3/\text{g}. \tag{14}$$

At 25 °C $[\eta]_0=20.9$ cm^3/g, noteworthily close to the value 20.1 cm^3/g from Tanford [26] obtained in 6 M guanidinium HCl.

From the calculated IV of native and denaturated myoglobin in Eqs. (12) and (14), respectively, we weigh the two states by the order parameter defined in Eq. (9). We put native IV proportional to the degree of *folding*, n, and denaturational IV proportional to the degree of *unfolding*, $(1-n)$, whereupon the total IV becomes

$$[\eta]_{\text{tot}}=[\eta]_1 n+[\eta]_0(1-n). \tag{15}$$

The two states, folded and unfolded protein, correspond to the limits $\lim_{n\to 1}[\eta]_{\text{tot}}=[\eta]_1$ and $\lim_{n\to 0}[\eta]_{\text{tot}}=[\eta]_0$, respectively. The order parameter n depends both on the temperature and on the chemical environments, as discussed in the section below.

IV. CALCULATIONS AND DISCUSSION

It seems to be reasonable to only incorporate hydration effect of the first solvation shell [30,23], according to Cohn and Edsall [31] who state that roughly one monolayer of

[3]Experimental data from Privalov *et al.* [3].

[4]Temperature here and in Figs. 2 and 3 in units of °C.

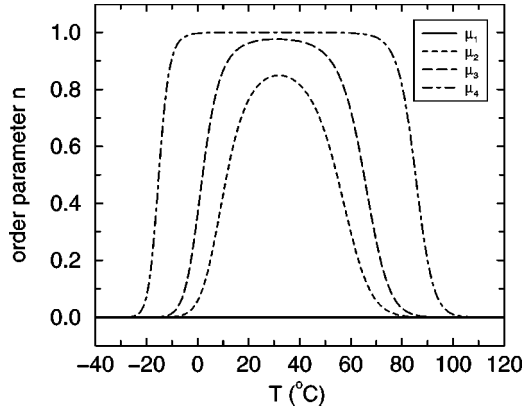

FIG. 3. Temperature dependence of the order parameter n defined in Eq. (9). All parameters correspond to Fig. 2. The order parameter measures the degree of folding. Thus, for μ_1 the protein is folded around physiological temperatures (30 °C), while for $\mu \lesssim \mu_4$ the protein is denaturated at all temperatures. Note that n is decreasing with decreasing μ.

water around apolar molecules is required to explain hydrodynamic data.

If we use an estimated value 8.4×10^3 Å^2 for the difference of the accessible surface area of the denaturated and the native apolar parts of myoglobin [15], together with an estimated value 9 Å^2 for the effective surface area of one water molecule [23], it is roughly 930 water molecules around the unfolded apolar regions of a myoglobin molecule in the first solvation shell. Let it be effectively one hydrogen bond per water molecule that forms or makes the ''freezing action'' in the hydration shell, thus $N=930$ in Eq. (8). Note that the latter number is the *excess* number of hydrogen bonds in the solvation shell compared to bulk water. Thus, N is not a very fundamental constant, merely a rough estimate. The estimated value of $\epsilon_{\mathrm{HB}}=5.5$ kJ/mol [19] implies $b=662$ K in Eq. (8).

Consequently, only two parameters remain ''adjustable'' in the protein model [see Eq. (7)] and thus in $[\eta]_{\mathrm{tot}}$ [see Eq. (15)], namely, a and μ. It is likely to believe that a change in the parameter μ is equivalent to a change in the chemical environments (pH, denaturant concentration, etc.), because μ is proportional to the chain-chain contact enthalpy ϵ_c, which reasonably depends upon, e.g., pH. On the other hand, the parameter a contains chain and water entropies in addition to the hydrogen-bond-bending constant ϵ_{HB}, which are presumably more stable parameters upon a change in the chemical environment compared to μ. Thus, we call the effective parameter μ the *chemical potential*.

In Fig. 2 we plot the intrinsic viscosity vs temperature for different μ and compare them to experimental data from Privalov *et al.* [3]. The curve corresponding to μ_1 exhibits the characteristic temperature dependence of an unfolded protein. In Fig. 3, where the corresponding order parameter vs temperature is plotted, one sees that μ_1 corresponds to $n=0$ for all T, i.e., it is only $[\eta]_0$ that contributes to $[\eta]_{\mathrm{tot}}$. This is nothing but the temperature dependent intrinsic viscosity of a free-draining coil expressed in Eq. (14). We note that the assumption of linear dependence on denaturational IV in Eq. (14) is a good approximation to experimental data.

The curve corresponding to μ_4 in Fig. 2 is horizontal in a broad temperature range from approximately 0 °C to 60 °C. This corresponds to $n=1$ as seen in Fig. 3. Thus, $[\eta]_{\mathrm{tot}} \rightarrow [\eta]_1$ implying a native protein in this temperature region, and is in fairly good correspondence to experimental data. However, in the experiments there seems to be a slight decrease of IV in the region discussed, probably due to a melting of the native structure analogous to the denaturational IV.

The curves μ_2 and μ_3 in Fig. 3 have both maxima $n<1$, thus only a *fraction* of the proteins are native. Here we note that an intermediate value of the order parameter, let us say $n=0.8$, does *not* mean that the protein is partly folded, but means *statistically* that 80% of an ensemble of proteins are folded, while 20% are unfolded. Actually, the curves corresponding to μ_2 and μ_3 is a crucial test of the validity of the model, because the corresponding experimental data clearly deviates from a straight line, as a consequence of the mixture of native and denaturated proteins that contribute to different intrinsic viscosities. In sum, our model seems to resemble the experimental data quite well.

The curve corresponding to μ_4 in Fig. 2 exhibits the characteristic temperature dependence of cold and warm destabilization. This is better seen in Fig. 3, where μ_4 corresponds to a native protein in an intermediate region around physiological temperatures (-10 °–80 °C), while it is denaturated outside this temperature region. Cold and warm unfolding is a common feature to small globular proteins [4,3,32]. The specific values of the chemical potential are all around $\mu=375$ K, which corresponds to $\epsilon_c=2900$ kJ/mol [see Eq. (8)]. It is interesting to compare this to the estimated values from Makhatadze and Privalov [15] on enthalpies of internal interactions $\Delta_N^U H^{\mathrm{int}}=7600$ kJ/mol, where van der Waals's (vdW) interactions contribute $\Delta_N^U H^{\mathrm{vdW}}=1200$ kJ/mol and hydrogen bonding contributes $\Delta_N^U H^{\mathrm{HB}}=6400$ kJ/mol. The latter three values are nearly constant between 5 °C and 100 °C. It is reasonable that $\Delta_N^U H^{\mathrm{HB}} > \epsilon_c > \Delta_N^U H^{\mathrm{vdW}}$, because in addition to disruption of the internal van der Waals's bonds the broken internal hydrogen bonds are likely to partly reappear as water-protein interactions. The latter enthalphy contribution is only partly because the specific water structure determines the possible hydrogen bond combinations towards the protein surface.

In a future expansion of the model it may be interesting to look at the apparent decreasing dimensionality with increasing temperature for both native and denaturational IV, which is more expressed for the latter. This may be attributed to a gradual melting of the structure due to some excitement of soft vibrational modes implying an effective smaller dimension [17]. If we were able to incorporate such interactions in the protein model, the parameter fit onto $[\eta]_0$ and $[\eta]_1$ may then turn out to be redundant—resulting in a more complete model.

To the author's knowledge regarding experiments on the temperature dependence of the IV, myoglobin is the only studied protein over such a broad temperature range and in different chemical environments as in Ref. [3]. Thus, we

hope that the present paper may stimulate experimental work on IV for other proteins, especially those that exhibit cold unfolding, in order to check the generality of the model.

V. SUMMARY

Single-domain proteins have thermodynamically two stable states, the native and the denaturated [1,11]. We apply a two-state description and incorporate the hydration effect upon unfolding by a model that mimics the ''icelike'' shell around the unfolded apolar surfaces as an increased number of hydrogen bonds compared to bulk water. By means of equilibrium statistical mechanics we calculate an order parameter (reaction coordinate) for the system, measuring the degree of folding.

In order to express the IV we do a linear fit onto experimental data on myoglobin of native and denaturational IV, respectively. The total IV for the native state is supposed to be linearly dependent on the order parameter and proportional to the degree of unfolding for the denaturated state.

The total IV exhibits good correspondence with experimental data from Privalov *et al.* [3]. For large chemical potentials the protein is native around physiological temperatures (30 °C), whereupon it becomes unstable at lower as well as higher temperatures. Cold and warm destabilizing action, common to small globular proteins, is a consequence of the model.

ACKNOWLEDGMENTS

I wish to thank Professor Alex Hansen for suggesting this work. I am also grateful to Professor Johan S. Høye for helpful discussions through the review process. The work was financially supported by the Research Council of Norway (Contract No. 129619/410).

[1] P. L. Privalov and N. N. Khechinashvili, J. Mol. Biol. **86**, 665 (1974); P. L. Privalov, Adv. Protein Chem. **33**, 167 (1979).
[2] F. G. Hopkins, Nature (London) **126**, 383 (1930); J. H. Clark, J. Gen. Physiol. **28**, 539 (1945); C. F. Jacobsen and L. K. Christensen, Nature (London) **161**, 30 (1948); R. B. Simpson and W. Kauzmann, J. Am. Chem. Soc. **75**, 5139 (1953); J. F. Brandts, *ibid.* **86**, 4291 (1964).
[3] P. L. Privalov, Yu. V. Griko, S. Yu. Venyaminov, and V. P. Kutyshenko J. Mol. Biol. **190**, 487 (1986).
[4] B.-lu Chen and J. A. Schellman, Biochemistry **28**, 685 (1989).
[5] A. Hansen, M. H. Jensen, K. Sneppen, and G. Zocchi, Eur. Phys. J. B **6**, 157 (1998).
[6] A. Bakk, J. S. Høye, A. Hansen, K. Sneppen, and M. H. Jensen, Biophys. J. **79**, 2722 (2000).
[7] A. Bakk, A. Hansen, and K. Sneppen, Physica A **291**, 60 (2001).
[8] H. S. Frank and M. W. Evans, J. Chem. Phys. **13**, 507 (1945).
[9] T. E. Creighton, *Proteins: Structures and Molecular Properties*, 2nd ed. (Freeman, New York, 1993), Chaps. 5 and 7.
[10] R. B. Bird, C. F. Curtiss, R. C. Armstrong, and O. Hassager, *Dynamics of Polymeric Liquids*, 2nd ed. (Wiley, New York, 1987), Vol. 2, p. 105; L. E. Wedgewood, Rheol. Acta **32**, 405 (1993).
[11] P. L. Privalov, J. Mol. Biol. **258**, 707 (1996).
[12] A. Hansen, M. H. Jensen, K. Sneppen, and G. Zocchi, Physica A **250**, 355 (1998).
[13] R. Zwanzig, Proc. Natl. Acad. Sci. U.S.A. **92**, 9801 (1995).
[14] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **84**, 7524 (1987); **30**, 177 (1990).
[15] G. I. Makhatadze and P. L. Privalov, Adv. Protein Chem. **47**, 307 (1995).
[16] B. Lee, Biopolymers **24**, 813 (1985); **31**, 993 (1991).
[17] P. L. Privalov and G. I. Makhatadze, J. Mol. Biol. **224**, 715 (1992).
[18] P. L. Privalov and S. J. Gill, Adv. Protein Chem. **39**, 191 (1988).
[19] G. Némethy and H. A. Scheraga, J. Chem. Phys. **36**, 3382 (1962).
[20] J. A. Pople, Proc. R. Soc. London, Ser. A **205**, 163 (1951); A. Bakk and J. S. Høye (unpublished).
[21] D. Eisenberg and W. Kauzmann, *The Structure and Properties of Water* (Oxford University Press, London, 1969), p. 167.
[22] L. Pauling, *The Nature of the Chemical Bond* (Cornell University Press, Ithaca, 1940), p. 304.
[23] S. J. Gill, S. F. Dec, G. Olofsson, and I. Wadsö, J. Phys. Chem. **89**, 3758 (1985).
[24] C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry. II. Techniques for the Study of Biological Structure and Function* (Freeman, San Francisco, 1980), p. 648.
[25] W. Hoppe, W. Lohmann, H. Markl, and H. Ziegler, *Biophysics* (Springer-Verlag, Berlin, 1983), p. 46.
[26] C. Tanford, Adv. Protein Chem. **23**, 121 (1968).
[27] F. Ahmad and A. Salahuddin, Biochemistry **13**, 245 (1974).
[28] P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, 1953), Chap. 14.
[29] D. A. Brant and P. J. Flory, J. Am. Chem. Soc. **87**, 2791 (1965); W. G. Miller, D. A. Brant, and P. J. Flory, J. Mol. Biol. **23**, 67 (1967).
[30] G. I. Makhatadze and P. L. Privalov, J. Chem. Thermodyn. **20**, 405 (1988).
[31] E. J. Cohn and J. T. Edsall, *Proteins, Amino Acids and Peptides* (Reinhold, New York, 1943).
[32] P. L. Privalov, Crit. Rev. Biochem. Mol. Biol. **25**, 281 (1990).

Paper 7

ELSEVIER

Physica A 304 (2002) 355–361

PHYSICA A

www.elsevier.com/locate/physa

Specific heat upon aqueous unfolding of the protein interior: a theoretical approach

Audun Bakk *, Johan S. Høye, Alex Hansen

Department of Physics, Norwegian University of Science and Technology, NTNU, NO-7491 Trondheim, Norway

Received 19 September 2001

Abstract

We study theoretically the thermodynamics, over a broad temperature range (5–125°C), related to hydrated water upon protein unfolding. The hydration effect is modeled as interacting dipoles in an external field, mimicking the influence from the unfolded surfaces on the surrounding water compared to bulk water. The heat capacity change upon hydration is compared with experimental data from Privalov and Makhatadze on four different proteins: myoglobin, lysozyme, cytochrome c and ribonuclease. Despite the simplicity of the model, it yields good correspondence with experiments. With some interest we note that the effective coupling constants are the same for myoglobin, lysozyme, and cytochrome c, although they are slightly different for ribonuclease.

PACS: 05.70.Ce; 87.14.Ee; 87.10.+e

Keywords: Protein folding; Protein thermodynamics; Hydration

1. Introduction

Proteins consist of 20 different amino acids with a great diversity with regard to size, polarity and charge. The understanding of water interactions seems to be important in order to understand protein folding in general, and the special feature of cold unfolding of several small globular proteins in particular [1–9].

We in the present work will represent the energy difference between the unfolded and folded interior, with regard to the water, by mimicking additional hydrogen bonds

* Corresponding author.

E-mail addresses: audun.bakk@phys.ntnu.no (A. Bakk), johan.hoye@phys.ntnu.no (J.S. Høye), alex.hansen@phys.ntnu.no (A. Hansen).

0378-4371/02/$ - see front matter
PII: S0378-4371(01)00569-6

from which we calculate the hydration heat capacity change upon protein unfolding. A justification of the model is the ability for water molecules to form an "ice-like" shell ("iceberg" in the terminology of Frank and Evans [10]) around apolar surfaces and thus create more hydrogen bonds. Reduction of both enthalpy [11–14] and entropy [15,16] upon apolar hydration seems to be well established [17].

However, the protein interior that becomes hydrated upon unfolding also consists of surfaces that has polarity, which means that the surface has permanent dipoles and charges. The heat capacity change upon purely polar hydration becomes surprisingly negative [18,19]. For apolar surfaces experiments show that the hydration contribution to the heat capacity upon solvation is positive. Also for proteins where part of the surface is polar this heat capacity is positive. Thus, for simplicity we will in this work use the apolar "ice-like" shell picture to make an effective model for the hydration effect upon protein unfolding. In this way, we may neglect some crucial features of polar solvation.

Finally, we apply equilibrium statistical mechanics to the model and calculate the hydration heat capacity increment, which we compare with experimental data from Privalov and Makhatadze [19] on four different proteins.

2. Hydration upon protein unfolding

We will use a refined version of a model first proposed by Hansen et al. [7,20,21]. The model studied here was applied by Bakk et al. [22] on a complete protein folding model, but they did not study the hydration effect separately. In this work we will study specifically this hydration upon protein unfolding.

Protein unfolding involves a cavity formation in water with a rearrangement of the water molecules surrounding the unfolded protein [23,24]. When estimating the solvation energy of exposing the interior of a protein to water, one has to calculate the energy difference between hydrated water, associated with the protein, and bulk water [19]. More precisely, the hydration is defined as the transfer of a solute from a fixed position in the ideal gas phase to a fixed position in the solvent [25], i.e., water in the present case.

In the solvation shell around the unfolded surfaces of the protein there will be forces that tend to orient the water molecules relative to these surfaces. Frank and Evans [10] introduced the term "iceberg" to describe the apparent "freezing" of the water molecules in the solvation shell around apolar molecules. This effective description is substantiated by the experimental fact that both enthalpy [11–14] and entropy [15,16] decreases upon apolar hydration.

In order to model the effect upon unfolding of a protein we use a simple model which we expect contains crucial physical features of apolar solvation. Thus, we model the water molecules as classical electric dipoles (which are not directly related to the actual dipole moments of water). In an electric field $\boldsymbol{\varepsilon}$, a dipole moment $\mathbf{s}$ has an energy $E = -\boldsymbol{\varepsilon} \cdot \mathbf{s}$. Assuming $|\mathbf{s}| = 1$ for simplicity, we have

$$E = -\varepsilon \cos\vartheta \,, \tag{1}$$

where ϑ is the angle between $\boldsymbol{\varepsilon}$ and $\mathbf{s}$. The $\varepsilon=|\boldsymbol{\varepsilon}|$, which is a bending distortion constant, represents directional forces and the angle ϑ represents orientations or bendings relative to the preferred direction for each water molecule. Eq. (1) is the hydration model used in the works by Bakk and Høye [26] and by Bakk [27], and it extends the interpretation of the hydration model applied by Hansen et al. [7,20] and Bakk et al. [28,29].

The idea of representing the solvent by dipoles in protein folding was introduced by Warshel and Levitt [30], and later in applications by, Russell and Warshel [31], Fan et al. [32], and Avbelj [33].

Evaluating the specific heat based on the energy of Eq. (1), one easily finds that it decreases monotonically. Among the proteins studied in the present case there is a weak maximum according to experiments. Thus, in addition to the energy due to the external field [see Eq. (1)] we will add a coupling term to model pair interactions between the water molecules. First of all, pair interactions of some kind are always present, but an additional reason to include them in the present case is the possibility to induce a maximum in the specific heat. The physical reason behind this is that such interactions effectively adds to the field ε of Eq. (1). This added field increases when the temperature is lowered. This again effectively leads to a "compression" of the temperature scale in some intermediate region in which a maximum in the specific heat can be created, as we actually find below in our results.

