Biophysical Chemistry, 45 (1992) 27–30
Elsevier Science Publishers B.V., Amsterdam

BIOCHE 01694

Noncoexisting structural elements in catalytic pre-messenger RNA's

Ariel Fernández [1]

Department of Biochemistry and Molecular Biology, Medical School, P.O. Box 016129, Miami, FL 33101-6129 (USA) and Department of Chemistry, University of Miami, Coral Gables, FL 33124 (USA)

(Received 25 May 1992; accepted 11 June 1992)

Abstract

We simulate the sequential folding of an autocatalytic pre-mRNA of group I revealing a scenario where core elements exert their function on intramolecular substrates as they are being generated. Our results indicate that the interactions shaping the 3′-substrate do *not* coexist with those shaping the 5′-substrate, but form after 5′-cleavage has occurred. This chronology of events is shown to be required for ribozyme function and quite universal in group I introns, since it is based on a competition of conserved helical stems. Preliminary probes rooted in site-directed mutagenesis are invoked to further validate the results.

Keywords: Pre-messenger RNA; Noncoexisting structural elements; Group I introns

1. Introduction

Even if all phylogenetically inferred structural elements of an autocatalytic pre-messenger RNA (pre-mRNA) of group I [1,2] had been probed combining site-directed mutagenesis with tests of ribozyme activity [3], one issue would still prevail: Is it certain that *all* such elements coexist in a single structure for a period of time, at least until until the flanking coding regions, the exons, are ligated? In this paper we compute the sequential folding of such pre-mRNA's thereby implementing a model which, unlike "single-structure" static models rooted in conformational free energy minimization [4], is able to address this question. Thus, we show the reply to the question posed to be negative, since the model must account for the fact that self-splicing demands a time-ordering in splicing events. For instance, the premature formation of the conserved P10 pairing between the internal guiding sequence (IGS) and the 3′-exon (Fig. 1) leads, when combined with the conserved adjacent interaction P9.0, to a strained backbone [1] resulting in hydrolysis at the 3′-extremity of the intron, *instead* of exon–exon ligation. We shall demonstrate that hydrolysis is precluded if the structure of the intron is searched for sequentially. This is achieved by delaying the formation of conserved helix P10 until 5′-cleavage had taken

Correspondence to: Dr. A. Fernández, Department of Biochemistry and Molecular Biology, Medical School, P.O. Box 016129, Miami, FL 33101-6129 (USA) and Department of Chemistry, University of Miami, Coral Gables, FL 33124 (USA). Fax: (305) 284-4571.

[1] Camille and Henry Dreyfus Teacher-Scholar.

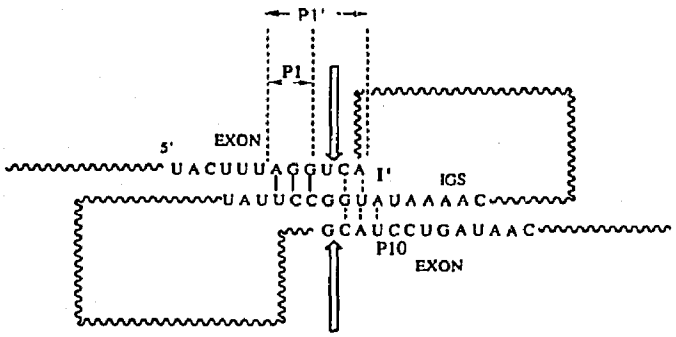

Fig. 1. Scheme of the competing base-pairing involving the internal guiding sequence (IGS) for the YCOB4 intron. The arrows indicate the splicing sites. The IGS–5′-intron helix, denoted I′, is part of the most probable but transient structure P1′ which forms during the initial stages of sequential folding. This structure will be dismantled when the extrinsic nucleophilic G-cosubstrate attacks the 5′-splicing site, directed by the recently formed catalytic core. Once the IGS has been disengaged, it becomes available for the long range P10 interaction which may form sequentially only when its associated 3′-exon segment has been synthesized.

place and the nucleophilic extremity of the already-spliced 5′-exon becomes available for attack at the 3′-junction. Since interaction I′, which competes for the same region of the IGS, gets disrupted only after 5′-splicing (see Fig. 1), it is only then that P10 is allowed to form. It is essential to confirm this scenario, since the competition between the two helices is universal within group I introns, as direct inspection of conserved sequences reveals [1].

