ELSEVIER

Biophysical Chemistry 74 (1998) 89–98

Biophysical Chemistry

The RNA folding problem: a variational problem within an adiabatic approximation

Ariel Fernández[a,b,*], Blanca Niel[a], Teresita Burastero[a]

[a] *Instituto de Matemática, Universidad Nacional del Sur, Consejo Nacional de Investigaciones Científicas y Técnicas, Avenida Alem 1253, Bahía Blanca 8000, Argentina*
[b] *The Frick Laboratory, Princeton University, Princeton, NJ 08544, USA*

Received 16 February 1998; accepted 30 April 1998

Abstract

Biopolymer folding 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 while focusing on the RNA context, 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 entropy cost for each folding event with concurrent maximization of the base pairing. This economy of the process is found to have kinetic consequences if we treat base-pairing contact patterns (BPPs) adiabatically, that is, as quasi-equilibrium states: 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 RNAs. In turn, this pathway is shown to yield all the phylogenetically-conserved structural features of the active conformation within biologically-relevant timescales.

Keywords: Biopolymer folding; Coarse resolution; Variational principle; Pseudoknot

1. Motivations for a variational treatment within an adiabatic approximation

The folding of 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 exhaus-

* Corresponding author. e-mail: inmabb@criba.edu.ar

0301-4622/98/$19.00
PII S0301-4622(98)00168-9

tive 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 (BPPs). It is precisely at 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 BPPs 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 (10^{-4}–10^{3} s) [1–14] on one hand, and relaxation timescales for torsional dihedral motions (10^{-12}–10^{-7} s), puckering (10^{-10}–10^{-7} s), rotation about glycosidic bonds (10^{-11} s) and vibrations 10^{-15} s) 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 functional RNAs (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. In Section 3 we show that catalytically-competent RNA species, the so-called ribozymes [13], actually fold according to the least action principle formulated in Section 2. Section 4 is devoted to a detailed characterization of the folding pathways which are extremals of the action. Particular attention is paid to showing that the catalytically-competent BPP already inferred by phylogenetic analysis is actually identical to the destiny BPP of the folding pathway singled out by the action or path functional formulated in Section 2.

2. The formulation of the variational principle

This section deals with natural RNAs in aqueous 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 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 under-

taking 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

$$\Omega(y) = \sum_{b_i \in B(y)} e^{b_i}[fn_i]^{-1} = \text{sum of mean escape times along pathway } y \quad (1)$$

Here $B(y)$ is the collection of kinetic barriers 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 intrachain 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 of 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 [20,21] to predict all plausible BPPs. 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. The time t at which an event belonging to $J(x,t)$ could actually begin to materialize is equal to the sum of the mean escape times for the events previously chosen, with x as the starting conformation. 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: $k_j(x,t)^{-1} \leq |I|$.

We now introduce a random variable $r \in [0, \sum_{j=1}^{n(x,t)} k_j(x,t)]$, 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

$$\sum_{j=0}^{j^*-1} k_j(x,t) < r^* \leq \sum_{j=0}^{j^*} k_j(x,t), \quad (k_0(x,t) = 0 \quad \text{for any } x,t) \tag{2}$$

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:

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

At this point, we must emphasize that the folding pathway generated through the stochastic process thus described is resolved at the level of transitions between BPPs because such a coarse description is essentially compatible with the adiabatic approximation, and this is precisely why we can actually compute transition rates.

Explicit values of the unimolecular rate constants require an updated compilation of the thermodynamic parameters at renaturation conditions [20,21]. 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 $-T\Delta S_{\text{loop}}$, where ΔS_{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 [21]. 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 a fixed L. The reader should observe in this regard that an early compilation of thermodynamic parameters [20] reveals differences in the free energies for different types of loops of the same size. This compilation has been revised and more recent calorimetric measurements prove to be insensitive within experimental error to the type of loop in agreement with our theoretical results [21]. On the other hand, the activation energy barrier associated with dismantling a stem is $-\Delta G(\text{stem})$, 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:

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

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 base pairing and stacking in the helix. Thus, the term $-G_h$ 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:

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

where $\Delta G_{loop} \approx -T\Delta S_{loop}$ is the change in free energy due to the closure of the loop concurrent with helix formation. Now n is the number of putative base-pairing, contacts that would form if the elementary event whose rate constant is given explicitly by Eq. (4) would take place. Thus, the collision factor, being itself proportional to the number of successful collisions, must be proportional to n. The actual estimations of collision factors and kinetic barriers involved imply that the elementary rate processes described are not diffusion-controlled. The dominant role of the solvent is by far that of offering different dielectric environments to different portions of the chain, according to the conformation adopted. Thus, because of the exponential dependence given in Eq. (4), the decisive influence of the solvent upon the rate stems from determining the conformational entropy cost associated to a folding event [16].

