

Relationship between protein folding thermodynamics and the energy landscape

Jaegil Kim,^{*} Thomas Keyes,[†] and John E. Straub[‡]

Department of Chemistry, Boston University, Boston, Massachusetts 02215, USA

(Received 23 September 2008; published 4 March 2009)

The origin of protein folding thermodynamics is examined in terms of the energy landscape, employing an off-lattice protein model with scaled non-native attractions, which is continuously tunable between a Go-like model and a highly frustrated system. Extensive statistical temperature molecular dynamics simulations, combined with inherent structure analysis, reveal the intimate connection between the global geometric properties of the energy landscape and the statistical temperature. The basin depth of the energy landscape is shown to play a key role in the first-order-like characteristics of the statistical temperature, which are easily identified by the squared modulus of the potential energy gradient in the microcanonical ensemble.

DOI: 10.1103/PhysRevE.79.030902

PACS number(s): 87.15.Cc, 05.70.-a, 87.15.A-

Protein folding—i.e., how a polypeptide chain reaches a unique native state for a given sequence of amino acids [1]—is one of the fundamental problems in biological science. While significant advances have been made [2–9], fundamental open questions still remain. For example, one would like to understand the remarkable observation that most single-domain proteins exhibit a cooperative, two-state folding transition with no detectable intermediates [2].

Energy landscape theory [3–6] provides a theoretical framework for protein folding, in which the connection between a proteinlike sequence and a minimally frustrated energy landscape is rationalized via a folding funnel [4]. A good folding sequence is characterized by a smooth potential energy landscape (PEL), with an overall funnel shape biased to the native state, while a poorly designed sequence like a random heteropolymer has a rugged PEL susceptible to long-lived kinetic trapping [7]. Many simulation studies have been devoted to elucidate the determinants of a proteinlike system, such as the ratio of the folding and glass transition temperatures [3], the stability gap between a native state and low-lying excited states [8], and the collapse cooperativity [9].

Recently, Mazzoni and Casetti [10] proposed the fluctuations of the energy landscape curvature, $K_R = \nabla^2 U$, U being the potential energy, as a determinant. A dramatic enhancement of the fluctuations, $\kappa(T) = \sqrt{(\langle K_R^2 \rangle - \langle K_R \rangle^2) / \langle K_R \rangle}$, in the vicinity of the folding temperature $T = T_f$ was observed in the proteinlike folding of well-designed sequences, where $\langle \cdot \rangle$ denotes an equilibrium average. The enhancement in $\kappa(T_f)$ is attributed to effective two-state dynamics between the native and denatured states with different curvatures [10].

Two-state protein folding is analogous to a first-order phase transition in finite-size systems [11–13]. Indeed, a recent study [14] reveals that the “backbending” (S bend) in the statistical temperature, or microcanonical caloric curve, characteristic of a first-order-like phase transition, fully characterizes protein folding and clearly distinguishes good folders from poor ones. The question naturally arises: how does the global geometry of the PEL translate into the behavior of

the statistical temperature? This paper addresses the issue by establishing a rigorous connection between the statistical temperature and various energy landscape measures, leading to the identification of the principal landscape determinants of two-state folding thermodynamics.

The key quantity in our study is the statistical temperature in the microcanonical ensemble, $T_S(U) = [\partial S(U) / \partial U]^{-1}$, $S(U)$ being the microcanonical entropy. It is well known that in a finite-size system such as a protein, the microcanonical ensemble is well suited to the analysis of phase transitions [11–14]. Around a first-order phase transition, $S(U)$ shows a convex dip, with $\partial^2 S / \partial U^2 > 0$, corresponding to backbending in $T_S(U)$ or a negative specific heat. Backbending is associated with an extremum condition in the Helmholtz free energy at T_f , $F(U, T_f) = U - T_f S(U)$, which has three roots U_i ($i = 1-3$), satisfying $T_S(U_i) = T_f$ ($U_1 < U_2 < U_3$). The multiple roots are associated with two-state transition behavior between free-energy minima at U_1 and U_3 , involving the crossing of a transition barrier at U_2 , while the transition becomes downhill-like without a barrier for the monotonic $T_S(U)$. Thus, all essential features of the transitions occurring in good and poor folders are completely characterized by the behavior of $T_S(U)$.

