Eur. Phys. J. D (2014)
DOI: 10.1140/epjd/e2014-50097-3

THE EUROPEAN
PHYSICAL JOURNAL D

Regular Article

Quantitative thermodynamic model for globular protein folding[⋆]

Alexander V. Yakubovich[a] and Andrey V. Solov'yov

MBN Research Center, Altenhöferallee 3, 60438 Frankfurt am Main, Germany

Received 2 February 2014 / Received in final form 27 March 2014
Published online (Inserted Later) – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2014

Abstract. We present a statistical mechanics formalism for theoretical description of the process of protein folding ↔ unfolding transition in water environment. The formalism is based on the construction of the partition function of a protein obeying two-stage-like folding kinetics. Using the statistical mechanics model of solvation of hydrophobic hydrocarbons we obtain the partition function of infinitely diluted solution of proteins in water environment. The calculated dependencies of the protein heat capacities upon temperature are compared with the corresponding results of experimental measurements for staphylococcal nuclease and metmyoglobin.

1 Introduction

Proteins are biological polymers consisting of elementary structural units, amino acids. Being synthesized at ribosome, proteins are exposed to the cell interior where they fold into their unique three dimensional structure. The process of formation of proteins three dimensional structure is called the process of protein folding. The correct folding of proteins is of crucial importance for their proper functioning.

In a course of tissue irradiation by swift ions emerges a cascade of diverse physical and chemical processes, which span over various temporal and spatial scales. These processes leading to biological tissue damage are utilized for the purposes of ion beam cancer therapy (IBCT) [1,2]. Recently it was shown that propagation of ions in the medium with high linear energy transfer leads to dramatic increase of the temperature in the vicinity of ion's trajectory, see reference [3] and references therein. Rapid increase of medium temperature can lead to direct breakage of chemical bonds in biological molecules, but also influence the conformational structure of proteins and DNA. Numerous works are devoted to the description of conformational changes in proteins resulting from temperature variation, i.e. temperature-induced protein folding and unfolding transitions. The current state-of-the-art in experimental and theoretical studies of the protein folding process are described in recent reviews, see references [4–8] and references therein.

In this paper we present a theoretical method for the description of the protein folding process which is based on the principles of statistical mechanics. Considering the process of protein folding as a first order phase transition in a finite system, we present a statistical mechanics model for treating the folding ↔ unfolding phase transition in single-domain proteins. Propagation of swift ion's in the biological medium is accompanied by increase of temperature and ion's concentrations in the vicinity of the trajectory. Therefore, the goal of this study is to reproduce with minimal number of assumptions thermodynamic behaviour of real particular proteins under the variation of external conditions, namely temperature and pH of the solvent. For the time being we do not account for any kinetic effects related to the temperature spikes in the vicinity of ions trajectories. The extension of the model towards the description of time-dependent processes can be developed in later works. The suggested method is based on the theory developed for the helix ↔ coil transition in polypeptides discussed in references [9–20] and applied for folding ↔ unfolding phase transition in single-domain proteins.

Many papers devoted to the description of thermodynamics of the protein folding process have been published since eighties of the previous century, for a comprehensive review see reference [21]. Here we do not intend to review all of them but rather refer to the most essential papers published in the field that are most closely related to the topic of our research.

A way to construct a parameter-free partition function for a system experiencing α-helix ↔ random coil phase transition *in vacuo* was studied in reference [9]. In reference [11] we have calculated potential energy surfaces (PES) of polyalanines of different lengths with respect to their twisting degrees of freedom. This was done within the framework of classical molecular mechanics. The calculated PES were then used to construct a parameter – free partition function of a polypeptide and to derive various

[⋆] Contribution to the Topical Issue "Nano-scale Insights into Ion-beam Cancer Therapy", edited by Andrey V. Solov'yov, Nigel Mason, Paulo Limão-Vieira and Malgorzata Smialek-Telega.

[a] e-mail: sasha@mbnexplorer.com

thermodynamical characteristics of alanine polypeptides as a function of temperature and polypeptide length.

We start our approach with the construction of the partition function of a protein *in vacuo*, which is the further generalization of the formalism developed in reference [12], accounting for folded, unfolded and prefolded states of the protein. Our model is based on a number of assumptions about the system. Most of the assumptions are necessary for the factorization of the partition function of the system. In principle, the factorization of the partition function implies the statistical independence of the corresponding subsystems, e.g. protein and water. In many cases it is a difficult task to estimate analytically the accuracy of a particular assumption allowing the partition function factorization. However, the most important assumptions that are used in our work have been already intensively discussed and thoughtfully analyzed in previous papers by Privalov and Makhatadze [22,23], Flory [24], Murphy and Freire [25], Go [26], Go and Abe [27], Nemethy and Scheraga [28], Lewis et al. [29], Kim and Baldwin [30] and Baldwin [31]. So in our work where appropriate we refer and rely on the results of these earlier investigations.

In references [32,33] was proposed a way to calculate the thermodynamic characteristics of the flexible molecules based on the construction of marginal probability density functions. The approach developed in that work utilizes the recursive application of the generalized Kirkwood superposition approximation. In references [32,33] it was shown that molecular fluctuations can be described to good approximation accounting only for low order correlations in the system.

The analysis of 76 Ising-like models for protein folding was performed in reference [34]. These models are based on different assumptions about the system. The relative performance of each assumption was evaluated using the rank sum statistics. The performed analysis revealed that simple models which only consider the trade-off between the loss of conformational entropy ans stabilization from native inter-residue contacts, are a surprisingly accurate predictor of two-stateness and relative folding rates of small single-domain proteins.

For the correct description of the protein folding in water environment it is of primary importance to consider the interactions between the protein and the solvent molecules. The hydrophobic interactions are known to be the most important driving forces of protein folding [35].

An extensive review of the hydrophobic effects in molecular solutions is presented in references [36,37]. The investigation of temperature dependence of the hydrophobic interaction in protein folding was performed in references [31,38]. In reference [39] it was shown that the denaturational heat capacity is composed of a large positive contribution arising from the exposure of apolar groups to the solvent and a significant negative contribution originating from the exposure of polar groups. The β propensities of various amino acids and their influence on the protein stability was studied in reference [40]. A rather simple but efficient model to calculate the intermolecular energy contribution to the protein stability was suggested in reference [41]. The driving forces for the protein dynamic and the kinetic cooperativity were investigated in detail for the well known protein-trypsin inhibitor by Kaya and Chan in reference [42].

In the present work we present a method allowing one to construct the partition function of a protein that accounts for the protein interaction with solvent, i.e. accounts for the hydrophobic effect.

We treat the hydrophobic interactions in the system using the statistical mechanics formalism developed in reference [43] for the description of the thermodynamical properties of the solvation process of aliphatic and aromatic hydrocarbons in water. The water molecules only form the protein's first solvation shell are considered to be interacting with the protein hydrophobic surface. The role of the water solvations shells on the rate of protein folding is discussed in reference [44]. However, accounting solely for hydrophobic interactions is not sufficient for the proper description of the energetics of all conformational states of the protein and one has to take electrostatic interactions into account. In the present work the electrostatic interactions are treated within a similar framework as described in reference [45].

All the aforementioned works include free parameters that are fitted to reproduce well experimental or molecular dynamics simulations. In the present paper we present a simple but quantitative model for the description of the protein's thermodynamic properties and focus on the fundamental physical effects that govern the protein folding process. To the best of our knowledge there is no such statistical mechanics approach, which can be applied for real proteins and reproduces well their heat capacity curves under various values of pH with only few parameters having clear physical meaning, which cannot be derived analytically.

