International Journal of Theoretical Physics, Vol. 30, No. 2, 1991

RNA Self-Splicing and Energy Localization

Ariel Fernández[1]

Received July 29, 1990

We establish a mechanism for energy localization in regions of the sugar-phosphate RNA backbone which leads to the formation of transesterification or hydrolysis hot spots. In particular, our results account for the site specificity of the covalent cyclization and cycle reopening in the catalytic intervening sequence (IVS) of a ribosomal RNA.

One of the most striking features associated with the recently-discovered catalytic activity of RNA is the site specificity of the nucleophilic attack on phosphodiester linkages (Zaug *et al.*, 1984; Cech *et al.*, 1981; Marsh and Pace, 1985). The attack might lead to transesterification or hydrolysis. Examples of the first family of reactions are cyclizations (Zaug *et al.*, 1984), exon–exon ligations with concomitant cleavage of an intron–exon juncture (Cech *et al.*, 1981), and intron excision with concomitant incorporation of a guanosine cosubstrate to the 5′ end of the intron (Marsh and Pace, 1985). The second family of reactions could be best illustrated by the cycle reopening of circular intervening sequence (C IVS) RNA. This species, in turn, is the transesterification product of intervening sequence (IVS) RNA, a subsequence of the ribosomal RNA of *Tetrahymena* (Zaug *et al.*, 1984): The free ribose 3′-OH of IVS RNA is the nucleophilic agent which primarily attacks at the 15–16 phosphodiester bond; thus, cyclization occurs concomitantly with the elimination of a fragment 15 nucleotides long. Not only is the attack site specific, but, in addition, the hydrolysis leading to cycle reopening occurs at the same site where the transesterification had occurred.

This site specificity suggests the existence of "hot spots" along the backbone whose activation is pH dependent. In this work we shall assess the importance of nonlinear coherent collective modes on the problem of energy localization leading to hot spot formation and indicate a mechanism

[1]Department of Biochemistry and Molecular Biology, Medical School, Miami, Florida 33101, and Department of Chemistry, University of Miami, Coral Gables, Florida 33124.

0020-7748/91/0200-0129S06.50/0

by which the titratable groups are responsible for turning the hot spot into a vulnerable site for nucleophilic attack.

The working hypothesis put forward in this work to account for hot spot formation is that the phosphodiester linkage is weakened whenever the ribose 2′-OH is able to participate through its electron-donor role, as shown in Figure 1, leading to a resonance form between the classical 3′-phosphate and a cyclic 2′-3′-phosphate. Thus, the activated ribose 2′-OH assists the exogenous nucleophilic attack. In turn, the 2′-OH is activated only when a dynamic coupling between two specific processes, as shown in Figure 2, becomes feasible. These processes are: (a) ribose C2′-endo↔ C3′-endo repuckering and (b) deprotonation and concomitant tautomerization of uracil.

Our hypothesis is supported by the following facts (Zaug *et al.*, 1984; Cech *et al.*, 1981).

I. In the splicing of IVS RNA from rRNA, the free guanosine is attached to the 5′ end of IVS RNA, while the newly-formed 3′ end of the coding sequence contains uracil as the base.

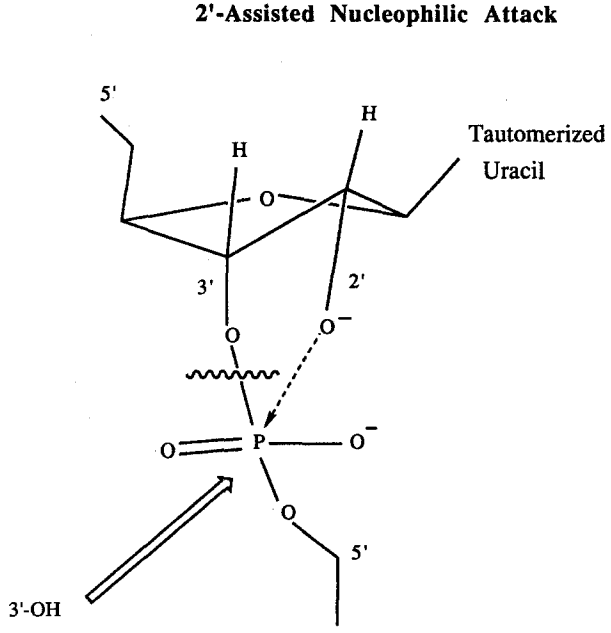

Fig. 1. Nucleophilic participation of the ribose 2′-OH in the phosphodiester bond cleavage at the hot spot. This participation implies a resonance between the classical 3′-phosphate and an activated cyclic 2′-3′-phosphate. The uracil is tautomerized concomitantly with the activation of the 2′-OH.

