Effect of solvation-related interaction on the low-temperature dynamics of proteins

Guanghong Zuo,[1,2] Jun Wang,[1,*] Meng Qin,[1] Bin Xue,[1] and Wei Wang[1,*]

[1]*Nanjing National Laboratory of Microstructure and Department of Physics, Nanjing University, Nanjing 210093, China*
[2]*T-Life Research Center, Department of Physics, Fudan University, Shanghai 200433, China*

The effect of solvation-related interaction on the low-temperature dynamics of proteins is studied by taking into account the desolvation barriers in the interactions of native contacts. It is found out that about the folding transition temperature, the protein folds in a cooperative manner, and the water molecules are expelled from the hydrophobic core at the final stage in the folding process. At low temperature, however, the protein would generally be trapped in many metastable conformations with some water molecules frozen inside the protein. The desolvation takes an important role in these processes. The number of frozen water molecules and that of frozen states of proteins are further analyzed with the methods based on principal component analysis (PCA) and the clustering of conformations. It is found out that both the numbers of frozen water molecules and the frozen states of the protein increase quickly below a certain temperature. Especially, the number of frozen states of the protein increases exponentially following the decrease in the temperature, which resembles the basic features of glassy dynamics. Interestingly, it is observed that the freezing of water molecules and that of protein conformations happen at almost the same temperature. This suggests that the solvation-related interaction performs an important role for the low-temperature dynamics of the model protein.

　　　　PACS number(s): 87.15.Cc, 87.15.kr, 87.15.hm, 87.15.ap

I. INTRODUCTION

Globular proteins regularly work in an aqueous environment. Their dynamics of folding and aggregations are generally coupled with large variations in their solvent-accessible surfaces. During these dynamic processes, the solvent-related electrostatic interaction, hydrogen bonds, and hydrophobicity contribute a lot to the variations of free energy of protein systems [1–5]. To clarify the effect of the solvent-related interactions in protein systems is thus essential to understand the dynamic behaviors of proteins.

The effects of the solvent water and its interaction with biopolymers around the physiological temperatures have been studied extensively for protein systems [5–17]. The solvent-related interactions have widely described in the hydrophobicity. The effective attraction of hydrophobicity between residues acts as the basis for the stability of proteins, as widely illustrated in many minimalist models [18–26]. During recent decades, some other features are pointed out for the desolvation processes beyond the consideration of energetic bias [10,22–28]. For example, an additional free-energy barrier is introduced in some researches to consider the effect of finite size and hydrogen-bond geometry of water molecules during the desolvation of water molecules out of a protein [10,22–26]. In some phenomenological descriptions for such kind of interaction, the attraction between solvated residues does not take a downhill profile, but have a barrier to cross. This kind of interaction is tightly related to the feature of solvent water, and is used to understand the foldability and cooperativity of proteins and the pressure-induced unfolding processes [10,22–26].

Besides the studies on the dynamics around physiological conditions, there are many interests in the effect of solvent on protein dynamics at low temperatures during the recent decade [29–33]. The hierarchical feature and the roughness (or ruggedness) or the landscape are discovered from the studies on the low-temperature behaviors of proteins [33–35]. It is found that there is a large change in the variation in mean-squared displacement for solvated proteins below a certain temperature T_D comparing with the physiological cases, but this kind of change could not be observed for dry sample of proteins [31–33]. This phenomenon is also observed for RNA systems [36,37]. It is believed that solvent water takes an irreplaceable role in protein dynamics at low temperatures [37–43]. This kind of phenomenon is recently suggested as a result of the variation in structural relaxation of proteins [44], and was previously related to the glass transition at low temperatures in simulations [40,42,45]. However, the all-atom-based simulations generally meet the sampling difficulty at such kind of low temperatures, and the models with only attractive solvent-related interactions are no sufficient to produce the diverse relaxation of protein systems. There are still no convincing theories to explain what happens in the microscopic scale at low temperatures. What kind of effects produces the diverse relaxation of proteins at low temperatures? What is the role of the solvent water take in this low-temperature dynamics? Facing with these questions, the studies to incorporate proper descriptions for the solvent-related interaction would be valuable to disclose the physical reasons how the water affects dynamic features of proteins at low temperatures. The origin of the special low-temperature dynamics of proteins could open another window for us to understand the interactions in protein systems and the source of the roughness of the folding landscape [42,43]. This knowledge may be helpful to build up a comprehensive picture for protein-water interactions and related folding dynamics at physiological conditions.

In our work, the simplified model with desolvation barriers is used to study folding behavior and low-temperature dynamics of a typical two-state protein, i.e., chymotrypsin

---

*Author to whom the correspondence shall be addressed; wangj@nju.edu.cn; wangwei@nju.edu.cn

inhibitor II (CI2) [46]. Based on an analysis on the geometries of the protein conformations, the locations of water molecules inside the protein are identified. The number of water molecules provides another coordinate to monitor the desolvation dynamics of the model protein. It is found that the number of water molecules decreases when the protein approaches to its native structure, and the folding processes are accompanied with a cooperative water-expelling event near the folding transition temperature T_f. However, below a certain temperature, the water molecules in the interior of the protein cannot be expelled even after 1000 times of expected folding time. This marks the freezing of water inside the model protein. At the same time, many misfolded conformations with water molecules embedded are observed. The landscapes of the model protein are build with the principal component analysis method based on an intensive sampling for misfolded conformations. The variation in the landscape below the temperature of freezing transition is also presented. The number of the nonergodic states increases roughly in an exponential manner following the decrease in temperature. This phenomenon is a typical feature of a glassy system. The concurrence of the freezing of water and protein structure implies the contributions of the desolvation barrier to the low-temperature dynamics. Interestingly, it is found that the temperature related to the onset of the ergodicity breaking is coincidental with that for the experimental dynamic transition of the protein. The relation between freezing transition and the desolvation-related interactions provides a possible explanation for the universality of the low-temperature dynamic transition for various biological polymers. These observations bring us a comprehensive picture about the desolvation-related freezing dynamics of proteins at rather low temperatures.

