SCIENCE CHINA
Life Sciences

• RESEARCH PAPER •

December 2014 Vol.57 No.12: 1197–1212
doi: 10.1007/s11427-014-4728-9

Statistical analyses of protein folding rates from the view of quantum transition

LV Jun[1*] & LUO LiaoFu[2]

[1]*Center for Physics Experiment, College of Science, Inner Mongolia University of Technology, Hohhot 010051, China;*
[2]*Faculty of Physical Science and Technology, Inner Mongolia University, Hohhot 010021, China*

Received November 21, 2013; accepted July 19, 2014; published online September 26, 2014

Understanding protein folding rate is the primary key to unlock the fundamental physics underlying protein structure and its folding mechanism. Especially, the temperature dependence of the folding rate remains unsolved in the literature. Starting from the assumption that protein folding is an event of quantum transition between molecular conformations, we calculated the folding rate for all two-state proteins in a database and studied their temperature dependencies. The non-Arrhenius temperature relation for 16 proteins, whose experimental data had previously been available, was successfully interpreted by comparing the Arrhenius plot with the first-principle calculation. A statistical formula for the prediction of two-state protein folding rate was proposed based on quantum folding theory. The statistical comparisons of the folding rates for 65 two-state proteins were carried out, and the theoretical vs. experimental correlation coefficient was 0.73. Moreover, the maximum and the minimum folding rates given by the theory were consistent with the experimental results.

quantum folding, protein folding rate, temperature dependence, number of torsion mode, folding free energy

Citation: Lv J, Luo LF. Statistical analyses of protein folding rates from the view of quantum transition. Sci China Life Sci, 2014, 57: 1197−1212, doi: 10.1007/s11427-014-4728-9

It is well known that a protein chain can spontaneously fold into its unique native structure [1,2]. To paraphrase Levinthal's paradox, if a protein were to attain its correctly folded configuration by sequentially sampling all the possible conformations, it would require a period of time longer than the age of the universe to arrive at its correct native conformation [3]. Apart from theory, it is a fact that experimentally measured times for spontaneous folding of single-domain globular proteins range from microseconds [4–6] to tens of minutes [7]. Thus, how configurations of proteins are determined and what makes them fold so quickly are questions that constitute a longstanding puzzle in molecular biology. While two prominent models have been proposed to study the protein folding mechanism, folding nucleus [8,9] and folding tunnel [10−14], the importance of topology and contact order in protein folding has been recognized over the last 15 years, and many new models to predict the protein folding rate have been published [15–29].

Curiously, it is notable that the rate at which proteins fold is highly sensitive to temperature, showing non-Arrhenius behavior, i.e., the temperature dependence of the rate constant is, in fact, not exponential for these reactions. The nonlinearity of logarithm folding rate on temperature $1/T$ has been conventionally interpreted by the nonlinear temperature dependence of the configurational diffusion constant on rough energy landscapes [30] or by the temperature dependence of hydrophobic interaction [31,32]. Another model was proposed more recently to interpret the difference between folding and unfolding by introducing the number of denatured conformation depending on temperature [33]. Recent experimental data indicated very different and unusual temperature dependencies of the folding rates

*Corresponding author (email: lujun@imut.edu.cn)

existing in the system of λ_{6-85} mutants [34] and in some *de novo* designed ultrafast folding protein [35]. These unusual Arrhenius plots of ultrafast folders provide an additional kinetic signature for protein folding.

Protein folding mechanism is commonly studied by use of classical molecular dynamics (MD) simulation as a main theoretical tool. However, when we observe the folding event at the molecular level, the application of quantum theory should be more reasonable. In fact, the classical approach is simply too limited to realize a complete solution of the protein folding problem, especially when the key goal is understanding of the fundamental physics underlying the folding mechanism. It is well known that fluorescence and phosphorescence are phenomena closely related to protein folding. However, since such phenomena can only be understood in the context of the quantum transition between molecules, the study of protein folding should not be divorced from the framework of quantum folding theory. Although the role of quantum decoherence was widely recognized in the last decades, the estimate of decoherence time for torsion degrees of freedom of a macromolecule does not preclude the possibility of its maintenance of quantum nature [36]. In recent years the quantum mechanical calculation for three-atom and four-atom reactions was worked out and the rate constant as a function of temperature was deduced [37]. However, to generalize the quantum calculation to the multi-atom system is still a difficult task. That protein folding is essentially a quantum transition between torsion states was proposed in recent work [36,38–41]. Based on this essential argument, the present work aims to apply the quantum transition theory to study the folding rate of two-state proteins, making a comprehensive analysis of the existing rate data and bringing more insight to the widely distributed and uniquely temperature-dependent folding rates.

1 Materials and methods

1.1 Datasets

Recently Garbuzynskiy and coworkers [42] collected folding rate data for 107 proteins—69 two-state proteins and 38 multistate proteins. Of the 69 two-state proteins, four (PDB code 1VII, 2PDD, 1PRB and 2A3D, respectively) will not be considered in our study because their folding experiments were carried out at high temperature, and extrapolation of these experimental results to 25°C inevitably contains a large error. The remaining 65 two-state proteins whose folding experiments were carried out at around 25°C constitute the dataset we used to compare the theoretical vs. experimental results (see details of the dataset as listed in Table 1). On the other hand, in studying the temperature dependence of folding rate, we used the experimental data of 16 proteins collected by Ghosh et al. [33] (Table 2).

Table 1 The folding rates and structural parameters for 65 two-state proteins[a)]

PDB code	Experimental data						Calculated parameters	
	Structure class	$\Delta G/RT^{\#}$	$\ln k_f^{\#}$	L^{*}	$L_\alpha{}^{\&}$	$L_\beta{}^{\&}$	$F^{\$}$	$N^{\$}$
1aps	α+β	7.4	−1.6	98	18	36	1	382
1avz	β	9.2	4.9	57	0	17	1	211
1ayi	α	4.9	7.2	85	47	0	25	332
1ba5	α	4.7	5.9	53	33	0	81	208
1bdd	α	7.5	11.7	58	36	0	81	212
1bf4	β	10.2	7.0	63	9	33	1	258
1c9o	β	7.6	7.2	66	0	40	1	254
1csp	β	4.2	6.5	67	0	36	1	255
1cun_16	α	10.8	4.8	106	91	0	81	442
1cun_17	α	11.2	3.4	107	94	0	81	424
1div_c	α+β	11.2	3.3	92	27	35	1	354
1div_n	α+β	6.5	6.6	56	19	11	1	229
1e0g	α+β	4.7	7.0	48	15	8	1	186
1e0l	β	2.0	10.6	37	0	12	1	128
1e0m	β	2.8	8.9	37	0	13	1	133
1e41	α	11.8	6.8	104	70	0	81	394
1fex	α	5.3	8.2	59	31	0	25	225
1fkb	α+β	9.7	1.6	107	8	41	1	413
1fnf_9	β	2.0	–0.9	90	0	49	1	335
1ftg	α/β	5.5	2.8	168	47	31	1	635
1g6p	β	10.3	6.3	66	0	28	1	263
1idy	α	7.0	8.7	54	30	0	25	184
1iet	α+β	4.4	3.0	94	6	6	1	346
1imq	α	9.2	7.3	85	45	0	25	315
1jmq	β	1.7	8.4	40	0	6	1	152
1jo8	β	5.2	2.5	58	0	23	1	221
1jyg	α	4.6	9.1	69	45	0	81	286
1k0s	β	19.5	7.4	143	14	63	1	565

(*To be continued on the next page*)

(*Continued*)

PDB code	Experimental data						Calculated parameters	
	Structure class	$\Delta G/RT^{\#}$	$\ln k_f^{\#}$	L^{*}	$L_\alpha{}^{\&}$	$L_\beta{}^{\&}$	$F^{\$}$	$N^{\$}$
1k85	β	8.4	1.4	88	0	47	1	269
1k8m	β	8.7	–0.9	93	0	33	1	352
1l2y	α	1.0	12.5	20	7	0	25	73
1l8w	α	10.5	2.0	271	154	4	25	947
1lmb	α	7.2	10.4	80	57	0	81	307
1lop	β	17	6.6	164	20	56	1	624
1m9s	β	5.7	4.0	76	0	13	1	301
1mjc	β	3.9	5.3	69	0	33	1	244
1n88	α+β	5.9	2.0	96	22	27	1	372
1nti	α	10.9	7.0	86	53	0	81	345
1o6x	α+β	7.2	6.8	81	22	15	1	275
1pgb_a	α+β	8.0	6.3	56	14	24	1	208
1pgb_b	β	0	12.0	16	0	10	1	58
1pin	β	5.1	9.3	34	0	12	1	129
1pnj	β	6.3	–1.0	86	0	24	1	301
1poh	α+β	8.9	2.7	85	29	23	1	320
1prs_c	β	7.3	–2.0	83	4	23	1	318
1prs_n	β	12.4	3.0	90	7	32	1	326
1qtu	β	11	0	109	6	53	1	426
1rfa	α+β	11.2	8.4	78	12	21	1	307
1ris	α+β	14.4	6.1	97	28	47	1	392
1shg	β	5.9	1.1	57	0	26	1	230
1spr	α+β	12.2	8.7	103	18	29	1	396
1srm	β	5.7	4.4	56	0	16	1	209
1t8j	α	–1.4	11.8	23	8	0	25	86
1ten	β	8.8	1.1	89	0	48	1	343
1u5p	α	10.7	11.0	110	93	0	81	441
1ubq	α+β	14.1	7.3	76	12	24	1	299
1urn	α+β	16.3	4.6	96	28	24	1	379
1w4j	α	6.0	12.3	51	23	0	25	164
1wit	β	8.6	0.4	93	0	48	1	349
256b	α	17.3	12.3	106	81	0	81	424
2acy	α+β	7.3	0.8	98	24	41	1	379
2ci2	α+β	16.1	5.8	64	11	14	1	250
2ptl	α+β	7.4	4.1	60	12	24	1	220
2wxc	α	4.6	11.2	47	20	0	25	153
a-helix	α	0	15.5	21	20	0	81	42

a) #, ΔG(J mol^{-1})/RT=ΔG(J)/k_BT where ΔG(J) is folding free energy per molecule defined by unfolded initial state energy minus folded final state energy, $\ln k_f$ the logarithm folding rate (k_f in unit s^{-1}). Both data are taken from [42]. *, The chain length L means the number of folded residues according to PDB. &, Secondary structure was assigned from Protein Data Bank [43] coordinates of proteins by using the program dssp ([44], http://swift.cmbi.ru.nl/gv/dssp/), which marks helical residues by symbols H and β-structural residues by symbols E. L_α means the number of α helical residues, and L_β the number of β-structural residues. \$, N is the number of torsion modes calculated following the rule given in section 1.2. F is the structural class parameter defined in the text eq. (11) and calculated following the rule: F=81 for $(L_\alpha-L_\beta)/L\geqslant 0.6$, 25 for $0.3\leqslant(L_\alpha-L_\beta)/L<0.6$, and 1 for $(L_\alpha-L_\beta)/L<0.3$. L_α and L_β are the number of residues in α helix and β sheet, respectively, and L is the number of folded residues according to PDB.