Between each pair of water molecules i and j there is thus a pair interaction

$$E_{ij} = -J_{ij}\,\mathbf{s}_i \cdot \mathbf{s}_j\,, \tag{2}$$

where J_{ij} is the coupling constant and $\mathbf{s}_i$ is the "dipole moment" of water molecule i. In a *mean field* solution [34,35] the interaction in Eq. (2) acts like an additional electric field that adds to ε in Eq. (1) to obtain an effective field of magnitude ε_{e}. In this way, the combination of Eqs. (1) and (2) for each water molecule results in the effective energy

$$E_{\mathrm{e}}(\vartheta) = -\varepsilon_{\mathrm{e}}\cos\vartheta + \tfrac{1}{2}bm^2\,, \tag{3}$$

where $\varepsilon_{\mathrm{e}} = \varepsilon + bm$ with $b = \sum_j J_{ij}$, and $m = \langle\cos\vartheta\rangle$ is the average dipole moment. Note that the sum over the values of J_{ij} (which becomes a spatial integral) are here not further specified as only the parameter b counts. E.g., the water molecules may or may not be on a lattice. Also note the addition of the $\frac{1}{2}bm^2$ term in Eq. (3). This term compensates the double counting of pair interactions in Eq. (2) when the mean field is evaluated [22,36].

Now for N such dipoles per protein the partition function for the total hydration contribution upon protein unfolding becomes

$$Z = \left[\int_0^{2\pi} \mathrm{d}\varphi \int_0^{\pi} \mathrm{d}\vartheta \sin\vartheta \exp(-\beta E_{\mathrm{e}}(\vartheta))\right]^N = \left[Z_{\mathrm{e}} \exp\left(-\frac{1}{2}\,\beta bm^2\right)\right]^N\,, \tag{4}$$

where

$$Z_{\mathrm{e}} = \frac{4\pi \sinh(\beta\varepsilon_{\mathrm{e}})}{\beta\varepsilon_{\mathrm{e}}}\,, \tag{5}$$

with $\beta = (k_B T)^{-1}$. Note that the free energy F is related to the partition function by $\ln Z = -\beta F$. From this the average dipole moment $m = \langle \cos\vartheta \rangle$ is obtained as (β is constant)

$$m = \frac{1}{N}\frac{\partial \ln Z}{\partial(\beta\varepsilon)} = \frac{\partial \ln Z_e}{\partial(\beta\varepsilon_e)}\frac{\partial \varepsilon_e}{\partial \varepsilon} + \frac{\partial(-1/2\ bm^2)}{\partial m}\frac{\partial m}{\partial \varepsilon}\,, \tag{6}$$

which yields

$$m = \frac{\partial \ln Z_e}{\partial(\beta\varepsilon_e)} = \coth\left(\frac{\varepsilon_e}{RT}\right) - \frac{RT}{\varepsilon_e}\,, \tag{7}$$

when the gas constant R is introduced to replace Boltzmann's constant k_B such that here and below ε and b becomes energies per mole.

The internal energy is now obtained as $U = -\partial \ln Z/\partial\beta$ (ε is constant) such that the total hydration heat capacity change per mole of proteins is ($k_B \to R$)

$$\Delta C = \frac{\partial U}{\partial T} = \frac{1}{RT^2}\frac{\partial^2}{\partial \beta^2}\ln Z\,. \tag{8}$$

We note that using mean field to this problem is an approximation. However, mean field is relative accurate and widely employed (away from critical points) [34,35,37]. In our case, the resulting parameter b is also an adjustable parameter, thus the mean field in itself should not introduce inaccuracies of importance in the present case. The largest uncertainty is expected to be the accuracy of the simplified model itself, that is used to represent a much more complex system into which one wants to obtain increased insight.

3. Discussion

We want to compare the heat capacity change upon unfolding solvation of the protein interior with experiments. The proteins considered are myoglobin (Mb), lysozyme (Lys), cytochrome c (Cyt), and ribonuclease (Rns) which we compare with experimental data from Privalov and Makhatadze [19] on the hydration contribution to the heat capacity change upon protein unfolding.

The hydration heat capacity change is shown in Fig. 1, and the parameter fit to the experimental data agrees quite well with these data. The heat capacity has a maximum around 25°C for Mb, Lys, and Cyt, while this maximum is shifted to around 50°C for Rns. Also in Table 1 a similar relation is reflected. With some interest we note from Table 1 that both the "electric field" constant ε and the coupling constant b are essentially the same for all of the four proteins, but there is a small deviation for Rns.

This small deviation for Rns reflects itself in the ratios between accessible polar and total surface area (reported from Makhatadze and Privalov [38]), $\Delta A_p/\Delta A_t$, which are almost equivalent for Mb, Lys, and Cyt, while this ratio is significantly larger for Rns. Hence, in this respect the parameters ε and b may be regarded as effective ones for the combined effect of apolar and polar surfaces as discussed in Section 1. Furthermore, the classification of apolar and polar surfaces is not very accurate in itself.

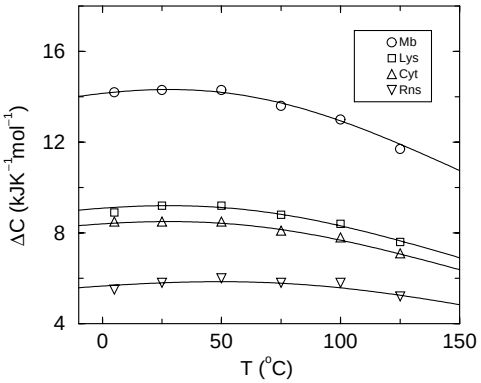

Fig. 1. The hydration heat capacity change upon unfolding of four different proteins. Theory is given by the continuous curves. Experimental data are from Privalov and Makhatadze [19]. Parameters, used to fit the experimental data are listed in Table 1.

Table 1
Parameters, according to Eqs. (1), (3), (4), and (8), used in Fig. 1 for the fitting to the experimental heat capacity hydration data from Privalov and Makhatadze [19][a]

Protein	ε (kJ mol^{-1})	b (kJ mol^{-1})	N	ΔA_t (Å^2)	$\Delta A_p/\Delta A_t$ (%)
Mb	2.05	8.2	1240	18250	36.5
Lys	2.05	8.2	800	14090	39.2
Cyt	2.05	8.2	740	11830	38.2
Rns	2.00	9.0	500	13300	44.8

[a]The difference in water accessible surface area between the unfolded and the folded protein ΔA_t is obtained from Makhatadze and Privalov [38]. $\Delta A_p/\Delta A_t$ is the ratio between the polar and total accessible surface area.

Since Rns differs a bit from the other three proteins considered, this may reflect its larger fraction of polar surfaces which then also can affect qualitative properties. Our model is more like an effective one for a mixed polar and apolar surface. Thus features specific for polar surfaces are not properly taken into account, but are more or less taken into account by adjusting available parameters. E.g., our present model may seem to give too small curvature on Fig. 1 for Rns. A reason for this may be the negligence of quantization, which will lower the specific heat and thus increase its curvature for decreasing temperatures. The polar (ionic) forces are relatively strong and the hydrogen atom is light, which both favor quantum effects. Earlier applications of a two-level system [28,39,40] for this kind of problem can thus reflect such quantization.

4. Conclusion

We have proposed a thermodynamical model for the hydration of the protein interior that becomes exposed to water upon unfolding. To our knowledge this is the first model

studied and compared to experimental protein data on the pure hydration heat capacity increment over such broad temperature range (5–125°C).

Hydration is modeled in an "ice-like" shell analogy, where the water molecules are represented by interacting dipoles in an external field. Compared with experimental data from Privalov and Makhatadze [19] for the four proteins myoglobin (Mb), lysozyme (Lys), cytochrome c (Cyt), and ribonuclease (Rns) the model fits quite well. The specific values of the field coupling constant ε and dipole coupling constant b [see Eq. (4)] are the same for Mb, Lys, and Cyt, while it is slightly different for Rns.

In a more refined model it can be necessary to distinguish between apolar and polar parts of protein surfaces. E.g., experimentally one finds that the heat capacity change is negative for hydration of purely polar surfaces [18,19], in contrast to the apolar surfaces where this heat capacity change is positive.

Acknowledgements

A.B. thanks the Research Council of Norway for financial support (Contract No. 129619/410).

References

[1] P.L. Privalov, Yu.V. Griko, S.Yu. Venyaminov, V.P. Kutyshenko, Cold denaturation of myoglobin, J. Mol. Biol. 190 (1986) 487–498.
[2] Yu.V. Griko, P.L. Privalov, S.Yu. Venyaminov, V.P. Kutyshenko, Thermodynamic study of the apomyoglobin structure, J. Mol. Biol. 202 (1988) 127–138.
[3] B.-lu Chen, J.A. Schellman, Low-temperature unfolding of a mutant of phage T4 lysozyme. 1. Equilibrium studies, Biochemistry USA 28 (1989) 685–691.
[4] P.L. Privalov, Cold denaturation of proteins, Biochem. Mol. Biol. 25 (1990) 281–305.
[5] F. Franks, Protein destabilization at low temperatures, Adv. Protein Chem. 46 (1995) 105–139.
[6] G. Graziano, F. Catanzano, A. Riccio, G. Barone, A reassessment of the molecular origin of cold denaturation, J. Biochem. 122 (1997) 395–401.
[7] A. Hansen, M.H. Jensen, K. Sneppen, G. Zocchi, Statistical mechanics of warm and cold unfolding in proteins, Eur. Phys. J. B 6 (1998) 157–161.
[8] P. De Los Rios, G. Caldarelli, Putting proteins back into water, Phys. Rev. E 62 (2000) 8449–8452.
[9] A. Bakk, J.S. Høye, A. Hansen, K. Sneppen, M.H. Jensen, Pathways in two-state protein folding, Biophys. J. 79 (2000) 2722–2727.
[10] H.S. Frank, M.W. Evans, Free volume and entropy in condensed systems. III. Entropy in binary liquid mixtures; partial molal entropy in dilute solutions; structure and thermodynamics in aqueous electrolytes, J. Chem. Phys. 13 (1945) 507–532.
[11] G. Olofsson, A.A. Oshodj, E. Qvarnström, I. Wadsö, Calorimetric measurements on slightly soluble gases in water. Enthalpies of solution of helium, neon, argon, krypton, xenon, methane, ethane, propane, *n*-butane, and oxygen at 288.15, 298.15, and 308.15 K, J. Chem. Thermodyn. 16 (1984) 1041–1052.
[12] H. Naghibi, S.F. Dec, S.J. Gill, Heat of solution of methane in water from 0 to 50°C, J. Phys. Chem. 90 (1986) 4621–4623.
[13] H. Naghibi, S.F. Dec, S.J. Gill, Heats of solution of ethane and propane in water from 0 to 50°C, J. Phys. Chem. 91 (1987) 245–248.
[14] B. Madan, K. Sharp, Molecular origin of hydration heat capacity changes of hydrophobic solutes: perturbation of water structure around alkanes, J. Phys. Chem. B 101 (1997) 11237–11242.
[15] E. Wilhelm, R. Battino, R.J. Wilcock, Low-pressure solubility of gases in water, Chem. Rev. 77 (1977) 219–262.

[16] S.F. Dec, S.J. Gill, Heats of solution of gaseous hydrocarbons in water at 25°C, J. Solution Chem. 13 (1984) 27–41.
[17] G.I. Makhatadze, P.L. Privalov, Energetics of protein structure, Adv. Protein Chem. 47 (1995) 307–425.
[18] P.L. Privalov, Physical basis of the stability of the folded conformations of proteins, in: T.E. Creighton (Ed.), Protein Folding, Freeman, New York, 1992 (Chapter 3).
[19] P.L. Privalov, G.I. Makhatadze, Contribution of hydration and non-covalent interactions to the heat capacity effect on protein unfolding, J. Mol. Biol. 224 (1992) 715–723.
[20] A. Hansen, M.H. Jensen, K. Sneppen, G. Zocchi, A model for the thermodynamics of globular proteins, Physica A 270 (1999) 278–287.
[21] A. Hansen, M.H. Jensen, K. Sneppen, G. Zocchi, Proteins top-down: a statistical mechanics approach, Physica A 288 (2000) 21–30.
[22] A. Bakk, J.S. Høye, A. Hansen, Heat capacity of protein folding, Biophys. J. 81 (2001) 710–714.
[23] B. Lee, The physical origin of the low solubility of nonpolar solutes in water, Biopolymers 24 (1985) 813–823.
[24] B. Lee, Solvent reorganization contribution to the transfer thermodynamics of small nonpolar molecules, Biopolymers 31 (1991) 993–1008.
[25] A. Ben-Naim, Y. Marcus, Solvation thermodynamics of nonionic solutes, J. Chem. Phys. 81 (1984) 2016–2027.
[26] A. Bakk, J.S. Høye, Microscopic argument for the anomalous hydration heat capacity increment upon solvation of apolar substances, Physica A 303 (2001) 286–294.
[27] A. Bakk, Two-state protein model with water interactions: influence of temperature on the intrinsic viscosity of myoglobin, Phys. Rev. E 63 (2001) 061906.
[28] A. Bakk, J.S. Høye, A. Hansen, K. Sneppen, Thermodynamical implications of a protein model with water interactions, J. Theor. Biol. 210 (2001) 367–373.
[29] A. Bakk, A. Hansen, K. Sneppen, Protein model exhibiting three folding transitions, Physica A 291 (2001) 60–70.
[30] A. Warshel, M. Levitt, Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme, J. Mol. Biol. 103 (1976) 227–249.
[31] S.T. Russell, A. Warshel, Calculations of electrostatic energies in proteins. The energetics of ionized groups in bovine pancreatic trypsin inhibitor, J. Mol. Biol. 185 (1985) 389–404.
[32] Z.Z. Fan, J.-K. Hwang, A. Warshel, Using simplified protein representation as a reference potential for all-atom calculations of folding free energy, Theor. Chem. Acc. 103 (1999) 77–80.
[33] F. Avbelj, Amino acid conformational preferences and solvation of polar backbone atoms in peptides and proteins, J. Mol. Biol. 300 (2000) 1335–1359.
[34] S.-K. Ma, Statistical Mechanics, World Scientific, Philadelphia, USA, 1985 (Chapter 27).
[35] C. Itzykson, J.-M. Drouffe, Statistical Field Theory, Vol. 1, Cambridge University Press, Cambridge, UK, 1989.
[36] J.S. Høye, G. Stell, Statistical mechanics of polar fluids in electric fields, J. Chem. Phys. 72 (1980) 1597–1613.
[37] H.E. Stanley, Introduction to Phase Transitions and Critical Phenomena, Clarendon Press, London, UK, 1971.
[38] G.I. Makhatadze, P.L. Privalov, Heat capacity of proteins. I. Partial molar heat capacity of individual amino acid residues in aqueous solution: hydration effect, J. Mol. Biol. 213 (1990) 375–384.
[39] B. Madan, K. Sharp, Heat capacity changes accompanying hydrophobic and ionic solvation: a Monte Carlo random network model study, J. Phys. Chem. 100 (1996) 7713–7721.
[40] G. Graziano, Hydration thermodynamics of aliphatic alcohols, Phys. Chem. Chem. Phys. 1 (1999) 3567–3576.

Paper 8

Mapping the non-directed polymer model to a non-linear growth equation of Burgers type

Audun Bakk[1] and Alex Hansen[2]
Department of Physics, Norwegian University of Science and Technology, NTNU, NO-7491 Trondheim, Norway

(February 25, 2002)

Abstract

We study the NDP model in the framework of a non-linear growth equation of Burgers type [Kardar-Parisi-Zhang equation with quenched noise (KPZQN equation)] by means of path integrals. The scaling exponents for the KPZQN equation are expressed in terms of the NDP model. In the strong-coupling regime, at low temperatures, the "tadpole" conformation seems to be reasonable for the polymer. The "tadpole" is discussed in the context of interfaces in a strong coupling regime where the noise dominates. We find that the "tadpole" behavior corresponds to structural "avalanches" of the interface, whereupon a totally new topology occurs. This restructuring is followed by periods of conservation of topology where the interface is "waiting" for new energetically more profitable structures.

PACS: 05.70.Ln, 68.35.Fx, 36.20.Ey, 05.40.Ca

Keywords: non-directed polymer, KPZ equation, quenched noise, tadpole

1 Introduction

The scaling properties of polymers have been studied for a long time. In 1987 Kardar and Zhang approached the problem through modeling polymers as directed [1], by assuming that polymers were stretched in a preferred direction without overhangs. In this work we will study a non-directed polymer model with quenched noise (NDP model) [2, 3]. We show that the model, using a path integral method, transforms to the Kardar-Parisi-Zhang equation [4] with quenched noise (KPZQN equation), whereupon we relate the scaling exponents in the two models. The "tadpole" is discussed in the light of diffusion of a population in a noisy environment. In the end we discuss the mapping between polymers and growing interfaces.

[1]E-mail: Audun.Bakk@phys.ntnu.no
[2]E-mail: Alex.Hansen@phys.ntnu.no

2 The NDP model

We study a non-directed polymer in a random medium [2,3]. The non-directness means that the polymer can choose every path in the space including self-crossings. The polymer is embedded in a d–dimensional random medium with a local uncorrelated interaction energy $V(\mathbf{x})$, which is meant to model the random fluctuations in the medium. Assuming Gaussian white noise, the first moment is

$$\langle V(\mathbf{x})\rangle = 0, \tag{1}$$

and the second moment of the noise is

$$\langle V(\mathbf{x})V(\mathbf{x}')\rangle \sim \delta^d(\mathbf{x}-\mathbf{x}'). \tag{2}$$

The polymer has a tension parameter γ, which models the energy cost through stretching by a term $\sim (d\mathbf{x}(t)/dt)^2$, where t is a coordinate along the polymer. We also note that this term may model entropic effects due to disorder [5, 6]. One may ask whether this model is realistic or not? One argument against the NDP model is that it does not include the energy related to *bending*, which must be related to a Laplacian term ($\sim \nabla_d^2\mathbf{x}$). Nevertheless, we ignore this bending energy.

We fix one end of the polymer at the origin and let the polymer of maximum length ℓ freely choose its conformation to the endpoint $\mathbf{x}(\ell)$, thus the Hamiltonian for a specific polymer conformation $\mathbf{x}(\ell)$ becomes

$$\mathcal{H}(\mathbf{x}(\ell)) = \int_0^\ell dt\, \left[\frac{\gamma}{2}\left(\frac{d\mathbf{x}(t)}{dt}\right)^2 + V(\mathbf{x}(t))\right], \tag{3}$$

We observe that without noise the polymer will find it energetically profitable to collapse at the origin, because stretching *per se* costs energy.

Discretizing Eq. (3) yields

$$\mathcal{H}(\mathbf{x}(\ell)) = \sum_{i=1}^{N-1}\left[\frac{\gamma}{2\epsilon}(\mathbf{x}_i-\mathbf{x}_{i-1})^2 + V(\mathbf{x}_i)\right], \tag{4}$$

where ϵ can be interpreted as a monomer size, or alternatively Kuhn's step length [7]. The maximum length ℓ of the polymer equals then

$$\ell = N\epsilon. \tag{5}$$

3 Mapping the NDP model to the KPZQN equation

In this section we transform the NDP model to the non-linear KPZQN equation and establish the relation between the scaling exponents in the two problems.