Our analysis linking $T_S(U)$ to the PEL is based on the thermodynamic identity between the statistical temperature and the configurational temperature $T_{conf}(U)$ originally introduced by Rugh [15] and later generalized by Jepps *et al.* [16]:

$$T_S(U) = T_{conf}(U) = \frac{\langle \|\nabla U\|^2 \rangle_U}{\langle \nabla^2 U \rangle_U} = \frac{\mathcal{F}(U)}{\mathcal{K}(U)}, \quad (1)$$

where the subscript “conf” implies that the configurational temperature is evaluated as the fraction of two separate microcanonical averages, $\mathcal{F}(U) = \langle \|\nabla U\|^2 \rangle_U$ and $\mathcal{K}(U) = \langle \nabla^2 U \rangle_U$ [16]. The significance of Eq. (1) is that it allows a rigorous mathematical connection between the purely statistical quantity $T_S(U)$ and the geometric structure of the energy landscape via the force, $\mathbf{F} = -\nabla U$, and the curvature, $K_R = \nabla^2 U$.

All-atom simulations considering the protein and explicit solvent are the most accurate, but still computationally demanding. Thus a simplified description, retaining the main features of real proteins, but reducing the number of degrees

^{*}jaegil@bu.edu

[†]keyes@bu.edu

[‡]straub@bu.edu

of freedom through a coarse graining, is especially appealing for the characterization of protein-folding mechanisms over long time scales. The newly suggested folding determinants [10,14] have been examined in coarse-grained models, with varied sequences exhibiting a different foldability.

Here we employ the scaled Honeycutt-Thirumalai (SHT) off-lattice model [17] to mimic the effect of sequence variations and to examine how two-state folding results from the global shape of the PEL and the emerging backbending behavior of $T_S(U)$. The original HT model [18] has a β -barrel native state and consists of 46 beads with three different types: hydrophobic (B), hydrophilic (L), and neutral (N). We modified the model to incorporate scaled, Go-like interactions by modulating the attractions between non-native BB contacts, which allows, for the original HT sequence $B_9N_3(LB)_4N_3B_9N_3(LB)_5L$, a continuous tuning from a highly frustrated PEL to a weakly rugged funnel. The potential energy is a sum of bond stretching (U_B), bending (U_A), torsion (U_T), and nonbonded interactions ($U_{NB,\lambda}$):

$$U(\mathbf{r};\lambda) = U_B + U_A + U_T + U_{NB,\lambda}, \quad (2)$$

where $U_B = \sum_i k_r (r_i - r_e)^2$, $U_A = \sum_i k_\theta (\theta_i - \theta_e)^2$, and $U_T = \sum_i \{A_i [1 + \cos(\phi_i)] + B_i [1 + \cos(3\phi_i)]\}$. The stretching and bending constants are $k_r = 200\epsilon/\sigma^2$ and $k_\theta = 10\epsilon/\text{rad}^2$, respectively, ϵ and σ being energy and length units, with $r_e = \sigma$ and $\theta_e = 105^\circ$. Torsion parameters are $A_i = B_i = 1.2$ for a quartet of successive beads with zero or one N bead, while otherwise $A_i = 0$ and $B_i = 0.2$.

The nonbonded term in Eq. (2) contains the scaling factor λ ($0 \leq \lambda \leq 1$) to modulate the non-native attractions as $U_{NB,\lambda} = \sum_{i < j-3} [U_{ij} + (\lambda-1)U_{ij}\delta(i,\alpha)\delta(j,\alpha)]$, $U_{ij} = 4\epsilon S_1 [(\sigma/r_{ij})^{12} - S_2(\sigma/r_{ij})^6]$, where $S_1 = S_2 = 1$ for BB pairs, $S_1 = 2/3$ and $S_2 = -1$ for LL and LB pairs, and $S_1 = 1$ and $S_2 = 0$ for all other pairs involving N . Here $\delta(i,\alpha) = 1$ if $\alpha = B$ and bead i belongs to the non-native contact set, and otherwise $\delta(i,\alpha) = 0$. As λ approaches zero, the second term in $U_{NB,\lambda}$ introduces a strong native bias in denatured states by prohibiting close contacts between non-native BB pairs. All 57 native contact pairs are identified with the cutoff 1.3σ , based on the global minimum structure of the original HT model ($U = -49.2635$).

