Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Identification of unusual RNA folding patterns encoded by bacteriophage T4 gene *60*

(RNA structure; mRNA translation; topoisomerase; coding gap bypass)

Shu-Yun Le[a], Jih-Hsiang Chen[b] and Jacob V. Maizel Jr.[a]

[a]*Laboratory of Mathematical Biology, Division of Cancer Biology and Diagnosis Centers, National Cancer Institute, NIH, Frederick, MD 21702, USA;* and [b] *Biomedical Supercomputing Center, Program Resources Inc., DynCorp, NCI/FCRDC, Frederick, MD 21702, USA. Tel. (301)846-5773*

SUMMARY

A 50-nucleotide (nt) untranslated region (coding gap sequence) that interrupts the amino acid coding sequence in T4 gene *60*, plus an additional 5 nt upstream and another 3 nt downstream from the gap sequence, shows unusual folding patterns according to RNA structure prediction. A predicted highly stable and significant hairpin structure in the 5′ half of the gap sequence and a plausible tertiary structural element computed in the 3′ part of the gap sequence seem significant by statistical tests on the wild-type (wt) sequence. This feature is absent in insertion, deletion and substitution variants of the gap sequence, in which template activities are markedly lower than that of the wt. The proposed feature is consistent with currently available data showing that the translational bypass of the coding gap is correlated with a stop codon involved in a stem-loop structure folded in the gap sequence. We suggest that the role of this segment in 'ribosomal bypass' of a portion of the mRNA sequence is a property of its special folded structure.

INTRODUCTION

It has been reported by Huang et al. (1988) that a 50-nt gap sequence interrupting aa coding region occurs within the coding sequence of bacteriophage T4 DNA topoisomerase-encoding gene *60*. Weiss et al. (1990) proposed a mechanism of translational bypass by ribosomal jumping across this segment during the translation of the 18-kDa subunit of bacteriophage T4 DNA topoisomerase. Efficient bypass of the untranslated segment requires the nascent peptide and a stop codon involved in a stem-loop structure.

Correspondence to: Dr. S.-Y. Le, Laboratory of Mathematical Biol., DCBDC, NCI, NIH, Bldg. 469, Rm. 151, Frederick, MD 21702, USA. Tel. (301)846-5576; Fax (301)846-5598; e-mail: shuyun@fcrfvl.ncifcrf.nci.gov

Abbreviations: aa, amino acid(s); bp, base pair(s); nt, nucleotide(s); SD, standard deviation; Sigscr, significance score; Stbscr, stability score; wt, wild type.

We know that intrinsic RNA components can be involved in the mediation of replication (Esteban et al., 1989), transcription (Landick et al., 1990; Feng and Holland, 1988; Malim et al., 1989a; Muesing et al., 1987), translation (Edery et al., 1989; Kozak, 1986; Starzyk, 1987) and ribosomal frameshifting (Brierley et al., 1989; Jacks et al., 1988; Moore et al., 1987). All of these functions require different RNA structures. It has been suggested that the role of these special regions involves RNA folding. At present, only a few X-ray crystal structures of small tRNA molecules are known at high resolution (Kim et al., 1974; Sussman et al., 1978; Westhof et al., 1985). Although much of the available structural information for large RNA molecules is crude (e.g., electron micrograph, cleavage by various enzymes at helical or single-stranded regions, and mutational studies), these data support the existence of folded structures and provide constraints for the determination of specific secondary or tertiary structures. Methods for precise prediction of RNA 3-dimensional structures by RNA folding are not

yet available. In this work we show that it is practical to find correlations between predicted thermodynamic stability of RNA folding and RNA function without knowledge of the precise molecular structures involved.

Methods for efficiently identifying distinct folding patterns of RNA secondary structure and/or tertiary structure that are significantly more stable than expected by chance have been developed (Chen et al., 1992; Le et al., 1990a,b). Several successful applications of the methods in viral systems show that the predicted, unusual folding patterns correlate with RNA functional elements of human immunodeficiency virus (Fenrick et al., 1989; Malim et al., 1989a,b). Theoretical analyses of significant folding patterns have helped experimental designs for the identification of biological functional elements in viral RNAs (Tiley et al., 1990).

The aim of present study was to determine the distribution of stable RNA folding regions in the mRNA sequence of bacteriophage T4 DNA topoisomerase gene 60. The thermodynamic stability and statistical significance of RNA folds in the sequence were assessed using independent random models and extensive Monte Carlo simulations. Our results indicate that both the most stable and the most significant segment in the sequence coincides with the untranslated gap sequence within the aa coding region of T4 gene 60. We attempted to analyse possible higher order structures formed in the distinct unusual folding region by our new method (Chen et al., 1992) and to detect significant tertiary interactions. These structural features are not predicted for insertion, deletion, and substitution variants of the gap sequence described in Weiss et al. (1990). Thus, correlation of predicted structure with functionality implies an important role for RNA highly order structure in efficient ribosomal bypass of the mRNA sequence.

