

LETTER TO THE EDITOR

# An approximation of loop free energy values of RNA H-pseudoknots

ALEXANDER P. GULTYAEV,<sup>1,2</sup> F.H.D. VAN BATENBURG,<sup>1</sup> and CORNELIS W.A. PLEIJ<sup>2</sup>

<sup>1</sup>Section Theoretical Biology and Phylogenetics, Institute of Evolutionary and Ecological Sciences, Leiden University, 2311 GP Leiden, The Netherlands

<sup>2</sup>Leiden Institute of Chemistry, Gorlaeus Laboratories, Leiden University, 2300 RA Leiden, The Netherlands

## ABSTRACT

**A set of free energy values is suggested for RNA H-pseudoknot loops. The parameters are adjusted to be consistent with the theory of polymer thermodynamics and known data on pseudoknots. The values can be used for estimates of pseudoknot stabilities and computer predictions of RNA structures.**

**Keywords:** helix–coil transitions; pseudoknot loops; RNA structure; RNA thermodynamics

## INTRODUCTION

RNA pseudoknots are widely occurring structural motifs and were shown to be essential for various RNA functions (for reviews, see Pleij, 1994; Deiman & Pleij, 1997). Therefore, elucidation of structural features of pseudoknots and reliable prediction of pseudoknotting using sequence data are very important for understanding structure–function relationships in many RNA molecules. While thermodynamic parameters for the structural elements of “classical” RNA secondary structure are measured or estimated with reasonable accuracy (SantaLucia & Turner, 1998; Xia et al., 1998), there is no systematic study available of pseudoknot thermodynamics.

Experimental measurements of thermodynamic parameters for some model pseudoknots have shown them to be only marginally more stable than the secondary structures involved and to be strongly dependent on ionic environment (Wyatt et al., 1990; Theimer et al., 1998). Also, the conformational transitions in pseudoknots may involve intermediate steps, mainly constituent hairpins, so that it is rather difficult to determine a contribution of pseudoknot loops to the stability of structures, especially because any pseudoknot contains at least two loops. For classical RNA secondary structure, the lack of detailed data on loop thermodynamics has been compensated by extrapolations

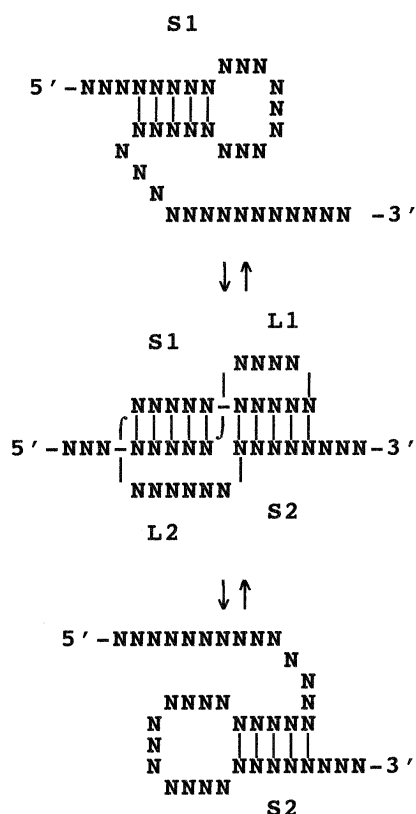
optimized by estimates of success of predictions that used various sets of parameters (Jaeger et al., 1989; Mathews et al., 1998). In the case of pseudoknots, optimization of parameters using computer predictions faces the problem that the most frequently used approach to predict RNA secondary structure by free energy minimization does not include them (Zuker, 1989). Here we present an attempt to estimate the pseudoknot free energy parameters, based on the general theory of polymer loop thermodynamics. The free energy values are adjusted to be consistent with available data on experimentally and/or phylogenetically proven pseudoknots. The suggested set of parameters can be used for estimates of pseudoknot stabilities and in the folding algorithms (e.g., Abrahams et al., 1990; Gultyaev et al., 1995; van Batenburg et al., 1995) that allow pseudoknotting.

## RESULTS AND DISCUSSION

### Assumptions on pseudoknot thermodynamics

The simplest pseudoknot, the classical or so-called H- (hairpin) pseudoknot (Pleij & Bosch, 1989), contains two stems (S1 and S2) and two loops (L1 and L2; Fig. 1). Such a pseudoknot is formed by the pairing of a hairpin loop (closed by S1) with the downstream nucleotides (S2 stem), or, alternatively, by interaction of the upstream nucleotides with the interior of S2 hairpin (S1 pairing). The free energy of an H-pseudoknot structure is mainly the sum of the free energies of stacking in the stems (stabilizing negative values) and the pos-

Reprint requests to: A.P.Gultyaev, Section Theoretical Biology and Phylogenetics, Institute of Evolutionary and Ecological Sciences, Leiden University, Kaiserstraat 63, 2311 GP Leiden, The Netherlands; e-mail: sgultyaev@rulsfb.leidenuniv.nl.



**FIGURE 1.** An equilibrium between an H-pseudoknot and alternate hairpins.

itive destabilizing loop values. The stacking energy can be calculated using the known nearest-neighbor model parameters of helix propagation (SantaLucia & Turner, 1998). For the loop energies some estimate is needed.

