Article No. jmbi.1999.2906 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 290, 967–982

JMB

Thermodynamics of Unpaired Terminal Nucleotides on Short RNA Helixes Correlates with Stacking at Helix Termini in Larger RNAs

Mark E. Burkard[1], Ryszard Kierzek[2] and Douglas H. Turner[1]*

[1]*Department of Chemistry RC Box 270216, University of Rochester, Rochester, NY 14627-0216, USA*

[2]*Institute of Bioorganic Chemistry, Polish Academy of Sciences, 60-704, Poznan Noskowskiego 12/14, Poland*

**Corresponding author*

Free energies for stacking of unpaired nucleotides (dangling ends) at the termini of oligoribonucleotide Watson-Crick helixes ($\Delta G^0_{37,\mathrm{stack}}$) depend on sequence for 3′ ends but are always small for 5′ ends. Here, these free energies are correlated with stacking at helix termini in a database of 34 RNA structures determined by X-ray crystallography and NMR spectroscopy. Stacking involving GA pairs is considered separately. A base is categorized as stacked by its distance from ($\leqslant$4.0 Å), angle with ($\leqslant 30°$), and overlap with the terminal helix base-pair. A base is unstacked if it does not satisfy one or more of these criteria.

Of the 36 unpaired bases in sequences with $\Delta G^0_{37,\mathrm{stack}}$ more favorable than −0.7 kcal/mol, 30 (83 %) are stacked on the adjacent base-pair, indicating a propensity for such sequences to stack in the 3D structure. Structures containing the strongly stacked sequence $\overset{\mathrm{CU}}{\underset{\leftarrow}{\overset{\rightarrow}{\mathrm{G}}}}$ ($\Delta G^0_{37,\mathrm{stack}} = -1.2$ kcal/mol) show that the amino group of C closely overlaps the carbonyl-4 of U. Thermodynamic measurement of U stacking on a 2-pyrimidinone-guanine base-pair, where the amino group of C is replaced by hydrogen, suggests that interactions with the cytosine amino group contribute ~0.5 kcal/mol to $\Delta G^0_{37,\mathrm{stack}}$.

For GA mismatches at helix termini, the nucleotide at the 3′ helix end is always stacked, and the nucleotide at the 5′ end is stacked in almost 90 % of occurrences. In available structures, non-Watson-Crick paired bases 3′ to an imino-hydrogen bonded GA are also always stacked; the GA provides a large platform for favorable stacking.

For the 56 sequences associated with $\Delta G^0_{37,\mathrm{stack}}$ less favorable than −0.4 kcal/mol, 19 (34 %) are stacked; these sequences have a propensity for not stacking on adjacent base-pairs. Phylogenetic conservation of weakly stacking sequences at 3′ ends may be a predictor of a backbone turn.

Keywords: dangling end; 2-pyrimidinone; inosine; electrostatics; RNA folding

Introduction

Base stacking and hydrogen bonding provide the major, sequence-dependent forces driving helix formation in nucleic acids (Turner *et al.*, 1988). The resulting Watson-Crick paired helixes are regular in their three-dimensional (3D) structures (Saenger, 1984) and have small, predictable variations (Gorin *et al.*, 1995; Young *et al.*, 1995). In non-Watson-Crick regions, base stacking and hydrogen bonding may also be important in determining folding, but structures are more heterogeneous, more sequence-dependent, and more difficult to predict. Even for well-studied GA mismatches there are no general rules for predicting which conformation predominates (Li *et al.*, 1991; Cheng *et al.*, 1992; Heus & Pardi, 1991; Wu *et al.*, 1997; Xia *et al.*, 1997; Pan *et al.*, 1999).

If stacking interactions are important in non-Watson-Crick regions, thermodynamic data for base stacking may correlate with 3D structure, allowing predictions about RNA folding. Stacking free energies have been estimated by com-

E-mail address of the corresponding author: TURNER@CHEM.ROCHESTER.EDU

0022-2836/99/300967–16 $30.00/0

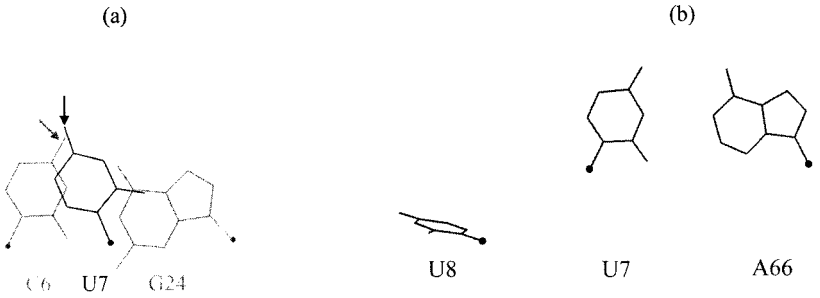

Figure 1. Stick models of (a) a stacked and (b) an unstacked dangling end; hydrogen atoms are not shown. (a) U7 from the Sarcin-ricin loop (Szewczak *et al.*, 1993) is stacked as a 3′ dangling end on an adjacent CG pair; the predicted free energy for $\overrightarrow{\substack{\mathrm{CU}\\\mathrm{G}}}$ is −1.2 kcal/mol. (One of the six final structures is shown.) Note the proximity of the uridine O-4 (black arrow) with the cytosine amino group (gray arrow). (b) U8 from tRNA^{Phe}, bound to *Thermus aquaticus* elongation factor Tu (Nissen *et al.*, 1995) is not stacked on the adjacent UA pair; the predicted free energy for $\overrightarrow{\substack{\mathrm{UU}\\\mathrm{A}}}$ is −0.1 kcal/mol. For each base, a filled circle identifies the position of ribose C-1′.

paring the ΔG^0_{37} for duplex formation of helixes with and without unpaired bases, or "dangling ends," on each end of a helix (Petersheim & Turner, 1983; Freier *et al.*, 1983, 1986b; Sugimoto *et al.*, 1987; Turner *et al.*, 1988). For certain sequences, a 3′ dangling end can stabilize the duplex by 0.8 kcal/mol or more, and such sequences are termed strongly stacking. For example, $\overrightarrow{\substack{\mathrm{CA}\\\mathrm{G}}}$ stabilizes by 1.7 kcal/mol. The 3′ dangling ends $\overrightarrow{\substack{\mathrm{UU}\\\mathrm{A}}}$ and $\overrightarrow{\substack{\mathrm{UC}\\\mathrm{A}}}$ as well as most 5′ dangling ends stabilize duplexes by less than 0.4 kcal/mol, and are termed weakly stacking.

Dangling end stability correlates with unpaired base stacking on helixes in yeast tRNA^{Phe}, and weakly stacking bases in yeast tRNA^{Phe} are associated with turns in the RNA backbone (Sugimoto *et al.*, 1987; Turner *et al.*, 1987b, 1988). Today, an expanded analysis is possible; the Protein Data Bank (PDB) and the Nucleic Acid Data Base (Berman *et al.*, 1992) contain many new experimentally determined 3D structures of RNAs. A preliminary examination of available X-ray structures suggests that trends observed with tRNA^{Phe} are general (Burkard *et al.*, 1999a). Here, we correlate free energy of stacking with the positions of bases adjacent to the helix ends in 34 structures that meet certain selection criteria. A base is defined as stacked if it meets distance, angle, and overlap requirements relative to the adjacent base-pair (see Materials and Methods). Additionally, we measure stabilities of duplexes with stacking on base-pairs of 2-pyrimidinone-guanine and cytosine-inosine to explore the interactions that make stacking strong in $\overrightarrow{\substack{\mathrm{CU}\\\mathrm{G}}}$.

Results

Fourteen crystal and 20 NMR structures were considered, providing a total of 1495 nucleotides. Table 1 lists the structures with their PDB accession codes and Table 2 presents some characteristics of the database. Although only 14 of the 34 structures (41 %) are from crystallography, these structures tend to be larger than those derived by NMR; thus 60 % of the nucleotides are from crystallography. Structures were classified as bound if a peptide or another non-RNA molecule was included in the NMR or crystal sample. About half of the structures and nucleotides are from bound RNAs. Over a third of the nucleotides in the database are from tRNA structures. The largest structure is the 158 nucleotide P4-P6 domain of the group I intron from *Tetrahymena thermophila* large subunit rRNA (Cate *et al.*, 1996).

Stacking on helixes

Figure 1 shows typical examples of a stacked base and an unstacked base at the end of a helix. For clarity, only the dangling base and the last Watson-Crick pair of the helix are shown. In Figure 1(a), a $\overrightarrow{\substack{\mathrm{CU}\\\mathrm{G}}}$ sequence has a stacked U, as expected from a measured free energy of stacking of −1.2 kcal/mol (Turner *et al.*, 1988). Figure 1(b) shows a $\overrightarrow{\substack{\mathrm{UU}\\\mathrm{A}}}$ dangling end from a crystal structure of tRNA^{Phe} bound to *Thermus aquaticus* elongation factor Tu. In this sequence, the 3′ dangling U is not stacked on the adjacent pair, consistent with the low free energy attributed to stacking, −0.1 kcal/mol. A similar dangling end is present for unbound tRNA^{Phe} (Burkard *et al.*, 1999a); protein binding does not affect stacking of this dangling end.

Table 3 contains the analysis of base stacking on helix ends, sorted by the ΔG^0_{37} for dangling end stacking in oligonucleotides. The following are excluded: (a) dangling bases that could make a GA pair; (b) the helix ends with non-canonical or non-hydrogen-bonded GC, AU, or GU pairs; (c) helix ends with stacking on wobble GU pairs, except $\overrightarrow{\substack{\mathrm{UA}\\\mathrm{G}}}$, which is the only such end with a measured free energy (Freier *et al.*, 1986a); (d) the helix ends in bulge, tetraloop, non-loop (exterior) motifs, and internal loops of 1 × 4, 1 × 6, 1 × 7, 1 × 11 and 2 × 7 nucleotides; (e) helixes in pseudoknots. Dramatically asymmetric internal loops are like bulges in that they may provide little space for base stacking. Although non-loop nucleotides are expected to conform to the stacking model, they are likely to be affected by packing forces in crystal structures, and are generally determined by few restraints in

Table 1. The available 3D structures that contain mismatches or unpaired bases in junctions, hairpins of five or more, and symmetrical and slightly asymmetrical internal loops