RESULTS AND DISCUSSION

(a) Unusual folding patterns in bacteriophage T4 DNA topoisomerase-encoding gene 60

Using the basic premise that some functional regions of RNAs involve a distinct folding or an extensive RNA structure, we set out to find those folding patterns that are unlikely to occur by chance and are highly stable relative to other possible folding segments in the sequence. Figs. 1 a to c show the distributions of thermodynamic stability (Stbscr) and statistical significance (Sigscr) of local RNA folding in bacteriophage T4 DNA topoisomerase gene 60. Calculations were done by choosing successive 50, 70, 100 and 150 nt 'windows' along the mRNA transcript from the 5' to 3' direction (the profile of window 100 nt is not depicted in Fig. 1). Fig. 1a
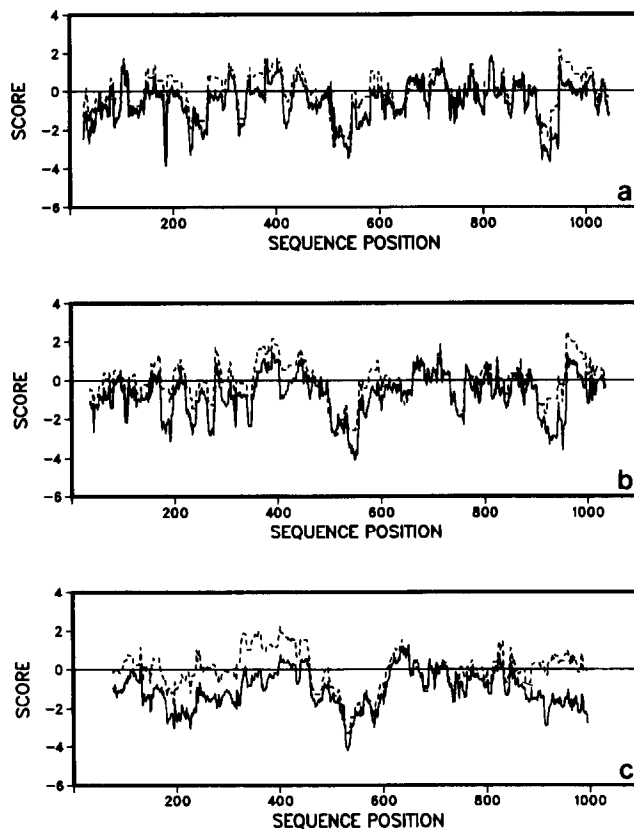

Fig. 1. Distributions of significance and stability scores in the bacteriophage T4 gene 60 sequence. (a) Profile for window length of 50 nt. (b) Window size is 70 nt and (c) window size is 150 nt. The map was obtained by plotting the Sigscr (continuous curve) and Stbscr (broken curve) of each segment against the position of the middle nt in the segment along the sequence. The Sigscr was calculated by dividing the difference between the actual and mean random stabilities by the SD of the random stabilities (Le and Maizel, 1989). The Stbscr was calculated by dividing the difference between the stability of the specific segment and the mean stability of all overlapping segments by the SD of the stabilities of the overlapping segments (Le et al., 1990a). The stability of RNA folding was computed using the Tinoco energy rules (Cech et al., 1983).

(window: 50 nt) presents two unusual folding domains, domain A (nt 514 to 565) and B (nt 901 to 956) in which Sigscr is less than −3.1 and Stbscr is less than −2.0. Among them, domain A is the only distinct folding region detected in other profiles of window lengths of 70, 100 and 150 nt. Domain A extends 162 nt from nt 451 to 612 in the sequence (Table I).

In addition to both highly significant and stable folding domains, two potentially significant folding patterns (nt 160 to 211 and 889 to 997) were also observed (Table I). The two folding patterns were more stable than random ones, but were not more stable than other segments in the sequence. Three potentially 'open' regions that were significantly less stable than shuffled sequences and less stable relative to other folding segments in the T4 gene 60 were identified at positions from 355 to 428, 695 to 746 and 790 to 845 (Table II). These significant non-

TABLE I

Significant runs of unusual folding pattern detected in bacteriophage T4 gene 60

Window[a] (nt)	Fraction of unusual patterns in sequence[b]	Length of run[c] (nt)	Region in the sequence	Mean of Sigscr[d]	Mean of Stbscr[d]	Mean energy of each nt[e] (kcal/mol)
50	0.019	3	160-211	−3.37	−1.93	−0.29
50	0.019	3	514-565	−3.38	−2.42	−0.32
50	0.019	7	901-956	−3.44	−2.41	−0.32
70	0.025	18	502-588	−3.61	−2.52	−0.33
100	0.040	7	156-261	−3.54	−1.67	−0.28
100	0.040	16	495-609	−3.57	−2.89	−0.34
100	0.040	4	512-614	−3.34	−1.95	−0.29
100	0.040	10	889-997	−3.47	−0.54	−0.23
150	0.014	13	451-612	−3.80	−3.13	−0.33

[a] Segment length used in the Monto Carlo simulation.

