Biophysical Chemistry, 34 (1989) 29–33
Elsevier

BIOCHE 01371

Pause sites and regulatory role of secondary structure in RNA replication

Ariel Fernández

Max-Planck-Institut für Biophysikalische Chemie, Am Fassberg, D-3400 Göttingen, F.R.G. and Institute for Nonlinear Science, B-039, University of California at San Diego, La Jolla, CA 92093, U.S.A.

Received 6 January 1989
Revised manuscript received 12 April 1989
Accepted 9 May 1989

Secondary structure; RNA replication; Pause site; Monte-Carlo simulation

We analyze the correlation between pause sites and changes in enzyme boundaries within the elongation complex in template-instructed MDV-1 RNA replication. A Monte-Carlo simulation is carried out to follow the refolding events in the replica and in the portion of template strand upstream of the site where nucleotide incorporation takes place. We introduce and verify the hypothesis that the refolding events upstream of the replication fork are involved in the regulation of replication by leading to a partial relaxation of interaction between the enzyme and the growing chain. This relaxation is carried out by replacing enzyme-product binding by intra-chain pairing of the bases involved.

1. Introduction

The extent of binding of RNA replicase to template and replica during RNA replication remains poorly understood [1,2]. An aspect of the replication kinetics directly relevant to this problem is the existence of pause sites in the replication process [1]. An analysis of other related systems, like the RNA transcription apparatus [3], has been invoked in order to shed some light on the reason why the enzyme Qβ-replicase makes significant pauses during replication. In the better-understood process of RNA transcription it has been determined that conspicuous pauses entail substantial changes in the character of the elongation-complex [3]. For instance, in the case studied in ref. 3, the first prominent pause site, at position +16 in the mRNA chain is correlated with the breaking of contacts by RNA transcriptase with the promoter, a process concomitant with a great reduction of the footprint. The portion of template DNA covered by the enzyme is reduced by approx. 40 DNA base-pairs, as revealed by DNA footprinting and methylation experiments.

A further motivation for investigating RNA transcription in order to understand the pausing in RNA replication is provided by the analogy in the elongation intermediates for the prominent pauses in both processes: The RNA fragments present a 3′-terminal hairpin (see ref. 1 and references cited therein). Inspired by the analogy, we shall attempt to obtain evidence in support of the following hypotheses: (a) There exists a partial relaxation of binding to Qβ-replicase in the region of the elongation complex upstream from the site where a new nucleotide is incorporated after the pause. (b) The footprint reduction is achieved by refolding events in the growing chain of the replica reinforced by parallel intra-chain refolding in the portion of template strand upstream of the

Correspondence address: A. Fernández, Department of Biochemistry and Molecular Biology, School of Medicine, PO Box 016129, Miami, FL 33101-6129, U.S.A.

0301-4622/89/$03.50 © 1989 Elsevier Science Publishers B.V. (Biomedical Division)

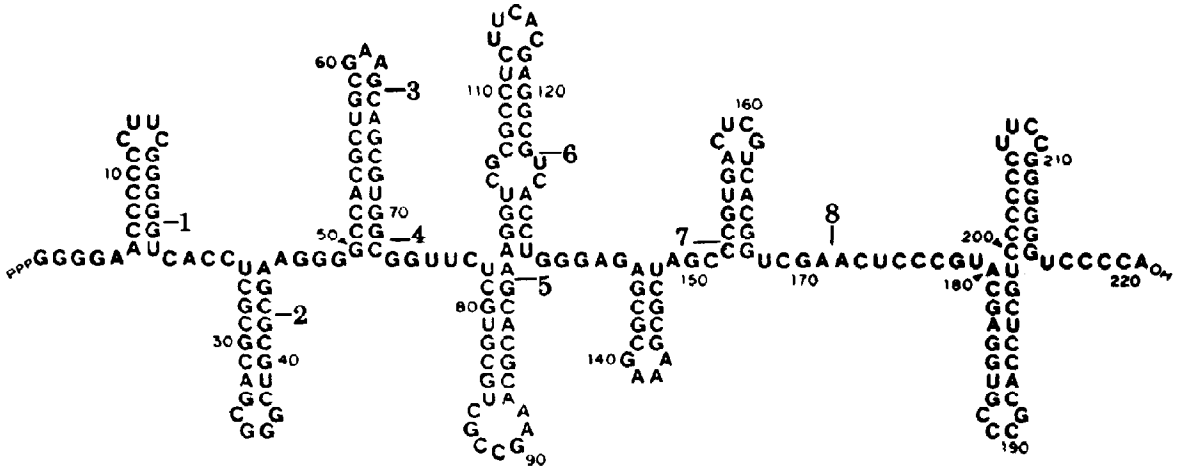

Fig. 1. Significant pause sites for the elongation in MDV-1 RNA replication.

elongation site. (c) These refolding events are the cause of the reduction of the footprint, thus enabling the enzyme to move forward along the replication fork, providing better solvation to the transition state in the subsequent formation of a phosphodiester bond.