We have applied the developed statistical mechanics model of protein folding for two globular proteins, namely staphylococcal nuclease and metmyoglobin. These proteins have simple two-stage-like folding kinetics and demonstrate two folding ↔ unfolding transitions, referred as heat and cold denaturation, see references [46,47]. The comparison of the results of the theoretical model with that of the experimental measurements shows the applicability of the suggested formalism for an accurate description of various thermodynamical characteristics in the system, e.g. heat denaturation, cold denaturation, increase of the reminiscent heat capacity of the unfolded protein, etc.

Our paper is organized as follows. In Section 2.1 we present the formalism for the construction of the partition function of the protein in water environment and justify the assumptions made on the system's properties. In Section 3 we discuss the results obtained with our model for the description of folding ↔ unfolding transition in staphylococcal nuclease and metmyoglobin. In Section 4 we summarize the paper and suggest several ways for a further development of the theoretical formalism.

2 Theoretical methods

2.1 Partition function of a protein

To study thermodynamic properties of the system one needs to investigate its potential energy surface with respect to all the degrees of freedom. For the description of macromolecular systems, such as proteins, efficient model approaches are necessary.

The most relevant degrees of freedom in the protein folding process are the twisting degrees of freedom along its backbone chain [9,12]. The degrees of freedom of a protein can be classified as stiff and soft ones. We call the degrees of freedom corresponding to the variation of bond lengths, angles and improper dihedral angles as stiff, while degrees of freedom corresponding to the angles φ_i and ψ_i are soft degrees of freedom [9]. The stiff degrees of freedom can be treated within the harmonic approximation, because the energies needed for a noticeable structural rearrangement with respect to these degrees of freedom are about several eV, which is significantly larger than the characteristic thermal energy of the system (kT), being at room temperature equal to 0.026 eV [16–18,48–50].

A Hamiltonian of a protein is constructed as a sum of the potential, kinetic and vibrational energy terms. Assuming the harmonic approximation for the stiff degrees of freedom it is possible to derive the following expression for the partition function of a protein in vacuo being in a particular conformational state j [9,51]:

$$Z_j = A_j (kT)^{3N-3-\frac{l_s}{2}} \times \int_{\varphi\in\Gamma_j} \ldots \int_{\psi\in\Gamma_j} e^{-\epsilon_j(\{\varphi,\psi\})/kT} d\varphi_1 \ldots d\varphi_n d\psi_1 \ldots d\psi_n, \quad (1)$$

where T is the temperature, k is the Boltzmann constant, N is the total number of atoms in the protein, l_s is the number of soft degrees of freedom, A_j is defined as follows:

$$A_j = \left[\frac{V_j M^{3/2} \sqrt{I_j^{(1)} I_j^{(2)} I_j^{(3)}} \prod_{i=1}^{l_s} \sqrt{\mu_i^{s(j)}}}{(2\pi)^{\frac{l_s}{2}} \pi \hbar^{3N} \prod_{i=1}^{3N-6-l_s} \omega_i^{(j)}} \right]. \quad (2)$$

A_j is a factor which depends on the mass of the protein M, its three main momenta of inertia $I_j^{(1,2,3)}$, specific volume V_j, the frequencies of the stiff normal vibrational modes $\omega_i^{(j)}$ and on the generalized masses $\mu_i^{s(j)}$ corresponding to the soft degrees of freedom [9]. ϵ_i in equation (1) describes the potential energy of the system corresponding to the variation of soft degrees of freedom. Integration in equation (1) is performed over a certain part of a phase space of the system (a subspace Γ_j) corresponding to the soft degrees of freedom φ and ψ. The form of the partition function in equation (1) allows one to avoid the multidimensional integration over the whole coordinate space and to reduce the integration only to the relevant parts of the phase space. ϵ_j in equation (1) denotes the potential energy surface of the protein as a function of twisting degrees of freedom in the vicinity of protein's conformational state j. Note that in general the proper choice of all the relevant conformations of protein and the corresponding set of Γ_j is not a trivial task.

One can expect that the factors A_j in equation (1) depend on the chosen conformation of the protein. However, due to the fact that the values of specific volumes, momenta of inertia and frequencies of normal vibration modes of the system in different conformations are expected to be close [12,52], the values of A_j in all conformations become nearly equal, at least in the zero order harmonic approximation, i.e. $A_j \equiv A$. Another simplification of the integration in equation (1) comes from the statistical independence of amino acids. We assume that within each conformational state j all amino acids can be treated statistically independently, i.e. the particular conformational state of ith amino acid characterized by angles $\varphi_i \in \Gamma_j$ and $\psi_i \in \Gamma_j$ does not influence the potential energy surface of all other amino acids, and vice versa. This assumption is known as Flory isolated pair hypothesis [24]. Despite the fact that isolated pair hypothesis is not always a good approximation (see e.g. [53]) it is still quite reasonable for the construction of the partition function of the native (i.e. rigid and thus harmonic) conformational state of the protein. Here we refer to the analogy with the Einstein's model for solids treating all the atoms of an ideal solid as statistically independent [54]. This model reproduces quite well the thermodynamic characteristics of solids. The very similar assumptions on the statistical independence of constituting atoms and, consequently, amino acids can be utilized for the description of the thermodynamical characteristics of a structural (native) or any compact tightly bound state of a protein. In unfolded states the flexibility of the protein backbone chain leads to a significant variation of distances between atoms, therefore the interaction between particular atoms changes substantially in different random coil conformations. This fact can lead to a considerable correlation in the motion of amino acids in the protein [53]. An accurate accounting (both analytical and computational) for the interaction between distant atoms in unfolded state of the protein is extremely difficult (for the analytical approach to the problem see Ref. [55]). In this work we assume that interaction between the distant backbone amino acids can be neglected in the the unfolded protein states. See Appendix and therein the discussion justifying this assumption.

With the above mentioned assumptions the partition function of a protein Z_p (without any solvent) reads as:

$$Z_p = A(kT)^{3N-3-\frac{l_s}{2}} \times \sum_{j=1}^{\xi} \prod_{i=1}^{a} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \exp\left(-\frac{\epsilon_i^{(j)}(\varphi_i, \psi_i)}{kT} \right) d\varphi_i d\psi_i, \quad (3)$$

where the summation over j includes all ξ statistically relevant conformations of the protein, a is the number of amino acids in the protein and $\epsilon_i^{(j)}$ is the potential energy surface as a function of twisting degrees of freedom φ_i and ψ_i of the ith amino acid in the jth conformational state of the protein. The exact construction of $\epsilon_a^{(j)}(\varphi_i, \psi_i)$ for various conformational states of a particular protein will be discussed below. We consider the angles φ and ψ as the only two soft degrees of freedom in each amino acid of the protein, and therefore the total number of soft degrees of freedom of the protein $l_s = 2a$.

Partition function in equation (3) can be further simplified if one assumes (i) that each amino acid in the protein can exist only in two conformations: the native state conformation and the random coil conformation; (ii) the potential energy surfaces for all the amino acids are identical. This assumption is applicable for both the native and the random coil state. It is not very accurate for the description of thermodynamical properties of single amino acids, but is reasonable for the treatment of thermodynamical properties of the entire protein. The judgment of the quality of this assumption could be made on the basis of comparison of the results obtained with its use with experimental data. Such comparison is performed in Section 3 of this work.