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 RNAs, we introduce the probability distribution P of total timespans of pathways with random coil and active BPPs as fixed endpoints. This distribution is computationally accessible within our coarse-grained representation via the following working formula:

$$\begin{aligned} P(\tau)\Delta\tau = {} & \text{probability that the overall folding} \\ & \text{time lies in the interval}[\tau,\tau+\Delta\tau] \\ = {} & \#\{\text{generated pathway with} \\ & \tau \le \Omega(y) \le \tau+\Delta\tau\}/ \\ & \#\{\text{all generated pathways}\}, \end{aligned} \tag{5}$$

where the symbol # denotes cardinal of a set.

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

The results of 10 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 RNAs 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, as our kinetically-controlled simulations have revealed*. The results represented by the time sequence of events are identical for all runs for each RNA species and show that a variational principle underlies the folding of 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).

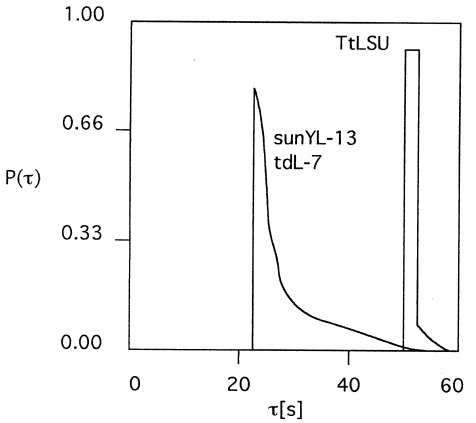

Fig. 1. Probability distribution $P(\tau)$ of folding time τ for three ribozymes of group I, sunYL-13, tdL-7 and TtLSU, obtained using a SMEL-based algorithm after 10^6 MC steps. Exact analytic expressions for the entropy loss associated to loop closure have been used (Eqs. (6),(7)). The only source of systematic error arises from uncertainty in the calorimetric parameters for enthalpic contributions [20,21]. Within this uncertainty the distribution obtained for the sunYL-13 and tdL-7 are identical.

4. Describing the brachistochrone folding pathway for specific ribozymes

4.1. The folding of ribozymes

Prior to identifying the folding pathways corresponding to brachistochrone trajectories, we need to discuss the kinetics of formation of the pseudoknot, a structural motif central to ribozyme function (see Figs. 2 and 3). As we shall presently show using orientational arguments to compute the entropic cost associated to pseudoknot formation, this structural motif cannot be planar, it must form a pocket or 'hinge-like' spatial motif which produces a cavity essential for catalysis.

In order to determine its activation barrier we first note that the entropy loss associated to closure of a loop of length L may be computed as [16]:

$$\Delta S = -\mu R \ln L - RL \ln 2, \quad (L < 18) \tag{6}$$

$$\Delta S = -\mu R \ln L, \quad (L \geq 18) \tag{7}$$

where the first term in the r.h.s. of Eq. (6), with $\mu \approx 1.75$, is the excluded volume contribution in good solvent and the second term is the orientational contribution meaningful for $L < 18$. The critical loop size depends on the solvent and is fixed at $L = 18$ for water. The orientational contribution has been studied only recently and results from the selective orientation of the charged phosphate groups towards the outer solvent domain defined once the loop of an object of rod-like dimensions, the RNA chain, has formed. For $L \geq 18$, the inner solvent domain may be regarded as bulk-like since the cluster dimensions allow for four solvation layers for each phosphate moiety pointing inwards and therefore, dielectric differences with the outer domain become so negligible that no orientational preference holds: there is no conformational entropy loss in this case. In other words, the cluster-like solvent domain confined within the loop of an RNA chain becomes dielectrically indistinguishable from the outer bulk solvent as a cluster of at least seven water molecules across is confined by the loop [16]. This fact, together with the actual effective dimensions of the RNA chain have enabled us to determine the critical loop size beyond which there is no orientational contribution to the entropy loss entailed by loop closure.

