11-1998

Conserved Geometrical Base-Pairing Patterns in RNA

Neocles B. Leontis
Bowling Green State University, leontis@bgsu.edu

Eric Westhof

Quarterly Reviews of Biophysics **31**, 4 (1998), pp. 399–455 Printed in the United Kingdom

Conserved geometrical base-pairing patterns in RNA

NEOCLES B. LEONTIS[1] AND ERIC WESTHOF[2]

[1] *Chemistry Department, Bowling Green State University, Bowling Green OH 43403. Tel. : 419–372–8663 ; Fax : 419–372–9809 ; E-mail : Leontis@bgnet.bgsu.edu*
[2] *Institut de Biologie Moléculaire et Cellulaire du CNRS, Modélisation et Simulations des Acides Nucléiques, UPR 9002, 15 rue René Descartes, F-67084 Strasbourg Cedex, France*
E-mail : Westhof@ibmc.u-strasbg.fr

1. INTRODUCTION

RNA molecules fold into a bewildering variety of complex 3D structures. Almost every new RNA structure obtained at high resolution reveals new,

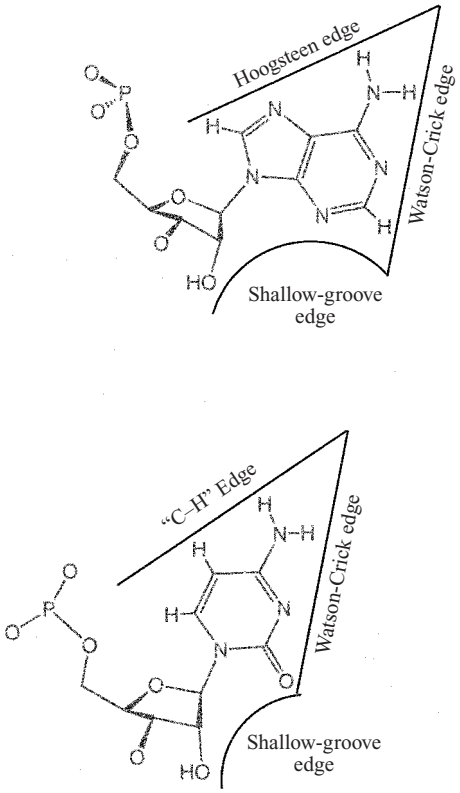


Fig. 1. The three edges on purine and pyrimidine bases that can participate in edge-to-edge interactions mediated by hydrogen-bonding. The Hoogsteen (purine) and 'C–H' (pyrimidine) edges are geometrically equivalent.

unanticipated structural motifs, which we are rarely able to predict at the current stage of our theoretical understanding. Even at the most basic level of specific RNA interactions – base-to-base pairing – new interactions continue to be uncovered as new structures appear. Compilations of possible non-canonical base-pairing geometries have been presented in previous reviews and monographs (Saenger, 1984; Tinoco, 1993). In these compilations, the guiding principle applied was the optimization of hydrogen-bonding. All possible pairs with two standard H-bonds were presented and these were organized according to symmetry or base type. However, many of the features of RNA base-pairing interactions that have been revealed by high-resolution crystallographic analysis could not have been anticipated and, therefore were not incorporated into these compilations. These will be described and classified in the present review. A recently presented approach for inferring basepair geometry from patterns of sequence variation (Gautheret & Gutell, 1997) relied on the 1984 compilation of basepairs (Saenger, 1984), and was extended to include all possible single H-bond combinations not subject to steric clashes. Another recent review may be consulted for a discussion of the NMR spectroscopy and thermodynamic effects

of non-canonical ('mismatched') RNA basepairs on duplex stability (Limmer, 1997).

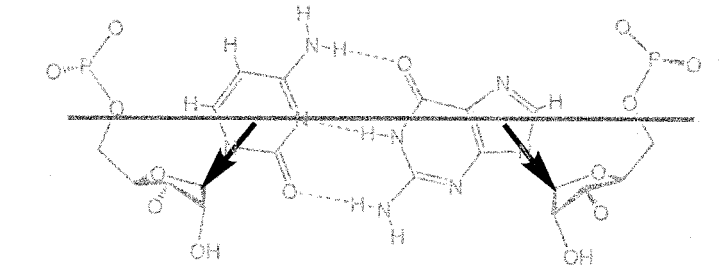
In the present review, the aim is to organize the available RNA crystallographic data into a coherent library of isosteric pairings that can substitute for each other in homologous RNA molecules. The underlying assumption is that the 3D structures of homologous RNA molecules are more strongly conserved than their individual sequences. Crystallographic data, therefore, become even more useful when correlated with comparative sequence analysis for the simple reason that isosteric pairs will substitute for each other in conserved regions or motifs. Conversely, sequence covariation data contain information that can potentially identify bases involved in tertiary interactions and can even indicate the most likely pairing geometry. Therefore, our goal in organizing the data in this way is to facilitate prediction of RNA tertiary structure from sequence. We anticipate, furthermore, that a properly conceived structural library will provide a framework for organizing the new data that are appearing at an accelerating pace from crystallographic as well as NMR studies of RNA molecules. We choose here to concentrate on X-ray crystallographic data because of their generally greater precision. Structures referred to in the text were obtained from the Nucleic Acid Database (Berman *et al*. 1992) unless otherwise noted (http://ndbserver.rutgers.edu/NDB/ndb.html).

2. DEFINITIONS

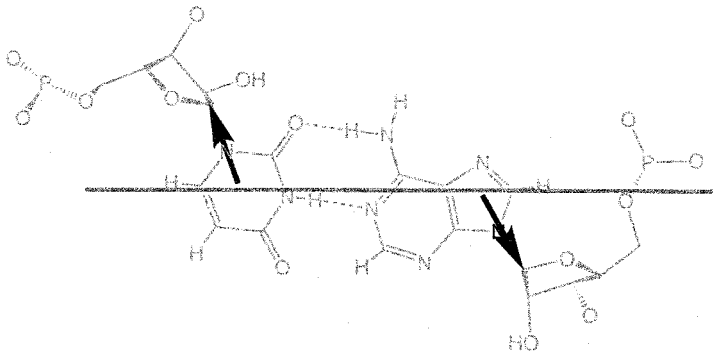
Nucleic acid bases interact either by stacking on each other or by abutting edge-to-edge. The edge-to-edge interactions are mediated by electrostatically driven hydrogen-bonding between complementary arrays of negatively and positively polarized atoms and, therefore, confer greater specificity than the stacking interactions, which, on the other hand, contribute more to the overall free energy stabilizing nucleic acid structures. Both purines and pyrimidines present three edges for interaction, as shown in Fig. 1. These are the Watson–Crick edge (used in canonical pairing), the Shallow-groove edge, and the Hoogsteen (for purines) or 'C–H' edge (for pyrimidines).

Nucleic acid bases are aromatic heterocycles with a large proportion of heteroatoms. This feature generates multiple modes of hydrogen-bonding interactions. The pairing patterns generated may be classified according to a limited set of conformational parameters. The relevant geometric parameters are listed in Table 1. The parameters are listed in the first column of Table 1 and their possible values appear in the second column. The canonical value for each parameter is the first one given in each entry of the second column. Canonical Watson–Crick (W.–C.)[1] pairs have both bases interacting at their W.–C. edges, locally anti-parallel strands, *anti* base-sugar conformations and glycosidic bonds oriented *cis* with respect to each other. The glycosidic bonds of two interacting bases are defined to be *cis* or *trans* with respect to an axis running parallel to and between the hydrogen bonds of the basepair, as shown in Fig. 2 for bases

[1] Abbreviations: Sh.G., shallow-groove; W.–C., Watson–Crick.



Cis orientation of the Glycosidic Bonds



Trans orientation of the Glycosidic Bonds

Fig. 2. Each edge-to-edge interaction can occur in *cis* or *trans* orientation of the bases. This is defined relative to a line running parallel to and in-between the hydrogen-bonds of the interaction, as shown schematically for two bases interacting using their Watson–Crick edges.

interacting with their W.–C. edges. Three of the four parameters in Table 1 determine the fourth. Thus, changing any one of the parameters in Table 1 changes at least one other parameter (Westhof, 1992). For example, when both bases remain in the preferred *anti* conformation (or the less favoured *syn* conformation), a *trans* W.–C./W.–C. basepair (wherein both bases interact with their W.–C. edges) necessarily has, at least locally, parallel-oriented strands. As a further example, when the W.–C. edge of a base interacts with the Hoogsteen edge of a second base, the pairing must be trans when the strands are locally anti-parallel and *cis* when the strands are locally parallel. Again, it is assumed that both bases remain in the same (by default *anti*) base-sugar conformation. If one base flips into the opposite range (so that the two paired bases occupy different torsions for the glycosyl bond), the relative orientations of the strands change.

It should be noted, however, that this rule must be modified for pairings involving the Shallow-groove (Sh.G.) edge of one or both of the bases. Thus, it is found that a W.–C./Sh.G. pairing has strands that are locally anti-parallel when the glycosidic bonds are in *cis*, and strands that are locally parallel when the

Table 1. *Geometric parameters for classifying nucleic acid base pairs. The first value given for each entry in the second column is the default value found for canonical W.–C. basepairs. R indicates A or G, and Y indicates C or U; Sh.G., is shallow-groove*

Parameter	Value
Base-sugar conformation	1. *Anti* 2. *Syn*
Glycosidic bond orientation	1. *Cis* 2. *Trans*
Interacting edges	1. W.–C. edge (RN1, GN2, AC2, GO6, AN6; YO2,YN3, CN4, UO4) 2. Hoogsteen edge (RN7, GO6, AN6; CN4, UO4, YC5, YC6) 3. Sh.G. edge (RN3, GN2, AC2, YO2, ribose O2′)
Local strand orientation	1. Anti-parallel 2. Parallel

Table 2. *Glycosidic bond and local strand orientations for Sh.G. pairings, assuming both interacting bases are either in anti (or syn) base-sugar conformations. If one of the bases adopts the syn (or, respectively, anti) conformation, the local relative orientations of the strands change*

Interacting edges	Glycosidic bond orientation	Local strand orientation
W.–C./Sh.G.	*Cis*	Anti-parallel
W.–C./Sh.G.	*Trans*	Parallel
Sh.G./Sh.G.	*Cis*	Anti-parallel
Sh.G./Sh.G.	*Trans*	Parallel
Hoogsteen/Sh.G.	*Cis*	Parallel
Hoogsteen/Sh.G.	*Trans*	Anti-parallel

glycosidic bonds are in *trans*. The relationships between the geometric parameters for the Sh.-G. pairings are summarized in Table 2. It is assumed in each case that both of the interacting bases in Table 2 remain in the *anti* conformation.

Besides the purely geometrical parameters listed in Table 1, additional parameters relating to hydrogen-bonding modes are forced upon us by new observations. These are collected in Table 3. The first is the observation of bifurcated hydrogen-bonds in multiple contexts. These are hydrogen bonds involving a single acceptor atom on one of the pairing bases (i.e. a carbonyl oxygen or imino nitrogen) and two donor atoms on the pairing partner. The donor atoms may be, for example, adjacent amino and imino protons or two amino protons

Table 3. *Schematic representation of hydrogen-bonding modes observed in RNA base–base interactions*

. Standard Ex.>N–H····N<
. Bifurcated Hydrogen Bonds Ex.>N–H O=C< >N–H
. Water-inserted or 'Open' Pairings Ex.>N–H····Ow····H–N<
. Hydrogen Bonds involving C–H bonds Ex.C–H····O or C–H····N —
. Hydrogen bonds involving Ribose oxygens Ex.>N–H····O2′–C2′ H

belonging to the same amino nitrogen. Several distinct base pairing geometries exhibiting bifurcated H-bonding have been observed and will be discussed below (see also Tables 4*a* and 4*b*).

A second unanticipated feature is the observation that water molecules participate directly in certain base-pairing geometries. We refer to these as 'water-inserted' or 'open' pairs because the insertion of a water molecule opens up the pairing toward one of the grooves of the helix. It should be noted that water molecules are also observed to bridge hydrogen-bonding sites on the interacting bases in bifurcated pairings, and perhaps in other geometries. A third important feature of base-pairing is the occurrence of C–H···O or C–H···N hydrogen bonds involving polarized hydrogen atoms covalently attached to aromatic carbon atoms (i.e. Purine H8, Adenine H2, and Pyrimidine H5 or H6). Auffinger *et al*. (1996) recently presented an overview of such hydrogen bonds. A fourth recurring feature is H-bonding between base and sugar protons, and even between sugar protons of the interacting strands. This is an integral component of pairings involving the Sh.G. edges of the interacting bases.

A fifth novel feature is the observation of the side-by-side pairing geometry involving bases on adjacent nucleotides in the same chain. This occurs in so-called adenosine platforms, which were predicted on theoretical grounds for DNA (Kuryavyi & Jovin, 1995), but were first observed in the crystal structure of P4–P6 of Group I (Cate *et al*. 1996*b*), file URX053 in the Nucleic Acids Database (NDB). Interestingly, the base-triple formed via interaction of a 'bulged G' with

Table 4*a*. *Summary of* cis *base-pairing geometries involving only W.–C. and Hoogsteen edges*

Basepair type	H-bonding	Isosteric upon reversal	Other isosteric pairs	Examples	Files	C1′–C1′ distance
Cis geometry						
Cis W.-C. C=G *Cis* W.-C. U–A	CN_4–$GO6$; CN_3–$GN1$; CO_2–$GN2$ UO_4–$AN6$; UN_3–$AN1$	Yes	Any *cis* W.–C. basepair	–	–	10.5 Å
Cis W.–C. A•G	$AN6$–$GO6$; $AN1$–$GN1$	Yes	*Cis* Water-inserted C•U	G26•A44(tRNA) A105•215	TRNA07 URX053	12.7 Å 12.5 Å
Cis wobble G○U *Cis* wobble A(+)○C	$GO6$–UN_3; $GN1$–UO_2 $AN6$–CN_3; $AN1(+)$–CO_2	No		G9○U20 A5○C12	AR0008 AR0001	10.3 Å 10.4 Å
Cis wobble U○U *Cis* wobble C(+)○C	UO_4–UN_3; UN_3–UO_2 CN_4–CN_3; $CN_3(+)$–CO_2	No	*Cis* wobble U•C(+)	U17○U5 C17.0○C3.0	PTR016 URX035	8.8 Å 8.5 Å
Cis W.–C./Hoogsteen U•A(syn)	UO_4–$AN6$; UN_3–$AN7$	No		U135•A187	URX053	8.2 Å
Cis W.–C./Hoogsteen C•G	CN_4–$GO6$; $CN_3(+)$–$GN7$			C260•G108	URX053	8.1 Å
Cis-W.–C./Hoogsteen U•U	UN_3–UO_4; **UO2–UC5**		C•C	U259•U107	URX053	8.6 Å
Cis W.–C./Hoogsteen A(+)•G(syn)	$AN6$–$GO6$; $AN_3(+)$–$GN7$	No	G•G	A27•G6	AR0006	10.5 Å
Cis W.–C./Hoogsteen G•A	$GO6$–$AN6$; GN_3–$AN7$			G257•A104	URX053	10.0 Å
Cis bifurcated G•U *Cis* bifurcated G•G *Cis* bifurcated A•C	$GN1/N_2$–UO_4 $GN1/N_2$–$GO6$ $AN1$–$CN_4(H1/H2)$	No	A•C A•A G•U	G102•U74 G100•G76 A38•C32	URL064 URL064 TRNA12	13.1 Å 13.6 Å 12.5 Å
Cis water-inserted ('open') G•A	$AN6$–$GO6$; $AN1$–O(water)–$GN1$	Yes		A101•G75	URL064	14.8 Å
Cis water-inserted ('open') C•U	CN_4–UO_4; CN_3–O(water)–UN_3	Yes	*Cis* W.–C./W.–C. A•G	C8•U7	Ar0005	11.8 Å
Cis water-inserted ('open') A•U	$AN6$–UO_4	Yes	G•C	A15.1•U16.1	URX035	13.2 Å

Table 4*b*. *Summary of* trans *basepairing geometries involving only W.–C. and Hoogsteen edges*

Basepair type	H-bonding	Isosteric upon reversal	Other isosteric pairs	Examples	Files	C1′–C1′ distance
Trans geometry						
Trans W.–C. G•C	CO2–GN1; CN3–GN2	Yes	–	G15•C48	TRNA09	10.5 Å
Trans W.–C. A•U	UO2–AN6; UN3–AN1; **UO4–AC2**	Yes	–	A15•U48	TRNA07	11.0 Å
Trans wobble G•U *Trans* wobble A•C	GN1–UO4; GO6–UN3 AN1–CN4; AN6–CN3	Yes		U120•G125	PR0005	11.9 Å
Trans W.–C. C•C	CO2–CN4; CN4–CO2	Yes		C16•C59	PR0004	9.7 Å
Trans wobble U•D	UO2–DN3; UN3–DO2	Yes	U•U	U59•D16	PTR012	8.9 Å
Trans W.–C. A•A	AN6–AN1; AN1–AN6	Yes	G•G	A151•A248	URX053	13.5 Å
Trans W.–C./Hoogsteen U•A	UN3–AN7; UO2–AN6	No	C•A	U103•A73	URL064	9.8 Å
Trans W.–C./Hoogsteen C•G *Trans* W.–C./Hoogsteen U•U	CN4–GN7; CN3(+)–GO6 **UO4–UC5**; UN3–UO4	No	C•A	C8•G12 U1A•U2C	UR0004 URF042	11.1 Å 11.2 Å
Trans W.–C./Hoogsteen A•A *Trans* W.–C./Hoogsteen A•G	AN1–AN6; AN6–AN7 AN1(+)/AC2–GO6; AN6–GN7	No		A7•A6 A46•G22	URL051 TRNA07	12.5 Å 11.8 Å
Trans W.–C./Hoogsteen G•G	GN1–GN7; GN2–GO6	No		7mG46•G22	TRNA09	10.9 Å
Trans Hoogst./Hoogst. A•A	AN6–AN7; AN7–AN6	Yes	A•G	A9•A23	TRNA09	10.9 Å
Trans Hoogst./Hoogst. C•G	CN4–GN7; **CC5–GO6**	Yes	A•C	1mG9•C23	TRNA12	11.5 Å
Trans bifurcated G•G *Trans* bifurcated A•A	GO6–GN1/N2 AN6(H1/H2)–AN1	No	G•U A•C	G15•G48 A14•A46	PR0004 PR0004	13.5 Å 12.8 Å
Trans bifurcated G•Ψ	GN1/N2– Ψ O4	No	G•U	G18•Ψ55	TRNA09	8.8 Å

Table 4*c*. *Summary of basepairing geometries involving Sh.G. edges*

Basepair type	H-Bonding	Isosteric upon reversal	Other isosteric pairs	Examples	Files	C1′–C1′ distance:
Shallow-Groove (Sh.G.) pairings						
Trans orientation	AN6–GN3; AN7–GN2;	No		A78•G98	URL064	9.6 Å
Hoogsteen/Sh.G.('Sheared'	AN6–GO2′			A113•A207	URX053	9.2 Å
locally anti-parallel)	AN6–AN3; **AN7–AC2**;		A•U	A38•Om–C32	TRNA09	11.1 Å
A•G, A•A, A•C, C•U	AN6–AO2′ AN6–CO2 CN4–UO2		C•C	C23•U7	UR0002	10.6 Å
Cis orientation Hoogsteen/Sh.G.	AN7–GO2′; AN6–GN3	No	A•A; A•Y	A142•G140	PR0005	7.9 Å
(locally parallel) A•G; A•C	AN7–CO2′; AN6–CO2			A261•C413	UR0003	8.6 Å
Trans orientation W.–C./Sh.G.	CN3–GN2; CN4–GN3	No	C•A	C124•G201	URX053	9.5 Å
(locally parallel) C•G; A•G;	AN1–GN2; AN6–GN3/GO2′	No	A•A	A24•G7	UR0004	9.2 Å
D•G	DO2–GN2; DN3–GN3; DO4–GO2′	No	U•R	D20•G15	PTR004	7.6 Å
U•U	UN3–UO2; UO4–UO2′	No	U•C	U106•U258	UR0003	8.2 Å
Cis orientation W.–C./Sh.G.	AN6–CO2; AN1–CO2′	No		A23•C15	UR0004	9.5 Å
(locally anti-parallel) A•Y; A•A	AN6–AN3; AN1–AO2′			A14•A21	TRNA07	9.3 Å
	AN6–UO2; AN1–UO2′			A14.0•U7.0	UHX026	9.4 Å
A•G	AN1–GN2	No		A27•G22	PRV022	8.3 Å
Trans orientation Sh.G./Sh.G.	AN3–GN2; **AC2–GN3**;					
(locally parallel) A•G	AN1–GO2′	Yes	G•G; A•A	A184•G212	URX053	8.1 Å
A•U	**AC2–UO2**; AN1–UO2′		A•C	A21•U8	PTR009	8.3 Å
Cis orientation Sh.G./Sh.G.	**AC2–UO2**; AN3–UO2′;	No		A152•U224	URX053	5.4 Å
(locally anti-parallel) A•U;	AO2′–UO2′;			A23•C10.4	UHX026	5.3 Å
A•C	**AC2–CO2**;					
	AN3–CO2′; AO2′–CO2′		A•A	A183•G110	URX053	5.8 Å
A•G	**AC2–GN3**; AN3–GO2′;					
	AO2′–GO2′		G•U; G•A	G200•C197	URX053	5.8 Å
G•C	GN2–CO2; GN3–CO2′; GO2′–CO2′					
Side-by-side pairing G•U	GN2–UO4	No		G10•U11	UR0002	6.9 Å
Side-by-side Pairing A•A	AN3–AN6		A•C	A225•A226	URX053	6.5 Å

the 3′-adjacent U in the sarcin/ricin loop of 23S rRNA (NDB file UR0002) is found to be isosteric with adenosine platforms, as will be described below. The U also forms a *trans* W.–C./Hoogsteen pair with an adenosine on the other strand. The side-by-side geometry gives rise to pairings that fall under the geometric category of the *cis* Hoogsteen/Sh.G. type (Table 2).