Amino acids in a protein being in its native state vibrate in a steep harmonic potential. Here we assume that the potential energy profile of an amino acid in the native conformation should not be very sensitive to the type of amino acid and thus can be taken as, e.g., the potential energy surface for an alanine amino acid in the α-helix conformation [11]. Using the same arguments the potential energy profile for an amino acid in unfolded protein state can be approximated by e.g. the potential of alanine in the unfolded state of alanine polypeptide (see Ref. [11] for discussion and analysis of alanine's potential energy surfaces). Indeed, for an unfolded state of a protein it is reasonable to expect that once neglecting the long-range interactions all the differences in the potential energy surfaces of various amino acids arise from the steric overlap of the amino acids's side chains. This is clearly seen on alanine's potential energy surface at values of $\varphi > 0°$ presented in reference [11]. But the part of the potential energy surface at $\varphi > 0°$ gives a minor contribution to the entropy of amino acid at room temperature. This fact allows one to neglect all the differences in potential energy surfaces for different amino acids in an unfolded protein, at least in the zero order approximation. This assumption should be especially justified for proteins with the rigid helix-rich native structure. The staphylococcal nuclease, which we study here has definitely high α-helix content. Another argument which allows to justify our assumption for a wider family of proteins is the rigidity of the protein's native structure. Below, we validate the assumptions made by performing the comparison of the results of our theoretical model with the experimental data for α/β rich protein metmyoglobin obtained in reference [47].

For the description of the folding ↔ unfolding transition in small globular proteins obeying simple two-state-like folding kinetics we assume that the protein can exist in one of three states: completely folded state, completely unfolded state and partially folded state where some amino acids from the flexible regions with no prominent secondary structure are in the unfolded state, while other amino acids are in the folded conformation. With this assumption the partition function of the protein reads as:

$$Z_p = Z_0 + \sum_{i=a-\kappa}^{a} \frac{\kappa!}{(i-(a-\kappa))!(a-i)!} Z_i, \tag{4}$$

where Z_i is defined in equation (1), Z_0 is the partition function of the protein in completely unfolded state, a is the total number of amino acids in a protein and κ is the number of amino acids in flexible regions. The factorial term in equation (4) accounts for the states in which various amino acids from flexible regions independently attain the native conformation. The summation in equation (4) is performed over all partially folded states of the protein, where $a - \kappa$ is the minimal possible number of amino acids being in the folded state. The factorial term describes the number of ways to select $i - (a - \kappa)$ amino acids from the flexible region of the protein consisting of κ amino acids attaining native-like conformation.

Finally, the partition function of the protein in vacuo has the following form:

$$Z_p = \tilde{Z}_p A (kT)^{3N-3-a}, \tag{5}$$

where

$$\tilde{Z}_p = Z_u^a + \sum_{i=a-\kappa}^{a} \frac{\kappa! Z_b^i Z_u^{a-i} \exp\left(iE_0/kT\right)}{(i-(a-\kappa))!(a-i)!} \tag{6}$$

$$Z_b = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \exp\left(-\frac{\epsilon_b(\varphi,\psi)}{kT}\right) d\varphi d\psi \tag{7}$$

$$Z_u = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \exp\left(-\frac{\epsilon_u(\varphi,\psi)}{kT}\right) d\varphi d\psi. \tag{8}$$

Here we omitted the trivial factor describing the motion of the protein center of mass, which is of no significance for the problem considered, $\epsilon_b(\varphi, \psi)$ (b stands for *bound*) is the potential energy surface of an amino acid in the native conformation and $\epsilon_u(\varphi, \psi)$ (u stands for *unbound*) is the potential energy surface of an amino acid in the random coil conformation. The potential energy profile of an amino acid is calculated as a function of its twisting degrees of freedom φ and ψ. Let us denote by ϵ_b^0 and ϵ_u^0 the global minima on the potential energy surfaces of an amino acid in folded and in unfolded conformations, respectively. The potential energy of an amino acid then reads as $\epsilon_{u,b}^0 + \epsilon_{u,b}(\varphi, \psi)$. E_0 in equation (6) is defined as the energy difference between the global energy minima of the amino acid potential energy surfaces corresponding to the folded and unfolded conformations, i.e. $E_0 = \epsilon_u^0 - \epsilon_b^0$. The potential energy surfaces for amino acids as functions of angles φ and ψ were calculated and thoroughly analyzed in reference [11].

In nature proteins perform their function in the aqueous environment. Therefore the correct theoretical description of the folding ↔ unfolding transition in water environment should account for solvent effects.

2.2 Partition function of a protein in water environment

In this section we evaluate E_0 and construct the partition function for the protein in water environment.

The partition function of the infinitely diluted solution of proteins Z can be constructed as follows:

$$Z = \sum_{j=1}^{\xi} \tilde{Z}_p^{(j)} Z_W^{(j)}, \tag{9}$$

where $Z_W^{(j)}$ is the partition function of all water molecules in the jth conformational state of a protein and $\tilde{Z}_p^{(j)}$ is the partition function of the protein in its jth conformational state, in which we further omit the factor describing the contribution of stiff degrees of freedom in the system. This is done in order to simplify the expressions, because stiff degrees of freedom provide a constant contribution to the heat capacity of the system since the heat capacity of the ensemble of harmonic oscillators is constant. Below for the simplicity of notations we put $\tilde{Z}_p \equiv Z_p$.

There are two types of water molecules in the system: (i) molecules in pure water and (ii) molecules interacting with the protein. We assume that only the water molecules being in the vicinity of the protein's surface are involved in the folding ↔ unfolding transition, because they are affected by the variation of the hydrophobic surface of a protein. This surface is equal to the protein's solvent accessible surface area (SASA) of the hydrophobic amino acids. The number of interacting molecules is proportional to SASA and include only the molecules from the first protein's solvation shell. This area depends on the conformation of the protein. The main contribution to the energy of the system caused by the variation of the protein's SASA associated with the side-chains of amino acids because the contribution to the free energy assosiated with solvation of protein's backbone is small [56]. Thus, in this work we pay the main attention to the accounting for the SASA change arising due to the solvation of side chains.

We treat all water molecules as statistically independent, i.e. the energy spectra of the states of a given molecule and its vibrational frequencies do not depend on a particular state of all other water molecules. Thus, the partition function of the whole system Z can be factorized and reads as:

$$Z = \sum_{j=1}^{\xi} Z_p^{(j)} Z_s^{Y_c(j)} Z_w^{N_t - Y_c(j)}, \tag{10}$$

where ξ is the total number of states of a protein, Z_s is the partition function of a water molecule affected by the interaction with the protein and Z_w is the partition function of a water molecule in pure water. $Y_c(j)$ is the number of water molecules interacting with the protein in the jth conformational state. N_t is the total number of water molecules in the system. To simplify the expressions we do not account for water molecules that do not interact with the protein in any of its conformational states, i.e. $N_t = \max_j\{Y_c(j)\}$.

To construct the partition function of water we follow the formalism developed in reference [43]. In Appendix we present the most essential details of the construction of the partition function of water molecules and derive the expressions for Z_w and Z_s, being the partition function of water molecules in the pure water and in the vicinity of the solute.

In our theoretical model we also account for the electrostatic interaction of protein's charged groups with water. The presence of electrostatic field around the protein leads to the reorientation of H_2O molecules in the vicinity of charged groups due to the interaction of dipole moments of the molecules with the electrostatic field. The additional factor arising in the partition function of water molecules (see Appendix for details) reads as:

$$Z_E = \left(\frac{1}{4\pi} \int \exp\left(-\frac{Ed\cos\theta}{kT} \right) \sin\theta d\theta d\varphi \right)^{\alpha}, \tag{11}$$

where E is the strength of the electrostatic field, d is the absolute value of the H_2O molecule dipole moment, α is the ratio of the number of water molecules that interact with the electrostatic field of the protein (N_E) to the number of water molecules interacting with the surface of the amino acids from the inner part of the protein while they are exposed to water when the protein is being unfolded (N_w), i.e. $\alpha = N_E/N_w$. Note that the effects of electrostatic interaction turn out to be more pronounced in the folded state of the protein. This happens because in the unfolded state of a protein opposite charges of amino acid's side chains are in average closer in space due to the flexibility of the backbone chain, while in the folded state the positions of the charges are fixed by the rigid structure of a protein.

