CABIOS

Vol.8. no.3. 1992
Pages 243 – 248

A procedure for RNA pseudoknot prediction

Jih-H.Chen, Shu-Yun Le[1,2] and Jacob V.Maizel[2]

Abstract

The RNA pseudoknot has been proposed as a significant structural motif in a wide range of biological processes of RNAs. A pseudoknot involves intramolecular pairing of bases in a hairpin loop with bases outside the stem of the loop to form a second stem and loop region. In this study, we propose a method for searching and predicting pseudoknots that are likely to have functional meaning. In our procedure, the orthodox hairpin structure involved in the pseudoknot is required to be both statistically significant and relatively stable to the others in the sequence. The bases outside the stem of the hairpin loop in the predicted pseudoknot are not entangled with any formation of a highly stable secondary structure in the sequence. Also, the predicted pseudoknot is significantly more stable than those that can be formed from a large set of scrambled sequences under the assumption that the energy contribution from a pseudoknot is proportional to the size of second loop region and planar energy contribution from second stem region. A number of functional pseudoknots that have been reported before can be identified and predicted from their sequences by our method.

Introduction

The three-dimensional structures formed by RNA molecules are crucial to their biological functions. It has been demonstrated that pseudoknots are important structural elements in RNA that can play an important role in the three-dimensional folding.

A pseudoknot is formed when bases outside a hairpin structure pair with bases within the hairpin loop to create a second stem and loop structure as depicted in Figure 1. The second stem can be stacked upon the first to form a quasi-continuous coaxial helix.

Pseudoknots have been predicted to be presented at the 3′ end of many plant-viral RNAs to mimic the structure, and therefore the function, of tRNA (Pleij *et al.*, 1985). There is also some chemical and enzymatic evidence for the existence of these structures (Rietveld *et al.*, 1982, 1983, 1984). The experiments by McPheeters *et al.* (1988) on the bacteriophage T4 gene 32 mRNA provide the first evidence of a pseudoknot motif in protein binding to an mRNA. The comparison of the T2, T4 and T6 gene 32 operator sequences gives phylogenetic evidence for conservation of the pseudoknot structure. Tang and Draper (1989) used direct protein–RNA binding measurements of an extensive set of site-direct mutageneses in the *Escherichia coli* α operon mRNA to confirm the pseudoknot folding, which has implications for the mechanisms of protein recognition and translational repression in this system. Varmus and co-workers have established that a stem–loop structure is required for frameshifting in Rous Sarcoma virus (RSV) *gag–pol* protein translation. For efficient frameshifting, a downstream stretch of 20 nucleotides is also essential; a tertiary interaction, perhaps a pseudoknot, is required (Jacks *et al.*, 1988). Brierley and co-authors (1989) provided strong experimental evidence that the pseudoknot structure is an essential element of the frameshifting signal for a non-retroviral system, avian coronavirus infectious bronchitis virus (IBV). In their experiments, mutations that disrupted either stem of the pseudoknot severely reduced frameshifting, while compensatory mutations that restored the pseudoknot also restored frameshifting.

The thermodynamic parameters of pseudoknots are unknown and the existing programs for energy minimization that predict secondary structure exclude the pseudoknot motif. In this report, we propose a method to predict a pseudoknot which is likely to be involved in some biological activity. From our method, we are able to predict the pseudoknot in IBV that is necessary for frameshifting, and the pseudoknot in bacteriophage T4 gene 32 mRNA that is involved in protein binding. We also predict four out of the five possible pseudoknots at the 3′ terminal noncoding region of tobacco mosaic virus (TMV) (Van Belkum *et al.*, 1985; Pleij *et al.*, 1987).

System and methods

The potential pseudoknot interactions of RNAs are located via the program RNAKNOT. Two different versions have been developed. One version is written in FORTRAN 77 on a Cray Y-MP supercomputer and runs in a Cray/Unicos (Cray Operating System) environment. The other one is written in FORTRAN 77 using DEC extensions and runs in a VAX/VMS environment. The programs are available from the authors upon request.

