Discretized torsional dynamics and the folding of an RNA chain

Ariel Fernández,* Rodolfo Salthú, and Hernán Cendra
Instituto de Matemática, Universidad Nacional del Sur, Consejo Nacional de Investigaciones Científicas y Técnicas, Avenida Alem 1253, Bahía Blanca 8000, Argentina

(Received 18 May 1998; revised manuscript received 30 November 1998)

The aim of this work is to implement a discrete coarse codification of local torsional states of the RNA chain backbone in order to explore the long-time limit dynamics and ultimately obtain a coarse solution to the RNA folding problem. A discrete representation of the soft-mode dynamics is turned into an algorithm for a rough structure prediction. The algorithm itself is inherently parallel, as it evaluates concurrent folding possibilities by pattern recognition, but it may be implemented in a personal computer as a chain of perturbation-translation-renormalization cycles performed on a binary matrix of local topological constraints. This requires suitable representational tools and a periodic quenching of the dynamics for system renormalization. A binary coding of local topological constraints associated with each structural motif is introduced, with each local topological constraint corresponding to a local torsional state. This treatment enables us to adopt a computation time step far larger than hydrodynamic drag time scales. Accordingly, the solvent is no longer treated as a hydrodynamic drag medium. Instead we incorporate its capacity for forming local conformation-dependent dielectric domains. Each translation of the matrix of local topological constraints (LTM's) depends on the conformation-dependent local dielectric created by a confined solvent. Folding pathways are resolved as transitions between patterns of locally encoded structural signals which change within the 1 ns–100 ms time scale range. These coarse folding pathways are generated by a search at regular intervals for structural patterns in the LTM. Each pattern is recorded as a base-pairing pattern (BPP) matrix, a consensus-evaluation operation subject to a renormalization feedback loop. Since several mutually conflicting consensus evaluations might occur at a given time, the need arises for a probabilistic approach appropriate for an ensemble of RNA molecules. Thus, a statistical dynamics of consensus formation is determined by the time evolution of the base pairing probability matrix. These dynamics are generated for a functional RNA molecule, a representative of the so-called group I ribozymes, in order to test the model. The resulting ensemble of conformations is sharply peaked and the most probable structure features the predominance of all phylogenetically conserved intrachain helices tantamount to ribozyme function. Furthermore, the magnesium-aided cooperativity that leads to the shaping of the catalytic core is elucidated. Once the predictive folding algorithm has been implemented, the validity of the so-called ''adiabatic approximation'' is tested. This approximation requires that conformational microstates be lumped up into BPP's which are treated as quasiequilibrium states, while folding pathways are coarsely represented as sequences of BPP transitions. To test the validity of this adiabatic ansatz, a computation of the coarse Shannon information entropy σ associated to the specific partition of conformation space into BPP's is performed taking into account the LTM evolution and contrasted with the adiabatic computation. The results reveal a subordination of torsional microstate dynamics to BPP transitions within time scales relevant to folding. This adiabatic entrainment in the long-time limit is thus identified as responsible for the expediency of the folding process. [S1063-651X(99)11607-6]

PACS number(s): 87.10.+e, 87.15.He, 87.15.Cc, 87.14.Gg

I. INTRODUCTION AND OUTLINE

This work is concerned with the theoretical underpinnings of the expedient by which natural biopolymers reach their active conformation under *in vitro* renaturation conditions within time scales incommensurably shorter than ergodic or thermodynamic times [1–7]. In simple generic terms, our aim is to determine the microscopic origin of the expediency of the folding process. In attempting to address this issue, we first note that there exists a vast gap between the time scales accessible to computer simulations of torsional dynamics ($1-10^3$ ps) and those relevant to folding events ($1\ \mu$s-10^2 s) [2,8]. At first sight, this fact by itself would render ours an impossible task. Nevertheless, we shall show that this is not so, at least if we focus on the RNA folding problem.

In this context, *ad hoc* simplifications of conformation space have been theoretically [2] and operationally [6] introduced to understand the folding expediency. These simplifications hinge upon basic structural tenets. (a) The RNA chain may be regarded as a torsionally complex flexible backbone with organic bases attached to it (one per unit or nucleotide). (b) The torsional constraints determined by the geometry of the RNA backbone make Watson-Crick base pairing the ubiquitous motif in intrachain contact patterns (CP's). Thus, each CP is actually realized as a base-pairing pattern (BPP). (c) Each BPP may be regarded as a quasiequilibrium state [2,4,6–9], whereby all torsional degrees of freedom are equilibrated or thermalized within the torsional constraints imposed by the particular BPP and within time

*Author to whom correspondence should be addressed. Electronic address: arifer@criba.edu.ar

scales incommensurably shorter than those associated with BPP transitions.

This set of assumptions represents an ''adiabatic ansatz,'' which considerably simplifies the folding problem. However, for all that these assumptions have been implicitly adopted in RNA folding algorithms [8], in the interpretation of kinetic data on folding pathways [6], and even in theoretical approaches to the problem [2], their microscopic foundations have remained elusive. Thus, we focus on this problem by first simplifying or coarse graining the torsional dynamics which we later resolve even more coarsely as a sequence of BPP transitions in order to compare with the adiabatic results.

A local analysis of the molecular dynamics (MD) of the RNA backbone [8,9] reveals that a binary or *cis-trans* resolution of the interval $[0,2\pi]$ of torsional states for each degree of freedom is valid within time scales larger than 10^2 ps. This means that each torsional state of the entire chain could be specified by a binary matrix which we shall call the local topological constraints matrix (LTM). The ith column in the LTM indicates the unit with contour number i along the chain, while the jth row indicates the jth degree of freedom for each unit.

Since each BPP imposes torsional constraints upon those units engaged in realizing the specified intrachain contacts, including the closure of the concurrent loops, the first issue that a microscopic treatment must deal with is to learn how to translate an LTM into a BPP. This correspondence is not one-to-one (injective), as shown below. Since a BPP can also be described as a matrix (denoted BPM) with its ij entry equal to 1 if units i and j are in contact and 0, otherwise, the translation operation becomes actually a projection map of one binary matrix space onto another one. Typically, several LTM's project onto a single BPM. This is so since all units which are not engaged in the formation of contacts or the closure of the concurrent loops are free to adopt any torsional isomeric state within the level of resolution adopted.

Given this picture, a microscopic treatment of the folding problem requires the following elements: (a) a judicious and realistic set of rules determining the time evolution of the LTM, (b) a means of translating or projecting the LTM evolution into the BPP dynamics, and (c) a means of contrasting the BPP dynamics obtained from the adiabatic ansatz with the rigorous one obtained by translating or projecting the coarse microscopic motion given by the LTM evolution. One notices that each translation of the LTM into a BPP introduces a new set of constraints in the chain which, in turn affect the way new LTM's will be generated. This bespeaks of the existence of a feedback loop or renormalization operation concurrent with the translation operation.

The semiempirical microscopic framework described above makes it possible to define a folding algorithm based on the long-time limit of torsional dynamics. In contrast with previous algorithms which are sequential, this one is inherently parallel, since the LTM→BPP translation is essentially a pattern recognition operation, reflecting the cooperative nature of the folding process. To warrant the reproducibility of the results expounded in this work, the parallel simulation must be implemented in a personal computer (PC) operating within a conventional sequential architecture, naturally more suitable for the previous sequential algorithms [7]. This difficulty is circumvented by realizing the parallel simulation as an iterative sequence of perturbation-translation renormalization (p-t-r) loops, as detailed in Sec. IV. The perturbation operation maps one LTM into another according to the dynamic flow defined by the rules of LTM evolution over a fixed time interval. This operation is justified since, in order to translate the LTM within a conventional sequential framework, we need to freeze the LTM in order to recognize the BPP inscribed in it, whereas, in a parallel computation, the translation, being a block operation, occurs concurrently with the time evolution of the LTM and the dynamic flow gets renormalized without any need to freeze it at fixed intervals in time. Thus, ultimately, in order to perform simulations on a conventional PC, while capturing the essentially cooperative nature of the folding process, we need to resort to the p-t-r iterative framework.

Following the above discussion, the outline of this work is as follows. In Sec. II, we describe the microscopic semiempirical foundations of the adiabatic ansatz and introduce a generic means of testing the adiabatic ansatz, contrasting its results with the projection of the torsional dynamics onto the BPP space. In Sec. III, we introduce our semiempirical microscopic model or coarse version of torsional dynamics and the parallel folding algorithm based upon this simplification of conformation space. Section IV is devoted to an implementation on a conventional sequential computer of the inherently parallel algorithm rooted in the microscopic model. The implementation requires adopting the p-t-r iterative framework. Section V reveals the actual predictive power of the algorithm by computing the folding pathway and functionally competent structure of a catalytic RNA species, a ribozyme [2]. Finally, in Sec. VI, we compare the semiempirical microscopic results with those obtained using the adiabatic approximation, thus revealing the validity of the latter simplified approach within the time scales associated with detectable folding events.

II. THE ADIABATIC ANSATZ

A. Semiempirical foundations of the adiabatic ansatz

We have shown in recent work specialized for natural RNA species [2] that the amount of Shannon information $-\sigma$ measured relative to a fixed coarse description of conformation space reaches its absolute maximum within experimentally relevant time scales. This result turns the information content into a good marker of the level of organization achieved by a folding chain. However, the result is not generic, and applies to RNA sequences which are targets of natural selection. Moreover, this result has been obtained making use of an ''adiabatic ansatz,'' whereby microscopic conformations are lumped up or coarsely resolved as contact or base-pairing patterns (BPP's) treated as quasiequilibrium states [2,7], while each elementary BPP transition is regarded as a single activated processes.

This adiabatic approach treats the transition probability between two BPP's as dependent on the kinetic barrier separating the respective valleys in the energy landscape [2,4,6,7]. From a purely combinatorial viewpoint, BPP's of an RNA chain folding onto itself are drawn upon the map of Watson-Crick base-pair complementarities (A-U,G-C), which in turn, is obtained for each sequence made up of the

four units denoted *A, U, G,* and *C* [7].

Since the present work focuses on the microscopic origin of the expediency of the folding process, we do not introduce the adiabatic ansatz as an *a priori* assumption. Then, the following question must be addressed: What is the microscopic origin of the expediency of the folding process which is revealed and monitored through the time evolution of the information content with respect to the BPP partition? Obviously, the answer to this question demands that we go beyond the adiabatic approximation.