Table 2 Data of temperature dependence of protein folding rate

Protein short name	PDB code	Mutants	T_c	T_f	$\Delta G_f/k_BT_f$	Reference
BdpA	1bdd	F13W/G29A	350	283	10.37	[45]
NTL9	1div$_n$	Wild type	350	283	7.81	[46]
FBP28	1e0l	ΔNΔC-Y11R/W30F	327	298	3.43	[47]
En-HD	1enh	Wild type	325	298	3.55	[6]
Apocytochrome b_5	1iet	Wild type	319	283	4.40	[48]
Trp^2-cage	1l2y	P12W	331	296	2.41	[49]
Trp-cage	1l2y	Wild type	315	284	1.91	[4]
λ-repressor	1lmb	Y22W	334	313	4.77	[34]
λ-repressor	1lmb	Y22W/A37G	328	313	2.80	[34]
λ-repressor	1lmb	Y22W/G46A/G48A	341	328	2.94	[34]
Pin WW domain	1pin	Wild type	332	313	3.01	[50]
Pin WW domain	1pin	S18G	330	312	2.65	[50]
Pin WW domain	1pin	N26D	311	302	1.35	[50]
Prb$_{7\text{-}53}$	1prb	K5I/K39V	372	348	4.06	[51]
α3D	2a3d	Wild type	346	318	3.45	[35]
Psbd41	2pdd	Wild type	326	314	1.75	[52]

1.2 Quantum folding model

We investigated protein folding rate as the quantum transition between torsion states on polypeptide chain. Compared with other dynamic variables, such as mobile electrons, chemical bonds and stretching-bending vibrations, also called fast variables, molecular torsion has the lowest energy and can be viewed as the slow variable of the macromolecular system. Assuming that "the slow variables slave the fast ones" and using the nonadiabaticity operator method, a formula for protein folding rate in analytical form was deduced in previous results [36,39], as

$$W_f = \frac{2\pi}{\hbar^2 \bar{\omega}'} I'_V I'_E,$$

$$I'_V = \frac{\hbar}{\sqrt{2\pi}\delta\theta} e^{\frac{\Delta G}{2k_B T}} e^{\frac{-(\Delta G)^2}{2\bar{\omega}^2 (\delta\theta)^2 k_B T \sum_j^N I_j}} (k_B T)^{1/2} \left(\sum_j^N I_j \right)^{-1/2}, \quad (1)$$

$$I'_E = \sum_j^M \left| a^{(j)}_{\alpha'\alpha} \right|^2 \cong M\bar{a}^2,$$

where W means protein folding rate at given temperature and denaturant concentration, I'_V is slow-variable factor and I'_E fast-variable factor, N is the number of torsion modes participating in a quantum transition coherently, I_j denotes the inertial moment of the atomic group of the j-th torsion mode (I_0 denotes its average), $\bar{\omega}$ and $\bar{\omega}'$ are the initial and final frequency parameters ω_j and ω'_j of torsion potential averaged over N torsion modes, respectively, $\delta\theta$ is the averaged angular shift between initial and final torsion potential (Figure 1), M is the number of torsion angles correlated to fast variables, $\bar{a}^2$ is the square of the matrix element of the fast-variable Hamiltonian operator, or, more accurately, its change with torsion angle, averaged over M modes, k_B is Boltzmann constant, T is absolute temperature and ΔG is the free energy decrease per molecule between initial and final states,

$$\Delta G = \Delta E + \lambda k_B T \quad \left(\lambda = \sum_{j=1}^N \ln \frac{\omega_j}{\omega'_j} \right), \quad (2)$$

where $\Delta E = \sum_{j=1}^N \delta E_j$, δE_j is the energy gap between initial and final states for the j-th mode. The substantial argument for the derivation of eq. (1) is the assumption on the torsion as slow variables and the adiabatic approximation applicable.

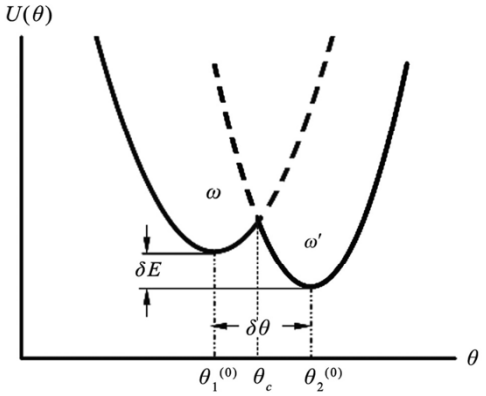

Figure 1 Typical potential curve for a given torsion mode j. The subscript j has been omitted for brevity. The case given in figure is for $\delta E>0$ and $\omega<\omega'$. The quantum state in the potential is defined by (k, n) where k=1,2 indicates the wave function localized around 1st or 2nd minimum of the potential and n refers to the vibration energy level.

To obtain quantitative result from eq. (1) one should calculate the number of torsion modes N in advance. N describes the coherence degree of multi-torsion transition in the folding. Based on the idea that the two-state protein folding is equivalent to a quantum conformational transition we assume that N is obtained by the numeration of all main-chain and side-chain dihedral angles on the polypeptide chain except those residues on its tail which does not belong to any contact. A contact is defined by a pair of residues at least four residues apart in their primary sequence and with their spatial distance no greater than 0.65 nm. Each residue in such contact fragment contributes two main-chain dihedral angles and, for non-alanine and -glycine, it contributes 1–4 additional side-chain dihedral angles [53]. To avoid repetitive enumeration, we assume that n=polypeptide chain length minus residues not contained in any contact fragment. Thus, the total number of main-chain dihedral angles in the polypeptide chain is 2n. The number of side-chain dihedral angles n' can be enumerated in the same way (Table 3, Figure 2).

Based on eq. (1), the following two problems on protein folding rate were studied.

1.2.1 Temperature dependence of protein folding rate

The temperature dependence of the transition rate reflects the folding dynamics of a protein. In principle, for any given protein, the problem can be solved starting from eqs.

Table 3 The number of side-chain dihedral angles for 20 amino acids

Amino acids	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile
n'	0	4	2	2	1	3	3	0	2	2
Amino acids	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
n'	2	4	3	2	2	1	1	2	2	1

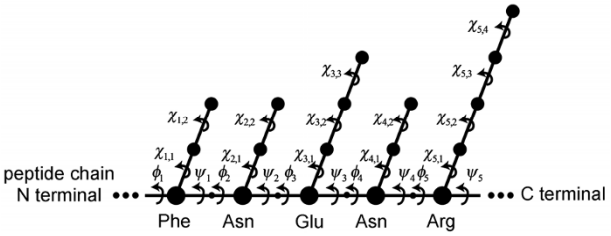

Figure 2 The sketch map of torsion angles (dihedrals) for a typical peptide chain 1ENH. The 20th amino acid Phe and the 24th amino acid Arg are a pair of contact residues of 1ENH. Five residues Phe, Asn, Glu, Asn and Arg compose a contact fragment. The main chain dihedrals φ_i and ψ_i for the i-th amino acid and the side chain dihedrals χ_{ij} (j=1,…,4) are labeled in the diagram. The total number of dihedral angles in this example is 23.

(1) and (2). In addition to the explicit T dependence in these equations, we should consider the association between free energy change ΔG and temperature. Assuming that the torsion potential is susceptible to temperature at melting point (T_c) since a protein may undergo a transition of structure near melting temperature [54], and setting

$$\Delta E(T) = \Delta E(T_c) + m(T - T_c), \tag{3}$$

$$\eta = \frac{\Delta E(T_c) - mT_c}{\Delta E(T_c)}, \tag{4}$$

we obtain

$$\Delta E(T_c) = \frac{\Delta G_f - k_B T_f \lambda}{\eta + (1-\eta)T_f / T_c}, \tag{5}$$

where ΔG_f is the measured value of folding free energy decrease at temperature T_f. Inserting eqs. (3)–(5) into eqs. (1) and (2), we obtain

$$\ln W_f = \frac{S}{T} + \frac{1}{2}\ln T - RT + const, \tag{6}$$

where S is the slope parameter and R—its modification on Arrhenius plot and *const* means a constant independent of temperature T,

$$S = \frac{\eta \Delta E(T_c)\left(1 - \dfrac{\eta \Delta E(T_c)}{\varepsilon}\right)}{2k_B}, \tag{7}$$

$$R = \frac{1}{2\varepsilon k_B T_f^2}[\Delta G_f - \eta \Delta E(T_c)]^2, \tag{8}$$

with $\varepsilon = \bar{\omega}^2(\delta\theta)^2 \sum I_j = NI_0\bar{\omega}^2(\delta\theta)^2$.

Thus, the relationship between rate slope and temperature on Arrhenius plot is

$$\frac{\mathrm{d}\ln W}{\mathrm{d}(1/T)} = S - \frac{1}{2}T + RT^2. \tag{9}$$

In a previous work [55] we deduced a similar relation for a simple system consisting of molecular torsions and only electrons as fast variables. It was demonstrated that, because the temperature dependences of folding rate caused mainly by torsion motion but not by fast-variables, only small changes occur (namely, the symbol of 1/2T term in eq. (9)) between two results and no important modification can be observed on the Arrhenius plot [36].