A general theory of phase transition in polymers (Jacobson & Stockmayer, 1950; Poland & Scheraga, 1970) allows us to calculate the entropy of a loop of  $N$  statistical units as

$$S(N) = R\{N \ln \Omega - [A + \frac{3}{2} \ln N]\},$$

where  $R$  is the universal gas constant, the constant  $A$  depends on how loop closure is defined, and  $R \ln \Omega$  is the conformational entropy of the free chain per statistical unit. This formula also describes the loop entropies in the helix-coil transitions of nucleic acids (Poland & Scheraga, 1966). If effects of excluded volume are taken into account, the coefficient  $\frac{3}{2}$  should be replaced by 1.75 (Fisher, 1966). With the equation  $\Delta G = \Delta H - T\Delta S$  ( $T$  is the temperature and  $\Delta H$  is the loop formation enthalpy), subtracting the term of free chain and assuming that loops have pure entropic nature ( $\Delta H = 0$ ), the free energy  $\Delta G$  of formation of a loop of  $N$  nucleotides can be approximated by the formula

$$\Delta G = RT(A_{\text{loop}} + 1.75 \ln N),$$

where  $A_{\text{loop}}$  is a constant that is defined by the loop type. Such logarithmic extrapolations are successfully applied for hairpin, internal, and bulge loops in RNA, making it possible to predict the secondary structures by folding algorithms (Jaeger et al., 1989). Although the recursive energy minimization algorithm requires a linear approximation for multibranch loops, the subsequent recalculation of free energies, also involving suboptimal structures, according to the logarithmic size dependence improves the quality of predictions (Walter et al., 1994; Mathews et al., 1998). Here we assumed that the Jacobson–Stockmayer function can be applied for pseudoknot loops as well. To take into account specific features of pseudoknot topology, we made several other assumptions.

1. The loop L1 spans the deep groove of RNA helix S2, whereas the L2 crosses stem S1 in the shallow groove. Therefore, the loops are not equivalent stereochemically and their features depend on the lengths of the corresponding stems (Pleij et al., 1985). This should be taken into account by introducing two variables  $A_{\text{deep}}(S2)$  and  $A_{\text{shallow}}(S1)$  for the loops L1 and L2, respectively.
2. The distances between phosphate atoms connected by the loops along the RNA grooves are minimal at S2 of 6–7 bp and at S1 of 3 bp (Pleij et al., 1985). This is also consistent with frequencies of natural occurrence of stem lengths (Deiman & Pleij, 1997). Therefore, we assumed that  $A_{\text{deep}}$  is minimal for S2 of 6 or 7 bp, and  $A_{\text{shallow}}$  is minimal for S1 = 3 bp.
3. Analysis of pseudoknot geometries (Pleij et al., 1985) also suggests the minimal sizes of loops possible for given stem lengths. In the deep groove, 7 bp can be bridged by a loop of 1 nt only. Bridging over the shallow groove requires at least 2 nt, and the distance to be crossed increases significantly with the length of the stem (Pleij et al., 1985). However, a bending and/or distortion of the RNA A-helical geometry is also possible, so the requirements for big stems may be less rigid. We assumed the minimally allowed size of loops L2 (shallow groove) as 2 nt for an S1 of 3 bp, 3 nt for 4 bp, 4 for an S1 of 5 or 6 bp, and a further increase of 1 nt for each increment of 2 bp. For the deep groove, a loop of 1 nt was allowed for stems of 4–7 bp, and loops of 2 nt for stems of 3 bp or more than 7 bp.
4. Instead of just using a logarithmic increase of entropy with loop size, we introduced the dependence on the difference between the loop length and the minimally allowed length. Although deviating from the dependence on the number of polymer statistical units in the loop (Jacobson & Stockmayer, 1950; Poland & Scheraga, 1970), such an approximation can partially reflect restrictions of conformational freedom imposed by the stem end-to-end distance.

Integrating these assumptions, we get the following two dependences for free energies ( $\Delta G_{37}^\circ$ ) of loops crossing deep and shallow RNA grooves respectively:

$$\begin{aligned}\Delta G_{L1} &= A_{\text{deep}}(S2) \\ &+ 1.75 RT \ln(1 + N - N_{\text{mindeep}}(S2)) \\ \Delta G_{L2} &= A_{\text{shallow}}(S1) \\ &+ 1.75 RT \ln(1 + N - N_{\text{minshallow}}(S1)),\end{aligned}$$

where  $N_{\text{mindeep}}$  and  $N_{\text{minshallow}}$  are functions defining the shortest loops, and the unity is added to make the logarithm equal to zero for the minimal value of  $N$ .

We have restricted ourselves to H-pseudoknots that contain two loops and not more than 1 nt at the junction between stems. For helix–helix interfaces, the coaxial stacking is known to have a strong stabilizing effect, about 1 kcal/mol more than the corresponding nearest-neighbor energy in a regular helix (Kim et al., 1996; Walter & Turner, 1994; Walter et al., 1994). A similar increase of stacking contribution has been suggested for junctions of pseudoknot stems on the basis of mutational analysis of the pseudoknot in the viral tRNA-like structure (Mans et al., 1992). Thus the bonus of 1 kcal/mol for helices without intervening nucleotides and the mismatch values for 1 nt at the junction were assumed to contribute to the free energies of pseudoknots. Such corrections for coaxial stacking were shown to improve predictions of secondary structures (Walter et al., 1994; Mathews et al., 1998). However, extrapolation of data on interfaces to pseudoknots has to be done with caution, because stacking may be influenced by changed twist of base pairs at the junction (Puglisi et al., 1990) or by bending due to a mismatch (Shen & Tinoco, 1995).