In the present work we develop a semiempirical microscopic model of folding in which the conformation soft-mode manifold X for the flexible RNA chain (the Cartesian product of as many circles as torsional degrees of freedom the chain possesses [8,9]) is coarsely resolved modulo torsional conformational isomers as the lattice $(\mathbf{Z}_2)^{3N-3}$, where $\mathbf{Z}_2$ represents the field of integers modulo 2 (made up of the two possible residue classes 0 and 1 resulting from the division of an integer by 2, with $1+1=0$), and N is the length of chain. Thus, each point in the lattice represents a complete set of torsional isomeric states (*cis* or *trans*) which coarsely represent the chain conformation. The implementation of the model hinges upon the identification of patterns of binary codified local signals, materialized through its translation into BPP's, followed by a renormalization of the problem according to the latest pattern generated. Ultimately, in order to actually elucidate the microscopic origin of the organizational expediency already revealed at the BPP level, the long-time torsional dynamics defined on the lattice $(\mathbf{Z}_2)^{3N-3}$ will be shown to yield the adiabatic dynamics when projected onto the BPP space.

A major stumbling block in the implementation of a semiempirical microscopic treatment of long-time torsional dynamics and its bearing on the folding process is due to the parallel nature of the exploration in conformation space. The concurrence of folding events taking place at the same time in different portions of the flexible chain precludes any meaningful isolation of a single ''reaction coordinate.'' To address this issue we shall first coarsely identify foldings as elements of $(\mathbf{Z}_2)^{3N-3}$, that is, as patterns of $3N$-3 locally encoded binary signals representing sets of $3N$-3 torsional states, each defined within a two-well or *cis-trans* flipping activated process.

We must first justify our coarse-graining of the manifold of soft internal degrees of freedom X upon which folding pathways are to be drawn. Our first problem becomes how to coarsely codify the local torsional states and local correlations of the flexible chain, and provide an effective dynamical picture introducing long-range correlations to account for its long-time behavior. Accordingly, to solve this problem we shall regard foldings as coarse patterns of locally encoded structural signals modeled as generated by two-state oscillators or spin flippers. Thus, we introduce a description based on a topological representation of the chain backbone. This is done by providing a binary codification of the soft-mode or torsional dynamics based on the local conformational restrictions that basically lead to a two-well or *cis-trans* flipping between torsional isomers subject to local and long-range correlations. Thus, each torsional potential basin represents a local topological state representing a local constraint. The geometry itself is immaterial within this level of description, since the latitude in the torsional potential basins (30° to 60°, [9]) yields vast conformational distortions which would make the conformations formed unrecognizable as BPP's.

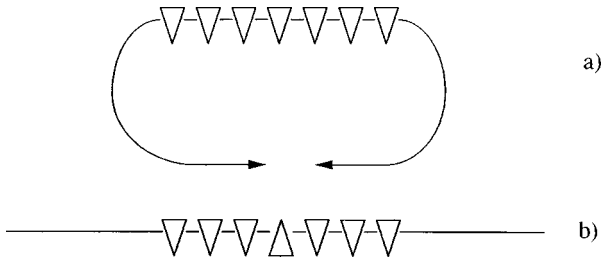

FIG. 1. Schematic representation of vertebral or row (1) consensus. When the vertebrae are correctly positioned, as in (a), loop closure becomes readily feasible, while it is precluded except for large loops $[N(\text{loop}) \gg 17]$ if vertebral consensus is not reached, as in situation (b).

Our treatment hinges upon a binary codification of local topological constraints associated to each structural motif, with each local topological constraint corresponding to a coarse local torsional state (local torsional isomer). Then, given such a representation, our treatment of the folding problem consists in identifying folding intermediates by recognizing patterns of local torsional states. As previously indicated, the latter states are defined according to a coarse description of the dynamics which identifies the local conformation state with the basin of attraction for each torsional degree of freedom, so that each basin corresponds to a *cis-trans* local torsional isomer (rotamer). Thus, the entire coarse torsional state of the chain is recorded as a time-dependent $3\times N$ binary matrix, denoted LTM, of local topological constraints. The time evolution of the LTM follows elementary kinetics for local basin transitions subject to renormalization as the folding of the chain evolves (see Sec. III). Once such kinetics are specified, and the set of local topological constraints associated with each folding motif is determined, translating an LTM into a BPP becomes straightforward.

For the sake of illustration let us consider a simple example. Suppose we want to identify or recognize a pattern within an LTM which translates into a single intrachain contact requiring the closure of a small loop of length L (L = number of unpaired bases in the loop). That means that if such a pattern arises at the time of the evaluation of the LTM, an intrachain contact will be recorded in the BPP matrix. To determine which pattern in the LTM corresponds to the intrachain contact requiring the L-loop closure, we first note (details are given in Sec. III) that each backbone unit is visualized as articulated through three torsional variables, two engaged in the orientation of the polar phosphate moiety and one acting as a vertebral torsion determining the feasibility of a local bending of the chain in the particular direction compatible with loop closure (see Fig. 1). Thus, the pattern in the LTM which translates into the contact consists of a $3\times L$ window in which the L vertebral entries [row (1)] designate all the same local rotamer needed to bend the chain in order to form the L loop, while the entries in the other two rows [(2) and (3)] designate the local torsional isomers required to orient all L phosphate groups towards the bulk solvent (as opposed to the loop interior) for better solvation.

The latter condition is needed for small loops, as shown in Sec. III. The example illustrates our computational strategy. Suppose all local topological constraints associated with a folding of the chain are fulfilled at some point in time; then a pattern within the LTM becomes recognizable and is translated into the BPP matrix where the folding pattern is recorded. Thus, the time evolution of the LTM actually represents the changes in local topological constraints, while a pattern in the LTM represents a set of constraints to be fulfilled in order to fold the chain in a particular way.

Our treatment enables us to adopt a relatively large computation time step of 512 ps, a value far larger than typical hydrodynamic drag time scales, without sacrificing accuracy within our level of description. Accordingly, the solvent can no longer be treated as the hydrodynamic drag medium, instead we incorporate its capacity for forming local conformation-dependent domains of different dielectric constants. Each evaluation of the LTM depends on the conformation-dependent local dielectric domains that the confined solvent will produce. As shown in Sec. IV, these local solvent environments determine constraints on the torsional freedom due to the orientational demands imposed on the charged phosphate groups of the RNA backbone [4].

Folding pathways are initially resolved as transitions between patterns of locally encoded structural signals which change within the 1/10 μs–100 ms time scale range. These coarse folding pathways are generated by a parallel search for structural patterns in the LTM. Each pattern is evaluated, translated and finally recorded as a BPP, an operation which is subject to a renormalization feedback loop. The renormalization operation periodically introduces long-range correlations on the LTM according to the latest BPP generated by translation. Nucleation and cooperative effects are accounted for by means of the renormalization operation which warrants the persistence of seeding patterns or kernels upon successive LTM evaluations.

In consonance with the first goal, our working strategy may be sketched as follows. (a) First, we introduce an ensemble of $3N$-3 locally correlated two-state oscillators or spin flippers to coarsely simulate torsional isomerizations, that is, a flipping between the two torsional wells for each internal degree of freedom of the RNA backbone. Then, we search for consensus regions of torsional isomers along the chain. By consensus we simply mean regions of the chain where the local topological constraints associated with the formation of a particular folding of the chain are satisfied. In this way, a consensus window emerges as a pattern of structural signals encoded locally along the sequence. The broad latitude in local torsional coordinates, or local correlation maps of the chain [9], and the vast structural distorsions it leads to implies that the binary codification cannot be implemented at the geometric level. Rather, the spin flippers are meant to mimick changes in the local topological constraints to which the flexible chain is subject in order to reach specific structural patterns. (b) We generate structural patterns as consensus regions within a matrix, the LTM, of local topological constraints (LTC's) of the chain. (c) We evaluate and translate such patterns into a contact matrix representing a BPP drawn upon the Watson-Crick map of compatible (A-U, G-C) base-pairing units along the chain evolving within the vast time scale range $512\,\mathrm{ps}$–$10^2\,\mathrm{s}$. Thus, the translation operation is actually a projection, henceforth denoted π, and becomes a pattern recognition and therefore, a parallel operation. Each pattern within an LTM emerges with a certain probability which is effectively computed as the number of evaluations of the LTM that yield the particular structural motif associated with the pattern divided by the total number of evaluations of the same LTM. (d) The translation operation is subject to a feedback loop, whereby a renormalization operation ρ readjusts the oscillator periods (or spin-flipping frequencies) according to the latest BPP translated, and the contour ranges of intrachain interactions and contour distances are renormalized relative to the latest contact pattern (CP) formed. In other words, the renormalization operation introduces long-range correlations on the LTM by slowing down or speeding up specific oscillators, depending on whether new interactions are formed or dismantled. (e) Nucleation steps and the cooperativity in the formation of secondary structure are accounted for by means of the renormalization operation. Suppose the LTM is evaluated at a given time and a short consensus window is detected. Then, the oscillators which generated this initial consensus window become endowed with frequencies which are lower than those of the neighboring residues and, consequently, the consensus region initially formed has a chance to grow upon successive evaluations of the LTM.

The fact that we characterize folding steps as BPP transitions does not imply that any ''adiabatic assumption'' has been introduced *a priori*, in the sense that no enslavement or subordination of the fast-evolving microscopic degrees of freedom to BPP transitions has been imposed. Equivalently, BPP's are not treated *a priori* as quasiequilibrium states by integrating out the relatively fast torsions as conformational entropy. Thus, the BPP is generated by a parallel search for consensus windows in the LTM. Defined in this way, an LTM represents a coarse microscopic realization of a BPP such that the consensus windows reflect the fulfillment of the LTC's determined by the BPP. This sketch of the operational tenets reveals that, although advantage is taken of the fact that there exists a wide separation between characteristic time scales associated to folding events (typically in the range 10^{-6}–10^{2} s) and chain torsions (typically in the range 10^{-11}–10^{-7} s), no adiabatic assumption is introduced.

In this way, a computational strategy is devised to provide theoretical underpinnings of the folding dynamics emerging as the evolution of patterns of locally encoded signals whose coherence reflects both cooperativity and nucleation effects. The key features of our approach are (a) the two-state coarse codification of local topological constraints of the flexible chain, (b) the renormalization of timescales for torsional isomerizations and their correlation decays relative to the successive stages of folding, thus introducing long-range correlations due to large-scale motions, (c) the identification of structural patterns with consensus regions in which specific topological constraints are fulfilled, and (d) the vast range of time scales 10^{-11}–10^{-2} s covered.

