3718–3725 Nucleic Acids Research, 1995, Vol. 23, No. 18

The influence of a metastable structure in plasmid primer RNA on antisense RNA binding kinetics

Alexander P. Gultyaev[1,2,+], F.H.D. van Batenburg[2] and Cornelis W.A. Pleij[1,*]

[1]Leiden Institute of Chemistry, Leiden University, PO Box 9502, 2300 RA Leiden, The Netherlands and [2]Institute of Theoretical Biology, Leiden University, Kaiserstraat 63, 2311 GP Leiden, The Netherlands

Received May 22, 1995; Revised and Accepted August 21, 1995

ABSTRACT

Replication of the ColE1 group plasmids is kinetically regulated by the interaction between plasmid-encoded primer RNA II and antisense RNA I. The binding is dependent on alternative RNA II conformations, formed during the transcription, and effectively inhibits the primer function within some time interval. In this paper, the folding pathways for the wild type and copy number mutants of ColE1 RNA II are studied using simulations by a genetic algorithm. The simulated pathways reveal a transient formation of a metastable structure, which is stabilized by copy number mutations. The folding kinetics of the proposed conformational transitions is calculated using a model of a multistep refolding process with elementary steps of double-helical stem formation or disruption. The approximation shows that the lifetime of the metastable structure is relatively long and is considerably increased in the mutants, resulting in a delay of the formation of the stable RNA II structure, which is the most sensitive to the inhibition by the antisense RNA I. Thus the effect of copy number mutations can be interpreted as a compression of the time window of effective inhibition due to an increased time spent by the RNA II in the metastable state. The implications of metastable foldings in RNA functioning are discussed.

INTRODUCTION

The copy number of *Escherichia coli* plasmid ColE1 is regulated by interaction between two plasmid-encoded RNA molecules (for reviews, see 1–3). One of these RNAs (RNA II) of 555 nucleotides serves as a primer for initiation of DNA replication by forming a RNA–DNA hybrid which is cleaved by RNase H. The primer formation is negatively regulated by interaction of RNA II with antisense RNA I of 108 nucleotides. The kinetic, rather than equilibrium, features play a key role in copy number regulatory circuits of ColE1 family plasmids and other antisense RNA-regulated plasmids (1–5). The interplay between the rates of various processes involved, like bimolecular interactions, conformational changes and RNA degradation, results in a system that maintains a steady state of plasmid copy number. In particular, the inhibitory effect of ColE1 RNA I on primer formation is mainly dependent on the kinetics of reaction with RNA II (6–8). RNA I is complementary to 108 nucleotides at the 5′-end of RNA II and the reaction results in RNA–RNA duplex formation preventing the formation of a functional RNA II structure. However, the initial steps of binding are determined by intramolecular secondary structures of both molecules. The reaction starts from relatively unstable loop–loop interactions ('kissing complex') that finally lead through a series of steps to an RNA–RNA hybrid along the entire length of RNA I. The kinetic parameters of some steps were studied experimentally (8) and allowed to develop a quantitative model of RNA I–RNA II binding kinetics (9).

The intramolecular conformational changes in RNA II are thought to play some role in RNA I–RNA II interaction kinetics. During its transcription, RNA II undergoes structure alterations involving the first 200 nucleotides, replacing one of the three stem-loop structures (stem III) by the alternative structure IV (Fig. 1) (7,8,10,11). This structural transition changes the binding mode of RNA II to RNA I and, consequently, the second-order rate constant of RNA II–RNA I interaction (7,8,10). There are several reasons to believe that the rate of the changes is of considerable functional importance. First, it is known that RNA I is an effective inhibitor only if it interacts with RNA II during a particular time window, while, although RNA I–RNA II interaction after this window still occurs, it does not inhibit RNA II–DNA hybrid formation (7). The second line of evidence for the importance of the RNA folding kinetics comes from the data on some copy-number mutations in the RNA II sequence. These mutations lead to an increased plasmid copy number due to defects in the inhibition by the RNA I (11,12). However, the nucleotide changes in some of the mutants influence neither the complementary region of RNA I and RNA II nor any of the proposed alternative RNA II structures (Fig. 1). This suggests the importance of some intermediate RNA II structures and the role of the RNA II folding pathway in the kinetics of RNA I–RNA II binding (11).

Possible free energy minimum structures for intermediate lengths of RNA II transcripts do show some differences between wild-type sequence and those of copy number mutants (11). However, the description of the RNA folding pathway as consecutive transitions between energy minimum states ignores

* To whom correspondence should be addressed

+On leave from Institute of Influenza, Laboratory of Molecular Virology, Prof. Popova Str. 15/17, 197376 St Petersburg, Russia

