Folding Rate of Protein and RNA Studied from Quantum Folding Theory

Liaofu Luo [1*] Jun Lv [2*]

[1] *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*

[*] corresponding author. email address 1, lolfcm@mail.imu.edu.cn ; 2, lujun@imut.edu.cn

ABSTRACT

Starting from the assumption that the protein and RNA folding is an event of quantum transition between molecular conformations，we deduced a folding rate formula and studied the chain length (torsion number) dependence and temperature dependence of the folding rate. The chain length dependence of the folding rate was tested in 65 two-state proteins and 27 RNA molecules. The success of the comparative study of protein and RNA folding reveals the possible existence of a common quantum mechanism in the conformational change of biomolecules. The predicted temperature dependence of the folding rate has also been successfully tested for proteins. Its further test in RNA is expected.

Keywords

RNA folding rate; protein folding rate; conformational change; quantum transition; molecular torsion; chain length; free energy.

INTRODUCTION

In a recent work Garbuzynskiy et al reported that the measured protein folding rates fall within a narrow triangle (called Golden triangle)[1]. Simultaneously, Hyeon et al reported that RNA folding rates are determined by chain length. [2] Both protein and RNA are biological macromolecules. They may obey the same dynamical laws and a unifying folding mechanism is expected [3]. We have proposed a quantum theory on protein folding [4] . Following the idea that the conformational change of biomolecule is essentially a quantum transition between conformational states we shall make comparative studies on two-state protein folding and RNA folding and give a unifying approach to find the folding dynamics of both molecules.

For a macromolecule consisting of n atoms there are $3n$ coordinates if each atom is looked as a point. Apart from 6 translational and rotational degrees of freedom there are $3n$-6 coordinates describing molecular shape. The molecular shape is the main variables responsible for conformational change. It has been proved that the bond lengths, bond angles and torsion (dihedral) angles form a complete set to describe the molecular shape. As compared with chemical bond energy (typically in several electron volts) , the vibrational energy of bond length and bond angle (in the range of 0.4-0.03 ev) and other forms of biological energies, the torsion vibration energy (about 0.03-0.003 ev) is the lowest and therefore constitutes the slow variables of the molecular biological

system. Moreover, different from stretching and bending the torsion potential generally has several minima with respect to angle coordinate that correspond to several stable conformations. Based on the idea that the molecular conformation is defined by torsion state and the folding/unfolding is essentially a quantum transition between them, through adiabatically elimination of fast variables we obtain a set of fundamental equations to describe the rate of conformational transition of macromolecule.[4] By use of these equations we have successfully explain the non-Arrhenius temperature dependence of the folding rate (the logarithm folding rate is not a decreasing linear function of 1/T） for each protein. Moreover, the statistical investigation of 65 two-state protein folding rates shows the fundamental equations are consistent with experimental data [5].

In the article, based on the rate equation deduced from quantum transition theory a unifying investigation on protein folding and RNA folding will be given. Firstly we shall deduce a relation between folding rate and chain length. Then we shall test the relation in 65 two-state protein dataset and use the relation to analyze the folding rate data of 27 RNA molecules. We shall also give the relation of folding rate with temperature and discuss its experimental implication.

MATERIALS AND METHODS

Datasets Recently Garbuzynskiy and coworkers collected folding rate data for 69 two-state proteins [1]. Of the 69 proteins, the folding rates of 65 two-state proteins are obtained at around 25 °C. They constitute a dataset used by us to compare the theoretical vs. experimental results (Table S1 in the Supplementary data). Hyeon and Thirumalai collected the folding rates of 27 RNA molecules [2]. They constitute the second dataset we shall use. In addition, the temperature dependence data of the folding rate for 16 proteins are given in Table S2 in the Supplementary data.

Theoretical model Suppose the quantum state of a macromolecule is described by a wave function $M(\theta, x)$ where $\{\theta\}$ the torsion angles of the molecule and $\{x\}$ the set of fast variables including the stretching-bending coordinates and the frontier electrons of the molecule, etc. The wave function $M(\theta,x)$ satisfies

$$(H_{tor}(q,\frac{\partial}{\partial q})+H_{fv}(x,\frac{\partial}{\partial x};q))M(q,x)=EM(q,x) \tag{1}$$

$$H_{tor}=\sum -\frac{\mathbf{h}^2}{2I_j}\frac{\partial^2}{\partial q_j^2}+U_{tor}(q) \tag{2}$$

where I_j denotes the inertial moment of the j-th torsion and the torsion potential U_{tor} is a function of a set of torsion angles $q=\{q_j\}$. Its form is dependent of solvent environment of the molecule. H_{fv} is fast-variable Hamiltonian. Because the fast variables change more quickly than the variation of torsion angles, the adiabatic approximation can be used. In adiabatic approximation the wave function is expressed as

$$M(q,\ x)=y(q)j(x,\ q) \tag{3}$$

and these two factors satisfy

$$H_{fv}(x,\ \frac{\partial}{\partial x};q)j_a(x,\ q)=e_a(q)j_a(x,\ q) \tag{4}$$

$$\{H_{tor}(q,\ \frac{\partial}{\partial q})+e_a(q)\}y_{kna}(q)=E_{kna}y_{kna}(q) \tag{5}$$

here α denotes the quantum number of fast-variable wave function j , and (k, n) refer to the conformational (indicating which minimum the wave function is localized around) and the vibrational state of torsion wave function y , respectively.

