SCIENCE CHINA
Life Sciences

• RESEARCH PAPERS •

April 2011 Vol.54 No.4: 386–392
doi: 10.1007/s11427-011-4158-x

The dynamical contact order: Protein folding rate parameters based on quantum conformational transitions

ZHANG Ying[1,2] & LUO LiaoFu[1*]

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

Received May 17, 2010; accepted August 9, 2010

Protein folding is regarded as a quantum transition between the torsion states of a polypeptide chain. According to the quantum theory of conformational dynamics, we propose the dynamical contact order (DCO) defined as a characteristic of the contact described by the moment of inertia and the torsion potential energy of the polypeptide chain between contact residues. Consequently, the protein folding rate can be quantitatively studied from the point of view of dynamics. By comparing theoretical calculations and experimental data on the folding rate of 80 proteins, we successfully validate the view that protein folding is a quantum conformational transition. We conclude that (i) a correlation between the protein folding rate and the contact inertial moment exists; (ii) multi-state protein folding can be regarded as a quantum conformational transition similar to that of two-state proteins but with an intermediate delay. We have estimated the order of magnitude of the time delay; (iii) folding can be classified into two types, exergonic and endergonic. Most of the two-state proteins with higher folding rate are exergonic and most of the multi-state proteins with low folding rate are endergonic. The folding speed limit is determined by exergonic folding.

moment of inertia, dynamical contact order (DCO), protein folding rate

Citation: Zhang Y, Luo L F. The dynamical contact order: Protein folding rate parameters based on quantum conformational transitions. Sci China Life Sci, 2011, 54: 386−392, doi: 10.1007/s11427-011-4158-x

Protein folding is a complex kinetic process by which a polypeptide changes from the denatured state to the native folding state. Although a great deal of theoretical work has been carried out, the fundamental physics underlying the folding remains unclear. The folding kinetics of a large number of proteins has been studied experimentally [1–3]. For many of these proteins, folding has been shown to be an all-or-none process with no clear intermediate state. These are called two-state proteins. However, some proteins require the accumulation of intermediates to complete the folding process. These are referred to as three-state or multi-state proteins. The rate of folding varies from milliseconds to hours. Some small proteins fold much faster, at rates in microsecond [2,3]. Experimental data has indicated that most of the ultrafast folders show a significant decrease in folding rates as the temperature is increased [4]. The wide range of folding rates and a possible folding speed limit are likely to be closely related to the inherent physics behind the phenomena. A deeper understanding of the folding mechanism and the accurate prediction of protein folding rates is an important topic in protein science.

Since 1998, when the relative contact order (RCO) was proposed by Plaxco *et al*. [5], it has been widely accepted that folding rates and mechanisms are largely determined by the topology of the native state. Much theoretical work based on this concept has been proposed. For example, Ivankov *et al*. [6] proposed the absolute contact order (ACO) and the more general size-modified contact order (SMCO).

*Corresponding author (email: lolfcm@imu.edu.cn)

They considered the effect of sequence length and secondary structure on the folding rate and proposed the effective length (L_{eff}) as an important folding parameter [7]. A variety of programs that predict the folding rate based on amino acid sequence integrated with other information have been published [8–13]. Recently, using Delaunay tessellation (DT) Ouyang and Liang [14] proposed a geometric contact ($N\alpha$) to replace the previous contact order. By operationalizing 3D proximities for the underlying polypeptide chain, Segal [15] proposed a novel topology for representing protein folds that can capture more chain deformation/structural information. This idea represents a noteworthy new development. However, all these investigations remain at the phenomenological or geometrical level.

Dynamic studies of protein folding are currently limited to molecular dynamics simulations of the ultrafast folding of some small proteins [16,17]. Apart from limitations due to computational capability currently available, molecular dynamics simulations are based on classical mechanics while protein folding is mainly a quantum mechanical process of conformational change between different torsion states of the polypeptide chain. Luo [18–21] was one of the first to study protein folding problem from the point of quantum transition theory and to calculate the protein folding rate based on quantum conformational dynamics. Luo's results explain why the time scale of the fundamental folding event is generally in the order of milliseconds to microseconds and show that the folding rate can be represented by several physical parameters related to the dynamics of the polypeptide chain. The relationships between the folding rate and chain length, moment of inertia, energy gap and temperature have been deduced [18,20,21]. Based on Luo's theory we now propose a new contact order, the dynamical contact order (DCO), and use it to kinetically study protein folding rates. This approach will be helpful in capturing the dynamic essence of the contact order and in increasing our understanding of the mechanisms of protein folding.