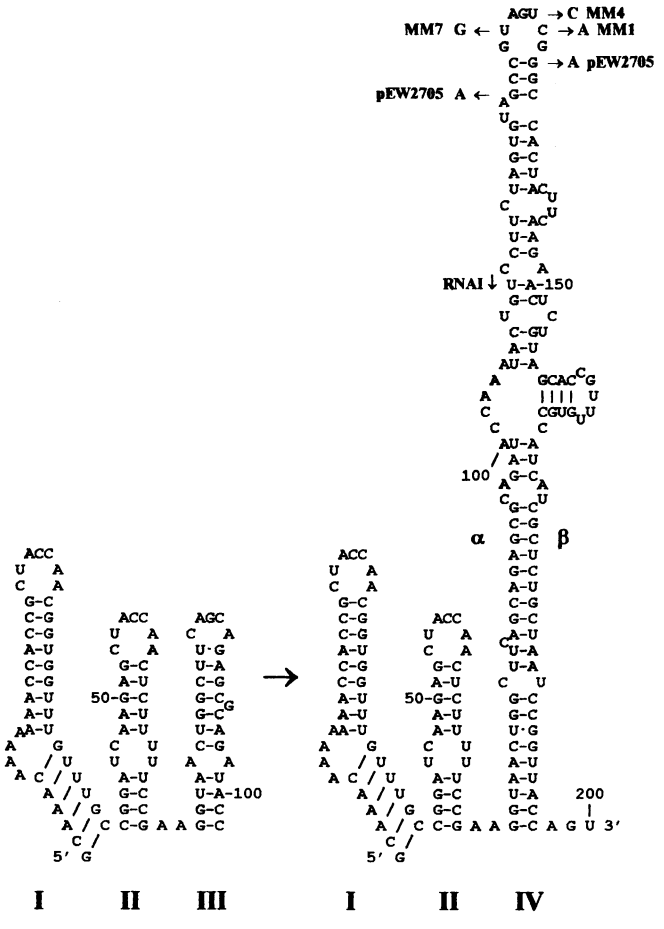

Figure 1. The main conformational transition in the region 1–200 of ColE1 RNA II (7,8,10,11). The copy number mutations (11,12) in the loop of the final structure are shown. The nomenclature of secondary structure elements (I–III) and the 3′-end of complementary region with antisense RNA I are indicated.

possible capturing of RNA in less stable structures. The assumption of a pathway going through the lowest energy structures means that the times of transitions depend on the transcription rate only and are ~0.1–0.2 s while the RNA I–RNA II interaction has a timescale of minutes (11). Thus the dramatic increase of plasmid copy number for certain mutants can not be explained and the current quantitative models for the ColE1 regulatory circuit (9,13,14) do not take into account RNA II folding kinetics.

The potential for intermediate RNA structures can be explored using the predictions of RNA folding by a so-called genetic algorithm (15,16). Instead of predicting the lowest energy state, the algorithm simulates the folding process for a growing RNA transcript. The advantage of the genetic algorithm compared to related methods for stepwise folding simulations (17–19) is the possibility to monitor intermediate foldings, that could be disrupted in the final structure and the procedure was shown to produce pathways close to the real ones (16). In this report we analyze the possible intermediate foldings for transcripts of wild-type and mutated sequences of RNA II molecules encoded by plasmids of the ColE1 family. The predictions clearly show the capturing of the RNA II molecule in metastable states, resulting in significant delays in the folding necessary for efficient interaction with RNA I. The timescale of the proposed conformational transitions is analyzed using a kinetic model (20) of RNA folding in the conformational space. The results show that the formation of the RNA II final structure is much slower than the transcription rate, and that the folding time is comparable with the time of RNA I–RNA II binding. This suggests that intramolecular folding kinetics should be taken into account for adequate quantitative descriptions of antisense RNA-regulated systems.

MATERIALS AND METHODS

The details of the method for RNA folding simulations using a genetic algorithm are given elsewhere (15,16). In summary, the algorithm generates a population of RNA structures for an intermediate length of transcript at every iteration. In the course of one iteration, new structures are generated by randomized disrupting and adding some stems in previous foldings and the new population is produced by selecting the most stable structures. The length of the RNA chain is gradually increased to simulate the folding of newly transcribed RNA. At every step the program displays the most stable folding found so far, which represents the simulated pathway through some local free energy minima. The used parameters for RNA secondary structure elements are as described in (21,22), with the exception of a parameter in the linear approximation for multibranched loops, which was changed from 0.4 kcal/mol per unpaired nucleotide (22) to 0.2 (16). The algorithm is implemented in the package STAR (18) for RNA structure predictions.

The folding model of Tacker *et al.* (20) was used to calculate kinetic curves for RNA conformational transitions. The model assumes the folding to occur in a stepwise way, where any reversible step is either formation or disruption of one stem. The rate constant for stem formation depends on the free energy of the loop to be formed while unfolding is determined by stem stacking value (20,23):

$$k_{folding} = A_1 \cdot \exp(-\Delta G_{loop}/RT)$$
$$k_{unfolding} = A_2 \cdot \exp(\Delta G_{stem}/RT).$$

Following the experimental data on hairpin formation rates (24) and earlier estimates given in (20), we set $A_1 = A_2 = 1.5 \times 10^8\ s^{-1}$. In this approximation, the folding consists of a set of monomolecular reactions. The time dependence for a change in the fraction of any possible folding is described by a differential equation according to the standard principle of mass action. Thus the folding scheme is represented by a system of differential equations:

$$d\mathbf{C}/dt = \mathbf{A} \cdot \mathbf{C},$$

where **C** is a vector of fractions (concentrations) of possible structures and **A** is a matrix of rate constants for structure formations with diagonal elements equal to negated sums of constants for structure disruptions. Because of the complexity of an analytical solution, kinetic curves were calculated by numerical integration.

RESULTS

The folding pathway of ColE1 RNA II

The genetic algorithm simulations of a folding pathway for the wild-type ColE1 RNA II did not show essential differences from

the pathway proposed (11) to pass through consecutive lowest energy structures. With the gradual increase in transcript length, the algorithm folded the first 102 nucleotides into the structure containing stems I–II–III, followed by a subsequent unfolding of stem III in favour of structure IV at an RNA length of 200 nucleotides (Fig. 2A and C). It should be noted that the genetic algorithm yielded slightly different pairings in the middle of the stem–loop structure IV (Fig. 2C) as compared with the lowest energy state (Fig. 1), namely, predicting a large internal loop in the region complementary to the 5′-end of antisense RNA I. The absence of pairing in this 'anti-tail' region is probably important for the RNA I–RNA II interaction (7,11).