Description	Name	PDB Accession code	Year	No. nt	Structures[a]	F/B[b]	C/N[c]	Reference(s)
16 S rRNA site w/aminoglycoside	16 S-aminogly	1PBR	1996	27	1	B	N	Fourmy *et al.* (1996)
5 S rRNA loop E domain	5 S loop E	354D/356D	1997	60		F	C	Correll *et al.* (1997)
Arg Aptamer RNA w/Arg	Arg-apt	1KOC	1996	25	1	B	N	Yang *et al.* (1996)
BIV TAR w/Tat	BIV TAR-Tat	1BIV	1995	28	5	B	N	Ye *et al.* (1995)
Citrulline aptamer	Cit-apt	1KOD	1996	25	1	B	N	Yang *et al.* (1996)
AA flanked by GA	CurveHelix	283D	1996	24		F	C	Baeyens *et al.* (1996)
FMN Aptamer RNA w/FMN	FMN-apt	1FMN	1996	35	5	B	N	Fan *et al.* (1996)
Free HIV-1 TAR	Free TAR	1ANR	1995	29	20	F	N	Aboul-ela *et al.* (1995)
All-RNA hammerhead	Hammerhead	1MME/299D	1995	41		F	C	Scott *et al.*, (1995, 1996)
HDV ribozyme hairpin	HDV-hp	1ATO	1997	19	10	F	N	Kolk *et al.* (1997)
HIV-1 TAR w/arginine	HIV 1 TAR	1ARJ	1995	29	20	B	N	Aboul-ela *et al.* (1995)
HIV-2 TAR w/argininamide	HIV 2 TAR	1AJU/1AKX	1997	31	1/20	B	N	Brodsky & Williamson (1997)
HIV-1 Rev - RRE	HIV Rev-RRE	1ETF/1ETG	1996	39	1/19	B	N	Battiste *et al.* (1996)
RNA hairpin w/U1A protein	hp-U1A	1URN	1994	21		B	C	Oubridge *et al.* (1994)
Iron-responsive element	IRE	1AQO	1997	29	15	F	N	Addess *et al.* (1997)
Loop-E/loop D of *E. coli* 5 S	LpE-NMR	1A4D	1997	41	1	F	N	Dallas & Moore (1997)
P4-P6 of group I intron	P4-P6	1GID	1996	158		F	C	Cate *et al.* (1996)
Free RRE with A bulge	RBE3	1EBR	1996	30	5	F	N	Peterson & Feigon (1996)
Free RRE without A bulge	RBE3A	1EBQ	1996	29	5	F	N	Peterson & Feigon (1996)
RRE3-A with Rev22	RBE3A-Rev22	1EBS	1996	29	5	B	N	Peterson & Feigon (1996)
HIV-1 Rev aptamer RNA	Rev apt	1ULL	1996	35	7	B	N	Ye *et al.* (1996)
Sarcin-ricin loop from 28 S RNA	Sarcin-ricin	1SCL	1993	29	6	F	N	Szewczak & Moore (1995)
Splice-leader from *C. elegans*	Splice-leader	1SLO/1SLP	1996	26	1/16	F	N	Greenbaum *et al.* (1996)
Tetraloop acceptor	Tetraloop acc	1TLR	1997	23	20	F	N	Butcher *et al.* (1997)
Theophylline-binding RNA	Theoph apt	1EHT	1997	33	10	B	N	Zimmermann *et al.* (1997)
$tRNA^{Asp}$ w/synthetase	$tRNA^{Asp}$-synth	1ASY/1ASZ	1991	75		B	C	Ruff *et al.* (1991); Cavarelli *et al.* (1993)
Free yeast $tRNA^{Asp}$	$tRNA^{Asp}$	3TRA	1988	73		F	C	Westhof *et al.* (1988)
$tRNA^{Gln}$ w/synthetase	$tRNA^{Gln}$	1GTR/1GTS	1991	74		B	C	Rould *et al.* (1991); Perona *et al.* (1993)
$tRNA^{Gln}$ w/mutant synthetases	$tRNA^{Gln}$ mut	1QRS/ 1QRT/1QRU	1996	74		B	C	Arnez & Steitz (1996)
Yeast initiator tRNA	tRNAi	1YFG	1991	74		F	C	Basavappa & Sigler (1991)
Free yeast $tRNA^{Phe}$	$tRNA^{Phe}$	1TRA/6TNA	1986/ 1978	76		F	C	Westhof & Sundaralingam (1986); Sussman *et al.* (1978)
$tRNA^{Phe}$ w/EF-Tu	$tRNA^{Phe}$ -EFTu	1TTT	1995	75		B	C	Nissen *et al.* (1995)
$tRNA^{Ser}$ w/synthetase	$tRNA^{Ser}$	1SER	1994	55		B	C	Biou *et al.* (1994)
UU mismatches flanked by GC	UU mm	280D	1996	24		F	C	Lietzke *et al.* (1996)

[a] Number of structures is given only for NMR. A 1 indicates a minimized, averaged structure. For crystallography, one structure exists for each PDB accession code.
[b] F, Free (unbound); B, bound to non-RNA molecule.
[c] C, structure from X-ray crystallography; N, structure from NMR.

Table 2. Characteristics of the structures

Characteristic	No. structures (%)	No. nucleotides (%)	Avg size (range)
From X-ray crystallography	14 (41)	904 (60)	64.6 (24 - 158)
From NMR	20 (59)	591 (40)	29.6 (19 - 41)
Unbound	17 (50)	785 (53)	46.2 (19 - 158)
Bound to protein[a]	10 (29)	505 (34)	50.5 (21 - 75)
Bound to non-protein	7 (21)	205 (14)	29.3 (25 - 35)
tRNA	8 (24)	576 (39)	72.0 (55 - 76)
Total	34 (100)	1495 (100)	44.0 (19 - 158)

[a] A polypeptide was co-crystalized with the RNA or, for NMR structures, a polypeptide was in solution with the RNA. Structures with a single amino acid are classified as non-protein.

Table 3. Comparison of base stacking in 3D structures with free energy increment for dangling end stacking shown in bold and numbered (bases in GA pairs are omitted)

Structure and motif[a]	F/B[b]	Sequence	HB[c]	$\Delta G^0_{37,stack}$[d] kcal/mol	S/U[e]	Structure and motif[a]	F/B[b]	Sequence	HB[c]	$\Delta G^0_{37,stack}$[d] kcal/mol	S/U[e]
A. *Strong stacking*											
Arg-apt (il)	B	**CG 38** GG	0	−1.7	S	Tetraloop acc (il)	F	**CU 5** GU	0.5	−1.2	S
Cit-apt (il)	B	**CG 38** GG	0	−1.7	S	Sarcin-ricin (il)	F	**CU 7** GC	0.3	−1.2	S
hp-U1A (hp)	B	**CA 6** GC	-[f]	−1.7	S	$tRNA^{Asp}$ (hp)	F	**CU 32** GC	0	−1.2	S
$tRNA^{Gln}$ (hp)	B	**CA 13** GA	1	−1.7	S	$tRNA^{Asp}$-synth (hp)	B	**CU 632** GC	0	−1.2	S
$tRNA^{Gln}$ (j)	B	**CA 26** GC	0	−1.7	S	UU mm (il)	F	**CU 6** GU	2	−1.2	S
$tRNA^{Gln}$ mut (hp)	B	**CA 13** GA	1	−1.7	S	Rev apt (il)	B	**CU 23** GC	0	−1.2	U
$tRNA^{Gln}$ mut (j)	B	**CA 26** GC	0	−1.7	S	16S-aminogly (il)	B	**GA 20** CA	0	−1.1	U
tRNAi (hp)	F	**CA 14** GA	0	−1.7	S	P4-P6 (il)	F	**GA 113** CA	1	−1.1	S
tRNAi (j)	F	**CG 26** GG	0	−1.7	S	tRNAi (hp)	F	**GA 54** CA	0	−1.1	S
$tRNA^{Phe}$ (hp)	F	**CA 14** GA	0	−1.7	S[g]	$tRNA^{Asp}$ (hp)	F	**UA 14** GA	1	−0.8	S
$tRNA^{Phe}$-EFTu (hp)	B	**CA 14** GA	0.3	−1.7	S	$tRNA^{Asp}$-synth (hp)	B	**UA 614** GA	1	−0.8	S
P4-P6 (j)	F	**GG 176** CG	0	−1.3	S	Tetraloop acc (il)	F	**UA 18** GA	0	−0.8	U
RBE3 (il)	F	**GG 24** CG	2	−1.3	U	P4-P6 (il)	F	**UA 248** GA	0	−0.8	U
RBE3A (il)	F	**GG 23** CG	2	−1.3	U	P4-P6 (il)	F	**UA 206** GA	0	−0.8	S
RBE3A-Rev22 (il)	B	**GG 23** CG	2	−1.3	S	Free TAR (hp)	F	**CC 30** GA	0	−0.8	S
HIV Rev-RRE (il)	B	**GG 71** CG	1	−1.3	S	HIV 1 TAR (hp)	B	**CC 30** GA	0	−0.8	S
HDV-hp (hp)	F	**CU 7** GC	0.5	−1.2	S	HIV 2 TAR (hp)	B	**CC 30** GA	0	−0.8	S[h]
P4-P6 (il)	F	**CU 224** GU	0	−1.2	S	Splice-leader (hp)	F	**CC 13** GA	0	−0.8	S
B. *Intermediate stacking*											
16S-aminogly (il)	B	**GU 5** CU	0	−0.6	S	$tRNA^{Phe}$ (hp)	F	**AC 32** UA	0	−0.5	S
$tRNA^{Asp}$ (j)	F	**AU 8** U	-	−0.6	U	$tRNA^{Phe}$-EFTu (hp)	B	**AC 32** UA	0.7	−0.5	S
$tRNA^{Asp}$-synth (j)	B	**AU 608** U	-	−0.6	U	16S-aminogly (il)	B	GA **CA 7**	0	−0.5	S
$tRNA^{Asp}$ (hp)	F	**GU 54** CU	0	−0.6	S	IRE (il)	F	GC **CA 23**	0	−0.5	S
$tRNA^{Asp}$-synth (hp)	B	**GU 654** CU	0	−0.6	S	P4-P6 (il)	F	GA **CA 207**	1	−0.5	S
$tRNA^{Gln}$ (j)	B	**AU 8** U	-	−0.6	U	Theoph apt (il)	B	GC **CA 7**	0	−0.5	U
$tRNA^{Gln}$ (hp)	B	**AU 32** UU	1	−0.6	S	$tRNA^{Gln}$ (j)	B	GC **CA 26**	1	−0.5	S
$tRNA^{Gln}$ (hp)	B	**GU 54** CU	0	−0.6	S	$tRNA^{Gln}$mut (j)	B	GC **CA 26**	1	−0.5	S
$tRNA^{Gln}$ mut (j)	B	**AU 8** U	-	−0.6	U	tRNAi (hp)	F	GA **CA 60**	0	−0.5	U
$tRNA^{Gln}$ mut (hp)	B	**AU 32** UU	1	−0.6	S	tRNAi (hp)	F	GC **CA 38**	0	−0.5	S
$tRNA^{Gln}$ mut (hp)	B	**GU 54** CU	0	−0.6	S	IRE (hp)	F	**GC 7** CA	0	−0.4	U
tRNAi (j)	F	**GU 8** CG	0	−0.6	U	Theoph apt (il)	B	**GC 27** CA	0	−0.4	U
$tRNA^{Phe}$ (hp)	F	**GU 54** CC	0	−0.6	S	$tRNA^{Gln}$ (j)	B	**GC 44** CA	1	−0.4	S
$tRNA^{Phe}$-EFTu (hp)	B	**GU 54** CC	0	−0.6	S	$tRNA^{Gln}$mut (j)	B	**GC 44** CA	1	−0.4	S
$tRNA^{Ser}$ (hp)	B	**GU 54** CU	0	−0.6	S	tRNAi (hp)	F	**GC 32** CA	0	−0.4	S
$tRNA^{Ser}$ (j)	B	**GU 8** C	-	−0.6	U						
BIV TAR-Tat (il)	B	**GU 10** CC	0	−0.6	U						

Table 3. *Continued*.

