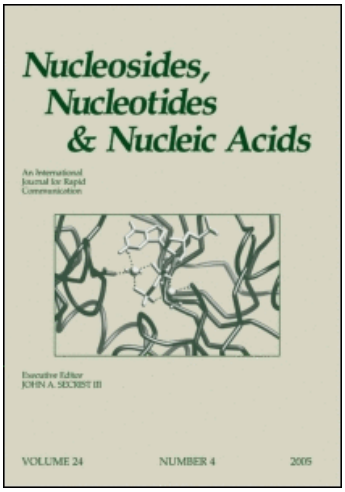

Nucleosides, Nucleotides and Nucleic Acids
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597286

Structures of Non-canonical Tandem Base Pairs in RNA Helices: Review

Hans A. Heus[a]; Cornelis W. Hilbers[b]
[a] NSR Center, Laboratory of Biophysical Chemistry, University of Nijmegen, Toernooiveld, Nijmegen, The Netherlands [b] NSR Center for Molecular Structure, Design and Synthesis, Laboratory of Biophysical Chemistry, University of Nijmegen, Toernooiveld, Nijmegen, The Netherlands

Online publication date: 09 August 2003

To cite this Article Heus, Hans A. and Hilbers, Cornelis W.(2003) 'Structures of Non-canonical Tandem Base Pairs in RNA Helices: Review', Nucleosides, Nucleotides and Nucleic Acids, 22: 5, 559 — 571
To link to this Article: DOI: 10.1081/NCN-120021955
URL: http://dx.doi.org/10.1081/NCN-120021955

NUCLEOSIDES, NUCLEOTIDES & NUCLEIC ACIDS
Vol. 22, Nos. 5–8, pp. 559–571, 2003

Structures of Non-canonical Tandem Base Pairs in RNA Helices: Review

Hans A. Heus* and Cornelis W. Hilbers

NSR Center for Molecular Structure, Design and Synthesis, Laboratory of Biophysical Chemistry, University of Nijmegen, Toernooiveld, Nijmegen, The Netherlands

ABSTRACT

The structures of tandem non-canonical base pairs, a frequently recurring motif in RNA molecules, are reviewed and analysed. The tandem non-canonical base pair motifs can be roughly divided in three groups, containing seven subgroups based on their base pairing patterns and local geometries. Structural details and helical parameters that can be used to numerically distinguish between the subgroups are tabulated. Remarkably, while the individual helical twists of the tandem and adjacent base pair steps can be substantially smaller or larger than the typical A-form value of 32.7°, the average value is close to A-form. This and other striking regularities resulting from compensating geometrical adjustments, important for understanding and predicting the configurations of non-canonical base pairs geometries are discussed.

Key Words: RNA; Structure; RNA geometry; Non-canonical tandem base pairs.

*Correspondence: Hans A. Heus, NSR Center, Laboratory of Biophysical Chemistry, University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands; Fax: +31-24-3652112; E-mail: hans@nmr.kun.nl.

DOI: 10.1081/NCN-120021955
1525-7770 (Print); 1532-2335 (Online)

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

INTRODUCTION

Unlike DNA, which is almost exclusively encountered as a double helix with canonical Watson-Crick base pairs, RNA contains many non-canonical base pairs. Formation of non-canonical base pairs in RNA is essential, because – in contrast to Watson-Crick base pairs in RNA – they often play an important role, either functionally, e.g., in the mechanism of catalytic RNA or structurally, e.g., in the stabilisation and formation of the RNA tertiary structure. Understanding the forces that dictate the stabilities and shapes of non-canonical base pairs in RNA molecules is therefore of prime importance.

A primary source of information for studying the effect of incorporating non-canonical base pairs in an RNA molecule lies in its structure determination, which enables us to visually examine the molecules and to extract important structural details. Evaluation of these can be used to formulate the principles that drive the molecules into their specific shapes to carry out their function.

Recently, the structure of a large number of large complex RNA molecules have been solved either in the free state or in complex with proteins, culminating in the high-resolution structures of the ribosome. Despite these major achievements the rules that govern the folding of an RNA molecule are still far from understood. A complicating factor is that the structures of non-canonical base pairs can be polymorphic, i.e., that their base pair configuration is dependent on the structure of nearby elements, e.g., adjacent base pairs or loops, or by a third interaction such as in a base triple. The current situation is that, despite the vast amount of structural data that is embodied within the structures of the ribosome or other large RNAs, many motifs have not yet been solved and it is still hard to predict the structure of a simple RNA motif, such as a helical element containing a non-canonical base pair.