Because $M\ (q,x)$ is not a rigorous eigenstate of Hamiltonian H_{tor} + H_{fv}, there exists a transition between adiabatic states that results from the off–diagonal elements

$$\langle k'n'a' \mid H' \mid kna \rangle = \int y^{+}_{k'n'a'}(q) \sum_j -\frac{\mathbf{h}^2}{2I_j} \{ \int j^{+}_{a'} (\frac{\partial^2 j_a}{\partial q_j^{\ 2}} + 2\frac{\partial j_a}{\partial q_j} \ \frac{\partial}{\partial q_j}) dx \} y_{kna}(q) dq \qquad (6)$$

Here H' is a Hamiltonian describing conformational transition. The nonadiabatic matrix element (6) can be calculated under the perturbation approximation. Through tedious calculation we obtain the rate of conformational transition [4]

$$W = \frac{2p}{\mathbf{h}^2 \overline{w}'} I_V^{'} I_E^{'}$$

$$I_V^{'} = \frac{\mathbf{h}}{\sqrt{2p}\, dq} \exp\{\frac{\Delta G}{2k_B T}\} \exp\{\frac{-(\Delta G)^2}{2\overline{w}^2 (dq)^2 k_B T \sum_j^N I_j}\} (k_B T)^{1/2} (\sum_j^N I_j)^{-1/2} \qquad (7)$$

$$I_E^{'} = \sum_j^M \left| a_{a'a}^{(j)} \right|^2 = M\overline{a}^2$$

where W means the rate of conformational transition at given temperature T and solvent condition, 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 hereafter), $\overline{w}$ and $\overline{w}'$ are the initial and final frequency parameters w_j and w_j' of torsion potential averaged over N torsion modes, respectively, dq is the averaged angular shift between initial and final torsion potential, ΔG is the free energy decrease per molecule between initial and final states, M is the number of torsion angles correlated to fast variables, $\overline{a}^2$ 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,

$$\overline{a}^2 = \frac{1}{M} \sum_j^M \left| a_{a'a}^{(j)} \right|^2 ,$$

$$a_{a'a}^{(j)} = \frac{i\mathbf{h}}{I_j^{1/2}} \frac{h_{a'a}^{(j)}}{e_a^{(0)} - e_{a'}^{(0)}} \qquad (a' \neq a)$$

$$h^{(j)}(x, \frac{\partial}{\partial x}) = (\frac{\partial H_{fv}(x, \frac{\partial}{\partial x}; q)}{\partial q_j})_0 \qquad (8)$$

$e_a^{(0)}, e_{a'}^{(0)}$ are the eigenvalues of $H_{fv}(x, \frac{\partial}{\partial x}; q_0)$.

Eqs (7)(8) are basic equations for conformational transition. To obtain quantitative result one

should calculate the number of torsion modes N in advance. N describes the coherence degree of multi-torsion transition in the folding. For two-state protein folding we assume that N can be obtained by 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 2 main-chain dihedral angles and, for non-alanine and -glycine, it contributes 1 - 4 additional side-chain dihedral angles (Table S3 in the Supplementary data). For RNA folding, we assume the quantum transition occurs between compact (yet disordered) intermediate and folding state [8] or between primary and secondary structures of the molecule [9]. The torsion number can be estimated by chain length. Following IUB/IUPAC there are 7 torsion angles for each nucleotide, namely

$$
\begin{aligned}
&a\,(\mathrm{O3'-P-O5'-C5'}),\\
&b\,(\mathrm{P-O5'-C5'-}C4'),\\
&g\,(\mathrm{O5'-C5'-}C4'\mathrm{-C3'}),\\
&d\,(\mathrm{C5'-}C4'\mathrm{-C3'-O3'}),\\
&e\,(C4'\mathrm{-C3'-O3'-P}),\\
&V\,(\mathrm{C3'-O3'-P-O5'})
\end{aligned}
$$

and

$$c\,(\mathrm{O4'-C1'-N1-C2})\ \text{(for Pyrimidine) or}\ \ c\,(\mathrm{O4'-C1'-N9-C2})\ \text{(for Purine),}$$

of which many have more than one advantageous conformations (potential minima). If each nucleotide has q torsion angles with multi-minima in potential then the torsion number $N=qL$, where L is chain length of RNA.