II. MODEL AND METHODS

A. Gō-like model

In this work, a typical two-state folder, the protein CI2, is used as the object for simulations. This protein has been successfully simulated with various Gō-like models [18,19,22,47]. The Gō-like potential [48] is a kind of interactions related to the funnel-like energy landscape with minimal frustrations. This potential has been widely used in simulations of protein for both lattice [45,49] and off-lattice models [18–21,47,50] and are successful in describing the folding, cooperativity, and the landscape of small single-domain proteins. As in some experiments [43], small two-state proteins behave with a single-exponential folding kinetics even at rather low temperatures, which supports the idea of minimal frustration. Therefore, the Gō-like potential would still work properly for the typical two-state folders even at low temperatures. With this consideration, a similar formalism as the regular Gō-like model is used for the present work.

Same as regular Gō-like models, the residues are represented by a chain of beads (centered at their C_α atoms). The interactions and the parameters for proteins are composed of the terms of bonds, angles, dihedrals, and nonbonded pairs similar as those in literature [19],

$$
\begin{aligned}
V_{\text{total}} &= V_{\text{bond}} + V_{\text{bond-angle}} + V_{\text{dihedral}} + V_{\text{nonbonded}} \\
&= \sum_{\text{bonds}}^{N-1} K_r(r-r_0)^2 + \sum_{\text{bond-angles}}^{N-2} K_\theta(\theta-\theta_0)^2 + \sum_{\text{dihedral}}^{N-3} \\
&\quad \times \{K_\phi^{(1)}[1-\cos(\phi-\phi_0)] + K_\phi^{(3)}[1-\cos 3(\phi-\phi_0)]\} \\
&\quad + \sum_{i<j-3}^{\text{native}} f(r_{ij}) + \sum_{i<j-3}^{\text{non-native}} \varepsilon\left(\frac{\sigma_0}{r_{ij}}\right)^{12},
\end{aligned}
\tag{1}
$$

where N is the number of residues, ε is the unit of energy, the variables r, θ, and ϕ measure bond lengths, bond angles, and dihedrals of consecutive residues, and the variable r_{ij} represents the distance between the ith and jth beads. For the interactions related to bonds, angles, and dihedrals, the parameters r_0, θ_0, and ϕ_0 all take the corresponding values in the native structure of the concerned protein. The corresponding parameters measuring the strengths, K_r, K_θ, $K_\phi^{(1)}$, and $K_\phi^{(3)}$, are set as 100ε, 20ε, ε, and 0.5ε, respectively. This kind of setup ensures the correctness of the local geometry and chirality. The nonbonded interactions are considered only between the beads i and j which are separated by at least four virtual bonds, namely, $i<j-3$. Base on the consideration of Gō model, only the pairs of beads which form contacts in the native structure, can contribute to the native stability in our coarse-grained model. This kind of contacts are defined as native contacts, which plays a vital role in the simulation of Gō-like model [18,22]. The interaction for the rest pairs is set as hard-core repulsions to mimic the exclusive volume effect. In our model, the function for native contacts, $f(r_{ij})$, is related to the desolvation barriers, and will be given in next subsection. Here all the native contacts in nonbonded interaction are defined when the distance between two nonhydrogen atoms from two residues is smaller than 5 Å. The parameter σ_0 for the nonspecific exclusive volume interaction is taken as 4 Å. Clearly, the Gō-like interactions are designed to emphasize the stability of the native structure, and could fulfill the requirement of funnel-like energy landscape.

B. Water-mediated Interaction

Water is the most intriguing ingredient in protein systems. The complex interactions between water and proteins are one of the main difficulties in protein modeling [51,52]. The hydrophobicity is the basic aspect of interactions related to water, and is considered as the most important part contributing to the protein stability [3]. As the study of the potential of mean force (PMF) between two methanelike particles, there are two minima, a contact minimum at the van der Waals distance between the particles and a solvent-separated minimum [25]. The two minima are separated by a desolvation barrier with a width of approximately one water molecule diameter. Such a desolvation barrier is also found in the interactions between the residues of protein [53,54]. And this kind of barrier has been modeled and applied in several theoretical and simulation studies [22–24,26]. The importance of the desolvation barrier has been illustrated in some pressure-related unfolding processes [24,25]. In our model, the solvation/desolvation barrier is only considered for the
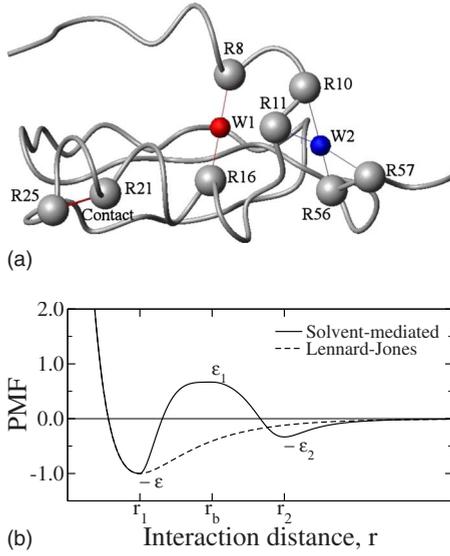
(a)

(b)

Interaction distance, r

FIG. 1. (Color online) (a) Schematic of contacts and pseudocontacts. Here two types of contact are formed. One is the native contact which is for example between residues R21 and R25. The other is the pseudocontacts. Residues R8 and R16 forms a pseudo contact with one water molecule W1 inserted between them. Residues R10 and R56, R11 and R57 are two pseudocontacts and these two pseudocontacts share the same one water molecule W2. (b) Solvent-mediated interaction and Lennard-Jones interaction. They are the native contact interactions for the solvation-related model and classical Gō-like model, respectively. The parameters for energies are extracted from Ref. [23], where $\epsilon_2 + \epsilon_1 / \epsilon_1 + \epsilon = 1.33$, $\epsilon_1 / \epsilon = 0.33$.

native contacts. Similar as the implementation in Ref. [23], a segmented function $f(r)$ is taken to fit the potential of mean force between two beads,

$$
f(r)
$$

$$
= \begin{cases} \varepsilon Z_k(r)(Z_k(r) - 2) & \text{when } r < r_1 \\ (\varepsilon + \varepsilon_b) Y_{2n}(r)(Y_{2n}(r) - 2) + \varepsilon_b & \text{when } r_1 \leq r < r_b \\ -\varepsilon_1 \dfrac{m(W_2(r) - p)+}{m(1 - p)+} \dfrac{(1 - q)}{(W_{2m}(r) - q)} & \text{when } r \geq r_b \end{cases}
$$