The model introduced reproduces this scenario and is rooted in multiprocessed Monte Carlo simulations of sequential folding. We regard refolding events as concurrent with the synthesis of the RNA molecule and, consequently, handle them concomitantly with polymerization events. Thus, the simulation mimicks a Markov process such that if at a given stage a refolding event has a larger transition rate than a polymerization event, the former is chosen, whereas, if the reverse holds, the chain grows by incorporation of one nucleotide. Exploration of conformation space is biased by a partial retention of upstream structure during transcription [5,6]. Thus, sequential folding might prove relevant to understand ribozyme function for those mitochondrial pre-mRNA's endowed with a catalytic structure generated under transcriptional time constraints provided the structure is preserved until the splicing process has been carried to completion. We intend to show that the fourth intron of the yeast apocytochrome *b* gene (YCOB4) falls into this category [7]. Note, however, that our discussion is limited to those introns that, requiring P10 as a motif to shape the 3′-splicing site, cannot function as independent enzymes once excised from the exons. This is so since, once the intron has been spliced, there is no 3′-exon present to produce P10 and, therefore, the G residue at the 3′-extremity of the spliced intron remains deactivated for further nucleophilic attack.

2. Results and discussion

Our algorithm, being multiprocessed, is ideally suited to study structurally deficient group I introns, such as the YCOB4, which require trans-acting factors to reach a catalytically competent structure. Thus, we are able to predict all dominant competing folding pathways and this enables direct determination of the perturbations of the *in vitro* pathway caused by trans-acting factors which are required to yield a catalytically competent structure [2].

In order to determine the stages of folding where perturbations are required, we first simulate the optimal *in vitro* folding pathway. This pathway is subsequently compared with the main folding pathway generated in the parallel extension of the algorithm. By "main folding pathway" we mean the pathway which finally leads to the most probable structure which coincides with the catalytically competent secondary structure [1,2]. The simulations reveal that there are two occassions where disruptions from the optimal pathway are required to generate the structure which carries the highest statistical weight: (a) A trans-acting factor is required to form the highly conserved helix P7 (Fig. 2), responsible for furnishing the G-binding site [1], and (b): An already-formed upstream structure of the transcript itself must interact with the 5′-splicing site in order to make