Advanced Scientific Computer Laboratory, Program Resources, Inc., NCI/FCRF, Frederick, MD 21702, USA, [1]Institute for Biological Sciences, National Research Council of Canada, Bldg. M-54, Ottawa, Ontario, Canada K1A OR6, and [2]Laboratory of Mathematical Biology, Division of Cancer Biology and Diagnosis, National Cancer Institute, NIH, Bldg. 469, Rm. 151, Frederick, MD 21702, USA

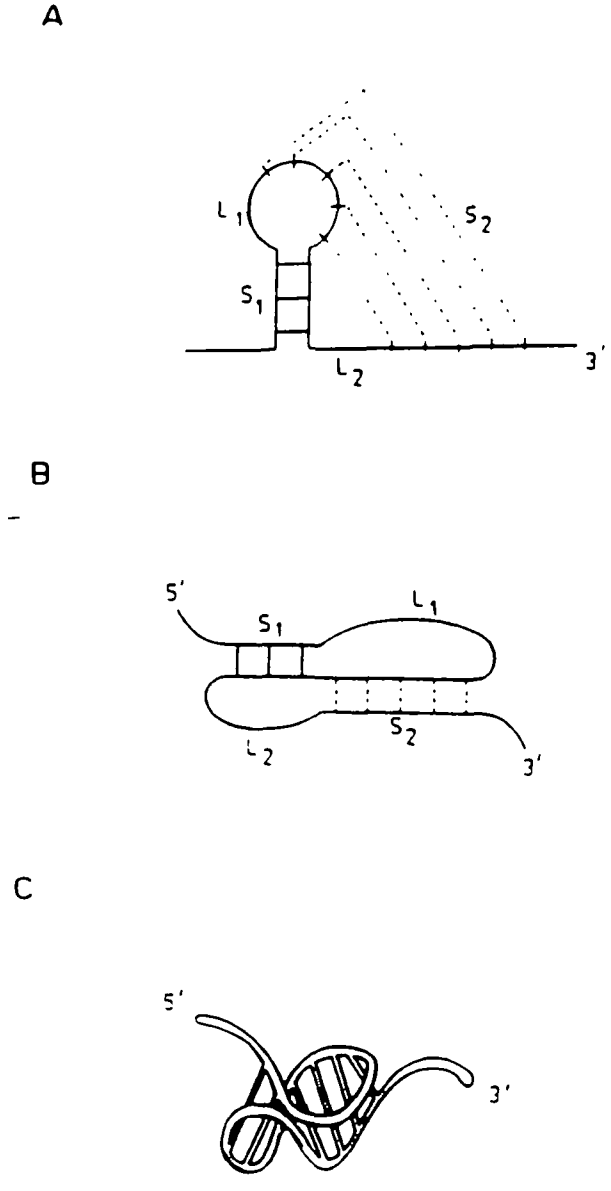

Fig. 1. Representation of pseudoknots. A pseudoknot is formed when a few bases at the 3' end pair with bases within the original loop (L_1) to form an additional stem (S_2) and loop region (L_2). (A) Conventional secondary structure presentation. (B) Schematic folding. (C) Three-dimensional drawing, showing coaxial stacking of the two stem regions to form a continuous helix.