Integrating equation (11) allows to write the factor Z_E for the partition function of a single H_2O molecule in pure water in the form:

$$Z_E = \left(\frac{kT \sinh\left[\frac{Ed}{kT}\right]}{Ed} \right)^{\alpha}. \tag{12}$$

This equation shows how the electrostatic field enters the partition function. In general, E depends on the position in space with respect to the protein. However, here we neglect this dependence and instead we treat the parameter E as an average, characteristic electrostatic field created by the protein.

Let us denote by N_s the number of water molecules interacting with the proteins surface in its folded state i.e. $N_t = N_s + N_w$; where N_t is defined in equation (10). We assume that the number of water molecules interacting with the protein (Y_c) is linearly dependent on the number of amino acids being in the unfolded conformation, i.e. $Y_c = N_s + iN_w/a$, where i is the number of the amino acids in the unfolded conformation and a is the total number of amino acids in the protein. Thus, the partition function (10) with the accounting for the factor (12) reads as:

$$Z = Z_s^{N_s} \sum_{j=1}^{\xi} \left(Z_b Z_w^{\frac{N_w}{a}} Z_E^{\frac{N_w}{a}} \exp\left(iE_0/kT\right) \right)^{i(j)} \times \left(Z_u Z_s^{\frac{N_w}{a}} \right)^{a-i(j)}, \tag{13}$$

where $i(j)$ denotes the number of the amino acids being in the folded conformation when the protein is in the jth conformational state. Accounting for the statistical factors for amino acids being in the folded and unfolded states, similarly to how it was done for the vacuum case (see Eq. (6)), one derives from equation (13) the following final expression:

$$Z = (Z_s)^{N_s} \left[Z_u^a Z_s^{N_w} + \sum_{i=a-\kappa}^{a} \frac{\kappa! \exp\left(iE_0/kT\right)}{(i-(a-\kappa))!(a-i)!} \times \left(Z_b Z_w^{N_w/a} Z_E^{N_w/a} \right)^i \left(Z_u Z_s^{N_w/a} \right)^{a-i} \right], \tag{14}$$

where the term in the square brackets accounts for all statistically significant conformational states of the protein.

Having constructed the partition function of the system we can evaluate with its use all thermodynamic characteristics of the system. In this work we analyze the dependence of protein's heat capacity on temperature and compare the predictions of our model with available experimental data.

3 Results and discussion

In this section we calculate the dependencies of the heat capacity on temperature for two globular proteins metmyoglobin and staphylococcal nuclease and compare the results with experimental data from [46,47].

The structures of metmyoglobin and staphylococcal nuclease proteins are shown in Figure A.1. These are relatively small globular proteins consisting of ∼150 amino acids. Under certain experimental conditions (salt concentration and pH) the metmyoglobin and the staphylococcal nuclease experience two folding ↔ unfolding transitions, which induce two peaks in the dependency of heat capacity on temperature (see further discussion). The peaks at lower temperature are due to the cold denaturation of the proteins. The peaks at higher temperatures arise due to the ordinary folding ↔ unfolding transition. The availability of experimental data for the heat capacity profiles of the mentioned proteins, the presence of the cold denaturation and simple two-stage-like folding kinetics are the reasons for selecting these particular proteins as case studies for the verification of the developed theoretical model.

3.1 Heat capacity of staphylococcal nuclease

Staphylococcal or micrococcal nuclease (S7 Nuclease) is a relatively nonspecific enzyme that digests single-stranded and double-stranded nucleic acids, but is more active on single-stranded substrates [57]. This protein consists of 149 amino acids. Its structure is shown in Figure A.1.

To calculate the SASA of staphylococcal nuclease in the folded state the 3D structure of the protein was taken from the Protein Data Bank (PDB ID 1EYD). Using CHARMM27 [50] forcefield and NAMD program [58] we performed the structural optimization of the protein and calculated SASA with the solvent probe radius 1.4 Å.

The value of SASA of the side-chains in the folded protein conformation is equal to $S_f = 6858$ Å^2. In order to calculate SASA for an unfolded protein state, the value of all angles φ and ψ were put equal to 180°, corresponding to a fully stretched conformation. Then, the optimization of the structure with the fixed angles φ and ψ was performed. The optimized geometry of the stretched molecule has a minor dependence on the value of dielectric susceptibility of the solvent, therefore the value of dielectric susceptibility was chosen to be equal to 20, in order to mimic the screening of charges by the solvent. SASA of the side-chains in the stretched conformation of the protein is equal to $S_u = 15\,813$ Å^2.

The change of the number of water molecules those interacting with the protein due to the unfolding process can be calculated as follows:

$$N_w = (S_u - S_f)n^{2/3}, \tag{15}$$

where $S_u = 15\,813$ Å^2 and $S_f = 6858$ Å^2 are the SASA of the protein in unfolded and in folded conformations, respectively and n is the density of the water molecules. The volume of one mole of water is equal to 18 cm^3, therefore $n \approx 30$ Å^{-3}

To account for the effects caused by the electrostatic interaction of water molecules with the charged groups of the protein it is necessary to evaluate the strength of the average electrostatic field E in equation (12). The strength of the average field can be estimated as $Ed = kT$, where d is the dipole moment of a water molecule, k is Bolzmann constant and $T = 300$ K is the room temperature. According to this estimate the energy of characteristic electrostatic interaction of water molecules is equal to the thermal energy per degree of freedom of a molecule. In Appendix we present the calculation of the number of water molecules interacting with a single charged group of a protein.

At physiological conditions staphylococcal nuclease has 8 charged residues [59]. The value of α for this protein varies within the interval from 1.29 to 31.27 for $\lambda_d \in [10 \ldots 30]$ Å, where λ_d is the Debye screening length of the charge in electrolyte. In our numerical analysis in equation (14) we have used the characteristic value of α equal to 2.5.

Note that number of molecules interacting with the electrostatic field N_E and the strength of the electrostatic field E should be considered as the *effective* parameters of our model. In this work we do not perform accurate accounting for the spatial dependence of the electrostatic field. Instead, we introduce the parameters α and E that

Table 1. Values of E_0 for staphylococcal nuclease ($E_0^{(S)}$) and metmyoglobin ($E_0^{(M)}$) at different values of solvent pH.

pH value	$E_0^{(S)}$ (kcal/mol)	$E_0^{(M)}$ (kcal/mol)
7.0	0.789	
5.0	0.795	
4.5	0.803	
4.10		1.128
3.88	0.819	
3.84		1.150
3.70		1.165
3.5		1.2
3.23	0.890	

can be interpreted as effective values of the number of H_2O molecules and the strength of the electrostatic field correspondingly. Let us stress that the number of water molecules α and the strength of the field E are not independent parameters of our model because by choosing the higher value of E and smaller value of α or vice versa one can derive the same heat capacity profile.

In this work we do not investigate the dependencies of the heat capacity profiles on the values of the parameters α and E. Below we focus on the investigation of the dependence of the protein heat capacity on the energy E_0 at the fixed value of α and E equal to 2.5 and 0.58 kcal/mol, respectively.

