Purine–purine mismatches in RNA helices: evidence for protonated G·A pairs and next-nearest neighbor effects

Sarah E. Morse⁺ and David E. Draper^{*}

Department of Chemistry, Johns Hopkins University, Baltimore, MD 21218, USA

Received July 12, 1994; Revised and Accepted November 8, 1994

ABSTRACT

Thermodynamic parameters are presented for 12 different RNA duplexes containing A·A, A·G, G·A and G·G mismatches flanked by C·G base pairs. UV melting studies were conducted under three different buffer conditions in order to evaluate the effects of salt concentration and pH on the stability of each mismatch-containing duplex. The main findings are: (i) the mismatches have a wide range of effects on duplex stability, decreasing ΔG°_{37} of denaturation by ~0–7 kcal/mol; (ii) the nearest-neighbor assumption commonly used to calculate helix stability breaks down for G·A mismatches; and (iii) G·A mismatches separated by 2 bp form a protonated structure.

INTRODUCTION

G·A, G·G, and A·A mismatches in RNA helices all potentially form hydrogen bonded pairs, and there is evidence that they contribute to the structure and stability of RNA helices in many circumstances. The crystal structures of tRNA^{Phe} showed that an A·m²G juxtaposition is hydrogen bonded and stacked at one end of the anticodon stem (1), and a statistical analysis of helices in conserved ribosomal RNA structures found that G·A juxtapositions at the ends of helices are found with unexpectedly high frequency (2). G·A pairs apparently substitute for Watson–Crick pairs in some phylogenetically conserved helices of ribosomal RNAs, and are highly conserved in some positions (3,4). Purine–purine substitution for Watson–Crick pairs occurs in other RNAs as well (5–7). Pairs of small subunit rRNA nucleotides which co-vary as A·A, G·G, or G·A have been noted (8). Recent NMR studies of an internal loop from 5S rRNA (9), a large subunit rRNA hairpin loop (10), and a protein recognition site from HIV (11,12) have provided structural evidence for G·A and G·G hydrogen bonding in contexts other than Watson–Crick helices.

The evidence that purine–purine juxtapositions have important roles in RNA structures is thus compelling. To investigate more thoroughly the potential contributions of G·A, A·A, and G·G

mismatches in RNA helices, we have synthesized these juxtapositions in a variety of sequence contexts and measured the thermodynamics of formation under several different solution conditions. G·A mismatches are found to have particularly unusual behavior, which can be rationalized in part by recent structural work from other laboratories. We also report the first evidence for protonated G·A pairs in RNA.

MATERIALS AND METHODS

RNA synthesis and purification

Oligoribonucleotides were synthesized on solid supports using the phosphoramidite method, by a Milligen Biosearch 8600 synthesizer. Monomers were supplied by Milligen Biosearch or Peninsula Labs. Other synthesis reagents were supplied by Milligen Biosearch or Cruachem. After synthesis, the RNA oligomer was removed from the support in pyridine:ammonia (1:1 by volume) overnight at room temperature. The solution was decanted, dried under vacuum, resuspended in 50% ethanol, and dried again. The RNA was resuspended in ethanol:ammonia (1:3 by volume) and heated to 55°C in a sealed tube for 4 h to remove base protecting groups. After drying thoroughly, the 2'-OH protecting group was removed by addition of 1 ml 1 M tetrabutylammonium fluoride in tetrahydrofuran (Aldrich). After 4 h at room temperature, 1 ml 0.5 M ammonium acetate was added. Most of the TBAF was removed by putting the solution over a Waters C18 Sep Pak cartridge, equilibrated with 0.1 M triethylamine bicarbonate, pH 7.0 (TEAB). After rinsing the column with 10 mM TEAB, the RNA was eluted with 50% acetonitrile in 10 mM TEAB. The eluate was extracted twice with ether and dried.

Further purification of the RNA was carried out by ion exchange HPLC on a BioRad DEAE-5PW column, using an NaCl gradient in 6 M urea, 20 mM Tris, pH 7.6. Fractions containing full-length oligomer were pooled and purified on a C4 reverse phase column (Rainin) using an acetonitrile gradient. Finally, the 5' trityl group was removed by treatment with 3% acetic acid. The purified oligomer was then buffered with TEAB and dried.

