Singly and Bifurcated Hydrogen-bonded Base-pairs in tRNA Anticodon Hairpins and Ribozymes

Pascal Auffinger and Eric Westhof

Communications et Simulations des Acides Nucleiques, UPR 9002, Institut de Biologie Moleculaire et Cellulaire du CNRS, 15 rue Rene Descartes 67084 Strasbourg Cedex France

The tRNA anticodon loops always comprise seven nucleotides and are involved in many recognition processes with proteins and RNA fragments. We have investigated the nature and the possible interactions between the first (32) and last (38) residues of the loop on the basis of the available sequences and crystal structures. The data demonstrate the conservation of a bifurcated hydrogen bond interaction between residues 32 and 38, located at the stem/loop junction. This interaction leads to the formation of a non-canonical base-pair which is preserved in the known crystal structures of tRNA/synthetase complexes. Among the tRNA and tDNA sequences, 93% of the 32-38 oppositions can be assigned to two families of isosteric base-pairs, one with a large (86%) and the other with a much smaller (7%) population. The remainder (7%) of the oppositions have been assigned to a third family due to the lack of evidence for assigning them into the first two sets. In all families, the Y32-R38 base-pairs are not isosteric upon reversal (like the sheared G:A or wobble G:U pairs), explaining the strong conservation of a pyrimidine at position 32. Thus, the 32-38 interaction extends the sequence signature of the anticodon loop beyond the conserved U-turn at position 33 and the usually modified purine at position 37. A comparison with other loops containing both a singly hydrogen-bonded base-pair and a U-turn suggests that the 32-38 pair could be involved in the formation of a base triple with a residue in a ribosomal RNA component. It is also observed that two crystal structures of ribozymes (hammerhead and leadzyme) present similar base-pairs at the cleavage site.

Introduction

Hydrogen bonds between bases constitute the cement of nucleic acid structures. Among nucleic acid base-pairs, the well-known Watson-Crick G:C and A:U pairs involve three and two hydrogen bonds, respectively. Besides these regular Watson-Crick interactions, a large array of possible base-pair interactions involving two hydrogen bonds have been enumerated and many of them have been observed in crystal structures (Tinoco, 1993; Dirheimer et al., 1995; Leontis & Westhof, 1999a). Such non-Watson-Crick interactions subvert the great architectural diversity of nucleic acids, and especially of RNA molecules where they are recurrently observed. Further, non-Watson-Crick pairs contribute to the structural flexibility of these molecules, since they are often strongly context dependent. A less recognized class of base-pairs comprises those in which the two bases are linked by a single or a bifurcated hydrogen bond.

In transfer RNAs, the anticodon loop plays a central role in processes associated with protein synthesis. On the one hand, it must interact specifically with its cognate synthetase, and on the other hand, it must fit in the common tRNA binding site of the ribosome for codon recognition on the messenger RNA. Over the years, rather extensive databases at the tRNA level (546 sequences) or at the tDNA level (2726 sequences) have been compiled (Sprinzl et al., 1998). Several years ago, it was noticed that the first (32) and last (38) residues of the seven-membered anticodon loop are involved in a bifurcated hydrogen bond contact (Quigley &