Algorithm

In addition to providing experimental evidence for the pseudoknot motif in IBV, Brierley and co-authors (1989) also reported that 14 out of 22 viral RNA sequences examined appeared to contain the potential for pseudoknot formation downstream of putative or established ribosomal frameshift sites. We previously reported that the stem–loop structures situated at the frameshift sites of retroviruses are both statistically significant and thermodynamically highly stable (Le *et al.*, 1989). It suggests that it may be reasonable to require the stem–loop structure involved in pseudoknotting to be both significant and stable. The stability and statistical significance of a given stem–loop relative to other possible structures in the sequence are assessed by two standardized scores, stability score and significance score (Le *et al.*, 1991). The significance score (*Sigscr*) and stability score (*Stbscr*) of a segment are defined as: $Sigscr = (E - Er)/SDr$ and $Stbscr = (E - Eb)/SDb$. In the two equations, E is the lowest free energy of the real biological sequence in the segment. Er and SDr are the mean and standard deviation of the lowest free energies from a large number of randomly shuffled sequences of the segment. Eb and SDb are the mean and standard deviation of the minimum free energies computed by folding all fragments of the same size within the given sequence. Even though there is a lack of energy parameters for pseudoknot motifs, it seems logical to assume that the stem region created from a pseudoknot, as in the planar interaction, stabilizes the structure. On the other hand, the loop region has a destabilizing effect on the structure if the single-strand connecting loops L_1 and L_2 (see Figure 1) pose no sterical constraints. Therefore, between two possible pseudoknots characterized by (S_2, L_2) and ($\tilde{S}_2$, $\tilde{L}_2$) with same stem–loop structure (S_1, L_1), we choose (S_2, L_2) over ($\tilde{S}_2$, $\tilde{L}_2$) without knowing the free energy contribution from a pseudoknot if $e(S_2) \leq e(\tilde{S}_2)$ and $L_2 \leq \tilde{L}_2$ where $e(S)$ is computed from the tables of Turner and co-workers (1989) and L is used to denote loop region and the size of loop region interchangeably. Because secondary structure contributes more to the free energy gain upon folding of an RNA molecule than does the pseudoknot (Wyatt *et al.*, 1990), it is also reasonable to assume that the bases in the predicted highly stable regions cannot participate in pseudoknot interactions. These two assumptions greatly reduce the number of possible pseudoknots.

The two loop regions in the pseudoknot are not equivalent: L_1 crosses the deep major groove of stem 2 and L_2 crosses the shallow minor groove of stem 1. Loop length requirements will vary depending on the number of base pairs in the stem regions. In our procedure, the minimum requirements for L_1 and L_2 are 1 and 2 respectively, and the minimum number of base pairs in each stem is 3 (Pleij and Bosch, 1989). To simplify the notation, a potential pseudoknot is also denoted by (S_2, L_2), the second stem and loop regions.

Therefore, our proposed procedure to search for pseudoknots in a given sequence is as follows:

1. Compute the significance and stability scores for every segment in the sequence. The calculations are carried out by sliding the segment one base at a time along the sequence for every segment length (window) ranging from 20 to 300 nucleotides in increments of 2 nucleotides (Le *et al.*, 1991).
2. Collect the regions (segments) that are both thermodynamically and statistically significant. We consider the segment with a significance score less than −3.1 to be statistically significant. The segment is considered to be ther-