* To whom correspondence should be addressed

⁺Present address: Department of Biochemistry, Biophysics, and Genetics, University of Colorado Health Sciences Center, Denver, CO 80262, USA

Figure 1. Typical melting behavior of RNA duplexes with G-A mismatches. (A) Absorbance (280 nm) versus temperature profiles for oligomer 9. Two different concentrations are shown. Absorbances of 0.4 and 0.15 have been subtracted from the upper and lower curves, respectively, to display the two curves on the same graph. Solid lines are least squares best fits as described in Materials and Methods. (B) concentration dependence of the melting temperature for several duplex RNAs, in buffer with 1 M NaCl. ●, oligomer 9; ○, oligomer 12; □, oligomer 11; ■, oligomer 12.

UV hyperchromicity data collection

Melting curves were measured at 280 nm using a Perkin Elmer Lambda 4 spectrophotometer with a thermostatted cell changer and computerized absorbance and temperature collection, as previously described (13). The heating rate for all the experiments reported here was 1°C/min.

Three different buffers were used for the melting studies: 1.0 M NaCl, 5 mM NaH₂PO₄, 0.5 mM Na₂EDTA, pH 7.0 (1 M buffer); 0.1 M NaCl, 5 mM NaH₂PO₄, 0.5 mM Na₂EDTA, pH 7.0 (0.1 M buffer); and 1.0 M NaCl, 5 mM NaH₂PO₄, 0.5 mM Na₂EDTA, pH 5.0 (pH 5 buffer). RNA samples were resuspended in the desired buffer, centrifuged briefly, placed in cuvettes, and heated briefly to 90°C to degas before carrying out the melting experiment.

Data analysis

Absorbance versus temperature profiles were fitted to a two state model for denaturation of self-complementary oligomers using the procedure described by Petersheim and Turner (14). Briefly, it is assumed that the equilibrium $2R \rightleftharpoons R_2$, between the single-stranded oligomer and its fully base paired dimer, fully describes the denaturation, and that the extinction coefficient of both the single- and double-stranded species are linear functions of temperature. The curve was fitted to six variables by a non-linear least squares program incorporating the Marquadt method (15); the variables are the slopes and intercepts for the single- and double-stranded RNA extinctions, and ΔH° and ΔS° which determine the position of the equilibrium at any temperature. The total concentration of strands (C_t) also enters into the calculation, and was determined at 80°C by the method of Borer (16).

For any one oligomer the melting experiment was repeated at 8–18 different total strand concentrations spanning a 70–130-fold concentration range. Thermodynamic values for duplex formation were then determined by either averaging the fitted ΔH° and ΔS° values, or by plotting T_m^{-1} as a function of $\ln(C_t)$; the slope of this graph is $R/\Delta H^\circ$ (R is the gas constant, T_m is the temperature at which the equilibrium constant is unity) (17). The

values given by the two methods agreed to 10% or better; values listed in Table 1 are from the T_m^{-1} vs. $\ln(C_t)$ plots. Also reported is ΔG°_{37} , the calculated free energy change at 37°C.

RESULTS

Design of oligomers and typical data

Thermodynamic parameters for the introduction of purine–purine mismatches were determined from the melting behavior of self-complementary RNA oligomers prepared by chemical synthesis (see Materials and Methods). As a check of the quality of the RNA, the oligomer ACUUAAGU was synthesized and its melting profiles compared with those of the same sequence RNA kindly provided by Dr D. Turner. Thermodynamic parameters calculated for the chemically synthesized oligomer were within 2% of those reported by Freier *et al.* (18).

Typical melting profiles are shown in Figure 1A. Fits of these curves to two-state melting transitions, and extraction of thermodynamic parameters from plots of the reciprocal of the melting temperature as a function of the logarithm of the total strand concentration (Fig. 1B), were done by the same procedures used by Turner and colleagues. Thus the data, which are summarized in Table 1, can be directly compared with the parameters compiled by these workers (19).