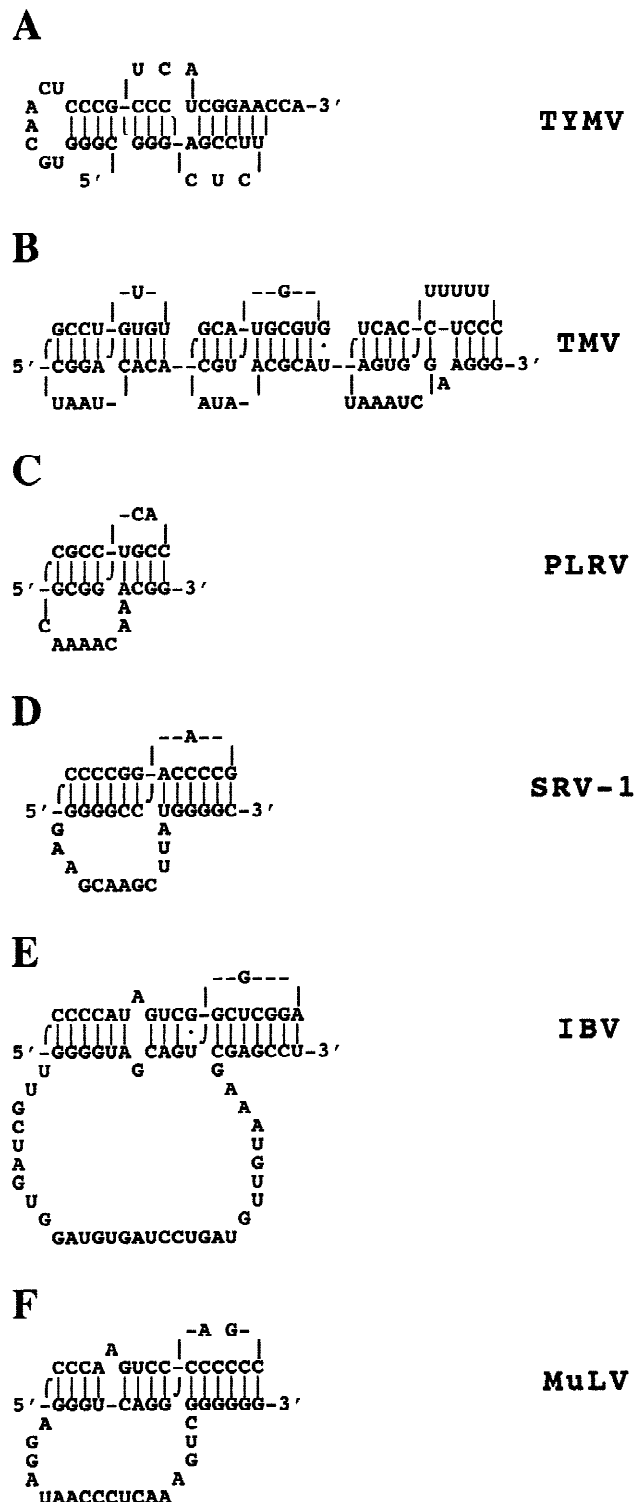
### Estimates of loop energies using proven pseudoknots

For proper estimates of the parameters in the formulas, we used some sequences of known pseudoknots that are evidenced by experiments and/or phylogenetic comparisons, assuming that the free energies of these pseudoknots are lower than those of corresponding hairpins formed by the pseudoknot stems. As seen in Table 1, these H-pseudoknots cover rather broad ranges of loop and stem sizes. The majority of pseudoknots, available for analysis, are found in viral RNAs (ten Dam et al., 1990; Deiman & Pleij, 1997) and representative structures are shown in Figure 2.

The large group of well-documented H-pseudoknots are found at the very 3' ends of plant viral RNAs (tymoviruses and tobamoviruses) as parts of tRNA-like structures (for review, see, e.g., Mans et al., 1991). These pseudoknots seem to have minimal loop sizes and show little variation in stem lengths, with a stem S1 of 3 bp and S2 of 4 bp in tobamoviruses and 5–6 bp in tymoviruses (Fig. 2A). Upstream of the tRNA-like structure in tobamoviruses, a conserved structure of three consecutive pseudoknots is located (Fig. 2B), which seems to be very important in regulation of viral multiplication and mRNA translation (van Belkum et al., 1985; Leathers et al., 1993). Those three pseudoknots or only two are duplicated in some tobamoviruses and satellite tobacco mosaic virus, sometimes with a considerable sequence variation (Gultyaev et al., 1994). Similar pseudoknot stalks are conserved in some hordei-, furo- and tobamoviruses and satellite tobacco necrosis viruses STNV-1 and STNV-2 (Pleij et al., 1986; Danthinne et al., 1991; Solovyev et al., 1996; Koenig et al., 1998), with different numbers of pseudoknots (Table 1). The biggest variation in the sizes of pseudoknot loops and stems is observed in the sites of pseudoknot-dependent

**TABLE 1.** The types of pseudoknot-forming sequences, used for estimates of free energy parameters. Accession numbers are given for representative sequences. The complete list of accession numbers is given in Materials and Methods.

Types	Number	S1 (bp)	L2 (nt)	S2 (bp)	L1 (nt)	Junction (nt)	Accession number
tRNA-like 3' ends of plant viral RNAs	47	3	2–3	4–6	2–3	0	J02415 X16378 J02415
PKstalks:							
PK1	26	3–8	2–6	4–7	1–5	0	
PK2	50	3–4	2–4	5–7	1–3	0	
PK3	48	3–6	2–9	4–8	1–5	0	
Luteoviral frameshift sites	7	4–6	6–9	3–5	1–2	0–1	D00530 Y07496
Retroviral frameshift sites	6	5–6	4–12	3–6	1–5	0–1	M16605
Coronaviral frameshift sites	2	11	30–32	7–11	1–2	0	M27472
Retroviral (type C) readthrough	5	8	17–18	6–7	1–3	0	K01803



**FIGURE 2.** Representative structures of H-pseudoknots, used for the estimates (see also Table 1). **A:** The 3' end tRNA-like structure of turnip yellow mosaic virus RNA. **B:** The pseudoknot domain of the 3'-UTR of tobacco mosaic virus RNA. **C–E:** Ribosomal frameshifting sites in RNAs of the luteovirus potato leafroll virus (PLRV), simian retrovirus-1 (SRV-1), and the coronavirus infectious bronchitis virus (IBV). **F:** Readthrough signal of murine leukemia virus.

ribosomal frameshift and readthrough (ten Dam et al., 1990; Brierley et al., 1991; Chamorro et al., 1992; Garcia et al., 1993; Kujawa et al., 1993; Wills et al., 1994). Although large pseudoknot loops, like those found in frameshift sites of coronaviruses (Fig. 2E), may have an interior structure, we included them in our analysis to test the performance of logarithmic extrapolation.