Structure and motif[a]	F/B[b]	Sequence	HB[c]	$\Delta G^{\circ}_{37,stack}$[d] kcal/mol	S/U[e]	Structure and motif[a]	F/B[b]	Sequence	HB[c]	$\Delta G^{\circ}_{37,stack}$[d] kcal/mol	S/U[e]
C. *Weak stacking*											
BIV TAR-Tat (il)	B	**GU CC 25**	0	−0.3	S	BIV TAR-Tat (il)	B	**UC AU 12**	0	−0.2	U
Hammerhead (j)	F	**UC AA 23**	0	−0.3	S	Free TAR (hp)	F	**CC GA 35**	0	−0.2	U
Hammerhead (j)	F	**UC AC 36**	0	−0.3	U	HIV 1 TAR (hp)	C	**CC GA 35**	0.1	−0.2	U
HDV-hp (hp)	F	**CU GC 13**	0.5	−0.3	S	HIV 2 TAR (hp)	B	**CC GA 35**	0	−0.2	U
tRNAi (j)	F	**G CC 48**	-	−0.3	U	HIV Rev-RRE (il)	B	**GG CG 48**	1	−0.2	S
Sarcin-ricin (il)	F	**CU GC 23**	0.3	−0.3	U	P4-P6 (j)	F	**GG CG 164**	0	−0.2	U
$tRNA^{Asp}$ (hp)	F	**CU GC 38**	0	−0.3	S	RBE3 (il)	F	**GG CG 6**	2	−0.2	U
$tRNA^{Asp}$-synth (hp)	B	**CU GC 638**	0	−0.3	U	RBE3A (il)	F	**GG CG 6**	2	−0.2	S
$tRNA^{Phe}$ (hp)	F	**GU CC 60**	0	−0.3	U	$tRNA^{Phe}$-EFTu (hp)	B	**CA GA 21**	0.3	−0.2	U
$tRNA^{Phe}$-EFTu (hp)	B	**GU CC 60**	0	−0.3	U	16S-aminogly (il)	B	**GU CU 23**	0	−0.1	S
Rev apt (il)	B	**CU GC 10**	0	−0.3	S	BIV TAR-Tat (il)	B	**UC 25 AU**	0	−0.1	S
$tRNA^{Ser}$ (j)	B	**C GC 48**	-	−0.3	U	Hammerhead (j)	F	**UC 6 AC**	0	−0.1	S
$tRNA^{Gln}$ (j)	B	**G CC 48**	-	−0.3	U	Hammerhead (j)	F	**UC 36 AA**	0	−0.1	U
$tRNA^{Gln}$ (j)	B	**CA GC 9**	0	−0.3	U	$tRNA^{Asp}$ (j)	F	**G CU 48**	-	−0.1	U
$tRNA^{Gln}$mut (j)	B	**G CC 48**	-	−0.3	U	$tRNA^{Asp}$-synth (j)	B	**G CU 648**	-	−0.1	U
$tRNA^{Gln}$mut (j)	B	**CA GC 9**	0	−0.3	U	$tRNA^{Asp}$ (hp)	F	**GU CU 60**	0	−0.1	U
$tRNA^{Phe}$ (j)	F	**G CC 48**	-	−0.3	U	$tRNA^{Asp}$-synth (hp)	B	**GU CU 660**	0	−0.1	U
$tRNA^{Phe}$-EFTu (j)	B	**G CC 48**	-	−0.3	U	$tRNA^{Gln}$ (hp)	B	**GU CU 60**	0	−0.1	U
$tRNA^{Phe}$ (hp)	F	**AC UA 38**	0	−0.3	S	$tRNA^{Gln}$mut (hp)	B	**GU CU 60**	0	−0.1	U
$tRNA^{Phe}$-EFTu (hp)	B	**AC UA 38**	0.7	−0.3	S	$tRNA^{Phe}$ (j)	F	**UU 8 A**	-	−0.1	U
RBE3A-Rev22 (il)	B	**GG CG 6**	2	−0.2	S	$tRNA^{Phe}$-EFTu (j)	B	**UU 8 A**	-	−0.1	U
Splice-leader (hp)	F	**CC GA 21**	0	−0.2	U	$tRNA^{Ser}$ (hp)	B	**GU CU 60**	0	−0.1	U
$tRNA^{Gln}$ (hp)	B	**CA GA 22**	1	−0.2	S	Arg-apt (il)	B	**CG GG 9**	0	0	U
$tRNA^{Gln}$mut (hp)	B	**CA GA 22**	1	−0.2	S	Cit-apt (il)	B	**CG GG 9**	0	0	S
$tRNA^{Gln}$ (hp)	B	**AU UU 38**	1	−0.2	S	P4-P6 (il)	F	**CU GU 249**	0	0	U
$tRNA^{Gln}$mut (hp)	B	**AU UU 38**	1	−0.2	S	Tetraloop acc (il)	F	**CU GU 19**	0.5	0	U
tRNAi (hp)	F	**CA GA 21**	0	−0.2	U	tRNAi (j)	F	**CG GG 9**	0	0	U
$tRNA^{Phe}$ (hp)	F	**CA GA 21**	0	−0.2	U	UU mm (il)	F	**CU GU 19**	2	0	S

[a] Motif: hp, hairpin; il, internal loop; j, junction.
[b] F, Free (unbound); B, bound to non-RNA molecule.
[c] The number of hydrogen bonds observed between the 5′ and 3′ unpaired bases, averaged per structure. A dash indicates that there is no opposing unpaired base, which occurs in junctions.
[d] The ΔG^{0}_{37} for stacking of a single dangling end in an oligonucleotide duplex (Turner *et al.*, 1988).
[e] S, Stacked; U, unstacked. Determination of stacked *versus* unstacked is made according to distance, angle, and overlap requirements, as defined in Materials and Methods.
[f] The coordinates of the 5′ C are not provided in hp-U1A, so hydrogen bonds are not available.
[g] The overlap criterion is met in the orthorhombic form, 6TNA, but not in the monoclinic form, 1TRA.
[h] The angle for 1AJU is 30.1°, just outside the 30° required by the stacking criterion. For 1AKX the angle is 16.9°, so average angle fulfils stacking requirement.

NMR structures. There is little difference in patterns between bound and unbound RNAs, and among internal loops, junctions, and hairpins of five or more nucleotides. Thus, all are included in Table 3.

Table 3 shows that 30 of 36 sequences (83%) with a stacking free energy $\leqslant -0.8$ kcal/mol are stacked in 3D structures. Thus a very favorable stacking free energy as a dangling end is a good predictor of stacking in a 3D structure. The six that are not stacked are often associated with a strong interaction for stabilizing the unstacked conformation: (1 and 2) in Rev-binding element structures, RBE3A and RBE3 (Peterson & Feigon, 1996), which are not bound to Rev, G23 is expected to be stacked with a $\Delta G^0_{37,stack}$ of −1.3 kcal/mol but G23 has a *syn* glycosidic torsion angle, forms a GG base-pair, and is over 4 Å away from the adjacent GC pair; this two-hydrogen bond GG pair apparently provides more than 1.3 kcal/mol of stabilization. Interestingly, the same G has an *anti* glycosidic torsion angle and is stacked when Rev is bound, in RBE3A-Rev22 (Peterson & Feigon, 1996) and in the similar G71 of HIV Rev-RRE (Battiste *et al.*, 1996). Perhaps the energy derived from stacking of this G facilitates binding of Rev. (3) The Rev aptamer RNA (Ye *et al.*, 1996) has an unusual internal loop where stacking of U23 on C22-G11 is disrupted; a C10-G25 Watson-Crick pair forms and U23 interacts with A26-U9. (4) In 16S-aminogly (Fourmy *et al.*, 1996), the bound paromomycin interacts with the internal loop, apparently disrupting the favorable $\overset{GA}{C}$ stacking interaction. (5 and 6) For the tetraloop acceptor of the group I intron, A248 in $\overrightarrow{\overset{UA}{G}}$ is unstacked in both the crystal structure (Cate *et al.*, 1996) and the NMR structure (Butcher *et al.*, 1997), even though its $\Delta G^0_{37,stack} = -0.8$ kcal/mol (Freier *et al.*, 1986a). In the crystal structure, there is an AA platform 5′ to the G, which stacks on the GU pair, preventing the 3′ A from direct stacking. In the NMR structure, the 5′ A also intervenes in the stack, but does not make an AA platform. The $\overrightarrow{\overset{UA}{G}}$ at position 206 of P4-P6 is stacked (see Table 3).

For stacking interactions in the range of −0.4 to −0.7 kcal/mol, there is less preference for stacking: only 21 of 32 bases at helix ends are stacked on the helix (66%). The most stable 5′ end, $\overrightarrow{\overset{G}{CA}}$ falls in this range and is stacked six of eight times.

Nucleotides with $\Delta G^0_{37,stack} \geqslant -0.3$ kcal/mol as dangling ends in oligonucleotides have the least preference for stacking; only 19 of 56 are stacked (34%). Thus there is a clear, but not absolute, correlation between dangling end stability and stacking in 3D structures.

Table 3 lists the number of hydrogen bonds between the 5′ and 3′ dangling bases. This number is fractional if the proton donor and acceptor geometry satisfies hydrogen bond requirements in only a fraction of the series of structures given by NMR or crystallography (see Materials and Methods). For weak stacking, there is a correlation between stacked bases and non-Watson-Crick hydrogen bonding. Of the 12 weakly stacking ends that have hydrogen bonds ($\geqslant 0.5$), ten are stacked (83%). In contrast, of 34 weakly stacking ends with an opposing base and <0.5 hydrogen bond between the 5′ and 3′ dangling bases only nine are stacked (26%). Thus hydrogen bonds appear to modulate stacking interactions. It is difficult, however, to predict hydrogen bonds between the dangling ends; UC, CA, GG, AA, and UU mismatches at weakly stacking ends sometimes have hydrogen bonds and sometimes do not.

Stacking by and on GA

Guanine-adenine pairs commonly cap helix ends (Gautheret *et al.*, 1994). In our analysis of a database of large and small subunit rRNA secondary structures (Gutell *et al.*, 1993; Gutell, 1994), 34% of the 34,100 helixes terminate in a GA. Thus it would be useful to predict stacking by and on GA pairs. The database of Table 1 contains GA pairs in both imino and sheared conformations (Burkard *et al.*, 1999b). For stacking of GA on adjacent Watson-Crick pairs, all 3′ bases and ~90% of 5′ bases of the GA are stacked (not shown).

For stacking on GA pairs, SantaLucia *et al.* (1991) showed that $\overrightarrow{\overset{GA}{A}}$ and $\overrightarrow{\overset{AG}{G}}$ dangling ends each favor duplex formation by about −0.6 kcal/mol. It is unclear whether these terminal GA pairs are in the imino or sheared conformation and the stacking interactions may depend on the pair conformation. In 3D structures, base stacking on imino and on sheared GA pairs were analyzed separately. Table 4 presents the results.

For imino GA pairs, the 12 3′ dangling ends are stacked regardless of sequence. A 3′ G is most common with eight occurrences; three times as $\overrightarrow{\overset{GG}{A}}$ and five times as $\overrightarrow{\overset{AG}{G}}$, so strong stacking is most certain for these ends. There are nine 3′ dangling ends adjacent to sheared GAs, of which seven are stacked. Of these, a sheared GA is followed by a reverse Hoogsteen AU pair (Burkard *et al.*, 1999b) in Sarcin-ricin (Szewczak & Moore, 1995) and twice in each NMR and crystal structure of the loop E sequence (Correll *et al.*, 1997; Dallas & Moore, 1997). Each reverse-Hoogsteen AU sequence is of the form $\overrightarrow{\overset{XGA}{YAU}}$, where XY is a Watson-Crick pair. This appears to be a common structural motif; in the hammerhead ribozyme, a sheared GA is followed by an AU that has one hydrogen bond of a reverse Hoogsteen pair (Scott *et al.*, 1995). In all these structures, the A in the Hoogsteen pair is stacked on the adjacent GA. Apparently, the purine-purine GA pair provides a large platform for favorable 3′ end stacking in both imino and sheared conformations.