To avoid placing mismatches too close to the ends of helices, and to be able to study very destabilizing mismatches, we used fairly long oligomers of 9–10 nucleotides. The T_m is less steeply dependent on RNA concentration for longer oligomers, and it becomes correspondingly more difficult to obtain highly accurate thermodynamic parameters. We nevertheless had average standard errors of 7% for ΔH° and ΔS° . This translates into a standard error for ΔG°_{37} of ~10%, but this is large enough that $\Delta(\Delta G^\circ)$ values obtained by comparison of two oligomers (e.g. with and without a mismatch) have quite large uncertainties associated. We have therefore limited our conclusions to the several very major effects that are observed.

Table 1. Thermodynamic parameters for denaturation of oligomers containing purine-purine mismatches

	Oligonucleotide	Buffer	ΔH°	ΔS°	ΔG°_{37}
1	CGCG <u>A</u> CGCG	1 M	61.1	166	9.6
		0.1 M	55.3	150	8.8
		pH 5	77.7	215	10.8
2	GCGC <u>A</u> GCGC	1 M	72.4	201	10.0
		0.1 M	49.3	130	8.7
		pH 5	57.3	155	7.9
3	CGCGG <u>C</u> GCG	1 M	76.9	207	12.5
		0.1 M	58.4	155	10.2
		pH 5	36.0	87.3	8.9
4	GCGCGG <u>C</u> GCG	1 M	59.9	157	8.5
		0.1 M	not a duplex		
		pH 5	57.8	155	9.6
5	GC <u>A</u> GCGCGC	1 M	59.9	163	9.5
		0.1 M	45.3	112	8.2
		pH 5	62.9	170	10.1
6	CG <u>A</u> GCGGCG	1 M	68.9	192	9.5
		0.1 M	45.4	122	7.6
		pH 5	45.3	112	8.2
7	CGCGG <u>C</u> AGCG	1 M	not a duplex		
		0.1 M	not a duplex		
		pH 5	49.8	134	8.4
8	CGC <u>A</u> GCGGCG	1 M	47.6	116	5.8
		0.1 M	21.2	50.6	5.5
		pH 5	100.0	288	11.3
9	GCGCG <u>A</u> GCGC	1 M	88.8	249	11.6
		0.1 M	65.1	245	10.9
		pH 5	96.2	272	11.5
10	CGCGG <u>A</u> CGCG	1 M	96.8	269	13.3
		0.1 M	84.5	232	12.6
		pH 5	65.7	177	10.8
11	GCGC <u>A</u> GGCGC	1 M	70.1	194	10.0
		0.1 M	65.8	182	9.4
		pH 5	83.3	229	12.3
12	CGCG <u>A</u> GCGCG	1 M	89.6	252	11.4
		0.1 M	87.6	277	11.2
		pH 5	49.6	133	8.4

The full buffer conditions are described in Materials and Methods, as are the calculation procedures. Standard errors for ΔH° and ΔS° are 7% and for ΔG°_{37} , 10%. 'Not a duplex' indicates that no cooperative melting transition was observed.

Hairpin formation is also a potential problem for longer oligomers. That this was not a problem for the data reported here was indicated by (i) linear plots of T_m^{-1} versus $\ln(C_1)$, as shown in Figure 1B, and (ii) good agreement between the enthalpies

calculated from $\ln(C_1)$ plots and the van't Hoff enthalpies derived from the individual melting curves.

All mismatches were introduced into the context of alternating GC sequences. As a control, the oligomer (CG)₄ was synthesized, and the measured ΔH° and ΔS° of melting found to be within 5% of the values predicted from tables of nearest neighbor thermodynamic parameters. (Measured values are $\Delta H^\circ = 76.5$ kcal/mol and $\Delta S^\circ = 205$ cal/°C mol, versus 74.6 kcal/mol and 194.5 cal/°C mol predicted (19). The difference in ΔG°_{37} is larger, 12.9 kcal/mol observed versus 14.3 kcal/mol predicted.) The nearest neighbor assumption is therefore valid for alternating GC sequences, as expected from several measurements of shorter alternating GC RNA oligomers (18). The predicted values for (GC)₄ are only slightly different: 80.0 kcal/mol and 210 cal/°C mol for ΔH° and ΔS° of denaturation, respectively, and a predicted ΔG°_{37} of 14.9 kcal/mol.