For each analyzed pseudoknot, the condition of lower free energy than that of either of two hairpins formed separately by stems S1 or S2 (Fig. 1) leads to the requirement that the sum of loop energies be less than some computable value. In these calculations, possible coaxial stacking on stems adjacent to the pseudoknots, as in 3'-terminal tRNA-like structures or pseudoknot stalks (Fig. 2A,B), was also taken into account. Possible extensions of stems, when they are formed separately, were also considered. Also, in case of relatively large pseudoknots (e.g., Fig. 2E,F) we compared the pseudoknot energies with other structures possible for a given region (between the 5' end of S1 and the 3' end of S2). We avoided structure predictions for longer sequences with flanking regions because of uncertainties about structures in these regions. Nevertheless, the inequalities derived from simple requirements for the pseudoknots to be more stable than alternative structures in the same regions yield rough, but reasonable estimates of upper limits of parameters used in the suggested formulas for loop free energies.

On the other hand, in some sequences we noted that an introduction of relatively small values of loop energies could lead to predictions of pseudoknots that are not consistent with the proven structures. Such inequalities estimate lower limits of values. Furthermore, the derived free energy parameters were corrected by comparisons with thermodynamic measurements (Wyatt et al., 1990; Theimer et al., 1998) for some model pseudoknots with relatively small loops (see also below).

Without a claim to be very precise, we suggest the set of free energy ( $\Delta G^\circ_{37}$ ) values given in Table 2, which is consistent with both the suggested formulas and all the pseudoknot data used for the adjustment. This set seems to be the best possible with a lack of systematic experimental results. Presumably it represents the loop energies in 1 M NaCl, because the values are adjusted to the nearest-neighbor parameters, mostly measured at these conditions (see, e.g., Xia et al., 1998). For minimal loop size at optimal stem lengths, the value of 3.5 kcal/mol is suggested (a loop of 1 nt bridging 6 or 7 bp in the deep groove or 2 nt connecting 3 bp in the shallow groove). The logarithmic increment, similar to loop size dependence, was implemented for the energies of the shortest loops crossing the stems with lengths deviating from the optimum [ $\Delta\Delta G = 1.75 RT \ln(1 + |d|)$ , where  $\Delta d$  is the difference between the stem length and the optimal one].



**TABLE 2.** The approximation for free energies ( $\Delta G^{\circ}_{37}$ ) of pseudoknot loops (kcal/mol). Values not given in the table can be extrapolated as described in the text.

stem size (bp)	loop size (nt)										
	1	2	3	4	5	6	8	10	15	20	30
(S2)	deep groove (L1)										
2	—	—	—	—	—	—	—	—	—	—	—
3	—	5.0	5.7	6.2	6.5	6.7	7.1	7.4	7.8	8.2	8.6
4	4.7	5.4	5.9	6.2	6.4	6.6	6.9	7.2	7.6	7.9	8.4
5	4.2	4.9	5.4	5.7	5.9	6.1	6.4	6.7	7.1	7.4	7.9
6	3.5	4.2	4.7	5.0	5.2	5.4	5.7	6.0	6.4	6.7	7.2
7	3.5	4.2	4.7	5.0	5.2	5.4	5.7	6.0	6.4	6.7	7.2
8	—	4.2	4.9	5.4	5.7	5.9	6.3	6.6	7.0	7.4	7.8
9	—	4.7	5.4	5.9	6.2	6.4	6.8	7.1	7.5	7.9	8.3
10	—	5.0	5.7	6.2	6.5	6.7	7.1	7.4	7.8	8.2	8.6
(S1)	shallow groove (L2)										
2	—	—	—	—	—	—	—	—	—	—	—
3	—	3.5	4.2	4.7	5.0	5.2	5.6	5.9	6.3	6.7	7.1
4	—	—	4.2	4.9	5.4	5.7	6.1	6.4	7.0	7.3	7.8
5	—	—	—	4.7	5.4	5.9	6.4	6.8	7.4	7.8	8.3
6	—	—	—	5.0	5.7	6.2	6.7	7.1	7.7	8.1	8.6
7	—	—	—	—	5.2	5.9	6.7	7.1	7.8	8.2	8.7
8	—	—	—	—	5.4	6.1	6.9	7.3	8.0	8.4	8.9
9	—	—	—	—	—	5.6	6.8	7.3	8.1	8.4	9.0
10	—	—	—	—	—	5.7	6.9	7.4	8.2	8.6	9.2

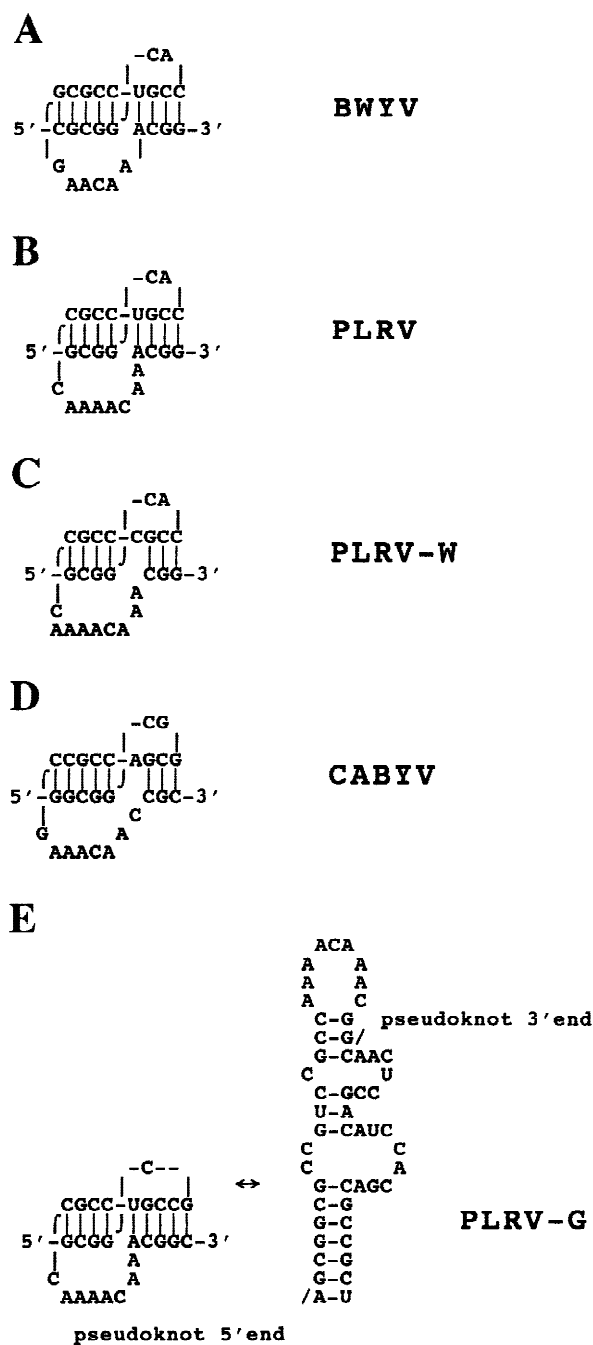
### Predicted stability of natural pseudoknots