One rational to predict the conformation of a non-canonical base pairs comes from geometric considerations, taking for instance the local helical shape or cross-strand P-O3′ distances into account. This has worked reasonably well for single base pairs, e.g., for the isosteric G-U and A^+-C base pairs, and recently isosteric matrices for base pairs have been itemised and discussed by Leontis and Westhof.[15–17]

In the order of increasing RNA complexity, after the single non-canonical base pairs the next motifs in line are the tandem non-canonical base pairs. There are 12 non-canonical base pairs, excluding polymorphic forms, hence there are also 12 so-called symmetric non-canonical tandem base pairs, i.e., of the type 5′XY3′/3′YX5′, where X-Y represents a base pair, and 66 so-called asymmetric non-canonical tandem base pairs, i.e., of the type 5′XW3′/3′YZ5′, where X-Y and W-Z both represent a base pair. Out of these, the symmetric UG and GA tandems are the most abundant and functional importance has been demonstrated for the GA/AG tandem.[23] Because of their abundance and the functional importance of single G-U and G-A base pairs, these motifs attracted attention early on and their occurrence, phylogeny and structural aspects have been reviewed and discussed before.[6,7,20,28] Also thermodynamic stabilities of single as well as symmetric and asymmetric tandems have been extensively investigated and discussed by Turner and co-workers.[12,29,32]

Recently, we described for the first time the structure of a helical element of polio viral RNA containing a 5′CU3′/3′UU5′ asymmetric tandem base pair motif.[18]

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

The structure of the non-symmetric tandem base pair motif revealed a pseudodyad character, which became apparent from an analysis of the helical twists and the virtual bond angle λ between the C1′-C1′ line and the glycosidic C1′-N1/N9 bonds. These structural details explained the configuration of the pyrimidine base pairs in the reported structure and related structures of RNA double helices containing tandem pyrimidine-pyrimidine mismatches. These results prompted us to re-evaluate structural aspects of tandem non-canonical base pair motifs observed in RNA molecules.

RESULTS AND DISCUSSION

For this study we restricted ourselves to tandem non-canonical base pairs flanked by at least one Watson-Crick base pair at either side of the tandem. Interestingly, these are underrepresented in the existing database; tandems flanked on one side by one or more other non-canonical base pairs, bulges or loops are more abundant. However, the structures of these latter motifs are more diverse and complicated, and need a more thorough analysis that goes beyond the present discussion. We also included relevant structures determined in solution by NMR-spectroscopy. Although these are occasionally excluded (e.g., [15,17]) with the argument that the crystal structures might be more reliable many RNA motifs were discovered by NMR spectroscopy before a high-resolution crystal structure was published, e.g., in case of the GNRA loop[8] and the tandem sheared GA/AG base pair motif.[25] The structural sources of the analysed motifs are listed in Table 1.

CLASSIFICATION BY BASE PAIR CONFIGURATION

The tandems can be roughly divided in 3 groups (vide infra) based on their base pair configurations (Table 2, Fig. 1). Not surprisingly, the number of solved structures of different tandem motifs roughly follows their frequency in nature, i.e., GU/UG $\gg$ GG/UU > GA/AG > AA/AG > AA/GA > GU/UG $\gg$ UU/UU > GG/AA. Therefore, the main groups designated Groups I and II contain G-A and G-U base pairs. A notable exception is the AG/GA tandem, which hardly occurs in the secondary structure models of natural RNAs, but has been investigated frequently because of its stability.[29]

Group I can be divided in two subgroups, containing either so-called sheared or imino-proton G-A base pairs (Fig. 1A,B). In the sheared G-A pair the guanosine N2H and N3 are hydrogen bonded to the adenosine N7 and N6H, respectively. In the imino proton base paired configuration, which is more frequently found as a loop closing base pair or as a single pair in helices,[6,26] the guanosine N1H and O6 are hydrogen bonded to the adenosine N1 and N6H, respectively. For the G-A base pair tandem it was found early on that its configuration is dependent on the identity of the flanking Watson-Crick base pairs. Pyrimidine-GA-purine duplexes contain sheared G-A base pairs, while purine-GA-pyrimidine duplexes contain the imino proton hydrogen bonded configuration. The AG/GA tandem is always in the imino

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Table 1. Structures analyzed in this study.

Motif[a]	Method	pdb-file	Remarks	Reference
CGAG	NMR	1FYV	8 bp oligoduplex	[25]
^{2873}CGAG[b]	X-ray	1FFK	23S rRNA	[1]
UGAA	NMR	–	12 bp oligoduplex	[9]
GGAC	NMR	1MIS	8 bp oligoduplex	[30]
CAGG	NMR	1MWG	8 bp oligoduplex	[31]
AUGU	NMR	1QET	8 bp oligoduplex	[22]
AUGU	X-ray	315D	8 bp oligoduplex	[2]
GUGC	NMR	1EKA	8 bp-oligoduplex	[5]
^{1535}CUGG	X-ray	1FFK	23S rRNA	[1]
^{81}CUGG	X-ray	1FFK	5S rRNA	[1]
^{817}CUGG	X-ray	1FJG	16S rRNA	[33]
^{650}CUGC/GGUG^{752}	X-ray	1FFK	23S rRNA	[1]
^{79}GUGU/CGUA^{94}	X-ray	364D	5S ribosomal RNA	[32,33]
CCA^{+}G	X-ray	402D	8 bp oligoduplex	[11]
GGUC	NMR	1GUC	8 bp oligoduplex	[21]
AGUU	NMR	1QES	8 bp oligoduplex	[22]
CGUG	NMR	1EKD	8 bp oligoduplex	[5]
UGUA	X-ray	332D	6 bp oligoduplex	[3]
^{118}GGUC	X-ray	1GID	Tetrahymena ribozyme	[4]
^{138}GA^{+}CC	X-ray	1MFQ	Signal recognition particle	[13]
^{146}AGGG/UUUC^{154}	X-ray	1GID	Tetrahymena ribozyme	[4]
CGGU/AUUG	X-ray	433D	12 bp oligoduplex	[27]
^{543}GGGC/CUUG^{610}	X-ray	1FFK	23S rRNA	[1]
CUUG	X-ray	280D	12 bp oligoduplex	[19]
CCUC/GUUG	NMR	1N66	Poliovirus 3′-NTR Y-domain	[18]
^{287}CAGC/GCAG^{362}	X-ray	1FFK	23S rRNA	[1]
^{131}AUAU/UGGA^{163}	X-ray	1MFQ	Signal recognition particle	[13]

