RAPID COMMUNICATIONS

PHYSICAL REVIEW A VOLUME 42, NUMBER 6 15 SEPTEMBER 1990

Activation-energy landscape for metastable RNA folding

Ariel Fernández
Department of Chemistry, University of Miami, Coral Gables, Florida 33124
and Department of Biochemistry and Molecular Biology, Medical School, Miami, Florida 33101

Eugene I. Shakhnovich
Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138
(Received 15 June 1990)

We consider the relaxation kinetics for folding intermediate structures in random uncorrelated RNA chains. Folding patterns correspond to intrachain secondary structures. We make use of a Monte Carlo simulation that mimics a Markov process to study the refolding dynamics. This is done to obtain the time-dependent behavior of the activation energies. We conclude that the mechanism of relaxation to the equilibrium secondary structure is compatible with the random-energy model as the thermodynamic limit is approached.

The glassy nature of folded structures in biopolymers has been emphasized from different perspectives.[1-3] However, an essential ingredient in such views is the quenched-disorder nature of the interactions[1] or of the primary structure itself.[2,3] These assumptions make the system tractable with tools such as the replica trick, borrowed from the theory of disordered condensed matter. The two important sources of quenched disorder are (a) the random distribution of "native" and "non-native"[1] interactions in polypeptide chains, or (b) the random uncorrelated primary RNA sequence.[2,3] The biological relevance of such systems is currently subject to intense scrutiny. Partial theoretical findings for disordered RNA chains[3] lead us to the belief that glassy behavior, with a highly degenerate ground state, holds as the thermodynamic limit is approached. The methods and results on relaxation kinetics for metastable RNA folding have been implemented assuming an underlying random uncorrelated primary sequence. The biological significance of such an assumption is apparent for biopolymers in general.[1,3,4,5] Recent results by one of us[4] and by Ptitsyn and Volkenstein[5] suggest that naturally occurring biopolymers might be more "disordered" than one might expect, that is, the dominance of specific folded forms which are biologically active appears to hold within vast domains in sequence space.

The complexity of the RNA backbone allows for the formation of intrachain secondary structures by means of Watson-Crick interactions. Most structures relevant to biological processes such as RNA replication and transcription are metastable or transient.[6] They are relevant precisely because their lifetimes are sufficiently long when compared with the time scales for the enzymatic reactions involved in their own assembling process. At the same time, their degree of folding is not maximal, a property that often renders the structure biologically inert.[7] An understanding of the relaxation dynamics for intermediate folded structures might prove essential to assess the molecular basis of regulation and control in RNA replication and transcription.

In this work, we shall make use of a Monte Carlo simulation that mimics the kinetically governed formation and dismantling of RNA transient structures.[7] The temperature of interest belongs to a neighborhood of the freezing transition temperature within a "static" random-energy model (REM).[8] However, only temperatures above that leading to the zero entropy phase are considered, since we are focusing on realistic biological conditions (cf. Ref. 1). The applicability of the REM will be substantiated in this paper. The aim of the work is to furnish a thorough characterization of the activation energy landscape, or, equivalently, of the spectrum of relaxation timescales for intrachain folding in disordered RNA. Our system constitutes a realization of Sompolinsky's early scenario concerning the hierarchy of relaxation time scales.[9]

For the sake of completeness, we shall describe the Monte Carlo simulation used to obtain the time-dependent probabilities for the transient secondary structures.[7] In previous simulations a definite RNA primary sequence was considered and the refolding events were concomitant with chain growth due to sequential incorporation of nucleotides. The situation we shall concentrate on in this work differs in that the primary sequence is randomly generated, the limit of relatively long chains is explored and the length of the chain is fixed throughout the simulation. The simulation mimics a Markov process in which, as new possibilities for folding arise, previously existing metastable structures are dismantled to allow for the formation of the emerging ones. The Markov process is comprised of two different kinds of elementary events: (a) intrachain partial helix formation and (b) intrachain helix decay. In addition, we have incorporated certain features absent in previous work: the possibility of guanine-thymine (G-T) and adenine-cytosine (A-C) mispairs and the possibility of looped-out bases in the process of helix formation. The transition time for each of the events in the Markov process is a Poissonian random variable. If an admissible helix formation happens to be the event favored, the inverse of the mean time for the transition will be given by

$$t^{-1} = fn \exp(-\Delta G_{\text{loop}}/RT), \tag{1}$$

where f is the kinetic constant for a single base-pair for-

mation [estimated at $10^8\ s^{-1}$, cf. (Ref. 7)], n is the number of base pairs comprising the helix and ΔG_{loop} is the change in free energy of the set of all loops due to the folding which leads to the new intrachain stem formation. The formation of new helices should always be topologically compatible with the pattern of existing ones in the sense that no knots can be allowed. This condition has been given proper combinatorial form, and as such is incorporated in the algorithm in a standard manner.

