

## Variational Approach to Relaxation in Complex Free Energy Landscapes: The Polymer Folding Problem

Ariel Fernández and Gustavo Appignanesi

*Instituto de Matemática–INMABB–Universidad Nacional del Sur, Consejo Nacional de Investigaciones Científicas y Técnicas,  
Avenida Alem 1253, Bahía Blanca 8000, Argentina*

(Received 12 September 1996)

A statistical theory of the folding dynamics of an ensemble of random heteropolymers is developed. The predicted relaxation is consistent with the glassy behavior of quenched disorder and is singled out by a generic least action principle: In processes of increasing difficulty, the logarithmic growth in time of the expected activation barrier yields the fastest relaxation pathway. [S0031-9007(97)02780-4]

PACS numbers: 87.15.He, 87.10.+e, 87.15.Da

The so-called logarithmic relaxation is found in statistical ensembles of equivalent entities with quenched disorder whose dynamics is mapped on a complex free energy landscape [1–3]. By logarithmic relaxation we mean an averaged behavior characterized by a logarithmic growth of the expected encountered activation barrier which at time  $t$  is of the order of  $\ln(t/\tau)$ , with  $\tau$  being a characteristic time scale. This slow growth has been found to be expedient in specific systems [1] which relax with an increasing level of difficulty in a rugged landscape. An illustration of this instance is provided by a collection of the order of Avogadro's number ( $\sim 10^{23}$ ) of polymer molecules of fixed length with a random primary sequence folding intramolecularly under *in vitro* solvent conditions [1–4]. The molecules undergo folding events in an asynchronous manner and only their statistical behavior is amenable to a physical treatment. Their relaxation towards foldings of increasing complexity has been accounted for by implementing a kinetic theory [5] rooted in the random energy model (REM) [6]. Our aim in this work is to cast this behavior in terms of a variational principle. Such a formulation has been shown to be essential to understand the expediency of the folding process, in particular to explain how the process circumvents the scenario of exhaustive random search in conformation space [7].

Focusing on relaxation processes of increasing difficulty, it becomes intuitively obvious that the fastest process would correspond to a logarithmic growth of activation barriers, since such a growth is the slowest possible. Then, the following question arises: Is it possible to derive a least action principle such that the glassy logarithmic relaxation described above is singled out as a brachistochrone or least over-all-time relaxation pathway?

In order to answer this question, we shall first develop a formal scheme with minimal detail that effectively reproduces the statistics of the folding dynamics for an ensemble of random heteropolymers. Then, we shall prove that the expected relaxation leading to increasing structural complexity obeys a variational principle which actually encompasses a broad range of phenomena involving systems with quenched disorder.

The search in conformation space performed by a polymer chain that forms intramolecular interactions under *in vitro* solvent conditions is neither a downhill process nor the result of a random exploration in conformation space [8–12]. To determine the statistics upon which the dynamics are built, we pick the enthalpy  $H$  ( $H \leq 0$ ) of a folded state as the relevant coordinate, assigning  $H = 0$  to the random coil conformation (RC). This choice is appropriate since enthalpy changes result from heat released and transferred to the statistical bath (the solvent) due to intramolecular contact formation and, consequently, the enthalpic content of a specific state depends directly on the contact pattern (CP) to which the state is associated. Thus, our statistical theory holds valid for RNA [1], in which each intrachain secondary contact is saturating, or for frustration-free model biopolymers [1,2,10], in which the enthalpic content can be lowered only by contact formation. Furthermore, the dynamics at the level of transitions between CP's are understood and have been effectively modeled [7,10]. Thus, our theory aims at defining the statistical dynamics along a single coordinate  $H$  as a projection of the dynamics within the CP space for random copolymers in the long chain limit.