[b] The fraction of the unusual pattern in the sequence was calculated by dividing the numbers of the significant folding segment identified by the total number of overlapping segments generated as window slides along the sequence from 5' to 3'. The unusual pattern was selected if the Sigscr of the segment is less than −3.1.

[c] The minimum lengths for a run of the unusual folding pattern which gives significance at the 0.01 level are 3, 3, 4 and 3 for window lengths of 50, 70, 100 and 150 nt, respectively. They were calculated using the analytical formula (Karlin et al., 1990): $\ln(1-0.01)/(f-1)/\ln(f)-\ln(N)/\ln(f)$, where f is the fraction of the unusual folding pattern detected in the sequence.

[d] The means of Sigscr and Stbscr were calculated by averaging the significance scores and stability scores of the overlapping segments in the region.

[e] The mean energy of each nt was computed by dividing the averaged energy of the overlapping segment in the region by the segment length. In the calculation, the Tinoco energy rules (Cech et al., 1983) were used.

TABLE II

Significant runs of open regions with little intramolecular base pairings in bacteriophage T4 Gene 60[a]

Window (nt)	Fraction of unusual patterns in sequence[b]	Length of run[c] (nt)	Region in the sequence	Mean of Sigscr	Mean of Stbscr	Mean energy of each nt (kcal/mol)
50	0.013	3	695-746	1.62	1.48	−0.048
50	0.013	7	790-845	1.67	1.66	−0.035
70	0.010	5	355-428	1.31	2.14	−0.054

[a] See footnotes a–e in Table I.

[b] The unusual open pattern with little intramolecular base pairing was selected if the Sigscr of the folding region is greater than 1.5 and folding energy greater than −2.5 kcal/mol for window size of 50 nt, and Sigscr greater than 1.0 and energy greater than −5.0 kcal/mol for windows of 70 nt, respectively.

[c] For a run of the open pattern, the minimum lengths which give significance at the 0.01 level is 3 for window lengths of 50, and 70.

bonded or weakly bonded segments are suspected to be vulnerable to single-stranded specific enzymes or resistant to double-stranded activities.

The exact location of the most distinct folding pattern in domain A was found using extensive simulations in the region from nt 300 to 900. In the simulation, the window length varied from 40 nt to 300 nt using 2 nt increments. The results were summarized in Fig. 2. The highlighted folding pattern in domain A that was both the most significant and the most stable relative to other possible foldings occurred at nt 515 to 572 (Sigscr −5.19, Stbscr −3.64) using a window of 58 nt. This unusual folding segment includes all of the 50 nt gap sequence and an additional 5 nt upstream and 3 nt downstream

from the gap sequence. Further studies indicate that the unusual folding segment is also the most stable and the most significant in the entire sequence for segment lengths of 40 to 300 nt.

The same procedure was applied to variants of the gap sequence that allowed gap bypass in a low level (from 2.8% to 39% relative to wt sequence) (Weiss et al., 1990). For the three substitution mutant sequences of BA9-2, BA9-3 and BA9-4, the most significant folding regions identified were not located at the gap sequence, though the most stable folding regions were detected in the gap sequence (see Fig. 3). The results indicate that the high structural stability of these three substituted variants of the gap sequence survived, but they are not the most
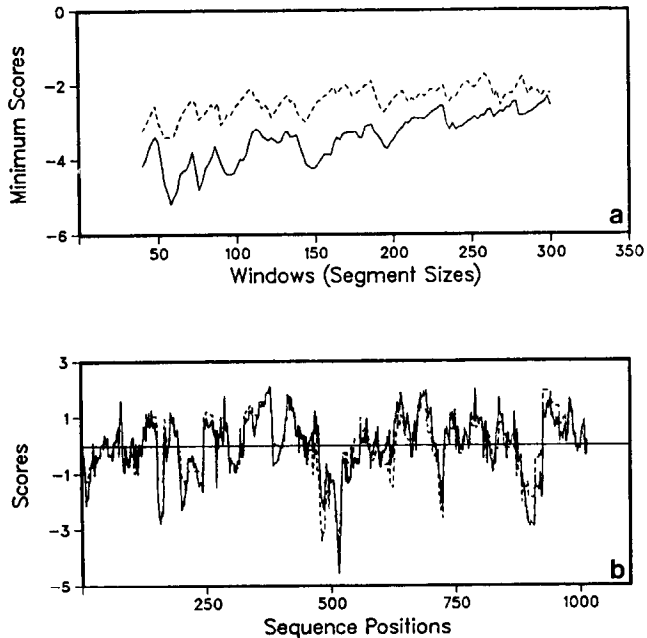
Fig. 2. Unusual RNA folding regions in the bacteriophage T4 gene *60* sequence. (a) Distributions of minimal Sigscr (continuous curve) and minimal Stbscr (broken curve) in the domain (nt 300 to 900) of T4 gene *60* sequence under different window sizes. The calculation was carried out repeatedly by increasing two nt to the window from 40 to 300 nt. For each window size, the calculation was carried out by sliding one nt along the sequence and the minimum scores were then plotted against the window size. The global minima of significance and stability scores were detected at window length of 58 nt. (b) Distributions of the Sigscr (continuous curve) and Stbscr (broken curve) in the T4 gene *60*. The map of both two scores was plotted against the position of the first nt of each segment along the sequence. In the calculation the window size was taken as 58 nt. The sequence positions of global minima of two curves are at the same position of 515, i.e., the highlighted unusual folding segment is located from position 515 to 572. A region that is both highly stable and significant (having lower Stbscr and Sigscr) is referred to as an unusual folding region and is assumed to have a significant folded form, implying a structural role for the sequence information (Le et al., 1990a,b).

