ELSEVIER

Biophysical Chemistry 61 (1996) 101–105

Biophysical Chemistry

Sequentially folded SV-11 RNA: metastability is relevant to biological function

Ariel Fernández [a,b], Alejandro Belinky [c]

[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*

[c] *Departamento de Economía, Universidad Nacional del Sur, Avenida Alem 1253, Bahía Blanca 8000, Argentina*

Received 8 January 1996; accepted 13 May 1996

Abstract

By simulating a Markov process comprising kinetically controlled elementary chain-growth steps and upstream refolding events, we compute the sequential folding of SV-11 RNA; a species capable of acting as a template instructing its own replication when folded into a metastable conformation. The predicted structure is endowed with available primers for replication and is identical within the regions of homology to the experimentally probed active structure of MNV-11 RNA, a species from which SV-11 RNA has evolved. The uneven rate of chain growth during SV-11 RNA replication and the pause sites along the replica sequence are determined. Such predictions are testable in pulse-chase kinetic experiments.

Keywords: RNA sequential folding; Recombinant RNA; Qβ replicase; Computer simulations

1. Introduction

In recent years, a number of RNA species have been shown to possess a biologically active secondary structure that differs from the global free-energy minimum to such an extent that the unsuitability of thermodynamics-based predictive algorithms cannot be merely attributed to uncertainties in the compilation of thermodynamic parameters [1–3]. This is not to say that such algorithms are not useful for finding stable conformations [4]: rather they should be complemented by other considerations whenever the active structure emerges within a stringent agenda that is incompatible with thermodynamic timescales.

Such considerations prompt us to compute the secondary structure of the RNA as it emerges from the synthetic machinery. The environmental complexity of such situations makes relevant computations impossible except in simple in-vitro systems. Thus, it has been shown that a sequential folding algorithm effectively predicts the active structure of MDV-1 RNA [1–3], a template that instructs single-stranded RNA replication as performed by Qβ-replicase [5]. The exploration of conformation space concurrent with the progressive synthesis of the MDV-1 replica can actually be simulated so that not only is the emerging active structure correctly predicted, but also the regions of slowing down in chain growth [5] are identified computationally [1].

In this work we focus on determining the active secondary structure of the 115 nucleotides (nts.)-long

0301-4622/96/$15.00 Copyright © 1996 Elsevier Science B.V. All rights reserved.
PII S0301-4622(96)02189-8

SV-11 RNA, a species known to be active or able to replicate only when it folds into a metastable secondary structure [2]. Our prediction for this secondary structure reveals that one of the two 5 nts.-long extremities of the molecule remains unbound and thus can serve as a primer for initiation of replication, a situation that entails a considerable enthalpic cost, thereby raising the overall free energy of the molecule. Making use of standard free energy compilations of conformational entropy losses that are due to loop closure and enthalpic losses that are due to intramolecular helix formation [4], we find that the free energy for the predicted structure shown in Fig. 1 is 11.8 ± 0.6 kcal mol^{-1} above that of the hairpin-like global free energy minimum. The latter structure is predictable using a standard free energy minimization algorithm [4].

The SV-11 RNA is a recombinant between the (+) and (−) strands of the natural Qβ template MNV-11 RNA (87 nts.). The evolution that leads to SV-11 under high ionic strength conditions will not be dealt with here. It is believed that a copy-shift mechanism might be involved in which the replication complex, say, (+) MNV-11 template-replicase replica, is destabilized and the budding replica adopts a new (−) MNV-11 template, whereby replication is pursued at a site closer to the 3′ end of the template. The net result of such recombination is a highly palindromic sequence whose most stable secondary structure is hairpin-like, as shown in Fig. 1 [2]. The predicted sequentially folded secondary structure obtained in this work is included in the figure for comparison. The most stable structure would not enable the SV-11 to serve as template since the primer regions are blocked from further recognition by Qβ-replicase.