The choice of  $H$  as the natural coordinate hinges upon the validity of an adiabatic scenario in which each CP is regarded as a quasiequilibrium state [1–5,7–12]. This means that, in dealing with folding events, the fast dihedral torsions are integrated out as conformational entropy of the chain (cf. [13]). This tenet is in full agreement with most treatments of the folding problem known to the authors [1–5,7–12]. The theory is built upon an  $H$ -level distribution compatible with the coarse CP description of conformation space. Within the CP space, the dynamics are determined following a general scheme [7,10]: The kinetic barrier  $B$  associated to a contact formation  $B = B(\text{loop})$  is entropic in nature since the transition state in the rate-limiting step entails a loop closure with the concurrent loss in conformational freedom [7]. Once the contact regions have been brought to proximity, bond formation becomes a downhill event since the entire cost has been borne up by the entropy:  $B(\text{loop}) \approx -T\Delta S(\text{loop})$ ,

where  $\Delta S(\text{loop})$  is the entropy loss associated to loop closure [7]. The computation of this quantity incorporates an estimate of the conformational constraints determined by polar group orientation towards bulk solvent, as well as excluded volume effects reflecting the self-avoiding nature of the chain [4,7]. On the other hand, the kinetic barrier associated to contact dismantling,  $B = B(\text{del})$ , is of enthalpic origin, since deletion of an intramolecular contact requires heat absorption in the rate-limiting step in the same amount as that released,  $\Delta H$ , upon formation of the contact. Once the intramolecular bonds are broken, the resulting gain in conformational freedom enables further loop-opening steps to occur at no additional cost [10]. Thus, we get  $B(\text{del}) \approx -\Delta H$ .

This kinetic treatment of CP transitions involving Arrhenius-like barriers is consistent with the approximation introduced in that we regard each CP as a quasiequilibrium state, thus validating the thermodynamic estimation of the kinetic barriers. This Arrhenius kinetics ansatz overlooks conformational detail, such as specific geometry, beyond that necessary to specify the thermodynamic quantities indicated above.

We introduce the density of conformations with enthalpy  $H$ :  $\Omega(H)/\Omega = F(H)$ , where  $\Omega(H)$  and  $\Omega$  are, respectively, the number of conformations compatible with enthalpy  $H$  and the total number of conformations. Thus, the entropy  $S = S(H)$  of a state with enthalpy  $H$  is  $S(H) = R \ln F(H)$ . In order to determine  $F(H)$ , we make use of the fact that there must exist a denaturation temperature  $T_c$ . Thus, at  $T = T_c$ ,  $G(H) = \Delta G(H)$  is identically zero irrespective of  $H$ . The quantity  $\Delta G(H)$  is the free energy change associated to the transition from the RC to the folded state with enthalpy  $H$ . Then, the following relations hold:

$$-RT_c \ln F(H) + H = 0; \quad (1)$$

$$F(H) = \exp(-|H|/s); \quad (2)$$

$$\Delta S = (R/s)\Delta H, \quad (3)$$

where  $s = RT_c$ . Since  $RT/s < 1$  for  $T < T_c$  and given the nature of the kinetic barriers involved in formation and dismantling of intramolecular contacts, Eq. (3) implies that the folding is mostly delayed due to dismantling of "misfolded" structure, in accord with current observations [8,9,11,12].

To determine the statistical dynamics, we define the probability  $P(H, t)$  of enthalpy  $H$  at time  $t$ , satisfying  $P(H, t) = N(H, t)/N$ , where an ensemble of copies of the system, assuming each one consists of an individual polymer molecule, and  $N(H, t)$  and  $N$  indicate, respectively, the number of molecules with enthalpy  $H$  at time  $t$  and the total number of molecules in the ensemble. To determine the time evolution of  $P$ , we introduce a Master equation which is valid for  $T \leq T_c$  (cf. [14]) that is, in the temperature regime where levels with  $H < 0$  may be populated. This equation determines the change in

$P(H, t)$ ,  $\partial P(H, t)/\partial t$ , as the cumulative result of elementary transitions to and from levels with enthalpy  $H$ . The integrals in which transitional contributions are added up have the value  $H$  as the upper or lower integration limit depending on whether the transitions involve states below or above  $H$ , respectively. Thus, the equation reads

$$\begin{aligned} \partial P(H, t)/\partial t &= \Omega(H) \int_{-\infty}^H \exp[(H' - H)/RT] P(H', t) dH' \\ &\quad - P(H, t) \int_{-\infty}^H \exp[(H' - H)/s] \Omega(H') dH' \\ &\quad + \Omega(H) \int_H^0 \exp[(H - H')/s] P(H', t) dH' \\ &\quad - P(H, t) \int_H^0 \exp[(H - H')/RT] \Omega(H') dH'. \quad (4) \end{aligned}$$