The KPZQN equation yields

$$\frac{\partial h(\mathbf{x},t)}{\partial t} = \nu \nabla_d^2 h(\mathbf{x},t) + \frac{\lambda}{2}\left[\nabla_d h(\mathbf{x},t)\right]^2 + \eta(\mathbf{x}), \tag{6}$$

where $h(\mathbf{x},t)$ is the height of the interface over the hyper-plane position $\mathbf{x}$ and t is the time. The position $\mathbf{x}$ is a d-dimensional space coordinate. Thus, $h(\mathbf{x},t)$ and $\mathbf{x}$ span a $(d+1)$-dimensional space.

The partition function $Z(\mathbf{x},\ell)$ for a polymer with one end fixed at the origin and the head located at $\mathbf{x}$, with the maximum length ℓ, becomes in the continuum case

$$Z(\mathbf{x},\ell) = \int \mathcal{D}^d\mathbf{x} \exp\left\{-\beta \int_0^\ell dt \ \left[\frac{\gamma}{2}\left(\frac{d\mathbf{x}(t)}{dt}\right)^2 + V(\mathbf{x}(t))\right]\right\}, \tag{7}$$

where $\beta = (k_{\mathrm{B}}T)^{-1}$. The k_{B} is Boltzmann's constant and T is the absolute temperature. The integral in Eq. (7) includes all possible paths between the origin and $\mathbf{x}$, where each path is weighted by a Boltzmann factor.

The energy account associated with each path is a competition between two terms. The energy penalty of stretching favors localized conformations, while the stochastic environment forces the polymers to wander in order to gain energy along the path. In Appendix A we map the NDP model to Eq. (25) by a Feynman path integral method [8]. The non-linear transformation

$$Z(\mathbf{x},\ell) = \exp\left\{\beta\widehat{\lambda}\gamma\,\widehat{h}(\mathbf{x},\ell)\right\}, \tag{8}$$

converts Eq. (25) into

$$\frac{\partial \widehat{h}(\mathbf{x},\ell)}{\partial \ell} = \widehat{\nu}\nabla_d^2\widehat{h}(\mathbf{x},\ell) + \frac{\widehat{\lambda}}{2}\left[\nabla_d\widehat{h}(\mathbf{x},\ell)\right]^2 + \widehat{\eta}(\mathbf{x}), \tag{9}$$

where we have defined

$$\widehat{\nu} = \frac{1}{2\beta\gamma}, \text{ and} \tag{10}$$

$$\widehat{\eta}(\mathbf{x}) = -\frac{1}{\widehat{\lambda}\gamma}V(\mathbf{x}). \tag{11}$$

Equation (9) is the KPZQN equation [Eq. (6)]. We can now map the physical parameters and quantities in the NDP model to the corresponding ones in the KPZQN equation, which are listed in Table 1. We observe that the requirement $|\nabla_d h(\mathbf{x},t)| \ll 1$ in the KPZQN problem is fulfilled in the NDP case by imposing $|\nabla_d\widehat{h}(\mathbf{x},t)| \ll 1$.

The scaling exponents are in general unknown for the KPZQN equation. Dynamic renormalization group treatments fail in all dimensions with respect to non-trivial fixed points, and thus fail in predicting the scaling exponents [7, 9].

Table 1: Mapping between the physical parameters and quantities in the NDP model and the KPZQN equation.

KPZQN	NDP
$h(\mathbf{x},\mathrm{t})$	$\widehat{h}$
t	ℓ
$\mathbf{x}$	$\mathbf{x}$
ν	$\widehat{\nu} = \frac{1}{2\beta\gamma}$
λ	$\widehat{\lambda}$
$\eta(\mathbf{x})$	$\widehat{\eta}(\mathbf{x}) = -\frac{1}{\widehat{\lambda}\gamma} V(\mathbf{x})$

However, Refs. [2, 3, 7] calculated the scaling exponents of the NDP problem in the low temperature limit, i.e., they ignored entropic effects. Using these results, one may predict the scaling exponents related to the KPZQN equation.

It is now important to interpret $\widehat{h}(\mathbf{x}, \ell)$ in the context of the NDP problem. The free energy for a canonical system with partition function Z is $\sim \ln Z$, thus we see from Eq. (8) that $\widehat{h}(\mathbf{x}, \ell)$ is proportional to the free energy, i.e., $\widehat{h}(\mathbf{x}, \ell) \sim -E_0(\mathbf{x}, \ell)$, where $E_0(\mathbf{x}, \ell)$ is the energy minimum, in the low temperature limit. For the KPZQN problem the surface width can be shown to obey the power law $\Delta h \sim t^b$ [10], where b is the growth exponent. In Ref. [3] they obtained $\Delta E_0 \sim \ell^\chi$, thus from Table 1 we see that the energy-fluctuation exponent χ for the NDP model equals the growth exponent in the KPZQN equation

$$\chi = b = 1 + \text{ log corrections.} \tag{12}$$

The finite size L of the KPZQN interface corresponds to $x_c(\ell)$ for the NDP model.[3] For the KPZQN problem one may define a crossover time t_x, which is the time when the surface crosses over from a time dependent surface width to a width determined by the system size. The crossover time (saturation time) yields the power law $t_x \sim L^z$ [10]. Considering Table 1 in connection with the scaling law $r \sim \ell^\zeta$ [3], thus

$$\zeta = \frac{1}{z} = 1 + \text{ log corrections.} \tag{13}$$

The Family-Vicsek scaling relation is $\alpha = bz$ [11], where the roughness exponent α is defined through the saturation width of a system of size L: $\Delta h_{\mathrm{sat}} \sim L^\alpha$. We then obtain from Eqs. 12 and 13

$$\alpha = \frac{\chi}{\zeta} = 1 + \text{ log corrections.} \tag{14}$$

Finally, we note that all these values for the scaling exponents are obtained by use of extreme statistics.

[3] Probably this is not obvious at the moment, but presumably it will become more clear after the next section about the "tadpole".

4 The "tadpole"

Let us imagine that the NDP model describes a growing polymer in the sense that the length is interpreted as a time, i.e., $\ell \to t$. Under these circumstances we ask:

How will the head $\mathbf{x}(t)$ *move during the growth of the polymer?*

In Appendix A we obtain Eq. (25) for the partition function of the NDP model. We observe that Eq. (25) has a close analogy to the following partial differential equation

$$\frac{\partial \Psi(\mathbf{x},t)}{\partial t} = \nabla_d^2 \Psi(\mathbf{x},t) + \lambda\, V(\mathbf{x})\, \Psi(\mathbf{x},t), \tag{15}$$

which models many interesting physical systems with diffusive character in random environments.

Zhang considered Eq. (15) as a prototype model for diffusion in a spatial potential [12, 13]. He investigated the evolution of unspecified species with an initially sharply localized distribution $\Psi(\mathbf{x},0) = \delta(\mathbf{x})$ where the species are able to move in space via diffusion, and found interestingly for localized distributions that they had the conjuncture of "hopping as a dynamical consequence of localization". Zhang found also that the localization center $x_c(t)$ behaved in all dimensions as

$$x_c(t) \sim \frac{t}{\sqrt{\ln t}}, \tag{16}$$

by using extreme statistics.

Let us now try to use this analysis in our own NDP model. The non-directed polymer starts growing at $\mathbf{x} = 0$, i.e., it is initially localized. The strong-coupling regime is considered where the random potential (noise) is dominating compared to the tension. Obviously, the polymer finds after some time that the origin is not the best attractor. Given enough time, which means increasing the total length ℓ of the polymer sufficient, some more distant attractors are more profitable. The result is that the head of the polymer (localization center) at a given time hops. This is analogous to "hopping as a dynamical consequence of localization" in Zhang's population conjuncture. Further it is important to understand that the hopping of the head of the polymer is not just an extension of the existing conformation of the polymer, but a totally new conformation compared to the previous conformation. The reason for this is that the whole path is minimized with respect to the free energy. This is schematically illustrated in Figure 1. Thus, Eq. (16) describes a sub-ballistic motion of the polymer head in a strong-coupling regime.

We now see a pattern of a polymer starting at $\mathbf{x} = 0$, but after a time it jumps to another location where it curls up in a local potential minimum. This is the best attainable potential well within a radius of the maximum length ℓ of

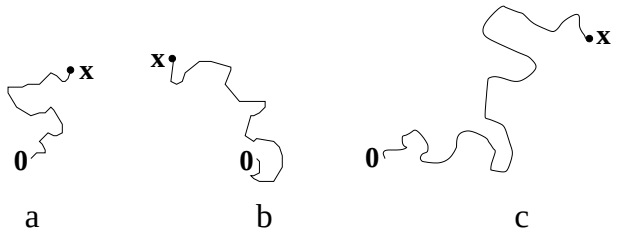

Figure 1: Schematic illustration of the non-directed polymer ("tadpole") having different maximum lengths $\ell_a < \ell_b < \ell_c$, which shows the "hopping" of the polymer. The polymer starts at 0 and ends at $\mathbf{x}$, where the head of the "tadpole" is indicated by the dot.

the polymer. The polymer will wait to grow for a longer body in order to reach new and better attractors, where the conformation of the polymer is given the descriptive name "tadpole".

Hansen *et al.* [3] considered the NDP model numerically. Using a transfer matrix they found that the polymer took the shape of a "tadpole" conformation, as we have discussed.

There seems to be no doubt about the "hopping-tadpole" behavior in the strong coupling regime. The question is for which value of the noise this behavior turns over to a diffusive one, i.e., when the localizing center moves as $x_c(t) \sim \sqrt{t}$? Another question is whether the path from the origin out to the head of the polymer $\mathcal{P}_a$ is fractal ? Let the length ℓ_a of this path $\mathcal{P}_a$ scale as

$$\ell_a \sim r^{\delta}. \tag{17}$$

If $\delta > 1$, then the path $\mathcal{P}_a$ is fractal. E.g., ref. [3] assumed $\delta = 1$.

Equation (25) describes a diffusive process. A good metaphor is that without a random potential $V(\mathbf{x}) = 0$ the polymer diffuses on a horizontal surface, but due to the random potential this surface becomes tilted and the average speed increases from random walk $x_c(t) \sim \sqrt{t}$ to sub-ballistic motion. This means that the Gaussian noise makes it more favorable to "speed up" in order to reach more attractive and deep potential wells. The path is in general fractal, with a fractal dimension that depends on the slope of the surface. This fact will obviously have impact on the scaling exponents found in Refs. [2, 3, 7].

5 What does the "tadpole" mean in the language of interfaces?

We have earlier shown that the NDP model transforms to the KPZQN equation. This means that it should be possible to interpret the "tadpole" conformation for the non-directed polymer in the framework of an interface growth model described by the KPZQN equation.

We have earlier interpreted the localization center $x_c(\ell)$ for the non-directed polymer to correspond with the finite size L of the interface. We also know that the noisy NDP exhibit the "tadpole" configuration in the strong coupling regime, which implies a hopping of the polymer depending on the maximum lengths. This hopping means that the *whole* polymer reorganizes its conformation to obtain an energetically favorable path.

Let us now think of an interface that has an *increasing* finite size $L(t)$. The "tadpole" for the NDP will then be analogous to the interface (described by the KPZQN) illustrated in Figure 2. An increase of the finite size from L_a to L_b will not change the structure of the interface from the original size which is illustrated by the continuous line in Figure 2. This is the same phenomenon as when the polymer curls up its head into a deep potential well for some time.

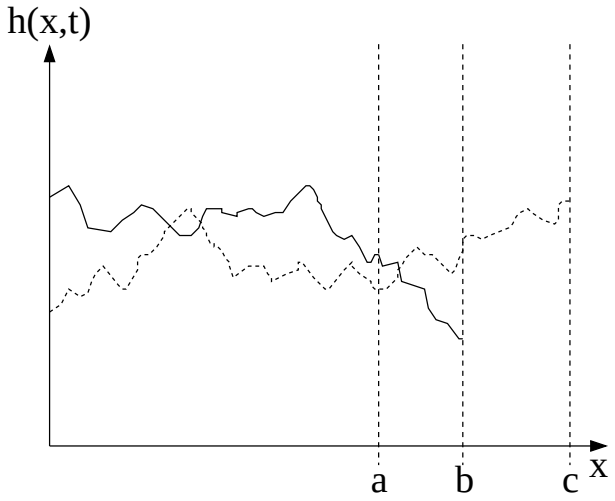

Figure 2: Schematic illustration of a (1+1)–dimensional interface in a strong coupling regime which shows the "hopping" of the structure. The interface is shown for different sizes $L_a < L_b < L_c$.

Now, imagine that we increase the finite size for the interface to L_c, which results in that the interface "converts" to a totally new structure. This corresponds to a "hop" from the continuous line to the dotted line in Figure 2. The structural transition of the interface is equivalent to the hopping of the "tadpole", where the NDP is long enough to take a new conformation and "dive" its head into a more distant potential well. In the context of the growing interface this means that for some time intervals it is waiting to reach a suitable finite size, whereupon an avalanche occurs and a totally restructured interface appears.

It is also reasonable that this structural hopping occurs even for a system of fixed size. We illustrate this in Figure 3 where the two diagrams illustrate an interface growth process within a short time interval in a weak coupling regime (a), and in a strong coupling regime (b). In the weak coupling regime the main

structure is conserved, while in the strong coupling regime we observe that the interface restructure itself totally during the growth. This is also analogous to the "tadpole" hopping, as for the increasing finite size discussed in the previous paragraph.

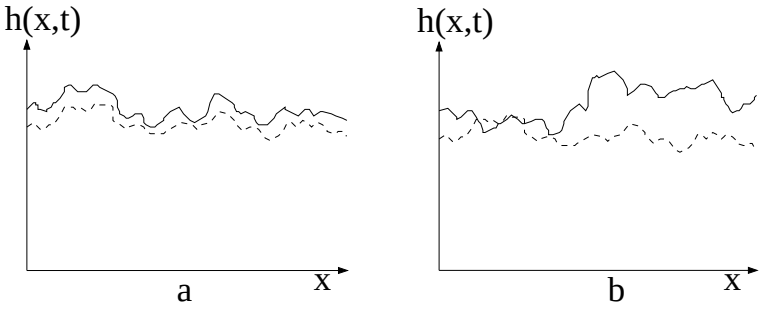

Figure 3: Schematic illustration of the interfacial structure during growth in a weak coupling regime (a) and in a strong coupling regime (b), considering a fixed system size.

6 Conclusion

We establish the mapping between the NDP model and the non-linear KPZQN equation of Burgers type [14], by means of the path integral method. We see the trend in the strong-coupling regime, where the noise dominates the tension, that the "tadpole" conformation is reasonable. We find that the growing "tadpole" corresponds to structural "avalanches" of the corresponding KPZQN interface, whereupon a topologically new interface appears. Between the interfacial "jumps" the topology is conserved, like a "sleeping tadpole".

Appendix A

The discrete version of Eq. (7) becomes

$$Z(\mathbf{x},\ell) = \left(\frac{\beta\gamma}{2\pi\epsilon}\right)^{\frac{(N-1)d}{2}} \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} dx_{1,1}\cdots dx_{N-1,1}\cdot\cdots\cdot dx_{1,d}\cdots dx_{N-1,d} \\ \times \exp\left\{-\frac{\beta\gamma}{2\epsilon}\sum_{i=1}^{N-1}\sum_{j=1}^{d}(x_{i,j}-x_{i-1,j})^2 - \beta\epsilon\sum_{i=1}^{N-1}V(\mathbf{x_i})\right\}, \tag{18}$$

where $x_{i,j}$ is the coordinate with indices $i \in \{0, N\}$ along the *path*, and $j \in \{1, d\}$ in the d–dimensional space. $\mathbf{x}_i = \sum_{j=1}^{d} \widehat{e}_j x_{i,j}$ is a space coordinate where $\widehat{e}_j$ is the unit vector in the j-direction.

Instead of directly calculate $Z(\mathbf{x},\ell)$ we first study the noiseless case $Z_0(\mathbf{x},\ell)$, i.e., $V(\mathbf{x_i})=0$ in Eq. (18). We start by calculating Z_0 in the j-direction. Let $x_{i,j}\to x_i$, the integral in Eq. (18) over x_1 yields

$$\begin{aligned} A_1 &= \frac{\beta\gamma}{2\pi\epsilon}\int_{-\infty}^{\infty} dx_1 \exp\left\{-\frac{\beta\gamma}{2\epsilon}\left[(x_2-x_1)^2+(x_1-x_0)^2\right]\right\} \\ &= \sqrt{\frac{\beta\gamma}{2\pi\cdot 2\epsilon}}\exp\left\{-\frac{\beta\gamma}{2\cdot 2\epsilon}(x_2-x_0)^2\right\}. \end{aligned} \tag{19}$$

Doing this recursively, we obtain after $N-1$ steps

$$A_{N-1} = \sqrt{\frac{\beta\gamma}{2\pi\ell}}\exp\left\{-\frac{\beta\gamma}{2\ell}(x_j(\ell))^2\right\}, \tag{20}$$

by substituting $\ell = N\epsilon$ [see Eq. (5)], and $x_j(\ell)=x_N-x_0$ is the $j-$coordinate of the head of the polymer. Thus, in d–dimensions

$$Z_0(\mathbf{x},\ell) = \left(\frac{\beta\gamma}{2\pi\ell}\right)^{\frac{d}{2}}\exp\left\{-\frac{\beta\gamma}{2\ell}\mathbf{x}^2\right\}, \tag{21}$$

where $\mathbf{x}^2=\sum_{j=1}^{d}(x_j(\ell))^2$. Eq. (21) describes the probability distribution of a non-directed polymer in a noiseless medium with mean $<\mathbf{x}>=0$, and standard deviation $\sigma=\sqrt{\frac{\ell}{\beta\gamma}}$. We note that Eq. (21) obeys the diffusion equation

$$\frac{\partial Z_0(\mathbf{x},\ell)}{\partial\ell} = \frac{1}{2\beta\gamma}\nabla_d^2 Z_0(\mathbf{x},\ell). \tag{22}$$

Let us now incorporate the noisy potential $V(\mathbf{x})$ in the partition function $Z(\mathbf{x},\ell)=Z(0,0;\mathbf{x},\ell)$, which describes a polymer starting at 0 and ending at $\mathbf{x}$ with maximum length ℓ. We are allowed to split the partition function in the sense

$$\begin{aligned} Z(0,0;\mathbf{x},\ell+\epsilon) &= \int_\Omega d^d\mathbf{x}_0\; Z(0,0;\mathbf{x}_0,\ell)Z(\mathbf{x}_0,\ell;\mathbf{x},\ell+\epsilon) \\ &\approx \int_\Omega d^d\mathbf{x}_0\; Z(0,0;\mathbf{x}_0,\ell)\, Z_0(\mathbf{x}_0,\ell;\mathbf{x},\ell+\epsilon)\exp\left\{-\beta\epsilon V(\mathbf{x_0})\right\}+\mathcal{O}(\epsilon^2), \end{aligned} \tag{23}$$

where $\mathbf{x}_0$ runs over a d–dimensional shell Ω of thickness ϵ, and $0<\epsilon\ll\ell$. We

then look at

$$\begin{aligned}
\frac{\partial Z(0,0;\mathbf{x},\ell+\epsilon)}{\partial \epsilon} &= \int_\Omega d^d\mathbf{x}_0\; Z(0,0;\mathbf{x}_0,\ell)\,\exp\{-\beta\epsilon V(\mathbf{x}_0)\} \\
&\quad\times\left[\frac{\partial Z_0(\mathbf{x}_0,\ell;\mathbf{x},\ell+\epsilon)}{\partial\epsilon} - Z_0(\mathbf{x}_0,\ell;\mathbf{x},\ell+\epsilon)\beta\, V(\mathbf{x_0})\right] \\
&\overset{(22)}{=} \int_\Omega d^d\mathbf{x}_0\; Z(0,0;\mathbf{x}_0,\ell)\,\exp\{-\beta\epsilon V(\mathbf{x_0})\} \\
&\quad\times\left[\frac{1}{2\beta\gamma}\nabla_d^2 Z_0(\mathbf{x}_0,\ell;\mathbf{x},\ell+\epsilon) - Z_0(\mathbf{x}_0,\ell;\mathbf{x},\ell+\epsilon)\beta\, V(\mathbf{x_0})\right] \\
&\approx \frac{1}{2\beta\gamma}\nabla_d^2\int_\Omega d^d\mathbf{x}_0\;\left[Z(0,0;\mathbf{x}_0,\ell)\,Z(\mathbf{x}_0,\ell;\mathbf{x},\ell+\epsilon)\right] \\
&\quad-\beta\, V(\mathbf{x})\int_\Omega d^d\mathbf{x}_0\, Z(0,0;\mathbf{x}_0,\ell)\,Z(\mathbf{x}_0,\ell;\mathbf{x},\ell+\epsilon) \\
&\overset{(23)}{=} \frac{1}{2\beta\gamma}\nabla_d^2 Z(0,0;\mathbf{x},\ell+\epsilon) - \beta\, V(\mathbf{x})\,Z(0,0;\mathbf{x},\ell+\epsilon).
\end{aligned} \tag{24}$$

By returning to the original notation
$Z(0,0;\mathbf{x},\ell) = Z(\mathbf{x},\ell)$, whereupon Eq. (24) becomes

$$\frac{\partial Z(\mathbf{x},\ell)}{\partial\ell} = \frac{1}{2\beta\gamma}\,\nabla_d^2 Z(\mathbf{x},\ell) - \beta\, V(\mathbf{x})\,Z(\mathbf{x},\ell). \tag{25}$$

Eq. (25) describes the probability distribution of a non-directed polymer in a random medium. Also note that Eq. (25) becomes a time-dependent Schrödinger equation by interpreting ℓ as imaginary time.