significant folding region in the entire sequence. Similar results were also obtained on the insertion (X46a) and deletion (XA1 and XA3) of mutant sequences in the gap sequence. On the variant X46a, 16 nt were inserted in the gap sequence. Again the most significant and most stable folding region was not found in the 66 nt gap sequence.

For domain B, a significant (Sigscr -3.67) and highly stable region (Stbscr -3.35) was found in nt 907 to 952 (window 46 nt) using the same procedure. Other significant folding patterns occur at nt 892 to 939 (window 48 nt, Sigscr -3.81, Stbscr -2.64) and 915 to 988 (window 74 nt, Sigscr -4.19, Stbscr -1.63).

(b) Possible RNA structures in identified unusual folding regions

All RNA secondary structural models were computed by the suboptimal structure method (Jaeger et al., 1989;

Zuker, 1989). In this calculation, all possible suboptimal secondary structures within 20% or 60% of the computed lowest free energy ('20-optimal' or '60-optimal' suboptimal structures) were derived (the minimum pairwise distance criterion was set to zero). For the highlighted segment of nt 515 to 572, 7 and 23 secondary structures were predicted to be within 10% and 20% of the minimum free energy. Two possible secondary structure models that consist of frequently recurring helices in the suboptimal folding are depicted in Fig. 4 a-b. The lowest free energies of the two secondary structures derived from Turner energy rules (Turner et al., 1987; Jaeger et al., 1989) are the same (-15.7 kcal/mol). Among them, hairpin I is conserved in all 23 suboptimal secondary structures. Hairpin I is identical with the structure proposed in the ribosome-jumping model previously (Weiss et al., 1990).

Moreover, hairpin I is also highly conserved even in 201 suboptimal structures computed within 60% of minimum free energy. However, a number of alternative suboptimal RNA folding structures can occur in the 3' half of the distinct RNA segment. Hairpins II and III are more stable than others that occur in the 3' half. It is noteworthy that the two stem-loops (hairpins II and III) can form with the same lowest free energy (-1.7 kcal/mol). There is a possibility that the unusual folding pattern at region nt 515 to 572 is dynamic and allows the formation of multiple structures in its 3' half.

RNA tertiary interactions based on the proposed secondary structures were searched and evaluated using two scores, n_1 and n_2, by our RNAKNOT program (Chen et al., 1992). The two scores are defined as the number of randomized sequences that have tertiary interactions thermodynamically more stable than the actual segment. n_1 and n_2 differ in that the free energy contributed by tertiary interactions is computed using different simulation rules. Three tentative tertiary structural elements were derived. Among them, two pseudoknots involving hairpin III and base-pairing between 5'-UUAA (5'-UAA) in loop and 5'-UUGG (5'-UUA) in the subsequence of 3' downstream from the hairpin III had large or moderately large values of $n_1 = 230$ and $n_2 = 342$ ($n_1 = 0$ and $n_2 = 76$) in a statistical test consisting of 1000 randomly shuffling sequences. The third one was composed of hairpin III and tertiary interactions between the 5'-AUU in the hairpin loop and upstream nt of AAU-3'. The tertiary structure had $n_1 = 0$ and $n_2 = 0$ signifying that the tertiary structure (see Fig. 4c) was favored to occur in the gap sequence.