The suggested set of parameters predicts that all considered 3'-terminal pseudoknots in the plant viral tRNA-like structures (e.g., Fig. 2A) are more stable than alternative hairpins, with differences in the range of 2–5 kcal/mol. The individual stabilities of consecutive pseudoknots in the stalks (Fig. 2B) are more difficult to determine, because of coaxial stacking contributions between them. Nevertheless, the values given in Table 2 clearly define the pseudoknot stalks as the most stable structures in these regions. The more conserved pseudoknots PK2 and PK3, which are probably more important (Leathers et al., 1993), are mostly more stable as well. However, the 5'-proximal pseudoknots are also predicted to be energetically more favorable than the hairpins formed by S2 stems, although the S1 stems in these pseudoknots are not stabilized by coaxial stacking on the 5'-side.

Ribosomal frameshift sites in the polymerase genes from luteoviruses (Fig. 3) provide an interesting test case for energy parameters. The presence of homologous pseudoknots in the analyzed sequences is supported by nucleotide base–base covariations in both pseudoknot stems. There is also experimental evidence of pseudoknot involvement in the frameshifting signals of beet western yellows virus (BWYV) and potato leafroll virus (PLRV) genes (Garcia et al., 1993; Kujawa et al., 1993). However, in one of the PLRV strains and in cucurbit aphid-borne yellows virus (CABYV), the pseudoknot seems to be destabilized by

a mismatch at the junction (Fig. 3C,D). Also, in a German isolate of PLRV (PLRV-G) an alternative stem-loop structure (Fig. 3E) was suggested (Prüfer et al., 1992). Our set of free energy values (Table 2) predicts that in all sequences, the pseudoknots are more stable than the alternate S1 or S2 hairpins, even in the two cases with mismatches. The alternative structure in the German strain is predicted to be only 3.7 kcal/mol more stable at 25 °C than the pseudoknot. However, its existence may be doubted, because it comprises an additional 24 nt downstream of the pseudoknot and could be disrupted because of competition with other downstream foldings. It is interesting that particularly in this PLRV variant (Prüfer et al., 1992) the pseudoknot is also stabilized by an additional G-C pair in the S2 stem (Fig. 3E).

Pseudoknots at the sites of ribosomal frameshifting and readthrough in animal viruses contain stems with many G-C pairs (Fig. 2D–F) and seem to be very stable. All analyzed pseudoknots are estimated to have considerably lower free energies than the alternate hairpins. In case of readthrough sites from type C retroviruses and frameshifting sites from coronaviruses, both the loops and the stems of pseudoknots are relatively big (Table 1), so that the pseudoknots comprise rather extended RNA regions that provide opportunities for other alternative foldings. We compared the estimated pseudoknot free energies with the predicted stabilities of such structures. It turned out, however, that only one (MoMuLV) of the five different readthrough sequences was folded into a structure with free energy equal to



**FIGURE 3.** The pseudoknots at the sites of ribosomal frameshifting of luteoviruses. **A:** Beet western yellows virus (BWYV). **B:** Potato leafroll virus (PLRV). **C:** PLRV, Wageningen strain; **D:** Cucurbit aphid-borne yellows virus (CABYV). **E:** Alternative folding in the RNA of PLRV, German strain.

that of the pseudoknot. For the coronavirus MHV, an alternative structure was found to be 1.9 kcal/mol more stable than the pseudoknot, whereas the infectious bronchitis virus (IBV) site pseudoknot is estimated to be 1.2 kcal/mol more stable than the closest alternative. It should be noted that these pseudoknots may be further stabilized by interior structures in large L2 loops. Nevertheless, even in the absence of such stabilizing

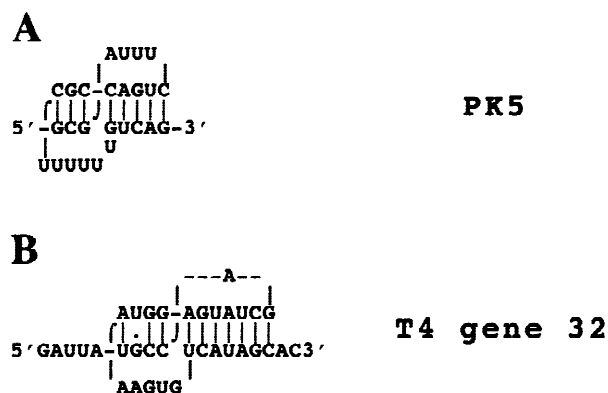
contributions our extrapolation for long loops predicts competitive free energies for the formation of these pseudoknots with loops up to 32 nt, as in the case of the IBV frameshifting site.

### Comparison with experiments

Systematic thermodynamic measurements on RNA pseudoknots with various sequences and/or stem and loop sizes are missing. Moreover, pseudoknot stabilities are known to be strongly dependent on ionic conditions, in particular, on the presence of  $Mg^{2+}$  (Wyatt et al., 1990; Puglisi et al., 1991), so that comparisons of data obtained at different conditions may be misleading. Nevertheless, some qualitative comparisons can be made.