(2)

in which the functions Z, Y, and W are defined as $Z_k(r) = (r_1/r)^k$, and $Y_l(r) = u^l$ with $u = \frac{r - r_1}{r_b - r_1}$, and $W_m(r) = v^m$ with $v = \frac{r - r_b}{r_2 - r_b}$. In this function, the energetic parameter ε_b, ε_1 represent the height of the desolvation barrier and the depth of the water-related pseudocontact, respectively, and the distances r_1, r_b, and r_2 correspond to the positions of the contact, the desolvation barrier, and the water-related pseudocontacts, respectively. The parameters p and q takes the form, $p = \frac{\varepsilon_b}{\varepsilon_b + \varepsilon_1}$, $q = \frac{\varepsilon_1}{\varepsilon_b + \varepsilon_1} = 1 - p$. In our work, r_1 takes the value of the distance in native contact, $r_2 = r_1 + 3$ Å, and $r_b = \frac{1}{2}(r_1 + r_2)$. The other parameters are set as $k = 6$, $n = 2$, $m = 3$, $\varepsilon_b = \frac{2}{3}\varepsilon$, $\varepsilon_1 = \frac{1}{3}\varepsilon$. These parameters are adapted from the Refs. [23–25]. The related profile of the function $f(r)$ is shown in Fig. 1(a). This kind of interaction introduces an additional barrier for two residues to bind together. Some examples of contacts and the water-

related pseudocontacts are demonstrated in Fig. 1(b) as a contact between residues R21 and R25, and a pseudocontact between residues R8 and R16 related to water W1. It is worth noting that the size r_2 consists with the geometry with a water in between two residues. This kind of geometry is stabilized as a local minimum in PMF profile. This kind of interaction favors the solvation of water inside the protein, and enlarges the barrier of desolvation. Indeed, other strengths for solvation-related interactions are also tested. The conclusions would not be qualitatively changed.

C. Langevin dynamics

With the coarse-grained interactions, the kinetics of a model protein could be simulated with a Langevin equation which takes the form

$$
m\dot{\mathbf{v}}(t) = \mathbf{F}_{conf}(t) - \gamma\mathbf{v}(t) + \mathbf{\Gamma}(t),
$$

(3)

in which $\mathbf{v}$, $\dot{\mathbf{v}}$, and m are the velocity, acceleration and mass of a bead, respectively. $\mathbf{F}_{conf} = -\nabla V_{total}$ is the force related to the potential V_{total}, γ is the friction coefficient, and the $\mathbf{\Gamma}$ is the random force which satisfies $\langle \mathbf{\Gamma}(t)\mathbf{\Gamma}(t') \rangle = 6\gamma k_B T \delta(t - t')$ where k_B is the Boltzmann constant, T is the temperature, t is the time, and the function $\delta(t)$ is Dirac-δ function. Numerically, the leap-frog algorithm is used to integrate the Langevin equation. Each component of the random force is implemented by a Gaussian random generator with a null average and the variation of $\sqrt{2m\gamma k_B T / h}$ where h is the integration step. It is noted that the shortest time scale of the dynamics in our model is related to the oscillation period ($\sim\sqrt{m/K_r} = \sqrt{m\Delta^2/k_B T}$) of the virtual bonds, which puts a constraint on the size of the integration step h. Here, m is the mass of a bead, and Δ is the amplitude of bond oscillation. At regular conditions, Δ is about one-tenth of the equilibrium bond length ($L = 3.8$ Å), and the temperature satisfies $k_B T \approx \varepsilon$. We thus define the time unit $\tau = \sqrt{mL^2/\varepsilon}$. In our simulations, we generally set the time step as $h = 0.005\tau$, and the friction parameter γ as $0.5\tau^{-1}$. To simplify this notation, other units are chosen as $m = 1$ and $k_B = 1$ in the simulation as used by Veitshans et al. [55]. An approximate correspondence between the simulation time and real folding time has been discussed by Thirumalai and co-workers [55–57]. Based on these settings, the kinetic processes are simulated starting from random extended conformations with null velocities for all beads.

D. Weighted histogram analysis method

The weighted histogram analysis method (WHAM) is used to calculate relevant thermodynamic quantities based on the statistical physics [19,58,59]. WHAM yields an optimal estimate of the density of state of the system. In the canonical ensemble at temperature T, the probability distribution, P, of potential energy E is given by

$$
P(E) = (1/Z_c)w(E)e^{-E/k_B T},
$$

(4)

and the probability distribution of other reaction coordinate, such as the radius of gyrate (Q), follows:

$$P(Q) = (1/Z_c) \sum_E w(Q,E) e^{-E/k_B T}, \qquad (5)$$

where w is the density of state of the system and is obtained by solving two self-consistent equations [59]. Z_c is the canonical partition function. The free energy $F(Q)$ is estimated by the logarithm of probability $P(Q)$ [19].

E. Principal component analysis for trajectory

The two-state feature of the dynamics at folding temperature suggests that there is an essential reaction coordinate in conformational space, which is regularly identified as the structural similarity RMSD [60] or the contact similarity Q with respect to the native structure. At low temperatures, the protein molecules would generally be trapped in a large number of local minima on the energy landscape due to the ruggedness of the landscape. There are apparent structural diversity for these local minima. The regular reaction coordinates (such as RMSD and Q) would not be valid to describe the feature of such kind of landscapes. To describe the basic characteristic of the conformational space, the principal component analysis (PCA) [61–63], a frequently used method, is employed in this work.

In our procedure of the PCA, a covariance matrix $\mathbf{D}$ for distance information, rather than for the coordinate set, is used to describe the structural feature of all concerned N conformations. Here, the elements of the matrix $\mathbf{D}$ have the form

$$D_{\alpha\beta,\mu\nu} = \langle (d_{\alpha\beta} - \langle d_{\alpha\beta} \rangle)(d_{\mu\nu} - \langle d_{\mu\nu} \rangle) \rangle, \qquad (6)$$

in which $d_{\alpha\beta}$ (or $d_{\mu\nu}$) is the distance between the residues α and β (or between residues μ and ν). $\langle .. \rangle$ represents the average over all concerned conformations. Therefore, the matrix $\mathbf{D}$ is a real symmetric matrix with the dimension of $N_{bead}(N_{bead}-1)/2$ which is the number of unique pairs between residues. This matrix describes the variance and correlation of the distances between various residues. It is easy to find out that the quantity $d_{\alpha\beta}$ is independent of the translations or rotations of conformations, which overcomes the insufficiency of the coordinate representation of conformations.

