Folding pathways as brachistochrones

Ariel Fernández and Blanca Niel *

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

ABSTRACT: The folding of natural biopolymers is an expeditious process taking place within timescales incommensurably shorter than ergodic times. Furthermore, its robustness suggests that the process must depend on a relatively coarse level of resolution of conformation space. To account for these features we derive a variational principle formulated within an adiabatic approximation obtained by integrating out fast-relaxing molecular motions. Folding pathways are generated by means of a stochastic process which begets a least effort principle reflecting a stepwise minimization of the conformational cost associated to contact formation. This economy of the process is found to have meaningful kinetic consequences if we treat base-pairing contact patterns (BPP's) adiabatically, that is, as quasi-equilibrium states: Our computations reveal that the probability distribution of overall folding timespans associated to the process resolved at the BPP level is maximized at the brachistochrone or overall least-time pathway for functionally-competent molecules. This result represents a mathematical footprint of evolution at the molecular level.

1- Motivations for a variational treatment within an adiabatic approximation

The folding of a natural biopolymers (RNA, proteins) under *in vitro* solvent conditions is an expeditious, efficient and reproducible process which represents the search in conformation space performed by a biopolymer chain that forms intramolecular contacts at the expense of losing conformational freedom. The difficulty in finding theoretical underpinnings of such phenomena is due to the fact that folding is neither an energetically-downhill process nor the result of an exhaustive random exploration of conformation possibilities [1-4], the extreme cases which would make the problem far more tractable. On the other hand, the robustness of the process suggests that a simplification might be plausible: The folding process cannot depend on atomic-level detail but must be definable at a coarser level of resolution, as the present work reveals.

Such a framework calls for an action principle which should single out folding pathways resolved within a description of conformation space compatible with the means of detection [2,3,6]. If we focus on RNA, the detection of folding events along a folding pathway is essentially rooted in RNA-DNA hybridization techniques [6] which can only resolve structure at the level of base pairing contact patterns (BPP's). It is precisely at

* e-mail: biniel@criba.edu.ar

this level of resolution that a central problem arises: Obviously, to assume that the exploration of conformation space is dictated by a sequence of BPP transitions governed by an action principle [5] implies first of all that an adiabatic approximation to folding dynamics holds in this context. Thus, in the spirit of the Born-Oppenheimer approximation in molecular physics, treating BPP's as quasi-equilibrium states implies the existence of slow and enslaving folding degrees of freedom: If we rely on known estimates of mean relaxation parameters [7-9], there appears to exist a separation of timescales between folding events resolved as BPP transitions $\left(10^{-4}-10^{3}\ s\right)$ [1-14] on one hand, and relaxation timescales for torsional dihedral motions $\left(10^{-12}-10^{-7}\ s\right)$, rotation about glycosidic bonds $\left(10^{-11}\ s\right)$ and vibrations $\left(10^{-15}\ s\right)$ on the other hand. However, there are caveats when handling this information: Spectroscopic measurements of timescales have been performed mostly for individual nucleotides (the RNA monomeric units), and often under *in vacuo* conditions.

Is the adiabatic assumption a valid one? If so, then fast microscopic motion could be integrated out as entropy, each BPP would represent indeed a state of quasiequilibrium, and forward and backward activation barriers associated to BPP transitions could be evaluated respectively taking into account either the conformational entropy loss or the enthalpy loss associated to intrachain contact formation [1]. This is indeed the picture we have adopted in this work for natural RNA's (ribozymes), and it reproduces the biologically-active BPP, as probed experimentally [6] and confirmed by phylogenetic analysis [15]. Specifically, the core question we address is *whether a variational principle may be formulated so that it singles out a folding pathway whose destiny BPP is biologically active and has been inferred independently by phylogenetic analysis.*

This work is organized as follows: Section 2 deals with the formulation of the least action principle and Section 3 is devoted to showing that catalytically-competent RNA species, the so-called ribozymes [3], actually fold according to the least action principle previously formulated.

2- The formulation of the variational principle

This section deals with natural RNA's in acqueous solution which search *in vitro* for their native conformations under renaturation conditions. We examine folding pathways generated by realizing a stochastic process whereby the preferred folding step at each stage of the process is chosen following the tenet of sequential or stepwise minimization of conformational entropy loss (SMEL) [16]. Thus, the optimal or most favored pathway is the one in which the RNA chain seeks to maximize with each folding event the number of contacts forming an intramolecular helix while minimizing the kinetic barrier (or loss in conformational entropy [16,17]) associated to the loop closure that leads to the helix formation. This implies a choice based on local features of the free energy landscape, reflecting a maximum economy by minimization of the loss in conformational freedom with each favored folding step.