In contrast with the wild-type pathway, the simulations for copy-number mutant RNAs (11,12) revealed new intermediate foldings. The intermediates did not differ essentially till RNA length reached the value of ~200 nucleotides. At this transcript length, all mutant sequences were folded in the structure with pairing of regions 121–136/178–196 (Fig. 2B). This folding left stem III intact whereas hairpin 117–141 (that should be present in the final structure within stem IV) was disrupted. The folded structure apparently created an obstacle for the formation of stem IV that requires the pairing of nucleotides 177–197 [the so-called β-region (1–3)] to the region 73–94 (α). Nevertheless, after several iterations the genetic algorithm disrupted this intermediate folding in favour of the final formation of stem IV (Fig. 2C). Thus the structure shown in Figure 2B could be considered as a transient intermediate only. It should be noted that the free energy values for the intermediate structure in all sequences studied were much higher than that for the final folding (–43.3; –45.7; –44.9; –47.6 and –45.1 kcal/mol for wild-type, MM7, MM4, MM1 and pEW2705 sequences, respectively, compared with –55 kcal/mol for all sequences but for pEW2705 where the interior folding of structure IV was slightly different with total energy of –54.3 kcal/mol).

It is interesting to note that all the copy-number mutations (11,12) stabilize the proposed intermediate folding by the formation of more stable base-pairs (Fig. 2B), increasing the possible energy barrier for its disruption. This should lead to a delay in the formation of stem IV. On the other hand, the intermediate structure would probably interact with RNA I slower than the final RNA II folding due to the intramolecular pairing of the 'anti-tail' (7,11) region complementary to the 5′-end of RNA I (Fig. 2B). An increase in the lifetime of the intermediate could explain the elevated copy number for mutant plasmids by the decrease of number of RNA II molecules available for the binding to RNA I during the time window of efficient RNA I–RNA II interaction. Thus this folding, despite its transient nature, supposedly influences the RNA II function considerably.

Simulation of the kinetic curve for ColE1 RNA II folding

The genetic algorithm simulations reveal only some intermediates in a folding pathway, but cannot show a timescale of folding. Thus it is impossible to estimate a delay in the folding of the final structure, which is caused by an intermediate. The evaluation of RNA folding kinetics can be carried out by desintegrating the overall process into a stepwise reaction where a single reversible step is either formation or disruption of a single double-helical stack (stem) (20,23). Such an approximation represents the RNA folding (or melting) as transitions along the edges of a n-dimensional hypercube, where n is the total number of possible stems. This model was applied by Tacker *et al.* (20) for a description of tRNA melting, assuming the formation of those stems only, that are present in the final structure, and was shown to be consistent with melting experiments.

The formalism proposed by Tacker *et al.* (20) can be applied for a folding pathway involving intermediate structures as well, with the only difference that some corners of a hypercube will not correspond to any structure and some edges will not represent any transition because of mutual incompatibility of some stems. Thus the construction of a folding scheme requires a definition of all stems that are taken into account. This set determines the structures to be considered as possible combinations of compatible stems. In the simple assumption of stem initiation by the first base-pair, the barrier for stem formation is the loop energy and that for unfolding is the stacking value (20,23). So, all rate constants for a given folding scheme can be calculated and kinetic curves for all structures in consideration can be computed by numerical integration of rate equations (see Materials and Methods).

To diminish the number of structures, taken for the analysis of ColE1 RNA II folding kinetics, and to avoid extremely long computations, we introduced several reasonable simplifications, resulting in the following reaction scheme. (i) Stem–loop structures I and II (Fig. 2) were not considered, because the conformational change does not involve them. Thus the refolding process was designed only for the region 71–200. (ii) Disruption of stem 158–161/167–170 was not considered because of the absence of favourable incompatible structural elements. Hairpin 107–120 was not included either, due to its low stability and compatibility with all structures in consideration, with the only exception of the 5′-end of hairpin 117–141. (iii) Formation of stem–loop structures 121–136/178–196 (suggested intermediate) and 73–94/177–197 (the final structure) was considered as a 'zippering' process (16), starting from the most short-range interaction only. Similarly, unfolding was assumed to start from long-range stems. (iv) Structures 73–102 and 117–141 were represented as single hairpins, because no stems incompatible with separate parts of them were considered. The stacking values for these hairpins were defined as sums of stacks, subtracted by destabilizing energies of internal loops. Similarly, the pairing 128–136/178–186 was considered as a single stem, the energy of the loop nucleotides 132 and 182 taken into account. (v) In addition to the stems shown in Figure 2, two other favourable hairpins, appearing transiently in some GA simulations, were taken into account (72–75/81–84 and 190–195/200–205).

Even with these simplifications, the reaction network requires the consideration of nine structural elements (Fig. 3) that could form 30 different structures. Being defined in such a way, the folding scheme included 55 reversible reactions. The numerical integration required a rather short time increment (0.5 μs) to take the fastest transitions into account.