The distance covariance matrix $\mathbf{D}$ would be analyzed by the regular diagonalization method, namely, by finding out a proper unitary transformation $\mathbf{W}$ which could transform the matrix $\mathbf{D}$ into a diagonal one ζ [64], $\mathbf{W}^\dagger \mathbf{D} \mathbf{W} = \zeta$. The diagonal elements $\{\zeta_m\}$ are the eigenvalues of the matrix $\mathbf{D}$, and the column of the unitary matrix $\mathbf{W}$ are the corresponding eigenvectors. Consequently, the matrix $\mathbf{D}$ could be represented with these eigencomponents. These eigencomponents (eigenvectors and eigenvalues) can be interpreted as the modes of the concerned distance fluctuations and their amplitudes. Using the modes with large amplitudes which are called as principal components, the matrix $\mathbf{D}$ could be largely simplified. The related modes with large amplitude act as the best coordinates to describe the system. Clearly, the PCA method here presents a linear analysis for the whole landscape.

F. Clustering of conformations

Clustering method is another useful method to analyze the diverse data and build up the high-dimensional landscapes, especially for the cases without any global information. For the conformations at very low temperatures, the clustering method shows a better validity to outline the features of landscapes.

In this work, the clustering is based on the similarity between the conformations, which is measured with the values of RMSD [60] between conformations. That is, two conformations are similar when the values of RMSD between them are rather small comparing with others. Similar conformations are grouped together to form a cluster which represents a state with distinguishing structural features on the landscape. The nearest-jointing method is employed in our clustering. In detail, in each step, two conformations with the largest similarity, namely with the smallest RMSD, are grouped together as a new state [65,66]. The similarities between this new state and another conformation $\mathbf{C}$ are measured as the average of the RMSD between the conformations in the state and the conformation $\mathbf{C}$. This kind of procedure is iterated and produces a treelike relationship for all the conformations. It is clear that the difference between the merged states is increasing following the iterations. With a presumed threshold on the similarity during merging, that is, the conformations with the RMSD larger than RMSD_T would not be merged, the iteration would be stopped after some certain steps. This similarity threshold generally represents the physical tolerance of the conformations to be combined (or be clustered) together as a state. Using such a procedure, the conformations are grouped into a series of states with similar structural features, and the set of these states gives out a description of heterogeneity of the concerned conformational space. In this work, we set the threshold as $\mathrm{RMSD}_T = 2$ Å. With this threshold, the conformations partitioned into the same state are rather similar. Some variations for the threshold are also evaluated (data are not shown). There are no apparent differences in the conclusions when using other values for the threshold.

III. RESULTS AND DISCUSSIONS

A. Free-energy profile and cooperativity

For the protein CI2, the free-energy profiles as a function of the fraction of native contacts (Q) at folding transition temperature T_f for solvation-related model and classical Gō-like model are obtained by WHAM, and shown in Fig. 2. It is found that both the free-energy profiles of these two models have a barrier between two basins, which is the character for the two-state folders. This is, both models could well reproduce the two-state folding behavior of CI2. There are some differences between the free-energy profiles for the two models. First, the positions of free-energy minima corresponding to the native state are different. The minimum for the classical Gō-like model locates at $Q \approx 0.85$ while the one for the solvation-related model at $Q \simeq 1$. This observation suggests that the native conformation in the solvation-related model is more rigid to produce the native basin rather close
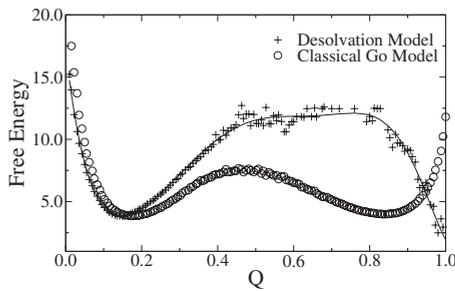
FIG. 2. Free-energy profiles for solvation-related model (+) and for classical Gō-like model protein (∘) at their T_f, respectively. The solid line is the fitting for the solvation-related model.

to the native conformation. Second, the free-energy barrier for the solvation-related model is higher and wider than that for the classical Gō-like model, with a rather shallow pit on the top of the peak. The feature for solvation-related model is consistent with the experimental observations for the free-energy landscapes for some typical proteins [67,68]. Comparing with the classical Gō model, the free-energy barrier (especially the part in the region with large Q) would comes from the desolvation barrier to expel the water molecules. The flat barrier over a wide range of Q for solvation-related model demonstrates the delicate balance between entropic and energetic factors, which introduces more physics into the solvation-related model. Consistently, the cooperative factor [69,70] is determined as $\kappa_2 = 0.94$ for the solvation-related model, and $\kappa_2 = 0.84$ for the classical Gō-like model. An apparent enhancement for the cooperativity is observed. This result clearly from the feature of wide and high free-energy barrier in the solvation-related model. The desolvation barrier could introduce more cooperativity into protein model, which meets better the realistic requirement for protein models [10,71].

B. Desolvation processes at different temperatures

In the present solvation-related model, the interaction between water molecules (namely, the pseudocontacts between residues) would be useful to characterize the desolvation processes of the concerned protein. By analyzing the number of pseudocontacts (and the number of the related water molecules), the movement of water molecules in the hydrophobic core during folding processes can be traced. Note that different pseudocontacts may share a same water molecule, e.g., in Fig. 1(a) the water molecule W2 is shared by the pseudocontacts of R10-R56 and R11-R57. The number of buried water molecules is probably smaller than the number of pseudo contacts. Practically, when the distance between two identified water molecules is less than a certain threshold related to the radius of a water molecule, these two water molecules are likely to be the same one. In our procedure, the threshold to distinguish two water molecules are set as $1.25r_{water}$. This length is chosen to tolerate the possible vibration of buried water molecules. Here, r_{water} is the radius of water molecule.