[a]Single-stranded sequences refer to self-complementary sequences from 5′ to 3′, e.g., CGAG refers to 5′-CGAG-3′/3′-GAGC-5′. In asymmetric cases the complete secondary structure of the duplex is given, with the complementary strand in the 3′-5′ direction.
[b]In case of large structures that contain more than one motif, the numbers added in superscript refer to the numbers of the 5′-residues in the published secondary structure model.

proton hydrogen bonded configuration. Out of these, the structures listed in Tables 1 and 2 have been actually determined. The structure of an important member of this group, the AGAU duplex, needed to support this rule is lacking.

Group II can be divided in three subgroups, in which the G-U base pair is found in only one configuration, the well known wobble base pair with the guanosine N1H and O6 hydrogen bonded to the uridine O2 and N3H, respectively (Fig. 1C). The subdivision of Group II is based on sequence, which results in different structural properties (*vide infra*) of the UG/GU, GU/UG and GG/UU motifs. This separation in three classes has been proposed before by Gautheret et al. (1995).[7] Tandems containing A^{+}-C base pairs (Fig. 1D), isosteric to the G-U wobble belong to group II by default.

Group III contains the remainder of tandems: the pyrimidine tandems and the asymmetric tandems containing two different non-canonical base pairs. The number

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Table 2. Structural characteristics of tandem non-canonical base pair motifs.

Motif[1]	Source	Base pair Segment[2]	Helical twists[1,3] N- N1-N2-N' | | | | N'-N4-N3-N			Cross-strand 5'P-O3' distances (Å) 5'P-N1-O3'-5'P-N2-O3' | | O3'-N4-P5'-O3'-N3-P5'				λ_{N1} λ_{N4}	λ_{N2} λ_{N3}
A-form RNA												
NCAN′/N′UGN	[24]	–	32.7	32.7	32.7	17.5	17.5	17.5	17.5	55.7 54.4	56.2 57.4
Ia. Sheared G-A base pairs motif												
CGAG	1FYV	8 bp	18	81	18	16.4	11.8	11.8	16.4	97 7	7 97
^{2873}CGAG	1FFK	7 bp	13	70	3	16.9	13.2	12.9	17.2	97 6	12 107
UGAA	–	12 bp	15	72	12	17.6	12.6	12.6	17.6	90 15	19 97
Ib. Imino proton G-A base pairs motif												
GGAC	1MIS	8 bp	36	21	36	18.1	20.4	20.4	18.1	39 52	52 39
CAGG	1MWG	8 bp	32	27	32	18.4	19.4	19.4	18.4	48 47	47 47
IIa. G-U wobble base pairs motif I												
AUGU	1QET	8 bp	32	39	32	17.8	17.5	17.5	17.8	70 48	48 70
AUGU	315D	8 bp	27	38	26	17.6	16.9	17.5	17.4	67 41	43 70
GUGC	1EKA	8 bp	25	39	25	17.8	16.7	16.6	17.9	71 47	46 70
^{1535}CUGG	1FFK	8 bp	27	36	25	17.2	17.2	17.1	17.2	68 45	40 66
^{81}CUGG	1FFK	6 bp	20	51	32	17.9	16.6	16.2	17.1	71 45	42 67
^{817}CUGG	1FJG	5 bp	32	34	30	17.3	17.6	17.7	17.6	65 43	45 66
^{650}CUGC/GGUG^{752}	1FFK	13 bp	26	39	27	17.9	17.5	17.7	17.3	66 42	41 67
^{79}GUGU/CGUA^{94}	364D	8 bp	26	42	23	16.8	16.6	16.0	17.4	62 26	41 79
CCA^{+}G	402D	8 bp	26	45	24	17.8	16.8	16.9	17.3	83 38	39 79
IIb. G-U wobble base pairs motif II												
GGUC	1GUC	8 bp	37	30	37	17.0	17.8	17.8	16.9	42 60	60 42
AGUU	1QES	8 bp	40	23	39	17.0	17.6	17.7	17.2	47 72	73 45
CGUG	1EKD	8 bp	38	17	38	16.8	17.5	17.5	16.8	58 87	87 58

(*continued*)

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Table 2. Continued.