Figure 4 shows the calculated kinetic curves over an interval of 1 min for the wild-type sequence and for the mutant MM7, which stabilizes the metastable structure by the formation of a G–C instead of an A–U pair (Fig. 2B). It is seen that initially the intermediate structure dominates in both cases. In the wild-type RNA II, the maximal amount of the intermediate structure (37.5%) is folded in ~20 s. At this moment only <25% of RNA forms stem IV of the final structure. Further kinetics shows the

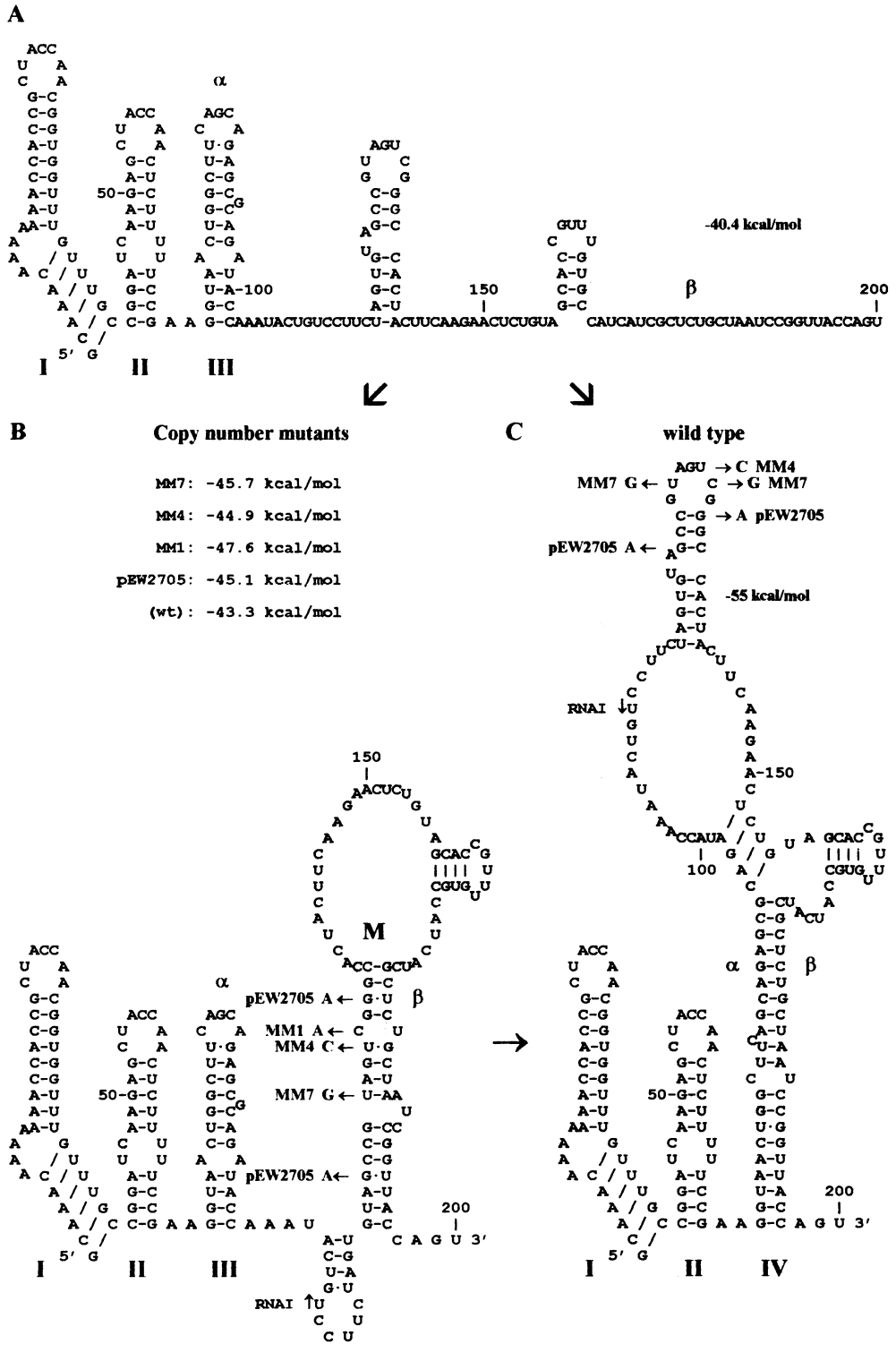

Figure 2. The simulated folding pathways for the wild type (A→C) and copy number mutants (A→B→C) of ColE1 RNA II. The stem predicted in the metastable structure is labelled as M, other structural elements are indicated as in Figure 1.

gradual disruption of the intermediate in favour of the structure with stem IV. Half of the molecules are predicted to form this stem after ~1 min, while ~28% yet retain the intermediate folding (Fig. 4). Thus the lifetime of the transient structure is rather long and this folding does represent a metastable state.

The lifetime of the metastable folding is considerably increased in mutant MM7: after 1 min, >70% of RNA II is predicted to have this structure whereas only 25% of the molecules are folded in the final state (Fig. 4). Thus at this moment the number of mutant molecules with the stable structure is approximately twice less as