1.2.2 Statistical investigations on the folding rates of 65 two-state proteins

To make the statistical investigation of the folding rate for two-state proteins, based on eqs. (1) and (2) a model on the folding rate can be established through the following steps.

(i) Study the relationship between folding free energy ΔG and N. At a given temperature, it was reported that the folding free energy ΔG of a polypeptide chain is approximately proportional to chain length [42,56]. Accordingly, we investigated ΔG with $\sqrt{N}$ for 65 two-state proteins and found a good linear relationship existing in these two quantities:

$$\frac{\Delta G}{k_B T} = b\sqrt{N} - c, \tag{10}$$

with b=0.709, c=4.08. Here ΔG (free energy decrease per molecule) is taken from the measured value $\Delta G/(k_BT)=\ln k_f - \ln k_u$ in the literature (k_f and k_u are experimental folding and unfolding rate, respectively) [42] at $T \neq T_c$ with T lower than T_c by an amount not too small. The data of ΔG and N for 65 proteins can be found in Table 1.

(ii) Study the relationship of fast-variable factor $M\bar{a}^2$ with N. The fast-variable factor $M\bar{a}^2$ is generally related to N. Detailed analysis indicates that it also depends on the structural class of the protein, e.g., the secondary content of the protein. We assume it takes a form of

$$M\bar{a}^2 = c_u F N^{-d}, \tag{11}$$

where c_u is a constant in the database which will be absorbed in c_0 of the next equation (12), F is a factor related to the secondary structure content of the protein, and d is a dimensional parameter describing how the fast-variable is related to torsion potential.

(iii) Give an expression for the folding rate of any two-state protein in the database. Inserting eqs. (10) and (11) into eq. (1), we obtain the relationship between folding rate and torsion mode number N, as

$$\ln W = \frac{b}{2}N^{0.5} + \frac{bc}{\rho}N^{-0.5} - \frac{c^2}{2\rho}N^{-1} - \frac{1}{2}\ln N + \ln(c_0 F N^{-d}). \tag{12}$$

Here, $\rho = I_0\bar{\omega}^2(\delta\theta)^2/(k_BT)$ is a torsion energy related parameter and c_0 is a constant independent of N, determined

by the minimization of the sum of the errors of folding rate between theoretical and experimental values for all proteins in the database, $\ln c_0 = \ln\left(\frac{\sqrt{2\pi}}{\hbar}\left\langle\frac{\bar{\omega}}{\bar{\omega}'}\right\rangle c_u\right) - \frac{1}{2}\ln\rho - \frac{b^2}{2\rho} - \frac{c}{2}$ ($\left\langle\frac{\bar{\omega}}{\bar{\omega}'}\right\rangle$ –the frequency ratio averaged over proteins).

2 Results

2.1 Results of temperature dependence of folding rate for 16 typical proteins

We have predicted a universal protein folding rate/temperature relationship given by eq. (6) or (9) on the Arrhenius plot. As is well known, the experiments on protein folding rate/temperature relationships exhibit the following characteristics of non-Arrhenius behavior. The folding rate universally decreases upon increase of temperature, and even the crossover occurs at high temperature from normal positive barrier to abnormal negative [4,6,34,35,49–56]. These characteristics can all be explained by the temperature-dependent terms in eq. (6). The last term, RT^2 in eq. (9) is the main terms contributing to curvature of the Arrhenius plot. To make a more quantitative comparison between theory and experimentation, we studied 16 proteins with currently available experimental data on temperature-dependent folding rate and folding free energy (Table 2). We find that eq. (6) is in good agreement with the rate/ temperature dependence for each protein (Figure 3). It is not surprising that the non-Arrhenius dependence occurs in the protein folding rate since $\ln W$ given by eq. (1) contains the square free energy term $(\Delta G)^2$ in addition to the linear term ΔG. Therefore, the curious non-Arrhenius temperature-dependence has been successfully explained in the proposed quantum folding model.

In our model the universal non-Arrhenius characteristics of folding rate are described by two slope parameters S and R and these parameters are related to the known folding dynamics. Through solving eqs. (7) and (8), we obtain $\eta\Delta E(T_c)$ and ε (or $\delta\theta\bar{\omega}$) for each protein since S and R have been determined by temperature-dependent folding rates, and the free energies ΔG_f have been measured at some temperature T_f. Then, for given η, the energy gap parameter $\Delta E(T_c)$ is obtained, and the frequency-ratio parameter λ is deduced. Thus, all parameters related to torsion potential defined in this theory can be determined. The results are summarized in Table 4. We notice that all conformational potential parameters can be calculated consistently with each other for all studied proteins. These torsion parameters will be able to give deeper insights into the understanding of folding mechanism.

Table 4 The folding temperature dependence and related torsion potential parameters for 16 proteins[a)]

PDB code	S	R	$\frac{\eta\Delta E(T_c)}{k_B T_f}$	$\frac{\varepsilon}{k_B T_f}$	$\delta\theta\bar{\omega}(\times 10^{11})$	$\lambda_{\eta=5.5}$	N
1bdd	–24669	0.2441	48.76	10.66	1.4153	–6.12	212
1div	–32076	0.2930	37.61	5.35	0.9649	–4.91	229
1e0l	–16241	0.1780	28.49	5.91	1.5995	–3.80	97
1enh	–33182	0.3345	31.23	3.84	0.8424	–4.25	227
1iet	–70322	0.7322	37.09	2.58	0.5463	–5.77	346
1l2y(p12w)	–14796	0.1602	20.98	3.64	1.4398	–3.21	73
1l2y(wt)	–18957	0.1774	11.44	0.90	0.7023	–1.10	73
1lmb(wt)	–83920	0.7613	49.06	4.11	0.7684	–6.67	307
1lmb(g46a)	–30292	0.3313	44.99	2.64	0.6155	–7.03	307
1lmb(sa37g)	–112897	1.0766	56.76	13.34	1.4154	–9.13	307
1pin(wt)	–69675	0.6812	44.28	4.00	1.1677	–7.17	129
1pin(s18g)	–77113	0.7565	42.79	3.41	1.0819	–7.02	128
1pin(n26d)	–27063	0.2990	23.64	2.75	0.9520	–3.52	129
1prb	–47886	0.3893	47.30	6.92	1.3737	–7.14	179
2a3d	–18486	0.1812	30.46	6.33	1.0186	–4.11	273
2pdd	–159403	1.5560	44.50	1.87	0.7321	–7.62	152

a) Column 1 gives PDB code of each protein. S and R in columns 2 and 3 are best-fit slope parameter of the folding temperature dependence. Columns 4–7 are torsion potential parameters which are calculated from eqs. (7) and (8). Column 7 gives λ at $\eta = 5.5$ that is near the estimate of frequency-ratio $\bar{\omega}/\bar{\omega}'$ from unfolding rate data. Column 8 gives the number of torsion modes of the polypeptide chain. In all calculations the average torsion inertial moment of atomic groups in polypeptide $\langle I_j \rangle = I_0 = 10^{-44}\,\mathrm{kg\,m^2}$ is assumed.

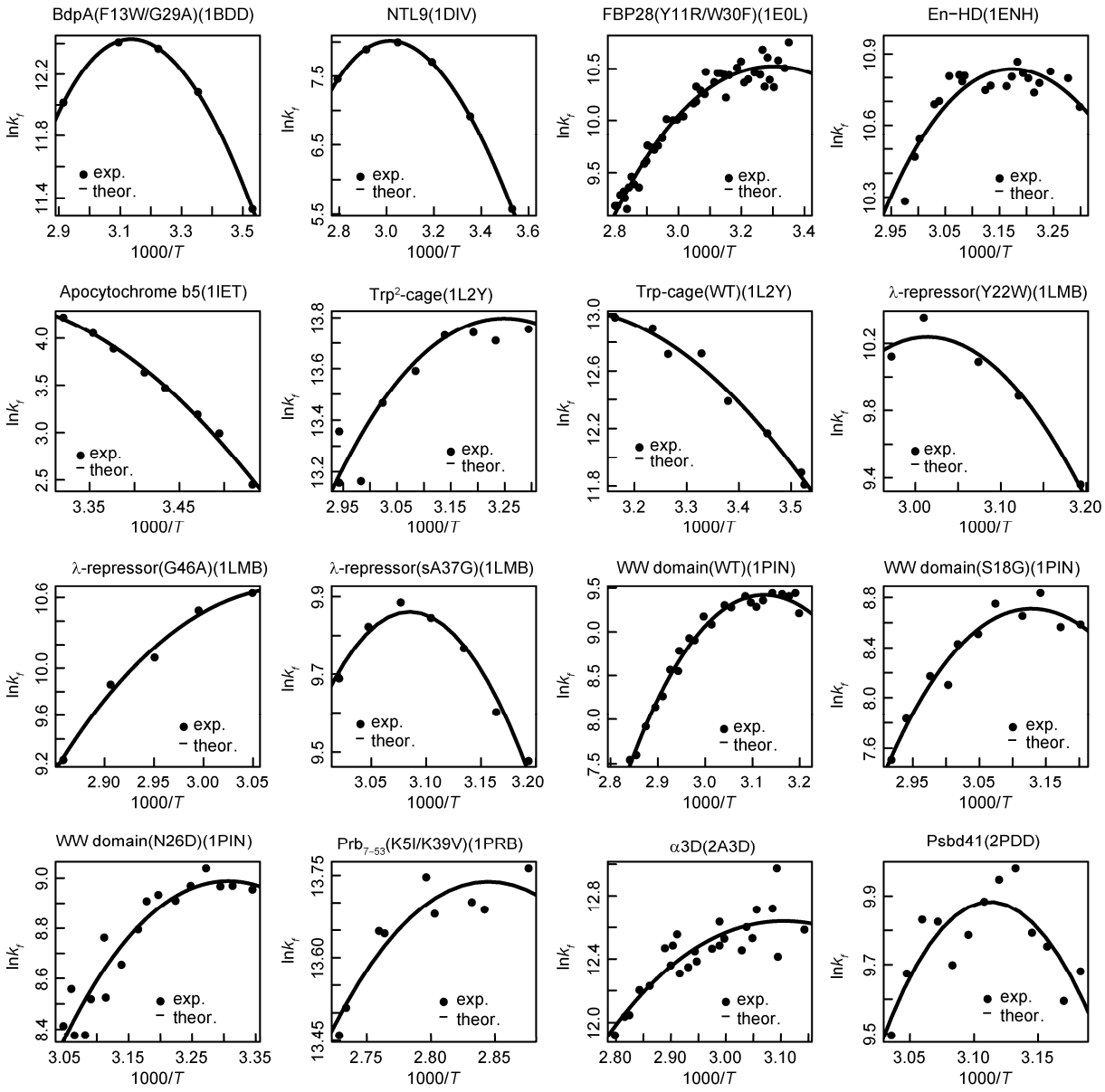

Figure 3 Model fits to overall folding rate k_f vs. temperature 1000/T for 16 proteins. Experimental logarithm folding rates are shown by "o", and solid lines are theoretical model fits to the folding rate (k_f in unit s^{-1}, T in unit Kelvin). PDB codes and references on the experimental temperature dependence for 16 proteins are given in Table 2.