We shall demonstrate that the coincidence between the pause sites and the sites along the growing chain at which a refolding event takes place is highly satisfactory.

One could in principle invoke other factors as playing some role in the pausing of the enzyme, for instance, it could be argued that, since the melting of intra-chain base-pairing in the downstream region of the template ('obstacle ahead') is required, this might affect the elongation rate. This possibility can be ruled out only after careful examination of the elongation intermediates for prominent pause sites. Since replica and template are complementary, the stop sites would occur at the beginning of a hairpin in the replica secondary structure rather than at the end, as shown in fig. 1 (even pause 7 corresponds to a terminal hairpin which is later dismantled when a more stable helix is formed [1]). That is so since a template double-stranded region must be melted before subsequent nucleotide incorporation can take place. Whatever the energy-transfer mechanism may be in this case (with the excitation provided by phosphate hydrolysis), it is sufficiently rapid for it to have no effect on the elongation rate.

In order to provide evidence for our hypotheses (a–c), we shall examine the results of Mills et al. [1] by carrying out a computer simulation of the refolding events in the growing chain, simultaneously allowing for events of chain elongation to occur concomitantly. The full replica for MDV-1 RNA on which the simulation is carried out is displayed in fig. 1. We shall make use of the algorithm presented in ref. 4.

2. Materials and methods

For the sake of completeness, we shall briefly summarize the nature of the simulation. Three elementary events can occur with a definite transition probability at a given instant: (1) intra-chain helix formation, (2) intra-chain helix decay and (3) chain growth by incorporation of one nucleotide. Each transition time is a Poissonian random variable τ_i ($i = 1, 2, 3$) with mean time $\bar{\tau}_i = k_i^{-1}$, where k_i is the kinetic constant for the transition. Thus [4],

$$k_1 = \chi N \exp(-\Delta G_1/RT) \qquad (1)$$

where χ is a kinetic constant ($\chi = 10^8\ s^{-1}$, cf. ref. 4) for the formation of a base-pair, N the number

of base-pairs in the intra-chain helix and ΔG_1 the free-energy change for the loops in the secondary structure due to helix formation. The values of the free energies for the components of secondary structures, helices, loops, etc., will be extracted from ref. 5. In addition, we have:

$$k_2 = \chi N \exp(-G_h/RT) \tag{2}$$

where $-G_h$ is the free energy of the helix [5]. The minimum admissible time span of a helix is taken as 5×10^{-1} s [4]. This cutoff value corresponds to the most fragile intra-chain helix which can be formed, creating a G-C pair. The loop entropies and stacking free energies used here were obtained from Salser's compilation at 25°C [5]. Recent revisions of the parameters used were obtained at 37°C by Turner and co-workers [6]. These results were extrapolated to 25°C by the authors and show no significant difference in the weight of the optimal kinetic secondary structure. The additional structures in the kinetic ensemble exhibit only minor weighting differences. Since our simulation only takes into account the most probable secondary structure at each stage, the subtle differences as a result of different parametrizations will not be pointed out here. The kinetic constant k_3 corresponds to a Michaelis-Menten elongation step with rate constant:

$$k_{MM} = \frac{k_{FP}[NTP]}{K_s + [NTP]} \tag{3}$$

where k_{FP} is the average phosphodiester linkage formation rate constant ($k_{FP} = 50\ s^{-1}$, cf. refs. 2 and 7) and K_s the average substrate association equilibrium constant. For this last constant we shall adopt the value 0.72×10^{-4}, given in refs. 7 and 8, which seems more accurate than previous estimations. For [NTP] we adopt the realistic value 0.5 mM [7]. We model the kinetic folding together with elongation by a Markov process in which one of the three elementary events is selected at each stage. The effective transition time for the process, τ, is a Poissonian random variable with mean $\bar{\tau} = k^{-1}$, where k is given by:

$$k = \sum_{i=1}^{F} k_1(i) + \sum_{j=1}^{D} k_2(j) + k_3 \tag{4}$$

where $i = 1, \ldots, F$ labels helices which can be formed in a way which is topologically compatible with the pattern of existing ones [4]. The latter are labelled by the dummy index $j = 1, \ldots, D$.

Let us now relabel the rate constants as follows:

$$k = \sum_{m=1}^{M} k'_m, \quad M = F + D + 1 \tag{5}$$

$$k'_1 = k_1(1), \ldots, k'_F = k_1(F), \; k'_{F+1} = k_2(1), \; \ldots$$
$$k'_{F+D} = k_2(D), \; k'_{F+D+1} = k_3 \tag{6}$$

This is done in order to derive the transition index m. To obtain this index we perform a Monte-Carlo simulation with a uniformly distributed random variable R, $0 \le R \le k$. If the value r of R lies in the interval

$$\sum_{m=1}^{m'-1} k'_m \le r \le \sum_{m=1}^{m'} k'_m \tag{7}$$

then the index m' has been chosen.