However, one crucial piece of information regarding the active structure of SV-11 has become accessible through experiment: its metastability. This fact has been confirmed by pulse-chase kinetic experiments where the band for the structure that serves as template decays with time [2]. This observation prompts us to determine the secondary structure that would emerge from kinetically controlled sequential folding [1,6] and verify that such structure is indeed biologically competent. The biological competence

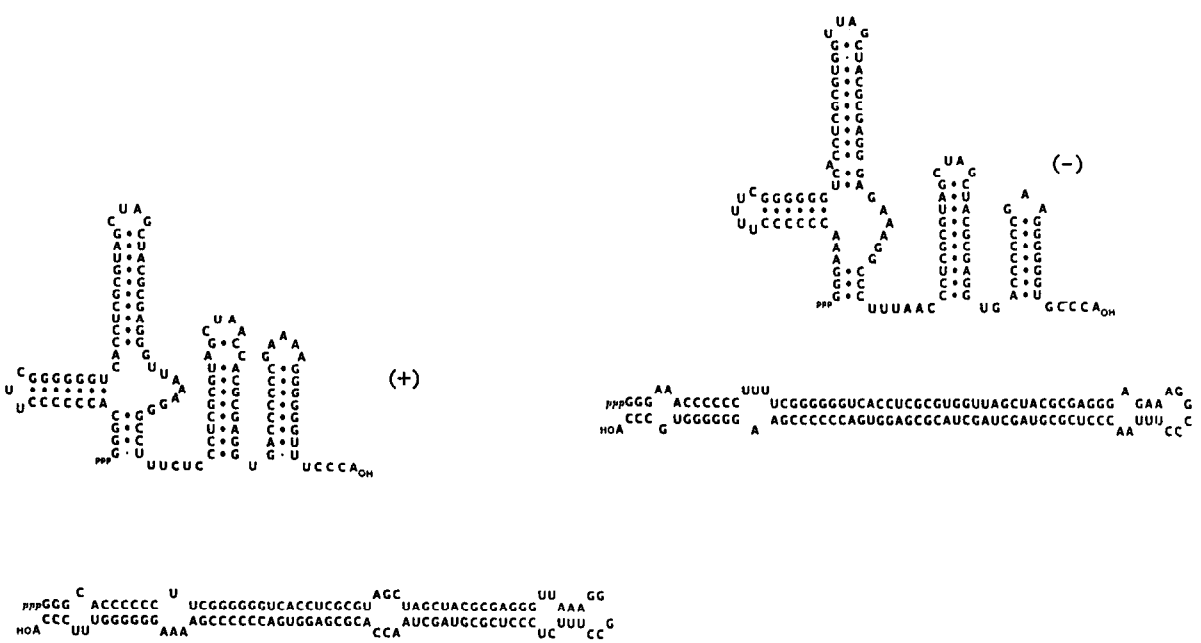

Fig. 1. The predicted sequentially folded and the most stable secondary structure of SV-11 RNA. The most stable structures are hairpin-like. The functional relevance of the predicted metastable structure has been inferred from two sources: philogenetic homological analysis and the awareness that the OH-extremity of the molecule is available for replicase recognition as it is intramolecularly unbounded, and thus can serve as primer.

of the predicted structure, in turn, will be assessed taking into account two key features: a) structural inferences resulting from the homology between SV-11 RNA and its evolutionary precursor, MNV-11 RNA, whose active structure has been probed experimentally [2]; and b) the availability of the OH-extremity serving as a primer for replication only if the SV-11 RNA molecule adopts its predicted sequentially folded secondary structure (Fig. 1). The importance of homological–philogenetic analysis as an auxiliary predictive tool stems from the simple observation that preserved regions of the RNA sequence are actually engaged in structure and therefore are functionally relevant, and this is precisely the reason why they must be conserved. The success of such methods is illustrated by the inference of the catalytically competent generic structure of group I ribozymes [7].