We will maintain throughout our review the following symbolic designations: G=C and U–A for canonical W.–C. pairs; GoU for wobble pairs; X•Y (e.g. A•A or U•A) for non-canonical pairs; and X⋆Y for non-identified pairs.

In Tables 4*a–c*, edge-to-edge interactions of RNA bases observed crystallographically (or inferred from sequence comparisons) are organized according to geometrical parameters and hydrogen-bonding patterns. The tables correspond to the following sub-divisions:

Table 4*a*
Cis pairings involving W.–C. and/or Hoogsteen edges;
Table 4*b*
Trans pairings also involving W.–C. and/or Hoogsteen edges;
Table 4*c*
Sh.G. pairings (including side-by-side pairings).

Our choice for the subdivisions is dictated by the search for logical and simplifying rules linking base covariations (and substitutions) and base pairing patterns in folded RNAs. Base–base interactions can occur between all combinations of the three edges shown in Fig. 1. For the Watson–Crick edge and the Hoogsteen edge in purines (or the equivalent 'C–H' edge in pyrimidines), it is geometrically most important to investigate whether the pairing ought to be *cis* or *trans*. In the course of RNA modelling, the choice of interaction within the Shallow-groove edge is often dictated by more complex stereochemical and folding considerations for which the *cis*/*trans* classification is less pertinent. Each pairing geometry is designated according to the interacting edges of the bases, with the exception of 'wobble' and 'bifurcated' pairs. Wobble pairs involve the W.–C. edges of both bases. Bifurcated pairs involve the W.–C. edge of one base and a single, exocyclic hetero atom of the partner base (e.g. AN6, GO6, UO4, CN4 and YO2). *Anti* configurations should be assumed for the bases as the default base-sugar configuration, unless otherwise noted because *syn* bases are rarely observed and then only for purines. The interacting atoms for each pairing type are given in the second column. For the *Cis* and *Trans* sections of Table 4, one can apply rules to determine the relative strand orientations, as described above (Westhof, 1992). For example, assuming *anti* base-sugar configurations, the *trans* W.–C./W.–C. pairings are found to have strands that are locally parallel. Of course, this does not exclude the possibility that these pairings also occur with one purine in the *syn*-configuration, which would in turn imply locally anti-parallel strands. A second example is provided by the *cis* W.–C./Hoogsteen pairings. The U•A pairing of this type, which is cited in Table 4, occurs with A in the *syn* configuration. This implies anti-parallel strands, whereas the cited C•G pairing of this type has the *anti* configuration (indicated by default), implying parallel strands. Again, it

Table 5. *Literature references and brief descriptions of RNA crystal structures cited in the text and in the tables. Coordinate files are from the Nucleic Acid Database* (*Berman* et al. *1992*)

NDB file	Brief description of structure	References
AR0001	Tandem wobble C•A(+)pairs	(Jang *et al.* 1998)
AR0005	Tandem water-inserted U•C pairs	(Tanaka *et al.* 1999)
AR0006	Non-adjacent A(+)•G(syn) *cis* W.–C./Hoogsteen pairs	(Pan *et al.* 1999)
AR0008	Adjacent wobble G•U pairs	(Trikha *et al.* In preparation)
PR0004	Cysteinyl tRNA with Synthetase	(Nissen *et al.* 1999)
PR0005	Hepatitis Delta Virus Ribozyme	(Ferre-D'Amare *et al.* 1998)
PTE003	*E. coli* $tRNA^{GLN}$ with Synthetase	(Rath *et al.* 1998)
PTR004	*T. thermophilus* $tRNA^{SER}$ with Synthetase	(Biou *et al.* 1994)
PTR012	Phe-tRNA/EF-Tu/GDPNP ternary complex	(Nissen *et al.* 1995)
PTR016	U2 RNA fragment/Spliceosomal U2″–U2A′ protein	(Price *et al.* 1998)
PRV022	HIV I RNA Pseudoknot/Reverse Transcriptase	(Jaeger *et al.* 1998)
TRNA07	Yeast $tRNA^{ASP}$	(Westhof *et al.* 1988)
TRNA09	Yeast $tRNA^{PHE}$	(Westhof *et al.* 1988)
TRNA12	Yeast initiator tRNA	(Basavappa & Sigler, 1991)
UHX026	Hammerhead Ribozyme	(Pley *et al.* 1994)
URF042	Overhanging 5′-U•U with inter-molecular *trans* W.–C./Hoogsteen pairing	(Wahl *et al.* 1996)
URL050	Tandem U•U wobble pairs	(Lietzke *et al.* 1996)
URL051	RNA Helix with tandem Sheared G•A and *trans* W.–C./Hoogsteen A•A	(Baeyens *et al.* 1996)
URL064	Bacterial 5S rRNA Loop E	(Correll *et al.* 1997)
UR0002	Sarcin/Ricin Loop of rat 28S rRNA	(Correll *et al.* 1998)
UR0003	Tetrahymena Group I Intron (5Å)	(Golden *et al.* 1998)
UR0004	RNA Pseudoknot (Ribosomal frameshifting)	(Lu *et al.* In press)
URX035	All-RNA Hammerhead Ribozyme	(Scott *et al.* 1995)
URX053	P4-P6 of Tetrahymena Group I Intron	(Cate *et al.* 1996*a*)
URX059	Mg^{2+}-soaked RNA Hammerhead Ribozyme	(Scott *et al.* 1996)
URX063	Helix A from Thermus Flavus 5S rRNA	(Betzel *et al.* 1994)

should be understood that each base in a given base pair could occur, at least in principle, in either the *syn* or *anti* configuration. For the Sh.G. section (Table 4*c*), one should refer to Table 2 for strand orientations.

Isosteric basepairs are grouped together in Tables 4*a*, *b*, and *c*, with no separating horizontal lines. Nearly isosteric pairs belonging to the same geometric category are separated from each other by a horizontal dotted line. Hydrogen-bonded atoms are given in the second column. Whether a pair is self-isosteric is indicated in the third column. Potentially isosteric pairs identified by covariation analysis and modeling are listed in the fourth column. Examples of each interaction are provided with reference to the respective NDB file, in which they are found in the fifth and sixth columns, and the observed C1′–C1′ distances are listed in the last column as a simple indicator of isostericity. NDB files referred to in this review are listed in Table 5 with a brief description of each structure and reference citations.

3. *CIS* BASEPAIRS

3.1 Cis *Watson–Crick/Watson–Crick*

3.1.1 *Canonical pairs*

The remarkable feature of the canonical W.–C. pairing geometry (*cis*, *anti/anti*, and anti-parallel strands) is the mutual isostericity of all four pairs A–U, U–A, G=C, and C=G. The biological consequence of this isostericity is the interchangeability of these pairings without any distortion of the canonical double helices (A-type in RNA) that they comprise. This is the basis of secondary structure determination by covariation analysis of homologous molecules (James *et al*. 1989; Michel & Costa, 1996).

Consecutive canonical W.–C. and wobble G○U pairings serve to define the secondary structure of an RNA molecule. This is also referred to as the 'two-dimensional structure' in reference to the planar representations of RNA secondary structure. Non-canonical pairs subtend and organize the tertiary or three-dimensional structure, even when they occur immediately adjacent to, or within, tracts of contiguous canonical pairs (in so-called 'internal loops', for example). W.–C. pairs can mediate 3D tertiary contacts as well. An example is provided by the isolated (and conserved) G19=C56 pair in tRNA. The tertiary motif called pseudo-knot implies W.–C. pairs between residues in a loop and another single-stranded region.

3.1.2 Cis *Watson–Crick/Watson–Crick A•G pairs*

Cis W.–C. A•G pairs form two standard hydrogen bonds (GO6–AN6 and GN1–AN1) analogous to A–U pairs. The C1′–C1′ distance in these pairs is necessarily longer than in canonical W.–C. pairs (*circa* 12·4 Å vs. 10·4 Å). It is not surprising, therefore, that they are usually observed at the ends of helices, although they can also occur within RNA double helices flanked by canonical W.–C. pairs. (For example, in the synthetic self-complementary oligonucleotide, 5′-CGCGAAUUAGCG-3′ – NDB file ARL048 – two isolated *cis* W.–C./W.–C.

Table 6. *Distributions observed for positions 26/44 (at the top of the anticodon stem) in Class I cytoplasmic elongator tRNAs. The first value in each entry is the observed number of occurrences in the database (Sprinzl* et al. *1998). The second value (in parentheses) is the statistically expected value calculated from the frequency counts using unbiased probability estimation (Chiu & Kolodziejczak, 1991)*

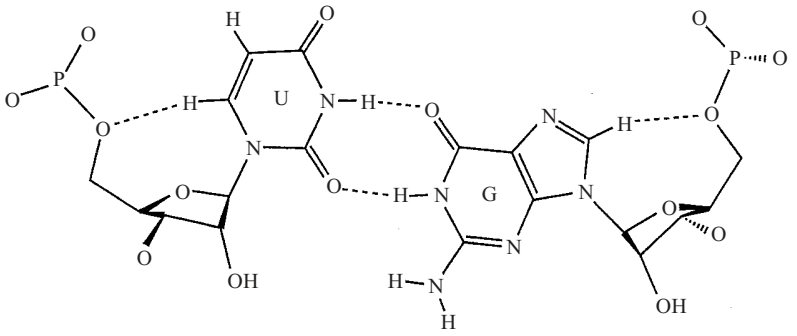
A44\G26	A	C	G	U
A	136	12	327	46
	(243)	(15)	(200)	(53)
C	76	8	22	43
	(70)	(4)	(57)	(15)
G	187	10	8	3
	(97)	(6)	(80)	(21)
U	89	19	45	13
	(78)	(5)	(64)	(17)

A•G pairs occur, but cause only minor distortion to the RNA double helix.) Examples of A•G pairs at the ends of helices in biological RNA molecules include the R26•R44 basepair at the interface between the D stem and anticodon stems of some tRNAs (e.g. TRNA09) and the A196•G126 pair in the crystal stucture of P4-P6, URX053 (Cate *et al.* 1996*a*). The *cis* W.–C. A•G pairing is self-isosteric (Table 4*a*, third column). Thus, one expects to observe A•G as well as G•A pairings at those positions in homologous molecules having this geometry (assuming no other constraint). A covariation analysis of position 26/44 of the class I elongator tRNA database (including tDNA sequences) is shown in Table 6 (Sprinzl *et al.* 1998). One observes that A•G and G•A pairs are statistically favoured, while A•A and G•G pairs are disfavoured. G•G occurs more rarely than A•A, perhaps because A•A can adopt the related wobble-type geometry, whereas G•G cannot, due to steric clash between the imino N1 of one G with the amino N2 of the other. W.–C. juxtapositions occur in a statistically neutral manner, whereas U•C and C•U pairs are statistically favoured. Since the C1′-C1′ distance in the water-inserted U•C and C•U pairs (discussed in more detail below) is significantly greater than that of canonical W.–C. pairs (11.8 Å vs. 10.4 Å) and approaches that of A•G pairs (12.5 Å), it is reasonable to suggest that water-inserted U•C/C•U may occur interchangeably with *cis* W.–C./W.–C. A•G/G•A at the ends of helices.

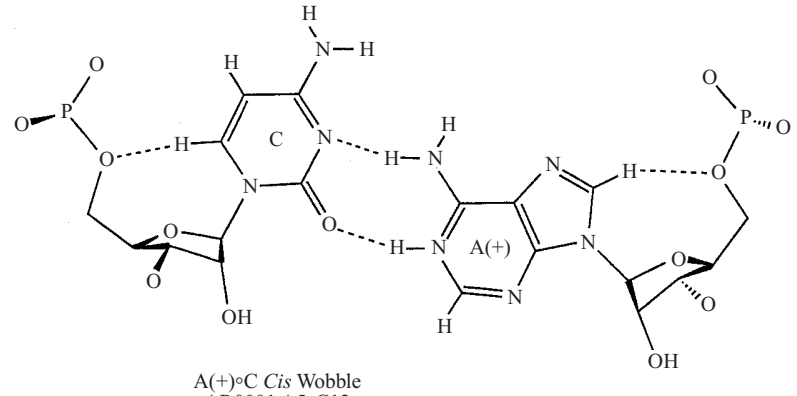
3.2 *Wobble pairings*

3.2.1 *G○U and A(+)○C wobble pairings*

The wobble pairs, the neutral G○U and the N1-protonated A(+)○C are isosteric with each other (see Fig. 3). However, they are not self-isosteric (Table 4*a*, column 3). That is to say, G○U is not isosteric with U○G, and the same applies



G∘U *Cis* Wobble
URL064 G96∘U80



A(+)∘C *Cis* Wobble
AR0001 A5∘C12

Fig. 3. Comparison of the mutually isosteric *cis* wobble basepairs (Table 4*a*), the neutral G∘U and N1-protonated A(+)∘C. Note that neither basepair is self-isosteric.

to A(+)∘C and C∘A(+). Hydrogen-bonding occurs between RN1 and YO2 and between the R6 position and YN3. The A(+)∘C pair probably has a protonated AN1, even at neutral pH: the AN1–CO2 distance observed at high-resolution (AR0001) in a crystal grown from neutral pH solution is comparable to the GN1–UO2 distance in G∘U wobble pairs (Pan *et al.* 1998). Wobble pairs frequently substitute for canonical pairs in RNA helices (Gautheret *et al.* 1995). In fact, isolated wobble pairs minimally distort the canonical RNA double helix.

Besides the C1′-C1′ distance in a basepair, the distance between P5′ of one base and O3′ of its pairing partner (when these interact so that their strands run in anti-parallel fashion) is a useful measure for judging how well adjacent basepairs interface each other. The RP5′–YO3′ distance in a *cis* wobble R∘Y pair is, on average, about 1 Å shorter than the YP5′-RO3′ distance, which closely matches that of canonical pairs. Thus, two wobble pairs interface ideally within a double-helical context when they occur with the sequence orientation 5′-YR-3′/3′-RY-5′. This arrangement brings together the shorter RP′–Y03′ distances of the adjacent wobble pairs so that the tandem unit fits more neatly into a regular RNA duplex. Not surprisingly, therefore, the 5′-UG-3′/3′-GU-5′ tandem occurs quite

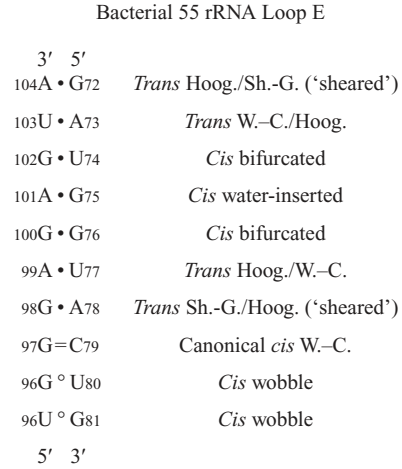


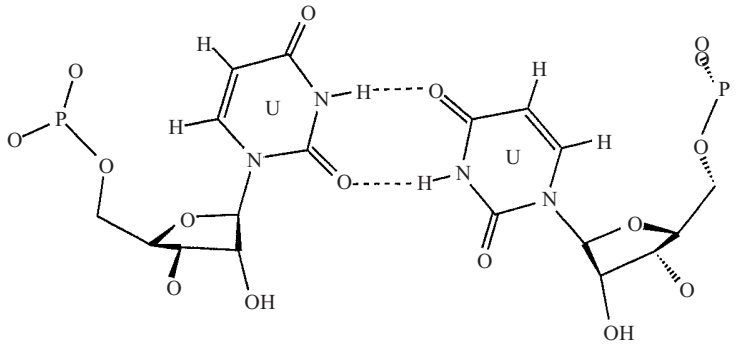
Fig. 4. Consensus sequence of Loop E of bacterial 5S rRNA. All bases are paired, although only one is a canonical pair. The pairing geometry is indicated beside each basepair.

Table 7. *Distributions observed for positions 81 and 95 in loop E of bacterial 5S rRNAs. The number of sequences having the indicated pair substituting for the consensus G81/U95 is shown in each cell. The database is the subset of 288 bacterial 5S RNA sequences described previously* (*Leontis & Westhof, 1999*). *The number of statistically expected occurrences of each pairing is shown in parentheses*

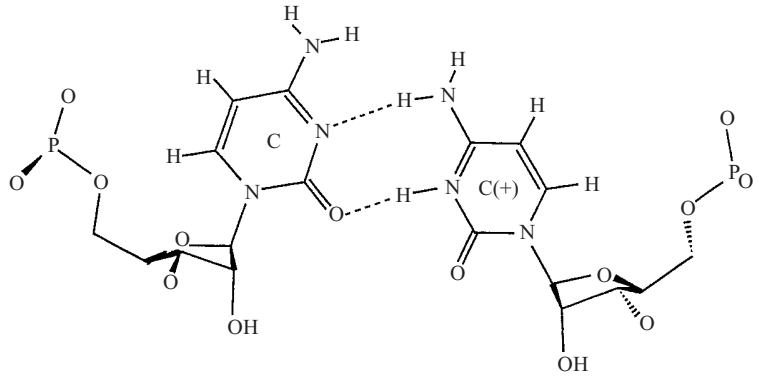
G81 U95	A	C	G	U
A	0 (0.2)	0 (0.1)	0 (1.2)	1 (0.5)
C	0 (5.4)	15 (3.6)	21 (41)	30 (15)
G	2 (0.2)	0 (0.2)	0 (1.8)	0 (0.7)
U	21 (18)	4 (12)	159 (134)	35 (50)

frequently in natural RNA molecules (Gautheret *et al.* 1995). Moreover, the 5′-UG-3′/3′-GU-5′ orientation has been found to be more stable thermodynamically than the reverse orientation 5′-GU-3′/3′-UG-5′ (Wu *et al.* 1995).

A tandem 5′-UG-3′/3′-GU-5′ motif is found at U80◦G81/G96◦U95 in loop E of the consensus sequence of bacterial 5S rRNA (NDB file URL064, shown schematically in Fig. 4) and in Helix A of the 5S rRNA of *Thermus flavus* (URX063). This motif appears to be stabilized by cross-strand purine–purine stacking (G81 on G96). The tandem 5′-CA-3′/3′-AC-5′ motif has also been



U∘U *Cis* wobble
PTR016 U5∘U17



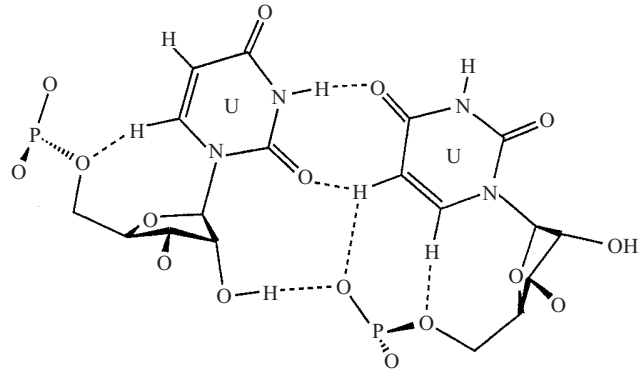
C∘C(+) *Cis* wobble
URX057 C3.0∘C17.0

Fig. 5. Comparison of the mutually isosteric *cis* wobble U∘U and C(+)∘C basepairs.