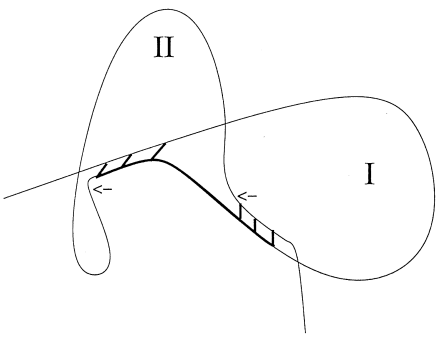

Fig. 2. Spatial representation of a pseudoknot comprising hairpin loop I followed by loop II. The region to be oriented concurrently with closure of loop II is displayed by a thin line between the arrows. The thick line region common to loops I and II has been already oriented upon formation of loop I and prevents both loops from being coplanar provided their respective size is below the critical size.

The activation barrier associated to pseudoknot formation can now be easily determined using orientational arguments: suppose the generic hairpin I has been formed (see Fig. 2) and we need to obtain the activation barrier associated with forming loop II, and suppose loops I and II are below critical range. We first observe that both loops cannot be coplanar since the common region (thick solid line in Fig. 2) has been already oriented towards bulk water upon formation of loop I. Should loop II form coplanarly, this would affect the hydration self-energy of the phosphates in the region common to loops I and II by drastically reducing the dielectric in which the phosphates were previously immersed. Thus, the two loops should form in the non-coplanar manner suggested by Fig. 2 and the orientational entropy contribution for loop II should be taken as $-RL \ln 2$, with L being the number of nucleotides exclusively belonging to loop II in between the arrows, as indicated in Fig. 2. In the case that the size of at least one of the loops comprising the pseudoknot were beyond critical range, there would be no orientational restriction

