J. Phys. Chem. B **2003,** *107,* 12009−12012

Free Volume Concept in Application to Folding Kinetics of Random Heteropolymers

Yaroslav E. Ryabov[†]

Department of Applied Physics, The Hebrew University of Jerusalem, Givat Ram, 91904, Jerusalem, Israel

Received: May 5, 2003; In Final Form: August 28, 2003

The presented work introduces a new model of folding kinetics of random heteropolymers, which is based on the free volume concept. This model reproduces the well-known saddlelike temperature dependence of folding time. The presented model is discussed in comparison to the random energy model which is widely employed for describing kinetics of this type and in relation to Levinthal's paradox.

I. Introduction

Folding is a process of considerable interest which is important for many scientific and practical reasons.[1] This process is intensively discussed in scientific literature in relation to experimental[2−5] investigations and theoretical modeling.[6−13] In this regard, perhaps the most important quantity which describes the folding process is the reaction rate k_f or the corresponding folding time $\tau_f = 1/k_f$. At lower temperatures, similar to the vast majority of chemical reactions, the temperature dependence of folding reaction rate k_f obeys[5,12,13] Arrhenius law which for the folding time reads as

$$\ln\left(\frac{\tau_f}{\tau_0}\right) = \frac{E_a}{T} \tag{1}$$

where E_a is the activation energy of the process and τ_0 is the fastest relaxation time corresponding to the limit $T \to \infty$ (it is assumed here and below that Boltzmann constant $k_B = 1$ and is dimensionless). Arrhenius law determines a decrease of τ_f with temperature increase. However, in the case of folding, further growth of T leads to a deviation from the Arrhenius law, and at a certain T_{opt} folding time, τ_f reaches a minimal value and its temperature dependence changes to the opposite tendency[1−10,12,13] (see also Figure 1). It is widely held now that this behavior is well-explained in the framework of the so-called random energy model (REM). This model was initially introduced by Derrida[14,15] to describe statistical properties of spin glasses. The essential idea of this model is the assumption that the density of energy states $n(E)$ is given by the Gaussian distribution

$$n(E) = \frac{\Omega}{\sqrt{2\pi\Sigma^2}}\exp\left\{-\frac{(E-\bar{E})^2}{2\Sigma^2}\right\} \tag{2}$$

where Ω is the normalization factor, $\bar{E}$ is a mean value and Σ^2 is the standard variance of the energy for the REM Gaussian energy distribution. Bryngelson and Wolyness in their pioneering work[6] applied REM to the folding kinetics and later with coauthors[7] derived the temperature dependence of $\ln(\tau_f)$ in the form of a second order polynomial with respect to the reciprocal temperature, widely to referred as "parabolic" dependence

$$\ln\left(\frac{\tau_f}{\tau_0}\right) = \frac{\bar{E} - E_0}{T} + \frac{\Sigma^2}{2T^2} \tag{3}$$

where E_0 is the energy of the local minimum state. The key result of the original REM is that at a certain temperature $T_c = \Sigma/\sqrt{2\ln\Omega}$ the entropy of the system vanishes. Thus, below this temperature the system is frozen. This REM transition is widely discussed in relation to the so-called heteropolymer freezing, which is the transition from the phase where many conformations dominate at equilibrium to one in which only a few conformations are statistically relevant. The REM ideology provided substantial progress in understanding of folding kinetics. However, there are many observations that are not easily understood in the original REM paradigm.[11,16] For example, it is clear that the "parabolic" dependence in (3) does not exhibit the Arrhenius limit in (1) at lower temperatures. To resolve this contradiction, one can note[6,7,8,13] that below T_c the "parabolic" formula is not valid due to the REM freezing transition. Thus, one obtains above T_c a "parabolic" like temperature dependence of τ_f while below T_c, using the mean field replica theory,[8] one derives an Arrehenius behavior of folding time as follows

$$\ln\left(\frac{\tau_f}{\tau_0}\right) = \frac{E^*}{T}\left(\frac{1}{T} - \frac{1}{T_c}\right)^2 \quad \text{for } T > T_c \tag{4a}$$

$$\ln\left(\frac{\tau_f}{\tau_0}\right) = \frac{E^*}{T} \quad \text{for } T > T_c \tag{4b}$$

where $E_c = \bar{E} - \Sigma\sqrt{2\ln\Omega}$ is the energy correspondent to the REM "freezing" transition at T_c. Nevertheless, the basic REM assumption—statistical independence of the energies of states over disorder—seems violated for many cases, for example for the proteins, and therefore, REM applicability is questionable.[11] Nevertheless the generalizations of REM are still actively discussed[17] in relation to the folding kinetics and to the interpretation of computer experiments which are simulating folding kinetics of random amino acid sequences.[13]

The nonmonotonic temperature dependency of folding time is also discussed from the point of view of a modified transition state theory. In this case, as it shown by Oliveberg, Tan, and Fersht,[3,18] the changes in heat capacity for a biopolymer with a hydrophobic core may also lead to non-Arrhenius folding

---

[†] Also at Institute for Mechanics and Engineering, Kazan, 420111, Russia. E-mail: Ryabov@vms.huji.ac.il.