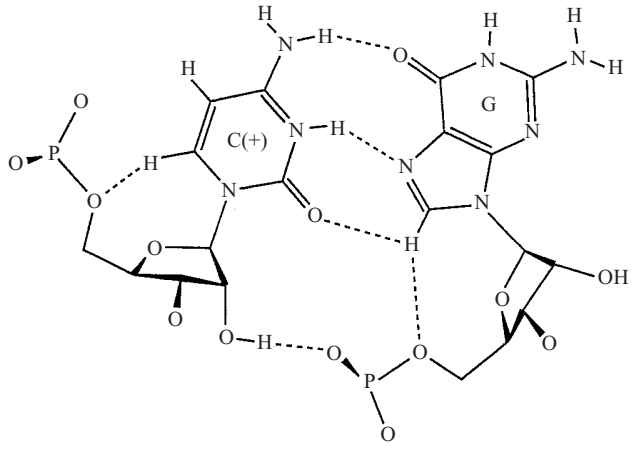
crystallized (AR0001), and has been found to be isosteric with the tandem U∘G motif. It displays identical cross-strand purine–purine stacking (A upon A). The sequential 5′-UU-3′/3′-GG-5′ motif has also been studied at high resolution (AR0008) and, as expected, does not show cross-strand stacking.

3.2.2 *Y∘Y wobble pairings*

The G81∘U95 pair in the 5′-UG-3′/3′-GU-5′ tandem motif in loop E of bacterial 5S rRNA is observed to covary with U•U, C•C, and U•C pairings as shown in Table 7 (Leontis & Westhof, 1999). It should be noted, however, that C•U substituting for G81∘U95 occurs at less than statistically expected frequency, whereas U•C (as well as U•U and C•C) occur at higher frequency. Tandem wobble U∘U pairings have been observed at high resolution (e.g. URL050) and each pair displays wobble-type geometries, although the C1′–C1′ distance is shorter than in G∘U wobble pairs. As illustrated in Fig. 5, UN3 of the first U, the one which substitutes for G, H-bonds with O2 of the second U, while O4 of the



Cis W.–C./Hoogsteen U•U
URX053 U259•U107



Cis W.–C./Hoogsteen C(+)•G
URX053 C260•G108

Fig. 6. Comparison of the mutually isosteric *cis* W.–C./ Hoogsteen pairings C(+)•G and U•U.

first U H-bonds with N3 of the second. Moreover, the same cross-strand stacking is observed between the Us substituting for Gs in these tandem pairs. C○C pairs can adopt the same geometry by H-bonding of (protonated) N3 of the first C with O2 of the second C, and N4 of the first C with N3 of the second C. Therefore, N3 of the second C remains unprotonated. It may thus be proposed that U○U or C○C substitute quasi-isosterically for G○U in tandem G○U motifs, and this is supported by the covariations observed in the bacterial 5S loop E (Leontis & Westhof, 1999). Furthermore, wobble U○C may form, with UO4 H-bonding to protonated CN3, and UN3 H-bonding with CO2. The reversed pairing, however, is not possible because of clash between CN4 and UN3. This may explain why U○C is observed to substitute for G81○U95 in preference to CU (Table 7). In summary, the following isosteric pyrimidine/ pyrimidine wobble pairs are capable

Table 8. *Examples of crystallographically observed basepairs having hydrogen bonds involving polarized C–H*

Pairing geometry	Pairing	H-bond	NDB file
Cis W.–C./Hoogsteen	U259•U107	UO2–UC5	URX053
Trans W.–C./W.–C.	U48•A15	UO4–AC2	TRNA07
Trans W.–C./Hoogsteen	U1A•U2C	UO4–UC5	URF042
Trans Hoogsteen/ Hoogsteen	1mG9•C23	GO6–CC5	TRNA12
Trans Hoogsteen/Sh.G.	A22•A13	AN7–AC2	PTE003
Trans Sh.G./Sh.G.	G212•A184	GN3–AC2	URX053
Trans Sh.G./Sh.G.	U8•A21	UO2–AC2	PTR009
Cis Sh.G./Sh.G.	U224•A152	UO2–AC2	URX053
Cis Sh.G./Sh.G.	C10.4•A23	CO2–AC2	UHX026
Cis Sh.G./Sh.G.	G110•A183	GN3–AC2	URX053

of substituting for wobble G○U (and by extension A(+)○C): U○U, C(+)○C, and U○C(+).

An example of a 'wobble' U○U closing a hairpin loop is provided by U5○U17 in a fragment of the U2 snRNA bound to the U2B–U2A protein (PTR016). The O4 of U5 projects into the deep groove (like the U in a wobble G○U pair). It is into the deep groove that the protein reaches to make specific contact, primarily with residues in the large hairpin loop. Interestingly, one contact involves this wobble U•U pair. The sidechain amino group of Lys 20 H-bonds to the O4s of both U5 and U17 in the deep groove of the RNA helix.

3.3 Cis *Watson–Crick/Hoogsteen pairings*

3.3.1 Cis *Watson–Crick/Hoogsteen U•A and C•G*

Examples of both U•A and C•G pairings are found in the group I ribozyme P4-P6 structure (URX053). The U135•A187 pair has *syn* A, hence, anti-parallel strands. This pair is flanked by a bulged base and by two W.–C. basepairs. The C260•G108 interaction belongs to a base triple (C260•G108=C213), with G108 in the standard *anti*-conformation. The strands bearing C260 and G108 are therefore *parallel*. The U•A and C•G pairs are isosteric. C260-N4 H-bonds to G108-O6, while U135-O4 H-bonds to A187-N6. U135-N3 H-bonds to A187-N7, while CN3 (most likely protonated) interacts with GN7.

3.3.2 Cis *Watson–Crick/Hoogsteen U•U*

The base triple U259•U107–A214 in P4–P6 includes a U•U *cis* W.–C./Hoogsteen pairing with parallel strands. It is isosteric with the *cis* W.–C./Hoogsteen U•A and C•G discussed in the previous paragraph, as shown in Fig. 6. The 'C–H' edge of U107, which is also paired canonically to A214, interacts with the W.–C. face of U259 via a standard H-bond between U259–N3 and U107–O4. In addition, an H-bond between U259–O2 and U107–C5 may be inferred from the short distance observed (shown in bold type in Table 4*a* and also listed in Table 8). One can also

expect to observe *cis* W.–C./Hoogsteen C•C isosteric to U•U with H-bonding between CO2 and CC5, and between CN3 and CN4. The YO2–YC5 hydrogen-bonds involve polarized C–H. Such H-bonds are observed mediating base-pairing interactions in a variety of pairing geometries (examples are collected in Table 8).

3.3.3 Cis *Watson–Crick/Hoogsteen R•R*

Cis A(+)•G pairings involving the W.–C. edge of A and the Hoogsteen edge of G have been observed in a self-complimentary oligonucleotide (AR0006). AN6 and GO6 are H-bonded. A second H-bond between protonated AN1 and GN7 is inferred from the distance (2.6Å) separating the two nitrogens. One predicts the isosteric *cis* W.–C./Hoogsteen G•A pair, with GO6–AN6 and GN1–AN7 H-bonds, and the nearly isosteric G•G pairing with N1–O6 and N2–N7 hydrogen bonds. In fact, the G•G pairing is observed by NMR in RNA aptamers that bind citrulline and arginine (Yang *et al*. 1996). A•A would not be expected due to steric clash of the N6 amine groups.

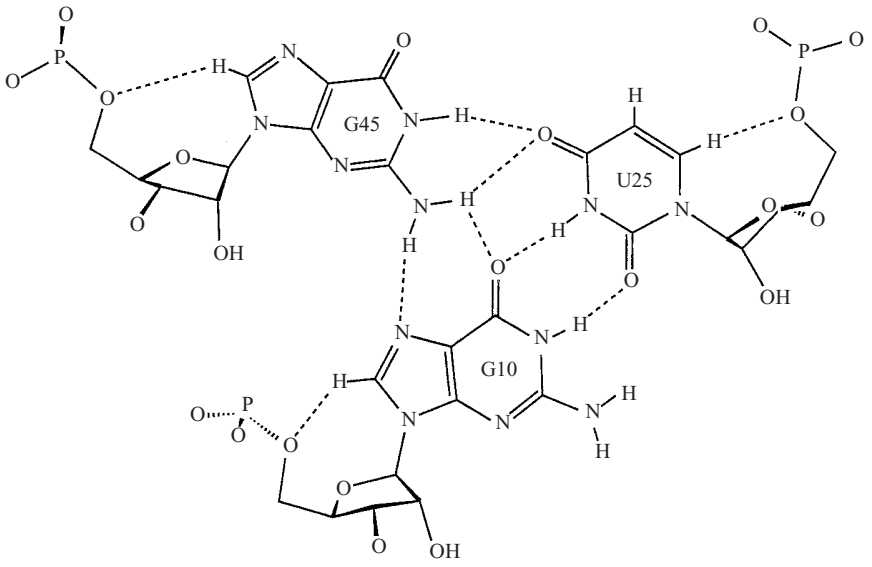
The R•R pairings have the same geometric parameters as the *cis* W.–C./Hoogsteen Y•Y and Y•R pairings discussed above, but are not isosteric to them. The C1′-C1′ distance is significantly longer in the R•R pairs than in the Y•R or Y•Y pairs. The A(+)•G pairings in Ar0006 occurs with the G in the *syn* conformation and thus have locally (and even globally) anti-parallel strands. Moreover, the C1′–C1′ distance is 10.54 Å, which allows this pair to fit neatly into the RNA double helix without distortion.

3.4 *Bifurcated pairings*

3.4.1 Cis *bifurcated G•G, A•A, G•U, and A•C*

Bifurcated pairings involving Gs were first observed in tRNA structures. The best example is the conserved G18•G Ψ55 pairing in which O4 of Ψ55 (corresponding to O2 of U) forms (bifurcated) H-bonds to both N1 and N2 of G18. The N3 of Ψ55 is exposed and interacts with a phosphate oxygen belonging to A58. The second example from tRNA is the tertiary interaction between G45 and U25 in yeast $tRNA^{Asp}$ (for example, TRNA07). The O4 carbonyl of U25 forms bifurcated H-bonds to the N1 and N2 positions of G45. Interestingly, U25 also forms a wobble pair with G10, as its W.–C. positions, UN3 and UO2, are available for H-bonding. The G45•U25•G10 base triple, featuring the bifurcated G45•U25 interaction, is shown in Fig. 7.

An identical G•U bifurcated pair was recently observed in Loop E of bacterial 5S rRNA (URL064), where it occurs within an internal loop context (G102•U74, see Fig. 8). The O4 H-bond acceptor of U74 H-bonds to both the N1 and N2 donors of G102. A water molecule can be seen in this high-resolution structure H-bonding to both N3 of U74 and N2 of G102. It is located at a position equivalent to the A58 phosphate oxygen that H-bonds to the Ψ55-N3 in the G18•Ψ55 bifurcated pair in tRNA, and to the G10-O6 atom in the G45•U25◦G10 base triple (Fig. 7). A second water molecule bridges between G102–o6 and U74–C5 in the pair from 5S rRNA loop E. Comparison of the G102•U74 pair in 5S rRNA and



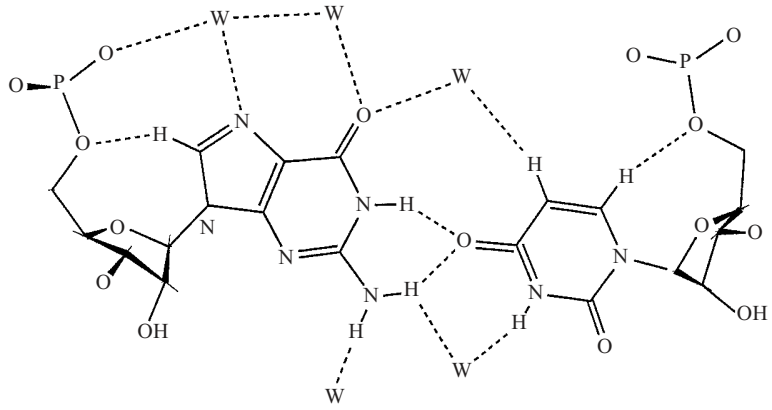
Cis bifurcated G45•U25 and *Cis* wobble G10•U25 TRNA07

Fig. 7. *Cis* bifurcated G45•U25 in tRNA showing the positioning of U25 to interact with its W.–C. edge with G10.

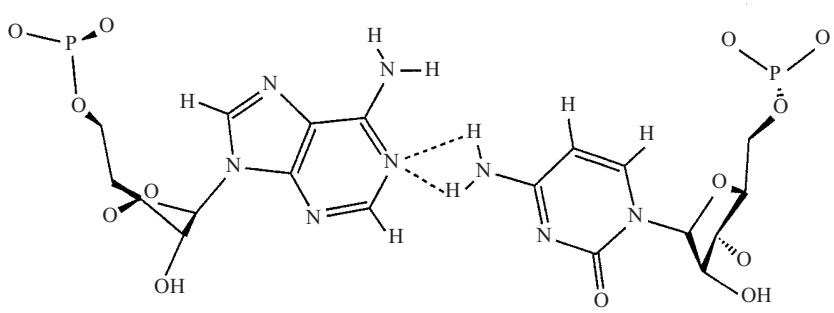
the G45•U25 pair in tRNA suggests that the U in the bifurcated pairing of loop E may interact with another RNA base (possibly belonging to 23S rRNA) using its W.–C. face.

The observation of locally parallel phosphodiester backbones in the G18•Ψ55 base pair suggests that bifurcated G•U pairs can occur equally well with a locally parallel orientation of the strands (with O2 of U interacting with N1 and N2 of G). Moreover, bifurcated G•Y could also occur with a locally anti-parallel orientation.

Bifurcated G•G also occurs in loop E of bacterial 5S rRNA (G76•G100). Bifurcated G•G, in which O6 of one G hydrogen-bonds with both N1 and N2 of the second G, is isosteric with bifurcated G•U (Leontis & Westhof, 1998*a*). This initially surprising result is due to the nearly equal distances between UO4 or GO6 and their respective C1′ atoms. Bifurcated G•U pairs are not isosteric with their reversed pairs (as is also the case for *cis* wobble pairs). The conservative substitutions observed for the G102•U74 pairing in the bacterial 5S rRNA database (Szymanski *et al.*, 1998) are A•A, A•C, A•U, G•A, G•C and G•G, of which A•A and A•C are statistically favoured (Leontis & Westhof, 1998*a*). The G76•G100 pairing covaries almost exclusively with A•A with a strong statistical bias against A•G pairs, giving a signature opposite to that of the water-inserted A101•G75 pair of loop E (see below). In addition, conservative A•C and G•A substitutions are also observed, whereas neither C•A nor A•G is. The same signature of covariations is observed for the G102•U74 pair. The conservative G•G substitutions for G102•U74 need no further discussion, as they demonstrate explicitly that bifurcated G•U and G•G are interchangeable. The G76•G100 pair



Cis bifurcated G•U: G102•U74 from URL064

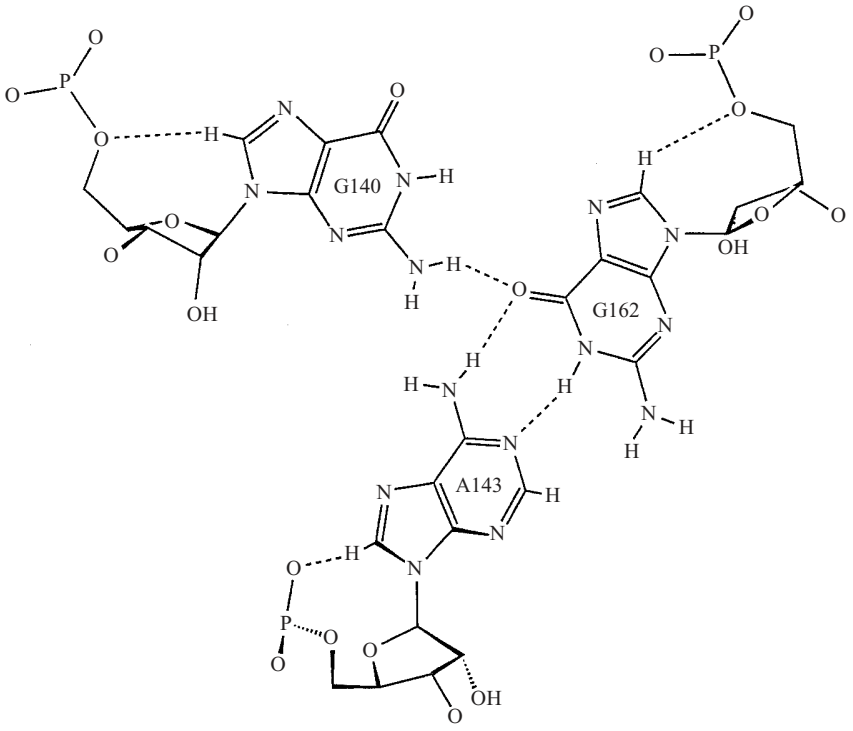


Isosteric *cis* bifurcated A•C: A38•C32 from TRNA12

Fig. 8. Comparison of isosteric *cis* bifurcated G•U and A•C. Note that neither basepair is self-isosteric.

may be superimposed almost exactly upon the G102•U74 in space: while the sugar phosphate backbones superimpose precisely, the G replacing U in the G•G pairing, extends further into the shallow groove.

The A•A substitutions for the 102/74 and 76/100 pairings can be accommodated by A74 (or equivalently A100) H-bonding via its N6 amino group to the N1 of A102 (equivalently A76). Bifurcated H-bonds result from the interaction of both amino N6 protons of one A with the N1 acceptor of the other. A slight reorientation can lead to the formation of an additional H-bond between N7 of A74 and N6 of A102. A•C is just as easily accommodated at these positions by using AN1 as the H-bond acceptor and CN4 as the H-bond donor in place of the GN1–UO4 H-bond. Again, bifurcated H-bonding results from the simultaneous interaction of both amino CN4 protons with AN1. A water molecule can potentially bridge from CN3 (or AN1 in the A•A pairing) to the polarized base proton H2 of A74 (or equivalently A102). An isosteric *cis* bifurcated pairing has, in fact, been observed at the end of the anticodon stem of yeast initiator tRNA (A38•C32 in TRNA12). The isosteric *cis* bifurcated G•U and A•C pairs are compared in Fig. 8.



Cis bifurcated G140•G162 and
Cis W.–C./W.–C. G162•A143 from PR0005

Fig. 9. *Cis* bifurcated geometry of G140•G162 from the hepatitis δ ribozyme showing the positioning of G162 to interact with its W.–C. edge with A143. Note the geometrical similarity to the G•U•G triple in Fig. 7.

The G•A substitutions for these pairings require geometrical adaptation, as the amino group of the A would otherwise be directed toward the imino of G. One possibility is that the amino N6 of A could pair with the carbonyl O6 of G. The A102•U74 pairing is geometrically possible as an isosteric replacement for G102•U74, but is expected to be less stable than G•U unless AN1 is protonated. Overall, a correlation is observed between conservative substitutions and isosteric pairings for the bifurcated pairings. Moreover, one sees an overlapping set of substitutions for the 102•74 and 76•100 pairings in loop E, which taken together clearly define the sequence variation signature that can identify these pairings in other contexts. For examples, see Leontis & Westhof (1998*a*).

In analogy to the G45•U25 bifurcated pair in tRNA, it may be anticipated that, in a bifurcated G•G pair, the G corresponding to U can also participate in W.–C. pairing with a third base. In fact, such a situation occurs in the hepatitis δ ribozyme structure (PR0005). The relative orientation of the bases in the G140•G162 pair is similar to that of the bifurcated G•G pair in the 5S loop E structure, except that G162-O6 appears, on distance and angle criteria, in this (relatively low-resolution) structure to H-bond only with G140-N2 (see Fig. 9).

The W.–C. face of G162 pairs with the W.–C. face of A143. The lateral shift of G162 relative to G140 may be induced by pairing with A143 to allow G162–O6 to H-bond with A143–N6. It should be noted that molecular dynamics simulation of $tRNA^{Asp}$ has demonstrated coordinated switching between alternative H-bonding patterns over a 500 ps time scale in the G45•U25○G10 triple containing the G4•U25 bifurcated pair (Auffinger *et al.* 1999). The wobble interactions between U25 and G10 are stable over this time scale, whereas G45–N2 alternately interacts with G10–N7 in one conformation and with U25–O4 in the other. Similar dynamic effects were observed for other base triples in the study cited and may be relevant to the hepatitis ribozyme triple G140•G162•A143.