2.2 Prediction on temperature dependence of unfolding rate for 16 typical proteins

It was indicated that the plots of ln$W$ versus 1/T are strongly curved for refolding of some proteins, but almost linear for their unfolding under denaturant [33]. However, in our model the folding and unfolding can be studied on the same foot. From eq. (1) the unfolding rate W_u for the reversed process is easily obtained by the replacement of ΔG by $-\Delta G$ and $\bar{\omega}'$ ($\bar{\omega}$) by $\bar{\omega}$ ($\bar{\omega}'$) in W_f, as eq. (13). It leads to

$$\ln W_u = \frac{S'}{T} + \frac{1}{2}\ln T - R'T + const\,, \tag{13}$$

$$S' = \frac{\eta\Delta E'(T_c)\left(1-\dfrac{\eta\Delta E'(T_c)}{\varepsilon'}\right)}{2k_B}, \tag{14}$$

$$R' = \frac{1}{2\varepsilon' k_B T_f^2}[\Delta G_f' - \eta\Delta E'(T_c)]^2. \tag{15}$$

In eqs. (14) and (15), $\Delta E'(T_c) = -\Delta E(T_c)$, $\varepsilon' = \bar{\omega}'^2(\delta\theta)^2 NI_0$, $\Delta G_f' = -\Delta G_f$. Therefore, in this theory the folding and unfolding rates are correlated with each other, needless of introducing any further assumption as given in [33].

From eqs. (7), (8), (14), and (15) one deduces

$$\frac{2k_BS'+\eta\Delta E(T_c)}{2k_BS-\eta\Delta E(T_c)}=s_\omega^2, \tag{16}$$

$$\frac{R'}{R}=r_\omega^2, \tag{17}$$

obeying

$$s_\omega=r_\omega=\bar{\omega}/\bar{\omega}'. \tag{18}$$

Thus, the temperature dependence of unfolding rate can fully be predicted from the temperature dependence parameters of folding rate under given frequency ratio $\bar{\omega}/\bar{\omega}'$. The frequency ratio can be estimated from the parameter λ in Table 4. We found that the predictions for 16 proteins are in good agreement with experimental data (Figure 4). s_ω and r_ω for each protein are given in Table 5, whose difference $\delta r_\omega=s_\omega-r_\omega$ is generally smaller than 15%, falling in the range of measurement error of rates. Moreover, that the calculated frequency ratios $\bar{\omega}/\bar{\omega}'$ for these 16 fast-folding proteins are all smaller than 1 (close to 1) explains the plots of ln$W$ versus 1/T less curved for unfolding than for folding.

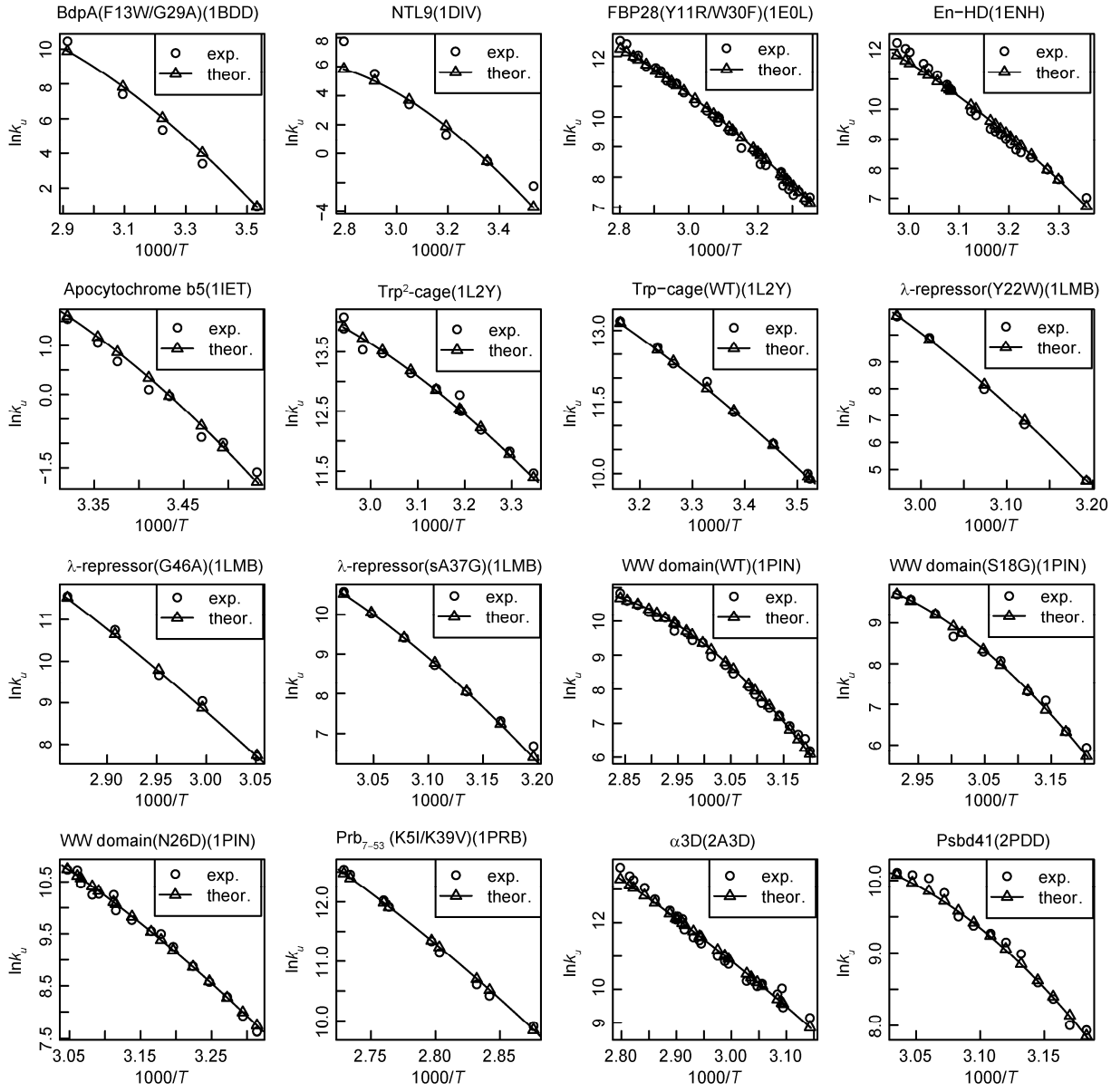

Figure 4 Model fits to overall unfolding rate k_u vs. temperature 1000/T for 16 proteins. Experimental logarithm unfolding rates are shown by "o", and theoretical model fits to logarithm unfolding rate are shown by "Δ" (k_u in unit s^{-1}, T in unit Kelvin).

Table 5 The parameters used in the prediction for temperature dependence of protein unfolding rate[a)]

PDB Code	S'	R'	s_ω	r_ω	ε'/k_BT_f	*MRE*
1bdd	−36742.43	0.2304	0.972	0.972	11.35	0.085
1div	−41209.80	0.2807	0.979	0.979	5.65	0.252
1e0l	−23226.90	0.1321	0.962	0.862	6.43	0.013
1enh	−41084.84	0.2779	0.981	0.911	3.97	0.021
1iet	−78345.63	0.7521	0.983	0.999	2.66	0.412
1l2y(p12w)	−19495.24	0.1318	0.957	0.907	3.97	0.006
1l2y(wt)	−21660.65	0.1422	0.985	0.895	0.97	0.004
1lmb(wt)	−95405.16	0.6422	0.978	0.918	4.31	0.012
1lmb(g46a)	−46610.77	0.2342	0.971	0.841	14.15	0.009
1lmb(sa37g)	−121636.30	0.9459	0.977	0.937	2.78	0.010
1pin(wt)	−75475.70	0.5714	0.946	0.916	4.49	0.012
1pin(s18g)	−81804.30	0.6353	0.946	0.916	3.85	0.011
1pin(n26d)	−33085.92	0.2220	0.972	0.862	3.44	0.008
1prb	−59773.65	0.3304	0.961	0.921	7.25	0.005
2a3d	−27437.16	0.1294	0.985	0.845	6.46	0.015
2pdd	−157517.20	1.3787	0.951	0.941	2.05	0.009

a) S' and R' in the 2nd and 3rd column are calculated by inserting folding parameters $\eta\Delta E(T_c)$, S and R into (16) and (17) and taking s_ω and r_ω from columns 4 and 5; s_ω in the 4th column is given by the frequency ratio $\bar{\omega}/\bar{\omega}'$ (eq. (18)) calculated from $\lambda_{\eta=5.5}$ in Table 4 and r_ω in the 5th column is supposed to be $s_\omega-\delta r_\omega$, where δr_ω is introduced to take the possible error existing in the unfolding rate measurement (the values of η and δr_ω have been chosen appropriately through the minimization of *MRE*); ε' in the 6th column is deduced from ε/s_ω^2; *MRE* in the 7th column is the mean error between theory and experiment defined by $MRE=\frac{1}{n}\sum_{i=1}^{n}|(\ln k_u^i-\ln W_u^i)/\ln k_u^i|$, k_u^i and W_u^i are experimental and theoretical unfolding rates respectively at the i-th temperature (i=1,…,n). Following this definition, if the unfolding rate of a protein takes a value near 0 at some temperature then the *MRE* will be abnormally large (as shown in the table for 1div and 1iet).