The two sources of probability represented by the first and third terms in the right-hand side of Eq. (4) give the positive rate of probability change due to elementary transitions  $H' \rightarrow H$ , from levels with enthalpy  $H'$  below and above  $H$ , respectively. The barrier associated to the former transition is  $B = H - H'$  ( $H' \leq H$ ), while the latter transition requires surmounting a barrier of entropic origin,  $B = RT(H' - H)/s$  ( $H' \geq H$ ). The rate contributions follow directly from Eq. (3), the computation of the kinetic barriers for elementary transitions within the space of CP's and the general form of the unimolecular rate constant  $r$  for a folding transition. This rate is computed as  $r = f \exp(-B/RT)$ , where  $f \approx 10^6 \text{ s}^{-1}$  is the rate constant for contact formation once the nucleating event of loop closure has taken place [7], and  $B$  is the kinetic barrier involved in the transition. On the other hand, the two sinks of probability, given by the second and fourth terms, correspond to transitions  $H \rightarrow H'$ . In this situation, whenever  $H' \leq H$  (second term) the barrier is entropic and it becomes enthalpic if  $H' \geq H$  (fourth term).

Equation (4) may be integrated numerically with the appropriate initial condition:  $P(H, 0) = \delta(H - 0)$ . This condition holds because folding is assumed to take place when a renaturation temperature  $T < T_c$  is reestablished, and thus the starting point of the process is the RC with  $H = 0$ . In order to monitor the dynamics, we follow the expected enthalpy  $\langle H(t) \rangle$  at time  $t$ :

$$\langle H(t) \rangle = \int_{-\infty}^0 H P(H, t) dH. \quad (5)$$

The results are displayed in Fig. 1 for specific reduced temperatures  $T'$ . The following notation has been adopted:  $\tau = f^{-1}$ ;  $T' = (T_c - T)/T_c$  ( $T \leq T_c$ ). The critical temperature has been fixed at  $T_c = 318 \text{ K}$ . The logarithmic time-dependent behavior of  $\langle H(t) \rangle$  fits into the physical picture of general relaxation dynamics for glassy disordered materials [1,6], thus corroborating the

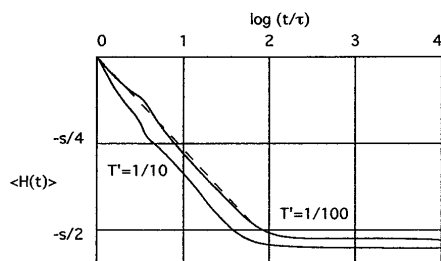


FIG. 1. Time-dependent behavior of the expected enthalpy  $\langle H(t) \rangle$ , as obtained by numerical integration of Eq. (4). The dashed line indicates ideal logarithmic relaxation. The abscissas are dimensionless and given in logarithmic form with scaling constant  $\tau = 1 \mu\text{s}$ . The ordinates are given in units of  $s = RT_c$ . Two plots are given corresponding to two choices of the reduced temperature:  $T' = 1/10$  and  $T' = 1/100$ .

validity of the approach presented in this work. This relaxation regime is invariably followed by a sudden asymptotic relaxation to a saturation enthalpy value  $H = H(\infty)$ . Within this new regime,  $\langle H(t) \rangle$  remains almost constant, satisfying  $|\langle H(t) \rangle - H(\infty)| \leq 10^{-5} H(\infty)$ .