The melting experiments were carried out in three different buffers. A salt concentration of 1 M NaCl has been standard for these kinds of thermodynamic analyses (19), but may be different from *in vivo* conditions; we therefore carried out measurements in both 1 M and 0.1 M NaCl (pH 7.0). There have been observations that G·A mismatches in DNA are sometimes stabilized by low pH (20,21), so we also carried out a series of experiments in 1 M NaCl at pH 5.0. The results are summarized in Table 1.

A·A and G·G mismatches

A·A and G·G mismatches were each studied in two different sequence contexts (oligomers 1–4, Table 1), and with one exception decrease the enthalpy of melting and destabilize the helix. The exception is CGCGGCGCG, which has the same stability as (CG)₄, within experimental error. Sequence context has a particularly large effect for G·G mismatches, as GCGCGGCGCG is ~4 kcal/mol less stable. These four oligomers show the same trend in stability when the salt concentration is lowered to 0.1 M NaCl.

When the pH was lowered from 7.0 to 5.0, the stability of CGCGGCGCG decreased by 3.6 kcal/mol, while the other purine-purine mismatches were affected to much smaller degrees or not at all. G nucleotides have pK values less than 3, so it is more likely that an alternative structure involving C⁺ residues (pK ≈ 4.2) is responsible.

Isolated G·A mismatches

A second series of RNAs were designed to look at the effects of isolated G·A mismatches. Each self-complementary RNA contains two mismatches in identical sequence contexts (oligomers 5–8). When the two mismatches are separated by 4 bp (oligomers 5 and 6) the duplex is destabilized, though the effect per mismatch is not as severe as for most of the A·A or G·G mismatches examined. Two additional molecules with G·A mismatches separated by only 2 bp were also studied (oligomers 7 and 8). If the nearest neighbor approximation applies, the thermodynamics of these two RNAs and the RNAs with G·A mismatches separated by 4 bp should be similar. (For instance, accounting for the difference in numbers of GC and CG sequences, oligomers 7 and 8 should both be only 1.4 kcal/mol less stable than oligomer 5.) However, these molecules are drastically affected: CGCGGCAGCG does not even form a duplex in the 1 M NaCl buffer, and CGCAGCGGCG has been destabilized by more than

7 kcal/mol. A G-A mismatch must introduce distortions into the helix which extend at least to next-nearest neighbors.

Reduction of the salt concentration destabilized all four of the G-A mismatch duplexes (oligomers 5–8) without much change in the relative stabilities. However, lowering the pH to 5 had a large effect on the RNAs with G-A mismatches separated by 2 bp. Oligomer 7 is now observed to form a duplex, and the enthalpy and free energy of denaturing oligomer 8 nearly doubles. The lower pH has little or no effect on the stabilities of G-A mismatches separated by 4 bp. This result again suggests that sequence context beyond nearest neighbor base pairs affects the structure of G-A mismatches.

Tandem G-A mismatches

A third set of RNA duplexes examined the effects of two adjacent G-A mismatches (oligomers 9–12). Compared to the other RNAs containing two separated G-A mismatches, the effects on helix stability are very modest, and in one case, CGCGGACGCG (oligomer 10), the duplex has about the same stability and a much larger enthalpy than (CG)₄. In another recent study GGAC was also found to be the most stable sequence context for tandem G-A mismatches (22). The same trend is observed in 0.1 M NaCl buffer. In contrast to the strong stabilization of G-A mismatches separated by 2 bp, low pH only affects oligomer 12 strongly, and destabilization is seen in this case. This is consistent with another report that the CD spectrum of a tandem mismatch in the context CGAG (the same as oligomer 9) is not affected by pH (23).