Structural analyses of seven mutant derivatives of the gap sequence show that a moderately stable stem-loop structure can form in the 3' part of the gap sequence, while the base-pairing structure can be different in each
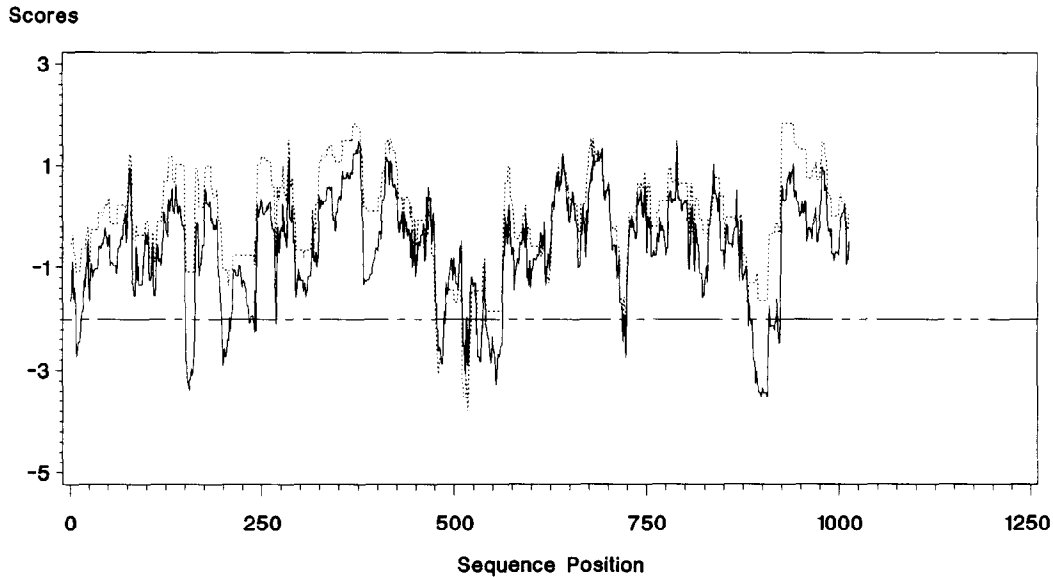
Scores

Fig. 3. Distributions of the statistical significance (continuous curve) and thermodynamic stability scores (broken curve) in the substitution variant of BA9–2. The nt 550 to 566 (5'-AUAAUUAAGAGAUUAUU) in the gap sequence of T4 gene *60* was substituted by 5'-CCCCCAUGCUCACCGGC (Weiss et al., 1990). The map of both two scores was plotted against the position of the first nt of each segment along the sequence. In the profile the window size was taken as 58 nt. The highlighted unusual folding segment detected in wt (see Fig. 2b) cannot be identified in variant BA9–2. Similar results for variants BA9–3, BA9–4, X46a, XA1, and XA3 are not shown.
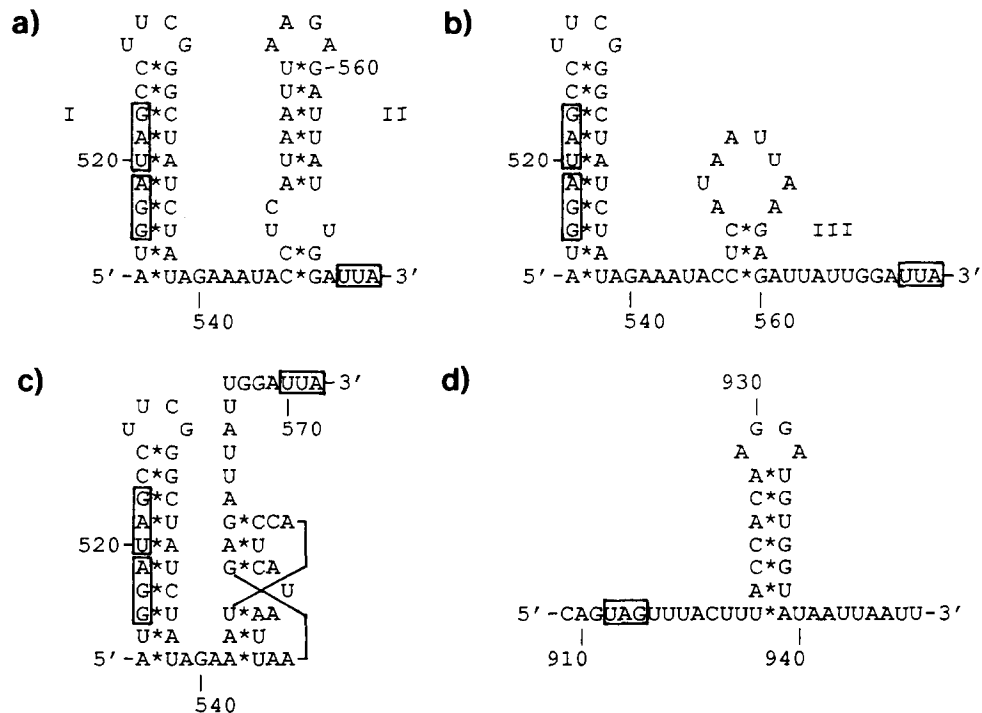
Fig. 4. Predicted RNA structures folded in the highlighted folding patterns of T4 gene *60*. The gap sequence is located at nt 520 to 569. The base-pairing in the structure is denoted by stars. (a) and (b) Two alternative RNA secondary structure models of the unusual RNA folding region (nt 515 to 572). Their lowest free energies are both -15.7 kcal/mol. Three distinct hairpins are denoted by I to III. The codon and stop codon are boxed. (c) Conventional representation of the tertiary structure model of the unusual RNA folding region (nt 515 to 572). (d) Highly conserved RNA hairpin (nt 922 to 939) predicted in the three unusual folding segments (nt 892 to 939, 907 to 952 and 915 to 988) in the 3' transcript of T4 gene *60*. The stop codon of T4 gene *60* is boxed and located at nt positions 912 to 914.