2.3 Results of statistical analysis of 65 two-state protein folding rates

Eq. (12) gives a relation for predicting the folding rate of any two-state protein. For a dataset of 65 two-state protein at temperature T_0=298 K, the prediction results are shown in Figure 5. The correlation between theoretical logarithm rate $\ln W_f$ and experimental $\ln k_f$ has attained 0.73. Much work on the folding rate prediction was published in the pure-empirical approach. They include the prediction model based on amino acid sequence [18,26–28], based on tertiary structure [15,29] and based on secondary structure [17], etc. The prediction accuracy is generally dependent on the size of database. For the same database of 65 two-state proteins, the prediction results are listed in Table 6.

From Table 6 we find the present prediction is better than other empirical models in the correlation coefficient R and it is comparable with model ACO, SMCO and L_{eff} but better than others in the standard error σ. Garbuzynskiy et al. [42] recently proved that the measured protein folding rates fall within a narrow triangle (called Golden triangle). Our results give an explanation for the origin of the Golden triangle.

It is worth pointing out that if the free energies ΔG are directly taken from the experimental data instead of using eq. (10) then the prediction accuracy on folding rate from (1) and (11) will be further increased, the correlation coefficient R between theoretical and experimental logarithm rate for 65 two-state proteins attained 0.78 [57].

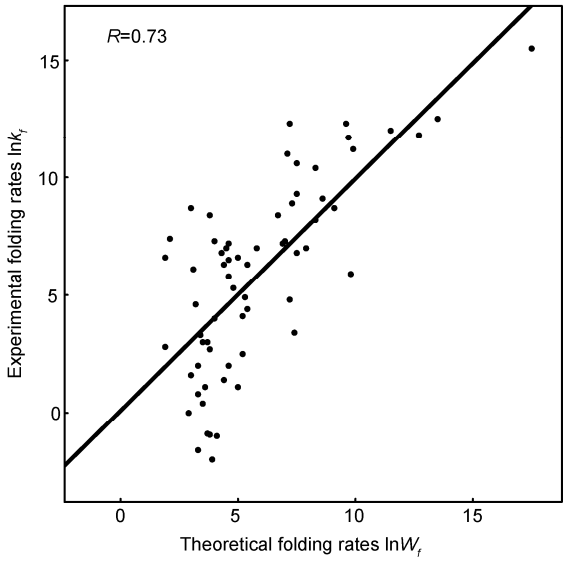

Figure 5 Comparison between theoretical and experimental folding rates for 65 two-state proteins. Theoretical predictions are calculated from eq. (12). In theoretical calculation, the parameters are chosen as follows: b=0.709, c=4.08, ρ=0.097, c_0=2.07×10^{13}, and d=5.5. If ρ is changed to 0.03 and d is changed to 4.2 then the correlation R is slightly increased to 0.74. F and N are taken from Table 1.

Table 6 Comparison among different folding rate prediction methods based on the same set of 65 two-state proteins

Models	R	P-value	σ
SFoldRate[a)]	0.53	5.81×10^{-6}	7.04
CIpred[b)]	0.61	5.49×10^{-8}	4.02
FOLD-RATE[c)]	–0.17	0.17	16.26
SWFoldRate[d)]	–0.10	0.42	7.18
RCO[e)]	0.16	0.19	4.06
ACO[e)]	0.67	9.33×10^{-10}	2.68
SMCO[f)]	0.72	1.47×10^{-11}	2.83
L_{eff}[g)]	0.65	3.88×10^{-9}	3.10
L[h)]	0.33	0.0069	3.78
lnL[h)]	0.53	6.89×10^{-6}	3.44
Present method	0.73	5.79×10^{-12}	2.78

All the models were tested on the set of 65 two-state proteins. R, P-value and σ are correlation coefficient, level of significance of linear regression (by F-test) and standard error between experimental folding rates and predicting folding rates, respectively. The standard error is defined by $\sigma=\sqrt{\sum_{i=1}^{n}(y_i-x_i)^2/(n-2)}$ (y_i the experimental folding rates, x_i predicting folding rates, n=65). In model RCO, ACO, SMCO, L_{eff}, L and ln$L$ the Jack-knife test was used in obtaining the predicting folding rates. a) Result from the Nα [18] web server at http://gila.bioengr.uic.edu/lab/tools/foldingrate/fr0.html. b) Result from the CI [26] web server at http://ibi.hzau.edu.cn/FDserver/cipred.php. c) Result from the Fold-Rate [27] web server at http://psfs.cbrc.jp/fold-rate/. d) Result from the SWFoldRate [28] web server at http://www.jci-bioinfo.cn/swfrate/input.jsp. e) Results of RCO [15] and ACO [29] from web server at http://depts.washington.edu/bakerpg/contact_order/. f) Size-modified contact order SMCO=RCO×$L^{0.7}$ [29], L, number of residues that have defined three-dimensional coordinates and contribute to the relative contact order (RCO) calculations. g) The effective length of the folding chain $L_{\text{eff}}=L-L_{\text{H}}+l_1N_{\text{H}}$ [17], where L_{H} is the number of residues in helical conformation, N_{H} is the number of helices, and l_1=3. Following [17], the folding rate is proportional to L_{eff}^{P} with P=0.1. The prediction is carried out by using the linear regression between L_{eff}^{P} and experimental rate and the Jackknife test. Secondary structure was assigned from Protein Data Bank coordinates of proteins by using the program DSSP, which marks helical residues by symbols H. h) L is number of residues that have defined three-dimensional coordinates.

3 Discussion

3.1 Temperature dependence of free energy and folding rate

In studying the temperature dependence of folding rate, we assume that the torsion potential parameter ΔE is a linear function of T by eq. (3). This is equivalent to assuming that the free energy change ΔG per molecule is linearly dependent on temperature by the relationship that exists between ΔE and ΔG, as in eq. (2). The linear relationship between ΔG and T is checked for 15 proteins under investigation (listed in Table 2, apart from 1iet whose free energy cannot be fully determined from the experimental data [48]), and we found that the correlation coefficients were higher than 0.99 for most proteins. Figure 6 gives two examples.

To solve the conformational parameters $\eta\Delta E(T_c)$, ε and $\delta\theta\bar{\omega}$ for each protein from slope parameters S and R on Arrhenius plot, one should use the free energies ΔG_f measured at temperature T_f. It is required that T_f be lower than melting temperature T_c by 10 degrees or more, because there is some ambiguity in the experimental determination of the free energy change ΔG as T near T_c. The reason is as follows. From eq. (1) the relation between unfolding rate W_u and folding rate W_f is

$$\ln\frac{W_f}{W_u}=\frac{\Delta G}{k_BT}+\frac{(\Delta G)^2}{2k_BT\varepsilon}\left(\frac{\bar{\omega}^2-\bar{\omega}'^2}{\bar{\omega}'^2}\right)+\ln\frac{\bar{\omega}}{\bar{\omega}'}$$

$$\cong\frac{\Delta G}{k_BT}+\ln\frac{\bar{\omega}}{\bar{\omega}'}\left(1+\frac{(\Delta G)^2}{k_BT\varepsilon}\right)\qquad(19)$$

$$\left(\left|\frac{\bar{\omega}'-\bar{\omega}}{\bar{\omega}}\right|<<1\right).$$

Eq. (19) means the thermodynamic equilibrium condition for chemical reaction ΔG=0 is slightly different from the dynamical equilibrium $W_f=W_u$ for protein folding due to the non-equal bias frequency distribution of $\{\omega_j\}$ and $\{\omega'_j\}$. Usually the free energy ΔG at given temperature T was measured through $\Delta G=k_BT\ln(k_f/k_u)$ in literatures. However, from eq. (19), this determination of free energy is not accurate as T near T_c, where $\ln(k_f/k_u)$ is a small quantity and the term proportional to $\ln\bar{\omega}/\bar{\omega}'$ cannot be neglected. Then, we study how the calculated $\eta\Delta E(T_c)$ will be changed if there exist some errors in the ΔG_f measurement. Given $\delta[\Delta G_f/(k_BT_f)]$ the change of $\eta\Delta E(T_c)/(k_BT_f)$, $\delta[\eta\Delta E(T_c)/(k_BT_f)]$ is plotted for various temperature T_f in Figure 7. We find $\delta[\eta\Delta E(T_c)/(k_BT_f)]$ increases rapidly as T_f approaches T_c. Thus, the small error in free energy measurement at temperature T_f near T_c would bring about the instability of some theoretical results.

Since the denaturant possibly changes the torsion force

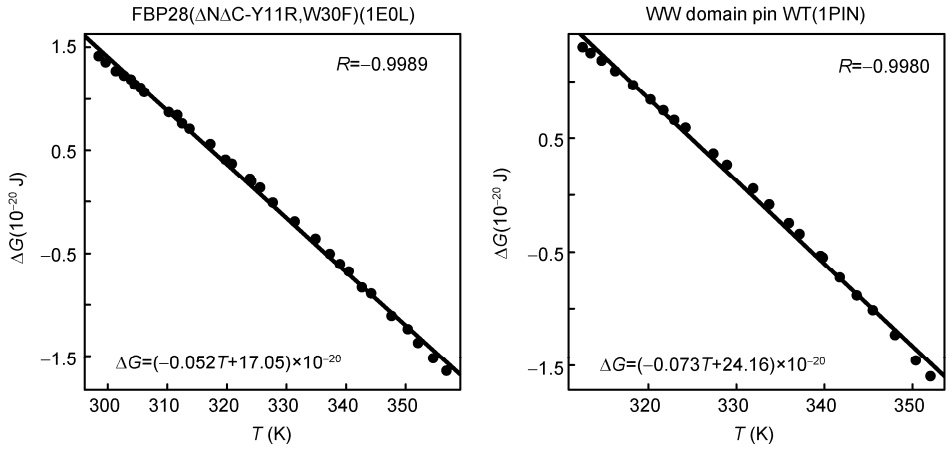

Figure 6 The linear relation of experimental folding free energy ΔG with temperature T for protein 1E0L and 1PIN.