2. Materials and methods

The sequential folding algorithm has been described elsewhere [1,3], thus, only the basic tenets are included here for completion. We model the exploration of conformation space concurrent with the assembly of the chain which takes place by progressive incorporation of nucleotides. This is done by determining a Markov process. Each realization of the process is a sequence comprised of elementary events of two basic types: a) elementary refolding events in the form of intramolecular helix formation or decay, modelled as unimolecular all-or-none elementary steps [1,6], for which rate constant expressions have been derived analytically; and b) chain growth events, the unimolecular rate of which is the rate of phosphodiester linkage formation [1,3] fixed at 50 s^{-1}.

For completness we shall display the expressions for the unimolecular refolding rate constants. Given a specific stage in the sequential folding process, if the jth admissible step or event happens to be a helix decay process, we obtain [1,3,6]:

$$k_j = fn\exp(G_h/RT) \tag{1}$$

where f is the kinetic constant for base pair formation (estimated at 10^6 s^{-1} [3]), n is the number of base pairs in the helix formed in the jth step and G_h is the (negative) free energy contribution resulting from stacking of the base pairs in the helix. Thus, the essentially enthalpic term $-G_h \approx -\Delta H(\text{stem})$ should be regarded as the activation energy for helix disruption. If an admissible hairpin formation happens to be the event designated by the jth step at a given stage of the process, the inverse of the mean time for the transition will be given by:

$$k_j = fn\exp(-\Delta G_{loop}/RT) \tag{2}$$

where $\Delta G_{\text{loop}} \approx -T\Delta S_{\text{loop}}$ is the change in free energy owing to the closure of the loop concurrent with helix formation. Loop closure, being the nucleating step [1,3,6], is the rate-limiting event, therefore $-T\Delta S_{\text{loop}}$ is essentially the activation energy of helix formation.

The compilation of rate constants is based upon a given primary sequence. This requires prior elucidation of all a-priori plausible no-knotted secondary structures associated to each length of the RNA chain, a relatively canonical combinational problem. Thus, the compilation of thermodynamic parameters [4] gives the compilation of unimolecular rate constants upon which the Markov process is constructed for a given sequence.

To generate a Markov chain, at each stage of the process a single choice of events is made based on the following mechanism. Let $j = 1,2,\ldots,n$ be the ordered collection of plausible events at the fixed stage, with $j = 1$ always labelling the rate constant for the chain growth event by incorporation of a single nucleotide. We introduce a random variable r which belongs to the real line interval $[0, \sum_{j=1}^{n} k_j]$, uniformly distributed over the interval. Let r^* be a realization of r produced by a random number generator. Then, there exists a value of j, denoted j^* such that

$$\sum_{j=1}^{j^*-1} k_j < r^* \leq \sum_{j=1}^{j^*} k_j \tag{3}$$

Given this fact, we choose the event $j = j^*$ at the given stage. Thus, we have a kinetically controlled process in which the event with the highest transition probability is the one with the largest unimolecular rate constant. A cutoff in the rates is imposed by the actual time frame I of the experiment ($I \approx 10$ s, the

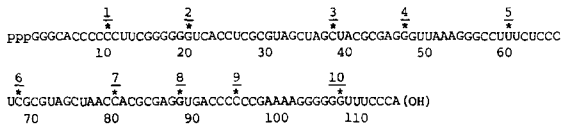

Fig. 2. The ten pause sites (indicated by asterisks) during template-directed synthesis of SV-11 RNA as determined by the sequential folding algorithm.

realistic replication turnover time [1,3]). Thus, no event j satisfying: $k_j^{-1} > I$ becomes admissible.

Sequential folding pathways resolved up to the secondary structure are realizations of the Markov process and reflect an opportunistic search in conformation space concurrent with the assembly of the chain. The kinetic control in the generation of pathways is compatible with the stringent schedule of enzymatic events in vitro.