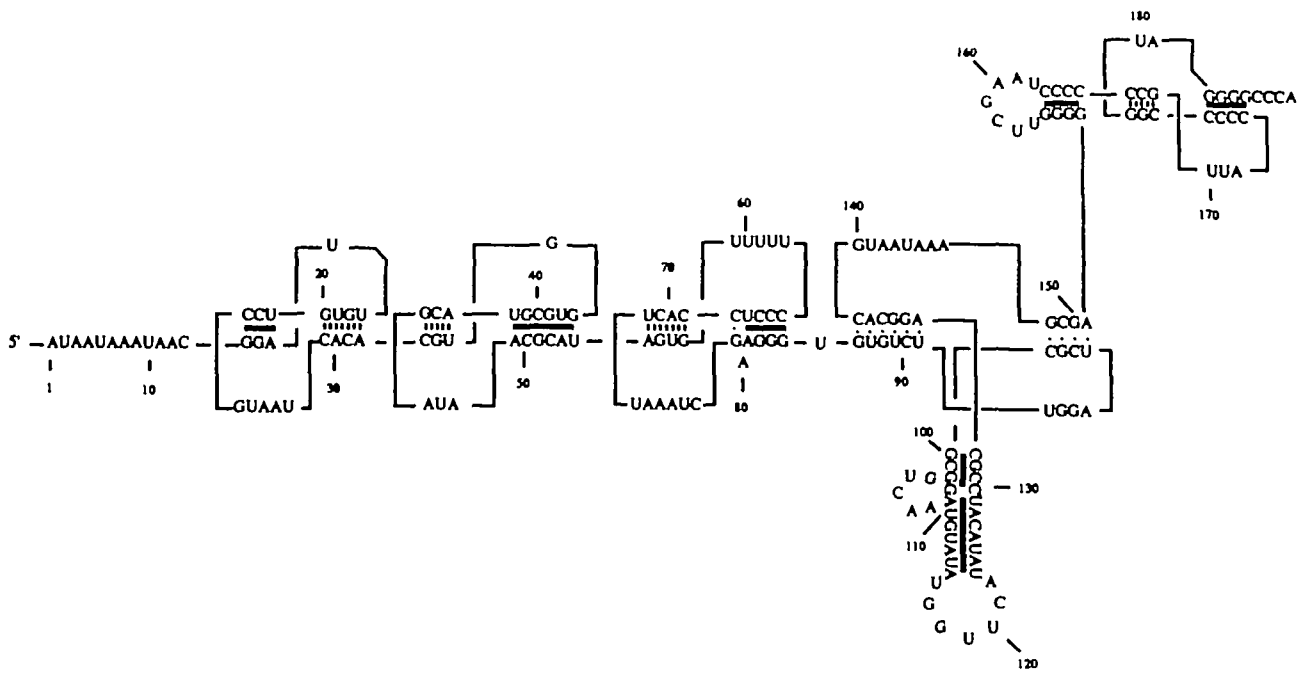

Fig. 2. The three-dimensional folding of the 3′ non-coding region of TMV RNA as proposed by Van Belkum *et al.* (1985). Base 1 corresponds to nucleotide 6206 of the entire RNA sequence. The black bars indicate the stems of the predicted compatible regions that are both significant and stable. The predicted tertiary interactions are hashed.

modynamically significant relative to others of same length in the sequence if there are no more than 5% of the segments of same length in the sequence that have a smaller stability score (Le *et al.*, 1991).

3. Select non-overlapped regions from (2).
4. Select non-overlapped regions that are thermodynamically significant (not necessarily statistically significant) and each of them is non-overlapped with regions in (3).
5. Search and form a list of all possible pseudoknots that satisfy sterical constraints and would not overlap with regions in (3) and (4) from both the 5′ and 3′ sides of any hairpin loop under consideration for each region in (3).
6. Eliminate the pseudoknot $(\tilde{S}_2, \tilde{L}_2)$ from the list if there is a pseudoknot (S_2, L_2) in the list such that $e(S_2) \leq e(\tilde{S}_2)$, $L_2 \leq \tilde{L}_2$ in which both are located at same side (3′ or 5′).

To evaluate the pseudoknot, we consider the subsequence 5′ upstream (or 3′ downstream) of the hairpin under consideration; we further calculate the following:

7. Compute $z = (nobs - rmean)/std$ for each pattern (stem-half of a pseudoknot) in the subsequence, where *nobs* is the number of times the pattern occurs in the subsequence, *rmean* is the average number of times the occurrence of the pattern in a set of scrambled sequences whose composition and length are identical to the subsequence, and *std* is the standard deviation. A large value in $|z|$ might be an indication of non-randomness in the pattern.

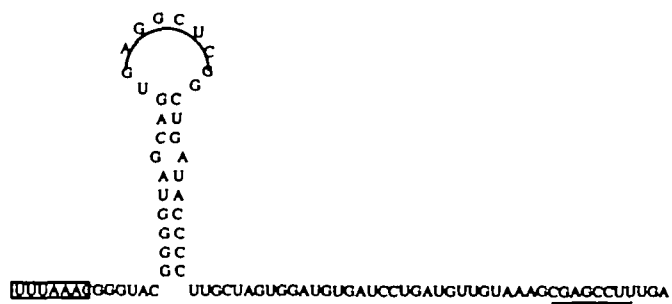

Fig. 3. This figure shows part of the IBV sequence of the RNA around the frameshift site, including the putative slippery sequence UUUAAAC (boxed), the predicted stem–loop structure downstream, and the predicted tertiary interactions (underlined).