DISCUSSION

G-A mismatches can potentially pair in a number of different ways, only two of which have been observed in RNA to date. An A(*anti*)-G(*anti*) pair with bonding between the two N1 positions is present in the tRNA^{Phe} crystal structures (1); the modified dimethyl N2 of the guanine limits the number of pairing possibilities in this case. Recently the same pairing has been found in an internal loop (11). There are now a number of examples of a second A(*anti*)-G(*anti*) pairing with hydrogen bonds between N6(A)-N3(G) and N7(A)-N2(G). This pairing is found as a closing pair for GNRA tetraloops (24) and in other loops with non-canonical pairings (9,10). More relevant to the present discussion is that this pairing occurs in the duplex formed by GGCGAGCC (25), which has tandem G-A mismatches in the same sequence context as oligomer 9.

A striking result of our thermodynamic measurements is that G-A mismatches have vastly different effects on helix stability depending on the number of base pairs separating them in the helix. Do the available G-A pair structures provide any rationale for this behavior? Gautheret *et al.* (26) have pointed out that the distance P(G)-O3'(A) for the second variety of G-A pair described is similar to that for standard A form helix (16.7 versus 17.5 Å), while the corresponding P(A)-O3'(G) distance is much shorter, 12.7 Å. This prevents a normal Watson-Crick pair from being attached to the 5' side of the A, which can explain the relatively stable structures formed by tandem 5'_{AG}^{GA}3' mismatch sequences. A requirement for a distorted backbone at the adjacent base pair could also account for non-nearest neighbor effects that we observe. However, the same structural distortion should render the tandem mismatch 5'_{AG}^{GA}3' particularly unstable, which it is not. It thus seems very likely that the two different tandem

mismatch sequences form different structures; whether they correspond to the two pairing schemes already observed in RNA will require NMR studies to decide.

The strong pH dependence of oligomers 7 and 8 suggests that the G-A mismatches in these sequences form a protonated pair. Such pairing has been observed in DNA. In an NMR study of a DNA duplex, lowering the pH below 5.5 switched a G-A pair from the G(*anti*)-A(*anti*) form seen in tRNA to G(*syn*)-A⁺(*anti*) with the N1(A) bonded to N7(G) (20). The same low pH pairing was also observed in DNA by crystallography (21,27). In these DNA studies the pairing scheme observed is highly dependent on the surrounding sequence. Evidently the same situation applies to RNA, and a G-A⁺ pair is only observed in certain contexts. The surprising result is that the context favoring a protonated pair was created only when the mismatches were separated by 2 bp.

Fewer instances of A-A and G-G mismatches have been studied. They are fairly rare in ribosomal RNAs (4), though protein S4 recognition depends on an A-A mismatch (28,29). Recent NMR studies of the Rev binding site within HIV RNA have revealed a G-G pair within an internal loop; pairing takes place via the N1 and O6 positions (11,12). Our results with A-A and G-G mismatches suggest that introduction of such a mismatch into a helix is at best thermodynamically neutral.

Our comparisons of duplex melting in 1 and 0.1 M NaCl reveal no major change in relative stabilities. Measurement of the salt dependence of the free energy of duplex formation for several different length oligomers show that the stability is a simple function of the oligomer length, as predicted by simple polyelectrolyte theory (30). On the basis of those experiments, it can be estimated that ΔG° should decrease by 1.8–2 kcal/mol in going from 1 to 0.1 M NaCl for the 9mers and 10mers studied here. The average observed decrease is only 1.0 kcal/mol, but given the uncertainties in ΔG°_{37} the discrepancy may not be significant.

In summary, we have shown that insertions of purine-purine mismatches into a helix have only minor effects on the overall stability in some situations, but that sequence context and solution pH play major roles in determining the final stability. The breakdown of the nearest neighbor approximation may be attributed to the fact that there are a variety of mismatch pairings available, most of which require some distortion of the normal A-form helix backbone. Communication of these distortions along the helix backbone make the sequence context particularly important.

ACKNOWLEDGEMENT

This work was supported by NIH grant GM37005.