For two model pseudoknot systems (Fig. 4), rather detailed melting experiments are available (Wyatt et al., 1990; Theimer et al., 1998). These studies provide measurements of pseudoknot stabilities by melting as monitored by NMR, UV absorbance and structure probing. For calculation of the pseudoknot melting points predicted by our set of  $\Delta G^{\circ}_{37}$  values for pseudoknot loops (Table 2), we used known enthalpy and entropy values for stacking in the stems and at coaxial stacking (Walter & Turner, 1994). As stated above, we assumed that the pseudoknot loops are purely entropic ( $\Delta H = 0$ ) so that their entropies can be directly calculated from Table 2 ( $\Delta G^{\circ}_{37} = -310 \Delta S$ ). The melting point of a pseudoknot can easily be calculated using the condition that at this temperature  $T_m$  the total pseudoknot free energy ( $\Delta G = \Delta H - T_m \Delta S$ ) is equal to the free energy of the most stable of alternative structures.

For the model pseudoknot PK5 (Fig. 4A), our parameters predict a melting temperature of 67.9°C in 1 M NaCl, which is only 4°C higher than the measured 64°C in 60 mM  $Na^+$  and 5 mM  $Mg^{2+}$  (Wyatt et al., 1990). Thus the suggested values slightly overestimate the



**FIGURE 4.** The model pseudoknots, studied experimentally. **A:** PK5 (Wyatt et al., 1990); **B:** bacteriophage T4 gene 32 pseudoknot (Theimer et al., 1998).

pseudoknot stability. It should be noted, however, that this relatively small discrepancy may be determined by reasons other than deficiencies in loop parameters, because actually the measured enthalpy of the pseudoknot formation ( $-50$  kcal/mol) is higher than the predicted sum of stacking contributions in the quasi-continuous helix ( $-72$  kcal/mol). It was suggested that this could be due to a distortion in stacking or positive enthalpic contributions of the loop regions (Wyatt et al., 1990) that can be sequence specific and are not taken into account by our approximation. It is interesting, however, that our parameters predict a single transition for the melting of this pseudoknot, with the melting point higher than those of both alternative hairpins, which is consistent with experiments (Wyatt et al., 1990).

A two-transition sequential pseudoknot melting was clearly detected for the pseudoknot (Fig. 4B) that is involved in the translational regulation of bacteriophage T4 gene 32. Using a 35-nt fragment, the melting profiles were monitored by measurements of UV absorbance at two wavelengths (280 and 260 nm), taking advantage of the spectroscopic silence of the pseudoknot-hairpin transition at 260 nm (Theimer et al., 1998). The melting points predicted using the values of Table 2 are consistent with the measurements: the predicted pseudoknot-hairpin transition has a midpoint of  $64.6^\circ\text{C}$  (experimental value in the presence of  $10\text{ mM Mg}^{2+}$  is  $67.1^\circ\text{C}$ ), whereas the S2 hairpin is predicted to melt at  $77.6^\circ\text{C}$  (experimental value:  $74.6^\circ\text{C}$ ). Here our parameters slightly underestimate the pseudoknot stability ( $\Delta T_m = 2.5^\circ\text{C}$ ).

Some discrepancies between measured and predicted values may be attributed to differences in ionic conditions. The values in Table 2 are adjusted at  $1\text{ M NaCl}$ , as in classic secondary structure parameters. It is estimated that for nucleic acids  $1\text{ M Na}^+$  is approximately equivalent to  $100\text{--}150\text{ mM Na}^+$  in the presence of  $10\text{ mM Mg}^{2+}$  (Williams et al., 1989; Xia et al., 1998). Melting of the PK5 pseudoknot (Fig. 4A) is measured in  $60\text{ mM Na}^+$  with  $5\text{ mM Mg}^{2+}$  (Wyatt et al., 1990); therefore a lowered stability may be expected. Gene 32 pseudoknot (Fig. 4B) melting was measured in  $100\text{ mM K}^+$  over the range of  $0\text{--}10\text{ mM Mg}^{2+}$ , with measured melting points between  $32.3$  and  $67.1^\circ\text{C}$  (Theimer et al., 1998); therefore the predicted  $T_m$  of  $64.6^\circ\text{C}$  corresponds to a  $\text{Mg}^{2+}$  concentration slightly lower than  $10\text{ mM}$ . Bearing in mind the accuracy of the melting experiments, it may be concluded that our parameters predict the pseudoknot melting curves very closely.

For a pseudoknot from ribosomal protein S15 mRNA, known to be in equilibrium with a mutually exclusive folding (Philippe et al., 1995), our parameters consistently predict a negligible difference between corresponding free energies (the pseudoknot is more stable by  $0.3\text{ kcal/mol}$ ). In this case the pseudoknot folding may be further stabilized by interior interactions in the

loop L2 of 31 nt. Although the three-dimensional modeling (Philippe et al., 1995) also suggests possible stabilizing contacts between this loop and the stem 1 in the minor groove, it seems that the logarithmic extrapolation for large loops is in remarkable agreement with the experimentally observed equilibrium.

### Accuracy and limitations of the estimates

Of course, the rough estimation presented does not take into account many additional effects that could be sequence-dependent. NMR studies of a few pseudoknot structures provide evidence for loop interactions with helical grooves, which may diminish destabilizing energies (Kolk et al., 1998). Possible positive enthalpic contributions in the loops were also suggested (Wyatt et al., 1990). This means that a logarithmic approximation of the size dependence for loop entropies could be rather simplistic. Also, the coefficient in the formula may slightly differ from  $1.75$ , derived for loop closure by a base pair (Fisher, 1966), due to different excluded volume and end-to-end distance effects in pseudoknot loops. Such effects are difficult to estimate because they should depend on a complex interplay between loop and stem dimensions, but they do not seem to lead to a considerable variation in the coefficient values. As RNA pseudoknots are tertiary structure elements,  $\text{Mg}^{2+}$  ion binding could be specific, with significant stabilizing effects (Puglisi et al., 1991) and complex concentration dependence (Theimer et al., 1998). Another important contribution to the stability is the stacking at the junctions between the coaxial stems, which could be influenced by some distortions. On the other hand, even in bent pseudoknots an unpaired nucleotide at the junction is stacked with neighboring bases (Shen & Tinoco, 1995).