Motif[1]	Source	Base pair Segment[2]	Helical twists[1,3] N- N1-N2-N' | | | | N'-N4-N3-N			Cross-strand 5'P-O3' distances (Å) 5'P-N1-O3'-5'P-N2-O3' | | O3'-N4-P5'-O3'-N3-P5'				λ_{N1} λ_{N4}	λ_{N2} λ_{N3}
UGUA	332D	6 bp	31	25	34	18.1	17.4	17.4	17.8	46	66
										68	42
^{118}GGUC	1GID	6 bp	39	23	46	16.7	17.1	17.3	15.9	43	70
										67	44
^{138}GACC/CCAG^{146}	1MFQ	8 bp	32	28	31	17.4	17.3	17.5	17.1	59	72
										60	56
IIc. G-U wobble base pairs motif III											
^{146}AGGG/UUUC^{154}	1GID	8 bp	40	37	25	16.6	16.9	16.4	17.4	44	43
										69	71
CGGU/GUUA	433D	8 bp	35	35	28	17.9	17.1	16.5	17.4	43	44
										67	73
^{543}GGGC/CUUG^{610}	1FFK	10 bp	39	36	27	17.5	17.4	17.4	17.6	39	42
										67	69
IIIa. Pyrimidine tandems											
CUUG	280D	12 bp	25	57	28	16.1	14.4	14.6	16.0	80	50
										50	80
CCUC/GUUG	1N66	9 bp	17	59	18	16.0	14.0	13.9	15.8	93	48
										55	93
IIIb. GA/GU mixed tandems											
^{287}CAGC/GCAG^{362}	1FFK	4 bp	30	31	29	17.3	17.5	17.8	17.5	56	59
										56	67
^{131}AUAU/UGGA^{163}	1MFQ	8 bp	29	38	26	17.6	18.3	18.3	17.6	60	42
										49	67

[1]N-N′ refers to any Watson-Crick base pair.
[2]The stretch of consecutive Watson-Crick base pairs and the non-canonical tandem.
[3]N1-N4 refers to the first 5′-N1-N4-3′ non-canonical base pair of the tandem motif, N2-N3 to the second non-canonical base pair, 3′-N2-N3-5′.

of examples for which the structure has been determined is surprisingly small, since about ~40% of all naturally occurring tandems belong to this class. The pyrimidine tandems, CUUG and CCUG/GUUC determined for this group contain U-U and C-U base pairs (Fig. 1E,F). In the U-U base pair the N3H and O2 of the 5′-uridine are hydrogen bonded to the O4 and N3H of the opposite 3′-uridine. In the C-U base pair the cytidine O2, and a potentially protonated N^+3H are hydrogen bonded to the uridine N3H and O4, respectively. The other two members of this class contain asymmetric tandems with mixed G-A and G-U/A^+-C base pairs. This extremely low number is again surprising, but also disappointing, because some general features that emerge from an evaluation of the symmetric tandems (*vide infra*) can not be compared with the asymmetric tandems.