As a result, different desolvation processes are observed at high and low temperatures. In Fig. 3, the relaxations of the
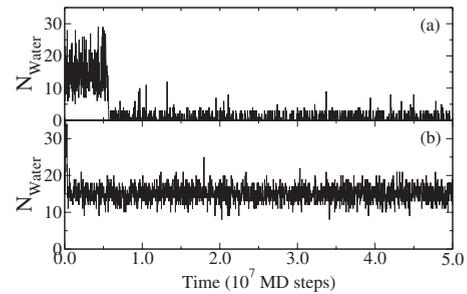

FIG. 3. Relaxation of the number of water molecules in the protein (a) $T = 0.94T_f$ and (b) $T = 0.47T_f$, respectively. Here, $T_f = 0.85$ is the temperature of folding transition.

number of water molecules inside the protein at the temperatures $T = 0.94T_f$ and $= 0.47T_f$ are presented. At the temperature around the folding transition (namely $T = 0.94T_f$), the water molecules which solvate with the protein chain in the unfolded state are almost all expelled outside the protein. The expelling process is cooperative, i.e., the water molecules (15 on average) are expelled almost at the same time [as shown in Fig. 3(a)]. To establish a statistical view on such a kinetic process, 1000 folding simulations are carried out. The temporal variation in the average number of embedded water molecules is given in Fig. 4. Three stages could be observed as indicated in Fig. 4. The first step happens in the unfolded state. During this stage, nearly 40% of all native contacts could be established, and the exclusion of water molecules is in a gradual manner. Referring to the kinetic trajectories, the quick variation for the number of water molecules [as shown in Fig. 3(a)] suggests that the water molecules can easily enter into or be expelled out of the concerned conformations. In this sense, this stage is related to the establishment of the partial compact unfolded conformations. The expulsion of water molecules is easy and the solvation-related interaction has a weak effect on the kinetics. The second stage is related to the crossing of the free-energy barrier referring to the free-energy profile (as in Fig. 1). During this stage, the number of water molecules inside the protein fluctuates around 10. The little invariance of the average number of water molecules indicates that there are no water molecules expelled in this stage. Combined with structural analysis, it is found that these embedded water molecules are highly buried inside the protein. These water
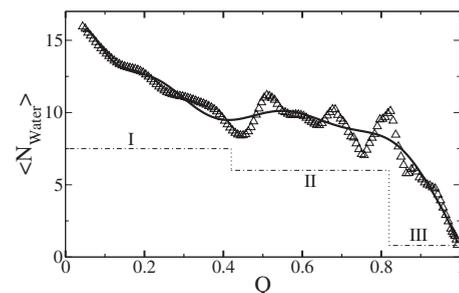

FIG. 4. Average number of the water molecules inside the protein versus Q. The sold line is the fitted for this data. The Roman numbers and auxiliary lines are used to indicate the three stages of the folding process.

molecules are generally interacted with the residues in core of the protein. Therefore, this stage is related to the slow diffusion of water molecules from the core to the outside of the protein, accompanied with the large scale reorganization of the protein chain to form the transition state. This process is important for both the folding event and the expulsion process. Clearly, the desolvation barrier largely affects the diffusion of the internal water molecules, and acts as a dominant factor in this stage. The last stage has an efficient exclusion of all the inside water molecules. It is related to the cooperative formation of the native conformation after the transition state. This kind of three-stage process is also observed in previous studies [23]. Together with all these studies, it is suggested that the exclusion of water molecules from the core of the protein is the key step during folding kinetics.

Differently, at a rather low temperature ($T=0.47T_f$), a large number of water molecules cannot be expelled out from the protein, even after a very long time (about one order of the typical folding time at the temperature $T=T_f$). In other words, these water molecules are trapped inside the protein system. In this trajectory, the structural analysis demonstrates that the conformation of the protein also varies slightly after a first quick compaction. This kind of slightly structural variation implies that the protein is in a frozen state. The frozen states would be sensitive to the initial conditions. The resultant conformations after various simulation runs are largely different. In this situation, the protein molecule is trapped in the local minima of the energetic landscape at this temperature, and the thermodynamic equilibrium between these minima cannot be established for such a kind of temperature. In such a condition, the water molecules frozen inside the protein prohibit the further folding of protein chain toward the native structure due to the energetic barrier related to solvation-related interactions. As a result, the desolvation barrier is one of the important factors to produce the frozen conformations of protein and to freeze the water molecules inside the protein. The water introduces another scenario for the protein system at low temperatures.

Clearly, the protein exhibits totally different behaviors at two kinds of temperatures though a same interaction is applied. The transition between these two kinds of behaviors would give out a quantitative measurement for the strength of the interactions. Therefore, to discover when the transition happens and how the freezing occurs would be very useful to understand the feature of interactions in protein systems and the folding dynamics at various temperatures. These are focused in the following sections.

C. Freezing of water molecules and frozen state of protein

As shown above, the inside water molecules could not be expelled out of proteins at low temperatures due to the disability of water molecules to overcome the desolvation barriers. The configurations and dynamics of the protein are tightly related to these water molecules. Are there any requirements for the number of water molecules inside the protein? Do these water molecules uniformly distributed inside the protein, or concentrate together? Are there any regulari-
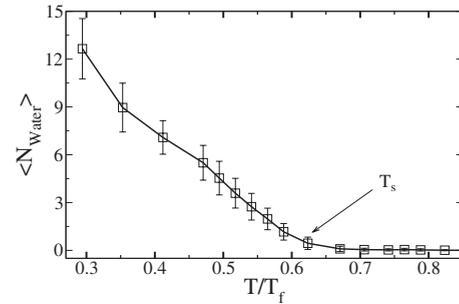

FIG. 5. Number of frozen water molecules in the protein at different temperatures. T_s indicates the freezing temperature of the water, which water molecules begins to freeze inside the protein.

ties for the conformations of the related protein? How do these properties change following the decrease in temperature? The answer to these questions would be interesting for the understanding on the low-temperature features of the protein system.