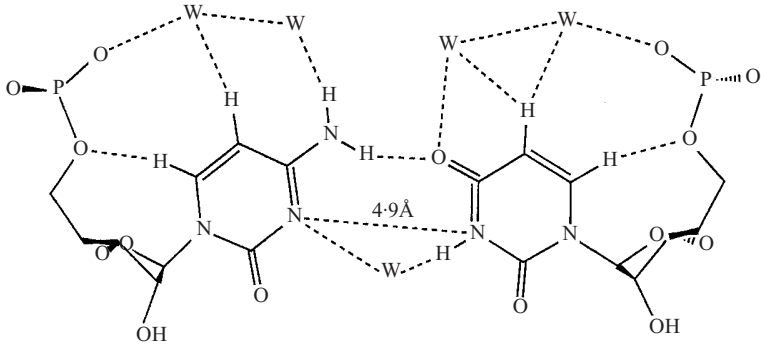
The bifurcated G188•U168 pairing in the P4–P6 domain of the Group I ribozyme (URX053) is an example of an isolated pairing mediating two tertiary interactions. The first is between G188 and U168, each of which belongs to a different arm of the P5abc three-way junction. The second tertiary interaction involves H-bonding between N3 of U168 and the O2′ of G111 in P4, in direct analogy with the previously described tertiary interactions involving bifurcated G•U, G•G, and G•Ψ. Uridines are frequently seen to H-bond with their imino N3 donor atom to the RNA backbone, as is commonly observed for U33 in the anti-codon loop. (Another example from P4–P6 is U177 interacting with a phosphate oxygen of G163 at the P5abc three-way junction.)

3.5 Cis *open and water-inserted*

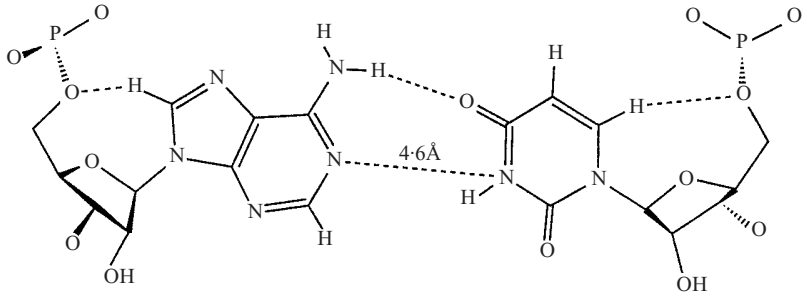
3.5.1 *G•A and U•C*

Water-inserted ('open') G•A and U•C pairings have been observed crystallographically at high resolution. The open G•A pair occurs in loop E of 5S rRNA (G75•A101) and is found centrally located between the isosteric G102•U74 and G76•G100 basepairs described above. In G75•A101, GO6 and AN6 H-bond directly, but a water molecule is inserted between GN1 and AN1. The C1′-C1′ distance is one of the largest observed for an RNA basepair (14.8 Å). The G•A pair covaries almost exclusively with A•G in the bacterial 5S rRNA database (Leontis & Westhof, 1998*a*). Symmetry alone indicates that the pair is self-isosteric. Although three A/A substitutions are also observed in the database of bacterial 5S sequences, there is a strong statistical bias against homopurine pairings. No case of G/G pairing is observed. A shift in geometry would be required to prevent clash of the amino groups in the A•A pair, leading to N6 of one base H-bonding with N1 of the other. Therefore, this is not an isosteric substitution. The isosteric A75•G101 pairing may be generated by simply rotating the G75•A101 basepair around the (pseudo-symmetric) axis passing between the bases, perpendicular to the axis of the double helix, after which the sugar phosphate backbones of the original and rotated pairs are found to superimpose exactly.

In the *cis* U•C pairing, UO4 and CN4 are H-bonded directly, while an inserted water molecule bridges UN3 and CN3. The C1′–C1′ distance (11.8 Å) is longer than that of canonical pairs, approaching that of the *cis* A•G W.–C. pairing, with



Water-inserted C•U: C8•U7 from AR0005



Water-inserted A•U (infererred from geometry):
A15·1•U16·1 from URX059

Fig. 10. Comparison of water-inserted U•C pair (AR0005) with the A15.1•U16.1 pair that occurs adjacent to the three-way junction in the hammerhead ribozyme (e.g. URX059). From the AN1–UN3 distance one infers the presence of an inserted water molecule in the hammerhead structure.

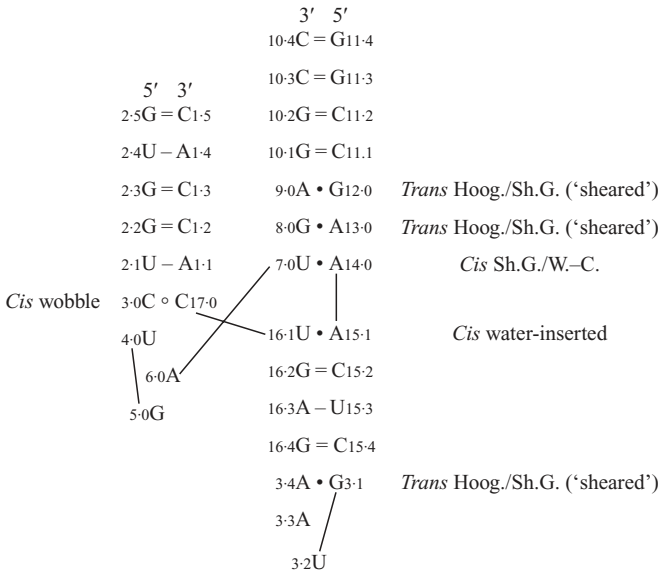


Figure 11. Secondary structure of hammerhead ribozyme corresponding to URX059, identifying non-canonical basepairs.

which it has been observed to covary, as discussed above. It has been observed crystallographically in oligonucleotide structures (e.g. AR0005). The water-inserted U•C pair is compared in Fig. 10 with the A15.1•U16.1 pair that occurs adjacent to the three-way junction in the hammerhead ribozyme (e.g. URX059). (The non-canonical pairs of the hammerhead ribozyme are labelled on the secondary structure in Fig. 11.) As shown in Fig. 10, the A15.1•U16.1 pair also opens toward the shallow groove and the distance between the imino nitrogens of the A and U bases is comparable to that observed in the U•C pair. Therefore, it is reasonable to suggest that a water molecule also bridges the imino nitrogens of the A15.1•U16.1 pair. However, the A•U and U•C pairings are not isosteric, as the C1′-C1′ distance is greater in the A•U pair (13.25 Å vs. 11.8 Å).

4. *TRANS* BASEPAIRS

4.1 Trans *Watson–Crick/Watson–Crick*

4.1.1 Trans *Watson–Crick/Watson–Crick A•U and G•C pairings*

The *trans* W.–C. pairs A•U and G•C are each self-isosteric, owing to their rotational symmetry about an axis perpendicular to the plane of the basepair. In the A•U pair, AN1 H-bonds to UN3 and AN6 to UO2, whereas in the G•C pair, GN1 H-bonds with CO2, and GN2 with CN3. Therefore, the two pairings are not isosteric with each other. Nonetheless, in the classic example of this pairing, the R15•Y48 'Levitt Pair' in tRNAs (Levitt, 1969), trans W.–C. A•U and G•C are found to occur interchangeably, as shown in the covariation table for Class I elongator tRNAs (Table 9). The isosteric *trans* 'wobble' pairs (A•C and G•U) also occur for this pair, although they are statistically strongly disfavoured. (They will be discussed below.) The G•G and A•A covariations that are observed will be discussed under *trans* bifurcated pairings.

4.1.2 Trans *Watson–Crick/Watson–Crick A•A and G•G pairs*

Trans W.–C. A•A pairs have been observed (Table 4*b*). An example is A151•A248 in P4–P6 (URX053), in which the bases are both anti, and, therefore, the strands are parallel. The N1 of one adenosine hydrogen-bonds with N6 of the other. This two-fold rotationally symmetrical pair is naturally self-isosteric. The isosteric G•G pair may be expected to occur, as H-bonding between O6 of one G with N1 of the other is also possible in a similarly symmetric and self-isosteric fashion. In the example given, A248 also pairs by *trans* Hoogsteen interaction with U224, while A151 belongs to a GAAA tetraloop.

The C1′-C1′ distance in the *trans* W.–C. A•A (or G•G) pairing is considerably longer than in the *trans* W.–C. A•U or C•G pairs (Table 4*b*). It has been proposed on the basis of NMR studies that the G•G pair in the Rev-binding element (RBE) of the HIV-1 Rev response element (RRE) adopts a *trans* geometry (Battiste *et al.* 1996). The NMR work has shown that the strands are locally parallel due to strand reversal at G71 caused by the presence of bulged base U72. The G•G interaction was identified by covariation analysis of artificial phylogenies created for the RBE using SELEX methodology (Bartel *et al.* 1991). G48•G71 in the RBE

Table 9. *Distributions of bases in the* trans-*W.–C. pairing R15/Y48 in Class I elongator tRNAs* (*refer to Table 3 legend for details*)

Y48\R15	A	C	G	U
A	3	0	7	2
	(3)	(0)	(10)	(0)
C	19	1	817	3
	(166)	(2)	(665)	(6)
G	1	0	3	1
	(1)	(0)	(5)	(0)
U	187	0	17	0
	(40)	(0)	(162)	(1)

covaries with A•A, C•A and U•G, leading to the proposal of the symmetric *trans* pairing. However, the *trans*-W.–C. geometry cannot accommodate all four pairings isosterically. We present an alternative geometry below (*trans*-bifurcated).

4.2 Trans *wobble pairs*

The *trans* wobble pairings A•C and G•U are expected to be isosteric with each other, as equivalent H-bonds may form: GN1-UO4 and AN1-CN4, GO6-UN3 and AN6-CN3. Note that protonation of AN1 is not required. Such *trans* wobble pairings are neither isosteric with *trans* W.–C. A•U nor with *trans* G•C, as their pyrimidines are laterally shifted to allow interaction of YN3 with GO6 or AN6. As mentioned above, the *trans*-wobble pairs are observed to covary (albeit weakly) with *trans* A•U and *trans* G•C in the Levitt pair in tRNA. The U120•G125 pairing in the Hepatitis δ ribozyme (PR0005) is *trans* wobble. Owing to the *syn* conformation of G125, however, the strands are anti-parallel.

Trans wobble U•D is observed in the complex of yeast $tRNA^{Phe}$ with elongation factor EF–Tu (PTR012). The isosteric *trans* U•U is expected to have symmetric UO2–UN3 H-bonds. *Trans* C16•C59, as observed in $tRNA^{Cys}$ (PR0004), is not isosteric with *trans* U•U, as it exhibits symmetric CN4-CO2 H-bonds (Nissen *et al*. 1999).

4.3 Trans *Watson–Crick/Hoogsteen pairs*

4.3.1 *Purine•Purine*

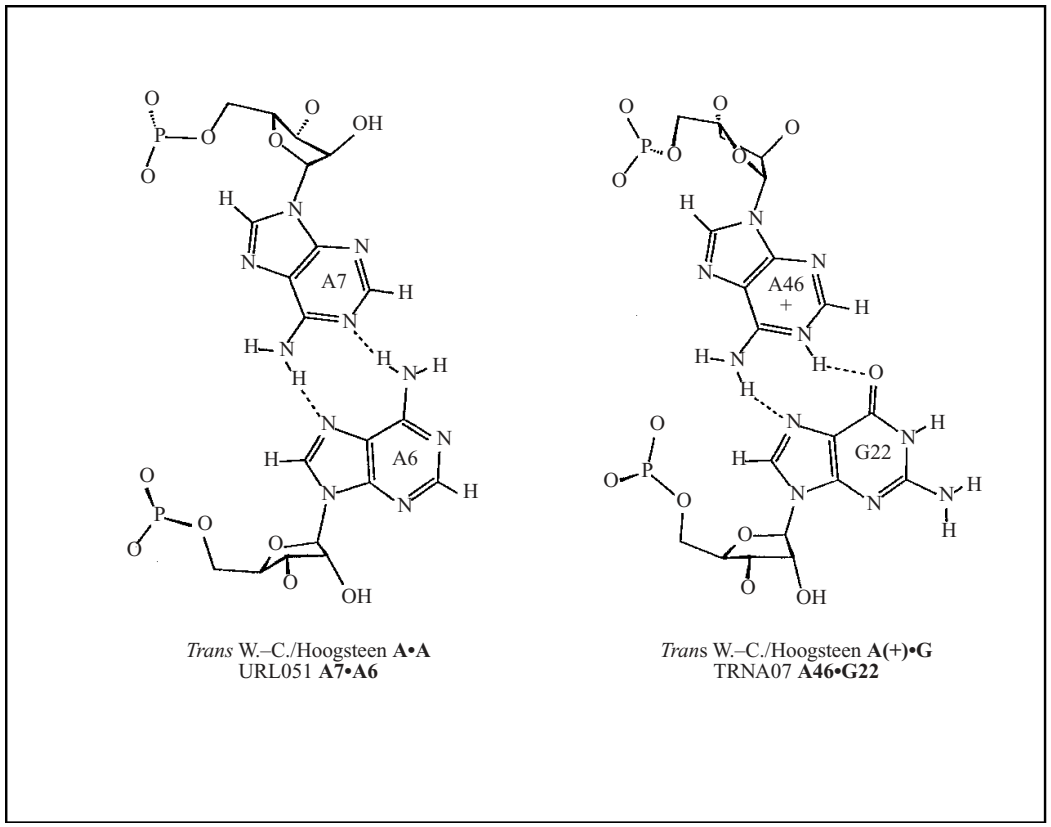
In the *trans* W.–C./Hoogsteen purine–purine pairs, one purine H-bonds via its W.–C. face and the other via its Hoogsteen face. Of the four possible combinations, three are observed. G•G is observed for the 46/22 tertiary interaction of yeast initiator tRNA (TRNA12) and of yeast $tRNA^{Phe}$ (TRNA09). *Trans* W.–C./Hoogsteen A•G, in which the W.–C. face of A interacts with the

Table 10. *Distribution analysis of R9 with the R13/A22 basepair in Class II tRNAs*

	A9	G9
A13•A22	31	48
	(9)	(70)
G13•A22	2	222
	(25)	(196)

Hoogsteen face of G, is observed for the 46/22 interaction of $tRNA^{Asp}$ (TRNA07). In addition, *trans* W.–C./Hoogsteen A•A is observed adjacent to 'sheared' (i.e. *trans* Hoogsteen/Sh.G.) G•A pair in a synthetic RNA oligonucleotide (URL051). (One should note the similarity of this motif to the 5′-U•A-3′/3′-AG-5′ motif having *trans*-Hoogsteen U•A adjacent to sheared G•A in bacterial 5S rRNA loop E and in the sarcin loop of 23S rRNA, as discussed below.) One would not expect trans W.–C./Hoogsteen G•A (W.–C. edge of G interacting with the Hoogsteen edge of A) to occur due to clash of GN2 and AN6, and, in fact, covariation analysis of tRNA bears this out. In Class II tRNAs, the last basepair of the D-stem (positions 13/22) is overwhelmingly A•A or G•A and, indeed, the crystal structure of *T. thermophilus* $tRNA^{Ser}$ (PTR004) shows that this is a sheared pair (see below) and not a *trans* W.–C./Hoogsteen pair. Interestingly, the W.–C. edge of G13 interacts with the Hoogsteen edge of G9 in the *trans* geometry. In Class II tRNAs G9/G13/A22 is observed to exchange with G9/A13/A22 and with A9/A13/A22. The A9/A13/A22 covariation is statistically favoured while A9/G13/A22 is disfavoured, as shown in Table 10. The sequence analysis thus supports the conclusion that three of the four *trans* W.–C./Hoogsteen are sterically allowed. Note that for *cis* W.–C./Hoogsteen geometry, all R•R pairs except A•A are sterically allowed, while for the *trans* W.–C./Hoogsteen geometry all pairs except G•A are allowed. The two geometries thus give distinctive sequence signatures.

In the *trans* W.–C./Hoogsteen A•R pairings, AN6 H-bonds with RN7 and AN1 with AN6 or GO6. In the A•G pair, AN1 is probably protonated. In the G•G pairing GN1 H-bonds with GN7 and GN2 with GO6. Thus, among the three allowed R•R pairings of this type, AG and A•A are mutually isosteric but are not isosteric to G•G. In the G•G pairing, the bases are displaced laterally to optimize H-bonding between N1 and N7 and between N2 and O6 (refer to Fig. 12). The fact that G•G and A•R *trans* W.–C./Hoogsteen pairs occur interchangeably in certain contexts (as in tRNA) once again indicates the adaptiveness of RNA tertiary interactions. This is shown also by the substitution of *trans* W.–C./Hoogsteen A•A for T54A58 (with A58 usually modified to 1 mA to prevent W.–C. pairing). The *trans* W.–C./Hoogsteen A•A and U(or T)•A pairings share the same geometrical parameters but are not isosteric. A•A is, in fact, the standard pairing at these positions in eucaryal initiator tRNAs.



Trans W.–C./Hoogsteen **A•A**
URL051 **A7•A6**

Trans W.–C./Hoogsteen **A(+)•G**
TRNA07 **A46•G22**

Trans W.–C./Hoogsteen **G•G**
TRNA09 **G22•7mG46**

Fig. 12. Comparison of *trans* W.–C./Hoogsteen purine/purine pairs. The A•A and A(+)•G pairs (shown in the box) are isosteric, whereas the G•G pair involves a small lateral shift.

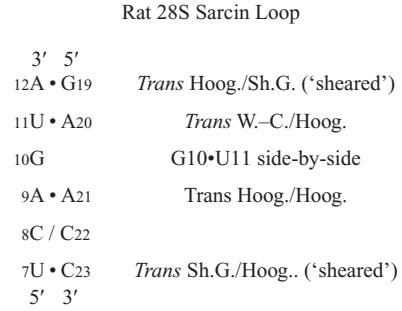


Fig. 13. Two-dimensional schematic of the sarcin/ricin loop from rat 28S rRNA.

4.3.2 Trans *Watson–Crick/Hoogsteen U•A*

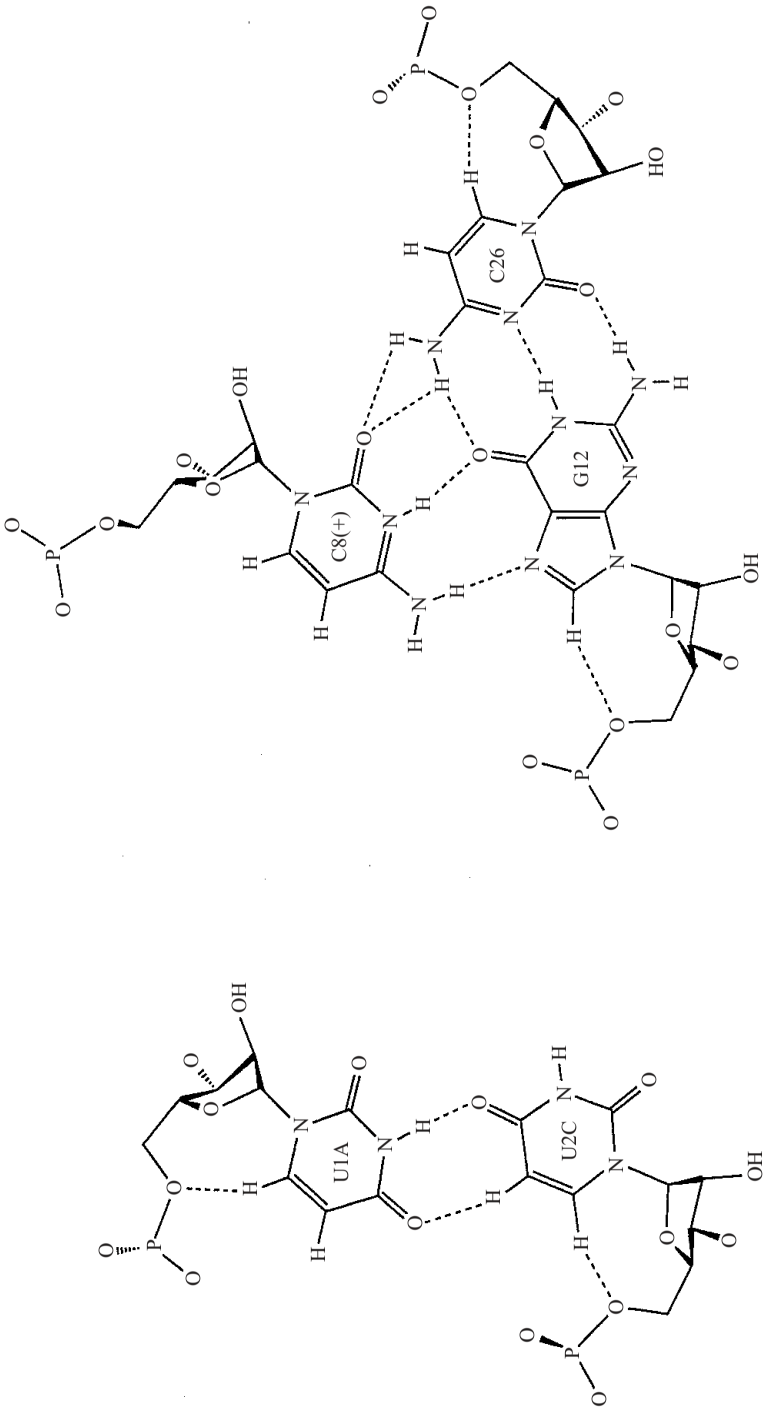
This U•A pairing occurs at the highly conserved U8•A14 and T54•1mA58 tertiary interactions in elongator tRNAs. The pairing also occurs adjacent to sheared R•A, as in loop E of 5S rRNA (U103•A73 and U77•A99 in URL064, Fig. 4) and in sarcin loop motifs (U11•A20 in UR0002, Fig. 13). UN3 H-bonds to AN7 and UO2 H-bonds to AN6. H-bonding between A•C8 and UO4 may also be suggested.