Keywords: Nucleic acid conformation; tRNA; anticodon; leadzyme; hammerhead

© 1999 Academic Press
C32 . . . N6(A38) in tRNA Phe. Interestingly, the
by a N1-H group and the C1 carbonyl oxygen atom and an amino group, i.e.
0-N1 link between the base and the sugar is replaced by a C1-C5 bond; see Figure 2). The 32 and 38 bases are linked by a bifurcated hydrogen bond established between a carbonyl oxygen atom and an amino group, i.e. O(Ψ32)...N4(C38) in tRNAASp and O2(C32)...N6(A38) in tRNAphe. Interestingly, the
carbonyl and amino groups as well as the glycosidic bonds of both base-pairs superimpose very well. Thus, although a significant out-of-plane deviation is observed (Figure 2), these 32-38 pairs can be considered as isosteric (Figure 2). It is noteworthy that the observed out-of-plane deviation does not affect the interaction scheme of the two bases or the inter-base C1′(32)...C1′(38) distance, close to 11 Å as in Watson-Crick pairs. Yet, the deviation of these isosteric pairs from a Watson-Crick arrangement is clearly illustrated by the large distance (≈7 Å) between the C1′ atoms of residue 38 in the bifurcated hydrogen bonded 32-38 pairs and in the corresponding Watson-Crick arrangement (Figure 2). Interestingly, a U-C pair identical with that observed in yeast tRNAASp has been observed in the crystal structure of the alpha-sarcin loop of 23 S rRNA (Correll et al., 1998). A different situation is observed in the crystal structure of yeast tRNAMet (Basavappa & Sigler, 1991), where the C32-A38 pair (Figure 3) is not isosteric to the 32-38 pairs described above. The authors reported that the resolution of the anticodon hairpin is very low. Hence, this interaction scheme is probably of poor structural relevance and will not be further discussed. This view is also supported by recent NMR studies of two transcripts of yeast tRNAMet and Escherichia coli tRNAmet anticodon hairpins (Schweisguth & Moore, 1997) from which it was concluded that the observed non-canonical C32-A38 NMR base-pairing are much closer to those observed in yeast tRNAphe crystal structures than the one observed in the yeast tRNAMet structure (Schweisguth & Moore, 1997). A recent crystal structure of the complex between E. coli tRNACys and the translation elongation factor EF-Tu reveals the arrangement of a Ψ32-A38 base-pair (pr0004; Nissen et al., 1999). Similarly to what is observed in the crystal structure of the complex of tRNAphe with EF-Tu (ptr012; Nissen et al., 1995), the anticodon loop of tRNAcys adopts a conformation close to that seen in the crystal structures of uncomplexed tRNAs with a non-Watson-Crick Ψ32-A38 pair. Again the Ψ-A pair is isosteric to the Ψ-C pair observed in

![Figure 1. Secondary structures of tRNA anticodon hairpins for which crystal structures are accessible in the NDB. The 32-38 base-pairs are shown in bold.](image-url)
yeast tRNA\textsuperscript{Asp} and to the C·A pair observed in yeast tRNA\textsuperscript{Phe} (Figure 2). Interestingly, in both tRNA\textsuperscript{Cys} and tRNA\textsuperscript{Phe} (trnA06) structures, a water molecule is seen in the vicinity of the N7 and N6-H groups of A38. In trnA09, there is a comparable water molecule which belongs, however, to the coordination sphere of a Mg\textsuperscript{2+}.

Hence, from the available tRNA crystal structures where the anticodon hairpin keeps its folded structure, it can be deduced that a set of non-canonical but isosteric base-pairs involving a bifurcated hydrogen bond between a carbonyl oxygen atom and an amino group is recurrently present at the junction between the stem and the loop of the anticodon hairpin. On a thermodynamic basis, base-pairs involving a single or bifurcated hydrogen bond should be more labile compared to those involving two or three hydrogen bonds. However, as described below, formation of specific tRNA/synthetase complexes, which involve the splaying out toward the exterior of five (33 to 37) of the bases of the loop, do not disrupt the direct hydrogen bonding contacts established between bases 32 and 38.

**Crystallographic structures of tRNA/synthetase complexes**

The coordinates of several crystal structures of tRNAs complexed with their cognate synthetase are available (Table 1). Upon formation of the tRNA/synthetase complexes, slight conformational rearrangements of the 32·38 base-pair are observed.

**Ψ32·C38 pairs.** Two structures of the complex of the wild-type yeast tRNA\textsuperscript{Asp}, including all the modified bases with its cognate synthetase (tRNA\textsuperscript{Asp}/RS) are available (Table 1). In these structures, the Ψ32·C38 pair adopts a conformation different from that observed in the uncomplexed tRNA structures (Figure 4). The C38 amino group forms a single hydrogen bond rather than a bifurcated hydrogen bond with the O4(Ψ32) atom. Further, a significant out-of-plane orientation of base C38 is observed. These arrangements appear less stable than those formed in uncomplexed tRNAs and are, possibly, stabilized by additional protein/tRNA interactions.

**U32·U38 and Um32·Ψ38 pairs.** No structure has been solved of an uncomplexed tRNA having a U32·U38 pair. However, in the structures of the transcripts of *E. coli* tRNA\textsuperscript{Gln}/RS complex (Table 1), a U32·U38 pair is observed (Figure 4). Again, a single hydrogen bond links the two bases. Note that this base-pair is approximately isosteric to a Watson-Crick pair (see the bottom drawing in right-hand column of Figure 6). Thus, its geometry
uncomplexed tRNA Phe. Furthermore, in the structure different from those observed in the structures of protein, a N3(C32) . . . N6(A38) is formed. This contact (cation). Beside contacts made by A38 with the pro-
et al., 1996a; S. Cusack, personal communication). In the structures of the two complexes of T. thermophilus tRNAPro synthetase with two tRNAPro isoacceptors, determined at 3.5 Å resolution, the occurrence of a non-standard U32-C9 pairs might be slightly lower since it has been