If the chosen elementary event happens to be the intrachain helix decay, the inverse mean time is

$$t^{-1} = fn \exp(G_h/RT) , \tag{2}$$

where G_h is the free energy of the helix.

The entropic contribution of the intrachain loops and the free-energy terms for partial helices are taken from the Turner parametrization[10] at 37 °C. In addition to the parametrization indicated, we shall impose a realistic cutoff value in the simulation: the minimum admissible time span of an intrachain helix is taken to be 5×10^{-1} s. The cutoff adopted is not arbitrary but corresponds to the minimum lifetime for the most fragile helix which can be formed involving a G-C pair.

The Markov process is simulated by selecting one of the two possible elementary events at each stage. The effective transition time for the chosen elementary event is a Poissonian random variable with mean k^{-1} where the effective rate constant k is given by

$$k = \sum_{i=1}^{F} k_1(i) + \sum_{j=1}^{D} k_2(j) . \tag{3}$$

The subindices 1,2 correspond to events (a) and (b), respectively. The indices $i = 1, \ldots, F$ label helices that can be formed so that they are topologically compatible with the pattern of existing ones. The latter ones are labeled by the dummy index $j = 1, \ldots, D$. In order to implement the simulation, we shall relabel the rate constants as follows:

$$k = \sum_{m=1}^{M} k'_m, \quad M = F + D ,$$

$$k'_1 = k_1(1), \ldots, k'_F = k_1(F) , \tag{4}$$

$$k'_{F+1} = k_2(1), \ldots, \quad k'_{F+D} = k_2(D) .$$

This is done in order to find the transition index m at each stage of the process. Thus, we consider a uniformly distributed random variable R, $0 \le R \le k$, so that if the value r of R lies in the interval

$$\sum_{m=1}^{m'-1} k'_m \le r \le \sum_{m=1}^{m'} k'_m , \tag{5}$$

then, the index m' has been chosen. The process has been repeated 10^5 times for a random uncorrelated chain of length $N = 512$, the maximal length considered. The actual time span for the simulation is 14 min Cray-1S time.

The time-dependent probability $U_n = U_n(t)$ for the most probable secondary structure, n, at time t is readily accessible from our simulations. This is particularly crucial since the kinetic barriers for interconversion between metastable secondary structures become also accessible: The activation energy for the transition between two structures is given by

$$E_a(n \to n+1) \propto |U'_n(t^*) - U'_{n+1}(t^*)|^{-0.33} , \tag{6}$$

where the prime denotes time derivative and t^* is the actual instant when one structure is superseded by another occurring with a subsequently higher probability. This fit was found empirically and finds justification in the statistical-mechanical treatment presented in this work. The qualitative feature one notices first is that the transitions are initially fast and they slow down as the chain grows.

In order to prove the existence of an "asymptotically glassy" behavior, that is, in order to show that the kinetic barrier between conformational substates presents a spin-glass-like behavior in the thermodynamic limit, we have examined the behavior of the logarithm of the "nonergodic" transition time, T_{nonerg}, for elementary events which are either of type (a) or (b), as a function of $N^{1/4}$. A spin-glass-like behavior[11] is observed. This is revealed by the linear plot in Fig. 1, where $M = N^{1/4}$, and the preexponential factor is $A = 1.12$ s. The *minimum* transition time for each chain length, T_{trans}, is the one considered in the ordinates of Fig. 1. Its logarithm scales with the length of the chain according to a power law equivalent to that found in spin glasses for the nonergodic relaxation times (N being in this case the size of the system). Thus, for N fixed, the variable T_{trans} will be denoted T_{nonerg}.

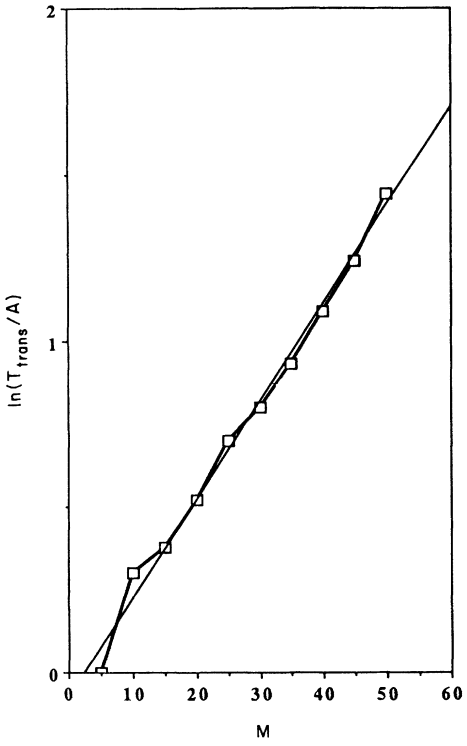

FIG. 1. The length dependence of the transition time, T_{trans} for intrachain helix formation or decay. The variable in the abscissas is $M = N^{1/4}$, where N is the length of the chain given in number of nucleotides. The preexponential factor A is fixed at 1.12 s. The straight line is an aid to the eye.