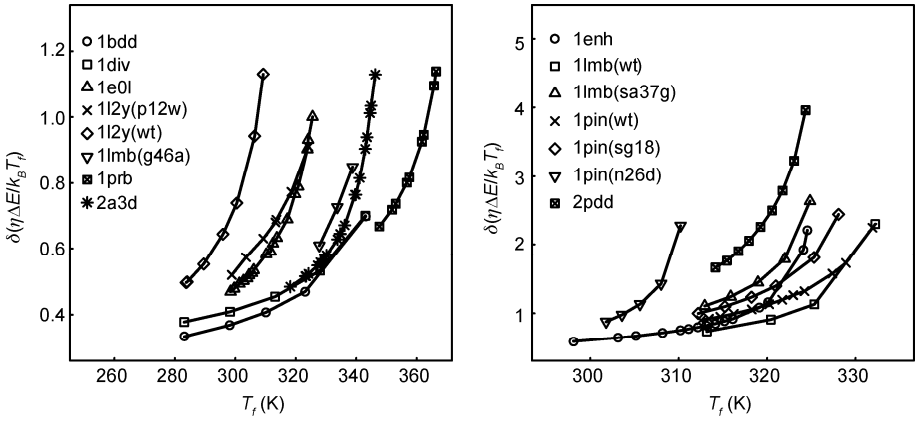

Figure 7 The change of $\eta\Delta E(T_c)/(k_BT_f)$, $\delta[\eta\Delta E(T_c)/(k_BT_f)]$, with temperature T_f for given $\delta[\Delta G_f/(k_BT_f)]=0.1$. The coordinate is the change of conformational energy parameter $\delta[\eta\Delta E(T_c)/(k_BT_f)]$, the abscissa is the temperature where ΔG_f is measured. The measurement errors of free energy at all temperatures are assumed to be 10%. The protein 1iet is not shown due to the scarcity of unfolding rate data.

field, the energy gap ΔE is a function of both variables, temperature and denaturant. Therefore, the free energy ΔG and the folding and unfolding kinetics depend not only on the temperature but also on the denaturant concentration. The denaturant dependence of folding rate has not been discussed in the article; however, the problem can be studied in the present theory because eq. (1) provides a basis for understanding the role of the influence of temperature and denaturant on the folding/unfolding kinetics.

3.2 The relation between free energy and N

The linear relation of free energy ΔG with $\sqrt{N}$, eq. (10), is used in the statistical analysis of protein folding rates. The relationship of ΔG with $\sqrt{N}$ can be understood by the Einstein's formula of random walk—distance square proportional to the number of walk steps—in his famous analysis of Brownian movement. The relation has been tested for 65 two-state proteins under investigation (Figure 8A). The regression analysis gives the correlation coefficient R between ΔG and $\sqrt{N}$ as 0.67 (P<0.0001, F-statistic test). Garbuzynskiy et al. [42] suggested another relation between free energy ΔG and length of chain L, $\Delta G = Lg + \sigma B_L L^{2/3}$. As a comparison we plot Garbuzynskiy's relation $\Delta G/(k_BT)$ vs. $(Lg + \sigma B_L L^{2/3})/(k_BT)$ in Figure 8B. We found the correlation coefficients R are nearly the same for both relationships of free energy with protein dimension.

To study the origin of the error occurring in the statistical relation between free energy and $\sqrt{N}$, we change the temperature at which the free energy is measured for 16 proteins in Table 2 and classify these proteins into five groups following the divergence of temperature $\delta T = \sqrt{\sum_{i=1}^{n}(T_i - \bar{T})^2/(n-1)}$ (n=16). Adding these 16 proteins into 65-protein dataset in Table 1, we obtain a new

dataset with 74 proteins after removing the repetitious ones. By making statistics on free energy with $\sqrt{N}$, we obtain the correlation coefficient for each group of 74 proteins. The correlation coefficient R and the diversity of temperature δT_f are listed in Table 7. The result shows that the correlation of $\Delta G_f/(k_B T_f)$ with $\sqrt{N}$ increases with decreasing δT_f for five groups. Therefore, the deviation from linear relation $\Delta G_f/(k_B T_f)$ with $\sqrt{N}$ at least partly comes from temperature diversity of proteins in the studied group.

3.3 The dimensional parameter *d* and the structure-related parameter *F*

The dimensional parameter d is introduced in the fast-variable factor $M\bar{a}^2$ of the folding rate formula, eq. (11). The relationship involves $M\bar{a}^2$ as it changes with N^{-d}. If the fast-variable factor were not related to N, then it could be shown that the folding rate would be an increasing function of N, thus conflicting with experimental data. From the general form of the potential function for peptides, $V\sim\sum\left(\frac{A_{ij}}{r_{ij}^{12}}+\frac{B_{ij}}{r_{ij}^{10}}+\frac{C_{ij}}{r_{ij}^{6}}+\frac{D_{ij}}{r_{ij}}\right)$ [58], it is reasonable to assume that the matrix element of the stretching-bending (as a part of fast-variables) Hamiltonian includes a factor of r^{-d_0}, in turn leading to $M\bar{a}^2$ including a factor of $N^{1-2d_0/3}$ for a globular protein. This explains why N^{-d} should be introduced in $M\bar{a}^2$. If d_0=10 (or 6), then d=5.66 (or 3), consistent with d=5.5 (or 4.2) by fitting experimental rates in Figure 5. In fact, to give a theory consistent with experimentation, the dimension parameter d should be chosen such that it is related to slow-variable parameter $\rho = I_0\bar{\omega}^2(\delta\theta)^2/(k_B T)$. From the statistical analysis of 65 two-state proteins we obtain a good linear relation existing between d and ρ^{-1} in the range $\rho^{-1}<55$ $\left(d=-0.0568\frac{1}{\rho}+6.0872\right)$. Therefore, one has to assume d according to the ρ value of the database. For example, we assumed d=5.5 (as $\rho=0.097$) in eq. (12) for 65-protein dataset. One may also assume d=4.2 (as $\rho=0.03$ with smaller $\bar{\omega}^2(\delta\theta)^2$) for the fast-folding protein dataset as those proteins listed in Table 2.

The structure-related parameter F is also introduced in the fast-variable factor $M\bar{a}^2$ of the folding rate formula. From experimental data analysis we found that F should be larger for a protein with more residues in α helix; for example, F takes a value 81 for pure α helix chain (Table 1). Alpha helix, or a protein with abundant α helices, may have a quite oblong or oblate ellipsoid, instead of spheroid, shape. In globular protein of ellipsoidal shape, many residue pairs have a smaller distance than the average distance of a pair in spherical proteins of the same volume. By factor r^{-d_0} (d_0=10~6) in the interaction potential, only those pairs with the smallest distance predominantly contribute to the fast-variable matrix-element $\bar{a}$. Moreover, the ellipsoidal protein will have an enhancing factor F in $M\bar{a}^2$. If the min-

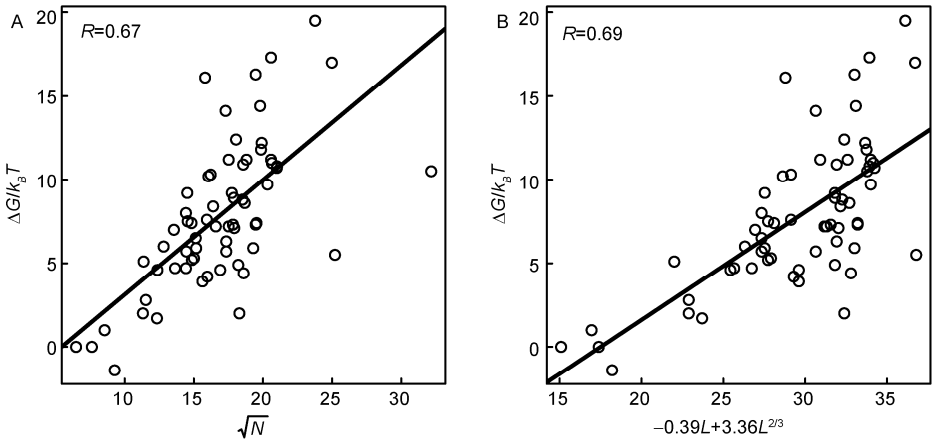

Figure 8 Relationship of free energy $\Delta G/(k_B T)$ with $\sqrt{N}$. A, Free energy $\Delta G/(k_B T)$ vs. $\sqrt{N}$ for 65 two-state proteins. B, Free energy $\Delta G/(k_B T)$ vs. $(Lg+\sigma B_L L^{2/3})/(k_B T)$ for 65 two-state proteins where parameters σB_L =3.36 $k_B T$ and the best fit g=–0.39 $k_B T$ are used, and L is the length of polypeptide chain defined in [42].

Table 7 The correlation between $\Delta G/(k_B T)$ and $\sqrt{N}$ increases with the lowering of temperature deviation in measurements

δT	24.3	18.0	15.0	12.9	12.3
R	0.6477	0.6687	0.6776	0.6887	0.6895

imal pair-distance decreases by a factor μ=1.25~1.45, then the enhancing factor is $F=\mu^{2d_0}$=87, which is close to 81, as required by experiments. Therefore, we assume the prediction rule: F=81 for $(L_\alpha-L_\beta)/L\geqslant 0.6$, F=25 for $0.3\leqslant(L_\alpha-L_\beta)/L<0.6$, and F=1 for $(L_\alpha-L_\beta)/L<0.3$ (L_α and L_β are the number of residues in α helix and β sheet, respectively, and L is the total number of folded residues). Our experience shows that the different choice of F-value in the intermediate region $0.3\leqslant(L_\alpha-L_\beta)/L<0.6$ is insensitive to the prediction results.