tRNA/RS complexes
ptr001 1QSG E. coli tRNAAsp 2.8 U - A Y
ptr002 1QTS E. coli tRNAAsp 2.8 U - A Y
ptr003 1QY S. Cusack, personal communication). In order to detect any sequence preferences for base-pairs at position 32-38, the tRNA gene database which contains up to 2726 tDNA sequences (Sprinzl et al., 1998) was analyzed. First, there is a very large proportion of pyrimidines at position 32 (≈98%, see Table 2) and a significant number of purines at position 38 (≈71%). Secondly, sequences which could form C-G (six occurrences), G-C (no occurrence) or A-U (ten occurrences) Watson-Crick pairs are very rare. Nevertheless, there is a large proportion of U-A pairs (≈18%). Note that the number of U-A pairs might be slightly lower since it has been

<table>
<thead>
<tr>
<th>Free tRNAs*</th>
<th>PDB code</th>
<th>Res. (Å)</th>
<th>32-38</th>
<th>Transcript</th>
<th>Coord.</th>
</tr>
</thead>
<tbody>
<tr>
<td>tRNA/EF-TU complexes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pr0004 1B23 E. coli tRNAPro</td>
<td>2.6</td>
<td>2.7</td>
<td>Cm - A</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>ptr012 1TTT yeast tRNAPro</td>
<td></td>
<td></td>
<td>Cm - A</td>
<td>Y</td>
<td></td>
</tr>
</tbody>
</table>

| tRNA/RS complexes | | | | | |
| ptr001 1QSG E. coli tRNAAsp | 2.8 | U - A | N | N |
| ptr002 1QTS E. coli tRNAAsp | 2.8 | U - A | N | N |
| ptr003 1QY S. Cusack, personal communication). In order to detect any sequence preferences for base-pairs at position 32-38, the tRNA gene database which contains up to 2726 tDNA sequences (Sprinzl et al., 1998) was analyzed. First, there is a very large proportion of pyrimidines at position 32 (≈98%, see Table 2) and a significant number of purines at position 38 (≈71%). Secondly, sequences which could form C-G (six occurrences), G-C (no occurrence) or A-U (ten occurrences) Watson-Crick pairs are very rare. Nevertheless, there is a large proportion of U-A pairs (≈18%). Note that the number of U-A pairs might be slightly lower since it has been

| 32-38 pairs. In the structures of the two complexes formed between Thermus thermophilus tRNAAsp synthetase and transcripts of E. coli and T. thermophilus tRNAAsp, the occurrence of a non-standard C32-A38 base-pair has been reported (Cusack et al., 1996a; S. Cusack, personal communication). Beside contacts made by A38 with the protein, a N3(C32)...N6(A38) is formed. This contact is different from those observed in the structures of uncomplexed tRNAPro. Furthermore, in the structures of the two complexes of T. thermophilus tRNAPro synthetase with two tRNAPro isoacceptors, determined at 3.5 Å resolution, the occurrence of a non-standard U32-C9 and of a Watson-Crick U32-A38 base-pair is mentioned (Cusack et al., 1998; S. Cusack, personal communication).

Phylogenetic analysis

dNA gene sequences. In order to detect any sequence preferences for base-pairs at position 32-38, the tRNA gene database which contains up to 2726 tDNA sequences (Sprinzl et al., 1998) was analyzed. First, there is a very large proportion of pyrimidines at position 32 (≈98%, see Table 2) and a significant number of purines at position 38 (≈71%). Secondly, sequences which could form C-G (six occurrences), G-C (no occurrence) or A-U (ten occurrences) Watson-Crick pairs are very rare. Nevertheless, there is a large proportion of U-A pairs (≈18%). Note that the number of U-A pairs might be slightly lower since it has been

shown that U → C editing can occur at position 32 (Beier et al., 1992).

The preferred base-pair sequence is C·A (≈48%), which is followed by the T·A (≈18%), T·T (≈11%), C·C (≈8%) and T·C (≈8%) pairs. Most of these pairs, with the exception of the T·A pair, cannot form Watson-Crick arrangements. Note that these proportions do not vary significantly if one considers the two following subsets: subset 1 (932 sequences of tDNA elongators) and subset 2 (352 sequences of eukaryotic tDNA elongators) of the whole 2726 tDNA sequences.