The scaling law for the lowest possible (nonergodic) activation energy is of the form familiar from spin-glass theory: $E_a \propto k_B T N^{1/4}$.

In principle, the relaxation of metastable secondary structures is completely characterized only if the activation energy landscape for transitions is fully described. Thus, the kinetic barrier encountered at any given instant should be calculated. However, only transitions which occur within a certain vicinity of the nonergodic time scale T_{nonerg} are accessible computationally. That vicinity corresponds to the range for the abscissas given in Fig. 2. Thus, transitions involving vast changes in secondary structure, with associated time scales of the order of ergodic times [$\approx \exp(N^{1/2})$] are not hitherto accessible. The time dependence of the (encountered) activation energies for refolding events is displayed in Fig. 2. The observed dependence of $\ln(T_{\text{relax}}/A)$ vs $\ln(t/T_{\text{nonerg}})$ may give some insight into the actual construction of the energy landscape of an RNA molecule. The most striking feature is that the plot in Fig. 2 presented in log-log scale is linear. This gives evidence that activation barriers grow in the course of folding. This may be explained by assuming that the molecule (or part of it) must first unfold in the course of the rearrangement in order to refold to a state with lower energy. Regarding the unfolded state as a "transition state" enables us to rationalize Fig. 2: The folding of a structure which is more favorable energetically (but still intermediate) would require overcoming a larger energy barrier. The detailed analysis of this mechanism of relaxation to equilibrium had been done.[8] This activation energy landscape in the range of time scales considered can be most adequately described in terms of the REM. Moreover, it had been shown in Ref. 8 that the energy of a molecule undergoing relaxation to equilibrium follows the logarithmic law

$$E(t) = -k_B T \ln(t/T_{\text{nonerg}}) . \tag{7}$$

This means that under the assumption that the transition state has the same energy $E^{\neq}$ for transitions between any two states, the barriers for relaxation $E^{\neq} - E$ will also follow the logarithmic law as is indeed observed in our numerical study. Thus, the results displayed in Fig. 2 reveal the validity of REM to treat the relaxation kinetics of metastable RNA structures.

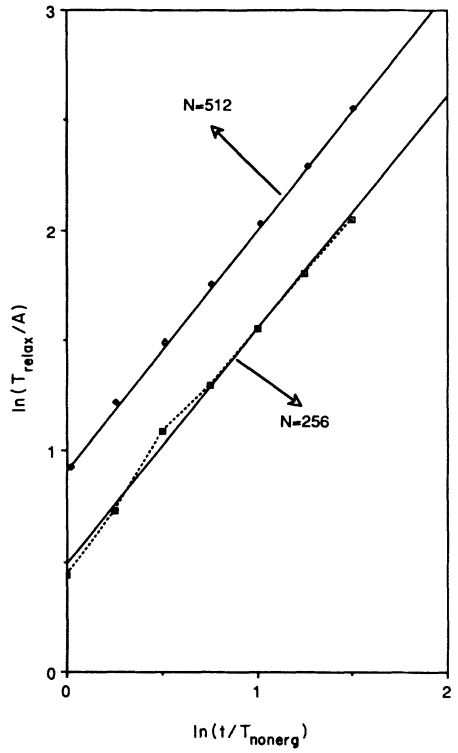

FIG. 2. Logarithmic growth in time of the activation energy E_a for transitions between metastable folded states. The ordinates are equal to E_a/k_BT. The minimum transition time T_{trans} for each given length has been conveniently labeled T_{nonerg}, to emphasize the analogy with spin-glass relaxation.

The work of A.F. was supported in part by the Camille and Henry Dreyfus Foundation. A.F. acknowledges the hospitality of D. Campbell during his visit to Los Alamos National Laboratory, where part of this work was performed.

[1] J. B. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. **84**, 7524 (1987).
[2] A. Fernández, Chem. Phys. Lett. **154**, 396–402 (1989).
[3] A. Fernández, Phys. Rev. Lett. **64**, 2328 (1990).
[4] E. I. Shakhnovich and A. M. Gutin, Nature (London) (to be published).
[5] O. B. Ptitsyn and M. V. Volkenstein, J. Biomol. Struct. Dyn. **4**, 137 (1986).
[6] S. E. La Flamme, F. R. Kramer, and D. R. Mills, Nucleic Acids Res. **13**, 8425 (1985).
[7] A. Fernández, Eur. J. Biochem. **182**, 161 (1989).
[8] E. I. Shakhnovich and A. M. Gutin, Europhys. Lett. **9**, 569 (1989).
[9] H. Sompolinsky, Phys. Rev. Lett. **47**, 935 (1981).
[10] S. M. Freier, R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson, and D. H. Turner, Proc. Natl. Acad. Sci. U.S.A. **83**, 9373 (1986).
[11] N. D. Mackenzie and A. P. Young, Phys. Rev. Lett. **49**, 301 (1982).