MS2 RNA has a Potential to Form an Unusually Large Number of Stable Hairpins

R. Nussinov, J. L. Sussman

Department of Structural Chemistry,
The Weizmann Institute of Science, Rehovot, Israel

AND

E. N. Trifonov

Department of Polymer Research
The Weizmann Institute of Science, Rehovot, Israel

(Received 25 October 1979, and in revised form 26 February 1980)

Several long nucleotide sequences were analysed to find out if any of them are unusual in terms of the possible formation of "hairpins" (pairs of complementary runs of nucleotides forming double-stranded structures) as compared to random sequences. The RNA of MS2 bacteriophage has more potential hairpins with short loops (up to 10 bases in a loop) than found in random sequences with the same length and base composition. Other analyzed nucleotide sequences (SV40, \(\phi X174\), Fd and 16S rRNA) behaved very much as their corresponding random ones. Most of the extra hairpins of the MS2 RNA are estimated to be thermodynamically stable. These potential hairpins might play some role in the function of the MS2 RNA.

1. Introduction

Since the determination of the first nucleic acid sequence (Holley et al., 1965) it was clear that hairpins, consisting of two complementary portions and a single-stranded loop (see Fig. 1), could be the basic element of the secondary structure of single-stranded molecules of nucleic acids. For tRNA structure this has been confirmed (Kim et al., 1974; Robertus et al., 1974). There are many other examples where specific hairpins appear to be correlated with some functions of particular regions of nucleic acids (Gilbert & Maxam, 1973; Maniatis et al., 1974; Sekiya & Khorana, 1974; Ravetch, Horiuchi & Zinder, 1977; Grosschedl & Hobom, 1979). By specific combinations of many complementary portions one can construct complicated structures as e.g. for MS2 RNA (Min Jou et al., 1972; Fiers et al., 1975; Fiers et al., 1976; Jacobson & Spahr, 1977).
The stability of any given hairpin is dependent on the base stacking interactions in the stem region as well as on the size of the loop region (Tinoco et al., 1973; Borer et al., 1974). For example, hairpins with stems containing four base-pairs of average composition are unstable if the loop size is more than about nine bases. If stable hairpins are of any importance for the secondary structure of single-stranded nucleic acids, one might expect some bias in the number of potential hairpins with short loop sizes. We therefore compared frequency distributions, i.e. the number of all the possible hairpins of different stem and loop sizes which can be formed by natural and random sequences with the same base compositions. Specifically, we checked the complete nucleotide sequences of the single-stranded DNA phages Fd (Beck et al., 1978) and \( \phi X174 \) (Sanger et al., 1977), 16S ribosomal RNA (Brosius et al., 1978), the MS2 RNA (Min Jou et al., 1972; Fiers et al., 1975; Fiers et al., 1976) and the DNA sequence of the eukaryotic virus SV40 (Reddy et al., 1978; Fiers et al., 1978). Surprisingly, we found that the distributions for all the sequences are indistinguishable from the respective random ones with the same base compositions. The only exception was the single-stranded RNA of the MS2 phage which showed a remarkably increased number of short loop hairpins.

2. Calculations

Starting from the 5'-end of a given sequence, for every successive run of three, four and five bases the nearest complementary run was recorded, as well as the length of the loop region between the complementary portions (see Fig. 1). Only these lengths of complementary runs were considered because longer ones are quite rare and therefore statistically insignificant, and shorter ones correspond to energetically unstable hairpins. Some bases can participate in the formation of several different partially overlapping hairpins. All of these were included in the tabulation, since it is not always...
known which one is preferred, i.e. we did not attempt to order them in terms of energies. Only classical Watson–Crick base-pairings were used (similar calculations were carried out including GU base-pairing, with essentially the same results). Clearly, from any long-stem hairpin one can generate a number of short-stem hairpins with longer loops. To avoid the trivial duplication we always tabulated only the hairpins with their longest possible stems. In this respect our classes of hairpins of stem-lengths three, four and five base-pairs are independent of each other. For each natural and its respective random sequence, we plotted the numbers of the hairpins as a function of their loop-sizes (≤90 bases). To diminish the noise level we used the averaged distributions of 100, independently generated, random sequences with the same base composition and length as the corresponding natural sequence.