1 Materials and methods

1.1 Protein folding rate data

The Ouyang and Liang [14] dataset of protein folding rates was used in this study. The dataset contains the experimental folding rates of 80 proteins or peptides of which 45 are two-state proteins and 35 are three-state or multi-state proteins. The proteins belong to different structural classes: 18 all alpha proteins, 32 all beta proteins and 30 alpha/beta proteins. In the dataset the difference in folding rates is more than eight orders of magnitude. All the data used in this study can be downloaded from http://gila.bioengr.uic.edu/resources/folding/Rate.html. We used the PDB identifiers to retrieve the structure data for these proteins from the PDB database (http://www.pdb.org/pdb/home/home.do) [22,23]. The "standard" set of 30 two-state proteins constructed by defining standard experimental conditions was also used in our study [1]. Following Segal [15], we have used 27 PDB-identified proteins from the "standard" set as an additional test set for our theory.

1.2 Dynamical contact order

Based on the theory of quantum conformational dynamics, Luo [18,20] calculated that the two-state protein folding rate is

$$W = 0.37\times10^{-87} e^{\frac{\Delta E}{2K_B T}} \left(\sum I_j\right)^{-1/2} \left(\sum \frac{b_j}{I_j}\right)^2, \qquad (1)$$

where W is the protein folding rate and the unit is s^{-1}. ΔE is the difference in the conformational potential minimum between the initial state (denatured state) and the final state (active state) of the protein. Because of the co-participation of many torsion angles in the transition, ΔE is the sum of the contribution of the different torsion degrees of freedom. K_B is the Boltzmann constant, T is the absolute temperature, and I_j is the moment of inertia of the jth mode (torsion angle) in g cm^2. b_j is the electronic quantum number, the square of the magnetic quantum number for the jth mode, for which the order of magnitude is one.

For multi-state protein folding we used the unified folding mechanism of non-two-state and two-state proteins proposed by Kamagata *et al*. [24]. We assumed that multi-state protein folding can be represented as the joining of several quantum transitions and that each quantum transition occurs at independent degrees of freedom for the torsion angles. Thus, the total collapse process can be approximately described as a non-radiative transition of many degrees of freedom [21]. The transition rate of multi-state protein can then be defined as

$$W' = 0.37\times10^{-87} e^{\frac{\Delta E}{2K_B T}} \left(\sum I_j\right)^{-1/2} \left(\sum \frac{b_j}{I_j}\right)^2 e^{-\tau}. \qquad (2)$$

The additional factor $e^{-\tau}$ in this equation indicates that, compared with the two-state protein folding rate, there may be a time delay in the multi-step process for non-two-state protein folding rates.

The complex movement of residues during protein folding is difficult to follow. To calculate the folding rate using eqs. (1) and (2), we propose a model based on the concept of contact order. To characterize a contact we assume the inertial moment of the polypeptide chain between contact residues and the torsion energy of the cooperative transition to be the feature variable. This we have named the dynamical contact order (DCO) model to emphasize the kinetic aspects of the contact. To investigate the consistency of the theoretical results and to ensure that the result does not de-

pend on the details of torsion motion, we have calculated the moment of inertia using three different algorithms detailed below.

The residue is regarded as a particle with a mass that is equal to the mass of the corresponding amino acid minus 18 u (the water molecule mass). The spatial location is determined by the coordinate of the main chain Cα atom. When the spatial distance between a pair of residues a_i and a_j (i and j denote the position of the residues in the sequence, assuming $i<j$) is not greater than 0.8 nm (threshold) and $j-i>1$, the residue pair is regarded as a contact pair s. The moment of inertia of the kth residue in contact pair s is calculated by

$$I_{sk} = m_{sk} r_{sk}^2, \tag{3}$$

where m_{sk} is the mass of the kth residue in the contact pair s and r_{sk} is the vertical distance between the kth residue and the rotational axis. Because, in a quantum transition, many torsion modes are cooperative the inertial moment of residue k should be summed. In eqs. (1) and (2), we see that the moment of inertia in transition occurs in two ways: one, as the direct sum of inertial moment of different modes called the series connection factor; and another, as the sum of reciprocal inertial moments called the parallel connection factor. Considering all contact residue pairs $s(i, j)$ in the sequence (the total number of contact pairs being N_c), we can define the dynamical contact order DCO as