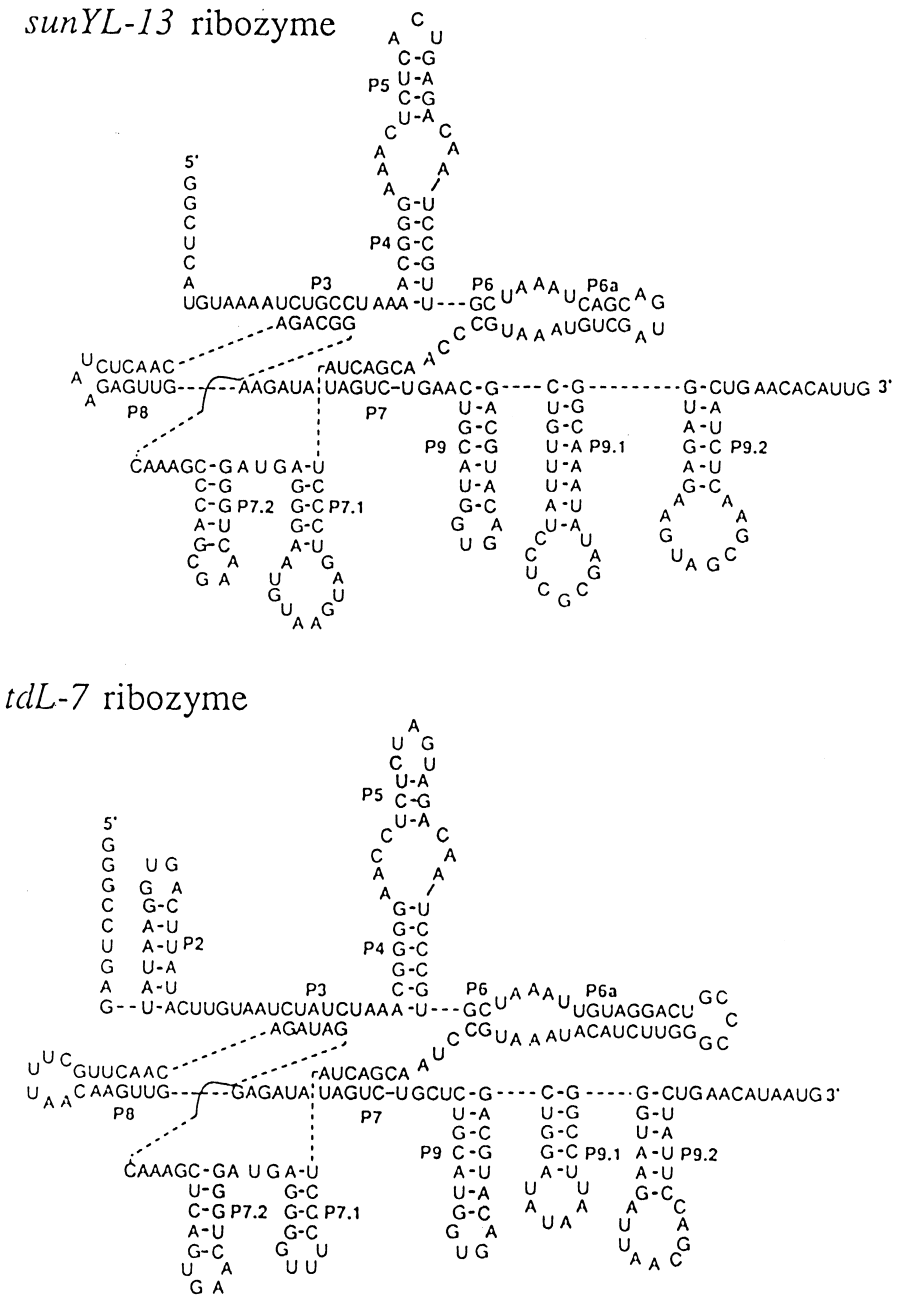

Fig. 3. Secondary structure for the sunYL-13 and tdL-7 group I ribozymes obtained from the SMEL-based sequential algorithms. Both structures are destinies of the folding pathways singled out by the functional Ω.

to coplanarity. However, in all known cases of ribozyme folding, this second situation does not arise, as indicated in Section 4.2.

An underlying assumption in the previous arguments is that the solvent is pure water. However, it is well known that the presence of Mg(II) ions play a decisive role in the kinetics of ribozyme folding [6], significantly lowering the activation

barriers associated to pseudoknot and loop formation. This effect must be quantified: It is well known that Mg(II) binds in a 'weak' form to adjacent phosphates of unpaired nucleotides by forming a chelate complex [22]. This fact must be encompassed within our approach by each pair of adjacent coordinated phosphates as a single orientational entity. Thus, in computing the folding pathways we shall adopt the working hypothesis that the orientational contribution to the kinetic barrier associated to loop closure is strictly reduced by one-half under realistic conditions for ribozyme folding.

In order to treat the kinetics of pseudoknot formation, the activation barriers of entropic nature associated to the concurrent loops I and II (Fig. 2) need to be determined. Since entropic parameters are particularly unreliable [20,21], we have determined analytically the entropic contributions for the two concurrent loops below criticality. Our results expounded in Section 4.2 reveal the role of pseudoknots in determining the actual bottlenecks in the formation of the catalytic core and the timings of such events are in solid agreement with kinetic experiments [6], thus supporting the accuracy of our kinetic parameters. Other attempts to determine the folding kinetics of pseudoknots have been performed [23]. However, such attempts are based on a combination of genetic algorithms and thermodynamic free energy parametrization, leading to a stepwise selection of the best fitted structures, simulating natural selection. Since such results do not allow for a direct estimation of the entropic contributions to the kinetic barriers of pseudoknot formation, an actual comparison with the kinetic experiments on ribozyme folding expounded in Zarrinkar [6] is not possible, in contrast with the treatment presented in this work. Furthermore, because the entropic contribution has not been previously obtained analytically, no inference on the non-coplanarity of the pseudoknot loops has been hitherto possible. Thus, to the best of the authors' knowledge, this is the first time that the spatial topology of the catalytic core has been inferred from purely entropic arguments which are also used to validate the theoretical results vis-a-vis actual kinetic experiments. Furthermore, besides the present work, we know of no other theoretical treatment that might lead to the quantitative estimation of the effect of Mg(II) ions upon the kinetics of pseudoknot formation.