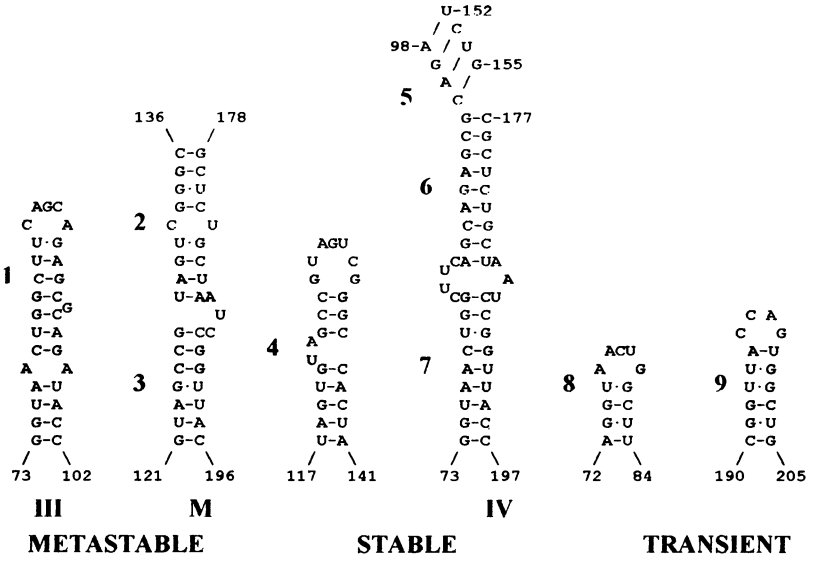

Figure 3. Structural elements (1–9) used in the approximation of RNA II folding kinetics.

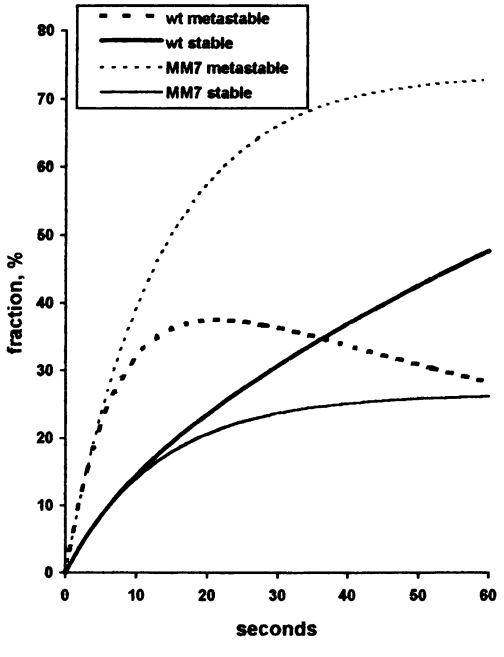

Figure 4. The calculated kinetic curves for the metastable (see Fig. 2B) and the stable structure (Fig. 2C) for the wild-type ColE1 RNA II and mutant MM7.

compared with the wild type. Although we did not calculate complete kinetic curves (numerical values can no longer be trusted due to the growing influence of downstream foldings), their qualitative behaviour apparently suggests that the difference could be even more pronounced at a longer time.

In terms of this approximation, the shift of folding in favour of the metastable structure for mutant MM7 is determined by the change in the stacking value of the stem that should be disrupted for the stable structure formation. Other copy number mutations, shown in Figure 2, also stabilize the metastable stem and consequently cause delays in the folding of the stable stem IV, similar to the mutant MM7 (not shown).

Possible metastable structures in RNA II molecules of ColE1-related plasmids

The structures I–II–III and I–II–IV (Fig. 1) can also be formed in RNA II molecules of other plasmids of the ColE1 family (3). We have analyzed the possibility of the proposed ColE1 metastable structure (Fig. 2B) in primer molecules of related plasmids by sequence comparisons of homologous regions. For this analysis the sequences of pBR322 (25), CloDF13 (26), RSF1030 (27), p15A (28), ColA and ColD (29) were taken. The alignments of these RNA II molecules are given in references 28 and 29. The comparisons show that the structures homologous to that shown in Figure 2B are rather unlikely in the related sequences, because some of its base-pairs cannot be formed. On the other hand, one can not expect a high extent of similarity here, bearing in mind that presumably this folding does not perform any specific function other than retarding the formation of stem IV.

In order to determine possible similar metastable foldings in RNA II molecules of other plasmids, we have carried out genetic algorithm simulations for the regions downstream of the stems homologous to stem III in ColE1 RNA II. Such an analysis allows one to reveal foldings that could retain stem III and do not contain stem IV. It turned out that such structures were folded in all RNA II molecules studied. The variety of produced structures (not shown) does not allow the derivation of a consensus folding. Nevertheless, the similarities between predicted structural elements in some primers can be noted. For example, the structures predicted for pBR322 and ColD RNA IIs (Fig. 5) involve the pairing of the β-region so that it creates a barrier for the interaction with the α-nucleotides necessary for formation of stem IV. Another conserved pairing was predicted in the 'anti-tail' region, complementary to the 5′-end of RNA I. This hairpin, corresponding to the hairpin 107–110/117–120 in ColE1 RNA II metastable structure (Fig. 2B), was predicted in almost all primers and a nucleotide covariation should be noted (Fig. 5). A different