The refolding dynamics this simulation accounts for is rationalized as follows: The elementary events of chain elongation and formation or decay of a helix are coupled in the sense that the sequential incorporation of nucleotides opens up new alternatives for secondary structure formation which require the dismantling of previously existing intra-chain folding. Thus, the refolding events are naturally regarded as transitions. We speculate that such transitions entail a reduction of the footprint, or region covered by the enzyme. The grounds for the assertion are furnished by the analysis of analogous processes, such as RNA transcription [3]. In this case the pausing is concomitant with a sudden reduction of the footprint and the reduction is not gradual; instead, it takes place at definite positions along the incipient RNA chain. Thus, it is to be expected that refolding events in the growing RNA chain, reinforced by equivalent events in the unwound DNA strands, are the cause of the relaxation of enzyme-product interaction. In other words, the refolding events in both transcription and replication serve the purpose of allowing RNA strands to escape from the enzyme environment. Such events entail the pausing of the enzyme and they determine the reduc-

tion of the footprint. This reduction has been verified in RNA transcription [3]. We are predicting its existence for RNA replication, as suggested in section 3. The crux of the argument is that the pauses occur exactly where refolding events take place.

3. Results and discussion

The number of repetitions of the Markov process is $M = 1000$ and the actual time span for the whole elongation process (excluding the replica-release event) is taken to be 15 s (cf. ref. 2).

The time-dependent probability for the most probable secondary structure at a given time t, $p = p(t)$, is shown in fig. 2. The refolding transition occurs when a new secondary structure emerges and its probability grows, surpassing that of the previous most probable structure. The transitions are indicated by numbers in fig. 2. The nucleotide in the chain at which the refolding takes place is located by finding the time t at which two probability curves intersect or, as in transition 7, when a new structure with a higher probability suddenly emerges. The correlation between these transitions and the most prominent pause sites is revealed by comparison of figs. 1 and 2. A direct inspection of the pause sites as revealed in fig. 1 shows that the elongation intermediates have a terminal hairpin or a hairpin followed by three or four dangling nucleotides. These characteristic structures strongly suggest a partial relaxation of the enzyme-replica binding by refolding of the growing chain. The footprint reduction is reinforced by folding of the template strand in the upstream region which should take place in a parallel fashion. Further analysis would require introducing a fourth kind of elementary event in the Markov process which would compete with intra-chain refolding and elongation: the template-replica recombination in the upstream region. Although such an event cannot be discarded a priori since it would also relax the enzyme-substrate binding, it is incompatible again with the terminal structure of the elongation intermediates, as determined experimentally in ref. 1. For that reason, they were excluded in our present analysis.

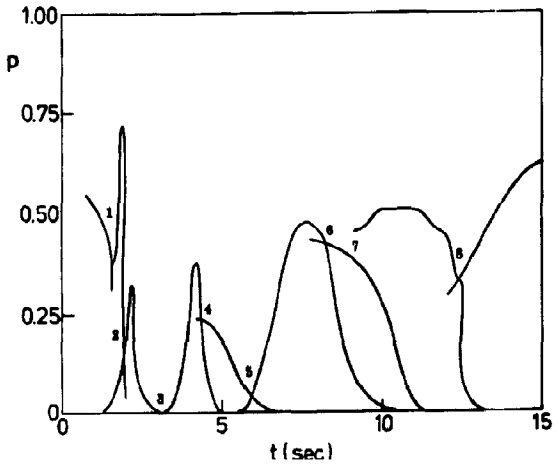

Fig. 2. Time dependence for the probability of the most probable secondary structures at any given time in the elongation process. The transition from one structure to another is labelled by a number. These transitions are correlated with the pause sites given in fig. 1.

To conclude, an additional test of the claims regarding footprint reduction at pauses can be suggested. Partial relaxation of binding between the enzyme and the growing chain due to intra-chain refolding should be detectable in methylation of the elongation complex. For instance, the G and C residues in the heads of the hairpins (see fig. 1) formed in progressive folding events should be protected before the pausing at the end of the respective hairpin and should be susceptible to methylation after the folding event has taken place. That must be so since the folding relaxes the binding of the enzyme to the subsequence which later on forms the hairpin.

Acknowledgement

Conversations and advice from Professor Manfred Eigen proved most useful.

References

1 D.R. Mills, C. Dobkin and F.R. Kramer, Cell 15 (1978) 541.
2 C.K. Biebricher, M. Eigen and W.C. Gardiner, Biochemistry 22 (1983) 2544.

3 A.J. Carpousis and J.D. Gralla, J. Mol. Biol. 183 (1985) 165.
4 A.A. Mironov and A. Kister, J. Biomol. Struct. Dyn. 4 (1986) 1.
5 W. Salser, Cold Spring Harbor Symp. Quant. Biol., 62 (1977) 985.
6 S.M. Freier, R. Kierzek, J.A. Jaeger, N. Sugimoto, M.H. Caruthers, T. Neilson and J.H. Turner, Proc. Natl. Acad. Sci. U.S.A. 83 (1986) 9373.
7 A. Fernández, J. Theor. Biol. 134 (1988) 419.
8 A. Fernández, J. Chem. Soc. Faraday Trans. I, 84 (1988) 1543.