B. The adiabatic approximation in RNA folding dynamics

This section is devoted to determining the expediency of the RNA folding process by determining the time evolution of the information content associated with the exploration in

conformation space. Conformations have been resolved as BPP's, each of which is regarded as a quasiequilibrium state according to an adiabatic ansatz. In simple terms, this means that microscopic degrees of freedom are integrated out as conformational entropy. The partition of conformation space X relevant to the computation of σ is the BPP collection, a family of mutually disjoint classes. The coarse information entropy measures the spreading of the probability distribution vector $P(t)=[P_1(t),\ldots,P_M(t)]$, where $P_j(t)$, $j=1,\ldots,M$ indicates the probability that a chain is folded into the BPP j at time t, and M is the total number of *a priori* possible BPP's for a fixed RNA sequence. These probabilities should be interpreted in a Gibbsian sense, as we have a statistically large number ($\sim 10^{20}-10^{23}$ per unit volume) of replicas of our system given by actual RNA molecules which are folding onto themselves as soon as renaturation conditions are established or recovered in the environment. Thus, the information entropy associated with the folding process resolved at the BPP level is

$$\sigma(t)=-\sum_{j=1,\ldots,M} P_j(t)\ln P_j(t). \tag{1}$$

A stochastic process governs the flow of probability [4,5]. This process is determined by the activation energy barriers required to produce or dismantle interactions that stabilize the BPP's. 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.

In order to compute the probability distribution at any given time and the resulting behavior of σ, we first discretize time t by adopting an elementary time interval length u such that $t=t'u$, where t' is dimensionless and u is the shortest possible mean BPP transition time. Then, if $U=U(u)$ represents the stochastic transition matrix at the BPP level, we get

$$P(t)=[U(u)]^{t'}P(u), \quad \text{with } t=t'u, \tag{2}$$

where the matrix element $[U(u)]_{ij}$ is given by

$$[U(u)]_{ij}=[k_{ij}/\Sigma_{j'\in J(i)}k_{ij'}]xp_{ij}(u). \tag{3}$$

In this equation, k_{ij} indicates the unimolecular rate constant for the BPP transition $i\rightarrow j$, $J(i)$ is the set of BPP's accessible from i through elementary transition steps involving surmounting a single kinetic barrier (see below), the factor $[k_{ij}/\Sigma_{j'\in J(i)}k_{ij'}]$ represents the probability for the transition $i\rightarrow j$ dictated by kinetic control within a timespan of the order of $\tau_{ij}={k_{ij}}^{-1}$, and $p_{ij}(u)$ is given by

$$p_{ij}(u)=\int_0^u Y(t-\tau_{ij})dt, \tag{4}$$

with $Y(t-\tau_{ij})$ a Gaussian distribution centered at the mean time τ_{ij} for the $i\rightarrow j$ transition with temperature-dependent dispersion. The dispersion parameter will be evaluated in Sec. IV within a realistic physical context.

Explicit values of the unimolecular rate constants require an updated compilation of the thermodynamic parameters at renaturation conditions [7]. These parameters are used to generate the set of kinetic barriers associated with 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 [2,4]. 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 ($L\geqslant 100$), 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 $-\Delta H(\text{stem})$, 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 k's. For clarity we shall drop the subindexing, since we shall focus each time on a specific BPP transition. If the transition happens to be a helix decay process, we obtain

$$k=fn\exp[G_h/RT], \tag{5}$$

where n is the number of base pairs in the helix formed in the jth step, $f\approx 10^6\ \text{s}^{-1}$ is the fixed effective frequency of successful collisions [2,4] and G_h is the (negative) free energy contribution resulting from stacking of the base pairs in the helix. Thus, the essentially enthalpic term $-G_h=-\Delta H(\text{stem})$ should be regarded as the activation energy for helix disruption. On the other hand, if the transition happens to be formation of a stabilizing interaction, the inverse of the mean time for the transition will be given by

$$k=fn\exp[-\Delta G_{\text{loop}}/RT], \tag{6}$$

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

Equations (1)–(6) should be regarded as the working equations of the adiabatic approximation. This approximation will be tested in Sec. VI by comparing the adiabatic computation of the information entropy with a more rigorous computation obtained from a more detailed level of description of the long-time dynamics of the chain.

III. DISCRETE SIMULATION OF RNA TORSIONAL DYNAMICS

The vast gap between the time scales accessible to molecular dynamics computations, typically in the range 1 ps–10 ns, and those inherent in transitions between contact

pattern BPP's, typically in the range $1\ \mu s - 10^2$ s, suggests the need for a semiempirical model judiciously simplifying the soft-mode or torsional dynamics. Thus, the problem becomes how to incorporate effective internal degrees of freedom of the chain whose dynamics translates or projects onto sequences of BPP transitions. The aim of this section is to introduce a matrix where local torsional states of the chain are codified in a simplified binary fashion, so that patterns of locally encoded structural signals may be recognized and translated as BPP's.

Since conflicting possibilities may arise yielding different evaluations or pattern recognitions, a probabilistic approach appears to be necessary. A realization of this concept is introduced in this work and materializes in a semiempirical model which deals mechanistically with the rich dynamic hierarchy of time scales determined by the different levels of structural resolution of the folding process and their interplay.

Thus, the exploration of conformation space results from the parallel occurrence of trails of folding events. The base-pairing matrix (BPM) evolving in the $1/10$ ns-10^2 s time scale and built upon the map of Watson-Crick (WC) antiparallel complementarities is generated by a search for consensus windows in the LTM which records the phases of $3N$ two-state oscillators or spin flippers evolving within a mean period range estimated at 10^{-11} s-10^{-5} s. Each oscillator represents the flipping between the two potential basins (or *cis-trans* isomers) for each dihedral torsional degree of freedom. The LTM represents a coarse microscopic realization of the BPP represented by the BPM such that a consensus window reflects the fulfillment of the local topological constraints determined by the putative intrachain contact: Specific dihedral torsions must be in the ''correct state'' required for contact formation. As emphasized in Sec. I, the local geometry itself is immaterial since the latitude of the torsional potential basins (30° to 60° [9]) yields vast conformational distortions which would make the conformations formed unrecognizable if translated as BPP's.

The local constraints are themselves imposed by the conformation-dependent confined-versus-bulk solvent environments of different dielectric, and by the steric restrictions determined by loop closure. In this way, a coarse long-time torsional dynamics of the chain becomes computationally accessible. The codification will be taken to be binary since each torsional degree of freedom flips between two potential wells representing two local torsional isomers subject to local correlations and constraints. As shown in this section, the computational time step at the typical folding temperature of 303 K is 512 ps, a value far larger than the hydrodynamic time scale of 15 ps used in the continuum soft mode analysis [8,10]. Accordingly, the solvent can no longer be treated as a hydrodynamic drag medium: Its capacity to form local conformation-dependent dielectric domains must be incorporated.

To illustrate how our model works, suppose a putative contact involves WC complementary regions of the chain which flank the consensus window and requires the closure of intrachain loops which define different dielectric environments determined by the confined clusterlike versus bulk solvent. The formation of such environments imposes constraints on the dihedral torsional states of the units forming the loop which must be fulfilled if the contact is to be formed and as such, registered in the BPM: Not only the backbone two-state ''vertebrae'' (see Fig. 1) must be correctly positioned for the loop to form, but the charged phosphate groups of the RNA backbone should face the best dielectric environment available for better solvation [4]. Now let us cast this situation within our the computational context: If all spin flippers or two-state oscillators are in the ''correct'' state specified above at the time of the reading, a contact is recorded in the BPM. In this way the coarse soft-mode dynamics of the chain evolves as a sequence of pattern recognitions.

These dynamical aspects are incorporated in our computations since they determine the way in which the LTM is translated into the BPM by means of a pattern recognition. In turn, this parallel operation must be represcribed—technically renormalized—after each BPP transition, since consensus evaluation depends on the last BPP generated: Once foldings have formed, the distance between any two specific units is no longer the contour distance along the chain, and the loops which must be closed to form a contact are of different lengths relative to a those formed upon the random coil. These facts imply that the dielectric constraints are different and thus, the patterns in the LTM are recognized differently with each folding step. This sketch reveals that long-range correlations are introduced in the LTM by means of the renormalization operation.

To summarize, the approach put forth in this section may be best described as a coarse-grained analysis of soft-mode (torsional angular motion) dynamics defined by the recognition of evolving patterns of local dihedral torsional constraints consistent with a simplified topological model for the RNA backbone. Long-range intramolecular interactions are then induced by the fulfillment of local constraints to which the chain dynamics is subject due to the capacity of the solvent to determine domains of different dielectrics. The essential premise in this analysis is that the local dihedral torsions within an intrachain loop must be constrained to remain in one potential basin (*cis* or *trans*) if the concurrent contact is to be formed, so that physically, charged phosphates within the loop face the highest dielectric environment. Thus, the fact that the solvent defines conformation-dependent dielectric environments is computationally assimilated into the fact that a specified pattern of fulfilled topological constraints at the time of a reading of the LTM gets translated into a BPP.

This sketch of the operational tenets reveals that, although advantage is taken of the fact that there exists a wide separation between characteristic time scales associated with folding events (typically in the range $1/10\ \mu s - 10^2$ s) and internal backbone motions, essentially realized as dihedral torsions (typically in the mean range $10^{-11} - 10^{-5}$ s) [9–11], no ''adiabatic assumption'' subordinating or enslaving microscopic degrees of freedom to BPP transitions is introduced.

The basic representational and operational tenets of the computational design for the semiempirical microscopic model sketched above are as follows.

(a) A LTM matrix simultaneously recording the state of each ''vertebral'' and phosphate orientation dihedral. Each effective torsional state is generated by a two-state oscillator whose period is chosen from Gaussian temperature-dependent distributions which are different depending on

whether the nucleotide is free or engaged in an intrachain contact and its concurrent loops. A new period is chosen for an individual oscillator after completion of the previous period, thus incorporating thermal fluctuations. The broad latitude (up to 30° to 60° [9]) in local torsional coordinates within local correlation maps of the RNA chain, and the vast structural distorsion it leads to, implies that the binary codification cannot be implemented at the geometric level. Rather, the spin flippers or oscillators are meant to mimic changes in the local topological constraints to which the flexible chain is subject in order to reach specific structural patterns.