Table 4. Stacking on GA pairs

Structure and motif	Sequence	HB[a]	S/U[b]	Structure and motif	Sequence	HB[a]	S/U[b]
A. *Imino hydrogen bonded GAs*							
5 S loop E	**GG 76**	1	S	$tRNA^{Phe}$	**AG 45**	0	S
(il)	**AG**			(j)	**G**		
LpE-NMR	**GG 76**	1	S	$tRNA^{Phe}$-EFTu	**AG 45**	0	S
(il)	**AG**			(j)	**G**		
FMN-apt	**AG 9**	0.6	S	5 S loop E	**GG**	1	S
(il)	**GG**			(il)	**AG 100**		
HIV Rev-RRE	**GG 48**	0	S	LpE-NMR	**GG**	1	S
(il)	**AU**			(il)	**AG 100**		
Rev apt	**GA 7**	0	S	FMN-apt	**AG**	0.6	S
(il)	**AA**			(il)	**GG 27**		
P4-P6	**AC 197**	0	S	HIV Rev-RRE	**GG**	0	U
(il)	**GA**			(il)	**AU 72**		
Theoph-apt	**AU 6**	0.5	S	Rev apt	**GA**	0	U
(il)	**GA**			(il)	**AA 29**		
$tRNA^{Asp}$-synth	**AG 645**	0.5	S	P4-P6	**AC**	0	U
(j)	**G**			(il)	**GA 125**		
$tRNA^{Asp}$	**AG 45**	0	S	Theoph apt	**AU**	0.5	S
(j)	**G**			(il)	**GA 28**		
tRNAi	**AU 45**	0	S	tRNAi	**AU**	0	S
(j)	**GC**			(j)	**GC 25**		
B. *Sheared GAs*							
5 S loop E	**GA[c] 73**	2	S	5 S loop E	**GA[c]**	2	S[d]
(il)	**AU**			(il)	**AU 77**		
LpE-NMR	**GA[c] 73**	2	S	LpE-NMR	**GA[c]**	2	U
(il)	**AU**			(il)	**AU 77**		
5 S loop E	**GA[c] 99**	2	S	5 S loop E	**GA[c]**	2	S[d]
(il)	**AU**			(il)	**AU 103**		
LpE-NMR	**GA[c] 99**	2	S	LpE-NMR	**GA[c]**	2	S
(il)	**AU**			(il)	**AU 103**		
CurveHelix	**GA 6**	2	U	CurveHelix	**GA**	2	S
(il)	**AA**			(il)	**AA 19**		
Hammerhead	**AA 23**	1	S	Hammerhead	**AA**	1	U
(j)	**GU**			(j)	**GU 10**		
P4-P6	**GC 165**	0	U	P4-P6	**GC**	0	U
(j)	**AC**			(j)	**AC 138**		
Sarcin-ricin	**GA[c] 20**	2	S	Sarcin-ricin	**GA[c]**	2	S
(il)	**AU**			(il)	**AU 11**		
$tRNA^{Ser}$	**GA 14**	1	S	$tRNA^{Ser}$	**GA**	1	U
(hp)	**AA**			(j)	**AA 21**		

[a] The number of hydrogen bonds observed between the 5′ and 3′ dangling bases, averaged over the series of structures.
[b] S, Stacked; U, unstacked.
[c] This AU pair is in the reverse Hoogsteen conformation.
[d] U77 and U103 meet the overlap criterion in 356D but not in 354D.

Functional groups and stacking interactions

The interactions that underlie stacking free energies are poorly understood. Several studies of dinucleoside monophosphate stacking by ultraviolet absorbance, circular dichroism, and NMR methods agree that dimers containing U stack less than other dinucleoside monophosphates (Frechet *et al.*, 1979; Davis & Tinoco, 1968). Stacking interactions are generally attributed to London interactions that require filled π electron orbitals; thus it is not surprising that U, with little aromaticity, stacks less than C. For dangling ends, however, stacking is stronger in $\overrightarrow{\substack{CU\\G}}$, where $\Delta G^0_{37,stack} = -1.2$ kcal/mol, than for $\overrightarrow{\substack{CC\\G}}$, where $\Delta G^0_{37,stack} = -0.8$ kcal/mol (Freier *et al.*, 1983). The orientation of the GC pair is important; for $\overrightarrow{\substack{CU\\G}}$ the $\Delta G^0_{37,stack}$ is −0.6 kcal/mol, half as large as for $\overrightarrow{\substack{GU\\C}}$. Inspection of $\overrightarrow{\substack{CU\\G}}$ in 3D structures suggests that a likely electrostatic interaction is the overlap of O-4 of uridine with the cytosine amino group (for example, see Figure 1(a)). To test this hypothesis, thermodynamics were measured for U stacking on PG and CI pairs, where P is 2-pyrimidinone and I is inosine. The analog, 2-pyrimidinone, has a hydrogen atom in place of the amino group at position 4 in cytosine, and inosine has a hydrogen atom in place of the amino at position 2 of guanine. Figure 2 compares these with the Watson-Crick CG pair. The PG and CI pairs can form two hydrogen bonds instead of the three expected

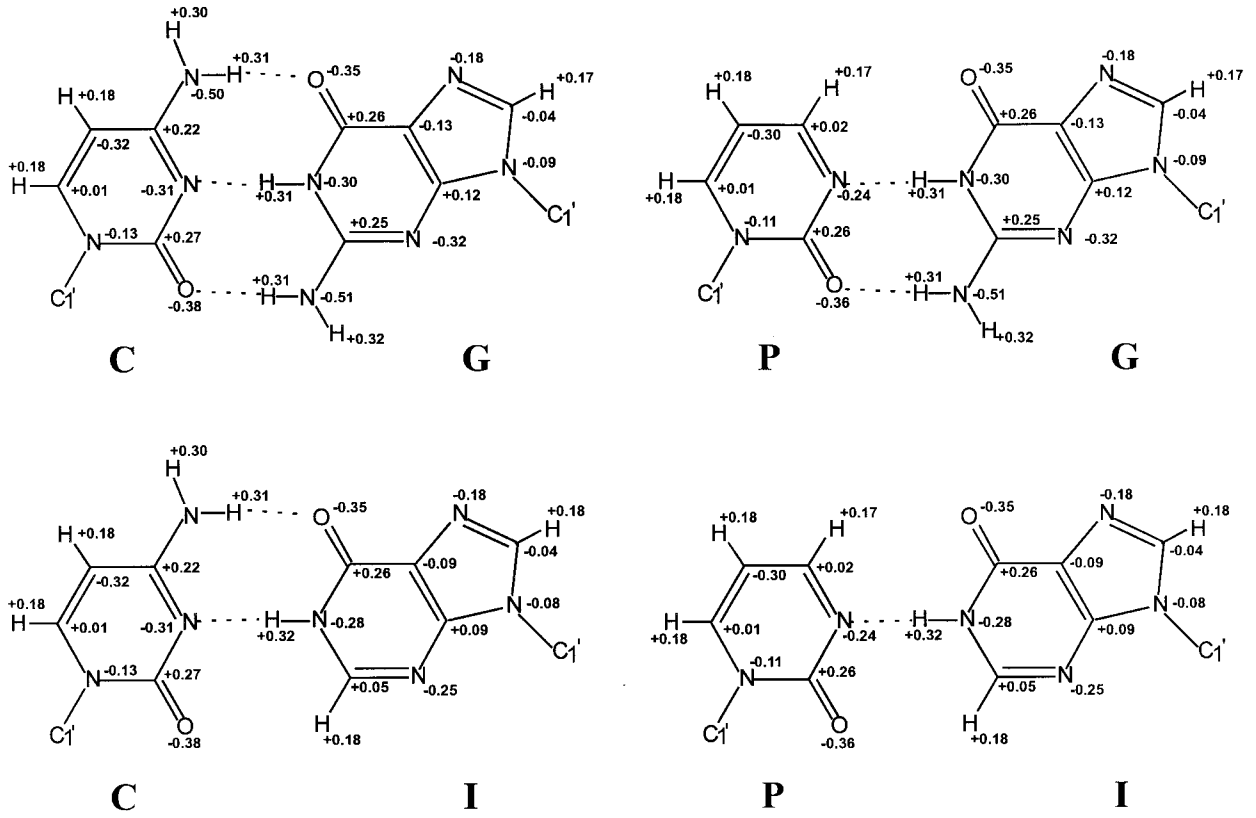

Figure 2. CG base-pair and analogs with 2-pyrimidinone (P) and inosine (I). These base analogs each lack an amino group for formation of a hydrogen bond. Löwdin atomic partial charges are given for each atom. Partial charges were calculated for individual bases, methylated at the sugar position (C-1′), using 6-31G* basis set with *ab initio* SCF methods (Schmidt *et al.*, 1993). The bond order of N3-C4 in C is 1.50, compared to 1.64 in P. Calculated dipole moments are 6.78, 6.41, 8.22, and 6.12 Debye for C, P, G, and I, respectively.

for CG. The partial charges in Figure 2 show that the amino group and the adjacent carbon atom have a net +0.33 charge for C, but for 2-pyrimidinone, the total charge of this region is reduced to +0.19 and does not extend as far from the aromatic ring. This will reduce the electrostatic interaction with a stacked base. The modification may also affect π orbital occupancy and therefore London attraction.

For each duplex, seven to nine melt curves were measured over ~100-fold range in duplex concentration. Figure 3 shows the van't Hoff plots of the UV melting data for self-complementary oligomers with dangling U on CG, PG, and CI along with the appropriate reference duplexes. The plots are linear, and melting temperature (T_m) increases with concentration as expected for a uniformly bimolecular transition. Table 5 contains thermodynamic parameters derived from these plots and from fitting individual curves with a two-state model (Petersheim & Turner, 1983; McDowell & Turner, 1996).

Table 6 summarizes the dangling end free energies calculated as:

$$\Delta G^{\circ}_{37,\text{stack}} = 1/2(\Delta G^{\circ}_{37,\text{danglingU}} - \Delta G^{\circ}_{37,\text{ref}})$$

where $\Delta G^{0}_{37,\text{danglingU}}$ is the Gibb's free energy change observed for the self-complementary duplex with two 3′ dangling U bases, and $\Delta G^{0}_{37,\text{ref}}$ is for the same duplex without the dangling U bases. For $(GCAUGC\underline{U})_2$, the $\overrightarrow{\underset{G}{CU}}$ dangling end stabilizes the duplex by 1.0 kcal/mol; this is slightly less favorable than the 1.3 and 1.1 kcal/mol observed previously for addition of a 3′-terminal Up to 5′-GGCCp-3′ and 5′-GCGC-3′, respectively (Freier *et al.*, 1983, 1986b). Without the cytosine amino group, the $\overrightarrow{\underset{G}{PU}}$ dangling end provides half as much stabilization, 0.5 kcal/mol. Evidently, this amino group contributes to the stacking interaction. The guanine amino group is slightly less important; the $\overrightarrow{\underset{I}{CU}}$ dangling end provides 0.7 kcal/mol stabilization. Modeling of the $\overrightarrow{\underset{G}{CU}}$ with *A*-form geometry shows that the G amino group is about 4.6 Å from uridine O-2, whereas the cytosine amino group overlaps uridine O-4 at a distance of 3.3 Å. The guanine amino group proximity with uridine O-2 also appears in crystal structures with this dangling end: P4-P6 (3.4 Å), $tRNA^{Asp}$ (4.27 Å), and UU mm (4.11 Å). In each case it is more distant than the cytosine amino