RESULTS AND DISCUSSIONS

Obtaining a relation of the fast-variable factor I'_E with respect to torsion number N.

For protein (and RNA) folding or other macromolecular conformational change not involving chemical reaction and electronic transition the fast variable includes only bond lengths and bond angles of the macromolecule. In this case an approximate relation of the fast-variable factor I'_E with respect to torsion number N can be deduced. When the kinetic energy in $H_{fv}(x, \frac{\partial}{\partial x}; q)$ is neglected as compared with interaction potential U one has

$$
\begin{aligned}
a_{a'a}^{(j)} &= \frac{i\mathbf{h}}{I_j^{1/2}} \frac{1}{e_a^{(0)} - e_{a'}^{(0)}} \int j_{a'}^{*}(\mathbf{r},\mathbf{r}')\left(\frac{\partial U(|\mathbf{r}-\mathbf{r}'|,q)}{\partial q_j}\right)_{q_0} j_a(\mathbf{r},\mathbf{r}')d^3\mathbf{r}d^3\mathbf{r}' \\
&= \frac{i\mathbf{h}}{I_j^{1/2}} \frac{1}{V(e_a^{(0)} - e_{a'}^{(0)})} \int \left(\frac{\partial U(|\mathbf{r}-\mathbf{r}'|,q)}{\partial q_j}\right)_{q_0} d^3(\mathbf{r}-\mathbf{r}')
\end{aligned}
\tag{9}
$$

In the above deduction of the second equality the fast-variable wave function $j_a(\mathbf{r},\mathbf{r}')$ has been assumed to be a constant and normalized in the volume V. As the energy and volume V are dependent of the size of the molecule one may assume energy $e_a^{(0)}$ and U proportional to the

interacting-pair number (namely N^2) and V proportional to N. However, because only a small fraction of interacting-pairs correlated to given $q_j\,(j=1,...,N)$, $(\frac{\partial U(|\mathbf{r}-\mathbf{r}'|,q)}{\partial q_j})_{q_0}$ does not increases with N. So, one estimate $a_{a'a}^{(j)} \approx N^{-3}$. On the other hand, the integral $\int(\frac{\partial U(|\mathbf{r}-\mathbf{r}'|,q)}{\partial q_j})_{q_0} d^3(\mathbf{r}-\mathbf{r}')$ may depend on the molecular structure. For example, the high helix content makes the integral increasing. It was indicated that a protein with abundant α helices may have a quite oblong or oblate ellipsoid, instead of spheroid, shape and this protein has higher folding rate[1][5]. Therefore, apart from the factor N^{-3} there is another structure-related factor in $a_{a'a}^{(j)}$. The latter is N-independent. Assuming M proportional to N, one obtains

$$M\overline{a}^2 = cfN^{-5} \tag{10}$$

where f is a structure-related shape parameter. It means the fast-variable factor I_E' is inversely proportional to N^5. With $\sum_j^N I_j \cong NI_0$ and Eq (10) inserted into Eq (7) we obtain the relation of logarithm rate with respect to N and ΔG

$$\ln W = \frac{\Delta G}{2k_BT} - \frac{(\Delta G/k_BT)^2}{2rN} - 5.5\ln N + \ln c_0 f \tag{11}$$

where

$$r = I_0\overline{w}^2(dq)^2/(k_BT) \tag{12}$$

is a torsion-energy-related parameter and $c_0 = \frac{\sqrt{2p}}{\mathbf{h}dq}(\frac{k_BT}{\overline{w}'^2 I_0})^{1/2}c$ is an N-independent constant.

The relationship of ln$W$ with N given by Eq (11) can be tested by the statistical analyses of 65 two-state protein folding rates k_f. We found that the theoretical logarithm rate ln$W$ is in good agreement with the experimental lnk_f (Fig 1). The correlation coefficient R has attained higher than 0.78.