10.1021/jp035216c CCC: $25.00

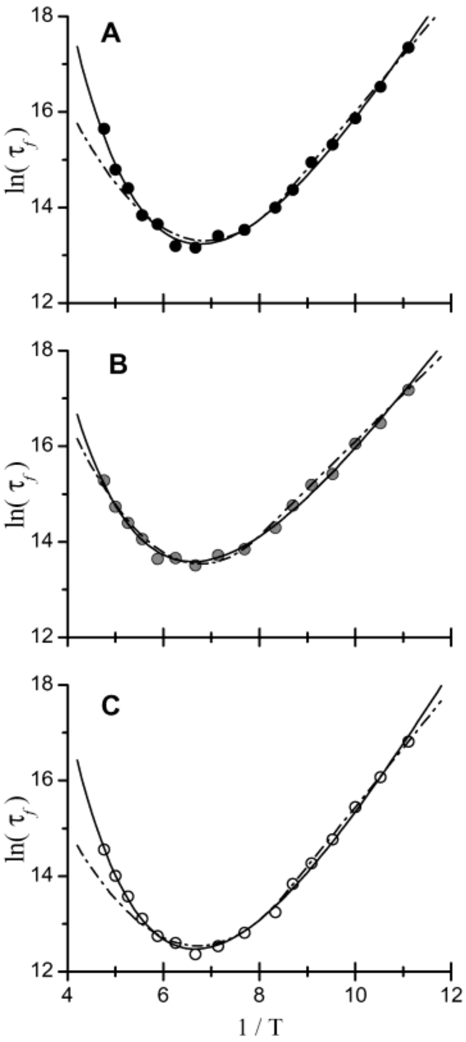

Figure 1. Arrhenius plot presenting the dependencies of ln(τ_f) vs reciprocal temperature. The symbols represent original data from ref 13 (panel A for sequence 1, panel B for sequence 4, and panel C for sequence 8). The dash-dotted lines correspond to model 4. The full lines correspond to model 12. The parameters of both models are listed in Table 1.

kinetics. The reason for this is apparent negative activation enthalpy occurring in the case when the heat capacities of the transition state and denatured state are different.[3,18]

The presented work is an attempt to describe folding kinetics of heteropolymers from an alternative point of view which is based on the free volume concept. This concept was elaborated by Fox and Flory to describe dynamic transition in polystyrene.[19−21] The main idea of this concept is that the probability to move a polymer molecule segment is related to the free volume available in a system. Later the concept of free volume was applied to the wider class of disordered solids by Doolittle[22] and Turnbull and Cohen[23] who suggested a similar relationship between the viscosity η, the self-diffusion coefficient D, and the free volume of an amorphous solid in the form

$$\ln \eta \sim - \ln D \sim - \frac{v_0}{v_f} \tag{5}$$

where v_0 is the volume of a molecule (a mobile unit) and v_f is the free volume per molecule (per mobile unit).

II. The Model

The folded conformation of a macromolecule is a state where all the beads in the chain are connected to the proper, so-called "native", neighbors to minimize the total energy of the conformation. Let us call p_b the probability to break a bond between the two neighboring beads and p_n the probability to find a native neighbor for a certain bond. Thus, if the chain contains N beads and if each bead in the chain has q bonds then the total probability p_f to convert this chain in the native state is

$$p_f = (p_b p_n^{N})^q \tag{6}$$

where p_b^q is the probability to let all q bonds for every bead be able to search for the native connection and p_n^{qN} is the probability to find the native contacts for all qN bonds. Let us assume that to break one bond between two beads one need to spend the energy E_b. This means that the probability p_b obeys the law

$$p_b = \exp\left(-\frac{E_b}{T}\right) \tag{7}$$

Next, let us discuss the probability p_n. Let us suppose that this probability obeys the free volume concept. However, in contrast to the original free volume concept eq 5, let us assume that this probability is related not to the actual volume occupied by a chain but to the relevant configuration volume in the space of chain conformations. Thus