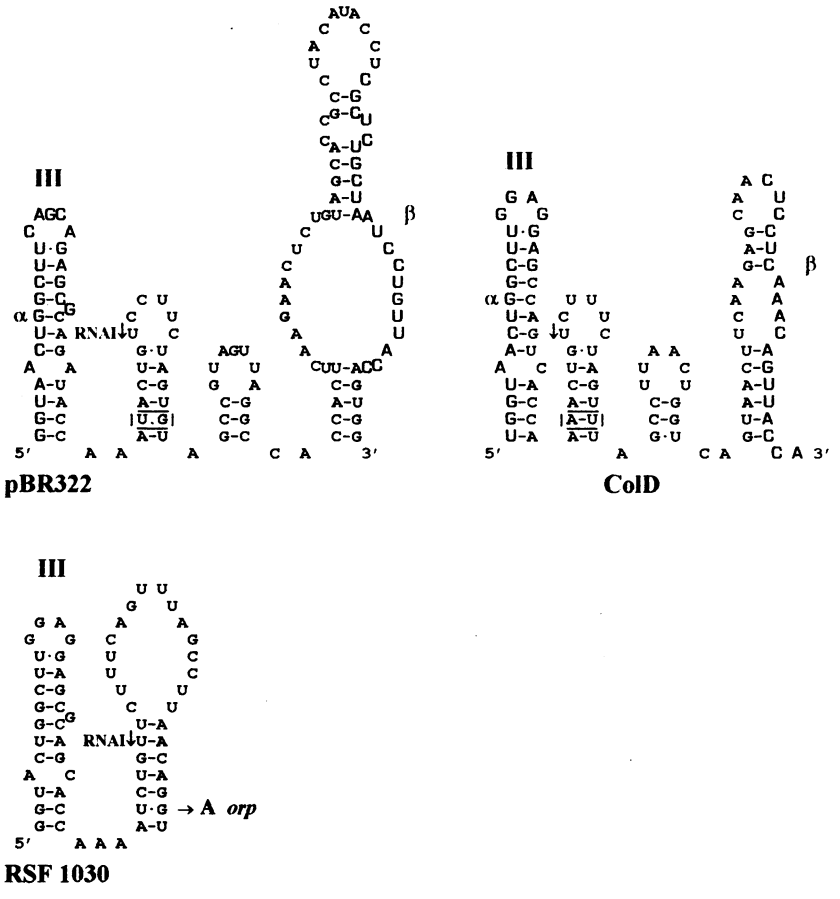

Figure 5. Possible intermediate structures, predicted by the genetic algorithm for RNA II of ColE1-related plasmids. The regions α and β in the pBR322 and ColD structures are shown by larger characters and the nucleotide covariation is boxed. The site of the copy number mutation *orp* in the RSF 1030 RNA 11(30) is indicated.

hairpin is predicted in the homologous region of RSF 1030 RNA II (Fig. 5). This hairpin also involves the anti-tail and is remarkably stabilized by a copy number mutation *orp* (30) which substitutes the U×G pair by the more stable U–A. Presumably anti-tail availability is important for the efficiency of antisense RNA I binding (7,11) and its intramolecular pairing could make the intermediate of RSF1030 RNA II (Fig. 5) less sensitive to RNA I inhibition than the final structure. The effect of RSF1030 *orp* mutation could then be explained by a delay in the disruption of the inactive structure, similar to the ColE1 mutations (Fig. 2B). A similar hiding of the anti-tail by a copy number mutation was proposed to occur in plasmid pMB1 primer structure (31). In summary, all RNA II molecules can be folded in structures that involve either the β-region or anti-tail nucleotides, while keeping stem III intact.

DISCUSSION

In general, RNA folding pathways simulated by the genetic algorithm reveal possible intermediate foldings which can be absent in the final structure. In the case of the ColE1 RNA II the main conformational transition occurs, along with the transcript elongation, between the lowest energy structures for chains of ~100 and 200 nucleotides (7,10,11). In addition, our algorithm predicts the formation of a transient structure in copy number mutant RNAs, folded at the same RNA length as needed for the main transition. This intermediate is not the state of minimal free energy for this transcript size and its formation can be explained by kinetic effects only. Assuming the temporary formation of the intermediate, stabilized by the mutations, their effect on plasmid copy number can be explained by some delay in the folding of the functional structure which in its turn is not influenced by the nucleotide changes.

Essentially, such a delay retards the folding of stem IV (Fig. 1) which is known to be important for both primer function and inhibiting antisense RNA I binding (1–3). However, for the primer activity, stem IV is not necessary by itself, but is only required to avoid alternative pairing of this part of the RNA with downstream regions and the nucleotides 1–200 are dispensable for the functional primer (32). The metastable structure, as suggested by the simulations (Fig. 2B), involves the same region and, consequently, does not seem to change the folding pathway of downstream regions as compared to the effects of stem IV. On the other hand, RNA I–RNA II binding does depend on RNA II conformational transitions in the region 1–200 (7,8). The inhibiting reaction with RNA I is enhanced by elongation of the RNA II transcript up to 200 nucleotides and more, presumably being dependent on the formation of stem IV (7,8). Thus the metastable conformation is able to retard an effective inhibition by RNA I.

Although the metastable structure is formed only temporarily, its lifetime influences the primer inhibition by RNA I considerably. The inhibition is effective during some time window and seems to be restricted by the irreversible formation of functional primer foldings in the downstream regions, which can no longer be altered by the RNA I binding (32). The metastable folding has no effect on the latter structures, so the increase of its lifetime in mutant RNAs (Fig. 4) only compresses the time window.

Presumably, the opposite effect of an extension of this interval determines the suppression of the copy number phenotype by secondary site mutations, found in ColE1 (10) and RSF1030 (30) plasmids. Such secondary mutations are dispersed in the functional primer region far downstream from the initial mutation and the suppression seems to be due to the reduced rate of primer functioning (30) rather than to long-range tertiary interactions (10). Indeed, one of the RSF1030 mutations, *sorp5*, which suppresses the effect of the mutation *orp* (30), shown in Figure 5, precisely corresponds to the mutation *pri6* (G370→A) in the ColE1 primer that confers a replication-defective phenotype (32). Thus the suppression of copy number phenotype seems to be the result of the shift in an endpoint of the window of effective inhibition by RNA I, restoring its duration, which was initially shortened by a stabilization of the metastable folding.