Thus some deficiencies in our approximations could compensate for each other. Without taking into account stabilizing or destabilizing sequence-specific energy contributions, we believe that the proposed parameters are valid with an accuracy of about  $\pm 1\text{ kcal/mol}$  and could be used as rough approximations of pseudoknot stabilities and for computer predictions of structures. Applying these parameters in the program STAR for RNA structure prediction (Abrahams et al., 1990; Gulyaev et al., 1995) did not result in an overrepresentation of pseudoknots in the predictions. Although this is indirect evidence, it indicates that the proposed values do not overestimate pseudoknot stabilities significantly. Compared to the previous estimate of the single value of  $4.2\text{ kcal/mol}$  for all pseudoknot loops (Abrahams et al., 1990), the current approximation suggests smaller or equal energies for short loops and greater values for large sizes, as expected. Thus this approximation can improve the prediction programs, being able to predict proven pseudoknots without incorrectly predicted pseudoknotting due to possible un-

derestimation of destabilizing effects in the loops by pseudoknot-including algorithms (Abrahams et al., 1990; Gultyaev, 1991; Gultyaev et al., 1995). Presumably, future experiments will improve these parameters and provide an opportunity to speculate about energies of more complicated pseudoknotted structures.

## MATERIALS AND METHODS

Sequences with the following database accession numbers were taken: J02415, X02144, AF012917, M34077, M25782, D12505, X72587, L35073, AB000709, U03387, U30944, D38444, AB003936, Z29370, D63809, E04305, AF033848, X78966, U34586, X72586, D13438 (tobamoviruses); X16378, AF035199, S97776, X54354, AF035201, AF035202, Y16104, J04374, AF035402, U91413, AF035633, AF035200, J04375 (tymoviruses); L07937, X81639, D30753, Z97873, U64512, Z66493, X78602, L07269, X99149, AJ223596, AJ223597, AJ223598 (furoviruses); U13916, M81486, M81487 (hordeiviruses); M64479, J02399 (STNV-1 and STNV-2); D00530, Y07496, X13063, L25299, L04573, X76931 (luteoviruses); M16605, AF033815, AF033818, AF033820, AF033807, M25381 (retroviral ribosomal frameshift sites); M27472, X73559 (coronaviral ribosomal frameshift sites); AF033811, M54993, K01803, M26927, D10032 (readthrough sites in type C retroviruses). Some database entries with identical pseudoknot sequences were excluded from consideration.

The nearest-neighbor and hairpin loop values were taken from the updated set of parameters by D. Turner and coworkers (<http://www.ibc.wustl.edu/~zucker/rna/energy/index.shtml>). When the stems contained bulges or internal loops, the loop and mismatch contributions were also taken into account. For plant viral RNAs, we used the parameters for 25 °C, for animal RNAs, those for 37 °C. For computer-assisted structure predictions, two algorithms, implemented in the package STAR, were used as described by Abrahams et al. (1990) and Gultyaev et al. (1995).

## ACKNOWLEDGMENTS

This work was supported by The Netherlands Foundation for Chemical Research (SON) and Foundation for Life Sciences (SLW) with financial aid from The Netherlands Organisation for Scientific Research (NWO).

Received October 7, 1998; returned for revision November 2, 1998; Revised manuscript received February 16, 1999