4.3.3 Trans *Watson–Crick/Hoogsteen C•G*

The *trans* W.–C./Hoogsteen C•G pair is observed between residues C8 and G12 in the recently published pseudo-knot structure, UR0004 (Lu *et al.*, 1999). The W.–C. face of C8 interacts with the Hoogsteen face of G12 (Fig. 14). The (most likely) protonated C8–N3 atom H-bonds to G12–O6 while C8–N4 H-bonds to G12–N7. G12 also forms a W.–C. pair with a second cytosine (C26). Exactly the same interaction is seen between residues C141 and G161 in the hepatitis delta ribozyme. In this case also, the G is W.–C. paired to a second cytosine, C144, demonstrating again the modular nature of RNA structure. The hydrogen bonding is identical in the two examples. The *trans* W.–C./Hoogsteen C(+)•G pair is not exactly isosteric with the *trans* W.–C./Hoogsteen U•A pair, since it is the N4 of C (and not N3) which H-bonds to the N7 of G. This is consistent with the failure to observe C•G pairs covarying with trans W.–C./Hoogsteen U•A pairings either in the tRNA or in the loop E or in the sarcin loop structures. However, one does observe a small number of C•A substitutions for the *trans* W.–C./Hoogsteen U•A pairs in loop E of 5S rRNA (Leontis & Westhof, 1998*a*). In the recent crystal structure of the TtLSU group I intron (UR0003), one finds the *trans* W.–C./Hoogsteen C255•A218 pairing (Golden *et al.*, 1998). CN4 H-bonds to AN7 while AN6 forms H-bonds to both CO2 and CN3. This (low-resolution) C•A pair appears to be geometrically closer to the C•G pair than to the U•A pair.

4.3.4 Trans *Watson–Crick/Hoogsteen U•U*

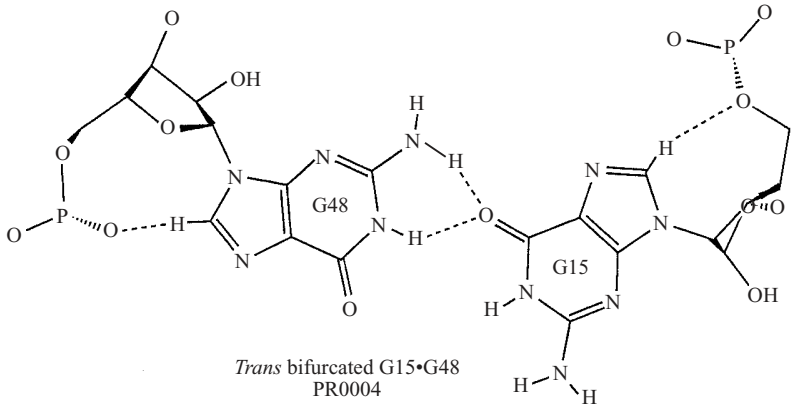
An example of this pairing is observed in the inter-molecular interaction between uridines on overhanging ends of synthetic oligonucleotide duplexes in the crystal



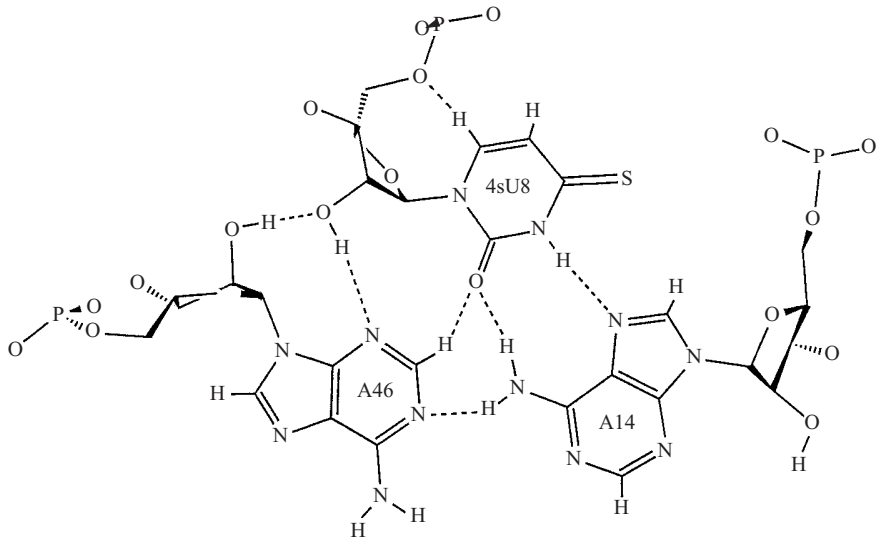
Trans W.–C./Hoogsteen C(+)•G
UR0005 C8•G12 = C26

Trans W.–C./Hoogsteen U•U
URF042 U1A•U2C

Fig. 14. Comparison of *trans* W.–C./Hoogsteen U•U with isosteric C•G. The example illustrates the use of the W.–C. face of G to form a triple interaction with another C.



Trans bifurcated G15•G48
PR0004



Trans bifurcated A14•A46
Trans W.–C./Hoogsteen 4sU8•A14
cis Sh.G./Sh.G. A46•4sU8
PR0004

Fig. 15. Comparison of *trans* bifurcated G•G and A•A pairings observed in tRNA.

structure URF042 (Wahl *et al.* 1996). H-bonding occurs between N3 of one uridine and O4 of the other. Moreover, as in the *cis* W.–C./Hoogsteen U•U pair discussed above, H-bonds involving C5 can be inferred from the short distance between O4 of the first uridine and C5 of the second (see also Table 8). It is observed that *trans* W.–C./Hoogsteen U•U is isosteric to C•G, as shown in Fig. 14, which also illustrates the triple interaction that makes use of the W.–C. face of G to form a canonical pair with another C.

Table 11. *Distribution analysis of* trans *Hoogsteen/Hoogsteen pairing in 8 sarcin/ricin loop motifs identified in 23S rRNA. Due to the symmetry of the interaction, occurrences of X/Y and Y/X are combined*

Bacteria

	A	**C**	**G**	**U**
A	803			
C	155	0		
G	290	1	0	
U	12	1	0	0

Archaea

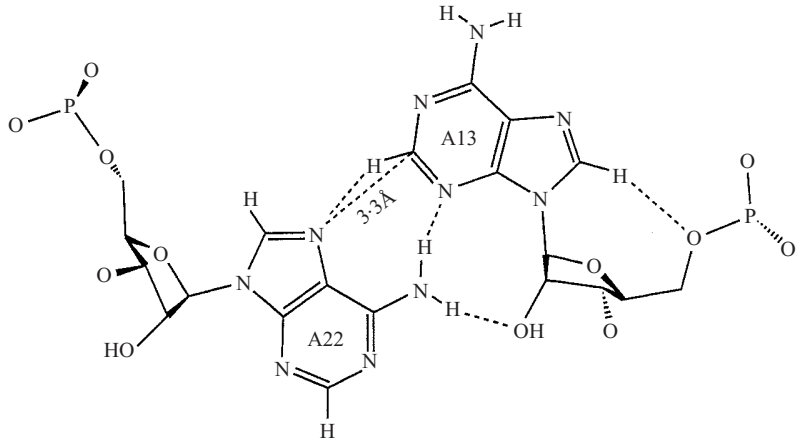
	A	**C**	**G**	**U**
A	144			
C	5	0		
G	29	0	1	
U	2	0	1	0

Eucarya

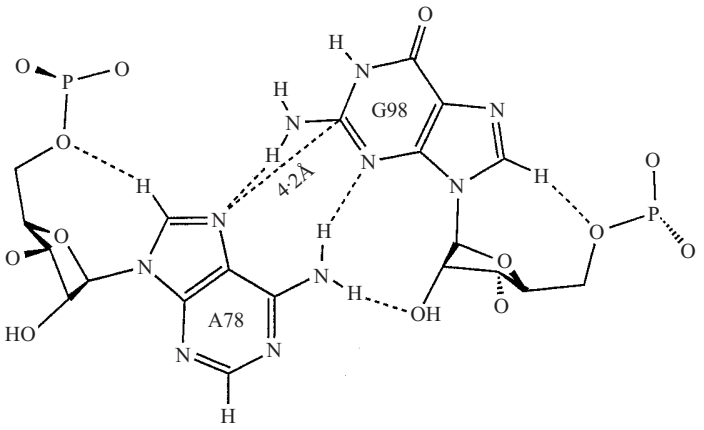
	A	**C**	**G**	**U**
A	536			
C	5	0		
G	17	1	0	
U	9	0	2	0

4.4 Trans *Hoogsteen/Hoogsteen pairs*

Such pairings, like *trans* W.–C./W.–C., are rotationally symmetric and, therefore, self-isosteric. When both bases are *anti*, the strands are, necessarily, locally parallel. The pairing occurs in the sarcin internal loop motifs (including loop E of eucaryal and certain archaeal 5S rRNAs) and in Class I tRNAs in the tertiary interaction involving positions 9 and 23. While in the rRNA internal loops the locally parallel orientation of the strands is compensated by reversal at the level of the sugar-phosphate backbone (due in part to the presence of a bulged base), in Class I tRNAs the strands are globally parallel because of the 3D fold. The most common pairing observed in Class I tRNAs is A•A, in which the N6 of each adenosine H-bonds symmetrically with the N7 of the other base. A9•A23 as well as G9•C23 and C9•G23 are observed in tRNA crystal structures. These are the most common covariations observed among Class I elongator tRNAs. Also observed are the covariations A/C, A/G, A/U and G/G. However, since base 9 can also interact with position 13, it is possible that some of these covariations do not represent actual interactions. There is evidence that this is the case for the A/U and G/G covariations. The occurrence of G9/G23 in the Class I database correlates with the occurrence of non-canonical pairings for 13/22. This is also the case for the A/U covariations. In fact, the A9/U23 covariation occurs in $tRNA^{Cys}$ (PR0004), but the two bases do not interact in the crystal structure. Instead, the



Trans Hoogsteen/Sh.G. ('Sheared') **A•A**
PTE003 A22•A13



Trans Hoogsteen/Sh.G. ('Sheared') **A•G**
URL064 **A78•G98**

Fig. 16. Comparison of *trans* Hoogsteen/Sh.G. ('sheared') A•G and A•A pairings showing small adjustments in structure to allow H-bonding between AN7 and AC2.

Hoogsteen face of A9 interacts with the W.–C. face of A13. A13 also interacts with the Hoogsteen edge of A22 using its shallow-groove edge in sheared fashion (see *trans* Hoogsteen/Sh.G. pairings below). Like the G9/G23 covariations, A9/U23 covariations in the Class I database correlate with non-canonical 13/22 pairs, and thus behave like Class II tRNAs.

Eight conserved sarcin loop motifs were identified in 23S rRNA (Leontis & Westhof, 1998*b*). A covariation analysis of the *trans* Hoogsteen/Hoogsteen position of all eight motifs, separated by phylogenetic group, is presented in Table 11. Due to the symmetry of the interaction, the data were folded across the diagonal. The data reveal A/N as the principle variants. Interestingly, the

isosteric G/C pairings observed for the 9/23 interaction in tRNA are not found in the sarcin loop motifs. The data suggest *trans* Hoogsteen/Hoogsteen A•C and A•G pairs. These can be constructed by H-bonding of AN7 with CN4, and of GN7 with AN6.

4.5 *Trans* bifurcated pairings

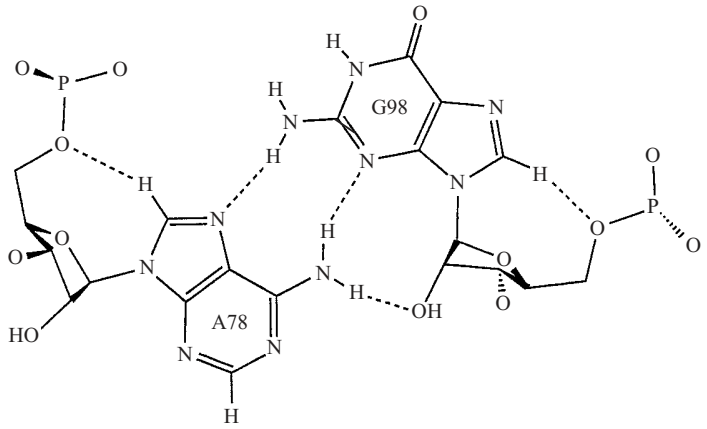
Trans bifurcated G•G and A•A are observed in the crystal structure of $tRNA^{Cys}$ (PR0004). The *trans* bifurcated pairing G15•G48 found in $tRNA^{Cys}$ substitutes for the more commonly occurring *trans* W.–C./W.–C. R15•Y48 pairing discussed above. The A46•A14 pairing in $tRNA^{Cys}$ is also *trans* bifurcated and essentially isosteric with G15•G48, just as the *cis* bifurcated G•G and A•A pairs are with each other. *Trans* bifurcated G•U and A•C may be modelled that are isosteric with G•G and A•A, just as is the case for *cis* bifurcated G•U and A•C. On the other hand, *trans* W.–C./W.–C. A•A and G•G have C1′-C1′ distances considerably longer than those of *trans* W.–C./W.–C. A•U or G•C (see Table 4*b*). It is therefore reasonable to propose that the G•G pairing in the RBE motif discussed above is also *trans* bifurcated to allow accommodation of the A•A, A•C, and G•U covariations observed in artificial phylogenies (Bartel *et al.*, 1991). Furthermore, it would be difficult to distinguish the *trans*-W.–C. from the *trans*-bifurcated geometries based on NMR data alone.

The *trans* bifurcated G•G and A•A pairings are compared in Fig. 13, which also shows the 4sU8 which forms a base triple with the *trans* bifurcated A46•A14 pair. The U8•A14 interaction is the conserved *trans* W.–C./Hoogsteen pairing, while the A46•4sU8 interaction is of the *cis* Sh.G./Sh.G. type discussed below. This illustrates the ability of the *trans* bifurcated geometry to mediate tertiary interactions, just as is observed for the *cis* bifurcated pairs.

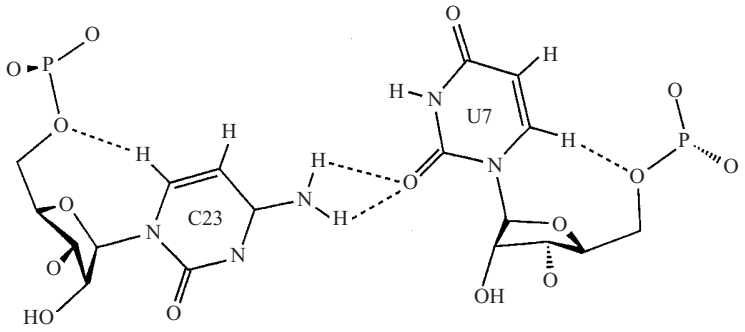
It should also be noted here that the conserved bifurcated G18•Ψ55 pair from tRNA discussed previously has *trans* orientation of the glycosidic bonds, implying parallel strands. It bears repeating that the bifurcated pairing geometry in this case also serves to position the imino proton of one of the bases (Ψ55) to form another interaction, in this case H-bonding to the phosphate of A58, as noted above.

5. SHALLOW-GROOVE PAIRINGS

These pairings involve the Sh.G. edge of one or both interacting bases. Hydrogen bonding groups on the Sh.G. edge include RN3, YO2, GN2 and AC2, and very frequently the O2′ of the ribose (Table 1 and Fig. 1). Interactions involving the Sh.G. edge of one base and the W.–C., Hoogsteen, or shallow-groove edge of the second base, have been observed in both the *cis* and *trans* orientations of the glycosidic bonds. Thus, all six possible geometries have been observed. In most cases existing examples allow us to predict other isosteric pairings. (The strand orientations corresponding to each of these six geometries were presented in Table 2, where it was assumed that each base retains the default *anti* base-sugar conformation.)



A•G *Trans* Hoogsteen/Sh.G. ('Sheared')
A78•G98 from URL064



C•U *Trans* Hoogsteen/Sh.G. ('Sheared')
C23•U7 from UR0002

Fig. 17. Comparison of *trans* Hoogsteen/Sh.G. ('sheared') U•C and A•G basepairs.

5.1 *Hoogsteen/Shallow-groove pairs*

5.1.1 *Hoogsteen/Shallow-groove :* trans *orientation and locally* antiparallel *strands* ('*Sheared*')

This is the prototypical shallow-groove pairing and one of the most commonly occurring. Adenosine-purine sheared pairings (A•A and A•G) have now been observed in many RNA crystal (and NMR) structures. H-bonding occurs between AN6 and RN3, and between AN7 and GN2 or AN7 and AC2. Consequently the AN7–RC2 distance is shorter in sheared A•A pairs than in sheared A•G (3·3 Å for A•A vs. 4·2 Å for A•G), as shown by comparison of A13•A22 from $tRNA^{Gln}$ (PTE003) with A78•G98 from loop E of *E. coli* 5S rRNA (Fig. 16). Nonetheless, the C1′–C1′ distances are little affected (*circa* 9·3 Å for A•A vs. 9·6 Å for A•G pairs), demonstrating that the two pairings are essentially isosteric and, in fact, occur interchangeably in many conserved motifs. The A•A

sheared pair also provides another example of a hydrogen bond involving polarized C–H (see Table 8). Sheared A•G pairs occur in 5′-GAAA-3′ tetraloops between the first and last bases of the loop, as first modelled for loop D of *X. laevis* 5S rRNAs on the basis of tRNA anticodon loops (Westhof *et al.*, 1989). The tRNA 13/22 pair (at the end of the D-stem in class II tRNAs, see Table 10) and the tetraloop examples illustrate the frequent occurrence of these pairs at the ends of canonical helices, where they always occur in the same orientation: The invariant A is always found on the 5′ end of the regular helix. This is due to the asymmetric geometry of the pairing: The RO3′–AP5′ distance is much shorter than the RP5′–AO3′ distance. The RP5′-AO3′ distance matches closely that of canonical basepairs. Not surprisingly, therefore, A•R sheared pairs also occur in tandem pairs, 5′-RA-3′/3′-AR-5′, within or adjacent to canonical helices. Examples include the hammerhead ribozyme (basepairs G12·0•A9·0 and A13·0•G8·0, Fig. 11) and the internal loop J4/5 in Group I introns (Cate *et al.* 1996*a*). Sheared A•R also occur in tandem pairs adjacent to U•A *trans*-Hoogsteen pairs. Such U•A pairs exhibit unusually short UO3′–AP5′ distances that closely match those of sheared A•R pairs. These tandem pairs are oriented 5′-UA-3′/3′-AR-5′. Examples are found in the crystal structures of loop E of bacterial 5S rRNA (URL064) and the sarcin/ricin loop of 23S rRNA (UR0002), as shown in Fig. 4 and Fig. 13. It is noteworthy that the tandem 5′-RA-3′/3′-AR-5′ and 5′-UA-3′/3′-AR-5′ motifs occur interchangeably in some contexts, an example being the J4/5 internal loop of Group I introns. Adjacent G•A pairs, 5′-GG-3′/3′-AA-5′, are much less common. An example occurs at the three-way junction P5abc found in some Group I introns (URX053), and testifies to the flexibility of the RNA backbone.