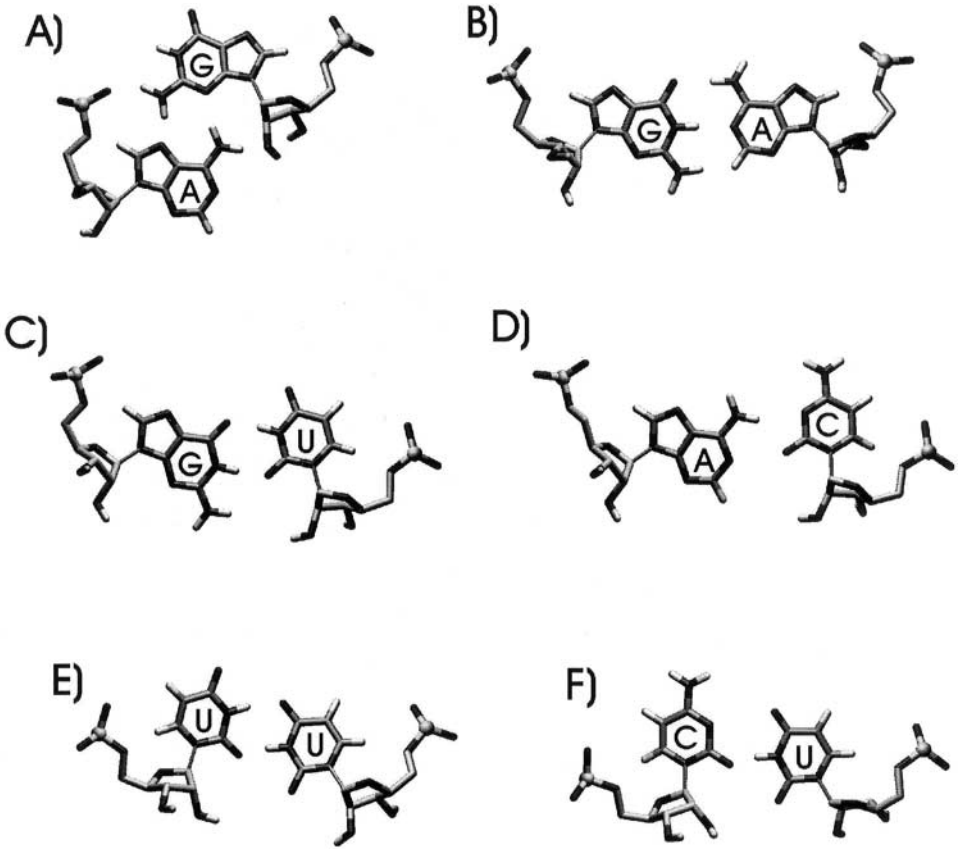

Figure 1. Structures of the non-canonical base pairs configurations observed in the different subgroups of tandems. A). Group Ia: Sheared G-A base pair (1FYV). B). Group Ib: Imino G-A base pair (1MWG). C). Group II: G-U wobble base pair (315D). D). Group II: A + C base pair (402D). E). Group IIIa: U-U base pair (280D). F). Group IIIa: C-U base pair (1N66).

STRUCTURAL DETAILS OF TANDEM BASE PAIR MOTIFS

Table 2 lists the structural details of the tandems, relevant for assessing the local helical geometry. To be more specific, we calculated the helical twists of the three base pair steps of the four base pair element containing the tandem and the virtual bond angle, λ, between the C1′-C1′ line and the nucleotide glycosidic C1′-N1/N9 bond. The λ angle is a useful measure to recognise isosteric base pairs. We also measured the cross-strand 5′P-O3′ distances adjacent to the individual base pairs, which is a useful indicator for judging how well a non-canonical base pair fits into the A-form helix.

In a regular A-form RNA helix the helical twist amounts to 32.7° and the cross-strand 5′P-O3′ distances are ∼17.5 Å. For Watson-Crick base pairs the λ angles are ∼55–57°. The similar λ angles, together with the nearly equal C1′-C1′ distances make the Watson-Crick base pairs isosteric and creates a pseudodyad axis in both of them,

which is the origin of the overall regular shape of the A-form helix with sequentially stacked bases. For most of the non-canonical tandems the λ angles of the individual base pairs are dissimilar and hence lack the pseudodyad axis, leading to different stacking patterns.

The conformational features of the sheared GA/AG (Group Ia) and pyrimidine tandems (Group IIIa) deviate the most from those of an A-form helix, be it in a similar manner (Table 2). For these tandems the twist angle at either side of the tandem is very small, while the helical twist at the central base pair step is very large. The cross-stand 5′P-O3′ distances are close to normal between the Watson-Crick and non-canonical base pairs, but very small at the central base pair step. This is a consequence of the short cross-strand C1′-C1′ distances resulting from shearing of the G-A base pair or the more limited dimensions of the pyrimidine base pairs. To orient the bases for proper pairing, the λ angles are large for the 5′-guanosines and pyrimidines and small for the 3′-adenosines, but normal for the 3′-pyrimidines. The combination of the altered twist and λ angles results in cross-strand stacking of the guanosines and adenines of the tandem G-A base pairs, as well as the 3′-residues in the pyrimidine tandem base pairs, but unstacking of the 5′-residues in the latter (Fig. 2). The small helical twists on both sides of the tandem results in enhanced sequential stacking at these sides.

As noted earlier for the sheared tandem GA base pair motif,[6] a single non-canonical base pair with a short C1′-C1′ cross-strand distance does not fit into an A-form helix because the 5′P-O3′ distances on either side of the base pair can differ by more than 4 Å. However, in a tandem motif the two non-canonical base pairs perfectly fit into an A-form helix because the asymmetry of the first base pair is compensated by the second. Hence the large 5′P-O3′ distances face the regular helix, while the two short 5′P-O3′ distances face each other (Table 2).

Due to the large cross-strand P-O3′ distance of ~20 Å on one side of the bulky imino proton G-A base pair it is more frequently found as a single base pair at the end of helices.[26] For the GGAC/CAGC tandem the λ angles of the 5′-guanosine are close to normal, but those of the 3′-adenines are somewhat smaller, and the helical twist alternates between normal – small – normal leading to regular sequential stacking between the helix and the tandem and enhanced stacking at the central GA/AG base pair step. Interestingly, in the reversed orientation, i.e., in the AG/GA tandem the alternation of the helical twists is not reversed, but remains normal – small – normal with similar stacking properties. This is brought about by a slightly smaller λ angle of all four residues.