mutant sequence. The stem-loop structures folded in variants BA9–2, BA9–3, BA9–4, XA1, and X46a are more stable than that folded in the wt. The distinct tertiary interactions detected in the wt sequence, however, could not be formed in the substitution variants BA9–2, BA9–3, BA9–4, deletion mutants XA1, XA3, XA7 and insertion mutant X46A of the gap sequence (Weiss et al., 1990). We speculate that the special configuration of the tertiary

structure folded in the 3' part of the gap sequence is important.

The significant RNA secondary structures near the 3' terminal of the T4 gene *60* were predicted in the same way. In the calculation, 12, 5, and 7 suboptimal RNA secondary structures were found within 20-optimal suboptimal folding in regions of nt 915 to 988, 892 to 939 and 907 to 952, respectively. The hairpin of 7 bp (nt 922 to 939) depicted in Fig. 4d is conserved in all predicted RNA secondary structures folded in the three overlapping segments. The hairpin structure is both highly stable and significant. This distinct hairpin is situated just 3' to the stop codon (nt 912 to 914) for T4 gene *60*. It shows similar features to the prokaryotic transcription termination signals, a RNA hairpin followed by polyuridine stretches (Landick et al., 1990). We suggest that the highly significant and stable RNA hairpin may cause RNA polymerase to pause in the transcription of T4 gene *60*.

(c) Possible function of gap sequence

The results presented suggest that a 50-nt gap sequence in the aa coding region of T4 DNA topoisomerase gene *60* is involved in a highlighted RNA folding pattern. The identified 58-nt segment also contains 5 nt upstream and 3 nt downstream from the gap sequence. It has been demonstrated that the unusual RNA segment can fold into a both highly stable and significant higher order structure. The detailed analysis of the RNA secondary structures in the segment shows that three stem-loop structures (see Fig. 4 a-b) are the most frequently recurrent in the RNA folding of this RNA segment. Moreover, the significant tertiary structural element including hairpin III (Fig. 4c) possibly form in the 3' part of the gap sequence.

Similar cases occur in the ribosomal frameshifting of many retroviruses (Brierley et al., 1991; ten Dam et al., 1990; Le et al., 1989; 1991; Chamorro et al., 1992). In retroviral frameshifting, a highly stable and statistically significant stem-loop structure with ability to form a pseudoknot occurs just 3' to the frameshift site. In the absence of this distinct RNA tertiary structural element much less ribosomal frameshifting occurs (Brierley et al., 1991; Chamorro et al., 1992). The mechanism by which the pseudoknot promotes frameshifting is not yet clear. Jacks et al. (1988) suggested that a significantly stable structure may stall elongating ribosomes so that they slip back one nt and pair with the codon in the −1 frame. Furthermore, Brierley et al. (1991) showed that the contribution of the pseudoknot is more dependent on its specific conformation rather than its energetic stability.

In the T4 gene *60* case, a stop codon at the 5' junction of the gap sequence is involved in the highly stable and statistically significant hairpin I. The occurrence of the

hairpin structure is not random. The significance and stability scores of this unusual folding pattern are less than −5.1 and −3.6 SD unit. A possible function of the distinct stem-loop may be to help the stop codon in the coding region to stall ribosomal translation entirely, where upon it promotes dissociation of the peptidyl-tRNA from codon GGA (nt 517 to 519) with the interaction between the protein and gap sequence. In this case the thermodynamic stability of the hairpin structure is important. Using mutational analysis, Weiss et al. (1990) demonstrated that the disruption of the hairpin structure can result in 10- to 30-fold decrease in ribosomal bypass in the translation of T4 gene *60*. Also, the effective gap bypass can be recovered by restoring the hairpin with compensatory base changes. The important role of the hairpin structure in the translational bypass of the gap sequence was confirmed in the bacteriophage T4 gene *60*. This structure alone, however, is not sufficient for ribosomal bypass.

In the 3' half of the gap sequence, the most frequently recurrent structures are hairpins II and III. They have the same moderately stable free energy (-1.7 kcal/mol). For five variants of the gap sequence the folded structures are more stable than that of the wt, but ribosomal bypass in these variants decreases to the level of 6.6% to 39% . Thus, the thermodynamic stability of RNA secondary structure folded in the 3' part of the gap sequence is not a key factor by itself.