The temperature dependence of the number of frozen water molecules is shown in Fig. 5. To practically judge the occurrence of freezing phenomenon, the upper limit of the simulation time is set as $\tau_{max}=5\times10^7$ MD steps, which is much longer than the folding time at the physiological temperature. If the first passage time (FPT) of a folding process is smaller than the above threshold τ_{max}, namely, $\tau_f < \tau_{max}$, all the water molecules can be expelled out of the protein. In other words, there are no frozen water molecules in this case. Differently, for the cases in which the native state is never reached in the whole simulation, the protein generally takes a frozen conformation containing some water molecules inside. The number of water molecules inside the protein is counted in statistics. Practically, a series of (generally 100) sample conformations are picked up in the successive simulations after τ_{max} (namely $\tau > \tau_{max}$). The number of the embedded water molecules at a certain temperature is averaged over the samples for the concerned trajectories. To eliminate the possible structural correlation in sampling, the samples are selected with the interval $\tau_i=1000$ MD steps. At each temperature, about 1000 simulations are carried out to achieve a significant statistics.

As suggested in the last section, there is a transition of the number of frozen water molecules versus temperature (as shown in Fig. 5). The transition is located at the temperature $T_s \simeq 0.62T_f$. When the temperature is higher than this transition temperature T_s, i.e., $T > T_s$, there are no water molecules frozen in the protein. Whereas when $T \leq T_s$, the freezing of water molecules appears. There are nonzero counts for the frozen water molecules. This marks the existence of a new scenario at low temperatures. Following the decrease in temperature, more and more water molecules are frozen inside the protein. This demonstrates that the expelling of the water molecules from hydrophobic core become harder. When the temperature is about $0.35T_f$, the average number of frozen water molecules is close to 9, which is the number of water molecules in the transition state. Referring to the expulsion process of water molecules near the transition temperature (as shown in Fig. 4), these trapped molecules are probable in the hydrophobic core of the protein, so that they cannot be
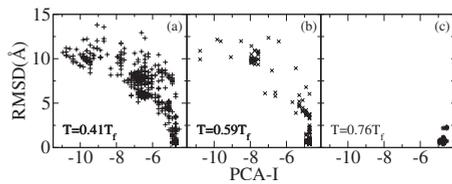
FIG. 6. The distribution of the conformations in the space spanned with the RMSD related to native conformation and the first vector of *PCA* analysis (*PCA*−*I*) at three temperatures, (a) *T* =0.41T_f, (b) *T*=0.59T_f, and (c) *T*=0.76T_f.

expelled at the initial collapse process. This reflects the capacity of the protein to contain water molecules internally. Detailed structural analyses show that the frozen water molecules are not uniformly distributed inside the protein. Actually, most of them are concentrated together in the gap between the loop of 10–14 and the loop 54–59. Some others are distributed in the gap between the N-loop and the α-helix as well as in the space between the loop of 51–56 and the turn of β-hairpin. These regions are generally related to the hydrophobic core of the protein CI2. Thus, this further confirms the previous conclusion on the water distribution. Structural analysis also illustrates that various embedded water molecules are generally related to different number of the pseudo contacts. That is, these water molecules would experience difference environments. Therefore, the expulsion of these water molecules would be a heterogeneous process with difference water molecules having diverse energetic barriers. This makes the low-temperature behavior of the whole system rather complex.

With the diverse energetic barriers for various frozen conformations at low temperatures, the landscape of the protein would not be as simple as the one at the folding transition temperature. Near the folding transition, the free-energy landscape has only two dominant basins. Meanwhile, the landscape would have a large number of metastable states, and become rather rugged. As a demonstration, the frozen protein conformations are projected to some certain coordinates (the RMSD related to native structure, and the first vector of PCA analysis). The distribution of frozen states on the two-dimensional landscape is given in Fig. 6. Three cases with different temperatures are given. Here, the final conformations of the trajectories are taken as the frozen conformations when no folding events happen within τ_{max} steps. The native structure is used for analysis when native state could be reached in the concerned simulation. 1000 runs are performed at each temperature. That is, there are 1000 points in every subfigure of Fig. 6. As shown in Fig. 6, at the temperature *T*=0.76T_f, all 1000 trajectories can fold to the native state. No freezing happens at this temperature. When the temperature drops to 0.59T_f, some trajectories are trapped. The 1000 frozen states are concentrated in some certain regions of the landscape. This illustrates the occurrence of the metastable states on the landscape. When the temperature is even low, say *T*=0.41T_f, most of trajectories cannot reach the native state, and stop at different places in the conformational space. Thus, the spots are diversely distributed on the landscape. The emergence of these metastable states and the related diverse energetic barriers, as well as the inaccessibil-
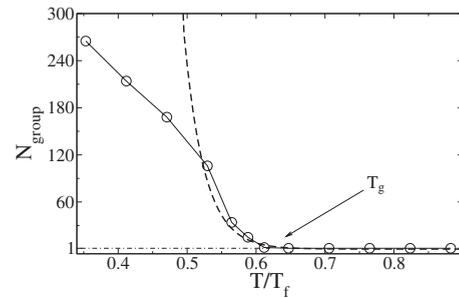

FIG. 7. The number of substates (groups) at various temperatures. T_g marks the temperature of the freezing transition for the conformations of the protein.

ity of the native conformation are generally related to the breaking of the ergodicity of the landscape. This would make the properties of the system sensitively depend on the distribution of the metastable state rather than on the features of the native state. This resembles the glass states in condensed matter physics, and marks a new macroscopic state of proteins at the low temperatures.

Besides the projection method to build up a visualization of the landscape, the clustering method offers another way to describe the variation in the landscapes of proteins. Here, 1000 frozen structures at each temperature are used to build up RMSD matrix (see the section of Model and Method). With the presumed threshold for RMSD, these structures are grouped into a series of conformational states. Based on the consideration of clustering method, the number of the conformational states, N_{group}, describes the diversity of states on the landscape. The quantity N_{group} is used as a monitor for the low-temperature transition. In Fig. 7, the numbers N_{group} for various temperatures are shown. There is a critical temperature T_g=0.61T_f separating the cases with a unique frozen state (namely the native state) and those with multiple metastable states. As shown in Fig. 7, at temperature lower than T_g, a number of freezing states are observed for different trajectories. Except the folded state, all the states include some water molecules in the core of the protein. As the temperature decreases, the number of such substates increased exponentially. This kind of increase in the states is consistent with the above analysis on the projected landscapes. The exponential growth of the population of stable states is another feature in the regular glass transition, which is also suggested in some previous studies [29,72,73]. This diverse distribution of misfolded conformations and the resultant barriers between the states would introduce a broad range of relaxation times, which is consistent with the declaration in recent experiments [44]. This feature may be the physical source of the dynamic transition of protein systems. It is interesting to find out that this kind of freezing is not observed in the models without the consideration for desolvation-related barriers, such as those with Gō interaction [65] or with MJ interactions [74]. In those models, only a few metastable states determines the low-temperature dynamics, and the dynamics are generally activation processes with Arrhenius dependence on the temperature. Differently, the freezing in the present model has diverse distribution of metastable intermediates, and the number of intermediates