Due to the high degree of frustration in the original HT model [18,19], an enhanced sampling technique is indispensable to obtain a correct conformational sampling for both the transition and low-temperature regions, overcoming ergodicity breaking. We applied the statistical temperature molecular dynamics (STMD) algorithm [19], which generates a flat energy sampling via the dynamic refinement of the statistical temperature (for details see Ref. [19]) with four different scalings at $\lambda = 1.0, 0.9, 0.5$, and 0 . All STMD simulations have been performed to sample the temperature range from 0.07 to 1.3 , with the average kinetic energy fixed at $T_0 = 1.3$. A total of 8×10^4 configurations for each simulation were collected to determine the canonical and microcanonical thermodynamics, and minimized using the conjugate-gradient algorithm to map inherent structures (ISs) [20].

As summarized in Table I, the folding (T_f) and collapse (T_θ) temperatures, identified by peaks in the T -dependent heat capacities and order parameter fluctuations in Figs. 1(a)

TABLE I. Characteristic temperatures of folding (T_f), collapse (T_θ), and curvature fluctuation (T_κ) and the collapse cooperativity $\sigma = (T_\theta - T_f)/T_\theta$ at different $\lambda = 1.0, 0.9, 0.5$, and 0 .

Scaling parameter λ	T_f	T_θ	T_κ	$\sigma = (T_\theta - T_f)/T_\theta$
1.0	0.19	0.67	0.68	0.71
0.9	0.49	0.65	0.64	0.25
0.5	0.61	0.67	0.62	0.09
0.0	0.59	0.62	0.60	0.05

and 1(b), respectively, illustrate that the original HT model (SHT with $\lambda = 1$) is a poor folder and the SHT model with $\lambda = 0.9$ is an intermediate folder, while $\lambda = 0.5$ and 0 correspond to good folders according to the classification based on the collapse cooperativity, $\sigma = (T_\theta - T_f)/T_\theta$ [9]. The validity of this classification is also seen in the profiles of the curvature fluctuations $\kappa(T)$ in Fig. 1(c). A marked peak shows up close to T_f for good folders, while the plots are featureless in poor and intermediate folders [10]. Note that the temperature T_κ , corresponding to the maximum curvature fluctuation in Fig. 1(c), coincides with T_f for good folders, implying that $\kappa(T)$ is also an effective folding indicator.

The statistical temperature $T_S(U)$, determined by STMD in Fig. 2, shows a dramatic change as the energy landscape changes with decreasing λ . The smooth, monotonically varying $T_S(U)$ in the original HT model ($\lambda = 1$) first starts to show a plateau, or weak backbending, and a substantial backbending appears as λ falls below 0.5 . Remarkably, the temperatures corresponding to the plateau and backbending regions in $T_S(U)$ with $\lambda = 0$ exactly match T_f and T_θ , respectively. The alteration of the folding mechanism from second-order-like collapse ($\lambda = 1$) into first-order-like folding ($\lambda = 0$) is clearly seen through the behavior of $T_S(U)$.

We find that the folding in Go-like models occurs through the transient formation of a three-strand core near T_θ and the immediate attachment of the terminal strand to attain the native state at $T_f (\approx T_\theta)$. A significant slowing down is observed in the folding of the original HT model with $T_f \ll T_\theta$ due to trapping in misfolded states. Since U is the external

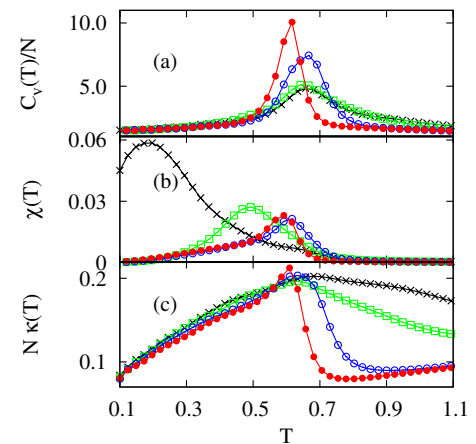


FIG. 1. (Color online) (a) Specific heats $C_v(T)$, (b) order parameter fluctuations $\chi(T)$, and (c) curvature fluctuations $\kappa(T)$ at different $\lambda = 1.0, 0.9, 0.5$, and 0 from bottom to top in (a) at $T = 0.6$.