$$\begin{aligned} &\text{DCO=DCO\_S+DCO\_P+}\Delta E_r \quad \text{(two-states)},\\ &\text{DCO=DCO\_S+DCO\_P+}\Delta E_r - \tau \quad \text{(multi-states)}, \end{aligned} \tag{4}$$

$$\text{DCO\_S} = -\frac{1}{2}\ln\left(\sum_{s=1}^{N_c}\sum_{k=i+1}^{j} I_{sk}\right),$$

$$\text{DCO\_P} = 2\ln\left(\sum_{s=1}^{N_c}\sum_{k=i+1}^{j} \frac{b_{sk}}{I_{sk}}\right),$$

where DCO_S is the series connection factor for the moment of inertia, DCO_P is the parallel connection factor for the moment of inertia, and the third term in eq. (4) is the conformational minimum potential energy difference in units $2K_BT$ and τ is the time delay for multi-state transitions. The values for ΔE_r and τ are listed in sections 2.3 and 2.4 of this article.

When eq. (4) is compared with eqs. (1) and (2), we find that, apart from a constant term, DCO is essentially ln$W$, the logarithm of protein folding rate.

As mentioned earlier, in calculating the inertial moment three algorithms, denoted C1, C2 and C3, were used.

C1, the link of a pair of contact residues a_i and a_j is taken as the axis of rotation. For a given contact pair, the rotational axis is fixed.

C2, for residue k between a contact pair a_i and a_j, the axis of rotation is defined by a line across i or j, perpendicular to the plane (k, i, j). The rotational axis between i and j changes with k.

C3, the inertial moment of the kth residue rotating relative to an axis across the ($k-1$)th residue and perpendicular to the link of k and $k-1$ is calculated. For a pair of contact residues a_i and a_j all the resultant moments are summed over k.

On account of the approximate proportionality of the torsion inertial moment of multipeptide chain with the sum of amino acid inertial moments between a pair of contact residues the protein inertial moments calculated from the above three models can reflect the relative magnitude of the torsion inertial moment of these multipeptide chains in the database.

2 Results

2.1 The correlation between the protein folding rate and the series connection factor of inertial moments

The correlation between DCO_S and the experimental values of lnk_f for folding rates was calculated for the 80 proteins in the data set. The results are summarized in Table 1. It can be seen that there is significant positive correlation between the series connection factor and the protein folding rate. The results are insensitive to the choice of rotational axis. The correlation coefficients for the three rotational axis cases are comparable to or higher than other predictions for the same test dataset.

2.2 The correlation between the protein folding rate and the parallel connection factor of inertial moments

Because of the difficulty in estimating the value of the elec-

Table 1 Correlation coefficients between DCO_S, DCO_P and experimental protein folding rates for 80 proteins and their comparison with other predictions[a)]

Parameter		Correlation coefficient					
		Two-state		Multi-state		All	
DCO_S	C1	0.81		0.84		0.85	
	C2	0.82		0.84		0.86	
	C3	0.83		0.85		0.87	
DCO_P[&]	C1	−0.78	−0.76	−0.65	−0.58	−0.78	−0.75
	C2	−0.82	−0.85	−0.85	−0.85	−0.87	−0.88
	C3	−0.85	−0.86	−0.84	−0.84	−0.88	−0.88
$N\alpha^*$		−0.86		−0.86		−0.83	
ACO*		−0.83		−0.64		−0.76	
RCO*		−0.53		0.06		−0.15	
L^*		−0.72		−0.79		−0.72	
Ln(L)*		−0.69		−0.84		−0.79	

a) &, two sets of correlation coefficients for the parallel connection factor were calculated for different values of the electronic quantum number b_j, namely b_j taking value stochastically between 0 and 1 and between 0 and 10, respectively. *, data are taken from Ouyang and Liang [14]. C1, C2 and C3 correspond to the three rotational axis cases.