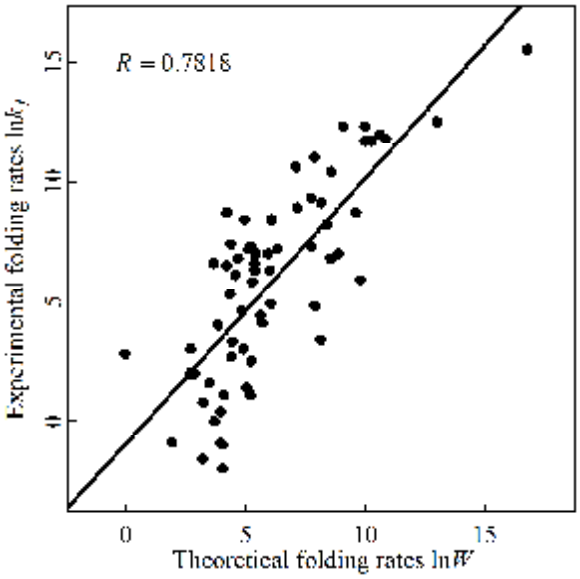

Figure 1 Comparison of theoretical folding rates ln$W$ with experimental folding rates lnk_f for 65 proteins. The experimental rates k_f are taken from Table S1. The straight line is the linear regression between ln$W$ and lnk_f . In calculation of theoretical ln$W$ by use of Eq 11 the shape parameter f is taken as follows:, f=81 for $(L_\alpha - L_\beta)/L \geq 0.6$, f=25 for $0.3 \leq (L_\alpha - L_\beta)/L < 0.6$, 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 is insensitive to the statistical result. The figure is plotted for $r = 0.097$ (with correlation R=0.7818 and slop of regression line 1.109). For any r between 0.06 and 0.1 the basically same results are obtained, for example, R=0.7537 and slop=1.044 for $r = 0.069$, R=0.7396 and slop=0.997 for $r = 0.06$.

Obtaining a relation of the free energy ΔG with respect to torsion number N and testing it in protein folding dataset To find the relation between free energy ΔG and torsion number N we consider the statistical relation of free energy combination $\frac{\Delta G}{2k_B T} - \frac{(\Delta G)^2}{2(k_B T)^2 r N}$ that occurs in rate equation (7) or (11). Set

$$\frac{\Delta G}{2k_B T} - \frac{(\Delta G)^2}{2(k_B T)^2 r N} = y, \qquad -\frac{1}{N} = x \tag{13}$$

The linear regression between y and x is given by $y = A + Bx$ where A and B are two statistical parameters describing free energy distribution in the dataset. We will test the linear relation in protein folding dataset. Due to the ignorance of the accurate ρ-value for each protein one can test the relation by using the single-ρ-fit (assuming a single ρ-value to deduce a linear regression) at first, then compare the fitting results and find the best-fit ρ-value and the corresponding free-energy statistical parameter A and B . The statistical results (correlation R and parameter A and B) in two-state protein dataset are listed in Table 1. From Table 1 we find the correlation R between y and x is near to 0.8 for ρ= 0.065~0.075 and reaches maximum at ρ= 0.069 where R=0.7966. Thus, by single-ρ-fit we obtain the best-fit statistical relation of free energy for two-state proteins as

$$y = 4.306 + 541.1x \qquad (\rho = 0.069) \tag{13a}$$

(Figure 2). In above discussion the single-ρ-fit has been used. Evidently, as the variation of ρ for

different proteins is taken into account the linear regression between free energy combination y and torsion number x will be further improved.

Table 1 Free energy parameters determined by linear regression

ρ	0.05	0.060	0.065	0.069	0.075	0.097
R	0.6573	0.7759	0.7934	0.7966	0.7919	0.7507
A	3.194	3.867	4.126	4.306	4.540	5.152
B	428.0	496.4	522.8	541.1	564.9	626.9

A and B - free energy parameter, R- correlation coefficient. R reaches maximum at ρ=0.069.

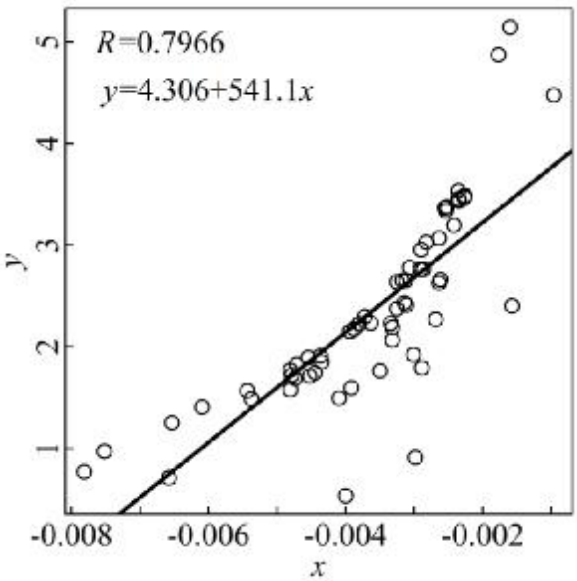

Figure 2 Statistical relation of free energy for two-state proteins. Experimental data are taken from 65-protein set (Table S1). Five proteins in the set denatured by temperature have been omitted in our statistics.

About the relationship of free energy ΔG with torsion number N two statistics were done in literatures. One was based on the assumption of linear relation existing between ΔG and $\sqrt{N}$, $\Delta G = a\sqrt{N} - b$ ($b \neq 0$) [5]. Another was based on the assumed relation of ΔG *vs* ($Lg + s\, B_L L^{2/3}$)(L - the length of polypeptide chain)[1]. By the statistics on 65 two-state proteins in the same dataset we demonstrated the correlation R between free energy and N-related quantity is 0.67 for the former and 0.69 for the latter [5], both lower than the correlation shown in Fig 2. The free energy combination of $\frac{\Delta G}{2k_B T} - \frac{(\Delta G)^2}{2(k_B T)^2 r N}$ occurring in the folding rate is an important peculiarity of the present theory. We shall use the statistical relation of the free-energy combination versus N in the following studies on RNA folding.