For an arbitrary RNA sequence, the local minimization of the mean escape time by undertaking the transition that entails the lowest barrier does not necessarily imply that the overall time involved in the favored folding pathway will be the minimum possible. *However, the results expounded in the next section for RNA sequences which are products of natural selection support the fact that the preferred pathway generated in a SMEL-based simulation is indeed the brachistochrone, or minimum overall time trajectory.* Thus, the results reveal a variational principle which governs pathways followed by natural biopolymers with random coil and active conformation as fixed endpoints: The pathway y^* that carries the highest statistical weight is actually a minimum of the functional

$$\Omega(y) = \sum_{b_i \in B(y)} e^{b_i} [f n_i]^{-1} =$$

$$= \text{sum of mean escape times along pathway } y. \qquad (1)$$

Here $B(y)$ is the collection of kinetic barriers in RT units to be surmounted along pathway y up to the point when the active conformation is reached; $f \approx 10^6\ s^{-1}$ is the rate constant for base pair formation [1,16] once the nucleation step leading to intra-chain helix formation has taken place and n_i is the number of base pairs in the ith intramolecular helix whose formation or dismantling-depending on the nature of the ith step-entails surmounting activation barrier b_i . The collection $B(y)$ is finite since a coarse-grained resolution of conformation space has been assumed (vide infra). Thus, each term in the sum on Eq. 1 is the reciprocal of the unimolecular rate constant for an elementary folding event [1,16,18]. The functional $\Omega(y)$ gives the total timespan for the overall transition along pathway y from the random coil to the biologically-active folding.

We shall start by specifying the degree of coarse graining of conformation space X. The RNA chain folds intramolecularly by base-pair association of complementary residues or nucleotides forming antiparallel stems with a concurrent loss in conformational entropy due to loop formation. *In our coarse description two RNA conformations are regarded as equivalent if they share the same BPP.* This equivalence relation determines a partition Z on X in mutually-disjoint equivalence classes. By BPP we not only mean secondary structure, incorporating all planar motifs resulting from hairpin, bulge or internal loops, but we also incorporate the pseudoknot motif [19]. A pseudoknot forms when the residues in a hairpin loop engage in base pairing with residues outside the hairpin, forming an additional stem and loop region.

A stochastic process ξ, whose realizations are coarse-grained folding pathways, has been defined assigning transition probabilities between elements of Z. To implement the process at the computational level, we first make use of current combinatorial algorithms (see, for example, [20]) to predict all plausible BPP's. Such algorithms incorporate the pseudoknot as a tertiary interaction motif and consider only base pairing and stacking as

stabilizing interactions in intramolecular structure. The stochastic process is determined by the activation energy barriers required to produce or dismantle stabilizing interactions. Thus, at each instant, the partially-folded chain undergoes a series of disjoint elementary events with transition probabilities dictated by the unimolecular rates of the events. The stochastic process is Markovian since the choice of the set of disjoint events at each stage of folding is independent of the history that led to that particular stage of the process.

The process is mechanistically constructed as follows: For each time $t \in I$, we define a map $t \to J(x,t) = \{j: 1 \le j \le n(x,t)\}$, where $J(x,t)$ = collection of elementary events representing conformational changes which are feasible at time t given that the initial conformation x has been chosen at time $t = 0$, and $n(x,t)$ = number of possible elementary events at time t. Associated to each event, there is a unimolecular rate constant $k_j(x,t)$ = rate constant for the jth event [16] which may take place at time t for a process that starts with conformation x. The mean time for an elementary refolding event is the reciprocal of its unimolecular rate constant. Thus, for a fixed time interval I, the only elementary events allowed are elementary refolding events that satisfy: $k_j(x,t)^{-1} \le |I|$.

We now introduce a random variable $r \in \left[0, \sum_{j=1}^{n(x,t)} k_j(x,t)\right]$, uniformly distributed over the interval. Let r^* be a particular realization of r arising in a simulation of the process, then there exists an index j^* such that

$$\sum_{j=0}^{j^*-1} k_j(x,t) < r^* \le \sum_{j=0}^{j^*} k_j(x,t) \quad , \qquad (2)$$

$$(k_0(x,t) = 0 \text{ for any } (x,t))$$

This implies that the event $j^* = j^*(x,t)$ is chosen at time t for the folding process that starts at conformation x. The map $t \to j^*(x,t)$ for fixed initial condition x constitutes a realization of the Markov process which determines the folding pathway ξ_X. In turn, the probability that the j^*-event is chosen at time t is