4.2. The chronology of events along the brachistochrone folding pathway

Making use of a SMEL-based algorithm we now elucidate the sequence of events which takes place along the brachistochrone folding pathway for the specific ribozymes. The kinetically-controlled Monte Carlo simulations are based on the working Eqs. (2)–(7). Stopping conditions are introduced so that any metastable folding that survives 10^6 MC iterations is assumed to be stable enough within a biological context so that its catalytic competence may be examined (see below, and [6]). Accordingly, once such a conformation is reached, no further search in conformation space is allowed. Ten runs for each natural RNA species of the so-called group I have been performed and the results reveal unambiguously that a single folding pathway prevails in all runs. Each folding step is assumed to be reversible, in accord with working Eq. (3) and Eq. (4). No phylogenetic constraints are imposed a-priori on the simulations, thus, non-conserved helix formation is in principle allowed. Comparison with phylogenetically-inferred structures has been performed only after the simulations have been carried out in an independent fashion.

As an illustration, the BPP transitions for the sunYL-13 and tdL-7 are computed [16]. The results concerning the times of formation of the different conserved helices are given in Table 1. On the other hand, Fig. 3 shows the BPPs for both ribozymes generated in 100 $s = 10^6$ MC steps. The structures are identical to the philogenetically-inferred ones [15]. We display the results for both ribozymes and discuss in detail the case of sunYL-13, since both cases are analogous. It becomes apparent from inspection of Table 1 that the role of Mg(II) is to reduce the kinetic barriers and make them compatible with the biologically relevant timescale and that the P3–P7 pseudoknot forms sequentially and cooperatively: The involvement of Mg(II) during completion of pseu-

Table 1
SMEL pathways for the sunYL-13 and tdL-7 ribozymes indicating the time of formation of different conserved domains

MC steps	Sun YL-13	tdL-7	Real time
	P9, P7.2	P9	16.0 μs
		P2, P7.2, P7.1	20.0 μs
10^3	P6.a	P6.a, P5	10^{-1} ms
	P8, P5, P4	P8, P9.1	8.0 ms
10^4	P9.1	P4	10.0 ms
10^5	P7.1, P9.2	P9.2	22.0 ms
	P7	P7	24.0 s
10^6	P3	P3	24.0 s

doknot formation, that is, in the occurrence of P3, essentially leads to a reduction of the rate-limiting step to the orientation of 11 phosphate-chelated complexes.

The sequence of events is described as follows (cf. Fig. 3): in the first stage, P9, P7.2 and P6a form since they involve the closure of the smallest unstrained loops for which the inner and outer domains are differentiated (tetraloops). For these loops, Mg(II) coordination reduces the entropic cost to the orientation of two entities, thus reducing the kinetic barrier by one-half. In the next stage, the strained triloops P8 and P5 and the domains P4 and P9.1 are formed. The formation of the latter involves the closure of loops of size seven and nine, respectively, which in the absence of magnesium would imply a great entropic cost. Phosphate chelation reduces the number of entities to be oriented to four and five, respectively, bringing the mean time of formation to within 10 ms. Even worse, the domains P7.1 and P9.2 involve closure of size 10 loops. Here Mg(II) chelation brings the timescale of formation to 22 ms. The occurrence of P8, P7.1 and P7.2 has a cooperative effect, shortening the size of the loop to be closed in order to form P7, the first stem of the pseudoknot. In this cooperative manner, the loop to be closed now falls well within critical range. In spite of this unfavorable conditions, the P7 stem forms in 24 s because of the assistance of magnesium ions, which lower the total orientational barrier to that of an equivalent loop of 11 orientational units. The formation of P7, the rate-limiting step of the entire folding process, would not take place within biologically-relevant timescales without Mg(II) participation. In turn, the occurrence of P7 easily induces the formation of P3. The total number of phosphates to be oriented is reduced to 15. This is so since eight phosphates common to P3 and P7 (see Fig. 3) have already been oriented. Thus, P3 formation entails the orientation of six groups overall, in turn divided into two regions of length three each, or an overall number of four orientational entities once Mg(II) participation is considered. The latter becomes a trivial task since its entropic cost is very low, thus revealing that the pseudoknot formation occurs sequentially and cooperatively.