An important parameter of the model is the energy difference between the two states of the protein normalized per one amino acid, E_0 introduced in equation (6). This parameter describes both the energy loss due to the separation of the hydrophobic groups of the protein which attract in the native state of the protein due to Van-der-Waals interaction and the energy gain due to the formation of Van-der-Waals interactions of hydrophobic groups of the protein with H_2O molecules in the protein's unfolded state. Also, the difference of the electrostatic energy of the system in the folded and unfolded states is accounted for in E_0. The difference of the electrostatic energy may depend on various characteristics of the system, such as concentration of ions in the solvent and its pH, on the exact location of the charged sites in the native conformation of the protein and on the probability distribution of distances between charged amino acids in the unfolded state. Thus, exact calculation of E_0 is rather difficult. It is a separate task which we do not intend to address in this work. Instead, in the current study the energy difference between the two phases of the protein is considered as a parameter of the model. We treat E_0 as being dependent on external properties of the system, in particular on the pH value of the solution. In this work the value of E_0 is fitted to reproduce the experimental measurements at different pH values.

Another characteristic of the protein folding ↔ unfolding transition is its cooperativity. In the model it is described by the parameter κ in equation (4). κ describes the number of amino acids in the flexible regions of the protein. The staphylococcal nuclease possesses a prominent two-stage folding kinetics, therefore only 5−10% of amino acids is in the protein's flexible regions. Thus, the value of κ for this protein is small. It can be estimated as being equal to $149 \times 7\% \approx 10$ amino acids.

The values of E_0 for staphylococcal nuclease at different values of pH are given in Table 1.

For the analysis of the variation of the thermodynamic properties of the system during the folding process one can omit all the contributions to the free energy of the system that do not alter significantly in the temperature range between −50 °C and 150 °C. Therefore, from the expression for the total free energy of the system F we can subtract all slowly varying contributions F_0 as follows:

$$\delta F = F - F_0 = -(kT\ln Z - kT\ln Z_0) = -kT\ln\left(\frac{Z}{Z_0}\right). \quad (16)$$

From equation (16) follows that the subtraction of F_0 corresponds to the division of the total partition function Z by the partition function of the subsystem (Z_0) with slowly varying thermodynamical properties. Therefore, in order to simplify the expressions, one can divide the partition function in equation (14) by the partition function of fully unfolded conformation of a protein (by $Z_u^a Z_s^{N_w}$) and by the partition function of N_s free water molecules (by $Z_w^{N_s}$). Thus, equation (14) can be rewritten as follows:

$$Z = \left(\frac{Z_s}{Z_w}\right)^{N_s}\left(1 + \sum_{i=a-\kappa}^{a}\frac{\kappa!\exp(iE_0/kT)}{(i-(a-\kappa))!(a-i)!} \times \left(\frac{Z_b}{Z_u}\right)^i\left(\frac{Z_w Z_E}{Z_s}\right)^{iN_w/a}\right). \quad (17)$$

Equation (17) is the final equation that is used for calculation of the partition function of the protein. In Appendix we present the exact expressions that we used for evaluation of the dependencies of heat capacity on temperature.

The dependence of heat capacity on temperature calculated for staphylococcal nuclease at different pH values are presented in Figure 1 by solid lines. The results of experimental measurements form reference [46] are presented by symbols. From Figure 1 it is seen that staphylococcal nuclease experience two folding ↔ transitions in the range of pH between 3.78 and 7.0. At the pH value 3.23 no peaks in the heat capacity is present. It means that the

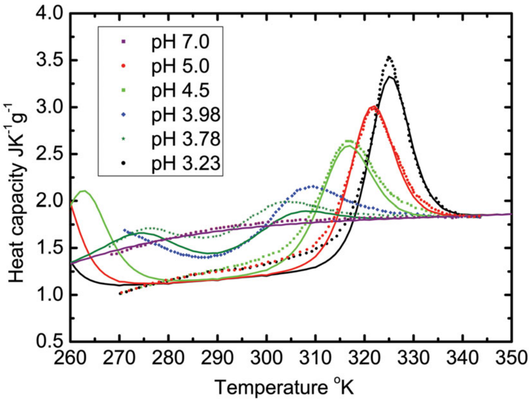

Fig. 1. Dependencies of the heat capacity on temperature for staphylococcal nuclease at different values of pH. Solid lines show results of the calculation, while symbols present experimental data from reference [46].

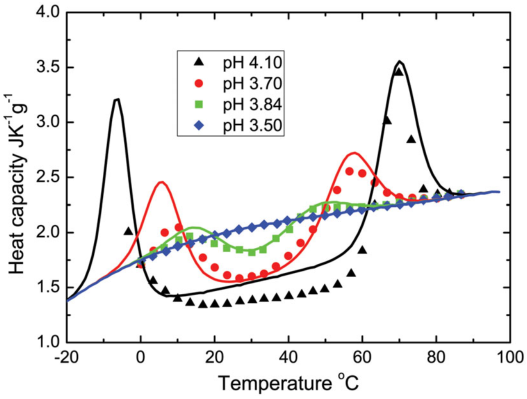

Fig. 2. Dependencies of the heat capacity on temperature for horse heart metmyoglobin at different values of pH. Solid lines show the results of the calculation. Symbols present the experimental data from reference [47].

protein exists in the unfolded state over the whole range of experimentally accessible temperatures.

Comparison of the theoretical results with experimental data shows that our theoretical model reproduces experimental behaviour better for the solvents with higher pH. The heat capacity peak arising at higher temperatures due to the standard folding $\leftrightarrow$ unfolding transition is reproduced very well for pH values being in the region 4.5−7.0. The deviations at low temperatures can be attributed to the inaccuracy of the statistical mechanics model of water in the vicinity of the freezing point.

The accuracy of the statistical mechanics model for low pH values around 3.88 is also quite reasonable. The deviation of theoretical curves from experimental ones likely arise due to the alteration of the solvent properties at high concentration of protons or due to the change of partial charge of amino acids at pH values being far from the physiological conditions.

Despite some difference between the predictions of the developed model and the experimental results arising at certain temperatures and values of pH the overall performance of the model can be considered as extremely good for such a complex process as structural folding transition of a large biological molecule.

3.2 Heat capacity of metmyoglobin

Metmyoglobin is an oxidized form of a protein myoglobin. This is a monomeric protein containing a single five-coordinate heme whose function is to reversibly form a dioxygen adduct [60]. Metmyolobin consists of 153 amino acids and its structure is shown in Figure A.1.

In order to calculate SASA of side chains of metmyoglobin exactly the same procedure as for staphylococcal nuclease was performed (see discussion in the previous subsection). SASA in the folded and unfolded states of the protein has been calculated and is equal 6847 Å^2 and 16 926 Å^2, respectively. Thus, there are 984 H_2O molecules interacting with protein's hydrophobic surface in its unfolded state.

The electrostatic interaction of water molecules with metmyoglobin was accounted for in the same way as for staphylococcal nuclease. The parameter α in equation (12) was chosen to be the same as for staphylococcal nuclease, i.e. equal to 2.5. With this we derive that 10 950 H_2O molecules involve in the interaction with the electrostatic field of metmyoglobin in its folded state. The strength of the field was chosen the same as for staphylococcal nuclease.

The parameter κ for metmyoglobin in equation (4), describing the cooperativity of the folding $\leftrightarrow$ unfolding transition, differs significantly from that for staphylococcal nuclease. The transition in metmyoglobin is less cooperative than the transition in staphylococcal nuclease because metmyoglobin has intermediate partially folded states [61]. Thus, while the rigid native-like core of the protein is formed, a significant fraction of amino acids in the flexible regions of the protein can exist in the unfolded state. We assume that 1/3 of metmyoglobin's amino acids are in the flexible region, i.e. the parameter κ in equation (4) equal to 50.

The values of E_0 in equation (6) differ from that for staphylococcal nuclease and are compiled in Table 1.