The saturation enthalpy may be easily determined: For a given enthalpy  $H$  we may determine the ratio  $y(H) = r \downarrow(H)/r \uparrow(H)$ , where  $r \downarrow(H)$  is the rate of downwards transition in the enthalpy spectrum with starting point  $H$ , and  $r \uparrow(H)$  is the rate of upwards transition:

$$y(H) = \frac{\int_{-\infty}^H \exp[(H' - H)/s] \Omega(H') dH'}{\int_H^0 \exp[(H - H')/RT] \Omega(H') dH'}. \quad (6)$$

In general  $y(H) > 1$  if  $H > H(\infty)$ , that is, starting at the RC ( $H = 0$ ), there is on the average a tendency to increase the folding complexity by forming contacts until the saturation enthalpy  $H = H(\infty)$  is reached. The saturation enthalpy is defined as satisfying the equation  $y(H) = 1$ . This gives

$$H(\infty) = \ln[(3RT - s)/2RT]/(1/RT - 1/s). \quad (7)$$

Since we get

$$\lim_{T \rightarrow T_c^-} H(\infty) = -s/2, \quad \text{while } H(\infty) = 0 \quad \text{if } T > T_c, \quad (8)$$

we obtain at  $T_c$  a first-order phase transition with latent heat  $s/2$ , in qualitative agreement with experimental findings rooted in calorimetric measurements of denaturation [15].

The time-dependent behavior of the expected enthalpy displayed in Fig. 1 reveals the existence of a metastable folded phase emerging as a dynamic equilibrium. This is so since the minimum free energy realized is  $G = H(\infty) - (RT/s)H(\infty)$ . On the other hand, since  $G = (1 - RT/s)H$ , the free energy could in principle decrease boundlessly in the limit of long chains considered. There-

fore, we may conclude that the metastable phase becomes dominant as a dynamic equilibrium in the range  $273 \text{ K} < T < T_c$ . Actually, no analysis is required below the freezing point of the solvent (273 K), since the folding process cannot take place.

The kinetic approach presented is consistent with the coarse description of conformation space partitioned in quasiequilibrium CP states. In this regard, our approach differs from other more detailed kinetic treatments of folding rooted in fundamental dynamical assumptions [16]. Our model is justified and corroborated on the following grounds: (1) The relaxation dynamics below criticality reproduces the known relaxation behavior of disordered glassy materials. (2) There exists a dynamically dominant *metastable* folded phase which undergoes a first-order phase transition with latent heat at the critical temperature, in accord with calorimetric experiments and with mounting evidence pointing towards the need for an alternative predictive tool radically different from free energy minimization algorithms.

At this point we rationalize the results within the context of a variational principle. Since we are dealing with ensemble-averaged relaxation, we shall consider a generic situation under the following constraints: (a) The free energy landscape is rugged because of the high degeneracy of enthalpy levels; (b) the expected barrier  $B(t)$  encountered at time  $t$  grows monotonically with  $t$ , as is the case for a system of increasing complexity in which successive relaxation steps become increasingly difficult; and (c) the progress of relaxation may be monitored by a single-valued function  $X = X(B)$ , which may be taken as proportional to  $\langle H \rangle$  in the folding context.

Under these tenets, the path integral functional giving the overall relaxation time is

$$\int dt = F^{-1} \int_0^b \exp(B/RT) (1 + X'^2)^{1/2} dB, \quad (9)$$

where  $X' = dX/dB$ ,  $(1 + X'^2)^{1/2} dB$  is the arc differential,  $F = RTf$  and  $F^{-1} \exp(B/RT)$  is the reciprocal of the velocity when the activation barrier is  $B$ . By solving the Euler-Lagrange equations we find the relaxation pathway that minimizes the functional. This pathway is defined by the equations

$$B = B(t) = RT/2 \ln[(ft)^2 + c^2]; \quad (10)$$

$$X = RT \arg \tan[ft/c].$$

Since  $X \leq 0$ , the integration constant  $c$  is less than zero, and for typical choices of  $c$  appropriate to represent the relaxation behavior within relevant long time scales ( $t > 100 \mu\text{s}$  in the folding context [1,3,5,7,10]), we get  $(ft)^2 \gg c^2$  or  $B(t) \approx RT \ln(ft)$ , which is precisely the logarithmic increase in barrier size with time, the signature of glassy relaxation under the given assumptions. This relaxation has been already singled out in the REM model which holds valid for random RNA copolymers [1], it has been established within a correlated landscape

in the spin-glass analogs of protein collapse [17], and it appears to represent a generic situation in rugged landscape models of helix-coil transitions, as pointed out by de Gennes [18]. Furthermore, the time behavior of the generic energy variable  $X$  along the brachistochrone relaxation pathway defined as the extremal of the path integral given in Eq. (10) is in excellent agreement with that of  $\langle H(t) \rangle$ , the relevant coordinate monitoring the progress of the folding process. This fact results from direct comparison of Figs. 1 and 2.