3.4 On the theoretical relationship of lnW vs. $\sqrt{N}$ for two-state protein

Figure 9 gives the theoretical relation of lnW vs. $\sqrt{N}$ for two-state protein. The curves are plotted following eq. (12). As ρ=0.097 and d=5.5, one has $W=W_{\max}=1.8\times10^8$ s^{-1} at $N=N_1=8$ and $W=W_{min}=3.26$ s^{-1} at $N=N_2=1283$. The maximum and minimum of folding rate, namely ln$W_{\max}$, ln$W_{\min}$, $\sqrt{N_1}$ and $\sqrt{N_2}$, changes with ρ and d but within a smaller range (as $\rho<0.65$ and $d<6$). The rate $W_{\max}$ is consistent with the known experimentally measured rate of conformational rearrangement of one amino acid residue [42]. On the other hand, although the rate $W_{\min}$ is much higher than the experimental lower limitation $W_{\min}$ (biol)=0.003 s^{-1} [42], the predicted minimum of the rate seems not conflict with any existing two-state protein data. For multi-state protein one may assume the folding is a mutual process of several quantum transitions in different domains and that some time delays exist between these transitions [40]. Therefore, it is reasonable to assume that multistate folding proceeds in a larger spatial dimension and needs more execution time.

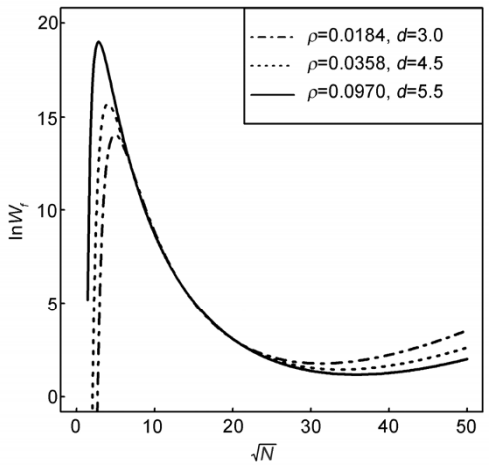

Figure 9 The theoretical relationship ln$W$ vs. $\sqrt{N}$. The curves are plotted following eq. (12) under parameter choices b=0.709, c=4.08 and F=1 for various ρ and d. As ρ=0.097 and d=5.5, one has $W=W_{\max}=1.8\times10^8$ s^{-1} at $N=N_1=8$ and $W=W_{\min}=3.26$ s^{-1} at $N=N_2=1283$. The maximum and minimum of folding rate changes with ρ and d but within a smaller range (as $\rho<0.65$ and $d<6$).

3.5 Generalization of folding rate prediction to the case of varying temperature

Eq. (12) is established based on the statistical analysis of protein folding rates at given temperature $T=T_0$=298 K. However, the equation can be generalized and it will be usable for the prediction of the folding rate of any two-state protein at different temperatures $T\neq T_0$. Assume the parameters b and c in eq. (12) are replaced by

$$b(T)=\frac{b_{T0}+b_1(T-T_0)}{T},\qquad c(T)=\frac{c_{T0}+c_1(T-T_0)}{T}, \tag{20}$$

and the dependence of ρ on T, namely $\rho=\rho_0\frac{T_0}{T}$, is taken into account. Then, from eq. (12) one can deduce

$$\ln W=\frac{S_{\text{pred}}}{T}-R_{\text{pred}}T+\frac{1}{2}\ln T+const. \tag{21}$$

with

$$S_{\text{pred}}=\frac{(b_{T0}-b_1T_0)\sqrt{N}-(c_{T0}-c_1T_0)}{2}\cdot\left[1-\frac{(b_{T0}-b_1T_0)\sqrt{N}-(c_{T0}-c_1T_0)}{N\rho_0T_0}\right], \tag{22}$$

$$R_{\text{pred}}=\frac{(b_1\sqrt{N}-c_1)^2}{2N\rho_0T_0}.$$

Notice that eq. (21) takes the same form of temperature dependence as given in eq. (6). Thus, eqs. (21) and (22) provide a useful tool for predicting protein folding rates at different temperatures. For 15 proteins (listed in Table 2, apart from 1iet) whose experimental data on the temperature-dependent rates and free energies had previously been available, the prediction results are shown in Table 8.

From Table 8 we find that $S_{\text{pred}}/R_{\text{pred}}$ is in good agreement with S/R for each of the 15 proteins and R/R_{pred} takes a value between 0.6 and 3 for most proteins. This gives basically consistent-with-experiment predictions for the temperature dependence of folding rates. The larger deviations of R/R_{pred} from one for some proteins are attributed to the common ρ_0 used in calculating R_{pred}. In fact, the diversity of $\bar{\omega}^2(\delta\theta)^2$ for different proteins (that can be found in Table 4) makes ρ_0 vary in the 15-protein dataset.

By using eqs. (20)–(22) with the same parameters, we are able to predict the folding rate for the proteins 1VII [59], 2PDD [52], 1PRB [51] and 2A3D [35] (collected in [42] but not considered in our 65-protein dataset) whose folding rates were measured at high temperatures (higher than T_0). The calculated temperature dependences of these proteins agree well with the experimental data. Theoretical logarithm

Table 8 The parameters for predicting temperature-dependent folding rates of 15 proteins[a)]

PDB code	$S/R(\times10^4)$	$S_{pred}/R_{pred}(\times10^4)$	R/R_{pred}
1bdd	−10.10	−10.65	0.987
1div	−10.95	−10.66	1.189
1e0l	−9.12	−10.09	0.687
1enh	−9.92	−10.66	1.357
1l2y (p12w)	−9.24	−9.71	0.605
1l2y (wt)	−10.69	−9.71	0.670
1lmb (wt)	−11.02	−10.65	3.129
1lmb (g46a)	−9.14	−10.65	1.362
1lmb (sa37g)	−10.49	−10.65	4.425
1pin (wt)	−10.23	−10.38	2.679
1pin (s18g)	−10.19	−10.37	2.974
1pin (n26d)	−9.05	−10.38	1.176
1prb	−12.30	−10.59	1.560
2a3d	−10.20	−10.67	0.741
2pdd	−10.24	−10.50	6.180

a) Fifteen proteins are taken from Table 2. S and R are taken from Table 4. S_{pred} and R_{pred} are calculated following eq. (22). In calculation $b_{T0}=0.709\times298$, $c_{T0}=4.08\times298$ are assumed, consistent with b=0.709 and c=4.08 used in Figure 5, and b_1=−2, c_1=1.5, $\rho_0=I_0\bar{\omega}^2(\delta\theta)^2/(k_BT_0)$ =0.03 are assumed after optimization.

folding rates ln$W$ differing from experimental lnk_f at different temperatures for each protein are only in the range of 1%–2%.

3.6 Comparison with molecular dynamics predictions

Recently, by using massively parallel supercomputer Anton, the atomic-level molecular dynamics simulations were performed for 12 fast-folding proteins and the folding times of these proteins were predicted [60]. The experimental measurements of the folding rate were carried out under higher temperatures. Of the 12 proteins, only Villin, NTL9 and Homeodomain are not included in the 65-protein dataset and have experimental rate data at lower temperature near 298 K. By using eq. (12) we predicted the folding times for Villin, NTL9 and Homeodomain are 8.5, 1287 and 70 μs respectively, near the experimental values 0.7 [61], 827 [62] and 13 μs [63]. The differences between predicted lnW_f and experimental lnk_f fall in the range as shown in Figure 5. While the molecular dynamics simulations by Shaw et al. gave the folding times 2.8, 29 and 3.1 μs respectively. But their simulation temperatures were assumed to be near 360 K, explicitly higher than experimental temperature. If the MD simulation is not strongly dependent on temperature, then the above two predictions, from MD simulation and from quantum folding, can be compared with each other. About the temperature dependence of folding rate, our approach has deduced a definite relation, eq. (21). We have checked the relation for proteins Villin [61], BBL [64] and α3D [35] whose experimental data on folding rate at more than one temperatures were published. The logarithm folding times at different temperatures are in accordance with eq. (21) and the corresponding ρ_0 for each protein is obtained, ρ_0 =0.0235, 0.0389 and 0.0334 respectively for three proteins. All parameters ρ_0 taking a value near 0.03 are consistent with the optimal ρ_0 used in the statistical analysis of 15 fast-folding-protein dataset (Table 8).

4 Conclusion

The temperature dependence of the folding rates for 16 fast-folding proteins is studied statistically from the view of quantum transition and the abnormal non-Arrhenius peculiarities have been explained. A statistical formula for the prediction of protein folding rates is proposed based on quantum folding theory. The formula is tested on a dataset of 65 two-state proteins and compared with other prediction models. The results obtained in this article support the conformational quantum transition theory of protein folding, giving a new approach to the exploration of the protein folding mechanism.

This work was supported by the Distinguished Scientist Award of Inner Mongolia Autonomous Region (2008). The authors thank Dr. Zhang Ying for his help in data collection. The authors also thank Dr. Zhao JuDong for numerous discussions in the writing of the paper and thank Wang JingFeng (Inner Mongolia University of Technology) for his help in graphic edits.