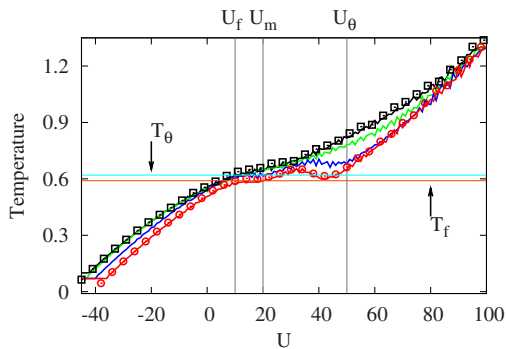


FIG. 2. (Color online) Statistical temperatures $T_S(U)$ (solid lines) and configurational temperatures $T_{conf}(U)$ (squares and circles) with different $\lambda=1.0, 0.9, 0.5,$ and 0 from top to bottom at $U=40$. Note that horizontal lines at $T_\theta=0.62$ and $T_f=0.59$ match the backbending and plateau regions of $T_S(U)$ at $\lambda=0$, respectively.

control parameter in the microcanonical ensemble (playing the role of the temperature in the canonical ensemble), we also identify the characteristic energies to describe the folding transition as a function of U , in the case of $\lambda=0$. In Fig. 2, both $U_f=10$ and $U_m=20$ approximately define the plateau region in $T_S(U)$, and both U_m and $U_\theta=50$ bound the collapse region associated with the backbending of $T_S(U)$, $T_S(U_\theta)=0.65$. The good folder with $\lambda=0$ starts to collapse as U crosses U_θ , and completes its folding once U passes through U_f .

Assuming that STMD samples all configurations with a given U , the microcanonical average in Eq. (1) is directly computed by binning in energy and collecting configurations corresponding to each bin. The configurational temperature $T_{conf}(U)$ with unit bin size shows good agreement with $T_S(U)$ in both $\lambda=1$ and 0 , illustrating that the thermodynamic identity, Eq. (1), is robust even with a smaller system size.

To illustrate how the PEL controls $T_S(U)$, and hence the folding, via Eq. (1), both $\mathcal{K}(U)$ and $\mathcal{F}(U)$ are plotted in Figs. 3(a) and 3(b), respectively. For all scalings, the curvature $\mathcal{K}(U)$ monotonically increases with decreasing U , showing that folded (denatured) states are associated with steeper

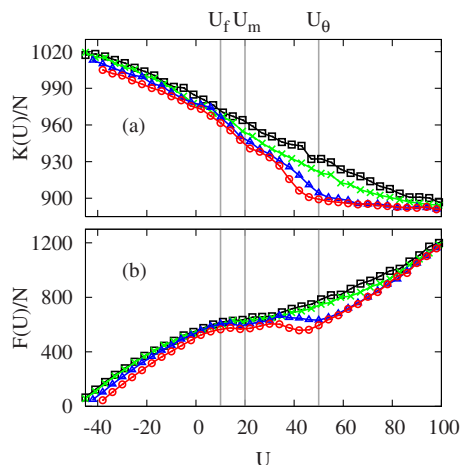


FIG. 3. (Color online) Microcanonical average of (a) $\mathcal{K}(U)$ and (b) $\mathcal{F}(U)$ with different $\lambda=1.0, 0.9, 0.5,$ and 0 from top to bottom around $U=40$.

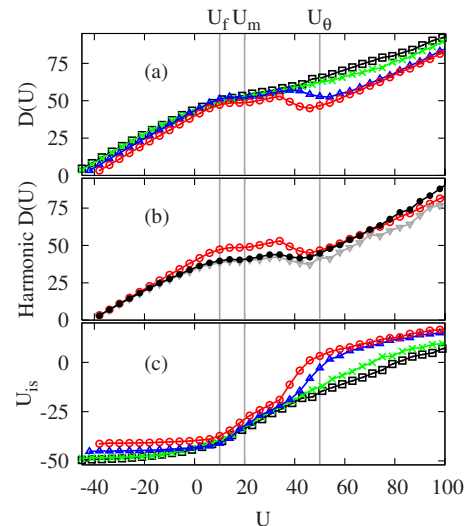


FIG. 4. (Color online) (a) Basin depth $\mathcal{D}(U)$, (b) harmonic basin depths $\mathcal{D}_{har}(U)$ (solid circles) and $\mathcal{D}_{har}^T(U)$ (triangles) at $\lambda=0$, and (c) average IS energy $U_{IS}(U)$. $\lambda=1.0, 0.9, 0.5, 0$ from top to bottom in (a).