Acknowledgments

A. B. thanks the Research Council of Norway for financial support (Contract No. 129619/410).

References

[1] M. Kardar, Y.-C. Zhang, Scaling of directed polymers in random media, Phys. Rev. Lett. 58 (1987) 2087-2090.

[2] M.E. Cates, R.C. Ball, Statistics of a polymer in a random potential, with implications for a nonlinear interfacial growth model, J. Phys. France 49 (1988) 2009-2018.

[3] A. Hansen, E.L. Hinrichsen, S. Roux, Non-directed polymers in a random medium, J. Phys. I France 3 (1993) 1569-1584.

[4] M. Kardar, G. Parisi, Y.-C. Zhang, Dynamic scaling of growing interfaces, Phys. Rev. Lett. 56 (1986) 889-892.

[5] S.S. Jang, W.H. Jo, Analysis of the mechanical behavior of amorphous atactic poly(oxypropylene) by atomistic modeling, Macromol. Theor. Simul. 8 (1999) 1-9.

[6] C.R. Cantor, P.R. Schimmel, Biophysical Chemistry. I. The Conformation of Macromolecules, Freeman, San Francisco, 1980 (p. 6).

[7] T. Nattermann, W. Renz, Diffusion in a random catalytic environment, polymers in random media, and stochastically growing interfaces, Phys. Rev. A 40 (1989) 4675-4681.

[8] R.P. Feynman, A.R. Hibbs, Quantum Mechanics and Path Integrals, McGraw-Hill, New York, 1965 (pp. 42).

[9] A. Bakk. Dynamic renormalization group analysis of the Kardar-Parisi-Zhang equation with spatial noise and its connection to the non-directed polymer model. M.Sc. thesis, NTNU (1999).

[10] A.-L. Barabási, H.E. Stanley, Fractal Concepts in Surface Growth, Cambridge University Press, New York, 1995 (Chapter 2).

[11] F. Family, T. Vicsek, Scaling of the active zone in the Eden process on percolation networks and the ballistic deposition model, J. Phys. A 18 (1985) L75-L81.

[12] T. Halpin-Healy, Y.-C. Zhang, Kinetic roughening phenomena, stochastic growth, directed polymers and all that. Aspects of multidisciplinary statistical-mechanics, Phys. Rep. 254 (1995) 215-414.

[13] Y.-C. Zhang, Diffusion in a random potential: hopping as a dynamical consequence of localization, Phys. Rev. Lett. 56 (1986) 2113-2116.

[14] J.M. Burgers, The nonlinear diffusion equation: asymptotic solutions and statistical problems, Riedel, Boston, 1974.

Paper 9

Viscosity and transient electric birefringence study of clay colloidal aggregation

Audun Bakk* and Jon O. Fossum
Department of Physics, Norwegian University of Science and Technology, NTNU, NO-7491 Trondheim, Norway

Geraldo J. da Silva†
Physics Institute, University of Brasília, Caixa Postal 04513, 70919-970 Brasília Distrito Federal, Brazil

Hans M. Adland, Arne Mikkelsen, and Arnljot Elgsaeter
Department of Physics, Norwegian University of Science and Technology, NTNU, NO-7491 Trondheim, Norway
(Received 30 May 2001; revised manuscript received 10 September 2001; published 23 January 2002)

We study a synthetic clay suspension of laponite at different particle and NaCl concentrations by measuring stationary shear viscosity and transient electrically induced birefringence (TEB). On one hand the viscosity data are consistent with the particles being spheres and the particles being associated with large amount bound water. On the other hand the viscosity data are also consistent with the particles being asymmetric, consistent with single laponite platelets associated with a very few monolayers of water. We analyze the TEB data by employing two different models of aggregate size (effective hydrodynamic radius) distribution: (1) bidisperse model and (2) log-normal distributed model. Both models fit, in the same manner, fairly well to the experimental TEB data and they indicate that the suspension consists of polydisperse particles. The models also appear to confirm that the aggregates increase in size vs increasing ionic strength. The smallest particles at low salt concentrations seem to be monomers and oligomers.

DOI: 10.1103/PhysRevE.65.021407 PACS number(s): 82.70.Dd, 78.20.Fm, 61.20.Lc, 66.20.+d

I. INTRODUCTION

Laponite [1–4] is a widely studied synthetic clay that belongs to the family of swelling 2:1 clays [5]. All dehydrated clays have a layered silicate mesostructure. The 2:1 clays (or smectites) thus consist of 1 nm thick and charged (negative surface charge and a smaller positive edge charge) mesosheets, which in the dehydrated state stack (like decks of cards) by sharing charge-compensating cations. Laponite is a particularly interesting model system because of the merely monodisperse size of the colloidal platelets (25–30 nm diameter, see Fig. 1). This is different from natural and other synthetic clays, which in general have a polydisperse distribution of micrometer sized platelets. Introductions describing the crystallographic structures and providing precise definitions of both natural and synthetic clays, including laponite (a synthetic hectorite), may be found in several books [5,6].

The addition of salt-containing water to these mesoscopic platelet systems gives rise to interesting colloidal dispersion ''phase'' diagrams. Four separate regions (phases) of physical complexity have been suggested from experimental observations of clay-electrolyte-concentration diagrams of laponite: isotropic liquid (IL), isotropic gel (IG), nematic gel, and flocculation [2].

Traditional theory of Derjaguin, Landau, Verwey, and Overbeek [7,8], where both van der Waals and double-layer forces are considered, provides the simplest available model capable of describing this complex behavior. Transitions and aggregate structures within stable phases may thus be discussed in terms of an interaction potential between individual platelets. This is achieved by adding the electrolyte-independent van der Waals attraction and the double-layer repulsion as characterized by an electrolyte-concentration-dependent Debye screening length [9]. The sum of these two forces yields different local potential minima, with regard to platelet-platelet interactions, which may be changed by varying the ionic strength.

The IL phase is a suspension of Brownian particles, and is made up of single platelets and/or larger aggregates of several laponite platelets suspended in water. The size and compactness of these aggregates may depend on electrolyte concentration via the Debye screening length. The aggregates in this phase are, in general, too small to scatter visible light appreciably, thus yielding a transparent liquid with a viscosity that may be changed by varying the salt-clay concentration. This phase can be made birefringent by applying high electric fields, as will be evident from the present work. At low concentrations the liquid seems to be Newtonian, whereas upon approaching the IL-IG line the IL phase can display non-Newtonian and thixotropic behavior. The aggre-

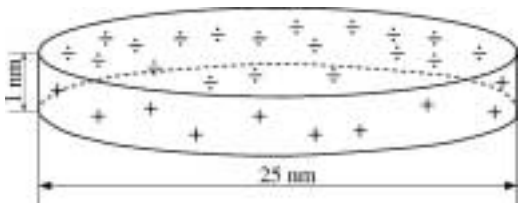

FIG. 1. Schematic illustration that shows the geometry and dimensions of a laponite platelet. The surface charges are indicated by negative charges (÷) and smaller positive charges (+) at the edges.

*Corresponding author.
Email address: Audun.Bakk@phys.ntnu.no
†Present address: Department of Physics, Norwegian University of Science and Technology, NTNU, NO-7491 Trondheim, Norway.

gate structure becomes a gel when the clay concentration becomes sufficiently large [4], possibly signaling a glass transition [10].

In the present paper we report on experimental studies of laponite samples within the low clay-concentration regime, i.e., in the IL phase. We study samples at different clay and NaCl concentrations by means of two techniques, viscometry, thus measuring the effective hydrodynamic volume of the suspended particles, and by transient electric induced birefringence (TEB) measurements, thus obtaining information about rotational diffusion. Combining our data from these two techniques we attempt to extract information about shape and size distribution of the laponite aggregates in the IL phase. The TEB technique is widely used in studying the rotational motion of macromolecules in solution [11], and has also been used for studying natural clays [12].

The goal of the present study is twofold: (1) extract information about aggregate volumes and shapes in laponite suspensions, and (2) investigate the possibilities and limitations of using TEB measurements combined with viscometry in order to characterize such complex colloid aggregate systems.

The paper is organized as follows: Section II gives a short introduction to intrinsic viscosity and in Sec. III we discuss some general aspects of rotational diffusion and birefringence. In Sec. IV we present the sample preparation and the experimental setup, and in Sec. V we discuss the experimental data from the viscosity and the TEB measurements in view of two different models with regard to the distribution of the aggregate sizes. Section VI is a summary.

II. INTRINSIC VISCOSITY OF RIGID PARTICLES

The *intrinsic viscosity* is defined as [13]

$$[\eta]=\lim_{c\to 0}\frac{\eta_{\mathrm{rel}}-1}{c}, \tag{1}$$

where c is the laponite concentration, and the relative viscosity equals

$$\eta_{\mathrm{rel}}=\frac{\eta'}{\eta}, \tag{2}$$

where η' is the macroscopic viscosity (water and laponite) and η is the water viscosity. Eq. (2) can be linearized to

$$\eta_{\mathrm{rel}}=1+[\eta]c, \tag{3}$$

which obviously has the correct limit of zero laponite concentration, where $\eta_{\mathrm{rel}}=1$.

It can be shown that $[\eta]$ for a compact macromolecule or aggregate of arbitrary shape can be described by the heuristic expression [13,14]

$$[\eta]=\nu(\delta\overline{V}_1+\overline{V}_2), \tag{4}$$

where ν is the dimensionless Simha factor containing all the shape dependence. The parameters $\overline{V}_1$ and $\overline{V}_2$ are the partial specific volumes of water (1.0 cm^3/g) and solute, respectively. The $\overline{V}_2$ is the change in solution volume per unit solute mass added, at the limit of zero solute concentration. The δ is the hydration ratio i.e., δ grams of bound water per gram solute (aggregate).

The Simha factor in Eq. (4) equals 2.5 for a sphere and increases with increasing asymmetry, i.e., $\nu\geqslant 2.5$. For the Laponite used in this present study $\overline{V}_2=0.37\,\mathrm{cm}^3/\mathrm{g}$. A change in $[\eta]$ can then arise as the result of a change in ν, changed hydration factor δ, or a combination of such changes. However, if the particle geometry is known, and $\overline{V}_1$ and $\overline{V}_2$ in Eq. (4) are known, the intrinsic viscosity is thus implicitly a measure of the hydration of the particles (aggregates).

It should be noted that it is not possible to determine the size of the particles only from a intrinsic viscosity measurement. For example, a mixture of spheres with different sizes will give the same $[\eta]$ as a monodisperse solution of spheres, provided that the hydration ratios of the spheres in the two cases are the same.

III. BIREFRINGENCE OF RIGID PARTICLES

When no external forces influence the orientation of the particles, they will be randomly oriented and in thermodynamic equilibrium. If the particles for some reason have a specific orientation at a given time, thermal (Brownian) motion will make the system decay to this equilibrium. It can be shown that the birefringence relaxation time of a dilute solution of identical particles, can have as many as five relaxation times [15]. All these decay times are known functions of the rotational mobility tensor [15,16], but for most particle geometries it is not possible, using experimental data, to obtain reliable estimates of more than two decay times.

In the analysis of the TEB data we will restrict ourselves to models with geometries that makes it adequate to use only one decay time. To induce the birefringence it is common to use electric field pulses that are rectangular as function of time, but in order to analyze the properties of the particle electric dipoles it is also useful to employ double pulses where the electric field of the second pulse is reversed relative to the first pulse [17,18].

In the case when the system is dominated by two distinct types of particles, i.e., a *bidisperse* model, each type of particles may have its own relaxation time. The birefringence signal at time t can then be described by

$$\frac{\Delta n(t)}{\Delta n(0)}=a_1\exp\left(-\frac{t}{\tau_1}\right)+a_2\exp\left(-\frac{t}{\tau_2}\right), \tag{5}$$

where $\Delta n(0)$ is the birefringence at time $t=0$, and a_1 and a_2 represent the relative contribution to the total birefringence from each of the two particle types, respectively. This model is used in Sec. V B. However, Eq. (5) can also be interpreted as representing a system of monodisperse anisotropic particles for which the five relaxation times of the particles is reduced to two times because of the particle symmetry.

For a rigid body, with one axis of rotational symmetry, the birefringence relaxation time is given by $\tau=(6D^{(\mathrm{rot})})^{-1}$

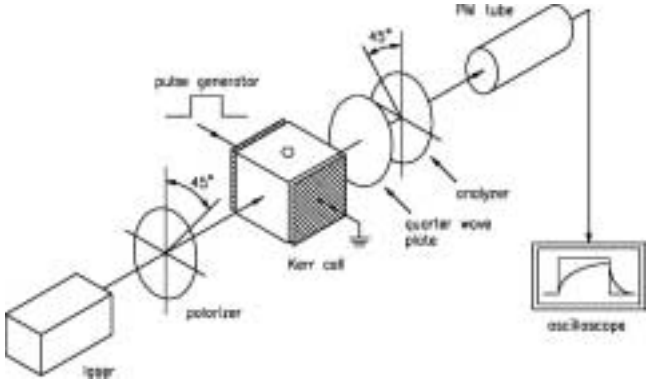

FIG. 2. Schematic illustration of the TEB experimental setup. Brief description in Sec. IV C; for further details see, e.g., Refs. [17,18,25].

[15,16], where $D^{(\mathrm{rot})}$ is the macroscopic rotational diffusion coefficient. The Nernst-Einstein relation [19], valid for dilute solutions, relates $D^{(\mathrm{rot})}$ to the rotational friction tensor $\zeta^{(\mathrm{rot})}$ [20]

$$D^{(\mathrm{rot})}=\frac{k_BT}{\zeta^{(\mathrm{rot})}}, \tag{6}$$

where k_B is the Boltzmann constant and T is the absolute temperature. The rotational friction for a sphere of radius r_s equals $\zeta_s^{(\mathrm{rot})}=8\pi\eta r_s^3$ [13], where η is the viscosity of water. This yields a birefringence relaxation time

$$\tau=\frac{\zeta^{(\mathrm{rot})}}{6k_BT}=\frac{4\pi\eta}{3k_BT}r_s^3. \tag{7}$$

Equation (7) relates the birefringence relaxation time to the effective hydrodynamic radius for rotation of a sphere.

IV. EXPERIMENT

A. Sample preparation

Laponite RD powder as purchased from Laporte Absorbents (UK), was added to NaCl containing water. The *p*H in the salt-water was adjusted to 10 before addition of laponite powder in order to prevent decomposition of the platelets themselves [2,21]. The samples were then stirred for two days, using a magnetic stirrer, before each sample was placed in a sample tube. Small sample portions were then taken from the tubes and the experiments reported here were performed. All investigated samples were 1–1.5 months old, with two exceptions. Samples AA and AB discussed below were both 6 months old.

It is important to note that laponite suspensions are known to show slow long-term aging effects [22]. This is probably due to decomposition of individual platelets, and thereby disintegration of the aggregates, when the suspension samples are not sealed from air. Such scaling was not carried out for the present samples. The effect of aging is, therefore, expected to be significant for the aggregates characterized here. Nevertheless, our samples were prepared and stored in a repeatable manner.

The present samples were not filtered, as was done and emphasized recently by Nicolai and Cocard [21] for their light scattering studies. Our samples could thus contain some large impurities reported in some cases to dominate static light scattering experiments [21], but both TEB and viscosity measurements are less sensitive to such impurities than static light scattering.

B. Viscometer

The viscosity was measured using a rotational conicylindrical viscometer (CONTRAVES Low Shear 30) with a *Couette* geometry, which consists of a static rod, measuring the torque, in a concentric rotating cup filled with the suspension. For a further introduction to the instrumental setup and the theoretical aspects of the rheology, see e.g., Van Wazer *et al.* [23]. The temperature in the suspension was fixed to 20 °C throughout the experiment by a thermostat (Haake D8). The torque signal from the rod was sampled and transformed to viscosity by a instrumentation data program (LABVIEW) [24].

C. TEB setup

The setup for the TEB instrument is shown in Fig. 2. The light source is an argon laser (Omnichrome 543-AF) operated at wavelength 488 nm. The monochromatic light is polarized at an angle of 45°, relative to the electric field, and passes through the Kerr cell where the sample is located. In the Kerr cell the aggregates are exposed to a pulsed electric field in the horizontal direction (see Fig. 2). The distance between the parallel electrodes is 4 mm. The optical anisotropy caused by the electric field makes the light exiting the Kerr cell, temporally, elliptically polarized.

When the principal axis of the analyzer is oriented perpendicular to the polarizer, the light intensity measured by the photomultiplicator (PM) is proportional to the square of the birefringence Δn (quadratic detection). The PM voltage is displayed on a digital storage oscilloscope (Tektronix TDS 620).