tronic quantum number b_j in eq. (1), we adopted a simplified view and assumed that the quantum number depends only on the nature of the amino acid residues. Thus, the parallel connection factor of inertia moment is defined as

$$\sum_j \frac{b_j}{I_j} = b_{\text{gly}} \sum_j \frac{1}{I_j^{(\text{gly})}} + b_{\text{ala}} \sum_j \frac{1}{I_j^{(\text{ala})}} + \dots .$$

This simplification limits the values of the electronic quantum number b_j to 20. Because b_j is in the order of magnitude O(1), two models are assumed: one in which b_j randomly takes 20 values between 0 and 1, and the other in which b_j randomly takes 20 values between 0 and 10. The correlation coefficients between the parallel connection factor of moment of inertia and the protein folding rate are also listed in Table 1. The results show that there is a negative correlation between them. Apart from a slightly lower correlation for the C1 case, the correlation coefficients for both models using a random choice of quantum numbers are all higher than earlier predictions. It is of note that there is no significant difference in the results between the two models. Indeed, if b_j=1 is used in all cases, the correlation coefficients between the protein parallel connection factor and the folding rate are –0.75, –0.84 and –0.86 for the three rotation axis cases, very close to the result for the two random models above. For simplicity we assumed b_j=1 in all further calculations on the comparison of folding rate among different proteins.

2.3 Conformational potential energy difference ΔE_r and exergonic/endergonic folding

We have compared the theoretical calculations of protein folding rates (DCO) with the experimental values of $\ln k_f$ to gain insights into the conformational potential energy difference ΔE_r and the time delay factor τ of multi-state folding.

To calculate the DCO, we first assume that the electronic quantum number is equal to 1. We then study the error as a result of this assumption. The deviation between the theoretical and experimental folding rate (DCO–$\ln k_f$) for each protein was used to estimate the consistency level between theory and experiment. Because the magnitude of the quantum number b_j ranges from 1 to 10, it is reasonable to allow uncertainty in the value of (DCO–$\ln k_f$) to be ±2. In other words, we took $|\text{DCO}-\ln k_f|<2$ as the condition for consistency.

To satisfy the condition $|\text{DCO}-\ln k_f|<2$ for as many proteins as possible, we made a simple choice for the conformational potential energy difference, namely ΔE_r=±4 or 0 for each protein. From the definition of ΔE_r, $\Delta E_r>0$ indicates that the folding process is exergonic while $\Delta E_r<0$ indicates an endergonic folding process. Because the specific torsion potential energy curve for each protein was unavailable, we divided ΔE_r into three categories, $\Delta E_r>0$, =0, and <0. The prediction results for the differentiation of proteins between exergonic and endergonic processes are given in Table 2. From these results, we find that the high rate folding of most two-state proteins (23 proteins) and a few multi-state proteins (six proteins) is exergonic while the low rate folding of most multi-state proteins (23 proteins) and a few two-state proteins (13 proteins) is endergonic. For the remaining 15 proteins in the dataset of 80 proteins, ΔE_r is 0, indicating that the difference in conformational potential energy between the final and initial states is small.

Using eq. (4) and the obtained DCO_S, DCO_P and ΔE_r for each of the proteins, we calculated the distribution of the deviation between theoretical and experimental values for the folding rate at τ=0 and τ=3.5. The statistical results are summarized in Table 3. Detailed discussion of the results for τ=3.5 is in section 2.4.

The results in Table 3 indicate that the differences between the theoretical and experimental folding rates fall mainly in the range from –2 to 2, indicating that the theory agrees well with experiments for exergonic/endergonic proteins when ΔE_r=±4. To give a more intuitive picture of the results, the data for the C2 rotational axis case for τ = 0 are plotted in Figure 1. The results are similar for the other two rotational axis cases. However, the best-fit time delay factor for multi-state proteins is τ =3.5 (see section 2.4).

2.4 Time delay factor of multi-state protein folding

There is a time delay factor τ in the DCO for multi-state protein folding. The factor was estimated by two approaches: one in which the number of proteins that satisfied the condition $|\text{DCO}-\ln k_f|<2$ was maximized, and the other when the correlation coefficient between DCO and $\ln k_f$ was at its maximum.