$$\left[k_{j^*}(x,t) / \sum_{j' \in J(x,t)} k_{j'}(x,t)\right].$$

Explicit values of the unimolecular rate constants require an updated compilation of the thermodynamic parameters at renaturation conditions [20]. These parameters are used to generate the set of kinetic barriers associated to the formation and dismantling of stabilizing interactions, the elementary events in our context of interest. Thus, the activation energy barrier for the rate-determining step in the formation of a stabilizing

interaction is known to be $-T\Delta S_{loop}$, where $T\Delta S_{loop}$ indicates the loss of conformational entropy associated to closing a loop. Such a loop might be of any of four admissible classes: bulge, hairpin, internal or pseudoknotted. For a fixed number L of unpaired bases in the loop, we shall assume the kinetic barrier to be the same for any of the four possible types of loops [20]. This assumption is warranted since the loss in conformational entropy is due to two overlapping effects of different magnitude: The excluded volume effect, meaningful for relatively large L ($L \geq 100$) and the orientational effect that tends to favor the exposure of phosphate moieties towards the bulk solvent domain for better solvation. Since both effects are independent of the type of loop, we may conclude in relatively good agreement with calorimetric measurements, that the kinetic barriers are independent of the type of loop for fixed L . On the other hand, the activation energy barrier associated with dismantling a stem is $-\Delta H(stem)$, the amount of heat released due to base-pairing and stacking when forming all contacts in the stem.

For completion we shall display the analytic expressions for the unimolecular rate constants [1,16]. If the jth step or event happens to be a helix decay process, we obtain:

$$k_j = f\, n\, exp\left[G_h / RT\right] \tag{3}$$

where n is the number of base pairs in the helix formed in the jth step and G_h is the (negative) free energy contribution resulting from stacking of the base pairs in the helix. Thus, the essentially enthalpic term $-G_h = -\Delta H(stem)$ should be regarded as the activation energy for helix disruption. On the other hand, if the formation of an admissible stabilizing interaction happens to be the event designated by the jth step, the inverse of the mean time for the transition will be given by:

$$k_j = f\, n\, exp\left[-\Delta G_{loop} / RT\right] \tag{4}$$

where $\Delta G_{loop} \approx -T\Delta S_{loop}$ is the change in free energy due to the closure of the loop concurrent with the helix formation.

The folding pathways are generated through Monte Carlo (MC) simulations with each run representing a realization of the stochastic process described above. To identify the variational principle underlying the search for the active structure in biologically-relevant RNA's, we introduce the probability distribution P of total timespans of pathways with random coil and active BPP,s as fixed endpoints. This distribution is computationally accessible within our coarse-grained representation via the following working formula:

$$P(\tau)\Delta\tau = \text{probability that the overall folding time lies in the interval } \left[\tau, \tau + \Delta\tau\right] = \tag{5}$$
$$= \#\{\text{generated pathways } y\text{'s with } \tau \leq \Omega(y) \leq \tau + \Delta\tau\}/\#\{\text{all generated pathways}\},$$

where the symbol # denotes cardinal of a set.

3- The variational principle actually holds in the folding of ribozymes

The results of 100 runs each consisting of 10^6 MC steps performed for each RNA species are displayed in Fig. 1 for three natural RNA species endowed with functional properties, sun YL-13, tdL-7 and TtLSU. These species are ribozymes or catalytically-competent RNA's of the so-called group I [15] whose active or functional BPP has been inferred by phylogenetic analysis. These species have lengths 386, 405 and 411, respectively. In all cases a distinctive feature becomes apparent (cf. Fig. 1): *The most probable folding time is the shortest time. The same conclusion was actually found to hold for all 87 functional RNA species of group I. The results reveal a variational principle that holds for those RNA sequences which are targets of natural selection: The most probable folding pathway is the brachistochrone trajectory in conformation space.* Thus, in choosing folding steps which entail the minimal loss in conformational freedom, real RNA chains fold so as to minimize the functional Ω defined by Eq. 1.