Solid lines in Figure 2 show the dependence of the metmyoglobin's heat capacity on temperature calculated using the developed theoretical model. The experimental data from reference [47] are shown by symbols.

Metmyoglobin experiences two folding $\leftrightarrow$ unfolding transitions at the pH values exceeding 3.5 which can be called as cold and heat denaturations of the protein. The dependence of the heat capacity on temperature therefore has two characteristic peaks, as seen in Figure 2. Figure 2 shows that at pH lower than 3.84 metmyoglobin exists only in the unfolded state.

The comparison of predictions of the developed theoretical model with the experimental data on heat capacity shows that the theoretical model is well applicable for metmyoglobin case as well. The good agreement of the theoretical and experimental heat capacity profiles over the whole range of temperatures and pH values shows that the model treats correctly the thermodynamics of the protein folding process.

Our theory includes a number of parameters, namely the energy difference between two phases E_0, strength of the electrostatic field E, number of interacting H_2O molecules α, the parameter describing the cooperativity of the phase transition κ, as well as other parameters introduced in reference [43] to treat the partition function of water. Three parameters, E, E_0 and κ, are dependent on the properties of a particular protein and on the pH of the solvent. We have adjusted the values of these parameters in order to reproduce the experimental data. All other parameters of the model describing the structure of energy levels of water molecules, their vibrational and librational frequencies, etc. are considered as fixed, being universal for all proteins.

In spite of the model features of our approach, we want to stress that the complex behavior and the peculiarities in dependencies of the heat capacity on temperature are all well reproduced by the developed model with only a few parameters. This was demonstrated for two proteins and we consider this result as a significant achievement. This fact supports our conclusion that the developed model can be used for the prediction of new features of phase transitions in various biomolecular systems. Indeed, from Figures 1 and 2 one can extract a lot of useful information on the heat capacity profiles: the concave bending of the heat capacity profile for a completely unfolded protein, the temperature of the cold and heat denaturation, the absolute values of the heat capacity at the phase transition temperature, the broadening of heat capacity peaks. Another peculiarity which is well reproduced by our statistical mechanics model is the decrease of the heat capacity of the folded state of the protein in comparison with that for unfolded state and asymmetry of the heat capacity peaks.

4 Conclusions

We have developed a novel statistical mechanics model for the description of folding ↔ unfolding processes in globular proteins obeying simple two-stage-like folding kinetics. The model is based on the construction of the partition function of the system as a sum over all statistically significant conformational states of a protein. The partition function of each state is a product of partition function of a protein in a given conformational state, partition function of water molecules in pure water and a partition function of water molecules interacting with the protein. The principles of the construction of the partition function of the system are thoroughly discussed and justified.

The introduced model relies on a number of physical parameters being responsible for certain characteristics and properties of the system. Most of the parameters have been determined from the available experimental data and only three of them (energy difference between two phases, cooperativity of the transition and the average strength of the protein's electrostatic field) are considered as being variable. Their choice of the variable parameters depend on a particular type of the protein and pH of the solvent.

The most prominent feature of the approach reported in the present paper distinguishing it from the earlier works is that it is developed for the real protein systems contrary to the various generalized and toy-models dealing with the analysis of the generalized features of the ideal protein-like systems and their folding characteristics.

We have compared the predictions of the developed model with the results of experimental measurements of the dependence of the heat capacity on temperature for staphylococcal nuclease and metmyoglobin. The experimental results were obtained at various pH of solvent. The suggested model is capable of reproducing well within a single framework a large number of features of the complex heat capacity profiles, such as the phenomena of cold and heat denaturations, reproduce correctly the corresponding maximum values of the heat capacities, the temperature ranges of the cold and heat denaturation transitions, the differences between the heat capacities of the protein folded and unfolded states.

The very reasonable agreement of the theoretical results with the results of experimental measurements demonstrates that the developed formalism can be used for the analysis of thermodynamical properties of many more biomolecular systems. With some advance and modification the model can be applied for the investigation of the influence of mutations on the protein stability, analysis of assembly and stability of protein complexes.

The developed theoretical model can be also used for the description of protein unfolding profiles as a result of temperature increase near the swift ion trajectories. Effects of medium ionisation resulting from ions propagation in the medium can be also included in the model via variation of pH of the solvent. However, accurate description should account also for the dynamical effects associated with spatial and temporal variation of temperature and ions/electrons concentrations near swift ion's trajectory. Such extensions of the model could be carried out in further works.

We acknowledge support of this work by the NoE EXCELL. A.Y. acknowledges Stiftung Polytechnische Gesellschaft for financial support.

Appendix

In the Appendix the details of the statistical mechanics formalism used for the construction of the partition function of proteins in water environment are presented. First, the influence of the long-range interactions on the thermodynamic characteristics of the unfolded polypeptide is discussed. Second, it is constructed the partition function of water molecules in pure water and in the vicinity of the hydrophobic solute. Third, it is presented an estimate of the number of water molecules interacting with charged groups of a protein at the finite concentration of ions.

Table A.1. Parameters of the partition function of water according to reference [43].

Number of hydrogen bonds	0	1	2	3	4
Energy level, E_i (kcal/mol)	6.670	4.970	3.870	2.030	0
Energy level, E_i^s (kcal/mol)	6.431	4.731	3.631	1.791	−0.564
Translational frequencies, $\nu_i^{(T)}$, cm^{-1}	26	86	61	57	210
Librational frequencies, $\nu_i^{(L)}$, cm^{-1}	197	374	500	750	750

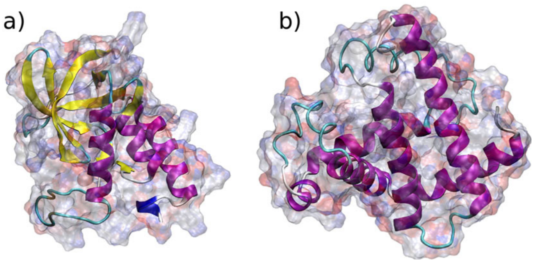

Fig. A.1. (a) Structure of staphylococcal nuclease (PDB ID 1EYD [62]), and (b) horse heart metmyoglobin (PDB ID 1YMB [63]). Images have been rendered using VMD program [64].

A.1 The influence of the long-range steric interaction on the entropy of the polypeptide chain in unfolded state

In this work we assume that in the unfolded protein states interaction between the distant backbone amino acids can be neglected. In order to support this assumption let us estimate the correction of the partition function for an unfolded protein state arising from from a steric overlap of distant segments of the polypeptide chain. The characteristic entropy change associated with unfolding of one amino acid ΔS can be estimated as $\sim$10 cal mol^{-1} K^{-1} [65], the persistence length of a protein in unfolded state l_p is $\sim$0.7 nm [66] and the length of one amino acid residue along the polypeptide chain (contour length) l_c is 0.35 nm. The latter value can be obtained for the geometry of the fully stretched conformation of metmyoglobin protein (*PDB ID 1YMB*). In three dimensions the number of final states after n-step random walk (RW) is equal to 6^n, while the number of final states after the n-step self-avoiding walk (SA) is 4.7^n [67]. The entropy of one segment of a polymer with the length equal to the persistence length can be estimated as follows:

$$S(\mu) = -\frac{k}{n}\sum_{i=1}^{\mu^n}\frac{1}{\mu^n}\ln\frac{1}{\mu^n}, \tag{A.1}$$

$$S(\mu) \approx k\ln\mu, \tag{A.2}$$

where k is the Boltzmann constant and μ equals 4.7 and 6 for the case of SA and RW, respectively. The relative difference in the entropy of a single amino acid calculated for RW and SA can be estimated as follows:

$$\delta = \frac{|S_{RW} - S_{SA}|}{\Delta S}\frac{l_c}{l_p}, \tag{A.3}$$

where S_{SA} and S_{RW} are equal to $S(4.7)$ and $S(6)$ from equation (A.2), respectively. Substituting the values for S_{SA} and S_{RW} to equation (A.3) one obtains that $\delta \approx 3\%$. This means that the accounting for SA is beyond the accuracy of our model, because even greater uncertainties arises from the determination of ΔS, see reference [65] and references therein. Note, that the diameter of the polypeptide chain in this estimate was is equal to the persistence length of a protein. The performed estimate illustrates that the thermodynamical characteristics of a polypeptide chain are not very sensitive to the steric overlap of distant amino acids. However, the correct description of the dynamical and spatial structure of a protein in unfolded state can only be achieved with accounting for the long-range interactions.