grows very quickly. The resultant kinetics is not simply activation processes. This phenomenon illustrates that the water-related interaction introduces many local frustrations and the desolvation-related barriers would be important for these freezing processes at low temperatures. Interestingly, the temperature of the water molecules beginning to freeze in the hydrophobic core, T_s, is almost equal to the temperature of the glass transition, T_g. This observation implies that the solvation-related interaction performs an important role in the glass transition of model protein system, which induces the concurrence of two kinds of freezing temperatures. In addition, the tight relation between the transition and the solvation-related interactions suggests a universality of the transition temperatures for various proteins, namely the invariance of the transition temperatures over proteins, since the transition is largely determined by the desolvation barrier and the barrier is insensitive to the types of proteins. This kind of feature of transition has been observed in experiments for various proteins. It is even observed that the RNA also shares the similar transition temperatures. The water-related interaction gives an explanation for the phenomenon of the common dynamic transition for various biopolymers. Besides, these two transition temperatures, T_s and T_g, is approximate to 198 K with the assumption on the temperature of folding transition $T_f = 320$ K. This accidentally equals to the temperature corresponding to the glass transition or the dynamics transition of proteins [31,33]. Even, the whole transition range (from $0.7T_f$ to $0.6T_f$) matches the experimental observations [37,75]. This reflects that the interaction and parameters derived from physiological experiments could be applied to low-temperature cases, and demonstrates the physical consistence between low-temperature freezing and high-temperature folding. This kind of consistence really indicates that this solvation-related model does capture some basic features of the protein systems and the low-temperature dynamics act as another way to understand this kind of interactions.

IV. CONCLUSION

In this paper, we applied a coarse-grained model with solvation-related interaction to study the folding behavior and low temperature dynamics of a model protein. The water molecules are identified based on pseudocontacts between residues. It is found that the desolvation efficient improves the folding cooperativity, and the water molecules are expelled from the hydrophobic core at the final stage in the folding process. At low temperature, however, the water molecules would be frozen in the hydrophobic core of the protein to trap the folding in glass state. The freezing process of water molecules and the protein conformations are described. A large number of nonergodic substates are observed in the glassy state of the protein. The solvation-related energy takes an important role in low-temperature dynamics of proteins.

[1] M. Lin, N. Fawzi, and T. Head-Gordon, Structure **15**, 727 (2007).

[2] Y. M. Rhee, E. J. Sorin, G. Jayachandran, E. Lindahl, and V. S. Pande, Proc. Natl. Acad. Sci. U.S.A. **101**, 6456 (2004).

[3] K. A. Dill, Biochemistry **29**, 7133 (1990).

[4] D. Eisenberg and A. D. McLachlan, Nature (London) **319**, 199 (1986).

[5] B. Baldwin, R. L. Baldwin, and D. Baker, *Peptide Solvation and H-bonds* (Academic Press, New York, 2006).

[6] J. England, D. Lucent, and V. Pande, Curr. Opin. Struct. Biol. **18**, 163 (2008).

[7] J. Juraszek and P. G. Bolhuis, Proc. Natl. Acad. Sci. U.S.A. **103**, 15859 (2006).

[8] A. M. Fernández-Escamilla, M. S. Cheung, M. C. Vega, M. Wilmanns, J. N. Onuchic, and L. Serrano, Proc. Natl. Acad. Sci. U.S.A. **101**, 2834 (2004).

[9] A. Cooper, Biophys. Chem. **115**, 89 (2005).

[10] M. Knott and H. S. Chan, Chem. Phys. **307**, 187 (2004).

[11] J. Chen, W. Im, and C. L. Brooks, J. Am. Chem. Soc. **128**, 3728 (2006).

[12] D. J. Felitsky, M. A. Lietzow, H. J. Dyson, and P. E. Wright, Proc. Natl. Acad. Sci. U.S.A. **105**, 6278 (2008).

[13] R. L. Baldwin, J. Mol. Biol. **371**, 283 (2007).

[14] L. Cruz, B. Urbanc, J. M. Borreguero, N. D. Lazo, D. B. Teplow, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **102**, 18258 (2005).

[15] E. J. Sorin, Y. M. Rhee, M. R. Shirts, and V. S. Pande, J. Mol. Biol. **356**, 248 (2006).

[16] D. Lucent, V. Vishal, and V. S. Pande, Proc. Natl. Acad. Sci. U.S.A. **104**, 10430 (2007).

[17] D. Rodriguez-Larrea, S. Minning, T. V. Borchert, and J. M. Sanchez-Ruiz, J. Mol. Biol. **360**, 715 (2006).

[18] G. Zuo, J. Wang, and W. Wang, Proteins: Struct., Funct., Bioinf. **63**, 165 (2006).

[19] C. Clementi, H. Nymeyer, and J. N. Onuchic, J. Mol. Biol. **298**, 937 (2000).

[20] N. Koga and S. Takada, J. Mol. Biol. **313**, 171 (2001).

[21] H. Kaya, Z. Liu, and H. S. Chan, Biophys. J. **89**, 520 (2005).

[22] H. Kaya and H. S. Chan, J. Mol. Biol. **326**, 911 (2003).

[23] M. S. Cheung, A. E. Garcia, and J. N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **99**, 685 (2002).

[24] N. Hillson, J. N. Onuchic, and A. E. García, Proc. Natl. Acad. Sci. U.S.A. **96**, 14848 (1999).

[25] G. Hummer, S. Garde, A. E. García, M. E. Paulaitis, and L. R.

Pratt, Proc. Natl. Acad. Sci. U.S.A. **95**, 1552 (1998).