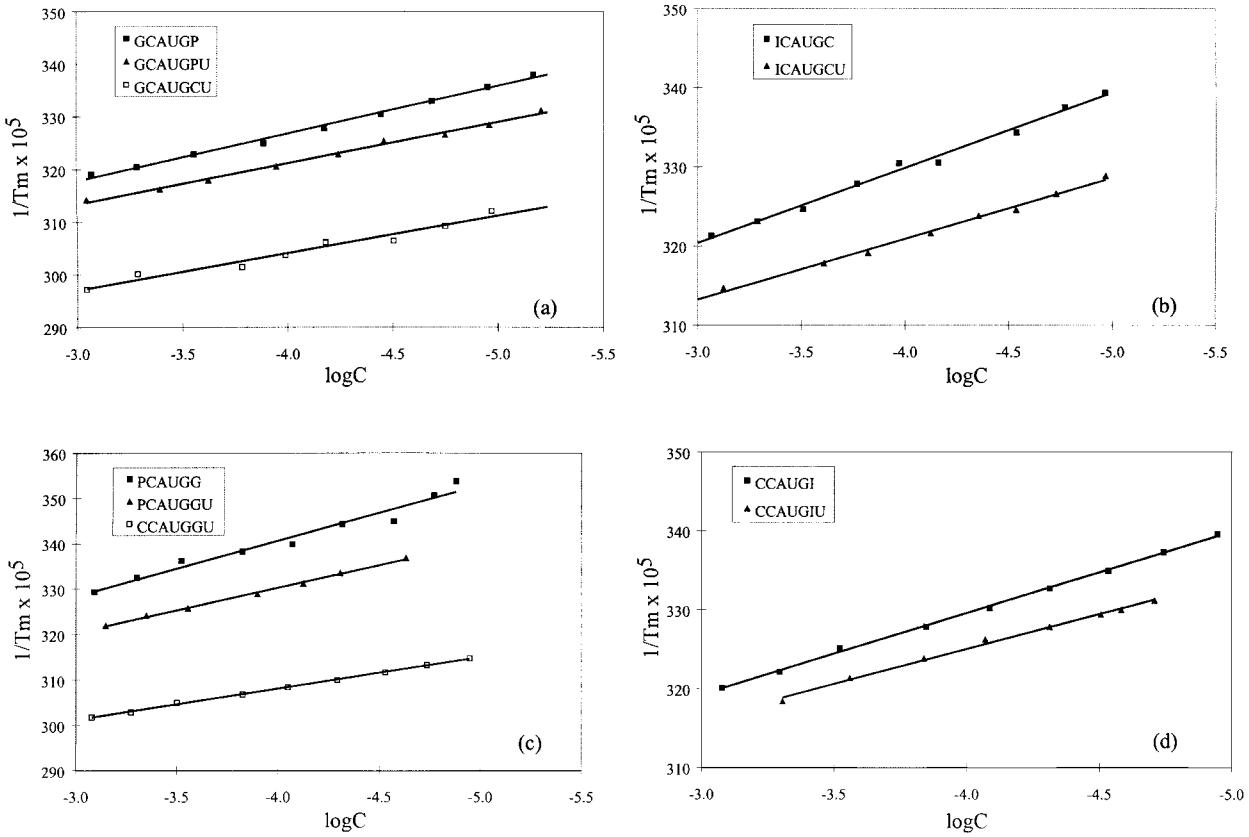

Figure 3. Representative van't Hoff plots from UV melting studies of the oligonucleotide duplexes in Table 5; inverse melting temperature *versus* log-concentration. Filled squares indicate duplex containing a base analog and no dangling U; filled triangles are for the duplex with a dangling U. Open squares represent duplexes with a dangling U and only AU or GC pairs. (a) $\overrightarrow{\substack{\text{PU}\\\text{G}}}$ (b) $\overrightarrow{\substack{\text{CU}\\\text{I}}}$ (c) $\overrightarrow{\substack{\text{GU}\\\text{P}}}$ (d) $\overrightarrow{\substack{\text{IU}\\\text{C}}}$.

group overlap with uridine O-4. In the Sarcin-ricin loop structure (Figure 1(a)), however, the G24 amino group does not overlap U7 O-2.

While the above thermodynamic results are consistent with the hypothesis that the cytosine amino group is important for an electrostatic stacking interaction with uridine, removal of an amino group can also have indirect effects on dangling end stability. Some possible sources are solvation energies, a change in base-pair conformation, and π orbital occupancy and thus London attraction. To control for these complications, a similar series of experiments was conducted for the $\overrightarrow{\substack{\text{GU}\\\text{C}}}$ dangling end (Tables 5 and 6). This dangling end provides ~0.7 kcal/mol stabilization to a duplex. The cytosine amino group is not important for this interaction; each $\overrightarrow{\substack{\text{GU}\\\text{P}}}$ dangling end stabilized the duplex by 0.5 kcal/mol. Absence of the G amino group led to a dangling end stability of 0.3 kcal/mol. This loss may arise from overlap of the G amino group with uridine O2; the distance is 3.9 Å in *A*-form geometry and 3.6-4.1 Å for G amino and O-2 of 5-methyluridine (ribothymidine) at position 54 in tRNA crystal structures (Sussman *et al.*, 1978; Westhof *et al.*, 1988; Biou *et al.*, 1994). The destabilizations from replacing G with I are nearly identical for $\overrightarrow{\substack{\text{CU}\\\text{G}}}$ and $\overrightarrow{\substack{\text{GU}\\\text{C}}}$, but the cytosine amino group is more important for stacking in $\overrightarrow{\substack{\text{CU}\\\text{G}}}$. These data support the hypothesis that a specific electrostatic interaction stabilizes the $\overrightarrow{\substack{\text{CU}\\\text{G}}}$ dangling end by a few tenths of a kcal/mol.

These thermodynamic measurements also provide estimates of hydrogen bond energy, as shown in Table 7. Removal of one hydrogen bond in the terminal base-pair destabilizes the duplex by 1.2-1.9 kcal/mol. This falls at the high end of the range of 0.4-1.6 kcal/mol expected from previous estimates of the free energy of a hydrogen bond in an RNA base-pair (Freier *et al.*, 1986b; Turner *et al.*, 1987a; SantaLucia *et al.*, 1991, 1992; Turner & Bevilacqua, 1993).

Interpretation of hydrogen bond energies is complicated by a pK_a of 2.5 at N-3 for 2-pyrimidinone (1-methyl derivative; Albert & Phillips, 1956) compared with 4.2 for C. This reflects the weakened ability of N-3 of 2-pyrimidinone to accept a

Table 5. Duplex thermodynamics in 1 M NaCl, pH 7, from optical melting

Duplex	T_m dependence on concentration[a] ΔG^0_{37} (kcal/mol)	ΔH^0 (kcal/mol)	ΔS^0 (eu)	T_m (°C) 0.1 mM	Curve fit[a] ΔG^0_{37} (kcal/mol)	ΔH^0 (kcal/mol)	ΔS^0 (eu)	T_m (°C) 0.1 mM
GCAUGC[b]	−7.38 ± 0.14	−62.3 ± 3.1	−177.2	45.7	−7.41 ± 0.23	−59.6 ± 6.2	−168.4 ± 19.8	46.3
GCAUGCU	−9.33 ± 0.29	−64.4 ± 5.0	−177.6 ± 15.3	55.7	−9.48 ± 0.24	−67.0 ± 1.0	−185.3 ± 2.6	55.7
GCAUGP	−4.99 ± 0.03	−50.1 ± 1.4	−146.9 ± 4.7	32.9	−5.00 ± 0.06	−52.9 ± 3.3	−154.7 ± 10.6	33.1
GCAUGPU	−5.92 ± 0.01	−58.9 ± 1.5	−170.9 ± 5.0	38.3	−5.93 ± 0.09	−56.4 ± 1.3	−162.8 ± 4.2	38.4
ICAUGC	−4.56 ± 0.04	−48.4 ± 1.6	−141.3 ± 5.4	30.0	−4.62 ± 0.13	−48.9 ± 4.5	−142.6 ± 14.7	30.4
ICAUGCU	−5.96 ± 0.02	−59.6 ± 1.9	−172.9 ± 6.2	38.5	−5.98 ± 0.09	−53.7 ± 2.8	−153.9 ± 9.1	38.8
ICAUGP[c]	(−2.9 ± 0.1)	(−65 ± 5)	(−200 ± 15)	24.3	(−3.7 ± 0.2)	(−39 ± 1)	(−115 ± 5)	22.0
CCAUGG[d]	−7.30 ± 0.1	−56.7 ± 2.2	−159.0 ± 6.7	46.3	−7.49 ± 0.2	−61.9 ± 5.1	−175.3 ± 11.0	46.4
CCAUGGU	−8.64 ± 0.04	−66.5 ± 0.8	−186.6 ± 2.5	51.5	−8.49 ± 0.05	−62.6 ± 1.7	−174.6 ± 5.5	51.6
PCAUGG	−3.58 ± 0.17	−37.3 ± 2.7	−108.7 ± 9.2	20.5	−3.52 ± 0.34	−39.2 ± 6.2	−115.1 ± 21.2	20.8
PCAUGGU	−4.57 ± 0.03	−45.6 ± 1.0	−132.1 ± 3.2	29.6	−4.76 ± 0.20	−39.6 ± 3.1	−112.4 ± 10.6	30.0
CCAUGI	−4.69 ± 0.02	−44.5 ± 0.6	−128.2 ± 1.9	30.3	−4.77 ± 0.12	−44.7 ± 4.9	−128.6 ± 16.0	30.9
CCAUGIU	−5.26 ± 0.02	−51.8 ± 1.4	−150.0 ± 4.6	34.5	−5.33 ± 0.10	−48.3 ± 3.1	−138.5 ± 10.1	34.8
PCAUGI[c]	(−3.2 ± 0.1)	(−46 ± 4)	(−139 ± 12)	21.2	(−3.5 ± 0.3)	(−38 ± 4)	(−110 ± 13)	19.6
PCAUGIU	−4.14 ± 0.03	−40.3 ± 0.6	−116.6 ± 2.2	25.6	−4.23 ± 0.18	−40.0 ± 4.1	−115.2 ± 13.6	26.2

[a] For T_m dependence on concentration, error limits are propagated from the errors in slope and intercept of the fit, according to the van't Hoff equation. For curve fit thermodynamics, the errors are the standard deviation of the values in the the seven to nine individual melt curves. Other error sources may make actual limits larger than shown.
[b] From Freier *et al.* (1986c).
[c] A large difference (>15%) between ΔH^0 from curve fit and from T_m dependence suggests that these duplexes deviate significantly from a two-state model.
[d] From McDowell *et al.* (1997).

hydrogen bond, compared with C, so energy per hydrogen bond is overestimated. The pK_a for I is 8.8, lower than the 9.2 for G, making it a slightly better hydrogen bond donor (SantaLucia *et al.*, 1991). These pK_a changes do not explain the differences of hydrogen bond estimates in Table 7, because for CG → PG, the hydrogen bond free energy is less than for CG → CI but for GC → GP the free energy is larger than for GC → IC. Apparently, other complicating effects such as hydration at the helix end are also significant.

Discussion

Thermodynamics have proved useful for prediction of RNA secondary structure (Mathews *et al.*, 1997, 1999). The present work demonstrates that thermodynamic parameters are also useful for predicting local 3D structure. In particular, bases are almost always stacked when they provide 0.8 kcal/mol or more stabilization. This includes $\overrightarrow{\substack{GA\\C}}$, $\overrightarrow{\substack{GG\\C}}$, and $\overrightarrow{\substack{CX\\G}}$, where X is A, C, G, or U. The $\overrightarrow{\substack{AG\\U}}$ motif has a free energy increment of −0.8 kcal/mol and is also expected to be stacked, although none is present in the current database. This correlation of thermodynamics and structure may be useful for modeling 3D structure. For example, the MC-SYM program described by Major *et al.* (1991) could choose to reduce the number of base conformations from ten to the two stacked conformations if there is a strong stacking interaction. Prediction of stacking can be supported further by phylogenetic conservation of a strongly stacking end.