8. Count the number, n_1, of randomized sequences that have a pseudoknot interaction $(\tilde{S}_2, \tilde{L}_2)$ such that $e(\tilde{S}_2) \leq e(S_2)$ and $\tilde{L}_2 \leq L_2$ in a set of scrambled sequences as in (7) for each pseudoknot (S_2, L_2) in the list. A large number of these randomized sequences might be an indication that the pseudoknot (S_2, L_2) is not thermodynamically favored to exist.
9. Count the number, n_2, of randomized sequences that can form a more stable pseudoknot than (S_2, L_2) under the assumption that the free energy contribution ΔG can be calculated explicitly by $\Delta G = a \times e(S_2) + b \times \ln L_2$ with $a,b > 0$. A necessary condition for the existence of a pseudoknot is $\Delta G < 0$, which is equivalent to $e(S_2) + c \times \ln L_2 < 0$ with $c = b/a$. Inspecting the pseudoknots from

the 3′ non-coding region of aminoacylatable plant viral RNAs (Pleij *et al.*, 1987), foot and mouth disease virus (FMDV) (Clarke *et al.*, 1987), bacteriophage T4 gene 32 (McPheeters *et al.*, 1988), IBV (Brierly *et al.*, 1989), and the pseudoknots from Puglisi *et al.* (1988) and from Wyatt *et al.* (1990), we found the minimal value for c is 1.30, which is obtained from tobacco rattle virus strain PSG where $e(S_2) = -2.1$ kcal/mol and $L_2 = 5$. In other words, $\Delta G < 0$ for every pseudoknot described above if $c < 1.30$. It is also easy to derive that a pseudoknot $(\tilde{S}_2, \tilde{L}_2)$ is more stable than a pseudoknot (S_2, L_2) if $[e(\tilde{S}_2) - e(S_2)] + c \times (\ln L_2 - \ln \tilde{L}_2) < 0$. Therefore, if $\Delta G < 0$ with $c = 2.5$ for pseudoknot (S_2, L_2) and if $n_2 = \max\{n_2(c = 0.9), n_2(c = 1.3), n_2(c = 2.5)\}$ is small, it may provide a supporting evidence for the existence of the pseudoknot (S_2, L_2) under the imposed condition.

Results

One of the best RNA sequences to test our procedure for the prediction of pseudoknots is the 200 nucleotide long 3′ terminal non-coding region of TMV RNA. The non-overlapped significant and stable regions, from our program, are regions 6215 ~ 6230, 6243 ~ 6260, 6268 ~ 6289, 6305 ~ 6338, 6356 ~ 6371 and 6375 ~ 6390 (see Figure 2). The possible pseudoknots occurring in these regions were extensively searched and evaluated.

There are two possible pseudoknot interactions with a hairpin in the region 6215 ~ 6230. One, located at 5′ side of the hairpin, had $e(S_2) = -1.6$ kcal/mol, $L_2 = 3$, $n_1 = n_2 = 521$ [the values in (8) and (9), n_1 and n_2, are calculated out of 1000 randomized sequences], and $\Delta G < 0$ only if $c < 1.46$ under the assumption in (9). The other, located at the 3′ side of the hairpin, had $e(S_2) = -5.7$ kcal/mol, $L_2 = 3$, $n_1 = 7$ and $n_2 = 18$, and $\Delta G < 0$ if $c < 5.17$. Therefore, we are in favor of the one located at the 3′ side of the hairpin for the region 6215 ~ 6230 as shown in Figure 2.