(shallow) potential wells. The primary feature differentiating good from poor folders in Fig. 3(a) is that good folders have nearly constant curvature for $U > U_\theta$, suggesting that their denatured states are structurally homogeneous. This finding is in accordance with our expectations for Go-like interactions. The main frustration of the original HT model arises from the conformational diversity of the collapsed states [18]. The native-centric $U_{NB,\lambda}$ in Go-like models significantly suppresses the frustration from hydrophobic mismatches of β strands during the collapse by reducing non-native interstrand BB contacts and also makes the unfolded landscape less structured. With decreasing U in good folders, the flat $\mathcal{K}(U)$ begins to increase steeply between U_f and U_θ , the backbending and plateau regions of $T_S(U)$, contributing to the large curvature fluctuation $\kappa(T_f)$ and to the bimodal energy distributions [10]. By contrast, in poor folders the slope of $\mathcal{K}(U)$ is nearly constant, causing no structure in $T_S(U)$ and corresponding to the fact that all systems studied have nearly identical $\mathcal{K}(U)$ at the highest and lowest U .

Turning to $\mathcal{F}(U)$, the most remarkable aspect of Fig. 3(b) is that $\mathcal{F}(U)$ retains all essential features of $T_S(U)$. Despite the discussion in the last paragraph, the dynamic range of $\mathcal{K}(U)$ is small and the variation of $T_S(U)$ is mostly encoded into $\mathcal{F}(U)$. What is the underlying physics relating $\mathcal{F}(U)$ to $T_S(U)$? Some insight is gained by noting that $\mathcal{F}(U)$ is a measure of the geometric distance between the sampled configurations at U and their associated ISs [20]. Indeed, the microcanonical basin depth of the PEL, quantified as $\mathcal{D}(U) = U - U_{IS}(U)$, $U_{IS}(U) = \langle U_{IS} \rangle_U$ being the average IS energy, shows a remarkable resemblance to $T_S(U)$ in Fig. 4(a). More quantitatively, in a harmonic approximation, the canonical basin depth obeys $\mathcal{D}_{har}^T(T) = N_f T / 2$, N_f being the number of degrees of freedom. Replacing the temperature T by $T_S(U)$ gives $\mathcal{D}_{har}^T(U) \approx (N_f / 2) T_S(U) = (N_f / 2) \frac{\mathcal{F}(U)}{\mathcal{K}(U)}$, which unveils a direct link translating the global geometry of the PEL into the proteinlike folding markers of $T_S(U)$. We checked the validity of the harmonic approximation by com-

paring $\mathcal{D}(U)$ and its harmonic counterparts, $\mathcal{D}_{har}^T(U)$ and $\mathcal{D}_{har}(U) = \frac{1}{2} \langle \nabla U \cdot \delta \mathbf{r} \rangle_U$ [21]. As shown in Fig. 4(b), both independently derived $\mathcal{D}_{har}^T(U)$ and $\mathcal{D}_{har}(U)$ display the essential characteristics of $\mathcal{D}(U)$.

The basin depth of the PEL as a function of U is most directly seen in the profile of $U_{IS}(U)$ in Fig. 4(c). A sharp drop in $U_{IS}(U)$, to a constant value, occurs for good folders between U_θ and U_f . This means that the system quickly reaches the native state basin as U crosses U_f , while $U_{IS}(U)$ shows a slow decay due to residual frustrations among collapsed states in poor and intermediate folders. The plateau regions in $U_{IS}(U)$ for $U > U_\theta$ reflect the homogeneous inherent structures in characteristic of the denatured states of good folders as observed in $\mathcal{K}(U)$ in Fig. 3(a). Our analysis suggests that the Go-like model has a near-Levinthal-like landscape [22], in which the denatured state landscape is gently

slanted, while the folded state landscape has a steep funnel to the native state, resulting from the native-centric interactions stabilizing the native state while homogenizing the denatured states.