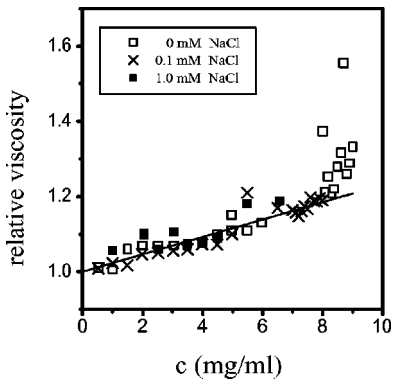

FIG. 3. Relative viscosity [see Eq. (2)] vs laponite weight concentration at three different NaCl concentrations. The concentration parameter c is weight laponite per unit volume of water. The straight line corresponds to a linear least-squares fit of the linear interval of the relative viscosity $\eta_{\mathrm{rel}}=1+0.23c$.

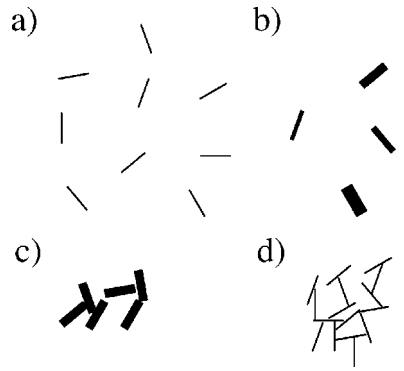

FIG. 4. Schematic examples of different models of laponite particle association. The different models shown are (a) dispersed suspension that consists of noninteracting single laponite platelets, (b) noninteracting stacks of platelets, (c) interacting stacks of platelets, and (d) ''house of cards'' configuration. For a further discussion of the different models see, e.g., van Olphen [5].

The temperature in the Kerr cell was fixed to 20 °C throughout the experiment by a thermostat (Haake D8). During the experiment the temperature was checked by a platinum thermometer. The temperature rise was found to be about 0.8 °C after a typical series of six pulses with 0.25 ms pulse length, 2.5 s intervals, and using an applied voltage of 850 V.

For further technical details on TEB experiments see Ref. [17,18,25].

V. DISCUSSION

A. Viscosity

In Fig. 3 we plot the relative viscosity [see Eq. (2)] vs the laponite concentration for three different NaCl concentrations (0, 0.1, and 1.0 mM). At low laponite concentrations ($c<8$ mg/ml) one sees a typical linear dependence of the relative viscosity, while for large Laponite concentrations ($c>8$ mg/ml) it raises abruptly. The latter is possibly due to a onset of gelation of the clay suspension [2]. In the linear regime we do a linear least-squares fit, where the inclination of the relative viscosity equals the intrinsic viscosity [see Eq. (3)]. We thus obtain $[\eta]=23\pm1\ \mathrm{cm}^3/\mathrm{g}$ for 0 and 0.1 mM [NaCl], and $[\eta]=26\pm3\ \mathrm{cm}^3/\mathrm{g}$ for 1.0 mM [NaCl]. As the difference is small between the various salt concentrations we obtain the value

$$[\eta]=23\pm1\ \mathrm{cm}^3/\mathrm{g}, \tag{8}$$

based upon the linear part of the relative viscosity data at all salt concentrations.

We may define a critical particle concentration c^* where there is one platelet per cube with side lengths that equals 25 nm, i.e., the diameter of a laponite platelet, and these cubes occupy the entire volume. This yields $c^*\approx85$ mg/ml. In the present experiments the largest Laponite concentration is 8 mg/ml, i.e., $c\ll c^*$.

One difficulty associated with using Eq. (4) is that we do not know the value of δ, because the hydration depends of the geometry of the associated particles. This association will again depend upon the laponite concentration and the ionic strength of the solution. Figure 4 gives examples of different models of particle association. The ''house of cards'' configuration in Fig. 4(d) has likely a large δ compared to, e.g., the single platelets in Fig. 4(a).

Assuming a spherical shape of the aggregates, the Simha factor ν in Eq. (4) equals 2.5, as stated in Sec. II. Inserting $[\eta]=23\ \mathrm{cm}^3/\mathrm{g}$, $\nu=2.5$, $\overline{V_1}=1.0\ \mathrm{cm}^3/\mathrm{g}$, and $\overline{V_2}=0.37\ \mathrm{cm}^3/\mathrm{g}$ into Eq. (4), yields a hydration ratio $\delta=8.8$, i.e., 8.8 g of water per gram of laponite platelets. The laponite platelet has a density 2.7 g/cm^3, thus $\delta=8.8$ yields that the hydrated water volume, associated to each laponite platelet, is on average approximately 24 times the volume of a platelet. In other words, the viscosity experiment, based upon a spherical geometry of the aggregates, shows that aggregates are associated with a large amount of water. The latter may indicate a kind of ''house of card'' [5] association of the platelets, as shown in Fig. 4(d), but we stress that this conclusion may be an artifact due to the assumed spherical shape of the aggregates.

In light of the measured intrinsic viscosity $[\eta]=23\ \mathrm{cm}^3/\mathrm{g}$ [see Eq. (8)], it is interesting to compare this to that of an aggregate system where each platelet *on average* occupies a cylindrical volume with diameter and height equivalent to its own volume, i.e., 25 nm. This means that one laponite platelet (see Fig. 1) is associated with a water volume that is 24 times its own volume, which is the same estimate as obtained from the spherical approximation used above.

However, the source of the intrinsic viscosity described by Eq. (4) may also be an asymmetric shape of the particles, which will give a larger Simha factor [13] and consequently a smaller hydration ratio δ in Eq. (4). Thus, asymmetry implies a more dense packing of the aggregates compared to spherical aggregates, when a constant intrinsic viscosity is assumed.

Assuming an asymmetric particle shape is further motivated by, e.g., Avery and Ramsay [26], Rosta and von Gunten [27], and Nicolay and Cocard [21] who concluded that the smallest particles in a laponite suspension were monomers [21,26] and/or oligomers [21,27]. In this respect it is interesting to investigate the maximum asymmetry that is reconcilable with a ''realistic'' minimum of hydration. We assume a monolayer of water molecules with thickness 2.5 Å around each laponite platelet, independent of whether it is single or in an aggregate. This gives a hydration ratio $\delta=0.2$. Inserting this $\delta=0.2$ into Eq. (4) implies a Simha factor $\nu=40$, which corresponds to an axial ratio 25 for an oblate ellipsoid [13]. This is interesting because an axial ratio 25 is approximately the ratio between the diameter and thickness of a single laponite platelet, which indeed can be approximated by an oblate ellipsoid.

Thus, the viscosity experiment does not exclude the possibility that the suspension consists of oblate particles that have axial ratios up to around 25, e.g., laponite monomers.

B. Bidisperse model (model 1)

The experimental TEB data for four different samples are shown in Fig. 5. It is assumed that the fluctuations of the data

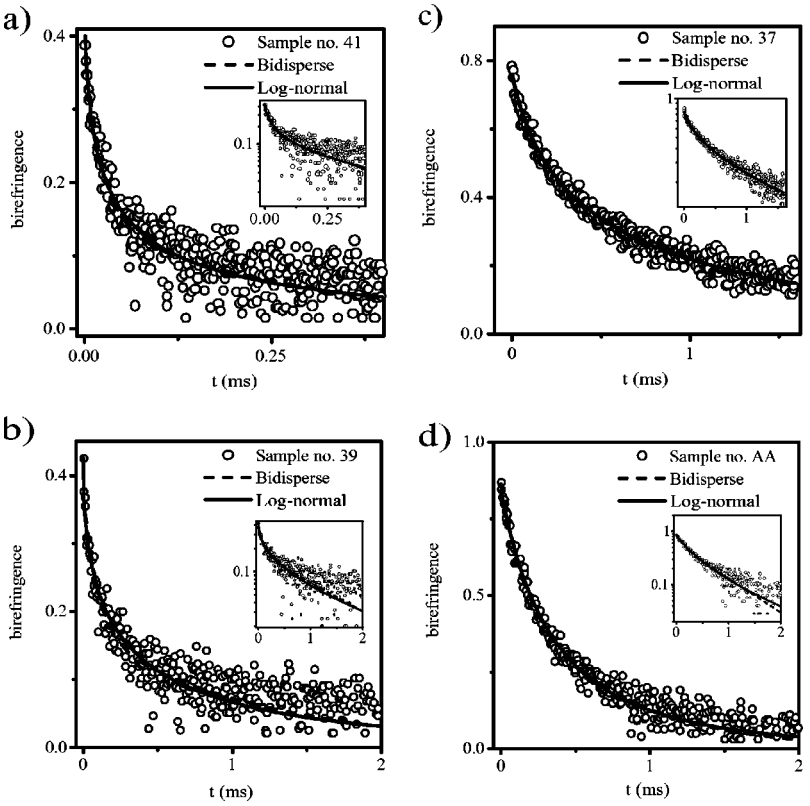

FIG. 5. Birefringence vs time t for four different samples. The experimental data are fitted by studying two different models: the bidisperse model in Sec. V B and the log-normal model in Sec. V C. The corresponding parameters are listed in Tables I and II. Note that for the inserted plots the axis of the birefringence is logarithmic, while the t-axis is linear.

in the TEB experiment are the a result of instrument noise, and not intrinsic to the samples. One source of error is the PM tube with an absolute error of ± 0.05 V. This will make the analysis for large times, where the relaxation signal is small, uncertain. The error estimates of the least-squares fit for the bidisperse model and log-normal model are presented in Table I and II, respectively.

The simplest possible fit to the TEB data is obtained by using two exponential functions with different birefringence decay times as in Eq. (5). The two relaxation times τ_1 and τ_2, correspond to two effective hydrodynamic radii r_1 and r_2, respectively. We have assumed a spherical aggregate geometry. The relation between τ and r for a sphere is given in Eq. (7). The obtained parameter estimates are presented in Table I. We find that increase of the salt concentration results in longer relaxation times, as seen in Table I.

We note from Table I that the standard deviation is small, despite the relatively large instrument error. This is probably due to the large amount of data available for each sample. However, samples numbers 28 and 29 exhibit a large standard deviation of the parameter r_2, but note that the corresponding amplitudes a_2 are small.

We now look at the shortest relaxation times at low salt concentrations ($[\mathrm{NaCl}] \leqslant 0.1$ mM). For spherical particles this corresponds to a mean value $\langle r_1 \rangle = 23$ nm. It would be interesting to calculate the corresponding size of a particle with an axial ratio that corresponds to a hydrated laponite platelet. If we assume one monolayer of water with thickness 2.5 Å attached to the platelet with diameter 25 nm and height 1 nm, this corresponds to an hydration axial ratio $p \approx 17$. Employing data given in Ref. [13] we find that the diameter of this particle is about 30 nm if we assume the shape to be that of an oblate ellipsoid.

By doing the same analysis as in the previous paragraph and using $p = 25$, which was the largest possible asymmetry calculated from the viscosity experiment, we also get a diameter around 30 nm. Thus, from the viscosity experiment together with the fit of the bidisperse model to the TEB data we may conclude that the smallest particles in the suspension have diameters in the range 20–35 nm, and that the diameter increases with increasing asymmetry of the particles. If we further assume that the individual platelets do not decompose, we conclude that the smallest particles at low salt content are individual platelets as Avery and Ramsay [26], and Nicolay and Cocard [21] concluded.

It is interesting to note that the effective hydrodynamic radius for the smallest aggregates are approximately one half of the larger ones for the corresponding pairs, i.e., $r_2/r_1 \approx 2$ in Table I. This means that we have some small aggregates and some large aggregates. Figure 5 shows the parameter fitting for some of the sample data. The fit is satisfactory in view of the simplicity of the model. However, it is more realistic to expect a *distribution* of the particle sizes. In the next section we will show that the TEB data also can be approximated by a log-normal distribution.

C. Log-normal distributed model (model 2)

The exponential fits in the preceding section suggest that there may be a broad distribution of the colloid particle size for some of the samples. Next we, therefore, tried to fit the TEB data, with regard to the size of the aggregates, onto a

TABLE I. Parameters associated with the various samples when a sum of two exponential decays is used (model 1). Parameters τ_i and r_i are the corresponding pairs of the relaxation times and the effective hydrodynamic radii, respectively, a_2/a_1 is the amplitude ratio of the birefringence, r_m is the weighted mean of r_1 and r_2, c equals mass laponite per unit volume of water, and [NaCl] is the salt concentration. The standard deviation (SD) is written as $\pm$SD.

Sample	τ_1 (μs)	r_1 (nm)	τ_2 (μs)	r_2 (nm)	a_2/a_1	r_m (nm)	c (mg/ml)	[NaCl] (mM)
21	3.8	16 $\pm$1	55	37 $\pm$2	1.3	28	1.0	0
20	2.4	13 $\pm$1	40	33 $\pm$1	1.4	25	1.5	0
18	2.5	13 $\pm$1	50	36 $\pm$2	0.67	22	2.5	0
11	68	40 $\pm$2	240	62 $\pm$2	1.5	53	6.0	0
41	24	29 $\pm$1	320	68 $\pm$2	0.58	43	8.0	0
30	55	37 $\pm$7	200	58 $\pm$3	3.2	53	1.0	0.1
29	14	25 $\pm$1	1010	99 $\pm$20	0.20	37	1.5	0.1
28	8.2	20 $\pm$1	850	94 $\pm$94	0.11	18	2.0	0.1
27	2.3	13 $\pm$1	59	39 $\pm$3	0.35	20	2.5	0.1
AB	29	30 $\pm$2	390	72 $\pm$2	2.1	58	2.5	0.5
39	97	46 $\pm$2	1170	100 $\pm$2	0.85	71	1.0	1.0
38	57	37 $\pm$1	790	92 $\pm$3	0.58	57	1.5	1.0
37	82	43 $\pm$1	650	86 $\pm$1	1.7	70	2.0	1.0
AA	140	52 $\pm$3	680	88 $\pm$1	1.8	75	2.5	4.0

Gaussian distribution, but we found that this did not work well in general. However, we found that the TEB data fit well onto a log-normal distribution, with regard to the radius of the aggregates. The log-normal probability distribution of a quantity r, with mean r_s, and standard deviation σ reads

$$p(r)=\frac{1}{\sqrt{2\pi}\sigma r}\exp\left(-\frac{\ln^2(r/r_s)}{2\sigma^2}\right), \tag{9}$$

which is asymmetric and has a longer tail than the normal distribution, as shown in Fig. 6. It should also be noted that the Gaussian distribution is somewhat unphysical in light of a probability larger than zero for $r<0$, in contrast to the log-normal distribution that is only defined for $r\geqslant 0$.

The log-normal distribution was fitted to the square root of the voltage from the PM, which is proportional to the birefringence $\Delta n(t)$, where t is the time. When we assume an exponential decay of the birefringence, aggregate radius r, and relaxation time τ_s, the birefringence reads

$$\Delta n(t)\sim\int_0^\infty dr\, p(r)e^{-t/\tau_s}, \tag{10}$$

and τ_s is the birefringence relaxation time for a sphere (s). The τ_s is given by letting $\tau\rightarrow\tau_s$ in Eq. (7).

The parameter fit is presented in Table II, from which it is clear that we have in general, a broad particle size distribution, i.e., large standard deviations σ. In Fig. 5 we see that the log-normal approximation is almost equivalent to the bidisperse model for these samples, with regard to the fit to the TEB data.

An interesting observation was that for sample number 39, an applied Kerr cell voltage of 850 V for 0.5 ms was not enough to align the aggregates, because the birefringence signal seemed not to be saturated at $t=0$, i.e., when the electric field was switched off. If we applied the same voltage for 1.5 ms the birefringence amplitude became saturated at $t=0$. However, the relaxation birefringence signal (after $t=0$) associated with a pulse length of 1.5 ms applied to sample number 39 appears to be almost equivalent to the birefringence signal associated with a pulse length of 0.5 ms. Nevertheless, it would in a future experiment be interesting to investigate the samples using different applied voltages and pulse lengths, which effectively can act as a particle filter.

It is worth noting that the log-normal distribution has also been applied to describe the polydispersity of the diameters of magnetic cores in ferrofluids [28]. Furthermore, Ivanov [29] showed that the continuous behavior of this particle distribution can be substituted by a bidisperse model, when the majority of the particles have the smaller radius.

One source of error in the experiments reported here is that we do not know whether the rotational motion of the aggregates in the Couette geometry deform the aggregates permanently, and thus giving rise to another configuration than before the experiment. In future experiments it would be interesting to first test the samples in the TEB device, whereupon one should measure the viscosity, and finally put the sample back into the TEB device in order to check the influence of the flow on the aggregates in the viscometer. It might also be interesting to check flow induced birefringence in order to investigate the latter effect.

The log-normal distribution for four different samples is drawn in Fig. 6, all showing the characteristic long tail. Because these distributions yielded the best fit to the TEB data, this may tell us that the samples have some large aggregates that make the distribution asymmetric compared to a Gaussian distribution [21]. In Fig. 6 we also see the effect of a relatively small standard deviation for sample number *AA* ($\sigma=0.24$), where in this particular case the log-normal distribution could be approximated by a normal distribution. We note for the log-normal distributed model, as for the bidisperse model, that for some of the samples the fit to experimental data for long times is only fair. Thus, we cannot rule out the presence of some large particles which we are not able to account for within our present models.

TABLE II. Parameters associated with the log-normal distribution (model 2). Parameter r_s is the mean value of the effective hydrodynamic radius, with a standard deviation σ. Parameters c and [NaCl] are explained in Table I. The standard deviation (SD) is written as $\pm$SD.

Sample	r_s (nm)	σ	c (mg/ml)	[NaCl] (mM)
21	24±2	0.71±0.09	1.0	0
20	23±1	0.65±0.06	1.5	0
18	17±1	0.73±0.03	2.5	0
11	52±1	0.23±0.01	6.0	0
41	33±2	0.70±0.05	8.0	0
30	53±1	0.18±0.03	1.0	0.1
29	32±4	0.7±0.1	1.5	0.1
28	27±4	0.9±0.2	2.0	0.1
27	14±1	0.72±0.04	2.5	0.1
AB	61±2	0.61±0.08	2.5	0.5
39	80±5	0.7±0.1	1.0	1.0
38	56±1	0.44±0.02	1.5	1.0
37	71±1	0.46±0.02	2.0	1.0
AA	75±1	0.24±0.02	2.5	4.0

The large standard deviation seen in Table II indicates that we have a polydisperse particle size distribution. This is substantiated by a look at the amplitude ratios in Table I, which fluctuates around 1 and where the corresponding pairs of radii have a ratio around 2. Thus, we come to the same conclusion as Nicolay and Cocard [21] in their static and dynamic light scattering experiments, i.e., that the laponite particles are polydisperse. Here we note that Rosta and von Gunten [27] come to the opposite conclusion, i.e., that the suspension is more or less monodisperse.

VI. SUMMARY AND CONCLUSION

We study a laponite clay suspension at different laponite and salt concentrations by shear viscometry and by TEB measurements. This is to the authors' knowledge the first reported TEB study of laponite. Two different models of the distribution of the aggregate sizes are considered: (1) bidisperse model and (2) log-normal distributed model.