Testing the relation of the folding rate with respect to torsion number N in RNA folding dataset

In virtue of Eqs (11) (13) we obtain an approximate expression for transitional rate ln$W$ versus

N for protein folding

$$\ln W = A - \frac{B}{N} - D\ln N + const \tag{14}$$

$$D=5.5, \quad A = 4.306, \quad B = 541.1$$

Here *const* is an N-independent constant but dependent of molecular shape (due to the factor f contained in it).

The quantum folding theory of protein is usable in principle for each step of the conformational transition of RNA molecule. Although recent experiments have revealed the multi-stages in RNA collapse, the final search for the native structure within compact intermediates seems a common step in the folding process. In the meantime it exhibits strong cooperativity of helix assembly [3][7]. In calculation of the transition from intermediate to native fold, as the collapse transition prior to the formation of intermediate is a fast process and the time needed for the latter is generally shorter than the former [7] , the calculation result can be directly compared with the experimental data of total rate. Moreover, for RNA folding the *const* term in Eq (14) is a real constant if the variation of the structure-related shape parameter f can be neglected in the considered dataset. By using $N=qL$ (L is the chain length of RNA) we have

$$\ln W = A - \frac{B'}{L} - D\ln L + c' \qquad (B'=B/q,\ c' = const - D\ln q) \tag{15}$$

Eq (15) is deduced from quantum folding theory and it predicts the relation of folding rate versus chain length: the rate W increasing with L, attaining the maximum at $L_{\max}=B'/D$, then decreasing with power law L^{-D}.

In a recent work Hyeon and Thirumala [2] indicated that the chain length determines the folding rates of RNA. They obtained a good statistical relation between folding rates and chain length L in a dataset of 27 RNA sequences. Their best-fit result is

$$\log W_H = 14.3 - 1.15 \times L^{0.46} \tag{16}$$

Both equations (15) and (16) give relation between RNA folding rate and chain length. Comparing the theoretical folding rates ln W or ln W_H with the experimental folding rates ln k_f in 27 RNA dataset the results are shown in Figure 3 . We find Eq (15) can fit the experimental data on RNA folding rate equally well as Eq (16). By using the best-fit value of B' and D the correlation between ln W (calculated from Eq 15) and ln k_f is R=0.9729 (Fig 3a)，while the correlation between ln W_H (calculated from Eq 16) and ln k_f is R=0.9752(Fig 3b). However, in Fig 3b the slope of the regression line is 1.03 and the line deviates from origin by -0.36 , while in Fig 3a the slope is 1.0001, very close to 1 and the line deviates from origin only by -0.0012. The reason is: although two equations have the same overall accuracy in fitting experimental data, but for large L cases the errors $Er = \left|\log W - \log k_f\right|$ calculated from Eq (15) are explicitly lower than $Er_H = \left|\log W_H - \log k_f\right|$

from Eq (16) (Table 2). It means the folding rate lowers down with increasing L as $L^{-D}(D \cong 5.5)$ at large L (a long-tail existing in the W-L curve) rather than a short tail as $\exp(-l\sqrt{L})$ assumed in [2].