The G-U wobble is easier to fit as a single base pair into a helix, because the λ angles are less dissimilar and the cross-strand 5′P-O3′ distance is close to that in the A-form. Nevertheless, in the G-U wobble the uridine always has a somewhat larger λ angle, while for the guanosine it is slightly smaller. The consequences of these differences are visible in the structural details of the three different GU tandem motifs. In the UG/GU tandem the 5′-residue has the largest λ angle and the helical twist alternates from small – large – small. These characteristics are reversed in the GU/UG tandem, where the 3′-residue has the large λ angle and the helical twist alternates from large – small – large. In the asymmetric GG/UU tandems a mixed situation obtains with the large λ angles at one strand of the helix and alternating large – regular – small twists angles.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

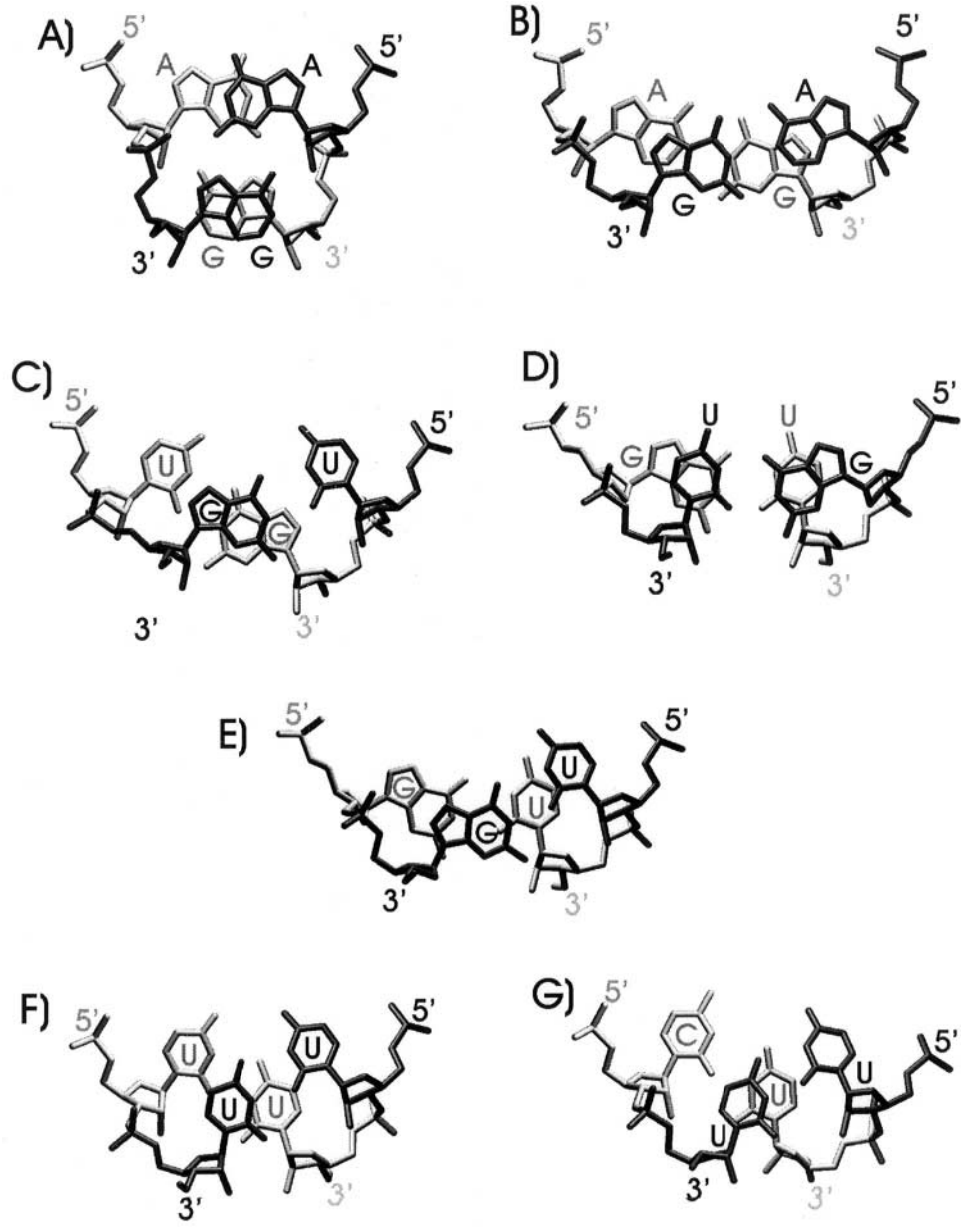

Figure 2. Views of representatives of tandem base pair motifs down the local helical axis showing the specific stacking patterns. Upper base pairs are in dark grey; lower base pairs are in light grey. A) GA/AG tandem (1FYV). B) AG/GA tandem (1MWG). C) UG/GU tandem (315D). D) GU/UG tandem (332D). E). GG/UU tandem (433D). F) UU/UU tandem (280D). G) CU/UU tandem (1N66).

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

In the single G-U wobble the 5′PG-UO3′ distance is ~1Å smaller than the 3′OG-UP5′ distance. Thus, as noted earlier by Leontis and Westhof,[15] one would expect the UG/GU tandem to fit better into the A-form helix, because the short distances face each other and the large distances face the A-form helix. The reported distances in Table 2 indeed follow this trend. Note also, as discussed above, that this trend is more outspoken for the tandem sheared G-A base pair and pyrimidine base pair motifs.

The remaining two mixed tandems, i.e., AG/CA and UA/GG, for which the structure has been determined, contain wobble and imino proton G-A base pairs. One sees that the deviating λ angles of the wobble base pairs are maintained, as well as a large cross-strand distance to accommodate the G-A base pair.