Thus, the variational formulation described by Eq. (10) not only validates the statistical theory of folding presented in this work, but also reveals that the logarithmic law which governs the glassy relaxation of several materials with quenched disorder actually represents the fastest mode of relaxation among pathways which entail a monotonic increase with time in the size of the expected activation barrier. The result has intuitive appeal: In any process where relaxation steps become increasingly diffi-

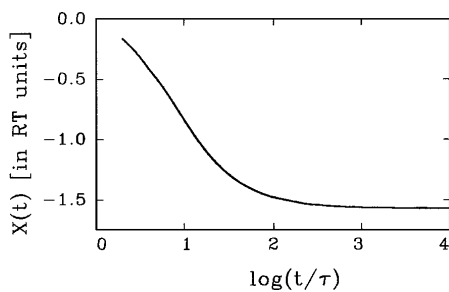


FIG. 2. Time-dependent behavior at  $T' = 1/100$  of the generic coordinate  $X$  along the brachistochrone relaxation pathway for processes with increasing size of the expected activation barrier. The choice of the integration constant  $c$  for the extremal of the functional given by Eq. (10) is  $c = -10$ . Since  $X = X(t)$  is an energy coordinate, it is proportional to  $\langle H(t) \rangle$  if the variational results are specialized to the context of the folding problem. Thus, the agreement in the relaxation behavior as revealed by comparison of Figs. 1 and 2 validates the variational formulation.

cult, the slowest possible growth of the activation energy barriers is provided by the logarithmic growth. For this reason, the variational formulation holds beyond the context of folding of disordered copolymers.

- 
- [1] A. Fernández and E.I. Shakhnovich, Phys. Rev. A **42**, 3657 (1990).
  - [2] J.D. Bryngelson and P.G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **84**, 7524 (1987); A. Fernández, Ber. Bunsen-Ges. Phys. Chem. **98**, 260 (1994).
  - [3] A. Fernández, Phys. Rev. Lett. **64**, 2328 (1990).
  - [4] H. Frauenfelder, S. Sligar, and P.G. Wolynes, Science **254**, 1598 (1991).
  - [5] A. Fernández, Physica (Amsterdam) **201A**, 557 (1993).
  - [6] B. Derrida, Phys. Rev. B **24**, 2613 (1981).
  - [7] A. Fernández, H. Arias, and D. Guerin, Phys. Rev. E **52**, R1299 (1995).
  - [8] R. Jaenicke, Angew. Chem. Int. Ed. Engl. **23**, 295 (1984).
  - [9] T. Creighton, J. Phys. Chem. **89**, 2452 (1985); T. Creighton, in *Protein Folding*, edited by L.M. Gierasch and J. King (American Association for the Advancement of Science Publishers, Washington, DC, 1990), pp. 157–170.
  - [10] A. Fernández, Ann. Phys. (Leipzig) **4**, 600 (1995).
  - [11] P. Zarrinkar and J. Williamson, Science **265**, 918 (1994).
  - [12] E.I. Shakhnovich, G. Farztdinov, A. Gutin, and M. Karplus, Phys. Rev. Lett. **67**, 1665 (1991).
  - [13] R.G. Palmer, D.L. Stein, E. Abrahams, and P.W. Anderson, Phys. Rev. Lett. **53**, 958 (1984).
  - [14] A. Fernández, Phys. Rev. Lett. **65**, 2259 (1990).
  - [15] P.L. Privalov, in *Protein Structure and Protein Engineering*, edited by E.-L. Winnacker and R. Huber (Springer, Berlin, 1988), pp. 6–15.
  - [16] E.G. Timoshenko, Y.A. Kuznetsov, and K.A. Dawson, Phys. Rev. E **53**, 3886 (1996).
  - [17] S. Plotkin, J. Wang, and P.G. Wolynes, Phys. Rev. E **53**, 6271 (1996).
  - [18] P.G. de Gennes, J. Stat. Phys. **12**, 463 (1975).