(b) A Watson-Crick map (WCM) of antiparallel windows upon which the BPM's are built by translation of the information encoded in the LTM.

(c) A built-in internal clock (C), incorporated in order to synchronize the LTM generation timing and its reading and subsequent evaluation at regular intervals. In accord with Shannon's information theory, the fixed beat period of C must be at most half of the shortest dihedral period.

(d) The folder (F) which evaluates the LTM and identifies consensus windows. This operation is matched with the WCM to generate BPM's and ponders the interrelationship between ''vertebral'' and phosphate-orientation consensus. In addition, F may dismantle or relax contact regions in the BPM whenever a significant consensus bubble arises within a previously-formed consensus window.

(e) The operation of F is feedbacked within a renormalization loop into the LTM generator, and the new evaluation of consensus windows is renormalized or prescribed according to the last BPM generated.

Taking into account that folding materializes in a statistical ensemble of RNA molecules and that conflicting consensus evaluations demand a probabilistic approach, we shall adopt an appropriate output representation. The statistical dynamics of consensus search is defined by the time evolution of a base-pairing probability matrix $B=B(t)$, representing the weighted overlap of different consensus evaluations.

The essential operation in our parallel algorithm, the folding operation, lies within a renormalization feedback loop and consists in the translation of information cast in terms of the state of internal microscopic degrees of freedom into a coarser representation, a BPP defined by a BPM. The latter matrix is, in turn, built upon the WCM. Thus, the algorithm and its underlying semiempirical model identify the BPP class to which an LTM belongs at certain time intervals according to a prescribed set of rules. The generation of the LTM is in turn affected by the last BPP transition that has taken place, since a new set of constraints arises with each new BPP formed, and the prescription for the evaluation operation itself is renormalized and thus depends on the last BPP which has occurred. Prior to defining the folding process explicitly within our model, we must specify the basic representational elements and their interrelationships with regard to the basic operations.

A. The LTM

This $3\times N$ matrix is a coarse representation of a microscopic realization of a BPP. Each entry adopts the value 1 or 0, representing two significant states of an RNA backbone torsion localized in a specific nucleotide (nt) or unit. The entry values are generated by two-state oscillators, one for each entry, whose period τ is automatically adjusted after one whole period has been completed from a fixed T-dependent distribution $w(\tau)$ according to renormalization specifications detailed below. Each column in the LTM represents a different nt, with the ith column ($1\leqslant i\leqslant N$) corresponding to the nt with contour value i along the chain. Each of the three rows represents a different reading space and their interrelationships are pondered each time the LTM is translated into the BPM, according to the size of the consensus window.

Each entry in the first row, denoted (1), indicates the dihedral spin state for a backbone ''vertebral'' torsion, as schematized in Fig. 1. A consensus window of consecutive spins in, say, state 1 in row (1) is a necessary condition for closure of a loop comprised of the sequence of nt's within the associated contour window. The physical interpretation of this vertebral consensus as a necessary constraint for loop closure is schematized in Fig. 1. Notice that vertebral consensus is not directly related to any geometric curvature condition for loop closure, which would make it impossible to have consensus subwindows flanked by WC complementary regions (it would be impossible to satisfy at the same time and in the same region of the chain geometric constraints for loop formation involving the whole region and those for any smaller loop involving a subregion). On the other hand, the possibility of different conflictive consensus evaluations is perfectly compatible with the probabilistic nature of our model. Depending on the size of the window, the existence of vertebral consensus at the time of a reading with flanking regions within the WCM may lead to the BPM recording of an intrachain long-range contact formation between the nt's flanking the consensus window. If $N(\text{loop})$ indicates the size of the consensus window, we may state that the necessary condition becomes sufficient if and only if $N(\text{loop})>N_c=17$ (see Refs. [4,12]), as shown below.

The entries in the second and third rows, denoted (2) and (3), of the LTM (see Fig. 2) indicate the dihedral spin states of those backbone torsions engaged in orienting the negatively charged phosphate group. Thus, for the sake of convention, spin state 1 in rows (2) and (3) for column i indicates that the phosphate of the ith nt faces bulk solvent whenever this nt is part of an intramolecular loop.

B. The interrelationship among the rows of the LTM

To specify the interrelationships between these different reading spaces in consensus evaluation, we first define an intramolecular (i,j) contact with $i<j$ as the WC base pairing engaging the two nt's with contour values i and j. The occurrence of the (i,j) contact is marked by a 1 in the ij entry of the triangular BPM. Suppose the initial BPP corresponds to the random coil, that is, there are no intramolecular contacts, and $N(\text{loop})=|j-(i+1)|>17$ (see Ref. [12]). Then, once a reading of the LTM takes place, an (i,j) contact will be produced after evaluation and recorded as such in the new BPM if all dihedral ''vertebral'' spin states for the segment of row (1) flanked by entries i and j are in the correct torsional conformation for folding, that is, they are in state 1 (see Fig. 1). The fact that the range of the putative (i,j) interaction must be larger than 17 to materialize with vertebral consen-

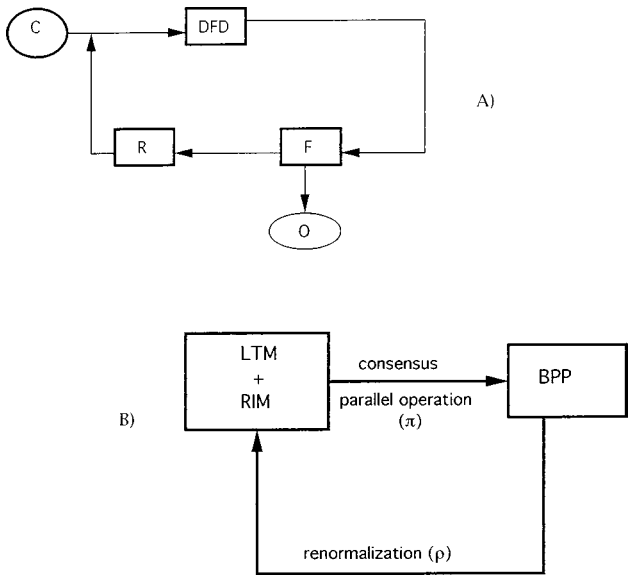

FIG. 2. (a) General scheme of the folding machine featuring its basic components: The inherent clock (*C*), the dihedral frequency distributor (DFD), the folder (*F*), the output displayer (*O*), and the renormalizer (*R*). (b) General scheme of the interrelation between the different representational elements: The dihedral spin state matrix DSSM, the reading instructions manual (RIM), and the base pairing matrix (BPM).

sus *only* can be justified as follows. Loop closure defines two solvent domains, an inner low-dielectric domain confined by the loop of rodlike dimensions and an outer high-dielectric or bulklike domain. If the loop is sufficiently large, encompassing more than 17 nt's, as demonstrated in Ref. [12], the dielectric difference between the inner and outer domain becomes negligible: Each charged phosphate group pointing to the inside of the loop admits four water-solvation layers for a loop of size 17 or larger. However, for smaller loops the consensus demands are higher and vertebral consensus is no longer sufficient: The phosphate groups must be oriented towards the bulk for better solvation. This argument leads us to qualitatively distinguish rows (2) and (3) from row (1): As more orientational constraints are associated to loop closure, more consensus is needed for it to materialize, so that for a putative loop smaller or equal to 17, that is, $|j-(i+1)| \leqslant 17$, consensus windows in rows (1), (2), and (3), flanked by columns i and j becomes the necessary and sufficient condition.

C. The dihedral frequency distributor (DFD)

In attempting to project dihedral torsions into the BPP space, where the folding process is conventionally recorded, the time scale limitations of molecular dynamics simulations must be circumvented. This explains the need to introduce models, such as the one presented in this work, in which activated molecular motions in the ns-to-μs range are considered [9–11]. Thus, faster diffusional-like unhindered torsions, such as the torsion around the glycosidic sugar-base bond in an unpaired nt are integrated out as conformational entropy of the state defined by the LTM representation. Such motions, well into the ps range, determine the rodlike shape of the RNA molecule when viewed within the time scale window between two LTM states. Precisely this interrelation between shape and time scale justifies the concept of inner and outer solvent domain defined by an intramolecular loop, as put forth in our consensus analysis of the entropic cost of loop closure. Thus, the microscopic mean time range relevant to LTM transitions is 1 ns–1 μs, covering the time scale for internal motions (1 ns–10 ns), of the order of the calculated diffusional displacements of flexible hinged domains [10], and, at the other end of the spectrum, the limiting value (1/10) to 1 μs for a localized helix-unwinding event leading to a bubble within a helix [11].

These considerations lead us to define a temperature-dependent normalized distribution of periods $w=w(\tau)$. In particular, the periods of unhindered dihedral oscillators are assigned from this distribution in such a way that the effect of thermal fluctuations on the formation of consensus and thus on structural transitions is incorporated. The distribution has three Gaussian peaks, each with dispersion $\sigma^2=gT$, where the constant g depends on the actual denaturation temperature T(denat) and on the consensus interpretation of denaturation, as shown below. The peaks occur, respectively, at mean periods 10 ps, 10 ns, and 1 μs. This distribution allows us to classify nt's in two classes: To class I belong all nt's with mean dihedral period 10 ps, while class II contains all nt's whose mean dihedral period is either 10 ns or 1 μs. The first class corresponds to internal dihedral torsions of the RNA chain of the type probed by fluorescence depolarization [10]. These torsions occur in free nt's, that is, unpaired nt's not belonging to a loop. Accordingly, the DFD in the folding machine establishes a lottery from which periods of dihedral spin oscillators for free nt's are assigned from within the period range centered at 10 ps. A new period is assigned from the lottery to each oscillator each time a whole previously assigned period has been completed. The frequency $f=1/\tau$ of a dihedral spin in any row of the DSSM corresponding to a nt not engaged in an intrachain interaction or loop satisfies the inequality

$$|f^{-1}-10\ \text{ps}| \leqslant |\tau'-10\ \text{ps}|,\ \text{with}\ \tau'\ \text{satisfying}$$

$$w(\tau')=\text{infimum}_{\tau}\{w(\tau) \geqslant 1/[3N-3]\}. \qquad (7)$$

The condition yielding the extreme period τ' arises from the fact that there are at most $3N-3$ free oscillators in the chain. At the typical folding temperature $T=303\,\text{K}$, we get $\tau' \approx 1.02\,\text{ps}$.