When ΔE_r was set to 4 or –4 for exergonic and endergonic proteins, respectively, we observed that the number of proteins satisfying $|\text{DCO}-\ln k_f|<2$ and the correlation coefficient between DCO and $\ln k_f$ both changed with τ. The results, shown in Figure 2, indicate a maximum value for τ of between 3 and 4 for the three axis cases. The maximum protein numbers that satisfy the condition $|\text{DCO}-\ln k_f|<2$ are 60, 68 and 65, and the maximum correlation coefficients between DCO and $\ln k_f$ are 0.84, 0.91 and 0.88, for C1, C2 and C3, respectively. The distribution of DCO–$\ln k_f$ for τ = 3.5 is given in Table 3.

2.5 Testing for consistency using the "standard" data set

In addition to the 80 protein dataset [14], we used 27 two-state folding proteins from the "standard" set given by Maxwell *et al*. [1] and Segal [15] to test the consistency of our results. Assuming ΔE_r=±4, we calculated the number of proteins that satisfy the condition $|\text{DCO}-\ln k_f|<2$ and found

Table 2 Prediction results for the exergonic/endergonic folding of proteins

PDB ID	Kinetics	PDB ID	Kinetics	PDB ID	Kinetics
29 exergonic folding proteins					
1BA5	Two	1FEX	Two	1PRB	Two
1BDD	Two	1G6P	Two	1RFA	Two
1C8C	Two	1IDY	Two	1UBQ	Multi
1C9O	Two	1IMQ	Two	1UZC	Multi
1CEI	Multi	1K9Q	Two	1VII	Two
1CSP	Two	1L2Y	Two	2A3D	Two
1DIV_n	Two	1LMB	Two	2ABD	Multi
1E0L	Two	1PGB_ab	Multi	2CRO	Multi
1E0M	Two	1PGB_b	Two	2PDD	Two
1ENH	Two	1PIN	Two		
36 endergonic folding proteins					
1ADW	Multi	1HEL	Multi	1QOP_b	Multi
1APS	Two	1HNG_n	Multi	1QTU	Two
1B9C	Multi	1I1B	Multi	1RA9	Multi
1BEB	Multi	1JOO	Multi	1TEN	Two
1BTA	Multi	1K8M	Two	1TIT	Multi
1DIV_c	Two	1L63	Multi	1WIT	Two
1DK7	Multi	1OPA	Multi	2A5E	Multi
1EAL	Multi	1PHP_c	Multi	2ACY	Two
1FKB	Two	1PHP_n	Multi	2BLM	Multi
1FNF_9	Two	1PKS	Two	2HQI	Two
1HCD	Multi	1PSE	Two	2RN2	Multi
1HDN	Two	1QOP_a	Multi	3CHY	Multi

Table 3 Distribution of the deviation between theoretical and experimental folding rates[a)]

τ		DCO–lnk_f				
		(–10, –6)	(–6, –2)	(–2, 2)	(2, 6)	(6, 10)
0	C1	3(0)	10(0)	54(26)	12(8)	1(1)
	C2	1(0)	8(0)	63(27)	7(7)	1(1)
	C3	2(0)	9(0)	60(27)	8(7)	1(1)
3.5	C1	3(0)	12(2)	60(32)	4(0)	1(1)
	C2	1(0)	10(2)	68(32)	1(1)	0(0)
	C3	2(0)	11(2)	65(32)	1(0)	1(1)

a) The numbers (–10, –6), (–6, –2), etc. in the table heading indicate the range of DCO–lnk_f. The numbers in the table indicate the number of all proteins in the given range with the number of multi-state proteins in parentheses. τ is the assumed time delay in the time delay factor exp(–τ) for multi-state protein folding. C1, C2 and C3 correspond to the three rotational axis cases.

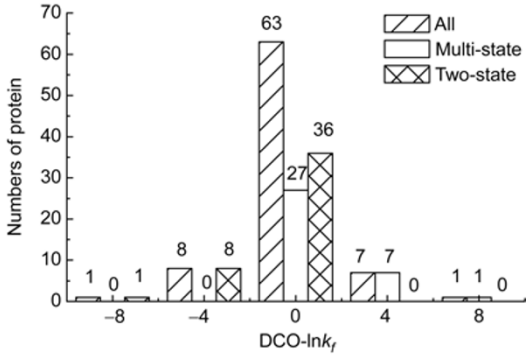

Figure 1 Distribution of the deviation between theoretical and experimental folding rates for the C2 rotational axis case for $\tau = 0$. The scale on the horizontal axis is the middle value of the DCO–lnk_f range. For example, 0 is in the range of (–2, 2) indicating the range $-2<$DCO–ln$k_f<2$.

for the C2 rotational axis case all 27 proteins satisfy the condition and 25 of the proteins satisfy the condition in the C1 and C3 axes cases. For the 27 proteins the maximum correlation coefficients between DCO and lnk_f are 0.88, 0.92 and 0.91 for C1, C2 and C3, respectively.