3. Results

Of the five different sequences analyzed only the loop-size distributions of the MS2 RNA hairpins (stems four and five base-pairs long) appear to be definitely different from their corresponding random sequences (Fig. 2). Figure 2(a) shows the loop-size distribution of three base-pair long stems of the MS2 RNA. The distribution appears to mimic the one for random sequences. In contrast, loop-size distributions of four and five base-pair long stems have an anomalously large number of short looped hairpins. This is seen most dramatically in a plot combining both of these distributions [Fig. 2(b)].

Loop-size distributions carried out for the DNA's of Fd, φX174, SV40 and for the 16S rRNA do not differ from their respective analogues for random sequences [very much like the example shown in Fig. 2(a)]. A few deviations found are not statistically significant (no more than 1–2 s.d.). In contrast, the difference seen for loop sizes ≤10 nucleotides in the MS2 RNA is about 6 s.d. from the expected random value which makes it extremely unlikely to occur by chance.

4. Discussion

Hairpins containing only three base-pairs in their stems are in general unstable, while longer stem hairpins are most stable if their loop sizes are in the range of about three to nine bases (Tinoco et al., 1973; Borer et al., 1974). The excess of hairpins with loop sizes in this range found in MS2 RNA for stems four and five base-pairs long and not for three base-pair stems seems to imply that the nucleotide sequence is specifically biased to
facilitate the formation of thermodynamically stable hairpins. The potential
to form hairpins and the corresponding distributions might be influenced by
the base composition of the sequence. We excluded this possible influence
by comparing each natural sequence with random ones with the same base
composition.

Why is this anomaly in distribution of short loop hairpins found only in the
MS2 RNA, and not in the other nucleotide sequences analyzed? In contrast
to other organisms the genetic material of RNA phages like MS2, Qβ and
R17 replicates without using a double-stranded intermediate. Although the
replication requires a "minus" RNA strand for the synthesis of the "plus"
strand, these two complementary strands appear not to form double-
stranded RNA molecules \textit{in vivo} (Robertson, 1975). What then prevents the

FIG. 2. Frequency distributions of the number of hairpins in the MS2 RNA as a function of
loop sizes. (A) Stems three base-pairs long. (B) The combined distribution for stems four and
five base-pairs long. Dashed lines correspond to average distributions of 100 random sequences
with the same composition and length as the MS2 RNA. ———, MS2; - - - - - , random.
formation of double-stranded molecules, such as those observed for the single-stranded DNA phage \( \phi X174 \)? The excess of stable, potential, hairpins in the RNA of bacteriophage MS2 seems to provide at least a partial answer. We estimated that the average length of single-stranded portions between adjacent stable hairpins in the RNA is only about 20 bases. This means that any free portion of the MS2 RNA, with the exception of the replication (or translation) site, could be folded into hairpins. Since the physiological conditions are far from being optimal for reassociation of the complementary strands (Wetmur & Davidson, 1968), the highly folded structure could provide at least a kinetic obstacle (Weissman, Feix & Slor, 1968) to the formation of plus–minus double-helical RNA. Additional obstacles could be provided by specific protein molecules bound to the RNA (Robertson, 1975).

Compared to other RNA's of comparable molecular weight, the RNA chains of the MS2, R17 and Q\( \beta \) phages are unusually compact (Strauss & Sinsheimer, 1963; Overby et al., 1966; Mitra, Enger & Kaesberg, 1963). If the high number of stable hairpins found in the MS2 RNA is in some way responsible for the compactness of these RNA chains in solution, we would predict that the same feature should be found in R17 and Q\( \beta \) RNA's as well.

A typical potential hairpin corresponding to the region of anomaly [Fig. 2(b)] has a stem four base-pairs long and an average loop size of about five bases. The free energy of this average hairpin estimated by using the known energy contributions (Tinoco et al., 1973; Borer et al., 1974) is about 1–3 kcal/mol, i.e. 1–5 times \( RT \). Thus, many of the potential hairpins are only marginally stable in physiological conditions. It is to be expected, since too stable hairpins would slow down the replication and translation processes. Calculations of the free energy distributions of the potential hairpins in the MS2 RNA are now in progress. It would be worthwhile to study further the functional role of the enlarged number of stable hairpins in the MS2 RNA and to analyze other nucleotide sequences to see if the same phenomenon occurs there.

We wish to thank J. Moult, S. Wain-Hobson, T. Jovin and C. Sander for critical reading of the manuscript, and the Weizmann Institute Computer Center. The work was supported in part by Israel Ministry of Immigrant Absorption.

REFERENCES