The other two peaks in the distribution correspond, respectively, to mean periods for nt's engaged in an *A*-*U* or *G*-*C* Watson-Crick base pair within a helix, or to nt's within loops. In the latter case the period range covers the entire bimodal distribution. Again, the same considerations apply with regard to the period assignment to oscillators for class II nt's. These rules imply that the nt's in loops concurrently formed with an intrachain helix adopt the same cadence as the helix nt's themselves. This is so, since the rate of helix dismantling is exclusively dependent on the size of the helix taken by itself [4,13], and determined by the formation of a significant consensus bubble among the class II nt's engaged in the helix. Furthermore, this local limiting event is fairly independent of concurrent microscopic events in the associated loops.

The mechanistic aspects of period distribution, as performed by the DFD, imply that this operation is subject to

renormalization with each BPP transition: A BPP determines which columns in the LTM correspond to free or class I nt's in the chain and which correspond to nt's engaged in an intrachain interaction, or, equivalently, they belong to class II. Thus, since a BPP transition reclassifies the nt's, it also dictates the range from which new period assignments are drawn. The period range for a specific nt remains the same as before the BPP transition if the transition does not alter the class of the nt, and changes if the BPP transition transfers the nt to a different class.

D. The built-in internal clock (*C*)

In order to satisfy the basic tenets of Shannon's information transmission while making all operations *C* synchronized, the intrinsic beat period of *C* should be taken to represent $1/2\tau'$, one half of the shortest possible period to be assigned to a dihedral spin oscillator. Since τ' depends on $w(\tau)$, it is itself *T* dependent. The time interval between two consecutive readings of the LTM τ(read) is taken as constant and, in order to lower the operational cost, it is fixed at the shortest time that a BPP transition could possibly take. Thus, $\tau(\text{read}) = 2^{3\times 3}\tau' \approx 512$ ps the shortest time to form the three loop, the smallest possible loop, engaging the fastest oscillators.

E. The folder (*F*)

The actual translation of the consensus evaluation of the LTM into the BPM is performed by the folder (*F*), with the aid of an updated version of the reading instructions manual (RIM). An updating takes place with each BPP transition marked by a change in the BPM. The folder might form or dismantle (relax) several intrachain helices in parallel and records in the BPM the consensus evaluation performed with the aid of the RIM, as indicated in the schemes displayed in Fig. 2. The consensus search or evaluation under unconstrained conditions, that is, with a RIM defined by the random coil BPP, has been partially delineated in Sec. III A for helix formation. The flow chart for this parallel operation is displayed in Fig. 3.

On the other hand, helix dismantling materializes and is recorded as such by deletion in the BPM whenever a consensus bubble forms among class II nt's engaged in base pairing. By ''consensus bubble'' we mean that in any of the three rows, a consecutive sequence within the set of dihedral spins of helix nt's of length 30% of the total helix length must be out of phase with the consensus value 1. In other words, the sequence within the helical region must adopt the state 0 at the time when the reading of the LTM takes place. Because of the renormalization loop, this transition at the BPM level immediately transfers a new set of constraints for the generation of new LTM's: The nt's previously engaged in the helix and in the concurrent loops are reclassified, being transferred from class II to the higher frequency class I.

Stacking effects [14] reflect themselves mechanistically in the formation of the consensus bubble: The larger the helix, the greater the improbability of finding a 30% out-of-phase subsequence of oscillators from class II. Furthermore, these considerations enable us to estimate the constant *g* which determines the effect of thermal fluctuations on the period distribution: At the denaturation temperature *T*(denat), every

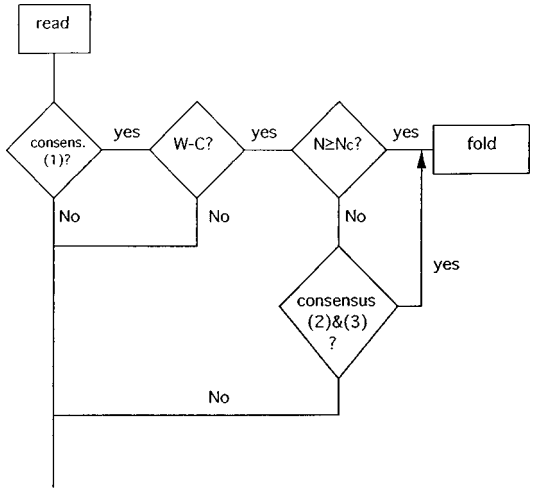

FIG. 3. The flow chart of the folding operation as a consensus evaluation after reading the three rows (1), (2), and (3) of the DSSM. The following notation has been adopted: ''consens. (1)?'' and ''consensus (2)&(3)?'' refer, respectively, to finding a consensus region on row (1) or finding it on rows (2) and (3) on the DSSM; ''W-C?'' refers to finding Watson-Crick complementarity in the regions flanking the consensus region and ''$N \geqslant N_c$?'' refers to deciding whether the size *N* of the consensus region is larger than the critical size N_c.

helix formed must develop a consensus bubble evaluated and recorded with the next reading of the LTM. Thus, if σ is ''large enough,'' the period distribution in the helix is broad enough that consensus cannot be preserved: The period range, of the order of σ, is such that a helix consensus cannot survive two consecutive readings. From these considerations, and taking into account our empirical estimate of the denaturation dispersion fixed at $\sigma = 0.6\,\mu$s, and the typical experimental $T(\text{denat}) = 312$ K for ribozymes such as the ones studied in this work [14], we get $g \approx 1.2\times 10^{-15}$ s^2/K.

F. The renormalizer (*R*)

The renormalizer has two simultaneous roles: It updates the RIM by determining contour distances *relative* to the latest BPP generated and, by readjusting frequencies, it places a new set of constraints on the generation of new LTM's based on the latest BPP translated. Thus, renormalization accounts for long-range correlations that develop on the LTM as a consequence of folding events recorded as BPP transitions.

Thus, we shall focus now on the RIM renormalization. The essential aspect of this operation is how to search for consensus in a chain folded already in a specific BPM. Three generic situations arise.

(a) A putative (*i*,*j*) consensus region contains inside previously-formed intrachain (*k*,*p*) contact, in such a way that $i<k-n<k<p<p+n<j$, where *n* is the length of the preexisting (*k*,*p*) helix. Then, if the number $\|j-i\|_{\text{BPM}}$ = contour distance between nt's *i* and *j* relative to the BPM determining the last RIM updating $=|k-n-(i+1)|+|j-(p+n+1)| \leqslant 17$, consensus must be searched for along rows (1), (2), and (3) for the regions defined by columns $i+1$ to $k-n-1$ and $p+n+1$ to $j-1$ of the LTM. On the other hand, if $\|j-i\|_{\text{BPM}} > 17$, then consensus should be

searched for only in the entries $i+1$ to $k-n-1$ and $p+n+1$ to $j-1$ of row (1), since no phosphate orientation is required to form the new loop.

(b) A putative (i,j) consensus region is contained within a preexisting loop determined by the (h,k) contact, so that $h<i-n<i<j<j+n<k$, with n being the length of the putative helix to be formed. If $\|k-h\|_{\mathrm{BPM}(1)} \leqslant 17$, with CP(1) representing the contact pattern that determined the previous to the last RIM updating, then the (i,j) contact materialization is only contingent upon Watson-Crick complementarity in the regions $[i-n,i]$ and $[j,j+n]$, and no further consensus evaluation is required. On the other hand, if $\|k-h\|_{\mathrm{BPM}(1)} > 17$, then two common regions must be evaluated separately, the $(i+1)$ to $(j-1)$ region in all three rows if $|j-(i+1)| \leqslant 17$ and only in row (1) if $|j-(i+1)|>17$ and the $(h+1)$ to $(i-n-1)$ and $(j+n+1)$ to $(k-1)$ region, again in all three rows if $|i-n-(h+1)|+|k-(j+n+1)| \leqslant 17$ and only in row (1) if $|i-n-(h+1)|+|k-(j+n+1)|<17$.

(c) Pseudoknot formation [14]: The putative contact (i,k) lies in a pseudoknotted position with respect to the preexisting (h,j) contact, so that $h-r<h<i-n<i<j<j+r<k<k+n$, with $r=$ length of the preexisting (h,j) helix, and $n=$ length of the putative (i,k) helix. We shall assume first that $\|j-h\|_{\mathrm{BPM}(1)} \leqslant 17$, then two situations may arise depending on whether $|k-(i+1)| \leqslant 17$ or >17. In the first case, the (h,j) and (i,k) loops cannot be coplanar since the phosphate orientation towards bulk solvent in the common $(i+1)$ to $(j-1)$ contour region concurrent with formation of the first loop precludes the new loop from turning the solvation environment of the common phosphates into a poor low-dielectric medium. On the other hand, if $|k-(i+1)|>17$, the two loops may be coplanar because the dielectric medium of the common phosphates is not significantly altered upon formation of the new loop [9]. In both cases, the region common to both loops, from columns $i+1$ to $j+r$ is excluded from consensus search involved in the formation of the (i,k) contact, and only the columns $j+r+1$ up to $k-1$ of the LTM are involved with rows (1) to (3) if $|k-(i+1)| \leqslant 17$ and row (1) only if $|k-(i+1)|>17$. An illustration of the ''most difficult'' case $\|j-h\|_{\mathrm{BPM}(1)} \leqslant 17$, $|k-(i+1)| \leqslant 17$, is shown in Fig. 4, and arises in the formation of the conserved $P3$-$P7$ pseudoknot, the crucial motif in RNA catalysis for group I ribozymes [2,14].

Suppose now that $\|j-h\|_{\mathrm{BPM}(1)}>17$, then again two situations arise. If $|k-(i+1)| \leqslant 17$, the activation of the (i,k) contact demands that consensus be searched for in all three rows for columns $j+r+1$ up to k and in rows (2) and (3) of the common region from columns $i+1$ to $j-1$. This peculiarity in consensus search is due to the fact that phosphates common to both loops have not been oriented upon formation of the first loop but must be oriented if the new (i,k) interaction is to materialize. On the other hand, if $|k-(i+1)|>17$, no phosphate orientational demands are imposed at all on the DSSM, and only row (1) consensus in the region flanked by columns $j+r$ and k is involved.