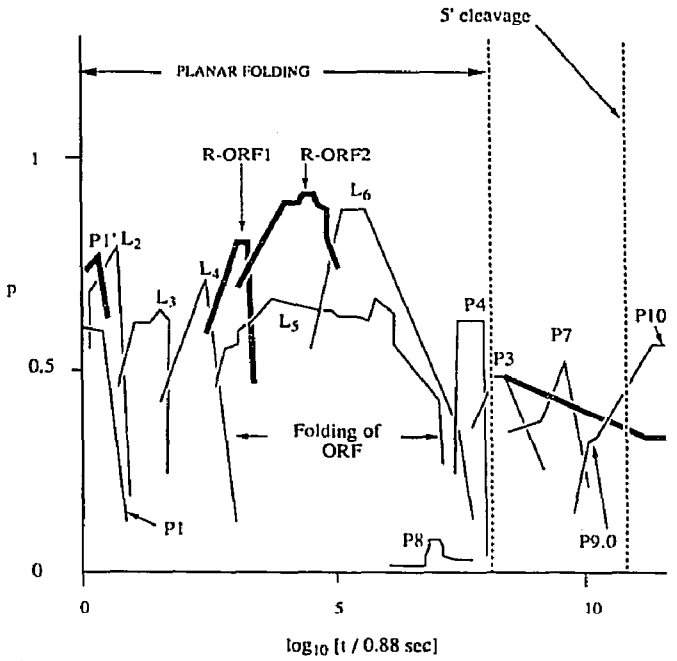

Fig. 2. Probability of transient structures along the main folding pathway. The thick lines represent structures which are not disrupted along the *in vitro* pathway. The thin solid line plot represent the main folding pathway occurring *in vivo*. A convenient logarithmic scale, monitoring the passing of time on the abscissas has been adopted. Conserved segments of the primary structure and conserved helical stems are labelled following standard notation [1,2]. The intramolecular loops that form sequentially as a result of the creation of conserved interactions are labelled L_2–L_6. The loop L_5 is complex and corresponds to the folding of the 1018 nucleotides-long ORF. As shown, the probability of a given structure increases as subsequent incorporation of nucleotides makes the structure more feasible and decreases when chain-elongation events lead to a better folding alternative. Each curve-crossing corresponds to a refolding event. The conserved segment R pairs initially to an internal open reading frame (ORF) which interrupts the intron sequence. This initial pairing forms R–ORF1 and this interaction is spontaneously dismantled at a later stage to form the more probable helix R–ORF2. Interactions P7 and P10 are absent in the *in vitro* pathway since they require prior disruption of existing structures. The terminal structure along the *in vitro* pathway is therefore inactive for splicing.

the IGS available for the long-range interaction P10. The folding pathway leading to the most probable structure is represented in Fig. 2. Direct inspection of the Fig. 2 leads us to conclude that the separation between the conserved segments R and S by a long open reading frame (ORF) makes the formation of the crucial G-binding-competent helix P7 possible only because, as we predict, an extrinsic factor disrupts the R–ORF interactions. This factor may be the translation machinery acting concomitantly with transcription or an intron-encoded maturase [2]. On the other hand, the occurrence of these interactions is intuitively obvious, since chances of partial complementarity between R and the 1018 nucleotides-long ORF are solid. Moreover, the absence of conserved interaction P8 along the main folding pathway favors the stacking of helices P3 and P7, thereby fulfilling a structural demand of the catalytic core. We have assigned the disruption of helix I′ to an interaction with part of the intron itself, acting as a perturbing agent. This is justified by following the *in vitro* pathway up until helix P4 is formed and, thereafter, simulating in parallel the branching from the *in vitro* pathway. In this case, we have found that helix I′ prevails in all competing pathways. This indicates that helix I′ will not be disrupted unless the disruption of the R–ORF helices had occurred previously. Thus, we may conclude that the disruption of helix I′ is not due to an extrinsic or trans-acting factor but to an intrinsic factor involving the catalytic core itself. Moreover, the competition between I′ and P10 is not accidental but necessary to prevent the premature formation of P10 which, as revealed in Fig. 2, forms strictly after 5′-cleavage.

Preliminary experimental work along these lines [3,8] is encouraging, since it reveals that the proposed scenario rationalizes the results obtained for a species closely related to YCOB4, the fifth intron (YCOB5) of the same gene: Destabilizing the interaction I′ by site-directed mutagenesis at the 5′-extremity of the intron does not affect significantly 5′-cleavage but greatly enhances the rate of 3′-hydrolysis over exon–exon ligation.

References

1 F. Michel and E. Westhof, J. Mol. Biol. 216 (1990) 585–610.
2 S.-H. Kim and T.R. Cech, Proc. Natl. Acad. Sci. USA 84 (1987) 8788–8792.

3 S. Partono and A.S. Lewin, Proc. Natl. Acad. Sci. USA 87 (1990) 8192–8196.
4 M. Zuker, Science 244 (1989) 48–52.
5 A. Fernández, Eur. J. Biochem. 182 (1989) 161–164.
6 A. Fernández, Phys. Rev. Lett. 64 (1990) 2328–2331.
7 A. Fernández, Chem. Phys. Lett. 183 (1991) 499–507.
8 A.S. Lewin, personal communication; B.W. Ritchings and A.S. Lewin, Nucl. Acids. Res., 20 (1992) 2349–2357.