Recently, we presented a sequence analysis of 'conservative' base-pair substitutions for the consensus sheared A•G pairings in loop E of bacterial 5S rRNA (Leontis & Westhof, 1998*a*). Sequence variations that were observed in the extensive 5S rRNA sequence database (332 bacterial species) were classified as 'conservative' or 'concerted'. 'Conservative' was used to denote the substitution of a single base or basepair in the consensus sequence by a potentially isosteric pairing without changes in the immediately adjacent pairs of a particular sequence of the database. Substitutions that were accompanied by changes in adjacent basepairs were classified as 'concerted' and were screened out in further analysis. The most common conservative substitution observed for the sheared A•G in loop E is the essentially isosteric A•A pair discussed above. In addition to the anticipated A•A, we observed the juxtapositions A/C, A/U, C/C and C/U substituting for sheared A•G. These covariations were all quite unexpected. It was found, however, that all of these substitutions did, in fact, produce plausible H-bonding geometries when modelled into the 3D structure of the A104•G72 bp of the loop E crystal structure URL064.

The question arose, therefore, whether these juxtapositions can actually produce sheared-type geometries in real RNA molecules. Re-examination of the structures of non-canonical basepairs in high-resolution RNA crystal structures in the Nucleic Acid Database revealed precedents for A•C, C•C, and C•U bp with

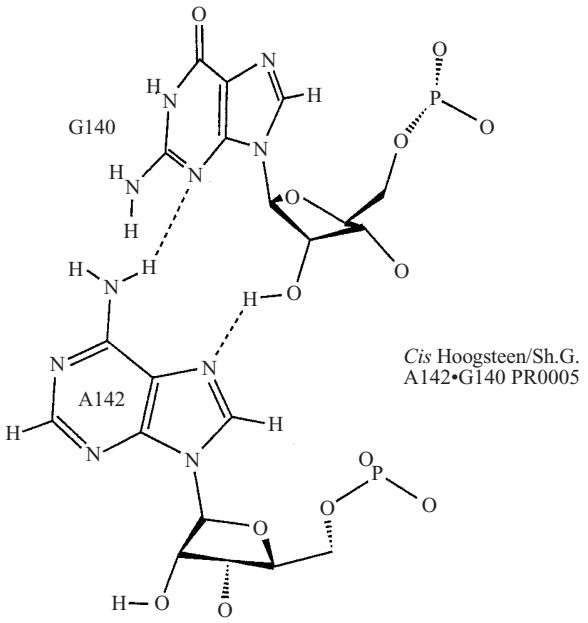


Fig. 18. Example of a *cis* Hoogsteen/Sh.G. basepair.

sheared type geometries. An example is the non-canonical C23•U7 pair that occurs in the recently solved crystal structure of the alpha-sarcin loop of 23S rRNA (UR0002). This pair occurs at the base of the sarcin 'internal loop' adjacent to the W.–C. pair C6•G24. The sheared A104•G72 of 5S rRNA loop E also occurs adjacent to a W.–C. pair (105/71) and with the same orientation. Superposition on the computer screen shows that in fact C23•U7 is isosteric with A•G, indicating that the C•U covariations observed for positions 104/72 in the bacterial 5S rRNA database may also adopt the sheared geometry. The *trans* Hoogsteen/Sh.G. C•U and A•G pairs are compared in Fig. 17. Note also the bifurcated H-bonds between UO2 and CN4.

The C38•Ψ32 pair at the base of the anticodon loop in yeast $tRNA^{Asp}$ (TRNA07) is essentially identical to C23•U7 of the sarcin loop. The corresponding A38•mC32 pair in yeast $tRNA^{Phe}$ (TRNA09) is also isosteric to sheared A•G. Bifurcated $NH_2\cdots O=C <$ interactions such as those observed in these C•U, C•Ψ, and A•C pairs have been observed to be stable in multiple molecular dynamics simulations (Auffinger & Westhof, 1996). A38•C32 in yeast initiator tRNA (TRNA12) displays a bifurcated H-bonding geometry isoteric to bifurcated G102•U74 observed in bacterial 5S rRNA loop E, as discussed above (see Fig. 8). However, the authors of this study report that the anticodon loop is not well defined in the TRNA12 structure (Basavappa and Sigler, 1991). The geometry of the U38•U32 pair observed in the complex of $tRNA^{Gln}$ with its cognate aminoacyl synthetase (PTE003) is very similar to bifuracted G•U but not exactly isoteric, due to the shorter C1′–C1′ distance.

Table 12. *Distribution analysis of positions 32/38 at the base of the anti-codon loop of class I tRNAs*

38\32	A32	C32	G32	U32
A38	13	**570**	0	**149**
	(13)	**(492)**	(1)	**(221)**
C38	0	**124**	0	**73**
	(4)	**(133)**	(0)	**(60)**
G38	1	6	0	4
	(0)	(8)	(0)	(4)
U38	4	4	0	**88**
	(2)	(65)	(0)	**(29)**

Table 13. *Distribution analysis of positions 400/374 in eucaryal 23S rRNA. In bacterial and archaeal sequences, sheared A•R pairing is inferred for this pairing*

G400\A374	A	C	G	U
A	1	33	0	2
	(6)	(18)	(0)	(1)
C	3	5	0	0
	(2)	(4)	(0)	(0)
G	9	0	0	0
	(2)	(5)	(0)	(0)
U	0	20	0	0
	(4)	(10)	(0)	(1)

Covariation analysis of the 38/32 positions in class I tRNA and tDNA sequences is shown in Table 12. Note the bias against W.–C. and wobble pairs. (A38•U32 can adopt a conformation identical to sheared A38•C32 and should not be counted as a canonical W.–C. pair.) Most data in Table 12 can be explained on the basis of the two observed geometries, *trans* Hoogsteen/Sh.G. (with bifurcated $NH_2 \cdots O=C$ H-bonds) for A38•Y32 and C38•Y32 oppositions and the singly bonded U32•U38 type (with an $O4 \cdots N3$ H-bond). A more detailed discussion of these pairings will soon be available (Auffinger and Westhof, in press).

As discussed above, the sarcin/ricin loop motif is recurrent. Eight conserved occurrences have been identified in 23S rRNA alone (Leontis & Westhof, 1998*b*). For most of these, the sheared A•R pairing is highly conserved (only A/A and A/G covariations occur). However, a richer pattern of covariations was observed for one of these motifs, which comprises sheared A374•G400 (*E. coli* numbering) and is found in a junction loop in Domain 1. In bacterial and archaeal sequences, the pairing is strictly conserved. However, in the eucaryal 23S rRNA sequence database the covariations shown in Table 13 are found. Notable is the statistically

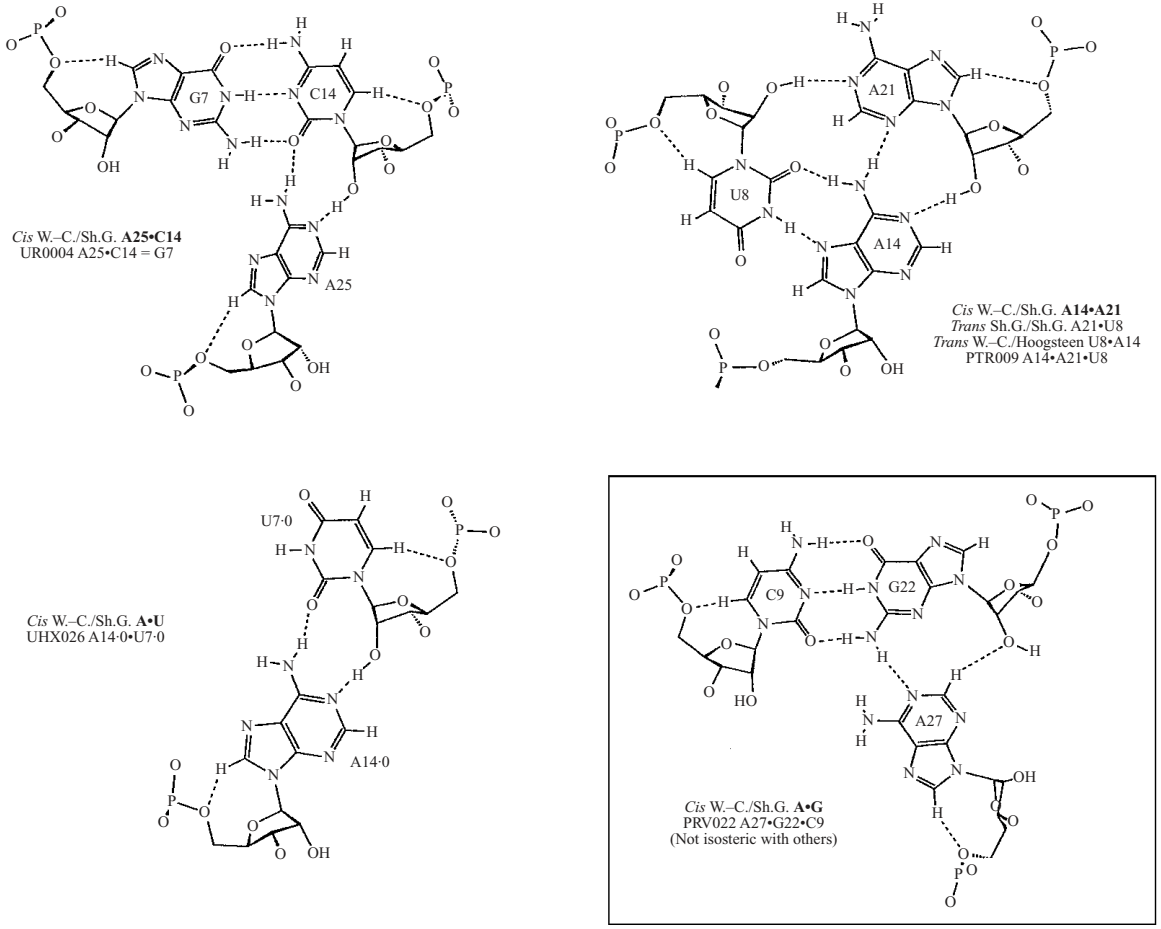


Fig. 19. Comparison of *cis* W.–C./Sh.G. pairings. Note that A•Y and A•A are isosteric whereas A•G is not.

significant fraction of C374/A400 and C374/U400 pairings substituting for sheared A374•G400. Taken alone, these covariations were, at first sight, quite baffling. However, when taken in light of the covariations observed for tRNA 32•38 and for 5S rRNA A104•G72, the 23S rRNA covariations paint a consistent picture of pairings isosteric to sheared R•A.

5.1.2 *Hoogsteen :* Cis *orientation and locally parallel strands*

As noted above, this geometry is found in side-by-side pairings, which will be discussed separately below. The *cis* Hoogsteen/Shallow-groove geometry can also occur between nucleotides that are not immediately adjacent in the primary sequence. An example is provided by the interaction between A142 and G140 in the hepatitis delta ribozyme. H-bonds are observed between AN7 and GO2′, and between AN6 and GN3, as shown in Fig. 18. An isosteric A•A pair may, therefore, be anticipated. Note the lack of co-planarity of the bases. In fact, the W.–C. edge of A142 interacts with the Sh.G. edge of C121, which in turn pairs canonically with G139. G139 also stacks on G140, which interacts with G162 in a cis bifurcated geometry (see above). Another example is provided by A261 in the Group I intron (UR0003), interacting via its Hoogsteen face with the Sh.G. of C203. A•U pairs of this kind can also be anticipated (Table 4*c*).

5.2 *Watson–Crick/Shallow-groove pairings*

5.2.1 *Watson–Crick/Shallow-groove*: Cis *orientation and locally* antiparallel *strands*

Two identical A•C pairings of this type (A23•C15 and A25•C14) are observed in the frameshifting pseudoknot structure (UR0004). Both C15 and C14 are canonically paired via their W.–C. faces (to G6 and G7 respectively). The adenosines belong to the third strand of the pseudoknot. In the A•C interaction, an H-bond occurs between AN6 and CO2, as also seen in the trans Hoogsteen/Sh.G. ('sheared') A•C pairs described above. A second H-bond forms between AN1 and the ribose O2′ of the cytosine. The same pairing is observed – but in a different context – in the Hepatitis δ ribozyme (PR0005, A142•C121 = G139), where again the interacting C is canonically paired to a G. (As mentioned above, A142 is also *cis* Hoogsteen/Sh.G. paired to G140, which stacks immediately above G139.) The anticipated isosteric A•U pair is observed in the hammerhead ribozyme, flanking the three-way junction (A14.0•U7.0, URX057). Nearly isosteric to these AY pairs are A•A pairs such as A14 •A21 at the end of the D-loop in tRNA (e.g. $tRNA^{Asp}$, TRNA09). In the A•A pairs, A-N6 H-bonds with A-N3, and A-N1 H-bonds with A-O2′ (Fig. 19).

The *cis* W.–C./Sh.G. A•G pair has also been observed (PRV022, A27•G22=C9), but is not isosteric to the A•Y and A•A pairings. The C1′–C1′ distance is shorter in the A•G pair (see Table 4*c*), due to a relative rotation of the interacting bases that is necessary to prevent steric clash between the amino groups of GN2 and AN6. Representative *cis* W.–C./Sh.G. pairs are compared in Fig. 19.

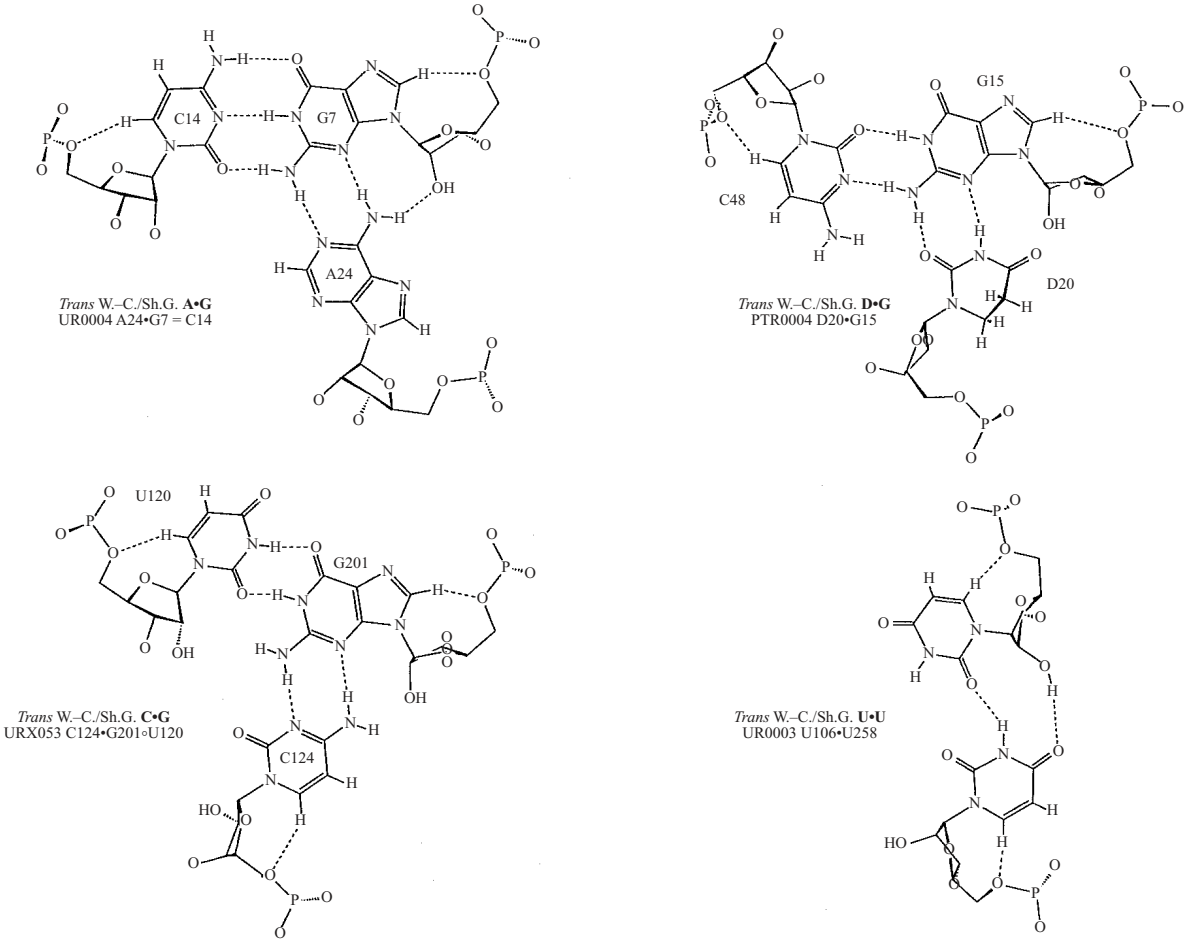


Fig. 20. Comparison of *trans* W.–C./Sh.G. pairings. Note that none of the pairs is isosteric.

5.2.2 *Watson–Crick/Shallow-groove :* trans *orientation and locally* parallel *strands*

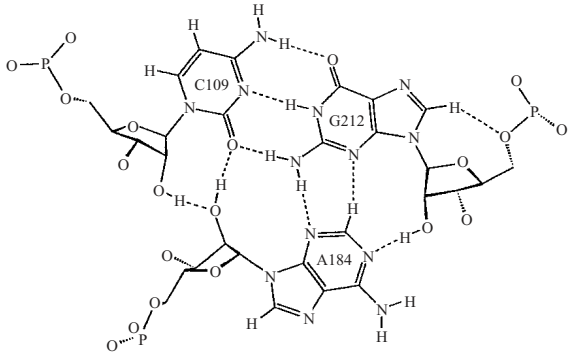
The *trans* W.–C/Sh.G. pairing was first suggested for GNRA hairpin loops interacting with helical G=C pairs (Jaeger *et al*., 1994; Michel and Westhof, 1990). The interactions actually observed crystallographically for these sites are the related *cis* and *trans* Sh.G./Sh.G. (see below). A24•G7 in the pseudoknot structure (UR0004) provides an example of a *trans* W.–C./Sh.G. pairing. The W.–C. edge of A24 interacts with the Sh.G. edge of G7, which in turn is canonically paired to C14. Both A24 and G7 are in *anti* configurations, the glycosidic bonds are *trans*, and the strands are, therefore, locally parallel (Table 2). AN6 H-bonds with GN3 and with GO2′ while AN1 H-bonds with GN2. An isosteric A•A pair may be anticipated, but neither G•G nor G•A is possible. The almost isosteric C•G pairing is observed in URX053, stabilizing the sharp bend between the P4 and P5 domains of the Group I intron (C124•G201). H-bonds occur between CN4 and GN3 and between CN3 and GN2. A third interaction of this type is observed in the *T. thermophilus* seryl-tRNA structure (PTR004) involving the W.–C. edge of D20 (dihydrouracil) and the shallow groove of G15. This pairing indicates that an isosteric U•G pairing of this type is possible. Interestingly, an almost isosteric U•U pairing is observed in the Group I ribozyme structure (U106•U258, UR0003). Examples of *trans* W.–C./Sh.G. pairings are shown in Fig. 20. Note that none is exactly isosteric, making it difficult to anticipate the range of possible pairings that can occur for this geometry. In this regard, the situation resembles that of the *cis* and *trans* W.–C./W.–C. and W.–C./Hoogsteen geometries, which can accommodate canonical as well as wobble-type pairings that are not exactly isosteric with each other by lateral shifts of the base-pairing partners that align complementary H-bond donors and acceptors.

5.3 *Shallow-groove/Shallow-groove pairings*

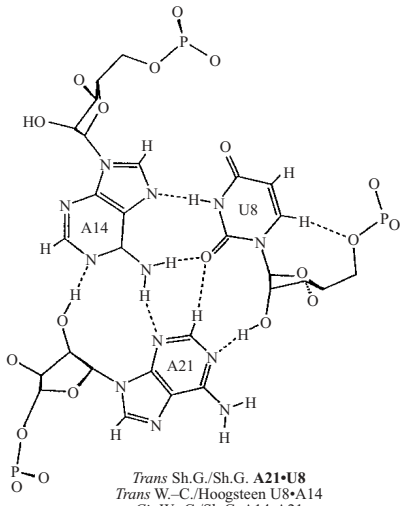
5.3.1 *Shallow-groove/Shallow-groove :* trans *orientation and locally* parallel *strands*

The purine/purine pairing of this type is symmetrical. It involves H-bonding between positions 2 and 3 of each of the interacting purine bases. When both bases are in *anti* configurations and the glycosidic bonds are oriented *trans*, the strands are locally parallel (Table 2). An example is G212A184 (Fig. 21) in the A-rich bulge of P4–P6 (URX053). H-bonds occur between GN2 and AN3, and between GN3 and AC2 (note the H-bond to a polarized C–H). Many examples of the A•G pairing have been observed. The G•G pairing involves two regular H-bonds and is observed in minor-groove crystal packing contacts between two DNA B-type helices (Wing *et al*. 1980). The A•A pair would be geometrically possible but would have two weak H-bonds (both C2H···N3).