IV. DISCRETIZED TORSIONAL DYNAMICS ON A PERSONAL COMPUTER

The aim of this section is to actually prescribe the implementation of the translation-renormalization $(\pi$-$\rho)$ loop operations on a personal computer (PC), thus introducing an algorithm for RNA structure prediction rooted in a semiempirical computation of torsional dynamics. A suitable PC is one roughly matching the following specifications: RAM: 256 MB; processor speed: 400 MHz; hard disk memory: 6 GB.

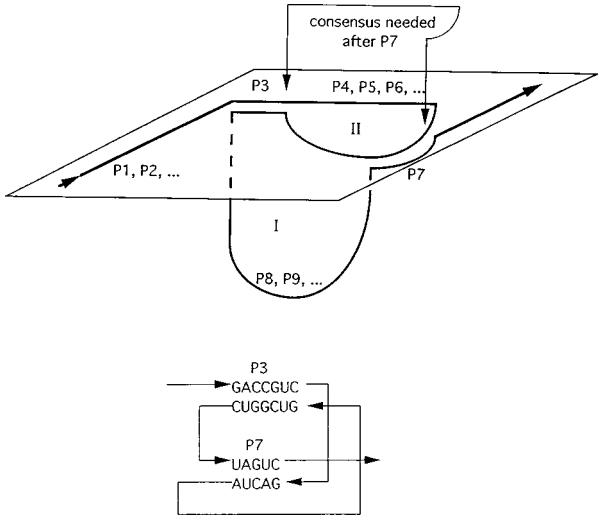

FIG. 4. Scheme of the pseudoknot shaping the catalytic core of the Tt.LSU ribozyme. This pseudoknot engages the conserved regions $P3$ and $P7$. The first helix to form is $P7$ as predicted by our computations, and it facilitates the formation of $P3$ by drastically reducing its consensus demands. That is, $P3$ may only form within realistic time scales in a cooperative fashion and within the pseudoknot.

In its original version described in Sec. III, the computation performed by $(\pi$-$\rho)$ loop iteration constitutes a parallel synchronous algorithm designed to predict the active structure of biomolecules reached within times incommensurably shorter than those required for thermodynamic equilibration. Such computations require a parallel evaluation of concurrent folding possibilities at regular intervals without any effective quenching of the folding dynamics. Each reading, in turn, defines the set of constraints to which the system is subject when undertaking the next folding stage.

A sequential computation on a conventional PC implies that we would need to quench the LTM every 512 ps, as indicated in Fig. 5, to sequentially perform a translation or pattern recognition. In turn, the recorded pattern determines via renormalization the starting conditions for the generation of the new sequence of LTM's to be read and evaluated only after another 512 ps have elapsed. As we recall, the LTM's evolve in time by having each two-state entry pick frequency (or period) from a Gaussian distribution determined by renormalization at every reading time step of 512 ps from a set of three distributions: one for free nt's one for nt's engaged in secondary structure and one for nt's engaged in tertiary interactions (the pseudoknot motif). Thus, renormalization assigns a distribution to each state flipper for each entry of the LTM, and the sequential computer mimics the dynamics resulting from the fact that a new flipper frequency is chosen from the same distribution after each period has been completed.

We may sequentially realize the $(\pi$-$\rho)$ loop as a perturbation-translation-renormalization cycle by quenching

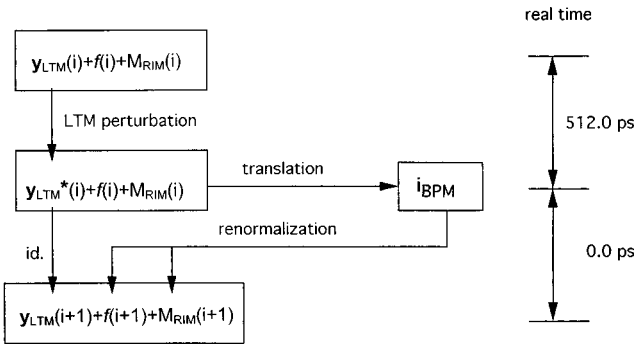

FIG. 5. The generic chart of operations for the PC realization of the ith $(\pi\text{-}\rho)$ loop iteration. The LTM perturbation requires one sequential computation for each LTM entry following the instructions given in Sec. IV. Each renormalization operation takes place while the LTM is quenched in the same state, LTM*, in which it was translated. This operation defines the vector $f(i+1)$ of mean frequencies, and the reading instruction manual $M_{\mathrm{RIM}}(i+1)$, to be used in the $(i+1)$th $(\pi\text{-}\rho)$ loop iteration.

the LTM dynamics during the last two operations, as shown in Fig. 5. In this figure, the ith iteration is displayed as preparing the LTM for the $(i+1)$th iteration. The LTM has been added two modules: a frequency N-row vector (f) and a reading instruction manual. Each entry in the frequency row may take one of three values $+1$, 0, -1, according to whether the mean frequency assigned by renormalization to the particular unit or residue is 10^{11}, 10^{8}, or 10^{6} Hz. The RIM determines the renormalized distances to be considered when searching for consensus in regions containing already windows of secondary of tertiary structure. Following the RIM, if a 0- or (-1)-frequency window is found within a window where consensus is being searched, it must excluded from the window being examined.

In order to inductively define the ith $(\pi\text{-}\rho)$ iteration, as shown in Fig. 5, we shall consider given all three modules $y_{\mathrm{LTM}}(i)$, $f(i)$ and $M_{\mathrm{RIM}}(i)$, where i refers to the ith iteration. The $y_{\mathrm{LTM}}(i)$ must be perturbed during 512 ps to obtain the $y^*_{\mathrm{LTM}}(i)$ at the end of the time step. This corresponds in real time to a 512 ps progress of the torsional dynamics since the $y^*_{\mathrm{LTM}}(i-1)=y_{\mathrm{LTM}}(i)$ has been translated. The $y^*_{\mathrm{LTM}}(i)$ is then ready to be translated into the i_{BPM}. The perturbation $y^*_{\mathrm{LTM}}(i)$ is determined exclusively by the $y_{\mathrm{LTM}}(i)$ together with the mean frequencies encoded in the vector $f(i)$. Thus, the perturbation actually corresponds to the running of the LTM dynamics for 512 ps, until the LTM is quenched again for pattern recognition. The pattern recognition or translation, in turn, defines the next vector $f(i+1)$ of mean frequencies and the next reading instructions manual $M_{\mathrm{RIM}}(i+1)$, both entities being produced by the ith renormalization operation. This information, together with the previously perturbed $y^*_{\mathrm{LTM}}(i)=y_{\mathrm{LTM}}(i+1)$ is all that is needed to now run the $(i+1)$th $(\pi\text{-}\rho)$ interation, which would involve first perturbing the $y_{\mathrm{LTM}}(i+1)$, then translating the $y^*_{\mathrm{LTM}}(i+1)$ and finally reassigning according to $i_{\mathrm{BPM}}(i+1)$, mean frequencies and the new RIM. The inductive definition of the sequential algorithm is now complete.

In practice, an actual bottleneck in the sequential computation is the pattern recognition or translation on the LTM, an inherently parallel operation. If N is not too long ($N \approx 100-350$) this operation may be accessible to a sequential machine engaged in column-by-column reading with concurrent memory storage. A state-of-the-art PC as specified above takes approximately 0.71 ms of real time to sequentially translate each LTM into a BPM with $N=300$, using the appropriate flow chart from Fig. 3 to detect and store nucleation folding events.

The perturbation LTM* of an LTM is the other aspect to be considered in a sequential computation. In order to determine the state of each entry in the LTM once a time step of 512 ps has been completed, one must consider the following scheme.

(1) Classify units according to whether their ith mean frequency value is $+1$, 0, or -1. If unit j has been previously assigned the 0 or -1 frequency value, store its actual frequency $F(j)$ chosen from the respective Gaussian distribution and record the number $Q_i(j)$ of $(\pi\text{-}\rho)$ loops that have taken place since the 0 or -1 values have been originally assigned until the ith pattern recognition is performed. Such units are given a different treatment, following instruction item (5). For the remaining or free units, instructions (2)–(4) are to be followed.

(2) Record the state $(+$ or $-)$ of each $y^*_{\mathrm{LTM}}(i-1)=y_{\mathrm{LTM}}(i)$ entry with mean frequency value $+1$, quenched throughout the $(i-1)$ translation or pattern recognition.

(3) For each unit or column j during the ith iteration, consider the Gaussian distribution peaked at $f_j(i)=10^{11}$ Hz. Using this distribution, generate by means of a Monte Carlo simulation as many frequencies f_k, $k=1,\ldots,n^*$ as necessary, so that the following inequalities hold:

$$\sum_{k=1,\ldots,n^*-1}(2\pi/f_k)<512\ \text{ps}\leqslant\sum_{k=1,\ldots,n^*}(2\pi/f_k). \quad (8)$$

(4) Determine the state of the entry at the time of the next reading as follows. It changes provided

$$3/2\pi\geqslant\left[512\ \text{ps}-\sum_{k=1,\ldots,n^*-1}(2\pi/f_k)\right]f_{n^*}\geqslant\pi/2. \quad (9a)$$

It remains the same as in the $(i-1)$th translation if

$$\left[512\ \text{ps}-\sum_{k=1,\ldots,n^*-1}(2\pi/f_k)\right]f_{n^*}<\pi/2\ \text{ or}\ >3/2\pi. \quad (9b)$$

(5) Suppose entry j has been assigned frequency value 0 or -1 a number $Q_i(j)$ of $(\pi\text{-}\rho)$ loops before the ith loop. Then, since $2\pi/F(j)>512$ ps, its state to be read in the ith pattern recognition remains the same if $[Q_i(j)512\,\text{ps}]F(j)<\pi/2$ or $>3/2\pi$, and changes otherwise. If a consensus bubble in a secondary or tertiary structure is formed by having 30% of out-of-phase spin flippers at the time of the ith translation, then the entire consensus units are reassigned the $+1$ frequency value by the ith renormalization operation.

In this way we define the perturbed matrix $y^*_{\mathrm{LTM}}(i)$, which is quenched during the ith translation-renormalization, as indicated in Fig. 5. This is precisely the matrix fed into the $(i+1)$ iteration, that is, $y^*_{\mathrm{LTM}}(i)=y_{\mathrm{LTM}}(i+1)$, which, in

turn, must be perturbed according to the reassignment $f(i+1)$ of frequencies which took place in the ith iteration.