In summary, we have shown that a sharp variation in the basin depth, $\mathcal{D}(U)$, from the denatured states to the native state, gives rise to the first-order-like folding in good folders. The basin depth is reflected in the statistical temperature $T_S(U)$ through the relations $T_S(U) = \mathcal{F}(U) / \mathcal{K}(U) \approx (2/N_f) \mathcal{D}(U)$. Two-state protein folding is associated with marked transitions of the potential energy gradient $\mathcal{F}(U)$, the average IS energy $U_{IS}(U)$, and the curvature $\mathcal{K}(U)$.

We are grateful to the National Science Foundation (Grant No. CHE-0750309) and the National Institutes of Health (Grant No. RO1 GM076688) for the generous support of our research.

-
- [1] C. Anfinsen, *Science* **181**, 223 (1973).
 [2] A. R. Fersht, *Curr. Opin. Struct. Biol.* **7**, 3 (1997); S. E. Jackson, *Folding Des.* **3**, R81 (1998).
 [3] J. D. Bryngelson and P. G. Wolynes, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7524 (1987).
 [4] P. E. Leopold, M. Montal, and J. N. Onuchic, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8721 (1992).
 [5] D. J. Wales, *Energy Landscapes* (Cambridge University Press, Cambridge, England, 2003).
 [6] K. A. Dill and H. S. Chan, *Nat. Struct. Biol.* **4**, 10 (1997).
 [7] H. Nymeyer, A. E. Garcia, and J. N. Onuchic, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5921 (1998).
 [8] A. Sali, E. I. Shakhnovich, and M. Karplus, *Nature (London)* **369**, 248 (1994).
 [9] D. K. Klimov and D. Thirumalai, *Phys. Rev. Lett.* **76**, 4070 (1996).
 [10] L. N. Mazzoni and L. Casetti, *Phys. Rev. Lett.* **97**, 218104 (2006); *Phys. Rev. E* **77**, 051917 (2008).
 [11] D. H. E. Gross and J. F. Kenney, *J. Chem. Phys.* **122**, 224111 (2005).
 [12] D. J. Wales and R. S. Berry, *Phys. Rev. Lett.* **73**, 2875 (1994).
 [13] C. Junghans, M. Bachmann, and W. Janke, *Phys. Rev. Lett.* **97**, 218103 (2006).
 [14] J. Hernandez-Rojas and J. M. Gomez Llorente, *Phys. Rev. Lett.* **100**, 258104 (2008).
 [15] H. H. Rugh, *Phys. Rev. Lett.* **78**, 772 (1997).
 [16] O. G. Jepps, G. Ayton, and D. J. Evans, *Phys. Rev. E* **62**, 4757 (2000).
 [17] J. Kim and T. Keyes, *J. Phys. Chem. B* **112**, 954 (2008).
 [18] J. D. Honeycutt and D. Thirumalai, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 3526 (1990).
 [19] J. Kim, J. E. Straub, and T. Keyes, *Phys. Rev. Lett.* **97**, 050601 (2006); *J. Chem. Phys.* **126**, 135101 (2007).
 [20] F. H. Stillinger and T. A. Weber, *Phys. Rev. A* **28**, 2408 (1983); *Science* **225**, 983 (1984).
 [21] In a harmonic approximation to the PEL around each IS configuration \mathbf{r}_{IS} , $U_{har}(\mathbf{r}) = U_{IS} + \frac{1}{2} \delta \mathbf{r}^T \mathbf{H}_{IS} \delta \mathbf{r}$, $\delta \mathbf{r} = \mathbf{r} - \mathbf{r}_{IS}$ and \mathbf{H}_{IS} is the Hessian at the IS. Denoting $\nabla U_{har} = \mathbf{H}_{IS} \delta \mathbf{r}$, $D_{har}(U) = \langle U_{har}(\mathbf{r}) \rangle_U - U_{is} = \frac{1}{2} \langle \nabla U_{har} \cdot \delta \mathbf{r} \rangle_U \approx \frac{1}{2} \langle \nabla U \cdot \delta \mathbf{r} \rangle_U$.
 [22] H. Kaya and H. S. Chan, *Proteins: Struct., Funct., Genet.* **52**, 510 (2003).