The viscosity data show, when assuming that the aggregates are spherical, that we have a ''house of cards'' [5] association of the laponite platelets, i.e., the aggregates are associated with a large amount of hydrated water. However, the viscosity data may also be interpreted as being due to asymmetrical aggregates with a smaller hydration ratio than the spherical aggregates. In this respect we show that it is possible to interpret the viscosity data as oblate ellipsoidal

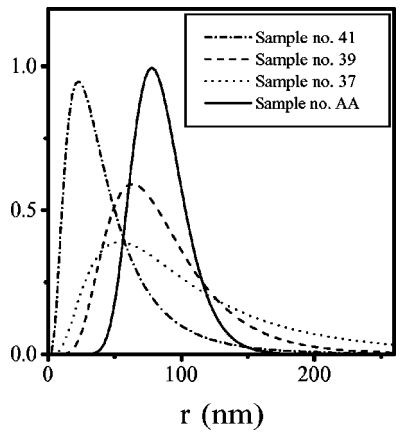

FIG. 6. Log-normal distribution of the samples that correspond to Fig. 5, where r is the effective hydrodynamic radius of the aggregates. The distribution has been normalized to unity.

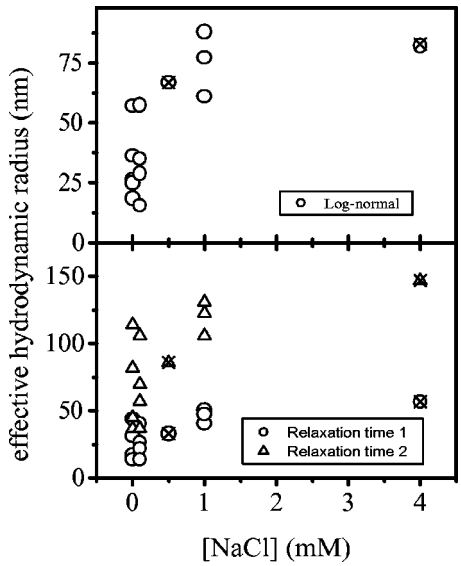

FIG. 7. Effective hydrodynamic radius for the log-normal distributed model of the aggregate sizes in the upper plot, and the effective hydrodynamic radius for the bidisperse model associated with the two relaxation times τ_1 and τ_2 in the lower plot. Data from Tables I and II are used. The experimental data points with ''$\times$'' correspond to samples *AA* and *AB*, which were 6 months old, as stated in Sec. IV A.

particles with an asymmetry up to 25, e.g., single laponite platelets.

The two models were fitted to experimental data obtained using TEB technique, and the models seem to fit the experimental data fairly well in most cases. This means, for the samples studied, that the simplistic bidisperse model is almost equivalent to the log-normal distribution. Most of the samples show, in view of the log-normal model, that we have a broad particle size distribution. This is the same conclusion as Nicolay and Cocard [21], and the opposite of Rosta and von Gunten [27], who performed light scattering experiments. However, from the available data it cannot be ruled out that the apparent lack of fits, for some long time tails, may suggest a particle size distribution that includes some larger laponite aggregates that cannot be covered within our present models.

Both models show that the aggregates increase in size, from a hydrodynamic point of view, when the salt concentration is increased, as seen in Fig. 7. Furthermore, our present experiments do not confirm or rule out the possibility of a correlation between particle size and laponite concentration. The smallest particles in the suspensions consist of only one or a very few laponite platelets at low salt concentrations, as also Avery and Ramsay [26], Rosta and von Gunten [27], and Nicolay and Cocard [21] concluded.

Finally we note that the present laponite samples were not made, filtered, or stored (scaled from air) in the same controlled manner as recent light scattering studies by Nicolai and Cocard [21].

In a future work it would be important to prepare the samples in a different way and also to study other synthetic clay suspensions such as fluorohectorite. In light of the broad asymmetric tails of the log-normal distribution for most of the samples reported here, it would be of considerable interest and importance to study suspension samples for which the large aggregates are filtered out successively.

ACKNOWLEDGMENTS

The authors would like to thank L. Wohlen for the drawing of Fig. 2. A.B. and J.O.F. thank NTNU/NFR SUP 115185/420 for financial support during the first part of the work. A.B. thanks the Research Council of Norway for financial support (Contract No. 129619/410) in the latter part of the work. H.M.A. thanks the Department of Physics, NTNU, for financial support. G.J. da S. acknowledges the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq/Brazil) for financial support.

[1] J.-O. Fossum, in *Soft Condensed Matter: Configurations. Dynamics, and Functionality*, edited by A. T. Skjeltorp and S. F. Edwards (Kluwer Academic, Netherlands, 2000), p. 269, and references therein.

[2] A. Mourchid, E. Lecolier, H. VanDamme, and P. Levitz, Langmuir **14**, 4718 (1998).

[3] J. P. C. Gabriel, C. Sanchez, and P. Davidson, J. Phys. Chem. **100**, 11 139 (1996).

[4] S. Cocard, J. F. Tassin, and T. Nicolai, J. Rheol. **44**, 585 (2000), and references therein.

[5] H. van Olphen, *An Introduction to Clay Colloid Chemistry*, 2nd ed. (Krieger Publishing Company, Florida, 1991).

[6] B. Velde, *Introduction to Clay Minerals* (Chapman and Hall, London, 1992).

[7] B. V. Derjaguin and L. Landau, Acta Physicochim. URSS **14**, 633 (1941).

[8] E. J. W. Verwey and J. T. G. Overbeek (with the collaboration of K. van Nes), *Theory of the Stability of Lyophobic Colloids: the Interaction of Sol Particles Having an Electric Double Layer* (Elsevier, New York, 1948).

[9] J. N. Israelachvili, *Intermolecular and Surface Forces*, 2nd ed. (Academic, London, 1992).

[10] D. Bonn, J. Tanaka, G. Wegdam, H. Kellay, and J. Meunier, Europhys. Lett. **45**, 52 (1999).

[11] P. J. Hagerman, Curr. Opin. Struct. Biol. **6**, 643 (1996); C. Houssier, Colloids Surf., A **148**, 3 (1999); R. Sasai, N. Ikuta, and K. Yamoaka, J. Phys. Chem. **100**, 17 266 (1996); J. K. Phalakornkul, A. P. Gast, and R. Pecora, Macromolecules **32**, 3122 (1999); F. Charney, Q. Rev. Biophys. **21**, 1 (1988).

[12] I. C. Hinds, P. J. Ridler, and B. R. Jennings, Clay Miner. **31**, 549 (1996).

[13] C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry* (Freeman, San Francisco, 1980), Vol. II, Chaps. 10 and 12.

[14] A. Bakk, Phys. Rev. E **63**, 061906 (2001).

[15] W. A. Wegener, R. M. Dowben, and V. J. Koester, J. Chem. Phys. **70**, 622 (1979).

[16] J. G. de la Torre, B. Carrasco, and S. E. Harding, Eur. Biophys. J. **25**, 361 (1997).

[17] E. Frederiq and C. Houssier, *Electric Dichroism and Electric Birefringence* (Clarendon, Oxford, 1973)

[18] A. Bjørkøy, A. Elgsaeter, and A. Mikkelsen, Biophys. Chem. **72**, 247 (1998).

[19] W. A. Wegener, R. M. Dowben, and V. J. Koester, J. Chem. Phys. **73**, 4086 (1980).

[20] R. B. Bird, C. F. Curtiss, R. C. Armstrong, and O. Hassager, *Dynamics of Polymeric Liquids*, 2nd ed. (Wiley, New York, 1987) Vol. 2.

[21] T. Nicolay and S. Cocard, Langmuir **16**, 8189 (2000).

[22] A. Mourchid and P. Levitz, Phys. Rev. E **57**, R4887 (1998); A. Knaebel, M. Bellour, J.-P. Munch, V. Viasnoff, F. Lequeux, and J. L. Harden, Europhys. Lett. **52**, 73 (2000).

[23] J. R. Van Wazer, J. W. Lyons, K. Y. Kim, and R. E. Colwell, *Viscosity and Flow Measurement-A Laboratory Handbook of Rheology* (Interscience, New York, 1963), Chap. 2.

[24] A. Bakk, M.Sc. project, NTNU, 1998.

[25] A. Bjørkøy, Ph.D. thesis, Norwegian University of Science and Technology, NTNU, 1997.

[26] R. G. Avery and J. D. F. Ramsay, J. Colloid Interface Sci. **109**, 448 (1986).
[27] L. Rosta and H. R. von Gunten, J. Colloid Interface Sci. **134**, 397 (1990).
[28] J. C. Bacri, R. Perzynski, D. Salin, V. Cabuil, and R. Massart, J. Magn. Magn. Mater. **62**, 36 (1986).
[29] A. O. Ivanov, Colloid J. USSR **59**, 446 (1997) [Kolloidnyi Zhurnal, **59**, 482 (1997)].

Paper 10

Bakk, A ; Høye, JS ; Hansen, A: Apolar and Polar Solvation Thermodynamics Related to the Protein Unfolding Process. *Biophys J* , 82(2002), 713-719

This paper is not included due to copyright restrictions.

Paper 11

Thermodynamics of Proteins: Fast Folders and Sharp Transitions

Audun Bakk [a,1], Paul G. Dommersnes [a,1], Alex Hansen [a,*], Johan S. Høye [a], Kim Sneppen [a], Mogens H. Jensen [b]

[a] *Department of Physics, Norwegian University of Science and Technology, NTNU, NO-7491 Trondheim, Norway*
[b] *Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen Ø*

Abstract

Several small globular proteins exhibit a simple two-state folding process (sharp transition). The rather short folding times of proteins (fast folders) indicate that folding is guided through some sequence of contact bindings. We discuss the possibility for reconciling a two-state folding event with a sequential folding process, i.e., a folding pathway in a schematic model of protein folding. We show that both single and multiple folding pathways can lead to an apparent two-state folding from a thermodynamic point of view. We also discuss water interactions in protein folding, leading to cold and warm destabilization of the protein.

Key words: protein folding, thermodynamics, folding pathway, van't Hoff enthalpy relation
PACS: 87.14.Ec, 05.70.Ce, 87.15.Aa

1. Introduction

Proteins appear to fold into a unique native conformation, in spite of an "astronomical" number of alternative configurations. This apparent paradox, usually attributed to Levinthal [1], is further sharpened in view of the fact that there is experimental evidence that the folding transition behaves nearly like a two-state system for many single domain proteins [2,3]. This means that for these proteins the transition from denatured (unfolded) to native (folded) state occurs rather directly without observed intermediates. One might think that such a two-state behavior excludes the possibility of guiding the protein to the native state. The purpose of the present work is to quantify the degree of guiding that is compatible with the observed two-state folding process.

The van't Hoff enthalpy relation relates the latent heat (Q) of a smoothed-out first order phase transition, taking place at T_c, to the height of the heat capacity peak ΔC [4]

$$Q^2 = \alpha k_B T_c^2 \Delta C, \tag{1}$$

where α is a dimensionless proportionality factor and k_B is the Boltzmann constant. A small α indicates a *sharp* transition. The T_c is defined as the temperature where the protein has equal free energies in the native state and in the denatured state.

When the transition is two-state it is known that $\alpha = 4$ [4] and when the transition has intermedi-

* Corresponding author.
Email address: Alex.Hansen@phys.ntnu.no (Alex Hansen).

[1] A.B. and P.G.D. thank the Research Council of Norway (NFR) for financial support.

ates then $\alpha > 4$ [5]. For the small single domain proteins ribonuclease, lysozyme, chymotrypsin, cytochrome c, and myoglobin Privalov and Khechinashvili [4] found experimentally $\alpha = 4.2$ to within 5 % accuracy, demonstrating that these transitions are very nearly two-state.

Here we describe some recent attempts of ours to reconcile these two simultaneous phenomena, i.e., coexistence of sharp transitions and short folding times. We first analyze a protein model with a single pathway. This analysis includes interactions between the protein and the surrounding water. This allows us not only to describe the usual unfolding transition, but also the cold unfolding transition, i.e., when the protein melts as the temperature is lowered [5]. We also discuss a generalization of the model to multiple pathways. This provides us with a different mechanism to destabilize the intermediate states [6].

2. Modeling the protein

2.1. *Single pathway*

In this section we show how one can model protein folding using a contact energy Hamiltonian. As a simple guiding principle, we adopt the sequential "zipper"-like [7] description of the process. We here view the zipper as an effective description of a unique folding pathway, i.e., a hierarchically ordered sequence of binding events between different parts of the protein [8,5,9].

One may visualize each binding event as closing of a specific contact between two different parts of the protein. Each of these events is characterized by the binary variable Ψ_i that indicates whether the contact i is closed ($\Psi_i = 1$) or open ($\Psi_i = 0$). The pathway implies the constraint

$$\Psi_i \geq \Psi_{i+1}, \tag{2}$$

because a folded contact i is not assumed to unfold when a contact $j > i$ is folded. This means that it is difficult to unfold a part of the protein within an already larger folded structure. In order to implement this into a Hamiltonian we introduce a second set of binary variables $\xi_i \in \{1, -B\}$, where $B \gg 1$. Thus, for a system of N contacts the chain-chain Hamiltonian becomes

$$H_{c1} = -\epsilon_c \sum_{i=1}^{N} \Psi_i \xi_i, \tag{3}$$

where ϵ_c is the energy gain to fold one contact [10].

From the constraints in Eq. (2) and requiring for simplicity $B \to \infty$, the Hamiltonian in Eq. (3) can be reformulated by the transformation $\Psi_i = \varphi_1 \varphi_2 \cdots \varphi_i$, where $\varphi_j \in \{0, 1\}$ are binary variables. In particular $\Psi_1 = \varphi_1$. The product terms meet the assumption about a folding pathway. Thus, Eq. (3) becomes

$$H_{c1} = -\epsilon_c(\varphi_1 + \varphi_1\varphi_2 + \cdots + \varphi_1\varphi_2 \cdots \varphi_N). \tag{4}$$

Furthermore, we assume that folding of the ith contact, i.e., that the i first φ-terms in (4) equals 1, is associated with a degeneracy f^{N-i}. The partition sum of the system becomes

$$\begin{aligned} Z &= \sum_{i=0}^{N} f^{N-i} \exp(\beta\epsilon_c i) \\ &= f^N \frac{1 - \exp\left((N+1)(\beta\epsilon_c - \ln f)\right)}{1 - \exp(\beta\epsilon_c - \ln f)}. \end{aligned} \tag{5}$$

The Z rapidly changes at $\beta_c = 1/T_c = \ln f/\epsilon_c$, corresponding to a smoothed-out first order phase transition at $T_c = \epsilon_c/\ln f$. The system described by the partition sum in Eq. (5) gives rise to a sharpness $\alpha = 12$ [see Eq. (1)].

On the other hand, if we consider a system that only changes energy when the protein is in the unique native state, this Hamiltonian becomes

$$H_{c2} = -N\epsilon_c \varphi_1 \varphi_2 \cdots \varphi_N, \tag{6}$$

Assuming a degeneracy $f^N - 1$ for the unfolded protein, the corresponding partition function becomes $Z = f^N - 1 + e^{\beta N \epsilon_c}$. This leads to a sharp phase transition where $\alpha = 4$ as expected [see Eq. (1)], because this is a description of a classical two-state system [3]. There is no guiding in the Hamiltonian in Eq. (6) since the ground state, $\{1111\cdots111\}$, is one out of the f^N possible states, while all the other $f^N - 1$ states are at zero energy. Thus, the time to find the ground state for such a two-state system will be very long.

The simple two-state Hamiltonian Eq. (6) results in a sharp folding transition, but seems incompatible with fast folding. On the other hand, the guided zipper-like Hamiltonian Eq. (4) is a fast folder, but gives a less sharp folding transition. We will in this section go into further detail on this question, and show possible ways of constructing a contact energy Hamiltonian which reconciles fast folding with sharp folding transitions.

We estimate the folding time by the one-step Monte Carlo Metropolis method [11]. The Monte Carlo time is not directly related to "real time", but should give a reasonable estimate of how the folding times scales with system size N. We define the folding time τ as the average number of Monte Carlo steps needed to go from the unfolded to the completely folded state with all Monte-Carlo variables $\varphi_i = 1$.

The folding time is widely different between the guided system in Eq. (5) and the two-state system in Eq. (6). For the true two-state model the folding time $\tau \propto f^N$. This is because no variable will be fixed at value 1 before all variables are 1, thus making an average probability of $1/f^N$ of reaching the ground state at each time step. Thus, the two-state system takes exponential times to fold, in accordance with the Levinthal paradox of astronomical folding times for unguided protein folding.

For the guided system governed by Eq. (5), the folding time scales as $\tau \propto N^2$. The reason for this is that at each time step only one variable can be fixed at the value $\varphi_i = 1$, the one where the previous variable equals 1 (i.e., $\varphi_{i-1} = 1$). Attempts to change other variables will either be energetically disfavored (for $j < i$) or likely to be subjected to reversals at later stages because these conformational changes are not associated with any energy changes. When each time step allows one variable to possibly change value, it typically takes $N/2$ time steps to fix the next φ on the pathway. Summed over all subsequent variables, this gives an overall folding time scaling as $\tau \propto N^2$. The exact prefactor to this folding time depends on temperature, as increased temperature enhances the probability that an already folded variable unfolds $(1 \to 0)$ again.

To reconcile that a large class of proteins behave as a two-state system with the necessity of being able to reach the ground state in a reasonable time, we now study a combination of the two Hamiltonians in Eq. (5) and Eq. (6)

$$H_p = \lambda_c H_{c1} + (1 - \lambda_c) H_{c2}. \tag{7}$$

The $\lambda_c \in [0, 1]$ is a dimensionless parameter that weighs the contributions from the Hamiltonians H_{c1} and H_{c2}, i.e., quantifying the relative amount of guiding in the system. This construction corresponds to guided folding, but with a large energy change for the last folding step.

The partition function associated with Eq. (7) is calculated as in (5). Thus, we can define a partial free energy $F(n)$ as

$$Z = \sum_{n=0}^{N} e^{-\beta F(n)}, \tag{8}$$

where n is the degree of folding. For a given temperature the partial free energy of states is $F(n \leq N-1) = n(T \ln f - \lambda_c \epsilon_c) - TN \ln f$, and $F(N) = -N\epsilon_c$.

In Fig. 1 we show $F(n)$ schematically for different temperatures T, where we set $\epsilon_c = 1$ here and in the following discussion. Each $F(n)$ exhibits a jump at $n = N$ corresponding to a free energy change $N(1 - \lambda_c) + \lambda_c - T \ln f$ for reaching the ground state. At low T, $F(n)$ is monotonically decreasing, reflecting a fast folding kinetics where the typical folding time grows as N^2, i.e., a guided system. At an intermediate $T = T_G = \lambda_c / \ln f$ all $n < N$ are equally probable.