The lottery for sequential frequency assignment to each chain unit leading to the perturbation of the LTM according to Eqs. (8) and (9), may be implemented in a Monte Carlo simulation. In real time, an LTM perturbation following instructions (1)–(5) takes 0.60 ms maximum. This upper bound is found in the case where all residues are free, and therefore, their frequencies are to be chosen from the distribution with the highest mean frequency. Thus, the most conservative estimation of the total PC computation time involved in the estimated $10^7(\pi\text{-}\rho)$ loop iterations, as required to satisfactorily penetrate relevant folding times, is 1.31×10^4 s.

V. RESULTS

The consensus evaluation on the LTM and subsequent generation of the BPM onto the Watson-Crick map are in essence parallel operations which may be mimicked on sequential computer routines by freezing the LTM dynamics according to the tenets expounded in Sec. IV. The output of the computation is given as a time-dependent base-pairing probability matrix $B=B(t)$ in the form of a triangular matrix whose ij entry $(i<j)$ is shaded to the extent of the probability of an (i,j) contact [see Figs. 6(a)–6(c)]. The $B=B(t)$ is regarded as an overlap of BPM's, the $M(i)$'s. Each $M(i)$ is weighted according to the number of LTM's created from the same distribution $w(\tau)$ whose consensus evaluation yields $M(i)$ at time t:

$$B(t)=\sum_i p(i,t)M(i), \quad p(i,t)=Z(i,t)/Z(t), \qquad (10)$$

where $Z(i,t)$ denotes the number of LTM's translated into the BPM $M(i)$ at time t, and $Z(t)$ is the total number of LTM's generated at time t from the period distribution $w(\tau)$. Thus, the time evolution of B is actually the relevant output we shall focus on not only for structure-prediction purposes but also to test the adiabatic approximation expounded in Sec. II. It provides the proper statistics on the ensemble of coexisting structures and its time development throughout the folding process within the time allotted by biological constraints, allowing us to compare the final or destination ensemble of BPM's with the phylogenetically inferred structural information [14,15].

In order to realistically simulate the folding process in accordance with the basic mechanistic tenets described in Sec. IV, we need to incorporate effectively the role of the divalent ion Mg II, known to expedite the folding process by significantly lowering the entropic cost of folding steps [2,14]. It is known that Mg II binds weakly to adjacent phosphates of unpaired nt's by forming a chelate complex [14]. This fact can be interpreted within the context of the orientational constraints imposed on loop formation by considering two adjacent coordinated phosphates as a single orientational rigid entity. In consonance with these facts, we may infer the effect of the presence of Mg II in solution upon the generation of consensus in the LTM: The Mg II ion couples the two dihedral spins of the same type engaged in the orientation of the two adjacent phosphates that serve as ligands

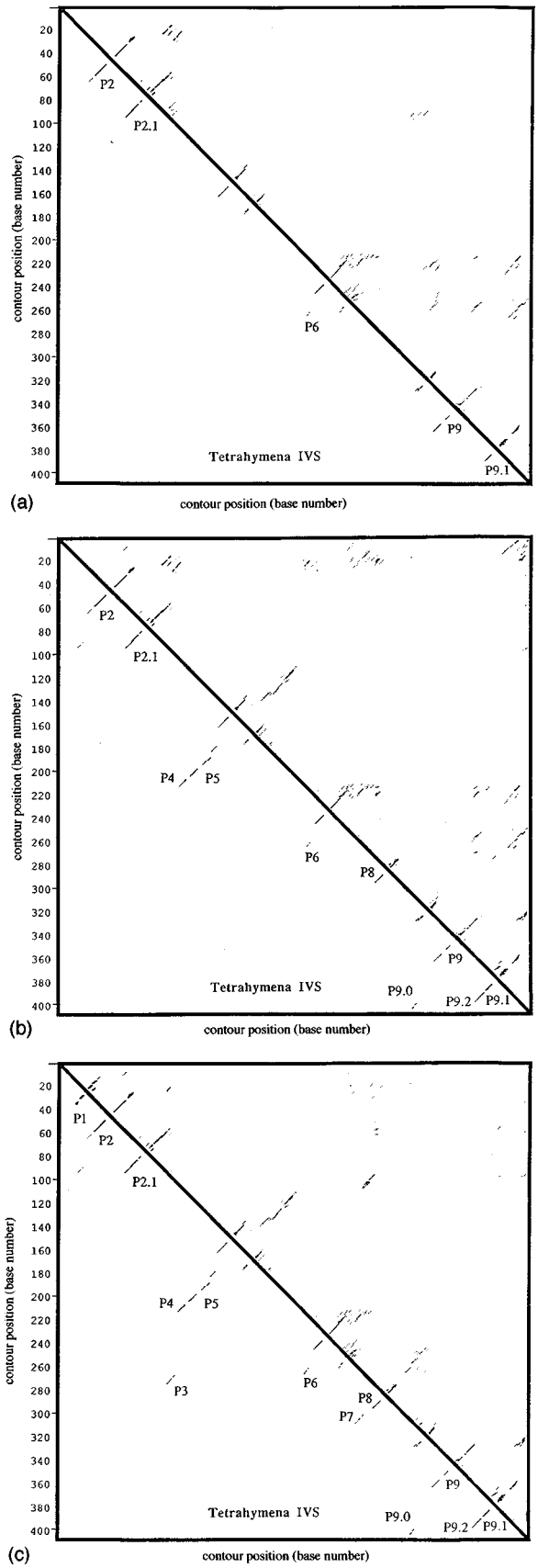

FIG. 6. Three snapshots of the base pairing probability matrix $B=B(t)$ taken, respectively, at (a) 100 μs, (b) 100 ms, and (c) 30 s. The upper right triangular matrix represents the base pair probabilities, while the phylogenetically conserved helices are indicated in the lower left triangular matrix.

for the divalent ion. The situation may be illustrated with the formation of a three loop: In the presence of the Mg II ion, consensus demands for loop formation engage $3+4=7$ spins, three from row (1) and 4 from rows (2) and (3) of the LTM, as opposed to nine engaged in the absence of Mg II. Thus, the time scale for achieving consensus is considerably shortened. The Mg-coupled entity is actually oriented by two single dihedral spins whose cadence is regulated periodically from the same distribution from which all dihedrals are periodically readjusted.

A scanning of the time-dependent evolution of $B=B(t)$ is displayed in Figs. 6(a)–6(c) for the intervening sequence (IVS) of tetrahymena, the basic study case for RNA catalysts which in standard notation is referred to as Tt.LSU, one of the most studied self-splicing introns of group I [14,15]. Figures 6(a)–6(c) display three snapshots of $B(t)$ taken, respectively, at 100 μs, 100 ms, and 30 s. As indicated before, the ij entry in the upper right triangular matrix ($i<j$) is shaded according to the probability of the (i,j) contact. The lower left triangular matrix is used to record only those interactions denoted $P1,P2,...,P10$, which are phylogenetically conserved as revealed by the sequence homologies found in all eighty seven introns of the group I family [14,15]. A significant feature emerging from the three snapshots of the time-evolving $B=B(t)$ is the fact that the ensemble of coexisting structures becomes more and more sharply peaked at a representative structure endowed with all conserved helical stems. This trend becomes apparent within the actual time frame alloted in biological contexts [2].

We shall focus now on the shaping of the $P3$-$P7$ pseudoknot (Fig. 4), the actual bottleneck in the kinetics of formation of the catalytic core [2,6,16]. After all peripheral conserved stems $P4$-$P6$ and $P8$ have formed within the 100 ms time scale, the formation of $P3$ is cooperatively expedited by the previous formation of $P7$ and complex loop I. This is so because of the concurrent orientation of the common loop region which facilitates the closure of loop II by decreasing its consensus demands, as indicated in Fig. 4: Within the preexisting BPM which has remained unaltered since the 100 ms snapshot, $P3$ formation requires the closure of a loop of size 17, a prohibitively difficult task even in the presence of Mg II, since the time scale associated with that single event would be 2^{35} ns, exceeding the overall experimentally probed folding time scale of 30 s [2,9]. Actually, $P3$ must wait until $P7$ has formed in 24 s and thereafter it may form cooperatively within the $P3$-$P7$ pseudoknot. Thus, $P7$ formation is the nucleating event lowering the consensus region of $P3$ from seventeen to seven unpaired nt's. Given the renormalized distance $\|j-i\|_{P7}=7$ between nt's i and j after $P7$ has formed, we infer that $P3$ forms 2^{7+8} ns $\approx 3.2\,\mu$s after $P7$ has formed (in the absence of Mg II this step would take 2^{21} ns). Thus, at least resolved up to the BPM level, the catalytic core is predicted to be shaped cooperatively and sequentially within experimental time scales, with $P7$ formation being the rate-determining step. Furthermore, the role of Mg II has been elucidated: Although it hastens all folding steps, it becomes particularly crucial in expediting the rate-determining step which would not materialize within relevant time scales in the absence of the divalent ion. These results help clarifying and are fully supported by recent kinetic experiments [6].

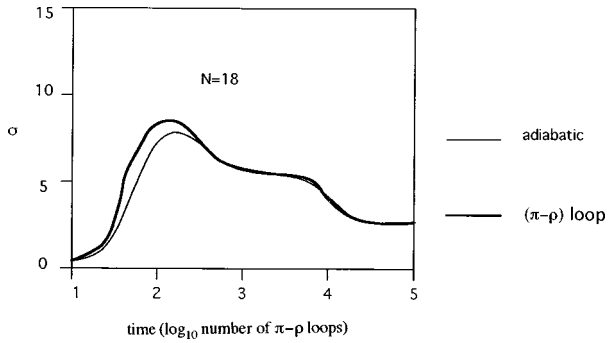

FIG. 7. Time dependence of the coarse information entropies $\sigma(t)$ and $\sigma_{(\pi\text{-}\rho)}(t)$, both relative to the BPP partition of conformation space for a specific randomly generated sequence of length $N=18$. The abscissas indicate time in No. $(\pi\text{-}\rho)$ loops units. Each LTM $(\pi\text{-}\rho)$ evaluation takes place at $\tau(\text{read})\approx 512$ ps.