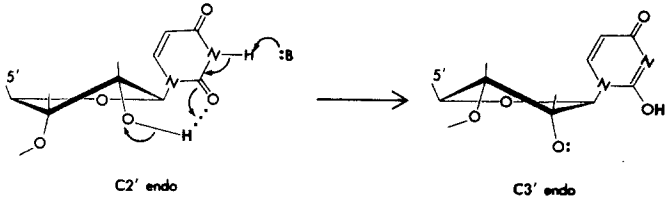

Fig. 2. Mode coupling leading to the participation of titratable groups (U-N3-H) in the creation of the hot spot. The sugar discommensuration might couple to the uracil tautomerization when the glycosidic torsional angle is able to depart from the constrained position in A RNA.

II. In the primary IVS RNA circularization, the base at position 15 is also a uracil. This residue will become the 3′ end of the excised oligomer.

III. In the alternative IVS circularization, the base at position 19 is also a uracil. This residue will become the 3′ end of the excised 19-nucleotide fragment.

IV. The cycle-reopening reaction occurs at exactly the same place where ligation took place, suggesting the existence of a memory effect. Most likely, this effect is given by the existence of a peculiar phosphate structure at the site of ligation. This structure adopts the form of a resonance between a 3′-phosphate and a cyclic 2′-3′-phosphate.

V. The reopening reaction is substantially enhanced (30–60 times faster) at pH 9, *precisely the pH at which deprotonation of the U-N3 H occurs*, when compared with the kinetics at pH 7.5.

In order to predict the hot spot region, we shall examine a "scenario" displaying the following features (Fernández, 1989*a*, 1991):

1. A coherent collective excitation in the form of a discommensuration or kink travels along the sugar-phosphate backbone and induces a C2′-endo↔C3′-endo repuckering of the ribose.

2. The discommensuration does not activate the ribose 2′-OH unless the glycosidic (ribose-uracil) bond torsional angle χ is able to depart from its value in the most probable "anti" configuration, which in double-stranded A RNA is approximately 13° (Cantor and Schimmel, 1980, and references therein).

Thus, in order to predict the hot spot location, we shall implement a statistical mechanical model for coherent excitations and prove that hot spots occur as a result of constraining the glycosidic angle χ in certain regions. *The constraints are therefore imposed by intrachain folding.*

Our model involves two variables per RNA unit: $u = u(x, t)$ and $\chi = \chi(x, t)$, where x denotes the contour variable, indicating the position on the chain, and u denotes the pseudorotational angle for the ribose pucker.

Thus, we introduce the following model Hamiltonian:

$$H_T = A_0 + B_0 + V + A_1(t) = H + A_1(t) \tag{1}$$

In equation (1), the first two terms are Hamiltonians which govern the dynamics of the sugar pucker and the swinging of the base, respectively. V describes a coupling between these two modes and A_1 is an external stochastic field. $A_0 = A_0(u)$ is given by

$$A_0 = \int \frac{dx}{l}\left[\frac{p(x)^2}{2m} + \frac{A}{2}u(x)^2 + \frac{B}{4}u(x)^4 + \frac{mc_0^2}{2}\left(\frac{du}{dx}\right)^2\right] \tag{2}$$

Here l is the lattice spacing and the contour variable x locates a sugar residue along the chain: $x = x_j = jl$; c_0 is the speed of sound. A is negative and B positive and they are chosen so that the depth $v = |A^2|/4B$ of the local potential wells ($u = \pm u_0$) corresponds to a barrier of 3 kcal between the two stable sugar configurations $u = +u_0$ (C3′-endo) and $u = -u_0$ (C2′2-endo) (Fernández, 1991; Bloomfield *et al.*, 1974).

B_0 is the contribution from the swinging of the base which involves only the torsional angle χ along the glycosidic bond between the base and the sugar pucker. This Hamiltonian contains the usual kinetic energy term and an elastic contribution, $K\{\partial\chi/\partial x\}^2$, that represents the stacking of the bases. The elasticity constant K should be estimated from the statistical thermodynamics of the separable Hamiltonian made up of the first two terms in equation (2): A reasonable denaturation temperature is obtained fixing K at 8×10^{-4} eV/deg^2 (Peyrard and Bishop, 1989).