There are four possible pseudoknots for the hairpin in the region 6243 ~ 6260. The one shown in the figure had $e(S_2) = -4.1$ kcal/mol, $L_2 = 2$, $n_1 = 37$ and $n_2 = 65$, and $\Delta G < 0$ if $c < 5.92$. All the other three possible pseudoknots had larger connecting loops and n values ($n_1 > 150$ and $n_2 > 550$). Only one of them had a better energy contribution [$e(S_2) = -5.2$ kcal/mol] for the stem formed by pseudoknot but with a considerable large loop size of 41.

Similarly, the predicted pseudoknot for the hairpin in the region 6268 ~ 6289 had $e(S_2) = -5.6$ kcal/mol, $L_2 = 8$, $n_1 = 61$ and $n_2 = 196$, and $\Delta G < 0$ if $c < 3.5$. The one with better stem energy contribution [$e(S_2) = -6.8$ kcal/mol] in addition to having a large loop size $L_2 = 26$ and $n_1 = 110$, $n_2 = 396$, was also part of a predicted pseudoknot for the hairpin in the region 6243 ~ 6260. The other four possible pseudoknots had considerably lower energy contributions from the stem and large n values.

There is only one possible pseudoknot which had $e(S_2) = -4.9$ kcal/mol, $L_2 = 3$, $n_1 = n_2 = 35$ and $\Delta G < 0$ if $c < 4.46$ for the hairpin in the region 6375 ~ 6390. For the remaining two significant and stable regions, 6305 ~ 6338 and 6356 ~ 6371, all of the possible pseudoknot interactions had large connecting loops, weak interactions [small $e(S_2)$ values] and large n_1 and n_2 values. These findings might indicate that there is no pseudoknot for these two regions. In summary, four out of the five pseudoknots proposed at the 3′ terminal non-coding region of TMV RNA reported by Van Belkum *et al.* (1985) were successfully predicted from our method.

We also applied our method to an experimentally supported structure, a pseudoknot in IBV. We considered a sequence of 600 nucleotides comprising 300 nucleotides each from upstream and downstream of frameshift site UUUAAAC. The stem–loop structure located at 6 bp downstream of UUUAAAC is one of the non-overlapped significant and stable structures. There are six possible pseudoknots. The one shown in Figure 3 with connecting loop size 32 had the strongest stem interaction [$e(S_2) = -14.5$ kcal/mol and $\Delta G < 0$ if $c < 4.18$] and smallest n_1 and n_2 values ($n_1 = 0$ and $n_2 = 34$). Moreover, the z value in (7) is 31.58, which is obviously large enough to render the pattern in this pseudoknot significant. These are strong indications for the existence of this pseudoknot. All of the other possible pseudoknots with smaller connecting loops had considerably weaker interactions [minimal $e(S_2) = -5.3$ kcal/mol], larger n_1 and n_2 values ($n_1 > 140$ and $n_2 > 500$) and insignificant z values ($|z| < 1.0$).

In the system of bacteriophage T4 gene 32, there is only one region that is both significant and highly stable in a sequence of 200 nucleotides consisting of the proposed translational operator region and extended to 20 nucleotides in the 5′ direction. The pseudoknot proposed by McPheeters *et al.* (1988) is the obvious choice from our procedure for the same reasoning as the case in IBV system.

Discussion

In our procedure, we only searched pseudoknot interactions for the hairpin loop in a region that was both statistically and thermodynamically significant. The program can be modified to include the searching of pseudoknots for the single-strand in a bulge, internal or multibranch loop. Furthermore, we can unwind the upper stem of a hairpin to increase the hairpin loop size if the hairpin stem is disrupted by one or more bulge and/or internal loop and the stem length is no less than five after unwinding. With these modifications, we may be able to predict a potential pseudoknot, as already described by Brierley *et al.* (1989), which involved a multibranch loop in RSV RNA, and a pseudoknot proposed by Tang and Draper (1989) in *E.coli* α operon RNA by unwinding the upper stem of the predicted hairpin. Also, with modification, the one in TMV that failed by the program may be predicted by extending the region 6305 ~ 6338 to 6291 ~ 6344.