GENERAL FEATURES AND COMPARISON OF TANDEM MOTIFS

From the above descriptions it becomes clear that the helical twist and the λ angles, which are mutually coupled, are useful measures for determining the stacking properties of the tandems. As noted above for the pyrimidine tandems, the large twist angle between the two non-canonical base pairs together with a large λ for the 5′-residue leads to cross-strand stacking of the 3′-residues and unstacking of the 5′-residues (See Fig. 2 and Table 2). Accordingly, in the GU tandems, those with a large twist and large 5′-λ angles, i.e., the UG/GU tandems, show similar stacking properties, albeit less outspoken in concurrence with the more moderate deviations of the helical twists and λ angles from A-form. The opposite occurs in the GU/UG tandems, which exhibit a small helical twist and large 3′-λ, where enhanced sequential stacking is observed.

The value of the λ angles is an easy aid in distinguishing between possible base pair configurations, i.e., the polymorphism of some canonical base pairs, as well. For instance, in a 5′XU3′/3′YU5′ consensus, where X-Y denotes any base combination, there are two possible U-U base pair configurations. One in which the O4 of the 5′-uridine protrudes into the major groove and the O2 participates in a hydrogen bond and, vice versa, one in which the O4 of the opposite 3′-uridine protrudes into the major groove and the O4 of the 5′-uridine accepts a hydrogen bond. The latter configuration is observed in the CUUG and CCUG/GUUC duplexes, which can easily be discerned from the large 5′-λ angle, because this swings the base towards the major groove.

A large 5′-λ angle, swinging the corresponding base into the major groove is also observed in the sheared G-A base pair. Conversely, in this case swinging of the 3′-adenines towards the minor groove leading to cross-strand stacking can easily be discerned from the very small 3′-λ angle.

Leontis and Westhof noticed that in loop E of bacterial 5S rRNA the 3′G-U5′ wobble in a UG/GU tandem covaries with U-U, 3′U-C5′ and C-C, but not with 3′C-U5′[14]. This is understandable from the observed λ angles, because in the G-U and isosteric base pairs U-U and U-C as well as the proposed isosteric C-C base pair[15] these are large (U:~70°) and small (G:~45°). Replacing the 3′-uridine in the

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

U-U base pair by a cytidine puts the uridine O2 and N3H opposite to the uridine N3 and N4H, respectively, which is incompatible with hydrogen bond formation.

Two striking regularities are observed in the RNA helices containing non-canonical tandem base pairs. 1) It appears that large helical twists between the non-canonical base pairs are compensated by small helical twists at the interfaces between the tandem and flanking A-form helices, or vice versa, leading to an overall twist of ~33°, i.e., close to A-form. 2) It is also striking that a large helical twist at the central base pair step is coupled with a short cross-strand 5′P-O3′ distance. The regularities of alternating helical twists and λ angles introduce a pseudodyad symmetry in the tandems, which can be used to understand or even predict non-canonical base pair configurations in tandems. For instance, a different configuration of a C-U base pair as found in the UUCG duplex[10] does not fit in the CU/UU tandem, due to its different twists and λ angles, leading to a near A-form cross-strand distance of 11.7 Å.

Thus, it appears that the conformation of RNA helices containing non-canonical tandem base pairs is determined by geometry. It seems that whenever possible the RNA helix tries to maintain its regularity and symmetry by compensating adjustments. This might be a more general phenomenon, which can be used to predict the conformation of non-canonical base pairs and to formulate additional principles that govern RNA folding.

REFERENCES

1. Ban, N.; Nissen, P.; Hansen, J.; Moore, P.B.; Steitz, T.A. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science **2000**, *289*, 905–920.
2. Biswas, R.; Wahl, M.C.; Ban, C.; Sundaralingam, M. Crystal structure of an alternating octamer r(GUAUGUA)dC with adjacent G × U wobble pairs. J. Mol. Biol. **1997**, *267*, 1149–1156.
3. Biswas, R.; Sundaralingam, M. Crystal structure of r(GUGUGUA)dC with tandem G × U/U × G wobble pairs with strand slippage. J. Mol. Biol. **1997**, *270*, 511–519.
4. Cate, J.H.; Gooding, A.R.; Podell, E.; Zhou, K.; Golden, B.L.; Kundrot, C.E.; Cech, T.E.; Doudna, J.A. Crystal structure of a group I ribozyme domain: Principles of RNA packing. Science **1996**, *273*, 1678–1685.
5. Chen, X.; McDowell, J.A.; Kierzek, R.; Krugh, T.R.; Turner, D.H. Nuclear magnetic resonance spectroscopy and molecular modeling reveal that different hydrogen bonding patterns are possible for G.U pairs: one hydrogen bond for each G.U pair in r(GGCGUGCC)2 and two for each G.U pair in r(GAGUGCUC)2. Biochemistry **2000**, *39*, 8970–8982.
6. Gautheret, D.; Konings, D.; Gutell, R.R. A major family of motifs involving G.A mismatches in ribosomal RNA. J Mol Biol. **1994**, *242*, 1–8.
7. Gautheret, D.; Konings, D.; Gutell, R.R. G.U base pairing motifs in ribosomal RNA. RNA **1995**, *1*, 807–814.
8. Heus, H.A.; Pardi, A. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. Science **1991**, *253*, 191–194.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