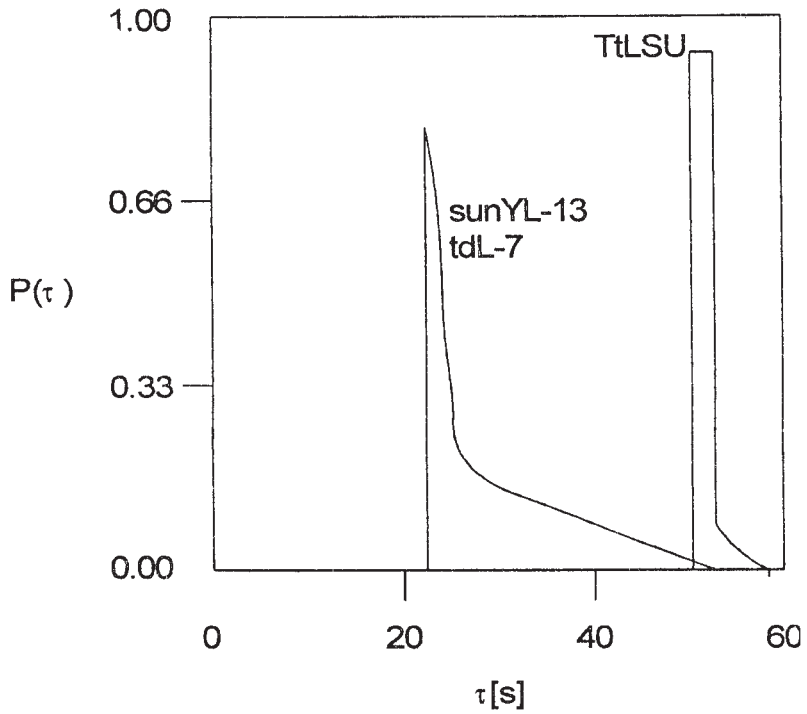

Fig. 1. Probability distribution $(P(\tau))$ of over-all folding times (τ) for different ribozymes.

4- Conclusion

The results presented in this work support the fact that catalytically-competent RNA species fold following a least action which singles out the brachistochrone or least overall time trajectory as the dominant folding pathway. This pathway is unambiguously defined within a level of resolution coarser than the atomic scale, precisely the level of description at which an adiabatic approximation holds: Each base pairing contact pattern (BPP) may be regarded as a quasi-equilibrium state when focusing on the kinetics of the folding process.

The variational principle accounts for the expediency of folding of a natural species and reflects the economy of the process, since the brachistochrone or least over-all time pathway involves only the folding steps that entail a minimal loss of conformational freedom. Futhermore, the least-action folding pathway has been probed by recent kinetic experiments.

To conclude we may state that by casting the folding problem in a variational context we have found a mathematical footprint of natural selection.

Acknowledgement

This research was carried out in part at the Frick Laboratory in Princeton University under a J. S. Guggenheim Memorial Foundation fellowship awarded to A. F.. The hospitality and fruitful discussions with Prof. H. Rabitz and his research group are gratefully acknowledged. Conversations with Prof. Jerome Percus from the Courant Institute of Mathematical Sciences, New York University, has proven instrumental for this research.

References

[1] A. Fernández, *Ann. Physik* **4** (1995) 600
[2] K. A. Dill, K. M. Fiebig and H. S. Chan, Proc. Natl. Acad. Sci. USA **90**, (1993) 1942
[3] E. Shakhnovich, G. Farztdinov, A. Gutin and M. Karplus *Phys. Rev. Lett.* **67** (1991) 1665
[4] R. Jaenicke, *Angew. Chem. Intl. Ed. Engl.* **23** (1994) 295
[5] T. Creighton, *J. Phys. Chem.* **89** (1985) 2452
[6] P. Zarrinkar and J. Williamson, *Science* **265** (1994) 918
[7] C. Cantor and P. Schimmel, "Biophysical Chemistry", Freeman and Co., New York (1980)
[8] W. Saenger, "Principles of Nucleic Acid Structure", Springer, New York (1984)
[9] C. Brooks, M. Karplus and M. Pettitt, "Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics", Wiley, New York (1988)
[10] A. Fernández and E. I. Shakhnovich, *Phys. Rev. A* **42** (1990) 3657
[11] J. D. Bryngelson and P. G. Wolynes, *Proc. Natl. Acad. Sci.* USA **84**(1987) 7524
[12] A. Fernández, *Phys. Rev. Lett.* **64** (1990) 2328
[13] H. Frauenfelder, S. Sligar and P. G. Wolynes, *Science* **254** (1991) 1598
[14] A. Fernández, *Physica A* **201** (1993) 557
[15] F. Michel and E. Westhof, *J. Mol. Biol.* **216** (1990) 585
[16] A. Fernández and H. Cendra, *Biophysical Chemistry* **58** (1996) 335
[17] A. Fernández, *J. Math. Chem.* **17** (1995) 401
[18] A. Fernández, *Phys. Rev. A - Rapid Comm.* **45** (1992) R8348
[19] C. W. Pleij, K. Rietveld, and L. Bosch, *Nucleic Acds. Res.* **13** (1985)1717
[20] J. A. Jaeger, D. H. Turner and M. Zuker, *Proc. Natl. Acad. Sci.* USA **86** (1989) 7706