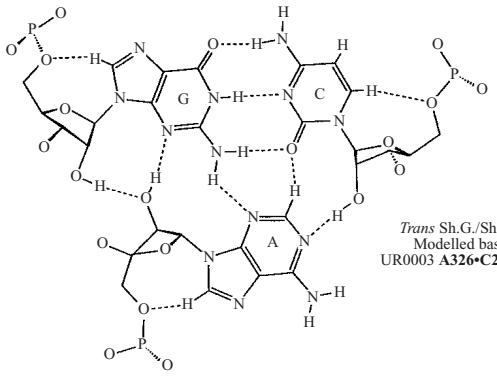
The isosteric *trans* Sh.G./Sh.G. A•U pairing occurs, in which the U also pairs in *trans* W.–C./Hoogsteen fashion to another A. An example is provided by the A21•U8 interaction in the A21•U8•A14 base triple in tRNA (PTR009 provides a



Trans Sh.G./Sh.G. **A184•G212**
URX053 A184•G212 = C109



Trans Sh.G./Sh.G. **A21•U8**
Trans W.–C./Hoogsteen U8•A14
Cis W.–C./Sh.G. A14•A21
PTR009 A21•U8•A14



Trans Sh.G./Sh.G. **A•C**
Modelled based on
UR0003 **A326•C203** = G118

Fig. 21. Examples of isosteric *trans* Sh.G./Sh.G. pairings.

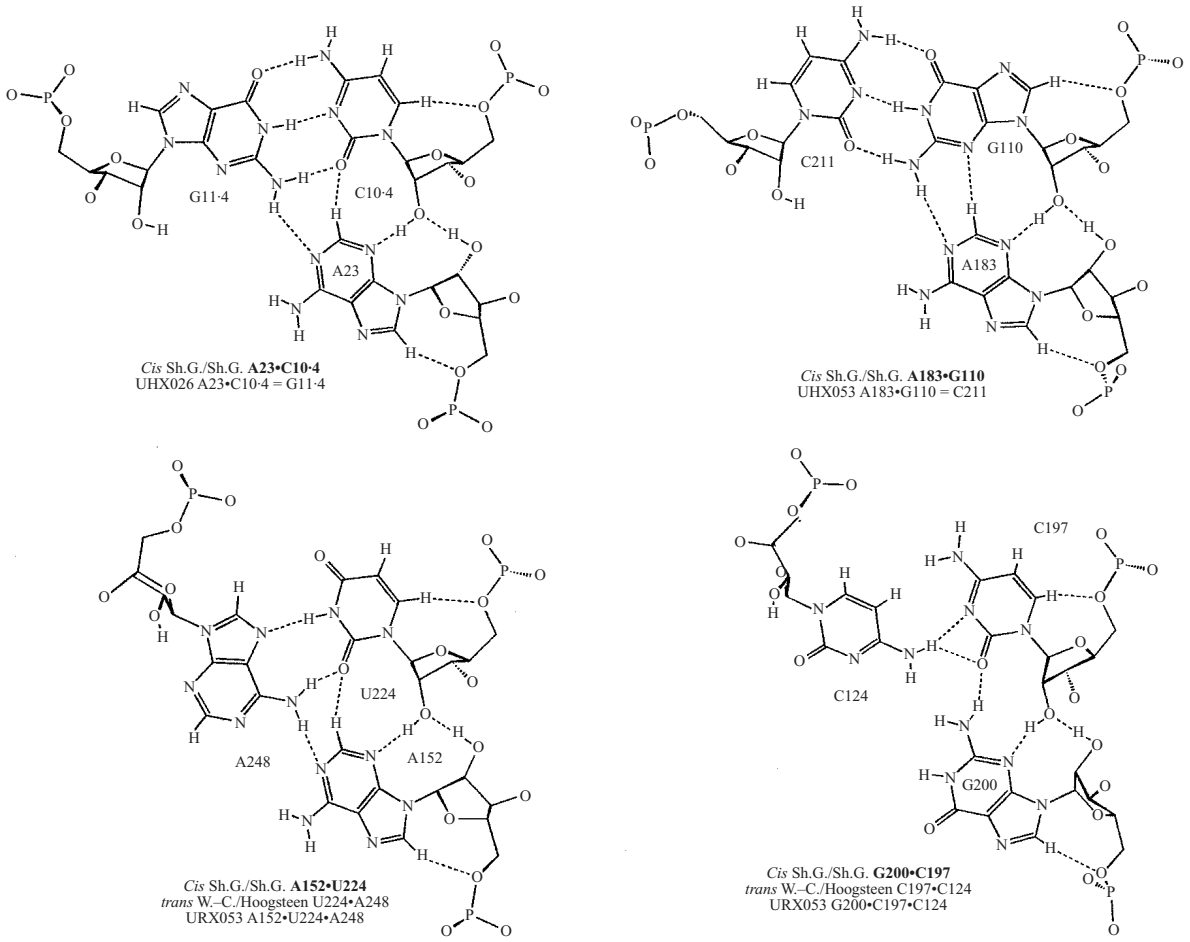


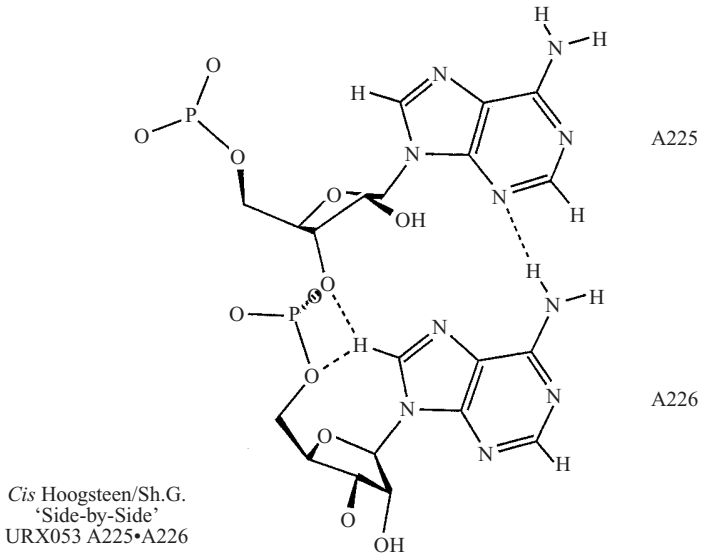
Fig. 22. Examples of isosteric *cis* Sh.G./Sh.G. pairings.

clear example, see Fig. 21). In this base triple, A21 forms a *trans* Sh.G./Sh.G. pair with U8, while U8 and A14 are *trans* W.–C./Hoogsteen paired. A14, in turn, forms a *cis* W.–C./Sh.G. pair with A21 (an interaction already discussed). The N6 amino group of A14 takes the place of the N2 amino group of the G in the A•G *trans* Sh.G./Sh.G. pairing discussed in the preceding paragraph. Thus, G=C is effectively replaced by *trans* W.–C./Hoogsteen U•A. This ability of a *trans* W.–C./Hoogsteen U•A pair to substitute for a canonical G=C pair to interact isosterically with A in *trans* Sh.G./Sh.G. fashion is notable (see Fig. 21).

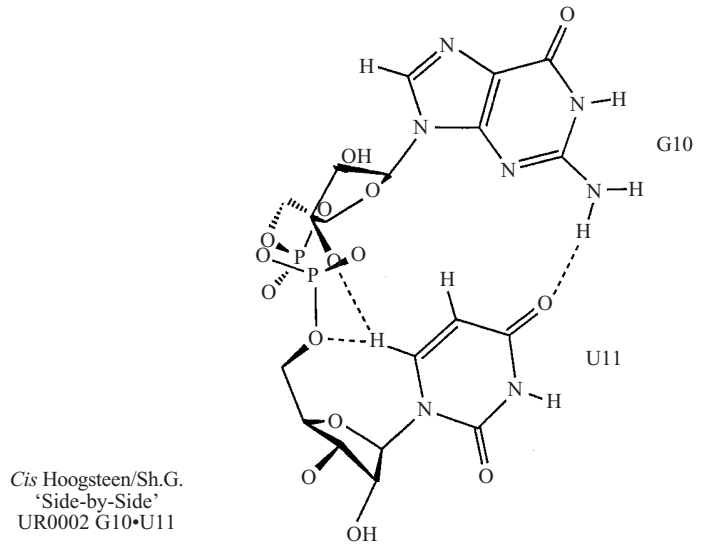
In a similar manner, substitution of a canonical G=C pair by a canonical C=G pair replaces the H-bond acceptor atom G-N3 by the acceptor C-O2, and the donor atom G-N2 by the N2 of the G replacing the C in the original pair. Superposition of the C1′ atoms of a G=C pair on those of a C=G pair shows that the corresponding GN2 atoms occur in nearly equivalent locations (Seeman *et al.*, 1976). Therefore, we can anticipate that a *trans* Sh.G./Sh.G. A•G pair can be replaced by a *trans* Sh.G./Sh.G. A•C pair if the C is canonically paired to a G. Such an A•C=G base triple appears to occur in the low-resolution tetrahymena Group I intron structure (UR0003), in the interaction of A326 in the GAAA tetraloop at the end of Helix P9b with the Sh.G. of P5 (A326•C203=G118). A model was constructed of this pairing and found to be exactly isosteric to the A•G=C and the A•U•A pairs discussed in the previous paragraph. This is also shown in Fig. 21.

Shallow-groove/Shallow-groove : Cis *orientation and locally* anti-parallel *strands*
Four isosteric examples of this geometry are given in Table 4*c*. Three of these involve the Sh.G. edge of A interacting with the Sh.G. edges of U, C, or G. An isosteric G•C pairing is also observed, in which G replaces A (G200•C197, URX053). The observation of the *cis* Sh.G./Sh.G. G•C pair suggests that isosteric G•U can also occur. The G200•C197 interaction serves to stabilize the sharp bend between the P4 and P5 domains of the Group I intron (URX053). In these pairings, equivalent H-bonds occur between the N3 positions of A or G and O2′ of the partner base. H-bonds also occur between O2′ of the partner base and O2′ of A or G, as well as between AC2/GN2 and YO2 or, equivalently, GN3. Examples of *cis* Sh.G./Sh.G. pairings are compared in Fig. 22.

The *cis* Sh.G./Sh.G. interaction frequently occurs adjacent to the *trans* Sh.G./Sh.G. pairing. Six examples from X-ray crystal structures are provided in Table 14, in which arrows are used to indicate the orientations of the three interacting strands. The bases forming the *cis* and *trans* shallow-groove pairs are underlined. All these examples involve two stacked adenosines in one strand interacting with two antiparallel, base paired strands, two of which feature non-canonical basepairs. In four of the six cases, the two stacked adenosines belong to hairpin loops and in other two (A-rich bulge-like) the adenosines belong to single-strands. In each case, the 5′ A forms a *cis* Sh.G./Sh.G. pair with the base of the strand antiparallel to it and the 3′ A forms a *trans* Sh.G./Sh.G. pair with the parallel-oriented strand. The *trans* Sh.G./Sh.G. pair is A•G in all cases but one



Cis Hoogsteen/Sh.G.
'Side-by-Side'
URX053 A225•A226



Cis Hoogsteen/Sh.G.
'Side-by-Side'
UR0002 G10•U11

Fig. 23. Examples of isosteric, adjacent *cis* Hoogsteen/Sh.G., 'side-by-side' pairings.

(UR0003). By contrast, for the *cis* Sh.G./Sh.G. pairing A•C, A•U, and A•G are observed.

The first example involves the third and fourth bases of a GAAA tetraloop interacting with its specific receptor, the so-called 11-nuceotide motif (Cate *et al.*, 1996*a*, *b*, Costa & Michel, 1995). A152 forms a *cis* Sh.G./Sh.G. pair with U224 which also participates in a W.–C./Hoogsteen pair with A248. This brings the 6-amino donor group of A248 into proximity to the N1 acceptor atom of A152.

The second example is the A-rich bulge that also comes from the Group I

intron (URX053). A closely related interaction ('A-rich bulge-like') is found in the core of the Hepatitis ribozyme (PR0005). In the A-rich bulge, the first A (residue 165) forms a *cis* Sh.G./Sh.G. pair with a G, whereas in the Hepatitis ribozyme this pair is A•C, where the C is also canonically paired to a G. In both cases, however, a favourable interaction between the amino N2 of a G and N1 of the shallow-groove A is possible (see Fig. 22), in analogy with the interaction between A152 and A248 in the GAAA receptor. Thus, at least in the context of a single strand interacting with a regular double helix, both G=C and C=G can interact in *cis* Sh.G./Sh.G. fashion with A. In both of the A-rich bulge-like examples, the *trans* Sh.G./Sh.G. pair remains A•G.

The last three examples involve hairpin loops interacting with helices. In the hammerhead ribozyme structure, UHX026, an intermolecular contact occurs that is representative of GNRA loop/helix interactions: the third residue (A23) of the terminal GAAA tetraloop interacts in *cis* Sh.G./Sh.G. fashion with C10.4, canonically paired to G11.4 (Table 14). This is identical to the A•C pair in the Hepatitis ribozyme (A165•C119).

The next example is from the interaction of a UUAAA hairpin loop with a regular helix (Conn *et al.*, 1999). In this loop the first and second As of the loop stack in a manner very similar to that found in GAAA loops. The first A forms a *cis* Sh.G./Sh.G. pair with a U which (unlike U224 in the GAAA receptor) is *canonically* paired, resulting in an A•(U-A) shallow-groove triple. The isoteric G•(U-A) interaction, in which G forms a *cis* Sh.G./Sh.G. pair with U, occurs within the context of GNRA loops interacting with the shallow grooves of regular double helices. This has been amply demonstrated by sequence analysis and experiment (Jaeger *et al.*, 1994; Michel & Westhof, 1990).

The final example in Table 14 is from the Group I ribozyme structure (UR0003), in which the A•G *cis* Sh.G./Sh.G. pair is observed albeit at low resolution. This interaction involves the third residue (A325) of a GAAA tetraloop and G119, wobble paired to U202. The wobble pairing is probably important for understanding this interaction, which is unexpected within a helical context (Jaeger *et al.*, 1994; Michel & Westhof, 1990). Wobble-pairs induce a significant underwinding of the helix relative to the basepair 3′ to the U (in this case G118=C203) (Masquida *et al.*, in press). Note also that the fourth residue of this tetraloop (A326) forms an unusual *trans* Sh.G./Sh.G A•C pair with C203 (paired canonically to G118).

In three of the *cis* Sh.G./Sh.G. interactions, the H-bond donors, GN2 or AN6, are oriented facing the AN1 acceptor atom, although the donor–acceptor distance is rather long (*circa* 3.6–3.8 Å). (In the 11-nucleotide motif, A248 also pairs, in *trans* W.–C./W.–C. fashion, with A151 of the GAAA tetraloop, and therefore A248 and A152 are not co-planar.) At the basepair level, at least, a range of substitutions are thus possible for the *cis* and even the *trans* Sh.G./Sh.G. pairings. Biochemical and phylogenetic data indicate, however, that the *three-dimensional context* plays a decisive role in determining which of the combinations identified as plausible by structural analysis at *the basepair level* actually occur in functional molecules.

Finally, as the last entry of Table 14, we provide, by way of contrast, an example of a related but distinctly different interaction. Again two adenosines in a single strand interact in the shallow-groove of a helix. However the strand has the opposite polarity and consequently the Sh.G./Sh.G. pairs are replaced by W.–C./Sh.G. pairs (discussed in the previous section).

In summary, all six possible shallow-groove pairing geometries have been observed, and the examples already in hand serve to suggest isosteric substitutions for each geometry. In the case of the *trans* W.–C./Sh.G. interactions, the pairings are not mutually isosteric. But in the five other families of Sh.G. pairings, the observed pairings are by and large isosteric. Moreover, certain pairings, notably *cis* and *trans* Sh.G./Sh.G., *trans* Hoogsteen/Sh.G. and *cis* W.–C./Sh.G. are highly recurrent and modular, occurring in many different contexts and mediating a variety of interactions.

6. SIDE-BY-SIDE BASES

Pairings between adjacent bases in the primary structure have been observed. They involve the shallow-groove edge of the 5′-base and the Hoogsteen edge of the 3′-base. An example is presented by A225•A226 (Fig. 23), which constitutes the 'AA Platform', an integral component of the recurrent eleven nucleotide motif binding GAAA tetraloops. An H-bond occurs between N3 of A225 and N6 of A226.

Very few single base changes are observed for the adjacent A•A of the platform motif. One example is the substitution of C for A226, which is found in *Dunaliella salina* SSU rRNA (GenBank M84320) and *Chlorella mirabilis* SSU rRNA (GenBank X74000). In the *Chlorella* sequence the bulged U249 is also changed to C (note that 'bulged' U249 forms a *cis* W.–C. pair with A226). The covariations are explained by the fact that an isosteric side-by-side pair can be formed with C226•A225, while neither G nor U can be accommodated at position 226. This is due to the amino group, CN4, which is positioned to interact with N3 of A225, just as A226-N6 does in the A226•A225 pairing. Furthermore, the variant C226•A225 was obtained from *in vitro* selection experiments (Costa & Michel, 1997).

The (formally) bulged G in the sarcin/ricin loop of 23S rRNA pairs with the 3′-adjacent U in a side-by-side fashion, essentially isosteric with A225•A226 of the 11-nucleotide motif (G10•U11 in UR0002, see Fig. 23). UO4 H-bonds with GN2. G•G may be anticipated, but not in this context, as U forms a highly conserved *trans* W.–C./Hoogsteen pair with an adenosine on the other strand in the sarcin motif.

7. DEFINING A LIBRARY OF ISOSTERIC PAIRINGS

The classification system presented in this review is intended to systematically organize the edge-to-edge, base-base interactions observed in RNA crystal structures, so as to provide a framework for identifying isosteric basepairs that can substitute for each other in homologous molecules. New interactions revealed by

Table 14. *Isosteric motifs comprising adjacent* cis *Sh.G./Sh.G. (A•U, A•G, and A•C) and* trans *Sh.G./Sh.G. A•G or A•C. Six examples are given, from the indicated NDB or PDB files. The bases participating in the pairing indicated in column 3 are underlined (column 2). The strand polarities are indicated by arrows in the 5′-to-3′ direction. The last example comprises adjacent* trans *and* cis *W.–C/Sh.G. pairs. Note the reversal of the relative strand polarities in this case*

NDB file	Motif	Interacting bases	Basepair type
URX053	GAAA tetraloop/ 11 nt-receptor	↓ A152 • ↑ U224•A248 ↓ ↓ A153 • ↑ C223=G250 ↓	*Cis* Sh.G./Sh.G. *Trans* Sh.G./Sh.G.
URX053	A-rich bulge	↓ A183 • ↑ G110=C211 ↓ ↓ A184 • ↑ C109=G212 ↓	*Cis* Sh.G./Sh.G. *Trans* Sh.G./Sh.G.
PR0005	A-rich bulge-like	↓ A165 • ↑ C119=G128 ↓ ↓ A166 • ↑ C118=G129 ↓	*Cis* Sh.G./Sh.G. *Trans* Sh.G./Sh.G.
UHX026	GNRA loop/Helix	↓ A23 • ↑ C10.4=G11.4 ↓ ↓ A24 • ↑ C10.3=G11.3 ↓	*Cis* Sh.G./Sh.G. *Trans* Sh.G./Sh.G.
1QA6 (PDB)	UUAAA loop/Helix	↓ A134 • ↑ U155-A104 ↓ ↓ A135 • ↑ C154=G105 ↓	*Cis* Sh.G./Sh.G. *Trans* Sh.G./Sh.G.
UR0003	GNRA loop/Helix with wobble pair	↓ A325 • ↑ G119°U202 ↓ ↓ A326 • ↑ G118=C203 ↓	*Cis* Sh.G./Sh.G. *Trans* Sh.G./Sh.G.
UR0004	Pseudoknot Sh.G. triple helix	↑ A24 • ↑ G7=C14 ↓ ↑ A23 • ↑ G6=C15 ↓	*Trans* W.–C./Sh.G. *Cis* W.–C./Sh.G.

crystallographic studies can be incorporated into this framework. A convenient way to do this is using Isostericity Matrices. This will serve, furthermore, as a test of the predictive power of the classification.