3 Discussion

3.1 DCO develops the concept of Plaxco contact order

Using the quantum theory of conformational dynamics we have developed the concept of Plaxco contact order and proposed a dynamical contact order (DCO). We suggest that the residue contact pair is the basic unit of quantum transitions and that coherent quantum transitions of various con-

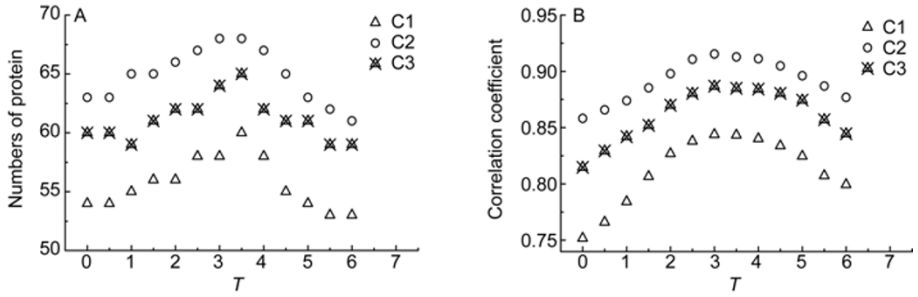

Figure 2 The change in the number of proteins satisfying $|\mathrm{DCO}-\ln k_f|<2$ with changing time delay factor τ (A) and the change in the correlation coefficient between DCO and $\ln k_f$ with changing τ (B). C1, C2, and C3 correspond to the three rotational axis cases.

tact pairs is the primitive process of protein folding. The consistency between the theoretical DCO and the experimental $\ln k_f$ supports our view that protein folding from the denatured state to the native state is a process of quantum transition. The moment of inertia is the main term that we use to define the DCO. Our calculations show that both the series factor and the parallel factor of the moment of inertia are significantly correlated with the folding rate. The Plaxco contact order is an empirical rule that is defined by the non-local contact number between residues and the distance between them in the sequence [5]. We propose that a pair of contact residues with greater sequence distance will have more couplings of residue inertial moments and will contribute more to the contact order. Further, torsion-rotational inertia is a dynamical factor that will influence the folding process. The DCO, thus, gives a deeper understanding of the dynamical essence of the Plaxco contact order.

3.2 Exergonic/endergonic folding and an upper limit for the folding rate

In the present study we emphasize the important contribution of the conformational potential energy difference ΔE to the DCO and successfully predict the exergonic and endergonic folding of proteins (Table 2). Other researchers have studied the ultrafast folding proteins [1,25] and an upper limit for the folding rate has been proposed [3]. As a general problem, the upper limit of the protein folding rate can be estimated by Kramer's model and polymer diffusion theory [26,27], by using the nucleation-condensation mechanism [28,29], by the thruway search model [4] or by other methods. None of these methods consider folding as a quantum transition process. In the quantum theory of folding that we now propose the main decision factor for ultrafast folding is the ΔE_r term in the DCO. We have shown that when $\Delta E_r>0$ the exergonic folding process is significantly accelerated. We have successfully classified 80 proteins into three categories according to their ΔE symbol (positive, negative or zero) and roughly determined the value of ΔE as equal to $+8K_BT$, $-8K_BT$ or 0. For a more rigorous approach, the values of ΔE_r should be calculated for each protein because even for proteins of the same category, the ΔE_r may be different. A more accurate determination will be carried out in future work. Here we compare our predicted exergonic folding proteins with the high-speed folding proteins listed by Kubelka *et al*. [3]. Of the 10 high-speed folding proteins listed, nine proteins have folding free energies of less than 0 ($\Delta G<0$). This is consistent with our prediction results shown in Table 2. The experimental data for these nine proteins are listed in Table 4. When comparing our results with experiments we should note that ΔE indicates the difference in torsion energy between the initial state and the final state while ΔG usually denotes the free energy difference between the final state and the initial state so that, $\Delta E>0$ corresponds to $\Delta G<0$. We should also note that the experimental free energy ΔG includes all changes of energy in the reaction and not only the torsion energy. For example, the conformation-adjusting energy after a coherent quantum transition between torsion states may be important for arriving at the native state. This energy is not included in ΔE, but will be included in the measured folding free energy ΔG. We found that the ΔG for the predicted endergonic protein, 2A5E, was less than 0 [30]. This discrepancy between the theoretical and experimental results may be because the conformation-adjusting energy after quantum transition is not included in ΔE.