$$p_n = \exp\left(-\frac{\omega_0}{\omega_f}\right) \tag{8}$$

where ω_0 is the number of chain conformations per bond and ω_f is the corresponding free volume in the configuration space. This relationship simply express the idea that the probability to find the "native" contact is bigger for greater values of ω_f. Obviously one should regard ω_0 as a constant which is equal to $\omega_0 = \Omega_0/(qN)$, qN is total number of bonds to arrange and $\Omega_0 = q^N$ is the total size of the configuration volume. If a bond is connected to a neighbor, not necessary "native", then it is excluded from the folding kinetics, and consequently, it does not occupy any space in the configurational volume. Thus, the free configuration volume per bond can be evaluated as $\omega_f = \Omega_0/N_d$, where N_d is the number of disconnected bonds which are searching for the "native" connection. To calculate N_d, let us recall eq 7, which states that every bond has the energy E_b. Thus, the number of disconnected bond obeys the Boltzmann law

$$N_d = qN \exp\left(-\frac{E_b}{T}\right) \tag{9}$$

and

$$\omega_f = \omega_0 \exp\left(\frac{E_b}{T}\right) \tag{10}$$

Combining together eqs 6, 7, 8, and 10, one finds that the total probability to convert a whole chain in the native state is

$$p_f = \exp\left(-\frac{qE_b}{T} - qN \exp\left\{-\frac{E_b}{T}\right\}\right) \tag{11}$$

It is important to note here that besides assumptions 7 and 8 used to derive eq 11, it was implicitly assumed that all the constituents of the macromolecular chain are equal; i.e., all of them have the same number of bonds and all the bonds have the same bonding energy. However, this cannot be the case for

TABLE 1: Parameters of Models 4 and 12 Together with the Reduced χ^2 Values

	model presented by eq 4					model presented by eq 12			
seq	$\ln \tau_0$	E_c	E^*	$1/T_c$	χ^2	$\ln \tau_0$	E_b	q	χ^2
1	4.20	−5.92	1.19	8.60	0.077	−1.18	0.49	3.36	0.010
4	6.10	−6.32	0.99	8.04	0.016	1.84	0.50	2.73	0.008
8	3.04[a]	−5.68	1.24	8.60	0.083	−2.14	0.49	3.40	0.006

[a] The original paper[13] reports a value of $\ln \tau_0 = 3.4$ for sequence 8. However this value is not absolutely fitted to the data. Most probably this value was misprinted. Substitution of $\ln \tau_0 = 3.04$ gives reasonable agreement with the data as presented in Figure 1.

real heteropolymers. Thus, let us take this fact into account and redefine that E_b and q are the averaged energy of a bond and the averaged number of bonds per bead, which is not necessary an integer. Thus, recalling the relationship $\tau_f \sim 1/p_f$ one can get the final expression

$$\ln\left(\frac{\tau_f}{\tau_0}\right) = \frac{qE_b}{T} + qN \exp\left\{-\frac{E_b}{T}\right\} \tag{12}$$

III. Comparison with Simulation Data

In the presented paper we are going to utilize the simulation data borrowed from the previous paper.[13] In this paper[13] 10 random amino acid sequences were investigated, however, only three of them were analyzed in the frameworks of eq 4. These are sequences 1, 4, and 8 (see Figure 1 and Figure 9 in ref 13). In this case for model 4 the original fitting parameters from previous work[13] (see Figure 9 in ref 13) were used, while for model 12 the fitting procedure was performed (see Table 1).

To establish a quantitative criteria of fitting goodness, one can use[24] the reduced χ^2 value

$$\chi^2 = \frac{1}{m - l}\sum_{i=1}^{m}(y_i - f_i)^2 \tag{13}$$

where m is the number of points to fit, l is number of adjustable parameters in a fitting model, y_i are values of a given function and f_i are correspondent values of the fitting function. For all analyzed sequences, $m = 16$ (see ref 13). Model 4 has four adjustable parameters: $\ln \tau_0$, E_c, E^* and $1/T_c$. Since the number of residues is fixed $N = 27$ (see ref 13), model 12 has only three adjustable parameters: $\ln \tau_0$, E_b, and q.