Weakly stacking dangling ends

The 3′ dangling ends $\overrightarrow{\substack{UC\\A}}$ and $\overrightarrow{\substack{UU\\A}}$ and 5′ dangling

Table 6. Free energy attributable to 3′ dangling U

Dangling end	$\Delta G^0_{37,\ stack}$ (kcal/mol)	Dangling end	$\Delta G^0_{37,\ stack}$ (kcal/mol)
CU G	−0.98 ± 0.16	GU C	−0.67 ± 0.05
PU G	−0.47 ± 0.02	GU P	−0.50 ± 0.09
CU I	−0.70 ± 0.02	IU C	−0.29 ± 0.01
		IU P	(−0.5)[a]

T_m-dependence ΔG^0_{37} values were used for calculations. Errors were calculated from $\sigma = \frac{1}{2}(\sigma^2_{de} + \sigma^2_{ref})^{\frac{1}{2}}$, where σ_{de} is the error in ΔG^0_{37} from the duplex containing two dangling ends, and σ_{ref} is the error in the reference duplex ΔG^0_{37}. The factor, $\frac{1}{2}$, arises because there are two dangling ends per duplex. Other error sources may make limits ~0.1 kcal/mol.
[a] The number in parentheses was derived from duplexes that exhibit non-two-state behavior.

Table 7. Estimate of hydrogen bond free energy

Modification	$\Delta\Delta G^0_{37}$ per hydrogen bond (kcal/mol)
CG → PG	1.20 ± 0.07
CG → CI	1.41 ± 0.07
CG → PI	(1.1)[a]
GC → GP	1.86 ± 0.10
GC → IC	1.30 ± 0.05
GC → IP	(1.0)[a]

T_m-dependence ΔG^0_{37} values were used for calculations. Errors were calculated from $\sigma = \frac{1}{2}(\sigma^2_{de} + \sigma^2_{ref})^{\frac{1}{2}}$, where σ_{de} is the error in ΔG^0_{37} from the duplex containing two dangling ends, and σ_{ref} is the error in the reference duplex ΔG^0_{37}. Other error sources may make limits ~0.1 kcal/mol.

[a] Numbers in parentheses were derived from duplexes that exhibit non-two-state behavior.

ends except $\overrightarrow{\substack{AC\\G}}$ and possibly $\overrightarrow{\substack{GA\\U}}$ have free energy increments less favorable than −0.4 kcal/mol (Turner *et al.*, 1988) and are thus considered weakly stacking. If a backbone turn is required at a helix end, a weakly stacking base is best for the role. In $tRNA^{Phe}$, all five weakly stacking sequences are unstacked and four of five are associated with turns (Sugimoto *et al.*, 1987; Turner *et al.*, 1988). In the 1999 database, however, about a third of weakly stacking bases at helix termini are stacked and an unstacked base does not necessarily produce a turn. Sequence comparison may be useful for identifying backbone turns at weakly stacking 3′ ends. If the 3′ dangling end has a weak stacking free energy across all known species, a turn at this location may be likely. On the other hand, if the 3′ dangling end often has a free energy more favorable than −0.7 kcal/mol, there is no turn. For example, in known structures of tRNAs, the backbone bends between U8 and the preceding base-pair (Sussman *et al.*, 1978; Westhof & Sundaralingam, 1986; Westhof *et al.*, 1988; Basavappa & Sigler, 1991; Rould *et al.*, 1991; Ruff *et al.*, 1991; Cavarelli *et al.*, 1993; Perona *et al.*, 1993; Biou *et al.*, 1994; Nissen *et al.*, 1995; Arnez & Steitz, 1996). In the sequence alignment of tRNAs provided by Sprinzl *et al.* (1998), U8 is highly conserved. When preceded by a GC pair, $\overrightarrow{\substack{GU\\C}}$ ($\Delta G^0_{37,stack} = -0.6$) occurs 156 times, whereas $\overrightarrow{\substack{CU\\G}}$ ($\Delta G^0_{37,stack} = -1.2$) occurs only 18 times. Although we would not term $\overrightarrow{\substack{GU\\C}}$ as weakly stacking, the rarity of a dangling end with a strong stacking interaction suggests a backbone turn.

Structural predictions for unpaired bases at 5′ helix ends are perhaps less useful than for 3′ ends. In canonical *A*-form geometry, a purine 5′ to a pair has little overlap and a 5′ pyrimidine has none, but a 3′ base has significant overlap. By the criteria used here, the 5′ purine is categorized as stacked, but it requires less perturbation than the 3′ end to become unstacked. Hydrogen bonding between the 5′ base and the opposing 3′ base may often drive 5′ end stacking (see Table 3).

The stacking pattern differs little between bound and unbound RNAs; in the database of structures, a strong base-stacking interaction is disrupted by a bound molecule only for the aminoglycoside-binding domain. It is not surprising to see such a small difference if binding depends on many weak contacts, as expected for achieving specificity (Herschlag, 1991). The weak interactions may influence overall RNA shape, but are unlikely to disrupt strong local interactions such as base stacking.

For the purposes of this work, base stacking is defined according to distance, angle, and overlap requirements. This definition is arbitrary and rigid, but it is useful for identifying stacking in 3D structures. The bases that we classify as stacked form a homogeneous population: the means and standard deviations are 3.29(±0.20) Å for the nearest distance and 11.4(±6.7)° for the stack angle. In contrast, those classified as unstacked are heterogeneous: distance is 6.32(±2.22) Å and angle is 48.1(±23.4)°. (Complete distances, angles, and overlap values are provided as Supplementary Material.) Although the unstacked population is heterogeneous, few bases were classified as unstacked because of a single distance, angle, or overlap slightly outside of the stacked range. Therefore, other definitions of base stacking would yield similar results. By most definitions, the purine bases will be considered stacked more often than pyrimidines simply because of their greater size, but this is acceptable because this is likely a source of favorable purine stacking.

Electrostatics of stacking

The free energy increment of −1.2 kcal/mol associated with the $\overrightarrow{\substack{CU\\G}}$ motif (Freier *et al.*, 1983, 1985, 1986b) is more favorable than expected from π-electron overlap. Inspection of 3D structures suggested a favorable electrostatic interaction between the C amino group and O-4 of U, and this is supported by thermodynamic experiments. Although the maximum effect is 0.5 kcal/mol, this is a significant fraction of the net stacking free energy. Our methods have the precision required to measure this difference, in part, because there are two dangling ends per duplex and errors are thereby halved. A similar interaction may be important for other RNA motifs. For example, the $\overrightarrow{\substack{GGUC\\CUGG}}$ motif is unusually stable when compared with tandem GU pairs in other contexts (He *et al.*, 1991; McDowell & Turner, 1996). Structures containing $\overrightarrow{\substack{GGUC\\CUGG}}$ (McDowell & Turner, 1996; Cate *et al.*, 1996) have the C amino group in the GC pairs overlapped with O-4 of U.

If electrostatics play an important role in stacking, it may be possible to design intercalators to enhance specificity for RNA binding. For example, $\overrightarrow{\substack{CU\\G}}$ is associated with a free energy increment of −1.2 kcal/mol at 37 °C and is stacked in known

3D structures (Table 3), whereas $\overrightarrow{\text{UU}}_{\text{A}}$ has $\Delta G^0_{37,\text{stack}} = -0.1$ kcal/mol and is not stacked. This suggests a ring system with a charge distribution similar to that of U could provide discrimination between terminal CG and UA pairs. Similarly, specific charge distributions could be used to target non-Watson-Crick pairs. Although intercalators are expected to recognize only adjacent bases, two or more such intercalators could be covalently linked to achieve specific binding to unique sequences or spatial orientations of Watson-Crick helixes.

The thermodynamic data highlight relative free energy contributions of functional groups for hydrogen bonds and for stacking. Elimination of a functional group involved in a hydrogen bond always destabilized the duplex by over 1 kcal/mol, whereas the stacking energy contribution of a single group was found to be a maximum of 0.5 kcal/mol. This is important for understanding the effects of functional group substitution. For example, interference mapping of RNA identifies chemical groups important for function. Interference from substitution of these groups is usually attributed to loss of hydrogen bonds or to steric interactions (Strobel & Shetty, 1997; Ortoleva-Donnelly *et al*., 1998). Our data suggest that interference mapping may additionally identify functional groups important for stacking interactions, but that these will show attenuated interference. Thus, stacking is significant enough to explain the observation reported by Ortoleva-Donnelly *et al*. (1998), that removal of amino-6 in A218 of the *Tetrahymena* ribozyme moderately interferes with activity even though this group apparently is not involved in a hydrogen bond.

Outlook

The catalog of 3D RNA structures containing non-helical motifs is expanding more slowly than the catalog of interesting RNA sequences. Thermodynamics of small model systems are useful for filling this gap between sequence and structure. Here, we have illustrated that thermodynamics, long used for prediction of RNA secondary structure, can be useful for predictions about 3D structure. Ultimately, a combination of thermodynamic data, molecular modeling, and a structural catalog should produce accurate predictions of 3D RNA structure.

Materials and Methods

RNA structures used in the present study meet three criteria: (1) they are provided in final form by the Protein Data Bank† or by the Nucleic Acid Data Base‡; (2) they contain both Watson-Crick pairs and bases not in Watson-Crick pairs; and (3) they are determined by X-ray crystallography or by NMR spectroscopy. The Image Library of Biological Macromolecules§ (Sühnel, 1996) proved invaluable for identifying the available structures of RNA. Structures were divided into bound and unbound categories: bound structures contain an interacting non-RNA molecule, often a protein. Secondary structures were obtained from the original publication and visually cross-checked with the 3D structures using the Insight II software package. Based on the secondary structure, unpaired regions were classified into five categories depending on the number of adjacent helixes (in parentheses): hairpins and non-loop (exterior) regions (one), bulges and internal loops (two), and junctions (three or more). Structures without internal loops, junctions, or hairpins of more than four nucleotides were eliminated. Additionally, structures containing pseudoknots and some containing asymmetric internal loops were removed.

Analysis

Unpaired or non-Watson-Crick-paired bases at the 5′ and 3′ ends of helixes were classified as either stacked or unstacked on the terminal base-pair. Three criteria define base stacking on the adjacent pair: (1) distance, the nearest approach of a non-hydrogen atom of the base to a non-hydrogen atom in the terminal pair is $\leqslant 4.0$ Å; (2) angle, the angle between the planes of the base and at least one base of the terminal pair is $\leqslant 30°$; (3) overlap, the base is stacked over, rather than being aligned next to, the adjacent base-pair. Specifically, at least one ring atom of the base overlaps with the outline defined by the ring atoms of either base of the terminal pair when these atoms are projected into the planes of any one of these three bases. If a base fails to meet one or more of these requirements, it is considered unstacked.

Potential hydrogen bonds between opposed 5′ and 3′ dangling bases were identified as a donor nitrogen atom with a covalently bound hydrogen atom, and an appropriate acceptor nitrogen or oxygen atom. Two criteria were required for hydrogen bonding: (1) distance between the hydrogen and the acceptor atom is no greater than 2.5 Å; (2) the angle defined by the three atoms is at least 120°. These criteria are consistent with theoretical solution geometries and observed geometries in crystal structures (Van Eijck *et al*., 1990).

The C++ program BaseStacks was developed to assess whether bases are stacked at the ends of helixes, and to identify hydrogen bonds between the 5′ and 3′ bases at helix termini. For input, BaseStacks takes PDB-format structure files and a list of terminal base-pairs in each structure. BaseStacks verifies that the last base in each helix is in a Watson-Crick, GU, or GA pair and then uses vector algebra to calculate distances and angles and check for overlap for the 5′ and 3′ dangling base at each helix end. BaseStacks identifies hydrogen bonds, if any, between the 5′ and 3′ dangling bases. Finally, BaseStacks produces output in tabular form.