In the work of Wyatt *et al.* (1990), it was demonstrated that pseudoknots offer only a marginal free energy gain over that of the secondary structure. In our previous reports (Le *et al.*, 1989, 1990, 1991), we showed that the secondary structure that is likely involved in some biological activity is both statistically significant and relatively stable compared to the others in the sequence. The validity of this approach has already been demonstrated in several experimental structures which have confirmed our predictions (Le *et al.*, 1989, 1990; Malim *et al.*, 1989a,b). These suggest that the prediction of a pseudoknot that is likely to have biological function may be approached in two stages; the local secondary structure that is both significant and stable is determined and then the pseudoknot can be predicted.

The work of Wyatt *et al.* (1990) also provides a strong support for our criterion (5), which is a criterion to avoid having unnecessary pseudoknot interactions during the search. We use the z value in (7) to characterize the pattern appearing in a pseudoknot. The distribution of the z value is unknown. We assume that the pattern is statistically significant if $|z| > 10.0$. From our experience in TMV, the z value might not be a sensitive measure. The z values for four predicted pseudoknots all have magnitudes < 3.0, but, a large $|z|$ can serve as supportive evidence for the existence of a pseudoknot. The large n_1 value in (8) may be an indication that the pseudoknot interaction occurs by chance. We use this information to filter out some unlikely pseudoknot interactions. Under the assumptions, $\Delta G = a \times e(S_2) + b \times \ln L_2$ with $a,b > 0$, in (9), the free energy gain or loss for a pseudoknot interaction (S_2, L_2) is reflected by the sign of $\Delta G/a = e(S_2) + c \times \ln L_2$ with $c = b/a > 0$. The c value in $\Delta G/a$ is a measure of the relative weight between the contribution from $e(S_2)$ and $\ln L_2$. The thermodynamic contributions from stem and loop regions formed by a pseudoknot interaction may also be reflected by the maximal value in c such that $\Delta G/a < 0$. The greater the maximal value in c, the more the contributions. Because of the uncertainty in c, the number n_2 is computed from three different c values.

Pseudoknot formation and stability, as pointed out by Tinoco and co-workers, depends on the thermodynamic contributions from stem size and sequence, and loop size and sequence (Wyatt *et al.*, 1990). The calculations of (7)–(9) are used as supportive evidence for the selection of a particular pseudoknot over various possibilities. They may not vindicate the existence of the pseudoknot. However, in our view, as the hairpin is both significant and stable, and has small n_1 and n_2 values ($<5\%$ of the randomized sequences), a large $|z|$ value (say, $|z| > 10.0$) and a $\Delta G < 0$ might be an indication for the formation of that pseudoknot.

Martinez (1990) has extended his program RNAFOLD to give it the capability of finding pseudoknots. It selects the one that has negative free energy with the smallest connecting loop (L_2) as the pseudoknot for a hairpin (constrained to have no bulges or inner loops) under consideration. The energy of a pseudoknot is calculated using the same base-pairing, stacking and destabilization energies as for orthodox stems and hairpins. The algorithm, developed by Abrahams *et al.* (1990) for prediction of RNA secondary structure, is similar to the one reported by Martinez (1984) in which one stem after the other is added to a nascent structure. During the simulation, the RNA is allowed to fold into a pseudoknot structure. The energy contribution of the stem formed by pseudoknot interaction is treated as the contribution from a planar interaction. Also, the two connecting loops with <15 nucletodies are assumed to have the same energy contribution, 4.2 kcal/mol, and to be independent of their loop sizes. The program of Abrahams *et al.* predicted three out of five possible pseudoknots at the 3′ terminal non-coding region of TMV. Our algorithm is quite different. First, we localize unusual folding regions that are both highly significant and stable in the sequence. A large number of stem–loop structures are filtered and the search for potential pseudoknots confined in these distinct folding regions, which are likely to be involved in biological functions in the sequence. Secondly, the selection of the pseudoknot in a distinct folding region is based on evaluation from an extensive Monte Carlo simulation.