The structural feature of the proposed tertiary structural element including hairpin III (Fig. 4c) is similar to classical pseudoknots described by Pleij and Bosch (1989). This tertiary structure type was also suggested as a RNA pseudoknot in the rRNA intervening sequence of *Tetrahymena* (Kim and Cech, 1987). The mediation of mRNA translation through specific recognition between proteins and RNA pseudoknots was also demonstrated in prokaryotic systems (McPheeters et al., 1988; Portier et al., 1990; Tang and Draper, 1989). Weiss et al. (1990) reported that the 3' half of the sequence also correlates with ribosomal bypass of the coding gap. Deletions of 4 (nt 543 to 546, mutant XA1), 12 (nt 538 to 549, XA3) and 26 nt (nt 541 to 566, XA7) in the 3' half of the gap sequence decreased translational activities to 31%, 27% and 2.8%, respectively. The insertion of 16 nt (variant X46a) and substitutions of 17 nt from nt 550 to 566 (BA9-2, BA9-3 and BA9-4) also caused a low gap bypass. All these variants mentioned above would eliminate or diminish the tentative tertiary interaction. It suggests that the special configuration of the significant tertiary structure formed in the 3' half of the gap sequence is important for the ribosomal bypass. If this tertiary structure functions in the coding gap bypass, we deem that the interaction between proteins and RNA may de-

stabilize the association of peptidyl-tRNA and codon GGA (nt 517 to 519) and promote the peptidyl-tRNA binding to the next GGA (nt 567 to 569) by skipping the untranslated region. The higher order RNA structure proposed in this paper may offer a binding target for proteins in a similar way to the process of autoregulation of S4-α mRNA expression (Tang and Draper, 1989). The results presented here add to the body of data showing a correlation of unusual structural properties detected by computer analysis with important regulatory functions in biological systems.

REFERENCES

Brierley, I., Digard, P. and Inglis, S.C.: Characterization of an efficient coronavirus ribosomal frameshifting signal, requirement for an RNA pseudoknot. Cell 57 (1989) 537–547.
Brierley, I., Rolley, N.J., Jenner, A.J. and Inglis, S.C.: Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. J. Mol. Biol. 220 (1991) 889–902.
Cech, T.R., Tanner, N.K., Tinoco Jr., I., Weir, B.R., Zuker, M. and Perlman, P.S.: Secondary structure of the Tetrahymena ribosomal RNA intervening sequence: structural homology with fungal mitochondrial intervening sequences. Proc. Natl. Acad. Sci. USA 80 (1983) 3903–3907.
Chamorro, M., Parkin, N. and Varmus, H.E.: An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. Proc. Natl. Acad. Sci. USA 89 (1992) 713–717.
Chen, J.-H., Le, S.-Y. and Maizel Jr., J.V.: A procedure for RNA pseudoknot prediction. CABIOS 8 (1992) 243–248.
ten Dam, E.B., Pleij, C.W.A. and Bosch, L.: RNA pseudoknots, translational frameshifting and readthrough on viral RNAs, Virus Genes. 4 (1990) 121–136.
Edery, I., Petryshyn, R. and Sonenberg, N.: Activation of double-stranded RNA-dependent kinase (dsI) by the TAR region of HIV-1 mRNA: a novel translational control mechanism. Cell 56 (1989) 303–312.
Esteban, R., Fujimura, T. and Wickner, R.B.: Internal and terminal cis-acting sites are necessary for in vitro replication of the L-A double-stranded RNA virus of yeast. EMBO J. 8 (1989) 947–954.
Feng, S. and Holland, E.C.: HIV-1 tat trans-activation requires the loop sequence within tar. Nature 334 (1988) 165–167.