A.2 The partition function of water

Following the formalism developed in reference [43] the partition function of a water molecule in pure water reads as:

$$Z_w = \sum_{l=0}^{4}\left[\xi_l f_l \exp(-E_l/kT)\right], \tag{A.4}$$

where the summation is performed over 5 possible states of a water molecule (the states in which water molecule has $4,3,2,1$ or 0 hydrogen bonds with the neighboring molecules). E_l are the energies of these states and ξ_l are the combinatorial factors being equal to $1,4,6,4,1$ for $l = 0,1,2,3,4$, respectively. They describe the number of choices to form a given number of hydrogen bonds. f_l in equation (A.4) describes the contribution due to the partition function arising to the translation and libration oscillations of the molecule. In the harmonic approximation f_l are equal to:

$$f_l = \left[1 - \exp(-h\nu_l^{(T)}/kT)\right]^{-3}\left[1 - \exp(-h\nu_l^{(L)}/kT)\right]^{-3}, \tag{A.5}$$

where $\nu_l^{(T)}$ and $\nu_l^{(L)}$ are translation and libration motions frequencies of a water molecule in its lth state, respectively. These frequencies are calculated in reference [43] and are given in Table A.1. The contribution of the internal vibrations of water molecules is not included in equation (A.4) because the frequencies of these vibrations are practically not influenced by the interactions with surrounding water molecules.

The partition function of a water molecule from the protein's first solvation shell reads as:

$$Z_s = \sum_{l=0}^{4}\left[\xi_l f_l \exp(-E_l^s/kT)\right], \tag{A.6}$$

where f_l are defined in equation (A.5) and E_l^s denotes the energy levels of a water molecule interacting with aliphatic hydrocarbons of protein's amino acids. Values of energies E_l^s are given in Table A.1. For simplicity we treat all side-chain radicals of a protein as aliphatic hydrocarbons because most of the protein's hydrophobic amino acids consist of aliphatic-like hydrocarbons.

It is possible to account for various types of side chain radicals by using the experimental results of the measurements of the solvation free energies of amino acid radicals from reference [23] and associated works. However, this correction will imply the reparametrization of the theory presented in [43] and will lead to the introduction of $\sim 20 \times 5$ additional parameters. Here we do not perform such a task since this kind of improvement of the theory would smear out the understanding of the principal physical factors underlying the protein folding ↔ unfolding transition.

A.3 The number of water molecules interacting with a point charge in electrolyte

The total number of water molecules N_E that interact with the electrostatic field of the protein can be estimated from the known Debye screening length of a charge in electrolyte λ_d as follows:

$$N_E = N_q \frac{4\pi\rho}{3}\lambda_d^3, \tag{A.7}$$

where N_q is the number of charged groups in the protein, ρ is the density of water and λ is the Debye screening length. Debye screening length of the symmetric electrolyte can be calculated as follows [68]:

$$\lambda_d = \sqrt{\frac{\epsilon\epsilon_0 kT}{2N_A e^2 I}}, \tag{A.8}$$

where ϵ_0 is the permittivity of free space, ϵ is the dielectric constant, N_A is the Avogadro number, e is the elementary charge and I is the ionic strength of the electrolyte.

The experiments on denaturation of staphylococcal nuclease and metmyoglobin were performed in 100 mM ion buffer of sodium chloride and 10 mM buffer of sodium acetate, respectively [46,47]. The Debye screening length in water with 10 mM and 100 mM concentration of ions is $\lambda_d = 30$ Å and $\lambda_d = 10$ Å at room temperature, respectively.

The described method allows to estimate the number of water molecules (N_E) interacting with electric filed created by the charged groups of a protein. It should be considered as qualitative estimate since we have assumed the average electric field as being constant within a sphere of the radius λ_d, but in fact it experiences some variations. Thus, at the distances ~15 Å from the point charge the interaction energy of a H_2O molecule with the electric field becomes equal to ~0.02 kT (for this estimate we have used the linear growing distance-dependent dielectric susceptibility $\epsilon = 6R$ as derived in Ref. [69] for the atoms fully exposed to the solvent). However, we expect that the more accurate analysis accounting for the spatial variation of the electric field will not change significantly the results of the analysis reported here, because it is based on the physically correct picture of the effect and the realistic values of all the physical quantities.

A.4 Calculation of the heat capacity

Having constructed the partition function of the system we can evaluate with its use all thermodynamic characteristics of the system, such as e.g. entropy, free energy, heat capacity, etc. The free energy (F) and heat the capacity (c) of the system can be calculated from the partition function as follows:

$$F(T) = -kT \ln Z(T), \tag{A.9}$$

$$c(T) = -T\frac{\partial^2 F(T)}{\partial T^2}. \tag{A.10}$$

In this work we analyze the dependence of protein's heat capacity on temperature and compare the predictions of our model with available experimental data.

With the use of equation (A.10) on can calculate the heat capacity of the system as follows:

$$c(T) = A + B(T - T_0) - T\frac{\partial^2 F(T)}{\partial T^2}, \tag{A.11}$$

where the factors A and B are responsible for the absolute value and the inclination of the heat capacity curve, respectively. These factors account for the contribution of stiff harmonic vibrational modes in the system (factor A) and for the unharmonic correction to these vibrations (factors B and T_0). The contribution of protein's stiff vibrational modes and the heat capacity of the fully unfolded conformation of protein is also included into these factors. In our numerical analysis we have adjusted the values of A, B and T_0 in order to match experimental measurements. However, factors A, B and T_0 should not be considered as parameters of our model since their values are not related to the thermodynamic characteristics of the folding ↔ unfolding transition and depend not entirely on the properties of the protein but also on the properties of the solution, protein and ion concentrations, etc.

In our calculations for the protein staphylococcal nuclease we have used the values of $A = 1.25$ J K^{-1} g^{-1}, $B = 6.25 \times 10^{-3}$ J K^{-2} g^{-1} and $T_0 = 323$ K in equation (A.11).

In our calculations for metmyoglobin we have used the values of $A = 1.6$ J K^{-1} g^{-1}, $B = 8.25 \times 10^{-3}$ J K^{-2} g^{-1} and $T_0 = 323$ K in equation (A.11).