[26] Z. Liu and H. S. Chan, J. Mol. Biol. **349**, 872 (2005).

[27] T. F. Miller, E. Vanden-Eijnden, and D. Chandler, Proc. Natl. Acad. Sci. U.S.A. **104**, 14559 (2007).

[28] A. Ferguson, Z. Liu, and H. S. Chan, J. Mol. Biol. **389**, 619 (2009).

[29] V. Lubchenko, P. G. Wolynes, and H. Frauenfelder, J. Phys. Chem. B **109**, 7488 (2005).

[30] M. Weik, U. Lehnert, and G. Zaccai, Biophys. J. **89**, 3639 (2005).

[31] A. L. Lee and A. J. Wand, Nature (London) **411**, 501 (2001).

[32] V. Réat, R. Dunn, M. Ferrand, J. L. Finney, R. M. Daniel, and J. C. Smith, Proc. Natl. Acad. Sci. U.S.A. **97**, 9961 (2000).

[33] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, Science **254**, 1598 (1991).

[34] J. N. Onuchic, Z. Luthey-schulten, and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).

[35] H. Frauenfelder, P. G. Wolynes, and R. H. Austin, Rev. Mod. Phys. **71**, S419 (1999).

[36] D. Thirumalai and C. Hyeon, Biochemistry **44**, 4957 (2005).

[37] G. Caliskan, R. M. Briber, D. Thirumalai, V. Garcia-Sakai, S. A. Woodson, and A. P. Sokolov, J. Am. Chem. Soc. **128**, 32 (2006).

[38] H. Frauenfelder, P. W. Fenimore, G. Chen, and B. H. McMahon, Proc. Natl. Acad. Sci. U.S.A. **103**, 15469 (2006).

[39] P. A. Papoian, J. Ulander, M. P. Eastwood, Z. Luthey-Schulten, and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **101**, 3352 (2004).

[40] R. Elber and M. Karplus, Science **235**, 318 (1987).

[41] Y. Joti, A. Kitao, and N. Go, J. Am. Chem. Soc. **127**, 8705 (2005).

[42] D. Vitkup, D. Ringe, G. A. Petsko, and M. Karplus, Nat. Struct. Biol. **7**, 34 (2000).

[43] B. Gillespie and K. W. Plaxco, Proc. Natl. Acad. Sci. U.S.A. **97**, 12014 (2000).

[44] S. Khodadadi, S. Pawlus, J. H. Roh, V. G. Sakai, E. Mamontov, and A. P. Sokolov, J. Chem. Phys. **128**, 195106 (2008).

[45] J. Wang, K. Fan, and W. Wang, Phys. Rev. E **65**, 041925 (2002).

[46] S. E. Jackson, M. Moracci, N. elMasry, C. M. Johnson, and A. R. Fersht, Biochemistry **32**, 11259 (1993).

[47] G. Zuo, J. Hu, and H. Fang, Phys. Rev. E **79**, 031925 (2009).

[48] H. Taketomi, Y. Ueda, and N. Gō, Int. J. Pept. Protein Res. **7**, 445 (1975).

[49] V. S. Pande and D. S. Rokhsar, Proc. Natl. Acad. Sci. U.S.A. **96**, 1273 (1999).

[50] W. Wang, W.-X. Xu, Y. Levy, E. Trizacc, and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **106**, 5517 (2009).

[51] V. Helms, ChemPhysChem **8**, 23 (2007).

[52] Y. Levy and J. N. Onuchic, Annu. Rev. Biophys. Biomol. Struct. **35**, 389 (2006).

[53] J. L. MacCallum, M. S. Moghaddam, H. S. Chan, and D. P. Tieleman, Proc. Natl. Acad. Sci. U.S.A. **104**, 6206 (2007).

[54] S. Vaitheeswaran and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **105**, 17636 (2008).

[55] T. Veitshans, D. Klimov, and D. Thirumalai, Folding Des. **2**, 1 (1997).

[56] D. K. Klimov and D. Thirumalai, Phys. Rev. Lett. **79**, 317 (1997).

[57] B. S. Gerstman and P. P. Chapagain, J. Chem. Phys. **123**, 054901 (2005).

[58] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).

[59] J. E. Shea and C. L. Brooks III, Annu. Rev. Phys. Chem. **52**, 499 (2001).

[60] S. K. Kearsley, Acta Crystallogr., Sect. A: Found. Crystallogr. **45**, 208 (1989).

[61] R. Abseher and M. Nilges, J. Mol. Biol. **279**, 911 (1998).

[62] A. Kitao and N. Gō, Curr. Opin. Struct. Biol. **9**, 164 (1999).

[63] G. G. Maisuradze, A. Liwo, and H. A. Scheraga, Phys. Rev. Lett. **102**, 238102 (2009).

[64] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney *et al.*, *LAPACK Users' Guide*, 3rd ed. (Society for Industrial and Applied Mathematics, Philadelphia, PA, 1999).

[65] J. Wang and W. Wang, Nat. Struct. Biol. **6**, 1033 (1999).

[66] D. W. Mount, *Bioinformatitics: Sequence and Genome Analysis* (Cold Spring Harbor Laboratory Press, New York, 2001), Chap. 9, pp. 421–423.

[67] S. E. Jackson and A. R. Fersht, Biochemistry **30**, 10428 (1991).

[68] M. Lindberg, J. Tångrot, and M. Oliveberg, Nat. Struct. Biol. **9**, 818 (2002).

[69] H. S. Chan, S. Shimizu, and H. Kaya, Methods Enzymol. **380**, 350 (2004).

[70] H. Kaya and H. S. Chan, Proteins **40**, 637 (2000).

[71] H. S. Chan, Proteins **40**, 543 (2000).

[72] S. Takada, J. J. Portman, and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **94**, 2318 (1997).

[73] J. D. Bryngelson and P. G. Wolynes, J. Phys. Chem. **93**, 6902 (1989).

[74] A. Gutin, A. Sali, V. Abkevich, M. Karplus, and E. I. Shakhnovich, J. Chem. Phys. **108**, 6466 (1998).

[75] S. Khodadadi, A. Malkovskiy, A. Kisliuk, and A. P. Sokolov, Biochim. Biophys. Acta **1804**, 15 (2010).