## REFERENCES

- Abrahams JP, van den Berg M, van Batenburg E, Pleij C. 1990. Prediction of RNA secondary structure, including pseudoknotting, by computer simulation. *Nucleic Acids Res* 18:3035–3044.
- Brierley I, Rolley NJ, Jenner A, Inglis SC. 1991. Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. *J Mol Biol* 220:889–902.
- Chamorro M, Parkin N, Varmus HE. 1992. 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* 89:713–717.
- Danthinne X, Seurink J, van Montagu M, Pleij CWA, van Emmelo J. 1991. Structural similarities between the RNAs of two satellites of tobacco necrosis virus. *Virology* 185:605–614.
- Deiman BALM, Pleij CWA. 1997. Pseudoknots: A vital feature in viral RNA. *Semin Virol* 8:166–175.
- Fisher ME. 1966. Effect of excluded volume on phase transitions in biopolymers. *J Chem Phys* 45:1469–1473.
- Garcia A, van Duin J, Pleij CWA. 1993. Differential response to frameshift signals in eukaryotic and prokaryotic translational systems. *Nucleic Acids Res* 21:401–406.
- Gultyaev AP. 1991. The computer simulation of RNA folding involving pseudoknot formation. *Nucleic Acids Res* 19:2489–2494.
- Gultyaev AP, van Batenburg FHD, Pleij CWA. 1994. Similarities between the secondary structure of satellite tobacco mosaic virus and tobamovirus RNAs. *J Gen Virol* 75:2851–2856.
- Gultyaev AP, van Batenburg FHD, Pleij CWA. 1995. The computer simulation of RNA folding pathways using a genetic algorithm. *J Mol Biol* 250:37–51.
- Jacobson H, Stockmayer WH. 1950. Intramolecular reaction in polycondensations. I. The theory of linear systems. *J Chem Phys* 18:1600–1606.
- Jaeger JA, Turner DH, Zuker M. 1989. Improved predictions of secondary structures for RNA. *Proc Natl Acad Sci USA* 86:7706–7710.
- Kim J, Walter AE, Turner DH. 1996. Thermodynamics of coaxially stacked helices with GA and CC mismatches. *Biochemistry* 35:13753–13761.
- Koenig R, Pleij CWA, Beier C, Commandeur U. 1998. Genome properties of beet virus Q, a new furo-like virus from sugarbeet, determined from unpurified virus. *J Gen Virol* 79:2027–2036.
- Kolk MH, van der Graaf M, Wijmenga SS, Pleij CWA, Heus HA, Hilbers CW. 1998. NMR structure of a classical pseudoknot: Interplay of single- and double-stranded RNA. *Science* 280:434–438.
- Kujawa AB, Dugeon G, Hulanicka D, Haenni AL. 1993. Structural requirements for efficient translational frameshifting in the synthesis of the putative viral RNA-dependent RNA polymerase of potato leafroll virus. *Nucleic Acids Res* 21:2165–2171.
- Leathers V, Tanguay R, Kobayashi M, Gallie DR. 1993. A phylogenetically conserved sequence within viral 3' untranslated pseudoknots regulates translation. *Mol Cell Biol* 13:5331–5347.
- Mans RMW, Pleij CWA, Bosch L. 1991. tRNA-like structures: Structure, function and evolutionary significance. *Eur J Biochem* 201:303–324.
- Mans RMW, van Steeg MH, Verlaan PWG, Pleij CWA, Bosch L. 1992. Mutational analysis of the pseudoknot in the tRNA-like structure of turnip yellow mosaic virus RNA: Aminoacylation efficiency and pseudoknot stability. *J Mol Biol* 223:221–232.
- Mathews DH, Andre TC, Kim J, Turner DH, Zuker M. 1998. An updated recursive algorithm for RNA secondary structure prediction with improved thermodynamic parameters. In: Leontis NB, Santa-Lucia J, eds. *Molecular modeling of nucleic acids*. Washington, DC: American Chemical Society. pp 246–257.
- Philippe C, Benard L, Portier C, Westhof E, Ehresmann B, Ehresmann C. 1995. Molecular dissection of the pseudoknot governing the translational regulation of *Escherichia coli* ribosomal protein S15. *Nucleic Acids Res* 23:18–28.
- Pleij CWA. 1994. RNA pseudoknots. *Curr Opin Struct Biol* 4:337–344.
- Pleij CWA, Abrahams JP, van Belkum A, Rietveld K, Bosch L. 1986. The spatial folding of the 3' noncoding region of aminoacylatable plant viral RNAs. In: Brinton MA, Rueckert R, eds. *Positive strand RNA viruses. UCLA symposia on molecular and cellular biology*, 54. New York: Alan Liss, Inc. pp 299–316.
- Pleij CWA, Bosch L. 1989. RNA pseudoknots: Structure, detection, and prediction. *Methods Enzymol* 180:289–303.
- Pleij CWA, Rietveld K, Bosch L. 1985. A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res* 13:1717–1731.
- Poland D, Scheraga HA. 1966. Occurrence of a phase transition in nucleic acid models. *J Chem Phys* 45:1464–1469.
- Poland D, Scheraga HA. 1970. *Theory of helix-coil transitions in biopolymers*. New York and London: Academic Press.
- Prüfer D, Tacke E, Schmitz J, Kull B, Kaufmann A, Rohde W. 1992. Ribosomal frameshifting in plants: A novel signal directs the –1 frameshift in the synthesis of the putative viral replicase of potato leafroll luteovirus. *EMBO J* 11:1111–1117.



- Puglisi JD, Wyatt JR, Tinoco I. 1990. Conformation of an RNA pseudoknot. *J Mol Biol* 214:437–453.
- Puglisi JD, Wyatt JR, Tinoco I. 1991. RNA pseudoknots. *Acc Chem Res* 24:152–158.
- SantaLucia J Jr, Turner DH. 1998. Measuring the thermodynamics of RNA secondary structure formation. *Biopolymers* 44:309–319.
- Shen LX, Tinoco I. 1995. The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. *J Mol Biol* 247:963–978.
- Solovyev AG, Savenkov EI, Agranovsky AA, Morozov SY. 1996. Comparisons of the genomic *cis*-elements and coding regions in RNA $\beta$  components of the hordeiviruses barley stripe mosaic virus, lych-nis ringspot virus, and poa semilatifolius virus. *Virology* 219:9–18.
- ten Dam EB, Pleij CWA, Bosch L. 1990. RNA pseudoknots: Translational frameshifting and readthrough on viral RNAs. *Virus Genes* 4:121–136.
- Theimer CA, Wang Y, Hoffman DW, Krisch HM, Giedroc DP. 1998. Non-nearest neighbor effects on the thermodynamics of unfolding of a model mRNA pseudoknot. *J Mol Biol* 279:545–564.
- van Batenburg FHD, Gultyaev AP, Pleij CWA. 1995. An APL-programmed genetic algorithm for the prediction of RNA secondary structure. *J Theor Biol* 174:269–280.
- van Belkum A, Abrahams JP, Pleij CWA, Bosch L. 1985. Five pseudoknots are present at the 204 nucleotides long 3' noncoding region of tobacco mosaic virus RNA. *Nucleic Acids Res* 13:7673–7686.
- Walter AE, Turner DH. 1994. Sequence dependence of stability for coaxial stacking of RNA helices with Watson–Crick base paired interfaces. *Biochemistry* 33:12715–12719.
- Walter AE, Turner DH, Kim J, Lyttle MH, Muller P, Mathews DH, Zuker M. 1994. Coaxial stacking of helices enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc Natl Acad Sci USA* 91:9218–9222.
- Williams AP, Longfellow CE, Freier SM, Kierzek R, Turner DH. 1989. Laser temperature-jump, spectroscopic, and thermodynamic study of salt effects on duplex formation by dGCATGC. *Biochemistry* 28:4283–4291.
- Wills NM, Gesteland RF, Atkins JF. 1994. Pseudoknot-dependent read-through of retroviral gag termination codons: Importance of sequences in the spacer and loop 2. *EMBO J* 13:4137–4144.
- Wyatt JR, Puglisi JD, Tinoco I. 1990. RNA pseudoknots: Stability and loop size requirements. *J Mol Biol* 214:455–470.
- Xia T, SantaLucia J, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH. 1998. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochemistry* 37:14719–14735.
- Zuker M. 1989. On finding all suboptimal foldings of an RNA molecule. *Science* 244:48–52.