References

1. E. Surdutovich, A.V. Solov'yov, J. Phys.: Conf. Ser. **373**, 012001 (2012)
2. I. Baccarelli, F.A. Gianturco, E. Scifoni, A.V. Solov'yov, E. Surdutovich, Eur. Phys. J. D **60**, 1 (2010)
3. E. Surdutovich, A.V. Yakubovich, A.V. Solov'yov, Sci. Rep. **3**, 1289 (2013)

4. V. Muñoz, Annu. Rev. Biophys. Biomol. Struct. **36**, 395 (2007)
5. K.A. Dill, S.B. Ozkan, M.S. Shell, T.R. Weikl, Annu. Rev. Biophys. **37**, 289 (2008)
6. J.N. Onuchic, P.G. Wolynes, Curr. Opt. Struct. Biol. **14**, 70 (2004)
7. E. Shakhnovich, Chem. Rev. **106**, 1559 (2006)
8. N.V. Prabhu, K.A. Sharp, Chem. Rev. **106**, 1616 (2006)
9. A.V. Yakubovich, I.A. Solov'yov, A.V. Solov'yov, W. Greiner, Eur. Phys. J. D **46**, 215 (2007)
10. A.V. Yakubovich, I.A. Solov'yov, A.V. Solov'yov, W. Greiner, Europhys. News **38**, 10 (2007)
11. I.A. Solov'yov, A.V. Yakubovich, A.V. Solov'yov, W. Greiner, Eur. Phys. J. D **46**, 227 (2008)
12. A.V. Yakubovich, I.A. Solov'yov, A.V. Solov'yov, W. Greiner, Eur. Phys. J. D **40**, 363 (2006)
13. A.V. Yakubovich, I.A. Solov'yov, A.V. Solov'yov, W. Greiner, Eur. Phys. J. D **39**, 23 (2006)
14. A.V. Yakubovich, I.A. Solov'yov, A.V. Solov'yov, W. Greiner, Khimicheskaya Fizika [Chem. Phys.] **25**, 11 (2006) (in Russian)
15. A.V. Yakubovich, I.A. Solov'yov, A.V. Solov'yov, W. Greiner, Eur. Phys. J. D **51**, 25 (2009)
16. I.A. Solov'yov, A.V. Yakubovich, A.V. Solov'yov, W. Greiner, J. Exp. Theor. Phys. **103**, 463 (2006)
17. I.A. Solov'yov, A.V. Yakubovich, A.V. Solov'yov, W. Greiner, Phys. Rev. E **73**, 021916 (2006)
18. I.A. Solov'yov, A.V. Yakubovich, A.V. Solov'yov, W. Greiner, J. Exp. Theor. Phys. **102**, 314 (2006)
19. A.V. Yakubovich, A.V. Solov'yov, W. Greiner, Int. J. Quantum Chem. **110**, 257 (2010)
20. A.V. Yakubovich, A.V. Solov'yov, W. Greiner, AIP Conf. Proc. **1197**, 186 (2009)
21. A.D. Robertson, K.P. Murphy, Chem. Phys. **97**, 1251 (1997)
22. P. Privalov, G. Makhatadze, Adv. Protein Chem. **47**, 307 (1995)
23. G. Makhatadze, P. Privalov, J. Mol. Biol. **232**, 639 (1993)
24. P.J. Flory, *Statistical Mechanics of Chain Molecules* (Wiley, New York, 1969)
25. K. Murphy, E. Freire, Adv. Protein Chem. **43**, 313 (1992)
26. N. Go, Annu. Rev. Biophys. Bioeng. **12**, 183 (1983)
27. N. Go, H. Abe, Biopolymers **20**, 991 (1981)
28. G. Nemethy, H. Scheraga, J. Chem. Phys. **36**, 3382 (1962)
29. P. Lewis, F. Momany, H. Scheraga, Proc. Natl. Acad. Sci. USA **68**, 2293 (1971)
30. P. Kim, R. Baldwin, Annu. Rev. Biochem. **59**, 631 (1990)
31. R.L. Baldwin, Proc. Natl. Acad. Sci. USA **83**, 8069 (1986)
32. S. Somani, B.J. Killian, M.K. Gilson, J. Chem. Phys. **130**, 134102 (2009)
33. S. Somani, M.K. Gilson, J. Chem. Phys. **134**, 134107 (2011)
34. E. Henry, W. Eaton, Chem. Phys. **307**, 163 (2004)
35. S. Kumar, C.-J. Tsai, R. Nussinov, Biol. Cyber. **41**, 5359 (2002)
36. L. Pratt, Annu. Rev. Phys. Chem. **53**, 409 (2002)
37. L.R. Pratt, A. Pohorille, Chem. Rev. **102**, 2671 (2002)
38. K. Murphy, P. Privalov, S. Gill, Science **247**, 559 (1990)
39. K. Murphy, S. Gill, J. Mol. Biol. **222**, 699 (1991)
40. F. Avbelj, R.L. Baldwin, Proc. Natl. Acad. Sci. USA **99**, 1309 (2002)
41. T. Lazaridis, M. Karplus, Biopolymers **100**, 367 (2003)
42. H. Kaya, H. Chan, J. Mol. Biol. **326**, 911 (2003)
43. J.H. Griffith, H. Scheraga, J. Mol. Struct. **682**, 97 (2004)
44. O. Collet, J. Chem. Phys. **134**, 085107 (2011)
45. A. Bakk, J.S. Hye, A. Hansen, Biophys. J. **82**, 713719 (2002)
46. Y. Griko, P. Privalov, J. Aturtevant, S. Venyaminov, Proc. Natl. Acad. Sci. USA **85**, 3343 (1988)
47. P. Privalov, J. Chem. Thermodyn. **29**, 447 (1997)
48. W. Scott, W. van Gunsteren, in *Methods and Techniques in Computational Chemistry: METECC-95*, edited by E. Clementi, G. Corongiu (STEF, Cagliari, Italy, 1995), pp. 397–434
49. W. Cornell et al., J. Am. Chem. Soc. **117**, 5179 (1995)
50. A. MacKerell et al., J. Phys. Chem. B **102**, 3586 (1998)
51. R.M. Levy, M. Karplus, J. Kushick, D. Perahia, Macromolecules **17**, 1370 (1984)
52. S. Krimm, J. Bandekar, Biopolymers **19**, 1 (1980)
53. R. Pappu, R. Srinivasan, Proc. Natl. Acad. Sci. USA **97**, 12565 (2000)
54. D. Yoshioka, *Statistical Physics. An introduction* (Springer-Verlag, Berlin, Heidelberg, 2007)
55. M. Cubrovic, O. Obolensky, A.V. Solov'yov, Eur. Phys. J. D **51**, 41 (2009)
56. A. Finkelstein, O. Ptitsyn, *Protein Physics. A Course of Lectures* (Elsevier Books, Oxford, 2002)
57. F.A. Cotton, J. Edward, E. Hazen, M.J. Legg, Proc. Natl. Acad. Sci. USA **76**, 2551 (1979)
58. J.C. Phillips et al., J. Comput. Chem. **26**, 1781 (2005)
59. H.-X. Zhou, Biophys. J. **83**, 2981 (2002)
60. J.P. Collman, R. Boulatov, C.J. Sunderland, L. Fu, Chem. Rev. **104**, 561 (2004)
61. D. Shortle, M.S. Ackerman, Science **293**, 487 (2001)
62. J. Chen, Z. Lu, J. Sakon, W. Stites, J. Mol. Biol. **303**, 125 (2000)
63. S. Evans, G. Brayer, J. Mol. Biol. **213**, 885 (1990)
64. W. Humphrey, A. Dalke, K. Schulten, J. Mol. Graph. **14**, 33 (1996)
65. G. Brady, K.A. Sharp, Curr. Opin. Struct. Biol. **7**, 215 (1997)
66. T. Fisher, A. Oberhauser, M. Carrion-Vazquez, P.E. Marszalek, J.M. Fernandez, Trends Biochem. Sci. **24**, 379 (1999)
67. I. Jensen, J. Phys. A **37**, 5503 (2004)
68. W. Russel, D. Saville, W. Schowalter, *Colloidal Dispersions* (Cambridge University Press, 1989)
69. B. Mallik, T.L.A. Masunov, J. Comput. Chem. **23**, 1090 (2002)