Special considerations

Crystal structures lack coordinates for hydrogen atoms, so to identify hydrogen bonds, BaseStacks infers their positions in imino and amino groups. Imino hydrogen atoms are placed 1.01 Å from the nitrogen atom in the plane of the base and at an equal angle from the two adjacent carbon atoms (C-N-H angles). Amino hydrogen atoms are placed 1.01 Å away from the nitrogen atom in

† URL: http://www.pdb.bnl.gov
‡ URL: httpp://ndb.rutgers.edu
§ URL: http://www.imb-jena.de/IMAGE.html

the plane of the base with each at a 120° C-N-H angle. A few crystal structures lacked coordinates for some nucleotides because it was difficult to locate the atoms in large protein-bound structures; we used only bases with experimentally determined coordinates.

Most of the NMR structures contain a series of optimized geometries rather than a single structure. For these, all the geometries were used to identify stacked bases. For the distance and angle requirements, the average distances and angles in the series of structures were used. For a base to be stacked, the overlap criterion had to be met in at least half the structures. Hydrogen bonds between the dangling bases were identified in each structure and their number was averaged over the entire series. Occasionally, a minimized averaged structure is provided as a separate structure from the series of optimized structures. For these, the averaged structure was weighted 1:1 with the entire series of converged structures.

Some RNAs have multiple known structures for the same sequence. For example, structures of $tRNA^{Phe}$ have been determined for both monoclinic and orthorhombic crystal forms (Sussman *et al.*, 1978; Westhof & Sundaralingam, 1986). Structures of other RNAs have been determined when bound to slightly different molecules. For example, $tRNA^{Gln}$ structures are available for complexes with three mutant synthetase proteins (Arnez & Steitz, 1996), and for wild-type synthetase (Rould *et al.*, 1991; Perona *et al.*, 1993). In these cases, we considered the structures together and used the average distance, angle, overlap and number of hydrogen bonds as with multiple NMR structures. The structures considered together nearly always agreed on base stacking; in Tables 3 and 4, we note the three exceptions. Identical RNA sequences were considered separately only if one is bound and one is unbound, as with the Rev-binding element, RBE3A and RBE3A-Rev22 (Peterson & Feigon, 1996). Thymidine and pseudouridine in tRNA structures were assumed to stack like uridine.

Experimental procedures

The 2-pyrimidinone phosphoramidite and support were synthesized according to published procedures. Briefly, the nucleoside was synthesized from pyrimidin-2-one base and 1,2,3,4 tetra-*O*-acetyl-β-D-ribofuranose similar to the method with 1-halogenated ribofuranoses used by Holý (1977). The 2-pyrimidinone nucleoside was protected with 5′ dimethoxytrityl (DMTr) and 2′ tetrahydropyranyl (thp) (Markiewicz *et al.*, 1984). The protected nucleoside was reacted with bis(diisopropylamino) β-cyanoethoxyphosphine in the presence of diisopropylammonium tetrazolide to yield the phosphoramidite (Kierzek *et al.*, 1986). For oligomers with P at the 3′ terminus, protected nucleoside was attached to Fractosil 500 support (Merck) according to established methods (Kierzek *et al.*, 1986).

RNA oligomers were synthesized on a 1 μmol scale using an Applied Biosystems 392 synthesizer, deblocked, and purified using standard methods (Kierzek *et al.*, 1986; Xia *et al.*, 1997). Phosphoramidites and support for A, C, G, and U were obtained from Glen Research (Sterling, VA); inosine phosphoramidite and support were from Chemgenes Corp. (Waltham, MA). Because the 2-pyrimidinone phosphoramidite was protected by 2′thp, rather than the 2′tert-butyl-dimethylsilyl (tBDMSi) groups of the commercial phosphoramidites, oligomers containing P were acid-deblocked at room temperature, pH 2.0 overnight before TLC purification (Kierzek *et al.*, 1986; SantaLucia *et al.*, 1991).

Oligomer purity was checked by HPLC after TLC purification (Xia *et al.*, 1997). Two oligomers containing P showed about 25% side-product that eluted shortly before the main oligomer product, similar to that described by Gildea & McLaughlin (1989) for DNA oligomers containing deoxy-2-pyrimidinone. Under mildly acidic conditions, like those we used for deblocking, P hydrolyzes in DNA oligonucleotides to leave an abasic site and the abasic product elutes before the non-hydrolyzed oligomer during HPLC (Iocono *et al.*, 1990).

The sequences containing abasic side product, 5′GCAUGP-3′ and 5′PCAUGGU-3′, were further purified by reverse-phase HPLC with slowly increasing acetonitrile against 100 mM aqueous triethylamine acetate (TEAA) buffer through a SuperCosil ABZ+ column (Supelco). The appropriate fractions were collected, lyophilized, and run through a Sep-Pak column (Waters) to eliminate remaining TEAA. The final ICAUGP and PCAUGI also had a significant amount of side products, but these are expected to produce duplexes with very low melting temperatures.

Oligomers used for melt curve experiments, except ICAUGP and PCAUGI, were at least 95% pure, as determined by HPLC, monitoring at 254 nm. Absorbance of duplexes was monitored at 280 nm as a function of temperature in 1 M NaCl, 20 mM sodium cacodylate, and 0.5 mM disodium EDTA at pH 7. For each duplex, melting curves were collected at seven to nine different concentrations, spread geometrically over ~100-fold range. Melting curves were analyzed as described (Xia *et al.*, 1997).

Partial charge determination

Löwdin partial charges and dipole moments were determined for 9-methyl guanine, 9-methyl inosine, 1-methyl-cytosine, and 1-methyl-2-pyrimidinone using the PC-GAMESS version 4.1 software package (Schmidt *et al.*, 1993). Geometries were optimized using natural internal coordinates (Fogarasi *et al.*, 1992) in the absence of solvent, with a 6-31G* basis set.

Acknowledgements

This work was supported by NIH grant GM22939. M.E.B. is a trainee in the Medical Scientist Training Program funded by NIH grant T32 GM07356.

References

Aboul-ela, F., Karn, J. & Varani, G. (1995). The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J. Mol. Biol.* **253**, 313-332.

Addess, K. J., Basilion, J. P., Klausner, R. D., Rouault, T. A. & Pardi, A. (1997). Structure and dynamics of the iron responsive element RNA: implications for binding of the RNA by iron regulatory binding proteins. *J. Mol. Biol.* **274**, 72-83.

Albert, A. & Phillips, J. N. (1956). Ionization constants of heterocyclic substances. II. Hydroxy-derivatives of nitrogenous six-membered ring-compounds. *J. Chem. Soc.* 1294-1304.

Arnez, J. G. & Steitz, T. A. (1996). Crystal structures of three misacylating mutants of *Escherichia coli* glutaminyl-tRNA synthetase complexed with tRNA(Gln) and ATP. *Biochemistry,* **35**, 14725-14733.

Baeyens, K. J., De Bondt, H. L., Pardi, A. & Holbrook, S. R. (1996). A curved RNA helix incorporating an internal loop with G·A and A·A non-Watson-Crick base pairing. *Proc. Natl Acad. Sci. USA,* **93**, 12851-12855.

Basavappa, R. & Sigler, P. B. (1991). The 3 Å crystal structure of yeast initiator tRNA: functional implications in initiator/elongator discrimination. *EMBO J.* **10**, 3105-3111.

Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D. & Williamson, J. R. (1996). Alpha helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. *Science,* **273**, 1547-1551.

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R. & Schneider, B. (1992). The Nucleic Acid Database: a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **63**, 751-759.

Biou, V., Yaremchuk, A., Tukalo, M. & Cusack, S. (1994). The 2.9 Å crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with tRNA(Ser). *Science,* **263**, 1404-1410.

Brodsky, A. S. & Williamson, J. R. (1997). Solution structure of the HIV-2 TAR-argininamide complex. *J. Mol. Biol.* **267**, 624-639.

Burkard, M. E., Turner, D. H. & Tinoco, I., Jr (1999a). The interactions that shape RNA structure. In *The RNA World* (Gesteland, R. F., Cech, T. R. & Atkins, J. F., eds), 2nd edit., pp. 233-264, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Burkard, M. E., Turner, D. H. & Tinoco, I., Jr (1999b). Appendix 1: Structures of base pairs involving at least two hydrogen bonds. In *The RNA World* (Gesteland, R. F., Cech, T. R. & Atkins, J. F., eds), 2nd edit., pp. 675-680, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Butcher, S. E., Dieckmann, T. & Feigon, J. (1997). Solution structure of a GAAA tetraloop receptor RNA. *EMBO J.* **16**, 7490-7499.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R. & Doudna, J. A. (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science,* **273**, 1678-1685.

Cavarelli, J., Rees, B., Thierry, J. C. & Moras, D. (1993). Yeast aspartyl-tRNA synthetase: a structural view of the aminoacylation reaction. *Biochimie,* **75**, 1117-1123.

Cheng, J.-W., Chou, S.-H. & Reid, B. R. (1992). Base-pairing geometry in GA mismatches depends entirely on the neighboring sequence. *J. Mol. Biol.* **228**, 1037-1041.

Correll, C. C., Freeborn, B., Moore, P. B. & Steitz, T. A. (1997). Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell,* **91**, 705-712.

Dallas, A. & Moore, P. B. (1997). The loop E-loop D region of *Escherichia coli* 5S rRNA: the solution structure reveals an unusual loop that may be important for binding ribosomal proteins. *Structure,* **5**, 1639-1653.

Davis, R. C. & Tinoco, I., Jr (1968). Temperature-dependent properties of dinucleoside phosphates. *Biopolymers,* **6**, 223-242.

Fan, P., Suri, A. K., Fiala, R., Live, D. & Patel, D. J. (1996). Molecular recognition in the FMN-RNA aptamer complex. *J. Mol. Biol.* **258**, 480-500.

Fogarasi, G., Zhou, X., Taylor, P. W. & Pulay, P. (1992). The calculation of ab initio molecular geometries: efficient optimization by natural internal coordinates and empirical correction by offset forces. *J. Am. Chem. Soc.* **114**, 8191-8201.

Fourmy, D., Recht, M. I., Blanchard, S. C. & Puglisi, J. D. (1996). Structure of the A site of *Escherichia coli* 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science,* **274**, 1367-1371.

Frechet, D., Ehrlich, R., Remy, P. & Gabarro-Arpa, J. (1979). Thermal perturbation differential spectra of ribonucleic acids. II. Nearest neighbour interactions. *Nucl. Acids Res.* **7**, 1981-2001.

Freier, S. M., Burger, B. J., Alkema, D., Neilson, T. & Turner, D. H. (1983). Effects of 3′ dangling end stacking on the stability of GGCC and CCGG double helices. *Biochemistry,* **22**, 6198-6206.

Freier, S. M., Alkema, D., Sinclair, A., Neilson, T. & Turner, D. H. (1985). Contributions of dangling end stacking and terminal base-pair formation to the stabilities of XGGCCp, XCCGGp, XGGCCYp, and XCCGGYp helixes. *Biochemistry,* **24**, 4533-4539.

Freier, S. M., Kierzek, R., Caruthers, M. H., Neilson, T. & Turner, D. H. (1986a). Free energy contributions of G·U and other terminal mismatches to helix stability. *Biochemistry,* **25**, 3209-3213.

Freier, S. M., Sugimoto, N., Sinclair, A., Alkema, D., Neilson, T., Kierzek, R., Caruthers, M. H. & Turner, D. H. (1986b). Stability of XGCGCp, GCGCYp, and XGCGCYp helixes: an empirical estimate of the energetics of hydrogen bonds in nucleic acids. *Biochemistry,* **25**, 3214-3219.

Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T. & Turner, D. H. (1986c). Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl Acad. Sci. USA,* **83**, 9373-9377.

Gautheret, D., Konings, D. & Gutell, R. R. (1994). A major family of motifs involving G·A mismatches in ribosomal RNA. *J. Mol. Biol.* **242**, 1-8.