Although a body of experimental evidence to probe RNA folding kinetics is still lacking, recent experimental work [6] is in accord with the predicted brachistochrone pathway. The kinetics of Mg(II)-induced folding of the group I Tetrahymena L-21 sca ribozyme has been probed by hybridization of complementary oligonucleotides. The formation of the P3–P7 pseudoknot was found to be the overall rate-limiting step, but, although P3 and P7 were formed approximately within the same timescale, it was not possible to decide experimentally whether the two domains formed cooperatively, sequentially or independently. This issue has now been elucidated in this work for species of the same family, and the specific role of Mg(II) in lowering the entropic cost associated to pseudoknot formation has been quantified and rationalized. Furthermore, the very same orientational argument that led us to the derivation of the entropic cost, also reveals that the pseudoknot cannot be a planar structure, thus shaping the catalytic pocket in accord with the fact that catalytic activity has been found to switch on at the same time at which the domains forming P3 and P7 become inaccessible to the probes.

5. Conclusion

The results presented in this work support the fact that catalytically-competent RNA species fold

following a least action principle 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 validity of the least-action principle underlying the search in conformation space is not a generic feature but actually holds for the very limited realm of RNA sequences chosen by natural selection: an adamant advantage of a naturally-selected RNA genotype stems directly from the fact that folding opportunistically by minimizing locally the transition time leads to the fastest route to materialize the phenotype or folded conformation. Since it is precisely at the phenotypic level that natural selection takes place, we may conclude that the local minimization of transition times, representing the local maximization in the economy of the process, becomes the expedient through which evolutionary pressure may be applied.

At a mechanistic level, the variational principle accounts for the expediency of the folding process and reflects the economy of the process since the brachistochrone or least overall time pathway involves only the folding steps that entail a minimal loss of conformational freedom. Furthermore, the least action folding pathway has been probed by recent kinetic experiments, and it leads to a destination structure that contains all phylogenetically-conserved elements that shape the catalytic core.

Acknowledgements

This research was carried out under the tenure of a J.S. Guggenheim Memorial Foundation fellowship awarded to A.F. This author thanks Prof. Jerome Percus from the Courant Institute of Mathematical Sciences, New York University and Prof. Herschel Rabitz from Princeton University for helpful discussions. A.F. is a principal investigator of CONICET, the National Research Council of Argentina.

References

[1] A. Fernández, Ann. Pkysik 4 (1995) 600.
[2] K.A. Dill, K.M. Fiebig, H.S. Chan, Proc. Natl. Acad. Sci. U.S.A. 90 (1993) 1942.
[3] E. Shakhnovich, G. Farztdinov, A. Gutin, 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, J. Williamson, Science 265 (1994) 918.
[7] C. Cantor, P. Schimmel, Biophysical Chemistry, Freeman and Co., New York, 1980.
[8] W. Saenaer, Principles of Nucleic Acid Structure, Springer, New York, 1984
[9] C. Brooks, M. Karplus, M. Pettitt, Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics, Wiley, New York, 1988.
[10] A. Fernández, E.I. Shakhnovich, Phys. Rev. A 42 (1990) 3657.
[11] J.D. Bryngelson, P.G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. 84 (1987) 7524.
[12] A. Fernández, Phys. Rev. Lett. 64 (1990) 2328.
[13] H. Frauenfelder, S. Sligar, P.G. Wolynes, Science 254 (1991) 1598.
[14] A. Fernández, Physica A 201 (1993) 557.
[15] F. Michel, E. Westhof, J. Mol. Biol. 216 (1990) 585.
[16] A. Fernández, H. Cendra, Biophys. Chem. 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, L. Bosch, Nucleic Acids Res. 13 (1985) 1717.
[20] J.A. Jaeger, D.H. Turner, M. Zuker, Proc. Natl. Acad. Sci. U.S.A. 86 (1989) 7706.
[21] A.E. Waiter, D.H. Turner, J. Kim, et al., Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 9218.
[22] T. Pan, D. Long, O.C. Uhlenbeck, in: R.F. Gesteland, J.F. Atkins (Eds.), The RNA World, Cold Spring, Harbor, N.Y., 12 1993, pp. 221–302
[23] A.P. Gultyaev, F.H. van Batenburg, C.W. Pleij, J. Mol. Biol. 250 (1995) 37.