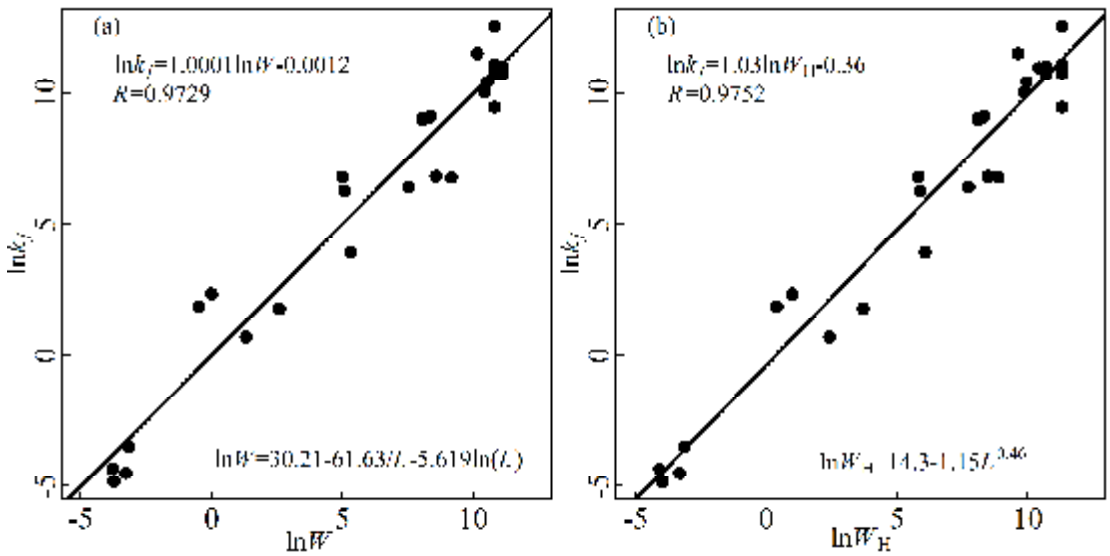

Figure 3 Comparison of experimental folding rates $\ln k_f$ with theoretical folding rates $\ln W$ (Fig 3a) or $\ln W_H$ (Fig 3b) for 27 RNA molecules. Experimental rates are taken from Table 1 in literature [2]. Theoretical rates are calculated from Eq (15) and (16) respectively.

Table 2 Errors of RNA folding rates in two theoretical models compared with experimental data

	1	2	3	4	5	6	7	8
L	125	160	205	225	368	377	409	414
k_f(s^{-1})	6	2	10.5	6.5	.03	.011	.008	.013
Er_H	0	0.17	3.27	3.37	1.52	0.71	1.06	1.66
Er	0	0.18	3.15	3.17	0.44	0.42	0.30	0.26

$Er_H = \left|\log W_H - \log k_f\right|$, $Er = \left|\log W - \log k_f\right|$,

1, hairpin ribozyme [9][10]; 2; P4-P6 domain (*Tetrahymena* ribozyme)[11]; 3, *Azoarcus* ribozyme [10][12]; 4, *B. subtilis* RNase P RNA catalytic domain[13]; 5, Ca.L-11 ribozyme[14]; 6, *E. coli* RNase P RNA[15]; 7, *B. subtilis* RNase P RNA[15] and 8, *Tetrahymena* ribozyme[10][16]. Experimental rates k_f's can be found in literatures [9-16]. The errors of eight RNAs with lengths larger than 120 in 27-RNA dataset are analyzed in the table. The errors of the first RNA hairpin ribozyme are normalized to zero in two models.

There are two independent parameters in RNA folding rate Eq (15), B' and D, apart from the additive constant. As seen from Fig 3a we obtain the best-fit D value D_f=5.619 on the 27-RNA dataset, close to D=5.5 predicted from a general theory of quantum folding. Simultaneously we obtain the best-fit B' value B'_f=61.63. The B'_f value derived from RNA folding can be compared with the B value from protein folding (Eq (14)). Notice that B or $B'=B/q$ represents the

contribution from free energy square term $\frac{(\Delta G)^2}{2(k_B T)^2 r N}$ in logarithm folding rate. The RNA folding free energy is typically 2 to 4 kcal / mol [7, 17,18] while the folding free energy for most proteins in 65-protein dataset is between 1 and 4.6 kcal /mol. (Table S1). On the other hand, the r - value varies from about 0.03 (for fast folds) to 0.1 as seen from the statistical analysis of protein data [5]. Although both $(\Delta G)^2$ and r changes with N the variation of $\frac{(\Delta G)^2}{r}$ with N in a given dataset may be more weak. It is plausible to assume the mean $\frac{(\Delta G)^2}{r}$ for protein differs that for RNA by a factor no larger than 1.5 to 2. If $\left\langle \frac{(\Delta G)^2}{r} \right\rangle_{prot}$ =(1.5~2) $\left\langle \frac{(\Delta G)^2}{r} \right\rangle_{RNA}$ then the B –value for RNA, B_{RNA} ,will be smaller than B_{protein}=541.1 and takes a value in the range 270-360. Comparing with B'_f =61.63 it leads to q=4.4 ~ 5.8, consistent with the theoretical upper limit q =7 for RNA molecule.

The 27-RNA dataset [2] contains data of experimental folding rates measured in different processes. The dataset is inhomogeneous, including several subsets, one subset of total folding rates for some RNAs and another subset of rates of secondary structure formation for other RNAs, etc. In using Eq (15) one should notice the parameter difference among subsets. All D's in different subsets are near to 5.5 by theoretical grounds. If the variation of B' (and additive constant) among subsets can be neglected then one can use Eq (15) with single parameter B' , D and additive constant to fit the experimental folding rates. We wonder why the simple formula Eq (15) can fit the experiments so well. Here gives an explanation. However, the dissimilarity of free energy parameters B (and A) and the variation of structure-related parameter f in different subsets do exist. For example, as the nucleotide G in the tetraloop hairpin UUCG is substituted by 8-bromoguanosine $\overline{G}$, the folding rate of gcUUC$\overline{G}$gc is 4.1-fold faster than gcUUCGgc.[8] Both samples were collected in the dataset and they have the same chain length L=8 but different rate k_f . From the present theoretical model, the difference comes from the variation of the free energy change ΔG and the structure-related parameter f in two samples (see Eq (11)). This explains the origin of the error of the model fit to the RNA folding rate by using Eq (15).