9. Heus, H.A.; Wijmenga, S.S.; Hoppe, H.; Hilbers, C.W. The detailed structure of tandem G.A mismatched base-pair motifs in RNA duplexes is context dependent. J. Mol. Biol. **1997**, *271*, 147–158.
10. Holbrook, S.R.; Cheong, C.; Tinoco, I. Jr.; Kim, S.H. Crystal structure of an RNA double helix incorporating a track of non-Watson-Crick base pairs. Nature **1991**, *353*, 579–581.
11. Jang, S.B.; Hung, L.W.; Chi, Y.I.; Holbrook, E.L.; Carter, R.J.; Holbrook, S.R. Structure of an RNA internal loop consisting of tandem C-A+ base pairs. Biochemistry **1998**, *37*, 11,726–11,731.
12. Kierzek, R.; Burkard, M.E.; Turner, D.H. Thermodynamics of single mismatches in RNA duplexes. Biochemistry **1999**, *38*, 14,214–14,223.
13. Kuglstatter, A.; Oubridge, C.; Nagai, K. Induced structural changes of 7SL RNA during the assembly of human signal recognition particle. Nat Struct Biol. **2002**, *9*, 740–744.
14. Leontis, N.B.; Westhof, E. The 5S rRNA loop E: Chemical probing and phylogenetic data versus crystal structure. RNA **1998a**, *4*, 1134–1153.
15. Leontis, N.B.; Westhof, E. Conserved geometrical base-pairing patterns in RNA. Q Rev Biophys. **1998b**, *31*, 399–455.
16. Leontis, N.B.; Westhof, E. Geometric nomenclature and classification of RNA base pairs. RNA **2001**, *7*, 499–512.
17. Leontis, N.B.; Stombaugh, J.; Westhof, E. The non-Watson-Crick base pairs and their associated isostericity matrices. Nucleic Acids Res. **2002**, *30*, 3497–3531.
18. Lescrinier, E.M.H.P.; Tessari, M.; van Kuppeveld, F.J.M.; Melchers, W.J.G.; Hilbers, C.W.; Heus, H.A. Structure of the pyrimidine rich internal loop in the poliovirus 3′-UTR. The importance of maintaining pseudodyad symmetry in RNA helices containing two adjacent non-canonical base pairs.
19. Lietzke, S.E.; Barnes, C.L.; Berglund, J.A.; Kundrot, C.E. 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 **1996**, *4*, 917–930.
20. Masquida, B.; Westhof, E. On the wobble GoU and related pairs. RNA **2000**, *6*, 9–15.
21. McDowell, J.A.; Turner, D.H. Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: Solution structure of (rGAGGUCUC)2 by two-dimensional NMR and simulated annealing. Biochemistry **1996**, *35*, 14,077–14,089.
22. McDowell, J.A.; He, L.; Chen, X.; Turner, D.H. Investigation of the structural basis for thermodynamic stabilities of tandem GU wobble pairs: NMR structures of (rGGAGUUCC)2 and (rGGAUGUCC)2. Biochemistry **1997**, *36*, 8030–8038.
23. Mckay, D.B. Structure and function of the hammerhead ribozyme: An unfinished story. RNA **1996**, *2*, 395–403.
24. Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
25. Santa Lucia, J., Jr.; Turner, D.H. Structure of $(rGGCGAGCC)_2$ in solution from NMR and restrained molecular dynamics. Biochemistry **1993**, *32*, 6026–6036.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

26. Traub, W.; Sussman, J.L. Adenine-guanine base pairing ribosomal RNA. Nucleic Acids Res. **1982**, *10*, 2701–2708.
27. Trikha, J.; Filman, D.J.; Hogle, J.M. Crystal structure of a 14 bp RNA duplex with non-symmetrical tandem G × U wobble base pairs. Nucleic Acids Res. **1999**, *27*, 1728–1739.
28. van Knippenberg, P.H.; Formenoy, L.J.; Heus, H.A. Is there a special function for U.G basepairs in ribosomal RNA?. Biochim Biophys Acta. **1990**, *1050* 14–17.
29. Wu, M.; McDowell, J.A.; Turner, D.H. A periodic table of symmetric tandem mismatches in RNA. Biochemistry **1995**, *34*, 3204–3211.
30. Wu, M.; Turner, D.H. Solution structure of r(GCGGACGC)$_2$ by two-dimensional NMR and the iterative relaxation matrix approach. Biochemistry **1996**, *35*, 9677–9689.
31. Wu, M.; SantaLucia, J.; Turner, D.H. Solution structure of r(GGCAGGCC)$_2$ by two-dimensional NMR and the iterative relaxation matrix approach. Biochemistry **1997**, *36*, 4449–4460.
32. Xia, T.; McDowell, J.A.; Turner, D.H. Thermodynamics of nonsymmetric tandem mismatches adjacent to G.C base pairs in RNA. Biochemistry **1997**, *36*, 12,486–12,497.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016