IV. Results and Discussion

As one can see from Figure 1, model 4 reasonably fits the data at the low-temperature region where the Arrhenius term works, however, at the high-temperature region where the "parabolic" REM term was applied, this model exhibits significant deviation from the simulation data. In contrast, model 12 fits the data over the whole temperature range. From Table 1 one can see that values of χ^2 for model 12 are about 1 order of magnitude less than those for model 4. This clearly indicates that model 12 is more relevant to the data presented in ref 13 than the model 4. Regarding model 12, note that the first time it was introduced to describe the dielectric relaxation of water molecules in a confined geometry[25] and later was proved for other systems.[26,27] This affirms the general importance of model 12 and shows that free volume arguments may be relevant not only for the actual volume of a system but for the configurational space as well.

From Table 1, one can observe that values of E_b for model 12 for all the analyzed random amino acid sequences are almost the same. This means that for all sequences one needs to spend roughly the same energy to break one bond between the beads in the macromolecule chain. Thus, the only parameter which affects the shape of the folding time temperature dependency is the number of bonds per bead q. Using this parameter one can estimate the number of contacts C between the beads in the native conformation. Actually, in work[13] the folding dynamic was simulated with a three-dimensional lattice model. Thus, the single bead can establish a maximum of four contacts to the neighbors (six bonds per particle where two of them are permanent connections in a macromolecule chain). Thus, $C = qN/4$ which for $N = 27$ and q from Table 1 gives $C = 22.7$ for sequence 1, $C = 18.4$ for sequence 4, and $C = 22.9$ for sequence 8. These estimated values are in fair agreement with the C values obtained directly from simulations $C = 22$, $C = 18$ and $C = 21$ for sequences 1, 4, and 8, respectively (see Table 1 in ref 13).

From eq 12 one can easily calculate the minimum in the folding time temperature dependency and find the so-called T_{opt} corresponding to the fastest folding time in the form

$$T_{opt} = \frac{E_b}{\ln N} \tag{14}$$

which gives $T_{opt} \approx 0.15$ for all analyzed sequences. Thus, with eq 14, one can expect a logarithmic dependency of inverse T_{opt} on the sequence length.

Substitution of eq 14 into eq 12 gives the fastest folding time

$$\tau_{fast} = \tau_0(eN)^q \tag{15}$$

where e ≅ 2.718 is the base of the natural logarithm. Equation 15 represents a so-called algebraic scaling for folding times discussed earlier.[12,28−33] It can be derived from the idea of Thirumalai[28−30] that τ_{fast} should have an approximate power-law dependence on N as in the case of the time scale for subdivision in subspace of compact structures. Similar dependencies have also been discussed based on phenomenological analysis of simulation results and experiential data.[12,31−33]

It is interesting to discuss eq 15 in relationship to the so-called Levinthal paradox. This paradox was raised for the first time by Cyrus Levinthal who mentioned[34] the contradiction between the astronomically large number of different possible conformations Ω_0 and the quite fast folding time of real macromolecules and biopolymers. Actually, if one implies that a macromolecule reaches the native conformations by random search over the entire configuration space $\Omega_0 = q^N$ and tests one conformation for time τ_0, then one immediately obtains that folding the time exponentially grows with length of a macromolecule $\tau_f = \tau_0 q^N$. There are many approaches to resolve this contradiction including the idea of so-called "funnels" of the energy landscape,[1] a quite elegant consideration that the processes uses to convert a correct bond into an incorrect one and to convert an incorrect bond into a correct one may have different probabilities,[35] the idea of nucleation mechanism leading to the so-called exponential scaling of folding times,[36,37] the considerations about cooperative (all-or-none) character of folding,[38] and even the claims that it has never been a paradox.[39,40]

In this regard, eq 15 suggests another point of view on the Levinthal's paradox based on the free volume concept and shows that the fastest folding time may have a power law dependency on N, which is much slower than the exponential growth implemented by the random search. The key heuristic idea here is since the total configuration space is constant the temperature increase leads to a decrease of relevant free configuration volume as presented by eq 10.

Acknowledgment. The author appreciates the extremely useful notes of Dr. A. Puzenko and Prof. Yu. Feldman, expresses his deep gratitude to The Hebrew University of Jerusalem for its hospitality, and notes candid discussion with Prof. E. Shaknovich.

References and Notes

(1) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins* **1995**, *21,* 167.

(2) Anfinsen, C. B.; Haber, E.; Sela, M.; White, F. H. *J. Proc. Natl. Acad. Sci. U.S.A.* **1961**, *47*, 1309.

(3) Tan, Y.-J.; Oliveberg, M.; Fersht, A. R. **1996**, *J. Mol. Biol. 264,* 377.

(4) Schindler, T.; Schmid, F. X. *Biochemistry* **1996**, *35*, 16833.

(5) Scalley, M. L.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94,* 10636.

(6) Bryngelson, J. D.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7524.