Obtaining a law on the temperature dependence of folding rate and testing it in protein dataset The free energy decrease ΔG in protein folding is linearly dependent of temperature T (Figure S2 in the Supplementary data). Inserting the linear relation into Eq(7) we obtain the temperature dependence of the transition rate as

$$\ln W(T) = \frac{S}{T} - RT + \frac{1}{2}\ln T + const. \tag{17}$$

It means the non-Arrhenius behavior of the rate–temperature relationships. The relation was tested for 16 two-state proteins whose temperature dependence data were available (Table S2 in the Supplementary data). The statistical analyses were made in [5] (see Figure S1 and Table S4 in the Supplementary data). Figure 4 gives an example. The strong curvature on Arrhenius plot is due to the R term in Eq (17) which comes from the square free energy $(\Delta G)^2$ in ln$W$. The good agreement between theory and experiments affords a support to the concept of quantum folding. Moreover, in

this 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. All parameters related to torsion potential defined in this theory (such as torsion frequency $\bar{w}$ and $\bar{w}'$, averaged angular shift dq and energy gap ΔE between initial and final torsion potential minima, etc) can be determined. They can be calculated consistently with each other for all studied proteins. Furthermore，in this theory the folding and unfolding rates are correlated with each other, needless of introducing any further assumption [5][19].

For RNA molecule the temperature dependence of folding rate for yeast $tRNA^{phe}$ was observed [20]. They measured the logarithm folding rate $\ln k_f$ versus $1/T$ between 28.5°C and 34.8°C. The Arrhenius plot shows a straight line in this temperature interval but large standard deviation existing at low temperature end. From the experiments on protein folding, the strong curve of the $\ln k_f$ - $1/T$ relation only occurs in a temperature interval of several tens degrees. We expect more accurate measurements within a large enough temperature interval will be able to exhibit the non-Arrhenius peculiarity of the temperature dependence of the RNA folding rate.

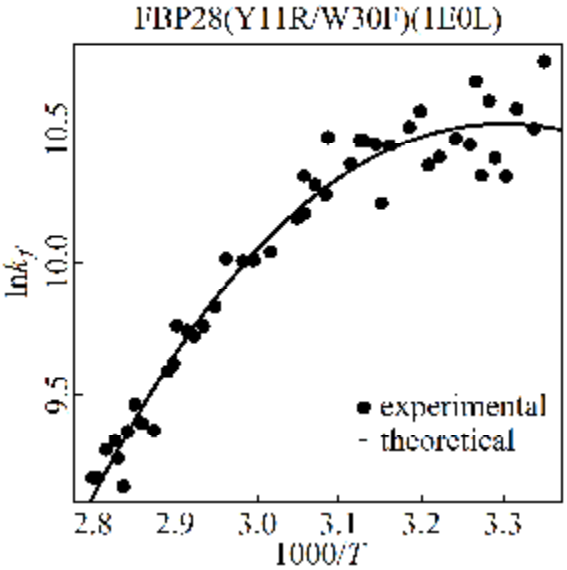

Figure 4 Model fits to overall folding rate k_f *vs* temperature $1000/T$ for protein FBP28(PDB code 1E0L). Experimental logarithm folding rates are shown by "o", and solid lines are theoretical model fits to the folding rate (k_f in unit s^{-1}, T in unit Kelvin). Experimental rates are taken from [22].

Remarks Eq (7) can be used for unfolding as well as for folding. The unfolding rate W(unfolding) is easily obtained by the replacement of ΔG by $-\Delta G$ and $\bar{w}'(\bar{w})$ by $\bar{w}$ $(\bar{w}')$ in W(folding). Thus we have

$$\ln\{\frac{W(\text{folding})}{W(\text{unfolding})}\} = \frac{\Delta G}{k_B T} + \frac{(\Delta G)^2}{2k_B T e}(\frac{\bar{w}^2 - \bar{w}'^2}{\bar{w}'^2}) + \ln\frac{\bar{w}}{\bar{w}'} \tag{18}$$

($e = \bar{w}^2 (dq)^2 \sum_j^N I_j$). Eq (18) means the condition of dynamical balance $W(\text{folding}) = W(\text{unfolding})$ for protein folding is slightly different from the usual equilibrium condition for

chemical reaction ΔG=0 due to the different bias samplings in frequency space of $\{w_j\}$ and $\{w'_j\}$, $\bar{w} \neq \bar{w}'$.

We have studied the folding of two-state proteins and RNA molecules from the point of quantum transition. Dislike two-state protein folding RNA folding is a multi-stage process. Both the folding from compact intermediate to native fold and the folding of secondary structure formation can be studied from quantum folding theory. However, to study the collapse transition as a whole for RNA molecule needs further development of the quantum model. The problem can be compared with the multi-state protein folding. 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 [21]. The like idea might be introduced in the study of the total folding rate of the RNA molecule.