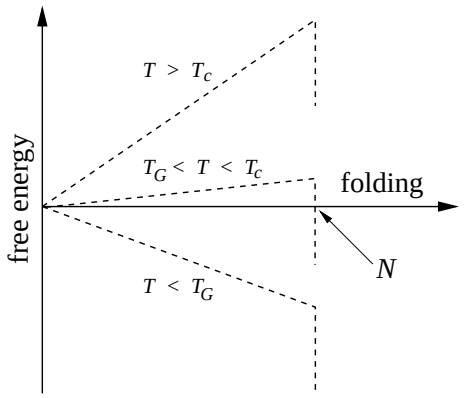

Fig. 1. Schematic drawing of the partial Gibbs free energy $F(n)$ as a function of the degree of folding n at three different temperatures [see Eq. (8)]. $F(0)$ is rescaled to 0. This figure is taken from Ref. [5].

Furthermore, T_G is lower than the folding-unfolding transition temperature T_c where the denatured state becomes thermodynamically favored. For $T_G < T < T_c$ the intermediate states are unstable (see Fig. 1), i.e., they form a barrier between the folded and denatured state – and the folding time scales exponentially with both T and N. At a higher $T = T_c = 1/\ln f$ the folded state becomes unstable, and the protein unfolds ($n \approx 0$). The fact that the free energy landscape changes with T means effectively that two-state folding around T_c is compatible with guiding and fast folding at low T.

2.2. *Multiple pathways*

The concept of two-state folding transitions and fast guided folding, can be explored further by a straightforward generalization of the strict guiding assumed in Eq. (4). In fact, many proteins show multiple pathways during folding and hence this is a necessary generalization of the model. The simplest way to incorporate *multiple* pathways is to assume that the folding variables φ can appear in any order in the Hamiltonian as

$$\begin{aligned} H = &- \sum_i a_i \varphi_i - \sum_{i \neq j} a_{ij} \varphi_i \varphi_j - ... \\ &... - \sum_{i \neq j \neq k ...} a_{ijk...} \varphi_i \varphi_j \varphi_k \cdots , \end{aligned} \tag{9}$$

where the indices $i, j, k, ...$ are chosen from the set $\{1, 2, ..., N\}$. The indices within each product term must be different and the parameters a weighs the different terms in the Hamiltonian.

For concreteness let us consider a representative example of the Hamiltonian in Eq. (9)

$$\begin{aligned} H = &- \varphi_1 - \varphi_5 - \varphi_4 \varphi_5 - \varphi_1 \varphi_2 \varphi_5 \\ &- \varphi_1 \varphi_2 \varphi_3 \varphi_4 \varphi_5 . \end{aligned} \tag{10}$$

The interpretation of this Hamiltonian is that folding of contacts 1 and 5 are independent starting points of the folding process. In order to fold contact 4, contact 5 must already be in place. However, this contact is independent whether contact 1 is formed. Contact 2, on the other hand, needs both contacts 1 and 5 in place. In the end, all five contacts are formed. A concrete example of multiple pathways has been reported for staphylococcal nuclease [12,13].

It can be shown [6] that Eq. (9) corresponds to a system with multiple folding pathways. By restricting ourselves to a class of Hamiltonians with only one term with one φ-variable, only one term with two φ-variables, etc., the system in Eq. (9) exhibits both two-state folding, and a fast folding time $\tau \propto f^{\sqrt{N}}$ at low temperatures. The key to this scaling is that after f^i folding steps, as long as i is small, typically i φ-terms are folded. When $i^2/2$ becomes comparable to N, overlap between subsequent steps becomes significant, and subsequent folding involves fewer new folding variables, and therefore folding becomes easier. For intermediate temperatures, we have as in the pure two-state model, a regime of slower folding, with folding time of order f^N.

3. How water affects protein folding

Protein-water interactions play an important role in protein folding [14–16]. We will here sketch a water model introduced by Bakk *et al.* [17,18] that is a refined version of a model first proposed by Hansen *et al.* [19]. This model is an expansion of the single pathway Hamiltonian in Eq. (4).

The basic idea of the protein-water interactions is that if the contact variable Ψ_i is closed ($\Psi_i = 1$), water has no access to that part of the protein, while if the contact is open ($\Psi_i = 0$), there is direct contact between protein and water in this region. Hence, if we call the water interaction energy E_w, the contact energy associated with the contact variable Ψ_i is

$$H_i = -\epsilon_c \Psi_i + E_w(1 - \Psi_i) . \tag{11}$$

The first part of the water interaction energy is modeled by using the simplified analogy of a classical electrical dipole in an external electrical field, whose energy is

$$E_{w1} = -\epsilon_w \cos\vartheta, \tag{12}$$

where ϵ_w is a bending distortion constant. The angle ϑ is the polar angle. Eq. (12) is the hydration model used in the works by Bakk *et al.* [17] and Bakk [20].

In addition to the energy due to the external field [Eq. (12)], we add a coupling term

$$E_{w2} = -\frac{1}{2}\sum_{i,j} J_{ij}\,\mathbf{s}_i \cdot \mathbf{s}_j \qquad (13)$$

that models pair interactions between the water molecules. The J_{ij} is the coupling constant between water molecules i and j, and $\mathbf{s}_i$ is the dipole moment of water molecule i. For simplicity we put $|\mathbf{s}_i| = 1$. It can be shown that the total water energy $E_w = E_{w1} + E_{w2}$ [Eqs. (12) and (13)] in a *mean field* solution [21] can be represented as [17]

$$E_w = E_{w1} + E_{w2} = -(\epsilon_w + bm)\cos\vartheta + \frac{1}{2}bm^2, \qquad (14)$$

where bm, with $b = \sum_j J_{ij}$, is the mean field coupling between a water molecule and its surrounding water molecules. The average dipole moment $m = \langle\cos\vartheta\rangle$ has to be determined self-consistently.

As mentioned earlier in this section, the water molecules are supposed to only interact with the unfolded regions of the protein. This protein-water interaction Hamiltonian may be written, after introducing the variables φ_i, as

$$H_w = \sum_{j=1}^{M}[E_w^{1j}(1-\varphi_1) + E_w^{2j}(1-\varphi_1\varphi_2) + ... + E_w^{Nj}(1-\varphi_1\varphi_2\cdots\varphi_N)], \qquad (15)$$

where E_w^{ij} is the energy of water molecule j in contact with the unfolded contact i, and M is the total number of water molecules associated with each contact. By summing the chain Hamiltonian H_c in Eq. (3) and the water Hamiltonian in Eq. (15), we obtain the total protein Hamiltonian

$$H = H_c + H_w. \qquad (16)$$

In Fig. 2 we show a typical plot of the heat capacity of a protein for different parameters, based upon the Hamiltonian in Eq. (16). The figure shows two characteristic peaks that correspond to cold and warm unfolding of the protein, which is common to several small globular proteins [22]. As evident from this work and other works [23–27], cold unfolding seems to be closely linked to the protein-water interactions.

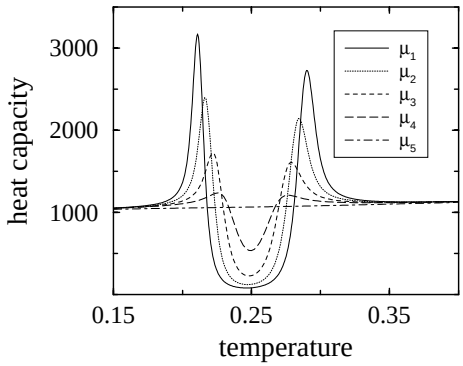

Fig. 2. Heat capacity in arbitrary units for different choice of parameters based upon Eq. (16). The μ is proportional to the contact energy ϵ_c in Eq. (3) and $\mu_i > \mu_{i+1}$. Note the characteristic two peaks, corresponding to cold and warm unfolding of the protein. This figure is taken from Ref. [17].

Finally we note that the present representation of the water is very simple. Experimentally one finds that the hydration contribution to the heat capacity is positive for apolar surfaces, while it surprisingly becomes negative for polar surfaces [28]. This important fact should be taken into account in order to refine the model further [29].

4. Summary

We have shown that it is possible to construct contact energy Hamiltonians displaying the same sharp phase transition as for a pure two-state system [van't Hoff coefficient $\alpha = 4$ in Eq. (1)], and yet allow for a fast relaxation to the ground state (fast folding time). This can be achieved either by a large energy change in the last folding step in a single folding pathway, or by constructing a Hamiltonian with multiple pathways. This first approach results in a folding time $\tau \propto N^2$, and the multiple pathways give $\tau \propto f^{\sqrt{N}}$, both in the low temperature limit. The folding time is in both cases dramatically reduced compared to that of the pure two-state system which folds in a time $\tau \propto f^N$.

Finally we show that introduction of water in our model gives two characteristic peaks for the heat capacity that correspond to cold and warm unfolding transitions of the protein.

References

[1] C. Levinthal, Are there pathways for protein folding?, J. Chim. Phys. Phys.-Chim. Biol. 65 (1968) 44-45.

[2] R. Lumry, R. Biltonen, Validity of the "two-state" hypothesis for conformational transitions of proteins, Biopolymers 4 (1966) 917-944.

[3] P.L. Privalov, N.N. Khechinashvili, A thermodynamic approach to the problem of stabilization of globular protein structure: A calorimetric study, J. Mol. Biol. 86 (1974) 665-684.

[4] P.L. Privalov, Stability of proteins. Small globular proteins, Adv. Prot. Chem. 33 (1979) 167-241.

[5] A. Bakk, J.S. Høye, A. Hansen, K. Sneppen, M.H. Jensen, Pathways in two-state protein folding, Biophys. J. 79 (2000) 2722-2727.

[6] P.G. Dommersnes, A. Hansen, M.H. Jensen and K. Sneppen, Parametrization of multiple pathways in proteins: Fast folding versus tight transitions, Cond-Mat/0006304.

[7] K.A. Dill, K.M. Fiebig, H.S. Chan, Cooperativity in protein-folding kinetics, P. Natl. Acad. Sci. USA 90 (1993) 1942-1946.

[8] A. Hansen, M.H. Jensen, K. Sneppen, G. Zocchi, A hierarchical scheme for cooperativity and folding in proteins, Physica A 250 (1998) 355-361.

[9] A. Bakk, A. Hansen, K. Sneppen, Protein model exhibiting three folding transitions, Physica A 291 (2001) 60-70.

[10] J.D. Bryngelson, P.G. Wolynes, Spin-glasses and the statistical-mechanics of protein folding, P. Natl. Acad. Sci. USA 84 (1987) 7524-7528.

[11] K. Binder, Applications of the Monte Carlo Method in Statistical Physics, Springer Verlag, Berlin, 1987.

[12] Y. Wang, D. Shortle, The equilibrium folding pathway of staphylococcal nuclease: Identification of the most stable chain-chain interactions by NMR and CD spectroscopy, Biochemistry US 34 (1995) 15895-15905.

[13] D. Shortle, Y. Wang, J.R. Gillespie, J.O. Wrabl, Protein folding for realists: A timeless phenomenon, Protein Sci. 5 (1996) 991-1000.

[14] W. Kauzmann, Some factors in the interpretation of protein denaturation, Adv. Protein Chem. 14 (1959) 1-63.

[15] D. Eisenberg, A.D. McLachlan, Solvation energy in protein folding and binding, Nature 319 (1986) 199-203.

[16] T. Lazaridis, M. Karplus, Effective energy function for proteins in solution, Proteins 35 (1999) 133-152.

[17] A. Bakk, J.S. Høye, A. Hansen, Heat capacity of protein folding, Biophys. J. 81 (2001) 710-714.

[18] A. Bakk, J.S. Høye, A. Hansen, Specific heat upon aqueous unfolding of the protein interior: a theoretical approach, Physica A 304 (2002) 355-361.

[19] A. Hansen, M.H. Jensen, K. Sneppen, G. Zocchi, Statistical mechanics of warm and cold unfolding in proteins. Eur. Phys. J. B 6 (1998) 157-161.

[20] A. Bakk, Two-state protein model with water interactions: Influence of temperature on the intrinsic viscosity of myoglobin, Phys. Rev. E 63 (2001) 061906.

[21] S.-K. Ma, Statistical Mechanics, World Scientific, Philadelphia, USA, 1985 (Ch. 27).

[22] P.L. Privalov, Cold denaturation of proteins, Biochem. Mol. Biol. 25 (1990) 281-305.

[23] P.L. Privalov, Yu.V. Griko, S.Yu. Venyaminov, V.P. Kutyshenko, Cold denaturation of myoglobin, J. Mol. Biol. 190 (1986) 487-498.

[24] Yu.V. Griko, P.L. Privalov, S.Yu. Venyaminov, V.P. Kutyshenko, Thermodynamic study of the apomyoglobin structure, J. Mol. Biol. 202 (1988) 127-138.

[25] B.-lu Chen, J.A. Schellman, Low-temperature unfolding of a mutant of phage T4 lysozyme. 1. Equilibrium studies, Biochemistry US 28 (1989) 685-691.

[26] F. Franks, Protein destabilization at low temperatures, Adv. Protein Chem. 46 (1995) 105-139.

[27] G. Graziano, F. Catanzano, A. Riccio, G. Barone, A reassessment of the molecular origin of cold denaturation. J. Biochem. 122 (1997) 395-401.

[28] G.I. Makhatadze, P.L. Privalov, Heat capacity of proteins. I. Partial molar heat capacity of individual amino acid residues in aqueous solution: hydration effect, J. Mol. Biol. 213 (1990) 375-384.

[29] A. Bakk, J.S. Høye, A. Hansen, Apolar and polar solvation thermodynamics related to the protein unfolding process, Biophys. J. 82 (2002) 713-719.

Paper 12

Heat capacities of solid state proteins: implications for protein stability in solution

Audun Bakk[1]
Department of Physics, Norwegian University of Science and Technology, NTNU, NO-7491 Trondheim, Norway

(February 25, 2002)

Abstract

We show that experimentally measured heat capacities of six different proteins (lysozyme, myoglobin, chymotrypsinogen, lactoglobulin, ovalbumine, and ribonuclease A) in their solid state, in the temperature range from 260 K to 420 K, can be well characterized as a sum of one Einstein mode, corresponding to the group vibrations, plus a constant, corresponding to the nearly excited skeletal vibrations. The relative root mean square deviations between experimental data and fitted data are less than 1%. We also show for lysozyme and ribonuclease A that the experimental values of the solid state heat capacities, corrected for intermolecular vibrational modes, deviate less than 15% relative experimental protein data in solution, where the hydration part of the latter data is excluded by a model compund evaluation [P.L. Privalov and G.I. Makhatadze, J. Mol. Biol. 224 (1992) 715]. Thus, analyzing solid state proteins may give important information to energetics and stabilization of proteins in solution.

Keywords: Protein; Solid state; Heat capacity; Vibrational modes

1 Introduction

Since the seminal work of Anfinsen [1], who established the "thermodynamical hypothesis" of proteins, a lot of experimental data have been accumulated on protein thermodynamics. Unfortunately, most of these data are on proteins in solution [2–4]. However, Makhatadze and Privalov have made important efforts to separate the hydration part of protein energetics [5–8]. Thus, subtracting the hydration part of the protein heat capacity from the total protein heat capacity, yields the intramolecular part of a protein's heat capacity. This intramolecular part of the heat capacity is likely to correspond to heat capacity of anhydrous proteins in some fashion. In the literature only a few measurements of heat capacities of anhydrous proteins exist [9–11]. In the recent years, however, the

[1]E-mail: Audun.Bakk@phys.ntnu.no

Advanced THermal Analysis System (ATHAS) [12] has contributed with data on some proteins, studied over a broad temperature range from 5 K to 420 K.

In this paper we will analyze the heat capacities of six different proteins, in their solid state form, at temperatures in the range from 260 K to 420 K. We show that it is possible to represent experimental data from Zhang et al. [13] and Di Lorenzo et al. [14], by a much simpler model than these authors applied. We represent heat capacities of solid state proteins simply by assuming that the vibrations of each protein is a sum of one Einstein mode [15] and a constant, corresponding to the group and skeletal contributions to the heat capacity, respectively. We also discuss relations between experimental data of solid state proteins and experimental protein data in solution.

2 The model

Proteins are huge macromolecules consisting of thousands of atoms. The vibrational spectra of the proteins are rather complex. The ATHAS [12] has made progress in analyzing the vibrational spectra of macromolecules, and now it contains a database of vibration spectra for over 200 linear macromolecules and several proteins. The ATHAS computational scheme is explained in, e.g., Ref. [16].

Briefly explained, in the ATHAS one splits vibrational spectra of macromolecules in group (g) vibrations and skeletal (s) vibrations. The group interactions represent vibrations from isolated groups along the (polypeptide) chain, while the skeletal vibrations account for larger intramolecular motions of the molecule. The latter also accounts for intermolecular couplings. Thus, the skeletal contribution to the energetics should not be very sensitive to molecular structure of the species.

The group vibrations are in the ATHAS represented as a sum of Einstein modes at different characteristic frequencies and box-distributions over narrow frequency ranges [17]. The box-distributions are non-analytic. In this work we will represent the group contribution by a single Einstein mode. The heat capacity at constant volume for $N_{\rm s}$ moles of group vibrators at temperature T reads [16]

$$C_{\rm V}^{\rm (g)} = \frac{N_{\rm g} R (\theta/T)^2 \, e^{\theta/T}}{(e^{\theta/T} - 1)^2}, \tag{1}$$

where $\theta = h\nu/k_{\rm B}$ is the characteristic temperature for a quantum mechanical oscillator with characteristic frequency ν. The parameters h and $k_{\rm B}$ are Planck's and Boltzmann's constants, respectively, and $R = 8.31434$ J/(K mol) is the molar gas constant.

The skeletal vibrations are in general exited at temperatures far below the characteristic temperatures (θ values) for the group vibrations. The ATHAS represents the skeletal vibrations by a Tarasov function [18], where Zhang et al. [13] and Di Lorenzo et al. [14] show that more than 83% of the skeletal modes

are exited at 260 K. It can be shown that all heat capcity functions normally used for vibration spectra of solids (Einstein modes, one-, two-, three-dimensional Debye functions) approach $N_i R$ at high temperatures [15], where N_i is the number of vibrators. As a simplification, we thus represent the heat capacity at constant volume due to skeletal vibrational modes by a constant

$$C_{\mathrm{V}}^{(\mathrm{s})} = N_{\mathrm{s}} R. \tag{2}$$

The experimental heat capacity data of Zhang et al. [13] and Di Lorenzo et al. [14] are given at constant pressure, i.e., C_{p}, while both Eqs. (1) and (2) are given at constant volume (C_{V}). To convert C_{V} to C_{p}, and *vice versa*, one may use the modified Nernst-Lindemann approximation that is proven applicable for polymers [19]

$$C_{\mathrm{p}} - C_{\mathrm{V}} = 3RA_0 C_{\mathrm{p}} \frac{T}{T_{\mathrm{m}}}, \tag{3}$$

where $A_0 = 3.9 \cdot 10^{-3}$ (K mol)/J, and T_{m} is an estimated equilibrium melting temperature that is set to 573 K for all proteins in this work.