In the course of compiling the data for this review, we repeatedly found evidence for isosteric pairings that we predicted, either in new structures that appeared during the course of our writing, or upon careful scrutiny of existing crystal structures. The predictions were made on the basis of comparisons of crystal structures with covariations observed in sequence databases of homologous molecules. Examples of pairs predicted, and subsequently observed, include the A•C and C•Y *trans* Hoogsteen/Sh.G. ('sheared') pairs, the A•C and A•A bifurcated pairs (both *cis* and *trans*), and the *cis* W.–C./Sh.G. A•A and A•U pairs, predicted on the basis of *cis* W.–C./Sh.G. A•C (e.g. A23•C15 in UR0004). For nearly every geometric group, additional basepairs may be predicted using these considerations, and these are listed in the fourth column of Tables 4*a*, *b–c*. Whether the predicted interactions will, in fact, be observed will provide a crucial test of this approach.

With regard to possible new families of edge-to-edge interactions, the present classification system predicts 12 geometrical groups. These result from pairwise

Table 15. *Isostericity Matrix for canonical W.–C. pairings. 'I' indicates the isosteric canonical pairs, whereas 'C' indicates the wobble pairs, which are compatible with, but not precisely isosteric to the canonical pairing geometry*

	A	C	G	U
A		C		I
C	C		I	
G		I		C
U	I		C	

interactions between three base edges in two possible orientations. This does not separately count the wobble, water-inserted, and bifurcated pairings, which may be considered variations of the 12 basic groups. (The wobble and water-inserted pairings are essentially variations of *cis* or *trans* W.–C./W.–C., whereas the bifurcated pairings are intermediate between W.–C./W.–C. and W.–C./ Hoogsteen or W.–C./Sh.G.) In the case of the shallow-groove pairings, examples of all six groups are provided in Table 4*c*. Of the pairings involving only W.–C. or Hoogsteen edges, Tables 4*a* and *b* contain examples of all possible geometries except *cis* Hoogsteen/Hoogsteen, which we have so far failed to observe.

A simple, graphical representation of the observed substitutions for each geometry may be proposed: The Isostericity Matrix. This is a 4×4 matrix in which the isosteric pairs identified experimentally, and/or predicted theoretically, are indicated. For example, the matrix shown in Table 15 summarizes the canonical W.–C. pairing.

'I' indicates the canonical pairs that are isosteric with each other, whereas 'C' indicates wobble pairs that are compatible with the canonical pairing geometry, although not precisely isosteric to it. (As discussed above, the wobble pairs are not all isosteric to each other.) As a second example, the matrix shown in Table 16 may be proposed for *cis* bifurcated pairs on the basis of the available crystallographic and phylogenetic covariation data.

As more crystallographic and phylogenetic data (from natural as well as artificial phylogenies) become available, one can anticipate completing Isostericity Matrices for each of the pairing geometries identified (including separate matrices for wobble, bifurcated, and open pairings), and any new geometries that emerge. One matrix suffices to represent all the pairings belonging to a given geometric group when these are, in fact, all isosteric to each other. This appears to be the case for *trans* Hoogsteen/Sh.G. pairings and for *cis* and *trans* Sh.G./Sh.G. pairings. For example, the matrix in Table 17 may be proposed for *trans* Hoogsteen/Sh.G. ('sheared') pairings.

In other cases, two or more matrices may be required. This is the case for the *trans* W.–C./Hoogsteen geometry, which requires a separate matrix for purine/purine pairs and for pyrimidine/ purine or pyrimidine/ pyrimidine pairs,

Table 16. *Isostericity Matrix for* cis *bifurcated pairings*

	A	C	G	U
A	I			
C	I			
G			I	
U			I	

Table 17. *Isostericity Matrix for* trans *Hoogsteen/Sh.G.* (‘*sheared*’) *pairings*

Hoog./Sh.G.	A	C	G	U
A	I	I	I	I
C		I		I
G				
U				

Table 18. *Isostericity Matrix for* trans *R•R W.–C./Hoogsteen pairings. The A•A and A•G pairings are exactly isosteric* (*I*), *whereas G•G is compatible with the other pairings* (*C*), *in certain contexts, although not isosteric*

W.–C./Hoog.	A	G
A	I	I
G		C

Table 19. *Isostericity matrix for* trans *W.–C./Hoogsteen Y•R and Y•Y pairings. A•C is intermediate in structure, and can substitute for either A•U, or for C•G and U•U*

Hoog./W.–C.	C	U
A	I1,I2	I1
C		
G	I2	
U		I2

as shown in Tables 18 and 19. The *trans* W.–C./Hoogsteen pyrimidine•purine and pyrimidine•pyrimidine pairings may be combined into one matrix (Table 19). I1 and I2 in Table 19 indicate two subfamilies of pairings related by a lateral shift

in H-bonding partners, as described in Table 4*b*. As indicated by this matrix, *trans* W.–C./Hoogsteen C•A is potentially isosteric with either U•A or with C•G and U•U. In fact, C•A covaries with U•A in 5S rRNA loop E, whereas C•G does not.

Nonetheless, the Isostericity Matrix approach should not be adhered to in a rigid and strict fashion. Base-pairing interactions involve non-covalent, primarily electrostatic forces, and as such, allow a flexibility of interaction within each geometric group, as has been repeatedly noted in this review. Therefore, bases can substitute for each other in homologous molecules that do not result in exactly isosteric pairs, as long as C1′–C1′ distances are within appropriate ranges. Furthermore, pairs that belong to different geometrical classifications may substitute for each other, as has been noted in the case of *cis*, *syn* W.–C./Hoogsteen A(+)•G substituting for canonical W.–C. basepairs, and also in the case of water inserted U•C substituting for *cis* W.–C./W.–C. A•G. Other examples will no doubt be found.

Moreover, non-isosteric basepairs may functionally substitute for each other to mediate a conserved RNA–RNA (or by extension RNA-protein) interaction. This was noted for shallow-groove interactions, in which *trans* W.–C./Hoogsteen U•A substitutes for canonical C=G or G=C by virtue of the essentially identical arrays of H-bond donors and acceptors that all three pairs present in the shallow groove. This example also shows that for these (and perhaps other) interactions, one base may interact with a pair of bases, which should thus be considered a single unit. The matrices that represent the possible covariations will have to be modified accordingly.

At the geometrical level of analysis, one should, thus, think flexibly about any system intended to organize basepairing structures into families of isosteric pairings for the purpose of predicting tertiary interactions. Other factors will have to be better understood, especially the role of stacking forces and hydration in modulating base–base interactions. In the meantime, we anticipate that a continuous process of proposing and revising classifications, always in a context-dependent manner, will be required. Context will prove crucial in successfully extending covariation analysis to predicting tertiary interactions because of the small number of basepair permutations and the large number of distinct geometrical modes of interaction. Furthermore, all possible covariations compatible with a given geometry will not necessarily occur in a given context, even when a large database of homologous RNA molecules is available. This is due to the influence of other factors and constraints, including the need to maintain interactions with other molecules, and the effects of other physico-chemical factors, such as base–base stacking interactions and hydration that also affect the base-pair interaction energy, but that we still understand incompletely.

8. CONCLUSIONS

From the present analysis of available high-resolution X-ray structures performed in the light of and with the concepts of sequence comparison analysis, the following points emerge.

(1) Eleven out of the twelve possible edge-to-edge, base–base interaction geometries have been observed in addition to intermediate geometries (i.e. bifurcated, wobble, and water-inserted). *Cis* Hoogsteen/Hoogsteen has not been observed to date.

(2) Two or more examples of each of the six possible pairings involving the shallow groove have been observed, allowing other isosteric pairs to be predicted and identified.

(3) In addition to canonical W.–C. pairings, certain other pairings, such as the *cis* bifurcated, the *cis* and *trans* W.–C./Hoogsteen, the cis and *trans* Sh.G./Sh.G., the *trans* Hoogsteen/Sh.G., and the *cis* W.–C./Sh.G. are highly recurrent and modular. They are observed in different contexts mediating a variety of interactions.

(4) Covariation of G•G and A•A pairs with G•U and A•C pairs can indicate bifurcated geometries which may occur in *trans* or *cis* orientations of the bases depending on local strand orientation. Neither *trans* nor *cis* bifurcated pairs are self-isosteric.

(5) Water inserted pairings may be more common than previously recognized, occurring, for example, adjacent to junctions, as observed for the A15.1•U16.1 pair in the hammerhead ribozyme.

(6) Covariation of A•N with C•U and C•C pairings suggests the *trans* Hoogsteen/Sh.G. ('sheared') pairing geometry.

(7) Covariation of A•G with G•A, and with U•C and C•U may indicate self-isosteric *cis* W.–C./W.–C. (A•G and G•A), and water-inserted geometries (U•C and C•U).

(8) *Cis* (or *trans*) W.–C./Hoogsteen R•R pairings are not isosteric with the corresponding *cis* (or *trans*) Y•R and Y•Y pairings. *Cis* and *trans* W.–C./Hoogsteen R•R pairings have distinctive covariation signatures.

(9) *Cis* wobble U•U, C(+)•C, and U•C(+), but not C•U, may substitute quasi-isosterically for wobble G•U or A(+)•C in certain contexts.

(10) Tandem motifs may indicate wobble or sheared (Hoogsteen/Sh.G.) geometries, stabilized by cross-strand stacking.

(11) Bifurcated and sheared pairs often serve to position one of the interacting bases to interact with a third base, or with the backbone atoms of a third strand, or even with non-RNA ligands.

The covariation rules described here were derived on the assumption that molecular evolution of RNA motifs samples sequence space so as to maintain homologous 3D structures. The assumption was shown to be meaningful since some of their covariation rules were successfully used for identifying unsuspected occurrences of the same RNA motifs in various RNA molecules (Costa & Michel,

1995; Leontis & Westhof, 1998*b*). It is now the hope that the Isostericity Matrices define molecular signatures for a given non-canonical base pair independently of the 3D motif in which the pair occurs.

9. ACKNOWLEDGEMENTS

We are grateful to Christian Massire for the use of the COSEQ program and to François Michel and Pascal Auffinger for numerous discussions and comments on the manuscript. Grant support to N.B.L. from NIH grant 1R15-GM/OD55898–01 is acknowledged. N.B.L. thanks Ikechukwuk Oguejiofor for assistance with figure preparation, Valentin Petrov for assistance with tRNA sequence analysis, and Vassiliki Leontis for editing assistance.

10. REFERENCES

Auffinger, P. & Westhof, E. (1998). Hydration of RNA base pairs. *Journal of Biomolecular Structure & Dynamics* **16**, 693–707.

Auffinger, P. & Westhof, E. (In press). Singly and bifuracted hydrogen bonded basepairs in tRNA anticodon hairpins and ribozymes. *J. Mol. Biol.*

Auffinger, P., Louise-May, S. & Westhof, E. (1996). Hydration of C-H groups in tRNA. *Faraday Discuss.* **103**, 151–173.

Auffinger, P., Louise-May, S. & Westhof, E. (1999). Molecular dynamics simulation of solvated yeast tRNA(Asp). *Biophys. J.* **76**, 50–64.

Baeyens, K. J., Debondt, H. L., Pardi, A. & Holbrook, S. R. (1996). A curved RNA helix incorporating an internal loop with G•A and A•A non-W.–C. base pairing. *Proc. natl. Acad. Sci. U.S.A.* **93**, 12851–12855.

Bartel, D. P., Zapp, M. L., Green, M. R. & Szostak, J. W. (1991). HIV-1 Rev regulation involves recognition of non-W.–C. base pairs in viral RNA. *Cell* **67**, 529–536.

Basavappa, R. & Sigler, P. B. (1991). The 3 Å crystal structure of yeast initiator tRNA: functional implications in initiator/elongator discrimination. *Embo J.* **10**, 3105–11.

Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D. & Williamson, J. R. (1996). α-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.

Betzel, C., Lorenz, S., FÅrste, J. P., Bald, R., Zhang, M., Schneider, T. R., Wilson, K. S. & Erdmann, V. A. (1994). Crystal structure of domain A of *Thermus flavus* 5S rRNA and the contribution of water molecules to its structure. *FEBS Lett.* **351**, 159–164.

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**, 1404–10.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R. & Doudna, J. A. (1996*a*). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678–1684.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Szewczak, A. A., Kundrot, C. E., Cech, T. R. & Doudna, J. A. (1996*b*). RNA tertiary structure mediation by adenosine platforms. *Science* **273**, 1696–1699.

Chiu, D. K. Y. & Kolodziejczak, T. (1991). Inferring consensus structure from nucleic acid sequences. *CABIOS* **7**, 347–352.

Conn, G. L., Draper, D. E., Lattman, E. E. & Gittis, A. G. (1999). Crystal structure of a conserved ribosomal protein–RNA complex. *Science* **284**, 1171–1174.

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.

Correll, C. C., Munishkin, A., Chan, Y. L., Ren, Z., Wool, I. G. Steitz, T. A. (1998). Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *Proc. natl. Acad. Sci. U.S.A.* **95**, 13436–41.

Costa, M. & Michel, F. (1995). Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* **14**, 1276–85.

Costa, M. & Michel, F. (1997). Rules for RNA recognition of GNRA tetraloops deduced by *in vitro* selection: comparison with *in vivo* evolution. *EMBO J.* **16**, 3289–3302.

Ferre-D'Amare, A. R., Zhou, K. & Doudna, J. A. (1998). Crystal structure of a hepatitis delta virus ribozyme. *Nature* **395**, 567–74.

Gautheret, D. & Gutell, R. R. (1997). Inferring the conformation of RNA base pairs and triples from patterns of sequence variation. *Nuc. Acids Res.* **25**, 1559–1564.

Gautheret, D., Konings, D. & Gutell, R. R. (1995). G•U base pairing motifs in ribosomal RNA. *RNA* **1**, 807–14.

Golden, B. L., Gooding, A. R., Podell, E. R. & Cech, T. R. (1998). A preorganized active site in the crystal structure of the Tetrahymena ribozyme. *Science* **282**, 259–264.

Jaeger, L., Michel, F. & Westhof, E. (1994). Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. *J. Mol. Biol.* **236**, 1271–1276.

Jaeger, J., Restle, T. & Steitz, T. A. (1998). The structure of HIV-1 reverse transcriptase complexed with an RNA pseudoknot inhibitor. *Embo J.* **17**, 4535–42.

James, B. D., Olsen, G. J. & Pace, N. R. (1989). Phylogenetic comparative analysis of RNA secondary structure. *Meth. Enz.* **180**, 227–239.

Jang, S. B., Hung, L. W., Chi, Y. I., Holbrook, E. L., Carter, R. J. & Holbrook, S. R. (1998). Structure of an RNA internal loop consisting of tandem C-A+ base pairs. *Biochemistry* **37**, 11726–31.

Kuryavyi, V. V. & Jovin, T. M. (1995). Triad-DNA: a model for trinucleotide repeats [letter]. *Nat. Genet.* **9**, 339–41.

Leontis, N. B. & Westhof, E. (1998*a*). The 5S rRNA loop E: chemical probing and phylogenetic data versus crystal structure. *RNA* **4**, 1134–1153.

Leontis, N. B. & Westhof, E. (1998*b*). A common motif organizes the structure of multi-helix loops in 16S and 23S ribosomal RNAs. *J. mol. Biol.* **283**, 571–583.

Leontis, N. B. & Westhof, E. (1999). Recurrent RNA motifs: analysis at the basepair level. In J. Barciszewki & B. F. C. Clark (eds), *RNA Biochemistry and Biotechnology*. Kluwer Academic Publishers, Boston.

Levitt, M. (1969). Detailed molecular model for transfer Ribonucleic Acid. *Nature* **224**, 759–763.

Lietzke, S., 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.

Limmer, S. (1997). Mismatch base pairs in RNA. *Prog. Nucl. Acid Res. and Mol. Biol.* **57**, 1–39.

Lu, S., Chen, L., Egli, M., Berger, J. M. & Rich, A. (1999). A minor groove RNA triplex in the crystal structure of a viral pseudoknot involved in ribosomal frameshifting. *Nat. Struct. Biol.* **6**, 2024–2029.

Masquida, B., Sauter, C. & Westhof, E. (In press). A sulfate pocket formed by three G◦U pairs in the 0.97 Å resolution X-ray structure of a nonameric RNA. *RNA*.

Michel, F. & Costa, M. (1996). Inferring RNA structure by phylogenetic and genetic analysis. In R. W. Simons & M. Grunberg-Manago (eds), *RNA Structure and Function*. Cold Spring Harbor Laboratory Press.

Michel, F. & Westhof, E. (1990). Modelling of the three-dimensional architecture of Group I catalytic introns based on comparative sequence analysis. *J. molec Biol.* **216**, 585–610.

Nissen, P., Kjeldgaard, M., Thirup, S., Polekhina, G., Reshetnikova, L., Clark, B. F. C. & Nyborg, J. (1995). Crystal structure of the ternary complex of phe-tRNAphe, EF-Tu and a GTP analog. *Science* **270**, 1464–72.

Nissen, P., Thirup, S., Kjeldgaard, M. & Nyborg, J. (1999). Crystal structure of Cys-tRNA:EF-Tu:GDPNP reveals general and specific features in the ternary complex and in tRNA. *Structure* **7**, 143–156.

Pan, B., Mitra, S. N. & Sundaralingam, M. (1998). Structures of a 16-mer RNA duplex r(GCAGACUUAAAUCUGC)2 with Wobble C•A+ mismatches. *J. molec. Biol.* **283**, 977–984.

Pan, B., Mitra, S. N. & Sundaralingam, M. (1999). Crystal structure of an RNA 16-mer duplex r(GCAGAGUUAAAUCUGC)2 with non-adjacent G(syn)•A+ (anti) mispairs. *Biochemistry* **38**, 2826–2831.

Pley, H. W., Flaherty, K. M. & McKay, D. B. (1994). Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**, 68–74.

Price, S. R., Evans, P. R. & Nagai, K. (1998). Crystal structure of the spliceosomal U2B″-U2A′ protein complex bound to a fragment of U2 small nuclear RNA. *Nature* **394**, 645–50.

Rath, V. L., Silvian, L. F., Beijer, B., Sproat, B. S. & Steitz, T. A. (1998). How glutaminyl-tRNA synthetase selects glutamine. *Structure* **6**, 439–49.

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

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.

Seeman, N. C., Rosenburg, J. M. & Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci.* (*USA*) **73**, 804–808.

Sprinzl, M., Steegborn, C., Håbel, F. & Steinberg, S. (1998). Compilation of tRNA sequences and sequences of tRNA genes. *Nucl. Acids Res.* **26**, 68–72.

Szymanski, M., Specht, T., Barciszewska, M. Z., Barciszewski, J. & Erdmann, V. A. (1998). 5S rRNA data bank. *Nucl. Acids Res.* **26**, 156–159.

Tanaka, Y., Fujii, S., Hiroaki, H., Sakata, T., Tanaka, T., Uesugi, S., Tomita, K. & Kyogoku, Y. (1999). A′-form RNA double helix in the single crystal structure of r(UGAGCUUCGGCUC) *Nucl. Acids Res.* **27**, 949–55.

Tinoco, I., Jr. (1993). Structures of base pairs involving at least two hydrogen bonds. In R. F. Gesteland & J. F. Atkins (eds), *The RNA World*, pp. 603–607. Cold Spring Harbor Press, Plainview, N.Y.

Trikha, J., Filman, D. J. & Hogle, J. M. (1999). Crystal structure of a 14 bp RNA duplex with non-symmetrical tandem G.U wobble base pairs. *Nucl. Acids Res.* **27**, 1728–1739.

Wahl, M. C. & Sundaralingam, M. (1997). C-H···O hydrogen bonding in biology. *TIBS* **22**, 97–102.

Wahl, M. C., Rao, S. T. & Sundaralingam, M. (1996). The structure of r(UUCGCG) has a 5′-UU- overhang exhibiting Hoogsteen-like *trans* U-U base-pairs. *Nature, Struct. Biol.* **3**, 24–31.

Westhof, E. (1992). Westhof's rule. *Nature.* **358**, 459–460.

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

Westhof, E., Romby, P., Romaniuk, P. J., Ebel, J. P., Ehresmann, C. & Ehresmann, B. (1989). Computer modelling from solution data of spinach chloroplast and of *Xenopus laevis* somatic and oocyte 5S rRNAs. *J. Mol. Biol.* **207**, 417–431.

Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R. E. (1980). Crystal structure analysis of a complete turn of B-DNA. *Nature* **287**, 755–758.

Wu, M., McDowell, J. A. & Turner, D. H. (1995). A periodic table of symmetric tandem mismatches in RNA. *Biochemistry* **34**, 3204–3211.

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.