REFERENCES

- 1 Rich, A. and RajBhandary, U.L. (1976) *Ann. Rev. Biochem.*, **45**, 805–860.
- 2 Traub, W. and Sussman, J.R. (1982) *Nucleic Acids Res.*, **10**, 2701–2708.
- 3 Woese, C.R., Gutell, R., Gupta, R. and Noller, H.F. (1983) *Microbiol. Rev.*, **47**, 621–669.
- 4 Gutell, R.R., Larsen, N. and Woese, C.R. (1994) *Microbiol. Rev.*, **58**, 10–26.
- 5 Pace, N.R., Smith, D.K., Olsen, G.J. and James, B.D. (1989) *Gene*, **82**, 65–75.
- 6 Michel, F. and Westhof, E. (1990) *J. Mol. Biol.*, **216**, 585–610.
- 7 Larsen, N. and Zwieb, C. (1991) *Nucleic Acids Res.*, **19**, 209–215.
- 8 Gutell, R.R. and Woese, C.R. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 663–667.
- 9 Wimberly, B., Varani, G. and I. Tinoco, J. (1993) *Biochemistry*, **32**, 1078–1087.

- 10 Szewczak, A.A., Moore, P.B., Chan, Y.-L. and Wool, I.G. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 9581–9585.
- 11 Battiste, J.L., Tan, R., Frankel, A.D. and Williamson, J.R. (1994) *Biochemistry*, **33**, 2741–2747.
- 12 Peterson, R.D., Bartel, D.P., Szostak, J.W., Horvath, S.J. and Feigon, J. (1994) *Biochemistry*, **33**, 5357–5366.
- 13 Laing, L.G. and Draper, D.E. (1994) *J. Mol. Biol.*, **237**, 560–576.
- 14 Petersheim, M. and Turner, D.H. (1983) *Biochemistry*, **22**, 256–263.
- 15 Bevington, P.R. (1969) *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill, New York, NY.
- 16 Borer, P.N. (1975) In G. D. Fasman (ed), *Handbook of Biochemistry and Molecular Biology: Nucleic Acids*. CRC Press, Cleveland, OH.
- 17 Borer, P., Dengler, B., Uhlenbeck, O.C. and Tinoco, I.J. (1974) *J. Mol. Biol.*, **86**, 843–853.
- 18 Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Nielson, T. and Turner, D.H. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 9373–9377.
- 19 Turner, D.H., Sugimoto, N. and Freier, S.M. (1988) *Ann. Rev. Biophys. Biophys. Chem.*, **17**, 167–192.
- 20 Gao, X. and Patel, D.J. (1988) *J. Am. Chem. Soc.*, **110**, 5178–5182.
- 21 Brown, T., Leonard, G.A., Booth, E.D. and Kneale, G. (1990) *J. Mol. Biol.*, **212**, 437–440.
- 22 Walter, A.E., Wu, M. and Turner, D.H. (1994) *Biochemistry*, **33**, 11349–11354.
- 23 SantaLucia, J., Kierzek, R. and Turner, D.H. (1990) *Biochemistry*, **29**, 8813–8819.
- 24 Heus, H.A. and Pardi, R. (1991) *Science*, **252**, 191–194.
- 25 SantaLucia, J., Jr and Turner, D. (1993) *Biochemistry*, **32**, 12612–12623.
- 26 Gautheret, D., Konings, D. and Gutell, R.R. (1994) *J. Mol. Biol.*, **242**, 1–8.
- 27 Brown, T., Leonard, G.A., Booth, E.D. and Chambers, J. (1989) *J. Mol. Biol.*, **207**, 455–457.
- 28 Stern, S., Wilson, R.C. and Noller, H.F. (1986) *J. Mol. Biol.*, **192**, 101–110.
- 29 Sapag, A. and Draper, D.E. (1994) submitted.
- 30 Freier, S.M., Petersheim, M., Hickey, D.R. and Turner, D.H. (1984) *J. Biomol. Struct. Dynam.*, **1**, 1229–1242.