In this report, we have demonstrated that a number of functional pseudoknots that were reported to be closely associated with the protein recognition and translational frameshift can be identified and predicted from their sequence data.

Acknowledgements

We would like to thank Ruth Nussinov and Bruce Shapiro for their helpful discussion. S.-Y.L. thanks Michael Zuker for his support. Research sponsored, at least in part, by the National Cancer Institute, DHHS, under contract NO1-CO-74102 with Program Resources, Incorporated. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commerical products, or organizations imply endorsement by the US Government.

References

Abrahams,J.P., Van Den Berg,M., Van Batenburg,E. and Pleij,C.W.A. (1990) *Nucleic Acids Res.*, **18**, 3035–3044.

Brierley,I., Digard,P. and Inglis,S.C. (1989) *Cell*, **57**, 537–547.

Clarke,B E., Brown,A.L., Curry,K.M., Newton,S.E., Rowlands,D.J. and Carroll,A.R. (1987) *Nucleic Acids Res.*, **15**, 7067–7080.

Jacks,T., Madhani,H.D., Masiarz,F.R. and Varmus,H.E. (1988) *Cell*, **55**, 447–458.

Le,S.-Y., Chen,J.-H. and Maizel,J.V.,Jr. (1989) *Nucleic Acids Res.*, **17**, 6143–6152.

Le,S.-Y., Malim,M.H., Cullen,B.R. and Maizel,J.V.,Jr. (1990) *Nucleic Acids Res.*, **18**, 1613–1623.

Le,S.-Y., Chen,J.-H. and Maizel,J.V.,Jr (1991) *Comput. Applic. Biosci*, **7**, 51–55.

Malim,M.H., Hauber,J., Le,S.Y., Maizel,J.V. and Cullen,B.R. (1989a) *Nature*, **338**, 254–257.

Malim,M.H., Bohnlein,S., Fenrick,R., Le,S Y., Maizel,J.V. and Cullen,B.R. (1989b) *Proc. Natl. Acad. Sci. USA*, **86**, 8222–8226.

Martinez,H. (1984) *Nucleic Acids Res.*, **12**, 323–334.

Martinez,H. (1990) *Methods Enzymol.*, **183**, 306–317.

McPheeter,D.S., Stormo,G.D. and Gold,L. (1988) *J. Mol Biol.*, **201**, 517–535.

Pleij,C.W.A., Rietveld,K. and Bosch,L. (1985) *Nucleic Acids Res.*, **13**, 1717–1731.

Pleij,C.W.A., Abrahams,J.P., van Belkum,A., Rietveld,K. and Bosch,L. (1987) *Positive Strand RNA Viruses*. Alan R.Liss, New York, pp. 299–316.
Pleij,C.W.A. and Bosch,L. (1989) *Methods Enzymol.*, **180**, 289–303.
Puglisi,J.D., Wyatt,J.R. and Tinoco,I.,Jr (1988) *Nature*, **331**, 283–286.
Rietveld,K., van Poelgeest,R., Pleij,C.W.A., van Boom,J.H. and Bosch,L. (1982) *Nucleic Acids Res.*, **10**, 1929–1946.
Rietveld,K., Pleij,C.W.A. and Bosch,L. (1983) *EMBO J.*, **2**, 1079–1085.
Rietveld,K., Linschooten,K., Pleij,C.W.A. and Bosch,L. (1984) *EMBO J.*, **3**, 2613–2619.
Tang,C.K. and Draper,D.E. (1989) *Cell*, **57**, 531–536.
Turner,D.H., Sugimoto,N. and Freier,S.M. (1988) *Annu. Rev. Biophys. Chem.*, **17**, 167–192.
Van Belkum,A., Abrahams,J.P., Pleij,C.W.A. and Bosch,L. (1985) *Nucleic Acid Res.*, **13**, 7673–7686.
Wyatt,J.R., Puglisi,J.D. and Tinoco,I. (1990) *J. Mol. Biol.*, **214**, 455–470.

Received on July 29, 1991; accepted on November 24, 1991