3. Results and discussion

The results of the simulation for SV-11 RNA are displayed in Figs. 1 and 2 and Table 1. The most probable sequential pathway has been chosen in 100% of the 28 runs. The sites along the replica (+) strand for which refolding elementary steps have been favored over chain growth events are displayed in Fig. 2. In accordance with previous results [1,3,5] the location of these sites is experimentally testable, since they must coincide with the loci where the replicase must pause to allow for an upstream refolding event in the replica that partially relaxes the replication complex [1,3]. Such events are responsible for the modulation of the enzyme footprint [1,3,5] which must occur at the sites indicated in Fig. 2.

Table 1
The complementary regions which bind to each other as the predicted pause sites are reached. Transient intra-chain helices are indicated by "(t)"

Pause site	Complementary regions
1	(t) [1-3]–[7-9]
2	[5-11]–[15-21]
3	(t) [25,26,28,29]–[33,34,36,37]
4	[24-33]–[38-47]
5	[1-4]–[56-59]
6	(t) [53-56]–[64-67]
7	(t) [68-69]–[74-75]
8	[65-72]–[81-88]
9	(t) [54-56,58]–[90,92-94]
10	[90-97]–[103-110]

Thus, the chain elongation intermediates predicted in this work (Fig. 2, Table 1) could be detected by pulse-chase kinetic experiments [2,5]. The Watson–Crick complementary regions which bind to each other during each pause are indicated in Table 1.

The most significant finding and the one that validates the sequential folding scenario and predictive algorithm for the SV-11 replicating RNA is the emerging structure revealed by the simulations. This structure is displayed in Fig. 1 and, within the regions of homology, is identical to the active structure of MNV-11 RNA (Fig. 3), the ancestor of SV-11. The active structure of MNV-11 displayed in Fig. 3 has been probed by nuclease digestion mapping and has been computed making use of the Zuker algorithm [2,4]. This algorithm proved very useful in the case of the MNV-11 since the active and most stable structures coincide. However, the homologous structure in SV-11 is metastable since this species is a highly palindromic recombinant and thus the active

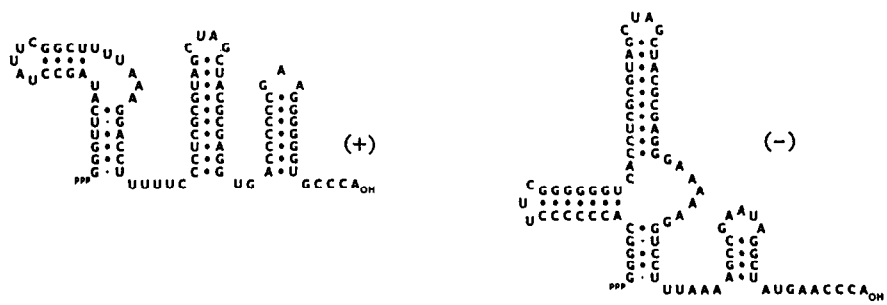

Fig. 3. The most stable secondary structure of MNV-11 RNA which coincides with the active structure.

structure is not predictable using thermodynamically controlled algorithms.

Acknowledgements

A.F. is a principal investigator of CONICET, the National Research Council of Argentina and his work has been supported by Fundación Antorchas (Argentina) and the John S. Guggenheim Memorial Foundation (USA).

References

[1] A. Fernández, Eur. J. Biochem., 182 (1989) 161–163.
[2] C.K. Biebricher and R. Luce, EMBO J., 11 (1992) 5129–5135.
[3] A. Fernández, Physica A, 201 (1993) 557–571.
[4] D.H. Turner, N. Sugimoto and S.M. Freier, Ann. Rev. Biophys. Biophys. Chem., 17 (1988) 167–184.
[5] D.R., Mills, C. Dobkin and F.R. Kramer, Cell, 15 (1978) 541–556.
[6] V.V. Anshelevich, V.A. Vologodskii, A.V. Lukashin and M.D. Frank-Kamenetskii, Biopolymers, 23 (1984) 39–57.
[7] F. Michel and J. Westhof, J. Mol. Biol., 216 (1990) 585–602.