Gildea, B. & McLaughlin, L. W. (1989). The synthesis of 2-pyrimidinone nucleosides and their incorporation into oligodeoxynucleotides. *Nucl. Acids Res.* **17**, 2261-2281.

Greenbaum, N. L., Radhakrishnan, I., Patel, D. J. & Hirsh, D. (1996). Solution structure of the donor site of a trans-splicing RNA. *Structure,* **4**, 725-733.

Gorin, A. A., Zhurkin, V. B. & Olson, W. K. (1995). *B*-DNA twisting correlates with base-pair morphology. *J. Mol. Biol.* **247**, 34-48.

Gutell, R. R. (1994). Collection of small subunit (16S- and 16S-like) ribosomal RNA structures: 1994. *Nucl. Acids Res.* **22**, 3502-3507.

Gutell, R. R., Gray, M. W. & Schnare, M. N. (1993). A compilation of large subunit (23S and 23S-like) ribosomal RNA structures: 1993. *Nucl. Acids Res.* **21**, 3055-3074.

He, L., Kierzek, R., SantaLucia, J., Jr, Walter, A. E. & Turner, D. H. (1991). Nearest-neighbor parameters for G·U mismatches: $^{5'GU3'}_{3'UG5'}$ is destabilizing in the contexts $^{CGUG}_{GUGC}$, $^{UGUA}_{AUGU}$, and $^{AGUU}_{UUGA}$, but stabilizing in $^{GGUC}_{CUGG}$. *Biochemistry,* **30**, 11124-11132.

Herschlag, D. (1991). Implications of ribozyme kinetics for targeting the cleavage of specific RNA molecules

in vivo: more isn't always better. *Proc. Natl Acad. Sci. USA,* **88**, 6921-6925.

Heus, H. A. & Pardi, A. (1991). Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science,* **253**, 191-194.

Holý, A. (1977). Preparation of acyl derivatives of pyrimidin-2-one nucleosides by the silyl variant of the Hilbert-Johnson reaction. *Coll. Czech. Chem. Commun.* **42**, 902-908.

Iocono, J. A., Gildea, B. & McLaughlin, L. W. (1990). Mild acid hydrolysis of 2-pyrimidinone-containing DNA fragments generates apurinic/apyrimidinic sites. *Tetrahedron Letters,* **31**, 175-178.

Kierzek, R., Caruthers, M. H., Longfellow, C. E., Swinton, D., Turner, D. H. & Freier, S. M. (1986). Polymer-supported RNA synthesis and its application to test the nearest-neighbor model for duplex stability. *Biochemistry,* **25**, 7840-7846.

Kolk, M. H., Heus, H. A. & Hilbers, C. W. (1997). The structure of the isolated, central hairpin of the HDV antigenomic ribozyme: novel structural features and similarity of the loop in the ribozyme and free in solution. *EMBO J.* **16**, 3685-3692.

Li, Y., Zon, G. & Wilson, W. D. (1991). NMR and molecular modeling evidence for a GA mismatch base pair in a purine-rich DNA duplex. *Proc. Natl Acad. Sci. USA,* **88**, 26-30.

Lietzke, S. E., Barnes, C. L., Berglund, J. A. & Kundrot, C. E. (1996). The structure of an RNA dodecamer shows how tandem U-U base pairs increase the range of stable RNA structures and the diversity of recognition sites. *Structure,* **4**, 917-930.

Markiewicz, W. T., Biala, E. & Kierzek, R. (1984). Application of the tetraisopropyldisiloxane-1,3-diyl group in the chemical synthesis of oligoribonucleotides. *Bull. Acad. Polon. Sci. Chem.* **32**, 433-451.

Mathews, D. H., Banerjee, A. R., Luan, D. D., Eickbush, T. H. & Turner, D. H. (1997). Secondary structure model of the RNA recognized by the reverse transcriptase from the R2 retrotransposable element. *RNA,* **3**, 1-16.

Mathews, D. H., Sabina, J., Zuker, M. & Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288**, 911-940.

Major, F., Turcotte, M., Gautheret, D., Lapalme, G., Fillion, E. & Cedergren, R. (1991). The combination of symbolic and numerical computation for three-dimensional modeling of RNA. *Science,* **253**, 1255-1260.

McDowell, J. A. & Turner, D. H. (1996). Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: solution structure of $(rGAG\underline{GU}CUC)_2$ by two-dimensional NMR and simulated annealing. *Biochemistry,* **35**, 14077-14089.

McDowell, J. A., He, L., Chen, X. & Turner, D. H. (1997). Investigation of the structural basis for thermodynamic stabilities of tandem GU wobble pairs: NMR structures of $(rGGA\underline{GU}UCC)_2$ and $(rGGA\underline{UG}UCC)_2$. *Biochemistry,* **36**, 8030-8038.

Nissen, P., Kjeldgaard, M., Thirup, S., Polekhina, G., Reshetnikova, L., Clark, B. F. & Nyborg, J. (1995). Crystal structure of the ternary complex of PhetRNAPhe, EF-Tu, and a GTP analog. *Science,* **270**, 1464-1472.

Ortoleva-Donnelly, L., Szewczak, A. A., Gutell, R. R. & Strobel, S. A. (1998). The chemical basis of adenosine conservation throughout the *Tetrahymena* ribozyme. *RNA,* **4**, 498-519.

Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. & Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature,* **372**, 432-438.

Pan, B., Mitra, S. N. & Sundaralingam, M. (1999). Crystal structure of an RNA 16-mer duplex $r(GCAGAGUUAAAUCUGC)_2$ with nonadjacent $G(syn)\cdot A^+(anti)$ mispairs. *Biochemistry,* **38**, 2826-2831.

Perona, J. J., Rould, M. A. & Steitz, T. A. (1993). Structural basis for transfer RNA aminoacylation by *Escherichia coli* glutaminyl-tRNA synthetase. *Biochemistry,* **32**, 8758-8771.

Petersheim, M. & Turner, D. H. (1983). Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp. *Biochemistry,* **18**, 256-263.

Peterson, R. D. & Feigon, J. (1996). Structural change in Rev responsive element RNA of HIV-1 on binding Rev peptide. *J. Mol. Biol.* **264**, 863-877.

Rould, M. A., Perona, J. J. & Steitz, T. A. (1991). Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. *Nature,* **352**, 213-218.

Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C. & Moras, D. (1991). Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). *Science,* **252**, 1682-1689.

Saenger, W. (1984). *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.

SantaLucia, J., Jr, Kierzek, R. & Turner, D. H. (1991). Functional group substitutions as probes of hydrogen bonding between GA mismatches in RNA internal loops. *J. Am. Chem. Soc.* **113**, 4313-4322.

SantaLucia, J., Jr, Kierzek, R. & Turner, D. H. (1992). Context dependence of hydrogen bond free energy revealed by substitutions in an RNA hairpin. *Science,* **256**, 217-219.

Schmidt, M. W., Baldridge, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S., Jensen, J. H., Koseki, S., Matsunaga, N., Nguyen, K. A., Su, S., Windus, T. L., Dupuis, M. & Montgomery, J. A., Jr (1993). General atomic and molecular electronic structure system. *J. Comput. Chem.* **14**, 1347-1363.

Scott, W. G., Finch, J. T. & Klug, A. (1995). The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell,* **81**, 991-1002.

Scott, W. G., Murray, J. B., Arnold, J. R. P., Stoddard, B. L. & Klug, A. (1996). Capturing the structure of a catalytic RNA intermediate: the hammerhead ribozyme. *Science,* **274**, 2065-2069.

Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A. & Steinberg, S. (1998). Compilation of tRNA sequences and sequences of tRNA genes. *Nucl. Acids Res.* **26**, 148-153.

Strobel, S. A. & Shetty, K. (1997). Defining the chemical groups essential for *Tetrahymena* group I intron function by nucleotide analog interference mapping. *Proc. Natl Acad. Sci. USA,* **94**, 2903-2908.

Sugimoto, N., Kierzek, R. & Turner, D. H. (1987). Sequence dependence for the energetics of dangling ends and terminal base pairs in ribonucleic acid. *Biochemistry,* **26**, 4554-4558.

Sühnel, J. (1996). Image library of biological macromolecules. *Comput. Appl. Biosci.* **12**, 227-229.

Sussman, J. L., Holbrook, S. R., Warrant, R. W., Church, G. M. & Kim, S. H. (1978). Crystal structure of yeast phenylalanine transfer RNA. I. Crystallographic refinement. *J. Mol. Biol.* **123**, 607-630.

Szewczak, A. A. & Moore, P. B. (1995). The sarcin/ricin loop, a modular RNA. *J. Mol. Biol.* **247**, 81-98.

Turner, D. H. & Bevilacqua, (1993). Thermodynamic considerations for evolution by RNA. In *The RNA World* (Gesteland, R. F. & Atkins, J. F., eds), pp. 447-464, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Turner, D. H., Sugimoto, N., Kierzek, R. & Dreiker, S. D. (1987a). Free energy increments for hydrogen bonds in nucleic acid base pairs. *J. Am. Chem. Soc.* **109**, 3783-3785.

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

Turner, D. H., Sugimoto, N. & Freier, S. M. (1988). RNA structure prediction. *Annu. Rev. Biophys. Biophys. Chem.* **17**, 167-192.

Van Eijck, B. P., Kroon-Batenburg, L. M. J. & Kroon, J. (1990). Hydrogen-bond geometry around sugar molecules: comparison of crystal statistics with simulated aqueous solutions. *J. Mol. Struct.* **237**, 315-325.

Westhof, E. & Sundaralingam, M. (1986). Restrained refinement of the monoclinic form of yeast phenylalanine transfer RNA. Temperature factors and dynamics, coordinated waters, and base-pair propeller twist angles. *Biochemistry,* **25**, 4868-4878.

Westhof, E., Dumas, P. & Moras, D. (1988). Restrained refinement of two crystalline forms of yeast aspartic acid and phenylalanine transfer RNA crystals. *Acta Crystallog. sect. A,* **44**, 112-123.

Wu, M., SantaLucia, J., Jr & Turner, D. H. (1997). Solution structure of (rGGC<u>AG</u>GCC)$_2$ by two-dimensional NMR and the iterative relaxation matrix approach. *Biochemistry,* **36**, 4449-4460.

Xia, T., McDowell, J. A. & Turner, D. H. (1997). Thermodynamics of nonsymmetric tandem mismatches adjacent to G·C base pairs in RNA. *Biochemistry,* **36**, 12486-12497.

Yang, Y., Kochoyan, M., Burgstaller, P., Westhof, E. & Famulok, M. (1996). Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. *Science,* **272**, 1343-1347.

Ye, X. M., Kumar, R. A. & Patel, D. J. (1995). Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex. *Chem. Biol.* **2**, 827-840.

Ye, X. M., Gorin, A., Ellington, A. D. & Patel, D. J. (1996). Deep penetration of an alpha-helix into a widened RNA major groove in the HIV-1 rev peptide-RNA aptamer complex. *Nature Struct. Biol.* **3**, 1026-1033.

Young, M. A., Ravishanker, G., Beveridge, D. L. & Berman, H. M. (1995). Analysis of local helix bending in crystal structures of DNA oligonucleotides and DNA-protein complexes. *Biophys. J.* **68**, 2454-2468.

Zimmermann, G. R., Jenison, R. D., Wick, C. L., Simorre, J. P. & Pardi, A. (1997). Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA. *Nature Struct. Biol.* **4**, 644-649.

Edited by I. Tinoco

(Received 15 March 1999; received in revised form 25 May 1999; accepted 25 May 1999)

http://www.academicpress.com/jmb

Supplementary Material is available from JMB Online.