The kinetic curves (Fig. 4) suggest that the disruption of the ColE1 RNA II metastable folding occurs in a scale of minutes, which is comparable with the time of RNA I–RNA II binding (6–9). This means that the dramatic increase of the ColE1 copy number in the mutants can be clearly described in these terms. The decrease in the fraction of RNA II molecules, efficiently interacting with RNA I, may be 2-fold or more (Fig. 4). Although absolute values for the timescale produced by the used approximation are probably reliable in order of magnitude only, it seems that folding time could exceed the time of RNA II transcription which occurs in seconds (13).

It is thought that the elongation of the primer up to ~360 nucleotides makes it resistant to the structural changes caused by RNA I binding. This is concluded from binding experiments in which RNA II was slowly elongated to produce transcripts of different lengths prior to the interaction and the RNA I inhibitory effects were measured after subsequent incubation of RNA I with these transcripts (7). However, the possible influence of folding kinetics suggests a particular extent of elongation to be a necessary, but not a sufficient, condition. Presumably, experiments with slow elongation of RNA II transcripts prior to RNA I binding (7) show mixed effects of folding time and RNA length on the interactions. The folding time estimates, derived from kinetic calculations (Fig. 3), are consistent with these experiments. The maximal inhibition effect is observed after 3–6 min of transcription with RNAs of up to 300 nucleotides, while the next 3–6 min lead to transcripts of 360–400 nucleotides where RNA I binding does not suppress the primer function (7). A rough extrapolation of the wild-type curve in Figure 4 also gives an estimate of several minutes for a complete folding of the 5′-proximal 200 nucleotides which are the target for RNA I. Thus in the conditions of normal transcription rate an elongation up to this length does not necessarily ensure the completeness of the folding for this region. Similarly, the kinetics of the decline in the inhibition probably depends on the time required to form some functional structure, resistant to RNA I binding, rather than on the time necessary to synthesize this part of RNA.

It should be noted that quantitative estimates based upon the presented kinetic model should be made with precautions only. The values of rate constants are rather sensitive to small changes in assumed energy barriers, which are approximated here as energies of loops and stacks respectively. The pre-exponential term in the formula for rate constants (see Materials and Methods) is not precisely determined either. Rather slow folding kinetics (compared with the transcription rate) suggests that some structures in the regions downstream of the one in consideration could change the calculations. However, the qualitative behaviour of conformational transitions in ColE1 primer RNA seems to be reasonably approximated by the kinetic curves and is consistent with the available experimental data.

The important role of metastable structures in plasmid regulatory RNAs suggests that some estimates for intramolecular transitions should be included in quantitative models for the plasmid copy number control (9,13,14). Such effects could be taken into account by introducing either apparent rate constants or fractions of particular RNA populations. Indeed, some empiric values of RNA fractions are included in such models, in addition to kinetic parameters. A study on the rates of intramolecular changes could quantify such fractions and the window of primer susceptibility to the inhibition by antisense RNA. The recent critical comparative analysis of the current models for copy number control (33) indicates that this window is one of the most essential parameters. Alternative structures have been suggested also for antisense RNA targets encoded by plasmids of pT181 (34) and IncFII (35) groups. The replication of these plasmids is also kinetically controlled (for review, see ref. 4), and the rates of conformational changes could be important in the copy number control.

The analysis of metastable structures, presented here, emphasizes the significance of RNA folding pathways in RNA functioning. The procedures for RNA folding simulations have been proved to be very useful when an RNA functional structure deviates from the one that is given by the search for the lowest energy state, for example due to the presence of pseudoknots (18,19,36,37). The genetic algorithm, used here, provides the additional advantage in that kinetically favoured metastable foldings are predicted. The metastable structures in the mutant ColE1 primer molecules (Fig. 2B) are revealed by the genetic algorithm, although their free energies are ~10 kcal/mol higher than that of the final folding in this region, which is predicted as well (Fig. 2C). Moreover, the estimate of folding kinetics (Fig. 4) demonstrates that the lifetime of the metastable structure (presumably non-functional) can essentially modulate the functional activity.

The presented example of the importance of RNA metastable structure does not seem to be unique. A role of folding dynamics in RNA function is probably underestimated due to the difficulties of metastable structure detection as compared to equilibrium studies. Strong evidences of metastable foldings are mostly restricted to those molecules where structural transitions are relatively slow (38) and/or a functional role of a metastable structure is clear from mutagenesis data (39,40). For the ColE1 primer RNA, the role of the metastable structure is manifested because it influences the inhibitor binding which is effective during restricted time interval. In this case the changes in the folding kinetics due to mutations have a very strong effect on the RNA activity. On the other hand, suboptimal structures seem to be functional in a number of RNAs (41), and in some cases an analysis of folding pathways could both explain the deviation of

RNA structures from the lowest energy states and help to understand RNA structural–functional relationships that cannot be interpreted in terms of equilibrium structures.

Note added in proof

One referee suggested a discussion in the recent article by Chao *et al.* (42) on the properties of RNA II transcribed by T7 RNA polymerase. It is shown that overexpressed RNA II molecules are non-functional due to the increased transcription rate as compared with *E.coli* RNA polymerase. Chao *et al.* (42) suggest that the rapid rate of transcription inteferes with proper folding of the functional primer. These results, similar to our calculations, demonstrate kinetic effects in RNA II folding into a non-functional (metastable) state, presumably in the region responsible for RNA–DNA hybrid formation.