Fenrick, R., Malim, M.H., Hauber, J., Le, S.-Y., Maizel Jr., J.V. and Cullen, B.R.: Functional analysis of the Tat trans activator of human immunodeficiency virus type 2. J. Virology 63 (1989) 5006–5012.
Huang, W.M., Ao, S.-Z., Casjens, S., Orlandi, R., Zeikus, R., Weiss, R., Winge, D. and Fang M.: A persistent untranslated sequence within bacteriophage T4 DNA topoisomerase gene 60. Science 239 (1988) 1005–1012.
Jacks, T., Madhani, H.D., Masiarz, F.R. and Varmus, H.E.: Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. Cell 55 (1988) 447–458.
Jaeger, J.A., Turner, D.H. and Zuker, M.: Improved prediction of secondary structures for RNA. Proc. Natl. Acad. Sci. USA 86 (1989) 7706–7710.
Karlin, S., Blaisdell, B.E. and Brendel, V.: Identification of significant sequence patterns in proteins. Methods Enzymol. 183 (1990) 388–402.
Kim, S.-H., Suddath, F.L., Quigley, F.L., McPherson, A., Sussman, J.L., Wang, A.H.J., Seeman, N.C. and Rich, A.: Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. Science 185 (1974) 435–440.
Kim, S.-H. and Cech, T.R.: Three-dimensional model for the active site of the self-splicing rRNA precursor of Tetrahymena. Proc. Natl. Acad. Sci. USA 84 (1987) 8788–8792.
Kozak, M.: Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. Proc. Natl. Acad. Sci. USA 83 (1986) 2850–2854.
Landick, R., Yanofsky, C., Choo, K. and Phung, L.: Replacement of the Escherichia coli trp operon attenuation control codons alters operon expression. J. Mol. Biol. 216 (1990) 25–37.
Le, S.-Y. and Maizel Jr., J.V.: A method for assessing the statistical significance of RNA folding. J. Theor. Biol. 138 (1989) 495–510.
Le, S.-Y., Chen, J.-H. and Maizel Jr., J.V.: Thermodynamic stability and statistical significance of potential stem-loop structures situated at the frameshift sites of retroviruses. Nucleic Acids Res. 17 (1989) 6143–6151.
Le, S.-Y., Chen, J.-H. and Maizel Jr., J.V.: Efficient searches for unusual folding regions in RNA sequences. In: Sarma, R.H. and Sarma, M.H. (Eds.), Structure and Methods, Human Genome Initiative and DNA Recombination, Vol. 1, Adenine Press, Schenectady, NY, 1990a, pp. 127–136.
Le, S.-Y., Malim, M.H., Cullen, B.R. and Maizel Jr., J.V.: A highly conserved RNA folding region coincident with the Rev response element of primate immunodeficiency viruses. Nucleic Acids Res. 18 (1990b) 1613–1623.
Le, S.-Y., Shapiro, B.A., Chen, J.-H., Nussinov, R. and Maizel Jr., J.V.: RNA pseudoknots downstream from the frameshift sites of retroviruses. Gen. Anal.: Tech. Appl. 8 (1991) 191–205.
Malim, M.H., Bohnlein S., Fenrick, R., Le, S.-Y., Maizel Jr., J.V. and Cullen, B.R.: Functional comparison of the Rev trans-activators encoded by different primate immunodeficiency virus species. Proc. Natl. Acad. Sci. USA 86 (1989a) 8222–8226.
Malim, M.H., Hauber, J., Le, S.-Y., Maizel Jr., J.V. and Cullen, B.R.: The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature 338 (1989b) 254–257.
Moore, R., Dixon, M., Smith, R., Peters, G. and Dickson, C.: Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus: two frameshift suppression events required for translation of gag and pol. J. Virol. 61 (1987) 480–490.
McPheeters, D.S., Stormo, G.D. and Gold, L.: Autogenous regulatory site on the bacteriophage T4 gene 32 messenger RNA. J. Mol. Biol. 201 (1988) 517–535.
Muesing, M.A., Smith, D. and Capon, D.: Regulation of mRNA accumulation by human immunodeficiency virus trans-activator protein. Cell 48 (1987) 691–701.

28

Pleij, C.W.A. and Bosch, L.: RNA pseudoknots, structure, detection and prediction. Methods Enzymol. 180 (1989) 289–303.

Portier, C., Philippe, C., Dondon, L., Grunberg-Manago, M., Ebel, J.P., Ehresmann, B. and Ehresmann, C.: Translational control of ribosomal protein S15. Biochim. Biophys. Acta 1050 (1990) 328–336.

Starzyk R.M.: A target site for translational regulation. TIBS 12 (1987) 415–416.

Sussman, J.L., Holbrook, S.R., Warrant, R.W., Church, G.M. and Kim, S.H.: Crystallographic refinement. J. Mol. Biol. 123 (1978) 607.

Tang, C.K. and Draper, D.E.: Unusual mRNA pseudoknot structure is recognized by a protein translational repressor. Cell 57 (1989) 531–536.

Tiley, L.S., Brown, P.H., Le, S.-Y., Maizel Jr., J.V. and Cullen, B.R.: Visna virus encodes a post-transcriptional regulator of viral structural gene expression. Proc. Natl. Acad. Sci. USA 87 (1990) 7497–7501.

Turner, D.H., Sugimoto, N., Jaeger, J.A., Longfellow, C.E., Freier, S.M. and Kierzek, R.: Improved parameters for prediction of RNA structure. Cold Spring Harbor Symp. Quant. Biol. 52 (1987) 123–133.

Weiss, R.B., Huang, W.M. and Dunn D.M.: A nascent peptide is required for ribosomal bypass of the coding gap in bacteriophage T4 gene 60. Cell 62 (1990) 117–126.

Westhof, E., Dumas, P. and Moras, D.: Crystallographic refinement of yeast aspartic acid transfer RNA.J. Mol. Biol. 184 (1985) 119.

Zuker, M.: On finding all suboptimal foldings of RNA molecule. Science 244 (1989) 48–52.