In addition to the vibration modes discussed here, at physiological temperatures some larger amplitude modes (rotamers) become more and more important. In this work, which involves a fitting of a model to experimental data, we incorporate these modes in the group modes and the skeletal modes, already discussed above. Thus, as a simplification we will represent the heat capacity of solid state proteins at constant pressure C_{p} as a sum of the group heat capacity [Eq. (1)] and the skeletal heat capacity [Eq. (2)]. When we correct for the Nernst-Lindemann approach [Eq. (3)], the heat capacity at constant pressure for one mole proteins in the solid state yields

$$\begin{aligned} C_{\mathrm{p}} &= \left(1 - 3RA_0 \frac{T}{T_{\mathrm{m}}}\right)^{-1} \left(C_{\mathrm{V}}^{(\mathrm{g})} + C_{\mathrm{V}}^{(\mathrm{s})}\right) \\ &= R \left(1 - 3RA_0 \frac{T}{T_{\mathrm{m}}}\right)^{-1} \left[\frac{N_{\mathrm{g}} \left(\theta/T\right)^2 e^{\theta/T}}{\left(e^{\theta/T} - 1\right)^2} + N_{\mathrm{s}}\right]. \end{aligned} \tag{4}$$

The strength of the model as stated in Eq. (4) is that it contains only three parameters that have to be determined in the fitting procedure for a given temperature T, namely N_{g}, N_{s}, and θ. Hopefully, by ignoring details in the vibrational spectra, as we have done, Eq. (4) can upon fitting to experimental data say something about the key features and generality of intramolecular interactions in proteins.

3 Results and discussion

We fit Eq. (4) to the experimental data of heat capacity for six different proteins in their solid state by a least squares error procedure. The experimental data

Table 1: Best fit parameters, according to Eq. (4), for the heat capacity of proteins in their solid state. M are molecular weights and rms are the relative root mean square deviations between the experimental data and the fitted data. Parameters $N_{\rm g}$ and θ, and $N_{\rm s}$ are explained in connection with Eqs. (1) and (2), respectively.

protein	M (kDalton)	$N_{\rm g}$ (10^3 #)	$N_{\rm s}$ (10^3 #)	θ (10^3 K)	rms (%)
Lysozyme	14.3	5.9	1.7	2.0	0.3
Myoglobin	17.0	5.7	2.0	1.8	0.3
Chymotrypsinogen	25.6	6.8	3.2	1.7	1.0
Lactoglobulin	18.4	5.8	2.4	1.8	0.3
Ovalbumine	42.8	13.0	5.0	1.8	0.6
Ribonuclease A	13.7	3.8	1.6	1.8	0.6

on lysozyme (chicken) and myoglobin (horse) are from Di Lorenzo et al. [14], and the data on $\alpha-$chymotrypsinogen (bovine), $\beta-$lactoglobulin (bovine milk), ovalbumine (chicken), and ribonuclease A (bovine pancreas) from Zhang et al. [13]. The optimal fitted parameters, according to Eq. (4), are listed in Table 1.

All the experimental data we use in this work are measured by a Perkin-Elmer DSC-7 (differential scanning calorimeter). Unfortunately, the authors who performed these experiments do not give an error estimate of their data points, but the standard deviations of the experiments of Zhang et al. [13] on chymotrypsinogen, lactoglobulin, ovalbumine, and ribonuclease A are 2.5% or less. However, this estimate does not, e.g., incorporate systematic errors. It is also up to discussion whether the heat capacity data contain contributions from denaturation of the proteins [13].

In Figure 1 we plot the experimental data together with the optimal fittings. In light of the simplicity of the model, the agreements between the experimental data and the fit are very good, with a relative root mean square deviation less than 1% between the experimental data and the calculated data based upon Eq. (4). This deviation between theory and experiment is better than the more detailed modeling of Zhang et al. [13] and Di Lorenzo et al. [14], who obtained a root mean square error around 3%. However, it should be noted that they studied the proteins in a broader temperature range (130 to 420 K) compared to us. The mean value of the characteristic temperature is $\langle\theta\rangle = (1.8 \pm 0.1) \cdot 10^3$ K (see Table 1), thus in light of the model the experimental data exhibit some generality with regard to θ.

In Figure 2, based upon the fittings listed in Table 1, we plot $N_{\rm g}$ and $N_{\rm s}$ [see Eqs. (1) and (2), respectively] versus the molecular weight (M). These two sets are fitted to two linear functions $\quad N_i = \phi_i \, M \quad$ for $\quad i = \{1, 2\}$, respectively. We obtain $\phi_1 = 0.31$ (Dalton)$^{-1}$ (relative root mean square error is 13%), correspond-

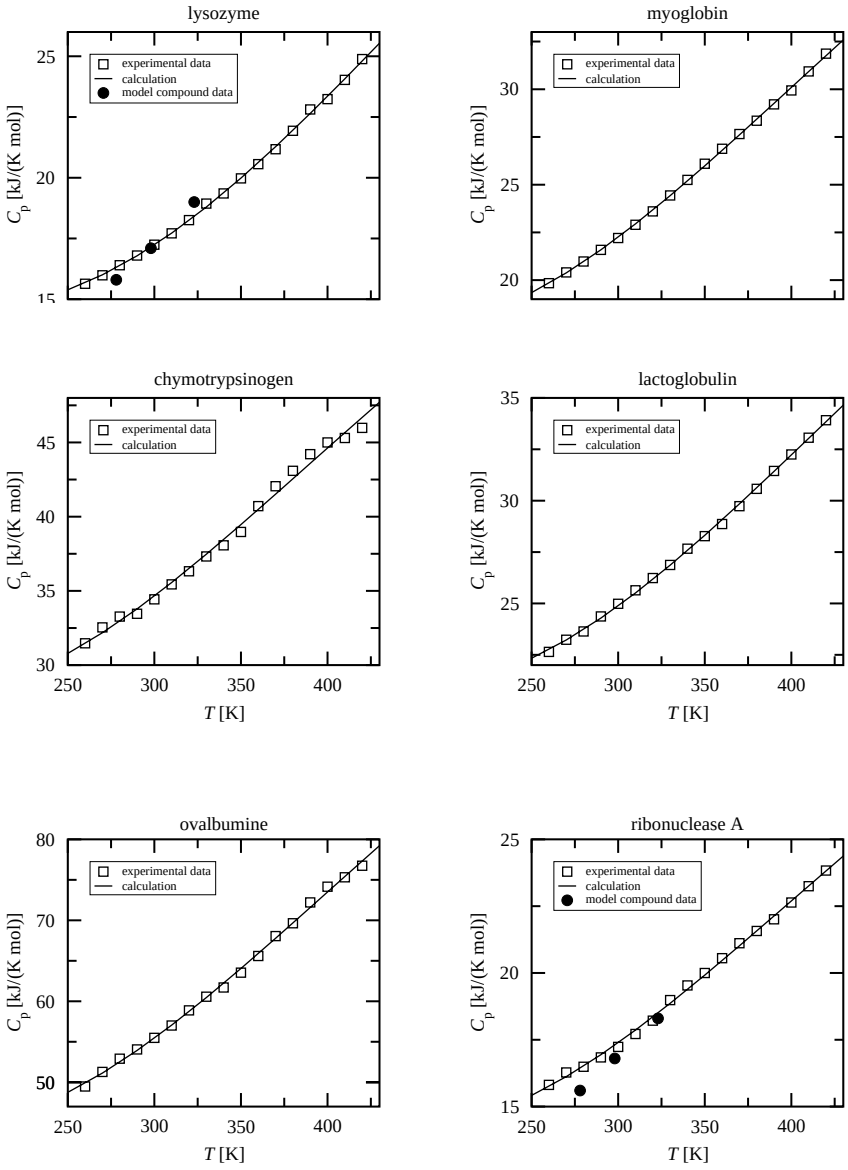

Figure 1: Heat capacity (C_p) per mole proteins in their solid state vs. temperature (T). Experimental data (□) for lysozyme and myoglobin are obtained from Di Lorenzo et al. [14] and experimental data for chymotrypsinogen, lactoglobulin, ovalbumine, and ribonuclease A are obtained from Zhang et al. [13]. Continuous lines (——) are best fits from the theoretical estimate in Eq. (4). Parameters are accordingly listed in Table 1. For lysozyme and ribonuclease A we have also inserted three data points (•), respectively. These latter data points represent an estimate of internal interactions based upon the difference between experimental protein data in solution and a model compound data evaluation of the hydration effect, both evaluated by Privalov and Makhatadze [6].

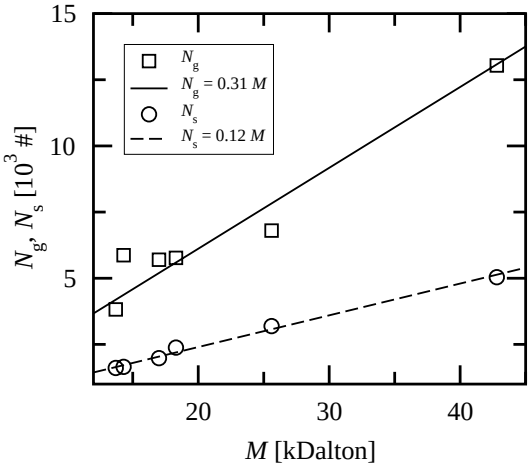

Figure 2: The number of oscillators N_g and N_s vs. the molecular weight M (see Table 1) for the six different proteins considered in this work. Both sets, corresponding to N_g and N_s, are fitted to a linear function, respectively.

ing to N_g versus M data, and $\phi_2 = 0.12$ (Dalton)$^{-1}$ (relative root mean square error is 4%), corresponding to N_s versus M data. This shows that the number of vibrators, to first approximation, increases linearly vs. increasing molecular weight. In particular, N_s has a small deviation from a linear dependence of the parameter M (4%). The reason for this may be that the N_s parameters correspond to the skeletal vibrations of the proteins. The ATHAS shows that the skeletal vibration modes are nearly excited at temperatures above 260 K [13, 14], and indicates that the skeletal vibrations correspond to long wavelengths. Thus, the skeletal vibration modes, and consequently the N_s parameter in our model should not be much affected by the molecular structure of the macromolecules. Consequently, this may explain the very near linear dependence of the parameter N_s with respect to the size of the protein, i.e., the molecular mass M.

If our present model, expressed by Eq. (4), for heat capacity of solid state proteins should become useful for proteins in solution, one may ask – what is the correspondence between thermodynamical data of proteins in solid state and proteins in solution? Protein stability depends upon internal interactions and water interactions. Up to now, there has been no direct way to experimentally separate these contributions from each other. However, Privalov and Makhatadze [6] show that it is possible to separate these contributions by use of model compound data, that are based upon transfer characteristics for the solvating process in water of more than 100 low molecular weight organic compounds [5]. The latter data are in accordance with the Ben-Naim definition of the solvation process of a molecule [20], i.e., transferring the molecule from a fixed position in the ideal gas phase into a fixed position in water, which only consider effects associated with insertion of the solute molecule into water. Thus, in the solvation process effects

associated with differences in translational motions of the molecules in the gas phase and in the water soluted phase are not included. Based upon the solvation data from these small organic substances and assuming that the hydration data of a given protein can be represented as a sum of these (known) low-molecular weight contributions to the heat capacity, Privalov and Makhatadze [6] evaluated the hydration contribution to the heat capacity in the folded (native) state for four different proteins, including lysozyme and ribonuclease A studied in this work. If we now subtract the hydration heat capacity, evaluated from model compounds, from the total experimental heat capacity data of proteins in solution, including both hydration effects and internal interactions, an estimate of the heat capacity of the internal interactions (HCII) is then left.

For two of the proteins discussed in this work, lysozyme and ribonuclease A, we plot three data points (corresponding to HCII as described in the previous paragraph) evaluated by Privalov and Makhatadze [6], respectively. We see that heat capacity in the solid state (HCSS) and the HCII agrees well.

The observation in the previous paragraph is interesting, however several aspects have to be taken into account. First of all, it is not obvious that the HCSS data and the HCII data should have a simple connection. One source of discrepancy between HCSS and HCII is that HCII includes possible changes in the internal interactions upon the solvation process. Here we assume that this contribution to thermodynamics is negligible, but this point should be further investigated. However, a more serious problem may be that HCSS includes intermolecular interactions in contrast to HCII. Based upon the ATHAS calculations of Zhang et al. [13] and Di Lorenzo et al. [14] it is possible to estimate the heat capacity contribution related to the skeletal vibrations. The skeletal vibrations are in these works described by a Tarasov function [18] where the lower frequency part is a three-dimensional Debye function, corresponding to intermolecular modes. The characteristic temperature of this function is typically around 100 K, thus the intermolecular modes are fully excited at physiological temperatures and consequently easy to calculate at these temperatures.

If we subtract the intermolecular part of HCSS, 1.8 kJ/mol for lysozyme and 1.5 kJ/mol for ribonuclease A, the HCII data become larger than the HCSS data with increasing difference vs. temperature. At 323 K the HCII data are 15% and 8% larger than the HCSS for lysosyme and ribonuclease, respectively. The reason for this difference may be that proteins in water have more excited rotamers at a given temperature compared to dehydrated crystallized proteins.

4 Summary and conclusion

We study the heat capacity in the solid state of the proteins lysozyme, myoglobin, chymotrypsinogen, lactoglobulin, ovalbumine, and ribonuclease A. We show, in the temperature region from 260 K to 420 K, that the heat capacity can be well

represented as a sum of one Einstein mode [Eq. (1)] and a constant [Eq. (2)]. This corresponds to the group contributions and the skeletal contributions to the total heat capacity in the ATHAS scheme, respectively [12]. Despite the simplicity of the resulting analytical model [Eq. (4)], the fits of the model to the experimental data are good, with a relative root mean square deviation less than 1% between experiment and theory.

We show that the experimental heat capacity data for two of the proteins in the solid state agree well with model compound data from Privalov and Makhatadze [6] of the same proteins. However, when the experimental data are corrected for intermolecular vibrational modes the model compound data are larger than the experimental ones for the heat capacity (up to 15%). This suggests that proteins in water may have more rotamers excited than crystallized proteins at a given temperature.

This simple description of proteins in solid state may serve as a "baseline" [14, 18] for a full thermodynamic characterization of proteins. By applying this description of vibrational heat capacity of proteins [Eq. (4)] together with a model that describes the hydration contribution to the energetics explicitly [21, 22], it may be possible to use this accumulated knowledge to predict thermodynamical stability of proteins in solution, and in particular to predict cold and warm destabilization of small globular proteins [23–27].

Acknowledgments

A.B. thanks the Research Council of Norway for financial support (Contract No. 129619/410).

References

[1] C.B. Anfinsen, Principles that govern the folding of protein chains, Science 181 (1973) 223-230.

[2] P.L. Privalov, Stability of proteins. Small globular proteins, Adv. Protein Chem. 33 (1979) 167-241.

[3] G.I. Makhatadze, P.L. Privalov, Energetics of protein structure, Adv. Protein Chem. 47 (1995) 307-425.

[4] G.I. Makhatadze, Heat capacities of amino acids, peptides and proteins, Biophys. Chem. 71 (1998) 133-156.

[5] G.I. Makhatadze, P.L. Privalov, Heat capacity of proteins. I. Partial molar heat capacity of individual amino acid residues in aqueous solution: hydration effect, J. Mol. Biol. 213 (1990) 375-384.

[6] P.L. Privalov, G.I. Makhatadze, Contribution of hydration and non-covalent interactions to the heat capacity effect on protein unfolding, J. Mol. Biol. 224 (1992) 715-723.

[7] G.I. Makhatadze, P.L. Privalov, Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration, J. Mol. Biol. 232 (1993) 639-659.

[8] P.L. Privalov, G.I. Makhatadze, Contribution of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration, J. Mol. Biol. 232 (1993) 660-679.

[9] J.O. Hutchens, A.G. Cole, J.W. Stout, Heat capacities from 11 to 305° K and entropies of hydrated and anhydrous bovine zinc insulin and bovine chymotrypsinogen A, J. Biol. Chem. 244 (1969) 26-32.

[10] A.R. Haly, J.W. Snaith, Calorimetry of rat tail tendon collagen before and after denaturation: the heat of fusion of its absorbed water, Biopolymers 10 (1971) 1681-1699.

[11] E.L. Andronikashvili, G.M. Mrevlishvili, G.Sh. Japaridze, V.M. Sokhadze, K.A. Kvavadze, Thermal properties of collagen in helical and random coiled states in the temperature range from 4° to 300°K, Biopolymers 15 (1976) 1991-2004.

[12] B. Wunderlich, The ATHAS database on heat capacities of polymers, Pure Appl. Chem. 67 (1995) 1019-1026.

[13] Ge Zhang, S. Gerdes, B. Wunderlich, Heat capacity of solid, globular proteins, Macromol. Chem. Phys. 197 (1996) 3791-3806.

[14] M.L. Di Lorenzo, G. Zhang, M. Pyda, B.V. Lebedev, B. Wunderlich, Heat capacity of solid-state biopolymers by thermal analysis, J. Polym. Sci. Pol. Phys. 37 (1999) 2093-2102.

[15] B. Wunderlich, H. Baur, Heat capacities of linear high polymers, Adv. Polymer Sci. 7 (1970) 151-368.

[16] B. Wunderlich, Thermal Analysis, Academic Press, London, UK, 1990 (Chapter 5).

[17] K.A. Roles, B. Wunderlich, Heat-capacities of solid copoly (aminoacid)s, J. Polym. Sci. Pol. Phys. 31 (1993) 279-285.

[18] Ge Zhang, B. Wunderlich, A new method of fitting approximate vibrational spectra to heat capacities of solids with Tarasov functions, J. Therm. Anal. 47 (1996) 899-911.

[19] R. Pan, M.V. Nair, B. Wunderlich, On the C_p to C_v conversion of solid linear macromolecules II, J. Therm. Anal. 35 (1989) 955-966.

[20] A. Ben-Naim, Solvation Thermodynamics, Plenum Press, New York, 1987.

[21] A. Bakk, J.S. Høye, A. Hansen, Specific heat upon aqueous unfolding of the protein interior: a theoretical approach, Physica A 304 (2002) 355-361.

[22] A. Bakk, J.S. Høye, A. Hansen, Apolar and polar solvation thermodynamics related to the protein unfolding process, Biophys. J. 82 (2002) 713-719.

[23] P.L. Privalov, Yu.V. Griko, S.Yu. Venyaminov, V.P. Kutyshenko, Cold denaturation of myoglobin, J. Mol. Biol. 190 (1986) 487-498.

[24] Yu.V. Griko, P.L. Privalov, S.Yu. Venyaminov, V.P. Kutyshenko, Thermodynamic study of the apomyoglobin structure, J. Mol. Biol. 202 (1988) 127-138.

[25] B.-lu Chen, J.A. Schellman, Low-temperature unfolding of a mutant of phage T4 lysozyme. 1. Equilibrium studies, Biochemsitry USA 28 (1989) 685-691.

[26] F. Franks, Protein destabilization at low temperatures, Adv. Protein Chem. 46 (1995) 105-139.

[27] A. Bakk, J.S. Høye, A. Hansen, K. Sneppen, M.H. Jensen, Pathways in two-state protein folding, Biophys. J. 79 (2000) 2722-2727.