The behavior of all 87 group I introns [15] is entirely analogous to the representative study case presented in this section. The following essential features appear to be common to the entire family: (a) convergence to a sharply peaked ensemble whose dominant folding presents all phylogenetically conserved structural elements [see Fig. 6(c)] and (b) the Mg-aided cooperativity leading sequentially to the shaping of the catalytic core with $P7$ formation acting as the nucleating event inducing $P3$ formation.

VI. THE VALIDITY OF THE ADIABATIC APPROXIMATION

The aim of this section is to determine the time range of validity of the adiabatic ansatz which has been described in Sec. II and primarily used in most structure-predictive algorithms [2,4,7,12–16]. In estimating the time range where BPP's may be treated as quasiequilibrium states we face the following problem: How do we compare the data generated by the semiempirical computation of discretized torsional dynamics with the results of the adiabatic approximation? The information-theoretic approach followed in Sec. II may only be adopted at the LTM level if we determine the expected entropy as

$$\sigma_{(\pi\text{-}\rho)}(t)=\langle S\rangle(t)=-\langle \ln p\rangle(t)$$
$$=-\sum_i p(i,t)\ln p(i,t), \quad p(i,t)=Z(i,t)/Z(t), \tag{11}$$

using the $(\pi\text{-}\rho)$ loop computation [see Eq. (10)] and contrast it vis-a-vis the adiabatic computation obtained using working Eqs. (1)–(6). The wanton complexity of the adiabatic transitional matrix drastically limits our computational capability and thus a limit chain size $N=18$ has been adopted. This is due to the exhorbitant cost in computing time required to iteratively multiply a matrix of order $M\times M$, with $M\sim e^N$. The results of both computations are given in Fig. 7 for an average over an ensemble sample of 24 randomly generated sequences of length $N=18$.

As direct inspection of Fig. 7 reveals, incipient helices formed in structure-nucleation events in the range

$5\times10^{-9}-5\times10^{-7}$ s are easily dismantled (a consensus bubble is more easily formed than in fully developed structures). This fragility of incipient structures causes the large pattern fluctuations marked by a large σ in the time range $5\times10^{-9}-5\times10^{-7}$ s, or $10-10^{3}(\pi\text{-}\rho)$ loops. A large plateau starting at 5×10^{-7} s marks the formation of a relatively stable kinetic intermediate which contains all structural motifs which may form *noncooperatively* [16]. That is, those motifs whose associated N(loop) lies within the favorable ranges of low conformational entropy cost: $3\leqslant N(\text{loop})\leqslant10$ [2,12,16]. At 10^{-6} s, cooperative events lead to other helices whose loops have favorable *renormalized* sizes, while their sizes relative to the random coil are unfavorable. On the other hand, the increase in class II nt's beyond the formation of the kinetic intermediate stabilizes the patterns determining the survival of the oscillator phase in consecutive LTM evaluations. This determines the relatively low fluctuations beyond 5×10^{-7} s.

For short time scales, $10^{-9}-10^{-7}$ s, fast-evolving internal degrees of freedom simulated as torsional oscillators are not yet *enslaved* or entrained by BPP transitions which evolve within typical time scales $5\times10^{-7}-10$ s for the chain length $N=18$. For this reason, within the range $10^{-9}-10^{-7}$ s, the level of exploration of conformation space due to uncorrelated or short-range correlated torsional excitations must be vastly larger than that resulting from an adiabatic process. However, as soon as the stabilized kinetic intermediate is formed [16], the long-range correlations coupling distant-row oscillators in the LTM, begin to develop, as cooperative effects occur upon short-range nucleating interactions. These long-range correlations are, in turn, induced by BPP transitions, in consonance with the nature of the renormalization operation. Thus, initial structure-nucleating steps involving uncorrelated or locally correlated motions do not demand as much enslavement of fast-evolving torsions as cooperative events, which entail long-range correlations. For this reason, we expect the adiabatic approximation to fit the rigorous results as soon as long-range correlations governed by BPP transitions occur. This is indeed what takes place, as the almost perfect coincidence of both the σ and $\sigma_{(\pi\text{-}\rho)}(t)$ plots beyond the 100 ns reveal (Fig. 7). This coincidence of the two independent computations validates the algorithmic dynamics of the $(\pi\text{-}\rho)$ loop, as it confirms the BPP dependence of the renormalization operation.

Both the adiabatic and the $(\pi\text{-}\rho)$-based plots reveal an almost perfect coincidence with higher than 98% agreement beyond 8×10^{2} $(\pi\text{-}\rho)$ interactions $\approx3\times10^{-7}$ s. The discrepancy between $\sigma(t)$ and $\sigma_{(\pi\text{-}\rho)}(t)$ raises to an upper bound of 8% within the time scale range $10^{-7}-5\times10^{-7}$ s. This is clearly due to the microscopic origin of fluctuations which becomes apparent at shorter time scales and is therefore only effectively captured by the $(\pi\text{-}\rho)$ dynamics over $(\mathbf{Z}_2)^{3N-3}$. An inspection of Fig. 7 reveals that the dynamics become entrained over the longer time scales ($\frac{1}{10}$ μs–1 ms) relevant to folding.

Summarizing, the initially large fluctuations observed in both computations of the Shannon entropy correspond to noncooperative misfolded structures, most of which are later dismantled to yield a fairly stable cluster of kinetically related structures [16]. The existence of such a dynamic intermediate state is confirmed by the existence of a plateau sustained within the $\frac{1}{10}$ μs–5 μs time scale range.

As observed in Fig. 7, σ does not tend to zero in the long-time dynamics relevant to the folding time scale frame. Rather, the coarse entropy decreases asymptotically to a plateau value $\sigma=3.4$ valid for $N=18$. This reflects the fact that folding into a unique structure, reaching a sharply peaked probability distribution within biologically relevant time scales is not a generic feature of the long-time chain dynamics. However, the almost perfect coincidence between the adiabatic and the projected torsional behavior is indeed a generic feature of RNA folding because it was obtained irrespective of natural selection, revealing the inherently binary structure of the soft-mode dynamics and the validity of a quasiequilibrium assumption over base-pairing patterns.

VII. CONCLUDING REMARKS

The semiempirical microscopic model of soft-mode dynamics and its associated predictive algorithm given in this work provide theoretical underpinnings to some issues which are central to understanding RNA folding as the long-time limit of torsional chain dynamics. In this regard, this work represents an attempt at elucidating the microscopic origin of the expediency of the RNA folding process. A vast time scale gap (1 ps to 1 ms) makes significant folding events inaccessible to molecular dynamics simulations of soft-mode (torsional) dynamics. Thus, we introduce a coarse description of the backbone torsional dynamics based on a binary ''*cis-trans*'' codification of local torsional states of the chain. This coarse states evolve in time, generating patterns which are recognizable as foldings of the chain. Such patterns are recorded in a base pairing contact matrix which, in turn, imposes new constraints on the time evolution of the coarse torsional states. This model captures the essential soft-mode dynamics of the chain, reproducing and explaining essential features such as the expediency of the folding process and the cooperative formation of the catalytic core in ribozymes or functionally competent RNA species [14–16]. The main issues addressed and incorporated in our semiempirical treatment are as follows.

(a) The parallel nature of folding, demanding a concurrent evaluation of folding possibilities at each stage of the process.

(b) The plethora of time scales inherent in our discretized version of torsional dynamics requiring a conformation-dependent hierarchy free from any *a priori* adiabatic ansatz whereby fast torsional motion would be treated as entrained by folding events.

(c) The role of the solvent which is not present merely as a statistical bath or a hydrodynamic drag term but rather as a set of conformation-dependent solvation environments which bear upon the structural development determining different conformational constraints [4].

(d) The role of Mg II ions which participate expediting the folding steps by simplifying the orientational constraints in the RNA backbone [14].

(e) The meaning and range of validity of the adiabatic approximation, widely used in RNA structure prediction [7,9,14,15], in which intrachain base-pairing patterns are regarded as quasiequilibrium states and transitions between

patterns are regarded as sequences of Arrhenius-like activated processes [6,7,9].

Our theoretical findings have been validated *vis-à-vis* kinetic experiments probing the favored folding pathways and phylogenetic inference of biologically competent structures [6]. Furthermore, following the results expounded in Sec. VI, we may attribute the expediency of the folding process to an adiabatic entrainment or subordination of the long-time torsional dynamics to a coarser dynamics involving the resolution of conformation space into contact patterns.

ACKNOWLEDGMENT

This work was produced with the help of the J. S. Guggenheim Memorial Foundation of New York City.

[1] T. E. Creighton, *Protein Folding*, edited by L. M. Gierasch and J. King (American Association for the Advancement of Science, Washington, DC, 1990), pp. 157–170.
[2] A. Fernández, Physica A **233**, 226 (1996).
[3] R. L. Baldwin, Proc. Natl. Acad. Sci. USA **93**, 2627 (1996).
[4] A. Fernández, H. Arias, and D. Guerín, Phys. Rev. E **52**, R1299 (1995); **54**, 1005 (1996).
[5] K. A. Dill, K. M. Fiebig, and H. S. Chan, Proc. Natl. Acad. Sci. USA **90**, 1942 (1993).
[6] P. Zarrinkar and J. Williamson, Science **265**, 918 (1994).
[7] J. A. Jaeger, D. H. Turner, and M. Zuker, Proc. Natl. Acad. Sci. USA **86**, 7706 (1989).
[8] M. Guenza and K. F. Freed, J. Chem. Phys. **105**, 3823 (1996); H. Cendra, A. Fernández, and W. Reartes, J. Math. Chem. **19**, 331 (1996).
[9] C. Cantor and P. Schimmel, *Biophysical Chemistry* (Freeman, New York, 1980), Vols. I–III.
[10] C. Brooks III, M. Karplus, and B. Montgomery Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics, Vol. LXXI of Advances in Chemical Physics* (Wiley, New York, 1988).
[11] A. Fernández, Z. Phys. B **79**, 255 (1990).
[12] A. Fernández, G. Appignanesi, and H. Cendra, Chem. Phys. Lett. **242**, 460 (1995).
[13] A. Fernández and G. Appignanesi, Phys. Rev. Lett. **78**, 2668 (1997).
[14] *The RNA World*, edited by R. F. Gesteland and J. F. Atkins (Cold Spring Harbor Press, New York, 1993).
[15] F. Michel and E. Westhof, J. Mol. Biol. **216**, 585 (1990).
[16] A. Fernández and G. Appignanesi, J. Phys. A **29**, 6265 (1996).