3.3 New understanding of the mechanism of multi-state protein folding

The mechanism of multi-state folding was without a clear

Table 4 Folding free energy of nine exergonic proteins

PDB ID	The natural logarithm of folding rate	Kinetic type	Protein Sequence length	$\Delta G^{a)}$ (kcal mol^{-1})
1E0L	10.37	Two	37	–1.7
1ENH	10.53	Two	54	–2.1
1L2Y	12.4	Two	20	–0.7
1LMB	8.5	Two	87	–3.0
1PIN	9.37	Two	34	–1.9
1PRB	12.9	Two	53	–2.6
1VII	11.51	Two	36	–0.6
2A3D	12.7	Two	73	–1.9
2PDD	9.69	Two	42	–1.1

a) Folding free energy ΔG data are from [3].

theoretical basis [31] until Kamagata *et al*. [24] indicated that the folding rates for non-two-state proteins show a similar dependence on the native backbone topological parameters as the folding rates for two-state proteins. Our statistical analyses of the correlation of DCO_S and DCO_P with the protein folding rate (Table 1) also show that there is no difference between two-state and non-two-state proteins. Following the work of Kamagata *et al*. [24] and from our own statistical analyses, we propose that multi-state protein folding can be described as a joint process of several quantum transitions that occur at different degrees of freedom of the torsion angle. The proposal is supported by its consistency with quantum transition theory [21]. The folding rate of a multi-state protein can be expressed by the formula for two-state folding with an additional factor $e^{-\tau}$ that indicates the time delay while in the intermediate state. The time delay factor $e^{-\tau}$ will be different for different proteins but, as a first approximation and to emphasize the difference between them and two-state proteins, we assume a common value for all multi-state proteins. The results of the statistical analyses described in section 2.4 support our understanding of the multi-state protein folding mechanism. From our calculations we obtained a best-fit value of τ=3.5 for the time delay caused by the intermediate state.

We have shown that using DCO to kinetically study protein folding rates gives a simple picture of multi-state protein folding that captures certain essentials of the mechanisms that are involved in the process. A more detailed study of the time delay parameter τ is required. Further, the threshold distance for a contact pair in multi-state folding requires further investigation because it may be different from, perhaps larger than, that for a contact pair in a two-state protein. These problems will be studied in subsequent research.

This work was supported by the Distinguished Scientist Award of Inner Mongolia Autonomous Region (2008), a Major Project Fund of Inner Mongolia University of Technology (Grant No. ZD200917), and a Project Fund of Inner Mongolia Natural Science (Grant No. 2010BS0104). We thank Dr. Lu Jun for his help in carrying out the computational and statistical analyses.