ACKNOWLEDGEMENT

This work has been performed under the auspices of the BIOMAC Research School of the Leiden and Delft Universities.

REFERENCES

1 Eguchi, Y., Itoh, T. and Tomizawa, J. (1991) *Annu. Rev. Biochem.*, **60**, 631–652.
2 Cesareni, G., Helmer-Citterich, M. and Castagnoli, L. (1991) *Trends Genet.*, **7**, 230–235.
3 Tomizawa, J. (1993) In Gesteland, R.F. and Atkins, J.F. (eds) *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 419–445.
4 Nordstrom, K. and Wagner, E.G.H. (1994) *Trends Biochem. Sci.*, **19**, 294–300.
5 Wagner, E.G.H. and Simons, R. (1994) *Annu. Rev. Microbiol.*, **48**, 713–742.
6 Tomizawa, J. and Som, T. (1984) *Cell*, **38**, 871–878.
7 Tomizawa, J. (1986) *Cell*, **47**, 89–97.
8 Tomizawa, J. (1990) *J. Mol. Biol.*, **212**, 683–694.
9 Perelson, A.S. and Brendel, V. (1989) *J. Mol. Biol.*, **208**, 245–255.
10 Wong, E.M. and Polisky, B. (1985) *Cell*, **42**, 959–966.
11 Polisky, B., Zhang, X.-Y. and Fitzwater, T. (1990) *EMBO J.*, **9**, 295–304.
12 Fitzwater, T., Zhang, X.-Y., Elble, R. and Polisky, B. (1988) *EMBO J.*, **7**, 3289–3297.
13 Brenner, M. and Tomizawa, J. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 405–409.
14 Brendel, V. and Perelson, A.S. (1993) *J. Mol. Biol.*, **229**, 860–872.
15 van Batenburg, F.H.D., Gultyaev, A.P. and Pleij, C.W.A. (1995) *J. Theor. Biol.*, **174**, 269–280.
16 Gultyaev, A.P., van Batenburg, F.H.D. and Pleij, C.W.A. (1995) *J. Mol. Biol.*, **250**, 37–51.
17 Martinez, H.M. (1984) *Nucleic Acids Res.*, **12**, 323–334.
18 Abrahams, J.P., van de Berg, M., van Batenburg, E. and Pleij, C.W.A. (1990) *Nucleic Acids Res.*, **18**, 3035–3044.
19 Gultyaev, A.P. (1991) *Nucleic Acids Res.*, **19**, 2489–2494.
20 Tacker, M., Fontana, W., Stadler, P.F. and Schuster, P. (1994) *Eur. Biophys. J.*, **23**, 29–38.
21 Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T. and Turner, D.H. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 9373–9377.
22 Jaeger, J.A., Turner, D.H. and Zuker, M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7706–7710.
23 Fernandez, A. and Shakhnovich, E. (1990) *Phys. Rev.*, **42**, 3657–3659.
24 Porschke, D. (1974) *Biophys. Chem.*, **1**, 381–386.
25 Sutcliffe, J.G. (1978) *Cold Spring Harbor Symp. Quant. Biol.*, **43**, 77–90.
26 Stuitje, A., Spelt, C.E., Veltkamp, E. and Nijkamp, H.J.J. (1981) *Nature*, **290**, 264–267.
27 Som, T. and Tomizawa, J. (1982) *Mol. Gen. Genet.*, **187**, 375–383.
28 Selzer, G., Som, T., Itoh, T. and Tomizawa, J. (1983) *Cell*, **32**, 119–129.
29 Zverev, V.V. and Khmel, I.A. (1985) *Plasmid*, **14**, 192–199.
30 Moser, D.R., Moser, C.D., Sinn, E. and Campbell, J.L. (1983) *Mol. Gen. Genet.*, **192**, 95–100.
31 Castagnoli, L., Lacatena, R.M. and Cesareni, G. (1985) *Nucleic Acids Res.*, **13**, 5353–5367.
32 Masukata, H. and Tomizawa, J. (1986) *Cell*, **44**, 125–136.
33 Merlin, S. and Polisky, B. (1995) *J. Mol. Biol.*, **248**, 211–219.
34 Novick, R., Iordanescu, S., Projan, S.J., Kornblum, J. and Edelman, I. (1989) *Cell*, **59**, 395–404.
35 Dong, X., Womble, D.D. and Rownd, R.H. (1987) *J. Bacteriol.*, **169**, 5353–5363.
36 ten Dam, E., Pleij, C.W.A. and Draper, D. (1992) *Biochemistry*, **31**, 11 665–11 676.
37 Gultyaev, A.P., van Batenburg, E. and Pleij, C.W.A. (1994) *J. Gen. Virol.*, **75**, 2851–2856.
38 Biebricher, C.K. and Luce, R. (1992) *EMBO J.*, **11**, 5129–5135.
39 Loss, P., Schmitz, M., Steger, G. and Riesner, D. (1991) *EMBO J.*, **10**, 719–727.
40 Qu, F., Heinrich, C., Loss, P., Steger, P., Tien, P. and Riesner, D. (1993) *EMBO J.*, **12**, 2129–2139.
41 Zuker, M., Jaeger, J.A. and Turner, D.H. (1991) *Nucleic Acids Res.*, **19**, 2707–2714.
42 Chao, M.Y., Kan, M.-C. and Lin-Chao, S. (1995) *Nucleic Acids Res.*, **23**, 1691–1695.