The statistical mechanics which stems from B_0 is trivial insofar as only a phononic response is plausible. On the other hand, the statistical mechanics stemming from the contribution A_0 may lead to a nonlinear response in the form of a domain wall of thickness $2[2]^{1/2}\xi = L$. The parameter ξ determines the familiar form of the soliton solution (Peyrard and Bishop, 1989) for the dynamics determined by A_0:

$$u = u_0 \tanh[(x - vt)/2^{1/2}\xi] \tag{3}$$

$$\xi^2 = m(c_0^2 - v^2)/|A| \tag{4}$$

A partition function associated with a Hamiltonian equivalent to A_0 has been obtained by making use of the transfer operator technique (Peyrard and Bishop, 1989). The same technique has been implemented to study phase transitions in polymer physics (Fernández, 1989*b*). One must now add to the previous contributions the crucial term V, responsible for hot spot formation. This term arises from the coupling between the conformational change in the sugar and the swinging of the base, as suggested by

the chemistry depicted in Figure 2. We shall treat V as a perturbation and adopt the following simplified expression (Fernández, 1991):

$$V = Ju\chi \tag{5}$$

where J is a coupling constant. The stochastic source $A_1(t)$ is introduced to simulate the full spectrum of an external fluctuating field. This field is comprised of excitations which result from random collisions with the solvent molecules and counterions. We adopt for the field the general form

$$A_1(t) = \sum_w e^{-iwt} A_1(w) \tag{6}$$

where w is the frequency of a particular extended normal mode. Since the dominant wavelengths for phononic excitations are the ones close to the thickness of the domain walls (Fernández, 1991), we shall adopt exclusively the phononic modes in the range 85–100 cm^{-1} for $J = 2.4 \times 10^{-6}$ eV/deg^2. The choice of parameters does not determine the *location* of the hot spot, as we have verified by altering the value of the coupling constant J.

In order to examine the response of the system to the fluctuating field, we shall adopt the density operator representation. Denoting by $\rho = \rho(t)$ the density operator for the system, the partition function Q is

$$Q = \mathrm{Tr}\, \rho \tag{7}$$

Since we are interested in the effect of the soliton which propagates in the u space, we find it convenient to introduce the variable s defined by

$$\sigma = \mathrm{Tr}_\chi \rho = \underline{\mathrm{Tr}}\, \rho \tag{8}$$

where Tr_χ denotes the trace operation over the torsional angle variable only. We shall present the results of the lowest Born approximation to the collision superoperators. This procedure gives

$$\frac{d}{dt}\sigma_1(t) = -i\tilde{A}_0\sigma_1(t) + \tilde{C}_0(t, \{\sigma_1\}) - i\tilde{A}_1(t)\sigma(0) + D_0(t) \tag{9}$$

where σ_1 denotes the linear part of σ, and the letters with tildes symbolize superoperators associated with the corresponding terms in the full Hamiltonian. The lowest Born approximation to the collision operator $\tilde{C}_0$ is given by

$$\tilde{C}_0(t, \{\sigma_1\}) = -\underline{\mathrm{Tr}}\, \tilde{V} \int_0^t d\tau \exp(-i\tilde{H}_0\tau)\, \tilde{V} f(B_0)\sigma_1(t-\tau) \tag{10}$$

where $H_0 = A_0 + B_0$ and $f(B_0) = \exp(-\beta B_0)/\mathrm{Tr}\{\exp(-\beta B_0)\}$.

The last term in (9) is given by

$$D_0(t) = -\underline{\mathrm{Tr}}\ \tilde{V} \int_0^t d\tau \exp(-i\tilde{H}_0\tau)\tilde{A}_1(t-\tau)f^{(1)} \tag{11}$$

where $f^{(1)}$ is the lowest order term in the expansion of $f(H)$ in powers of V.

We shall calculate the thermal average $\langle\chi(x, t)\rangle$, making use of working equation (9) and the statistical mechanics for H_0, obtained explicitly from the Φ^4-model of Krumhansl and Schrieffer (see Peyrard and Bishop, 1989). In order to predict the existence of hot spots, we shall constrain the torsional angle for nucleotides 1–108 in the chain by introducing a 5 kcal/mole barrier for the interconversion: $\chi = 13° \leftrightarrow \chi = 8°$. We shall simulate on the system depicted in Figure 3, with a chain length of 500 nucleotides. The departure from the torsional value χ_0 (=13°) corresponding to the intrachain A RNA is shown in Figure 4, where $|\Delta| = |[\langle\chi\rangle - \chi_0]/\chi_0| \times 10^2$ is plotted against the contour variable. The results reveal sugar–base coupling at positions 108–115, leading to activation of the ribose 2′-OH and the production of a hot spot. The maximum coupling is achieved at the end of the hairpin, at position 108. Therefore, we expect primarily a transesterification or hydrolysis at the 108–109 phosphodiester linkage. Alternatively, the attack could occur, albeit with lower probability, at any position between nucleotides 109–115. The situation is strictly equivalent to the cyclization reaction for IVS RNA (Zaug *et al.*, 1984). In this case, the nucleophilic attack by the