1 Maxwell K L, Wildes D, Zarrine-Afsar A, *et al*. Protein folding: Defining a ''standard'' set of experimental conditions and a preliminary kinetic data set of two-state proteins. Protein Sci, 2005, 14: 602–616
2 Qiu L, Pabit S A, Roitberg A E, *et al*. Smaller and faster: The 20 residue Trp-cage protein folds in 4 μs. J Am Chem Soc, 2002, 124: 12952–12953
3 Kubelka J, Hofrichter J, Eaton W A. The protein folding 'speed limit'. Curr Opin Struct Biol, 2004, 14: 76–88
4 Ghosh K, Ozkan B, Dill K A. The ultimate speed limit to protein folding is conformational searching. J Am Chem Soc, 2007, 129: 11920–11927
5 Plaxco K W, Simons K T, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol, 1998, 227: 985–994
6 Ivankov D N, Garbuzynskiy S O, Alm E, *et al*. Contact order revisited: Influence of protein size on the folding rate. Protein Sci, 2003, 12: 2057–2062
7 Ivankov D N, Finkelstein A V. Prediction of protein folding rates from the amino acid sequence-predicted secondary structure. Proc Natl Acad Sci USA, 2004, 101: 8942–8944
8 Shen H B, Song J N, Chou K C. Prediction of protein folding rates from primary sequence by fusing multiple sequential features. J Biomed Sci Eng, 2009, 2: 136–143
9 Huang J T, Tian J. Amino acid sequence predicts folding rate for middle-size two-state proteins. Proteins, 2006, 63: 551–554
10 Huang L T, Gromiha M M. Analysis and prediction of protein folding rates using quadratic response surface models. J Comput Chem, 2008, 29: 1675–1683
11 Capriotti E, Casadio R. K-Fold: A tool for the prediction of the protein folding kinetic order and rate. Bioinformatics, 2007, 23: 385–386
12 Gromiha M M, Thangakani A M, Selvaraj S. FOLD-RATE: Prediction of protein folding rates from amino acid sequence. Nucleic Acids Res, 2006, 34: W70–W74
13 Zhang L, Sun T. Folding rate prediction using *n*-order contact distance for proteins with two- and three-state folding kinetics. Biophys Chem, 2005, 113: 9–16
14 Ouyang Z, Liang J. Predicting protein folding rates from geometric contact and amino acid sequence. Protein Sci, 2008, 17: 1256–1263
15 Segal M R. A novel topology for representing protein folds. Protein Sci, 2009, 18: 686–693
16 Jang H, Hall C K, Zhou Y. Protein folding pathways and kinetics: Molecular dynamics simulations of β-strand motifs. Biophys J, 2002, 83: 819–835
17 Schonbrun J, Dill K A. Fast protein folding kinetics. Proc Natl Acad Sci USA, 2003, 100: 12678–12682
18 Luo L F. Conformation transitional rate in protein folding. Int J Quant Chem, 1995, 54: 243–247
19 Luo L F. Theoretic-physical Approach to Molecular Biology. Shanghai: Shanghai Scientific & Technical Publishers, 2004. 388–402
20 Luo L F. Protein folding as a quantum transition between conformational states. arXiv: 0906.2452v1, 2009, URL: http://arxiv.org/abs/0906.2452
21 Luo L F. Protein folding as a quantum transition between conformational states: Basic formulas and applications. arXiv: 1008.0237, 2010, URL: http://arxiv.org/abs/1008.0237
22 Bernstein F C, Koetzle T F, Williams G J, *et al*. The Protein Data Bank: A computer-based archival file for macromolecular structures. Eur J Biochem, 1977, 80: 319–324
23 Berman H M. The Protein Data Bank: A historical perspective. Acta Cryst, 2008, A64: 88–95
24 Kamagata K, Aral M, Kuwajima K. Unification of the folding mechanism of non-two-state and two-state proteins. J Mol Biol, 2004, 339: 951–965
25 Zhu Y, Alonso D O V, Maki K, *et al*. Ultrafast folding of α_3D: A de novo designed three-helix bundle protein. Proc Natl Acad Sci USA, 2003, 100: 15486–15491
26 Hagen S J, Hofrichter J, Szabo A, *et al*. Diffusion-limited contact formation in unfolded cytochrome c: Estimating the maximum rate of protein folding. Proc Natl Acad Sci USA, 1996, 93: 11615–11617
27 Jacob M, Schindler T, Balbach J, *et al*. Diffusion control in an elementary protein folding reaction. Proc Natl Acad Sci USA, 1997, 94: 5622–5627
28 Fersht A R. Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications. Proc Natl Acad Sci USA, 1995, 92: 10869–10873
29 Debe D A, Goddard III W A. First principles prediction of protein folding rates. J Mol Biol, 1999, 294: 619–625
30 Tang K S, Guralnick B J, Wang W K, *et al*. Stability and folding of the tumour suppressor protein p16. J Mol Biol, 1999, 285: 1869–1886
31 Galzitskaya O V, Garbuzynskiy S O, Ivankov D N, *et al*. Chain length is the main determinant of the folding rate for proteins with three-state folding kinetics. Proteins, 2003, 51: 162–166