Conclusion

A formula on protein and RNA folding rate dependent of torsion number（chain length）is deduced from quantum folding theory of macromolecule. The theoretical prediction is in accordance with the experimental data on two-state-protein and RNA folding. A law on the temperature dependence of folding rate is also deduced which can explain the non-Arrhenius peculiarity of protein folding. What encourages us is: the partial success of the present study on protein and RNA folding from a simple unified theory reveals the existence of a common quantum mechanism in the conformational transition of biomolecules.

AUTHOR CONTRIBUTIONS

L.L proposed the theoretical model and deduced the folding formula. J.L. made the statistical analyses on protein and RNA folding.

ACKNOWLEDGMENT

The author would like to thank Drs Zhao Judong and Zhang Ying for numerous discussions on protein data analyses，and Dr Bao Yulai for his help in literature searching.

REFERENCES

1 Garbuzynskiy S. O., Ivankov D. N.,Bogatyreva N.S., and Finkelstein A.V. 2013. Golden triangle for folding rates of globular proteins. *Proc. Natl. Acad. Sci. USA*. 110: 147-150.

2 Hyeon C. and Thirumarai D. 2012. Chain length determines the folding rates of RNA. *Biophys. J*. 102:L11-L13.

3 Thirumarai D. and Hyeon C. 2005. RNA and protein folding: Common themes and variations. *Biochemistry*. 44:4057-4969.

4 Luo L. F. 2014. Quantum theory on protein folding. *Sci. China – Phys. Mech. Astron*. 57:458-468.

5 Lv J. and Luo L.F. 2014. Statistical analyses of protein folding rates from the view of quantum transition. *Sci. China Life Sci*. 57:1197-1212.

6 Bhuyan, M. S. I. and Gao X.. 2011. A protein-dependent side-chain rotamer library. *BMC*

Bioinformatics. 12: S10.

7 Woodson S. A. 2010. Compact intermediates in RNA folding. *Annu. Rev. Biophys*. 39:61-77.

8 Proctor D.J., Ma H. , Kierzek E., Kierzek R., Gruebele M. , and Bevilacqua P.C. 2004. Folding thermodynamics and kinetics of YNMG RNA hairpins. *Biochemistry*.43:14004-14014.

9 Tan E., Wilson T.J.,…, and Ha T. 2003. A four-way junction accelerats hairpin ribozyme folding via a disrecte intermediate. *Proc Natl Acad Sci USA*. 100**:** 9308-9313.

10 Sosnick T.R. and Pan T. 2004. Reduced contact order and RNA folding rates. *J Mol. Biol*. 342:1359-1365.

11 Deras M.L., Brenowits M.,…and Woodson S.A. 2000. Folding mechanism of the Tetrahymena ribozyme P4-P6 domain. *Biochemistry* 39:10975-10985.

12 Rangan P., Masquida B.,…,and Woodson S.A.. 2003. Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme. *Proc Natl Acad Sci USA*. 100**:** 1574-1579.

13 Fan X., Pan T. and Sosnick T.R.. 1999. Mg-dependent folding of a large ribozyme without kinetic traps. *Nat. Struct. Biol*. 12:1091-1095..

14 Xiao M., Leibowitz M.J. and Zhang Y. 2003. Concerted folding of a Candida ribozyme into the catalytically active structure posteror to a rapid RNA compaction. *Nucl Acids Res*. 31:3901-3908.

15 Zarrinkar P.P., Wang J. and Williamson, J.R.. 1996. Slow folding kinetics of RNase P RNA. *RNA* 2:564-573.

16 Sclavi B., Suilivan M.,…and Woodson S.A.1998. RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting. *Science* 279:1940-1943.

17 Sattin B.D., Zhao W., Travers K., Chu S and Herschlag D. Direct measurement of tertiary contact cooperativity in RNA folding. *J. Am. Chem. Soc*. 130:6085-87.

18 Lambert D., Leipply D., Shiman R. and Drapper D.E. 2009. The influence of monovalent cation size on the stability of RNA tertiary structure. *J. Mol. Biol*. 390:791-804.

19 Ghosh K., Ozkan B. and Dill K.A. 2007. The ultimate speed limit to protein folding is conformational searching. *J. Am .Chem. Soc*. 129: 11920-11927.

20 Maglott E.J. and Glick G.D. 1997. A new method to monitor the rate of conformational transitions in RNA. *Nucl Acid Res*. 25:3297-3301.

21 Zhang Y. and Luo L.F. 2011. The dynamical contact order: protein folding rate parameters based on quantum conformational transition. *Science China Life Sciences* **54**:386–392.

22 Nguyen H., Jager M., …and Kelly J.W. 2003. Tuning the free-energy landscape of a WW domain by temperature, mutation, and truncation. *Proc. Natl. Acad. Sci. USA*. 100: 3948-2953.