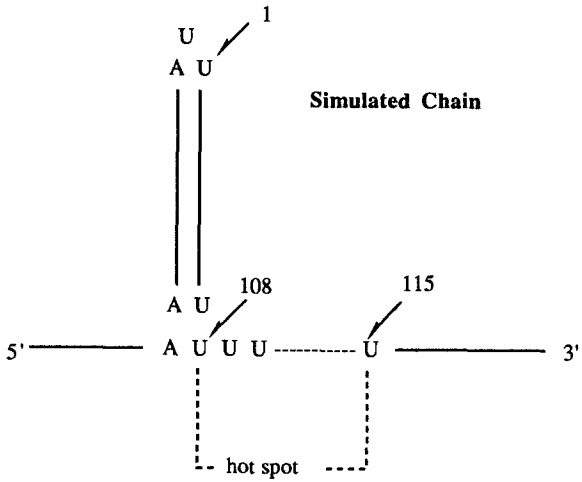

Fig. 3. Model RNA sequence on which the simulation is performed. The torsional angle constraints are imposed by the secondary structure.

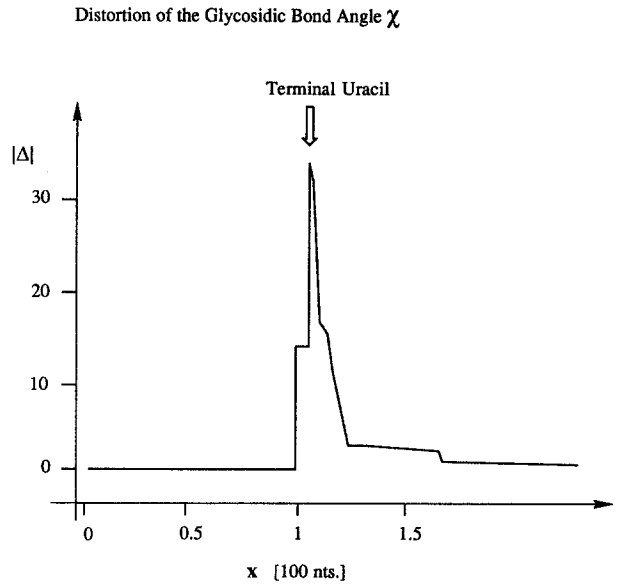

Fig. 4. Relative distortion of the glycosidic torsional angle χ as a function of the position on the chain, given by the contour variable x. The relative distortion is measured by $|\Delta| = |[(\chi(x, t)) - \chi_0]/\chi_0| \times 10^2$.

3′ terminus occurs primarily between the 15th and 16th nucleotides; the base at position 15 is a uracil and the 15th nucleotide is the terminal nucleotide of a hairpin. Moreover, the alternative nucleophilic attack occurs at the 19–20 linkage, with a uracil occupying the position 19, with this uracil being the nearest accessible uracil from the site of the primary attack.

We can conclude by stating that the creation of hot spots for transesterification and hydrolysis in catalytic RNA appears to be a consequence of the interplay between intrachain folding and the propagation of a sugar discommensuration along the backbone.

ACKNOWLEDGMENT

The author is a Camille and Henry Dreyfus Distinguished New Faculty Awardee.

REFERENCES

Bloomfield, V. A., Crothers, D. M., and Tinoco, I. (1974). *Physical Chemistry of Nucleic Acids*, Harper and Row, New York.

Cantor, C. R., and Schimmel, P. R. (1980). *Biophysical Chemistry*, Part 1, Chapter 6, Freeman, New York.

Cech, T. R., Zaug, A. J., and Grabowski, P. J. (1981). *Cell*, **27**, 487–492.
Fernández, A., (1989*a*). *Naturwissenschaften*, **76**, 469–471.
Fernández, A. (1989*b*). *Chemical Physics Letters*, **154**, 396–402.
Fernández, A. (1991). Coherent collective modes for catalytic RNA, *Zeitschrift für Physik B* (*Condensed Matter*).
Marsh, T. L., and Pace, N. R. (1985). *Science*, **229**, 79–81.
Peyrard, M., and Bishop, A. R. (1989). *Physical Review Letters*, **62**, 2755–2758.
Zaug, A. J., Kent, J. R., and Cech, T. R. (1984). *Science*, **224**, 574–578.