(7) Bryngelson, J. D.; Wolynes, P. G. *J. Phys. Chem.* **1989**, *93*, 6902.

(8) Shakhnovich, E. I.; Gutin, A. M. *Europhys. Lett.* **1989**, *9*, 569.

(9) Miller, R.; Danko, C. A.; Fasolka, M. J.; Balazs, A. C.; Chan, H. S.; Dill, K. A. *J. Chem. Phys.* **1992**, *96,* 768.

(10) Socci, N. D.; Onuchic, J. N. *J. Chem. Phys.* **1994**, *101,* 1519.

(11) Pande, V. S.; Grosberg, A. Yu.; Joerg, C.; Tanaka, T. *Phys. Rev. Lett.* **1996**, *76*, 3987.

(12) Gutin, A. M.; Abkevich, V. I.; Shaknovich, E. I. *Phys. Rev. Lett.* **1996**, *77,* 5433.

(13) Gutin, A.; Sali, A.; Abkevich, V.; Karplus, M.; Shaknovich, E. I. *J. Chem. Phys.* **1998**, *108,* 6466.

(14) Derrida, B. *Phys. Rev. Lett.* **1980**, *45,* 79.

(15) Derrida, B. *Phys. Rev. B* **1981**, *24,* 2613.

(16) Yue, K.; Fiebig, K. M.; Thomas, P. D.; Chan, H. S.; Dill, K. A.; I Shaknovich, E. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92,* 325.

(17) Papoian, G. A.; Wolynes, P. G. *Biopolymers* **2003**, *68,* 333.

(18) Oliveberg, M.; Tan, Y.-J.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92,* 8926.

(19) Fox, T. G.; Flory, P. J. *J. App. Phys.* **1950**, *21,* 581.

(20) Fox, T. G.; Flory, P. J. *J. Phys. Chem.* **1951**, *55,* 221.

(21) Fox, T. G.; Flory, P. J. *J. Polym. Sci.* **1954**, *14,* 315.

(22) Doolittle, A. K. *J. Appl. Phys.* **1951**, *22,* 1471.

(23) Turnbull, D.; Cohen, M. H. *J. Chem. Phys.* **1960**, *34,* 120.

(24) Korn, G. A.; Korn, T. M. *Mathematical Handbook for Scientists and Engineers*; MacGraw-Hill book company Inc.: New York, 1961; pp 617−618.

(25) Ryabov, Y.; Gutina, A.; Arkhipov, V.; Feldman, Yu. *J. Phys. Chem. B* **2001**, 105, 1845.

(26) Frunza, L.; Kosslik, H.; Frunza, S.; Fricke, R.; Schönhals, A. *J. Non-Cryst. Solids* **2002**, *307−310*, 503.

(27) Frunza, L.; Kosslik, H.; Frunza, S.; Fricke, R.; Schönhals, A. *J. Phys. Chem. B* **2002**, *106,* 9191.

(28) Camacho, C. J.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90,* 6369.

(29) Thirumalai, D. *J. Phys. I (Fr.)* **1995**, *5,* 1457.

(30) Li, M. S.; Klimov, D. K.; Thirumalai, D. *J. Phys. Chem. B* **2002**, *106*, 8302.

(31) Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94,* 6170.

(32) Plaxaco, K. W.; Simons, K. T.; Ruczinski, I.; Baker, D. *Biochemistry* **2000**, *39*, 11178.

(33) Koga, N.; Takada, S. *J. Mol. Biol.* **2001**, *313*, 171.

(34) Levinthal, C. In *Mossbauer spectroscopy in biological systems, Proceedings of a meeting held at Allerton house, Monticello, IL*; DeBrunner, J., Tsibris, J. C. M.; Münck, E., Eds.; University of Illinois Press: Urbana, IL, 1969; pp 22−24.

(35) Zwanzig, R.; Szabo, A.; Bagghi, B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89,* 20.

(36) Finkelstein, A. V.; Badretdinov, A. Y. *Fold. Des.* **1997**, *2,* 115.

(37) Finkelstein, A. V. *J. Biomol. Sruct. Dyn.* **2002**, *20,* 311.

(38) Grosberg, A. *J. Biomol. Sruct. Dyn.* **2002**, *20,* 317.

(39) Rooman, M.; Dehouck, Y.; Kwasigroch, J. M.; Biot, C.; Gilis, D. *J. Biomol. Sruct. Dyn.* **2002**, *20,* 327.

(40) Fernandez, A.; Belinky, A.; Boland, M. D. J. *Biomol. Sruct. Dyn.* **2002**, *20,* 331.