1 Anfinsen CB, Haber E, Sela M, White FH Jr. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc Natl Acad Sci USA, 1961, 47: 1309–1314
2 Anfinsen CB. Principles that govern the folding of protein chains. Science, 1973, 181: 223–230
3 Levinthal C. Are there pathways for protein folding? J Chim Phys Chim Biol, 1968, 65: 44–45
4 Qiu L, Pabit SA, Roitberg AE, Hagen SJ. Smaller and faster: the 20

residue Trp-cage protein folds in 4 μs. J Am Chem Soc, 2002, 124: 12952–12953
5 Kubelka J, Hofrichter J, Eaton WA. The protein folding 'speed limit'. Curr Opin Struct Biol, 2004, 14: 76–88
6 Mayor U, Johnson CM, Daggett V, Fersht AR. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. Proc Natl Acad Sci USA, 2000, 97: 13518–13522
7 Reader JS, Van Nuland NA, Thompson GS, Ferguson SJ, Dobson CM, Radford SE. A partially folded intermediate species of the beta-sheet protein apo-pseudoazurin is trapped during proline-limited folding. Protein Sci, 2001, 10: 1216–1224
8 Phillips DC. The three-dimensional structure of an enzyme molecule. Sci Am, 1966, 215: 78–90
9 Fersht AR. Nucleation mechanisms in protein folding. Curr Opin Struct Biol, 1997, 7: 3–9
10 Leopold PE, Montal M, Onuchic JN. Protein folding funnels: a kinetic approach to the sequence-structure relationship. Proc Natl Acad Sci USA, 1992, 89: 8721–8725
11 Wolynes PG, Onuchic JN, Thirumalai D. Navigating the folding routes. Science, 1995, 267:1619–1620
12 Dill KA, Chan HS. From levinthal to pathways to funnels. Nat Struct Biol, 1997, 4: 10–19
13 Wolynes PG. Folding funnels and energy landscapes of larger proteins within the capillarity approximation. Proc Natl Acad Sci USA, 1997, 94: 6170–6175
14 Bicout DJ, Szabo A. Entropic barriers, transition states, funnels, and exponential protein folding kinetics: A simple model. Protein Sci, 2000, 9: 452–465
15 Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol, 1998, 227: 985–994
16 Baker D. A surprising simplicity to protein folding. Nature, 2000, 405: 39–42
17 Ivankov DN, Finkelstein AV. Prediction of protein folding rates from the amino acid sequence-predicted secondary structure. Proc Natl Acad Sci USA, 2004, 101: 8942–8944
18 Ouyang Z, Liang J. Predicting protein folding rates from geometric contact and amino acid sequence. Protein Sci, 2008, 17:1256–1263
19 Segal MR. A novel topology for representing protein folds. Protein Sci, 2009, 18: 686–693
20 Chang L, Wang J, Wang W. Composition-based effective chain length for prediction of protein folding rates. Phys Rev E Stat Nonlin Soft Matter Phys, 2010, 82(5 Pt 1): 051930
21 Chiti F, Taddei N, White PM, Bucciantini M, Magherini F, Stefani M, Dobson CM. Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. Nat Struct Biol, 1999, 6: 1005–1009
22 Muñoz V, Eaton WA. A simple model for calculating the kinetics of protein folding from three-dimensional structures. Proc Natl Acad Sci USA, 1999, 96: 11311–11316
23 Ivankov DN, Finkelstein AV. Theoretical study of a landscape of protein folding-unfolding pathways. Folding rates at midtransition. Biochemistry, 2001, 40: 9957–9961
24 Alm E, Morozov AV, Kortemme T, Baker D. Simple physical models connect theory and experiment in protein folding kinetics. J Mol Biol, 2002, 322: 463–476
25 Garbuzynskiy SO, Finkelstein AV, Galzitskaya OV. Outlining folding nuclei in globular proteins. J Mol Biol, 2004, 336: 509–525
26 Ma BG, Guo JX, Zhang HY. Direct correlation between proteins' folding rates and their amino acid compositions: an *ab initio* folding rate prediction. Proteins, 2006, 65: 362–372
27 Gromiha MM, Thangakani AM, Selvaraj S. FOLD-RATE: prediction of protein folding rates from amino acid sequence. Nucleic Acids Res, 2006, 34(web server issue): W70–74
28 Cheng X, Xiao X, Wu ZC, Wang P, Lin WZ. Swfoldrate: predicting protein folding rates from amino acid sequence with sliding window method. Proteins, 2013, 81: 140–148
29 Ivankov DN, Garbuzynskiy SO, Alm E, Plaxco KW, Baker D, Finkelstein AV. Contact order revisited: influence of protein size on the folding rate. Protein Sci, 2003, 12: 2057–2062
30 Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins, 1995, 21: 167–195
31 Chan HS, Dill KA. Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. Proteins, 1998, 30: 2–23
32 Akmal A, Munoz V. The nature of the free energy barriers to two-state folding. Proteins, 2004, 57: 142–152
33 Ghosh K, Ozkan B, Dill KA. The ultimate speed limit to protein folding is conformational searching. J Am Chem Soc, 2007, 129: 11920–11927
34 Yang WY, Gruebele M. Rate-temperature relationship in λ-repressor fragment $\lambda_{6\text{-}85}$ folding. Biochemistry, 2004, 43: 13018–13025
35 Zhu Y, Alonso DO, Maki K, Huang CY, Lahr SJ, Daggett V, Roder H, DeGrado WF, Gai F. Ultrafast folding of α3D: a *de novo* designed three-helix bundle protein. Proc Natl Acad Sci USA, 2003, 100: 15486–15491
36 Luo LF. Quantum theory on protein folding. Sci China Phys Mech Astron, 2014, 57: 458–468
37 Chakraborty A, Truhlar DG. Quantum mechanical reaction rate constants by vibrational configuration interaction: the $OH+H_2->H_2O+H$ reaction as a function of temperature. Proc Natl Acad Sci USA, 2005, 102: 6744–6749
38 Luo LF. Conformation transitional rate in protein folding. Int J Quant Chem, 1995, 54: 243–247
39 Luo LF. Protein folding as a quantum transition between conformational states. Front Phys, 2011, 6: 133–140
40 Zhang Y, Luo LF. The dynamical contact order: protein folding rate parameters based on quantum conformational transitions. Sci China Life Sci, 2011, 54: 386–392
41 Luo LF. Protein photo-folding and quantum folding theory. Sci China Life Sci, 2012, 55: 533–541
42 Garbuzynskiy SO, Ivankov DN, Bogatyreva NS, Finkelstein AV. Golden triangle for folding rates of globular proteins. Proc Natl Acad Sci USA, 2013, 110: 147–150
43 Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The protein data bank. A computer-based archival file for macromolecular structures. Eur J Biochem, 1977, 80: 319–324
44 Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers, 1983, 22: 2577–2637
45 Dimitriadis G, Drysdale A, Myers JK, Arora P, Radford SE, Oas TG, Smith DA. Microsecond folding dynamics of the F13W G29A mutant of the B domain of staphylococcal protein A by laser-induced temperature jump. Proc Natl Acad Sci USA, 2004, 101: 3809–3814
46 Kuhlman B, Luisi DL, Evans PA, Raleigh DP. Global analysis of the effects of temperature and denaturant on the folding and unfolding kinetics of the N-terminal domain of the protein L9. J Mol Biol, 1998, 284: 1661–1670
47 Nguyen H, Jager M, Moretto A, Gruebele M, Kelly JW. Tuning the free-energy landscape of a WW domain by temperature, mutation, and truncation. Proc Natl Acad Sci USA, 2003, 100: 3948–2953
48 Manyusa S, Whitford D. Defining folding and unfolding reactions of apocytochrome b5 using equilibrium and kinetic fluorescence measurements. Biochemistry, 1999, 38: 9533–9540
49 Bunagan MR, Yang X, Saven JG, Gai F. Ultrafast folding of a computationally designed Trp-cage mutant: Trp2-cage. J Phys Chem B, 2006, 110: 3759–3763
50 Jäger M, Nguyen H, Crane JC, Kelly JW, Gruebele M. The folding mechanism of a beta-sheet: the WW domain. J Mol Biol, 2001, 311: 373–393
51 Wang T, Zhu YJ, Gai F. Folding of a three-helix bundle at the folding speed limit. J Phys Chem B, 2004, 108: 3694–3697
52 Spector S, Raleigh DP. Submillisecond folding of the peripheral subunit-binding domain. J Mol Biol, 1999, 293: 763–768
53 Richardson JS. The anatomy and taxonomy of protein structure. Adv Protein Chem, 1981, 34: 167–339
54 Liu F, Gruebele M. Tuning λ6-85 towards downhill folding at its

melting temperature. J Mol Biol, 2007, 370: 574–584
55 Luo LF, Lu J. Temperature dependence of protein folding deduced from quantum transition. arXiv: 1102.3748 [q-bio.BM], 2011, Available from: http: //arxiv.org/abs/1102.3748
56 Ghosh K, Dill K. Cellular proteomes have broad distributions of protein stability. Biophys J, 2010, 99: 3996–4002
57 Zhao JD, Luo LF. A theory of quantum conformation transition for tubulin vibration. Biosystems, in press
58 Kostrowicki J, Scheraga HA. Application of the diffusion equation method for global optimization to oligopeptides. J Phys Chem, 1992, 96: 7442–7449
59 Wang M, Tang Y, Sato S, Vugmeyster L, McKnight CJ, Raleigh DP. Dynamic NMR line-shape analysis demonstrates that the villin headpiece subdomain folds on the microsecond time scale. J Am Chem Soc, 2003, 125: 6032–6033
60 Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How fast-folding proteins fold. Science, 2011, 334: 517–520
61 Kubelka J, Chiu TK, Davies DR, Eaton WA, Hofrichter J. Sub-microsecond protein folding. J Mol Biol, 2006, 359: 546–553
62 Horng JC, Moroz V, Raleigh DP. Rapid cooperative two-state folding of a miniature α-β protein and design of a thermostable variant. J Mol Biol, 2003, 326: 1261–1270
63 Gillespie B, Vu DM, Shah PS, Marshall SA, Dyer RB, Mayo SL, Plaxco KW. NMR and temperature-jump measurements of *de novo* designed proteins demonstrate rapid folding in the absence of explicit selection for kinetics. J Mol Biol, 2003, 330: 813–819
64 Neuweiler H, Sharpe TD, Rutherford TJ, Johnson CM, Allen MD, Ferguson N, Fersht AR. The folding mechanism of BBL: plasticity of transition-state structure observed within an ultrafast folding protein family. J Mol Biol, 2009, 390: 1060–1073