Is the sequence at position 32·38 specific for a given tRNA type? In order to detect a possible correlation between the sequence of the 32·38 pair and either the amino acid type of the tRNA or the phylogenetic group in which they occur, the family of tDNAs containing a T32·T38 pair was analyzed (Sprinzl et al., 1998). The T·T pairs are found in tDNAs coding for 14 different amino acid residues, in eukaryotic, eubacterial, mitochondrial, and chloroplastic tRNAs (in archaea, only one occurrence is reported). Further, no correlation between the “base type” at positions 37 and 38 could be noted (146 A37·T38 and 140 G37·T38 sequences).

As in the case of tDNAs containing a T32·T38 pair,
Figure 5 is different from the U.C pairing types will be considered.

A-A (1.1 %, see Table 2), G-U (0.6 %), and G-A (0.1 %) pairs are roughly isosteric to C-A pairs (family I of Figure 5), while U-G (2.7 %), C-U (1.3 %), and A-U (0.4 %) pairs are approximately isosteric to U-U pairs (family II shown in Figure 5). For the first family, the A13-C22 pair, found in the crystal structure of E. coli tRNA^Gln^ (pte003), is close to the bifurcated C-A pair. For G-U pairs, no structures close to that of the C-A pair have been observed and, thus, the G-U pair shown in Figure 6 is suggested. The crystal structure of the 5 S rRNA loop E (ur1064; Correll et al., 1997) contains a sheared G98-A78 pair superposable to the singly hydrogen bonded C-A pair. For the second family, the G-U and A-U pairs shown in Figure 6 have

Table 2. Percentages of occurrence of the four A, G, C and T bases at the non-canonical 32-38 base-paired position for all 2726 tDNAs contained in the tRNA database (Sprinzl et al., 1998), a subset of 932 elongator tDNAs, and a subset of 352 eukaryotic elongator tDNAs (percentages larger than 5 % are idicated in bold)

<table>
<thead>
<tr>
<th>Base 32</th>
<th>A</th>
<th>G</th>
<th>C</th>
<th>T</th>
<th>Tot.</th>
<th>R/Y*</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>≈0</td>
<td>≈0</td>
<td>≈0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>≈6</td>
<td>1</td>
<td>1</td>
<td>56</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>48</td>
<td>-</td>
<td>8</td>
<td>11</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>18</td>
<td>3</td>
<td>3</td>
<td>30</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Total:</td>
<td>66</td>
<td>3</td>
<td>16</td>
<td>13</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>R/Y*</td>
<td>69</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All tDNA elonators (932)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>≈0</td>
<td>≈0</td>
<td>≈0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>54</td>
<td>1</td>
<td>12</td>
<td>1</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>13</td>
<td>7</td>
<td>10</td>
<td>11</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>Total:</td>
<td>68</td>
<td>2</td>
<td>18</td>
<td>11</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>R/Y*</td>
<td>69</td>
<td>1</td>
<td>11</td>
<td>30</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>All tDNA eukaryotic elonators (352)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>≈0</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>58</td>
<td>14</td>
<td>2</td>
<td>74</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>2</td>
<td>8</td>
<td>15</td>
<td>25</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>Total:</td>
<td>89</td>
<td>22</td>
<td>17</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R/Y*</td>
<td>61</td>
<td>39</td>
<td>-</td>
<td>100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Indicates the percentages of purine (R) and pyrimidine (Y) bases at position 32 and 38.

no apparent correlation exists for tDNAs displaying other types of 32-38 pairs.

**Model structures for the 32-38 interaction in tRNAs**

C-A, U-A, U-C, C-C and U-U base-pairs. From the preceding structural and phylogenetic data, it is possible to propose models for base-pairs located at position 32-38 of tRNA anticodon hairpins (Figure 5). As the vast majority of 32-38 pairs are of the C-A type, the arrangement observed in the crystal structures of yeast tRNA^Phe^ (see Figure 2) seems the most plausible (family I). The same configuration can be adopted by U-A pairs which may be linked by a bifurcated rather than by two regular hydrogen bonds as in Watson-Crick pairs (Figure 5). As shown before, the much rarer U-C and C-C pairs could adopt conformations isosteric to C-A and U-A pairs. Interestingly, the shallow groove patterns of C-A and U-A pairs are similar, as are the shallow groove patterns of U-C and C-C pairs. Note that the U-C pair shown in Figure 5 is different from the U-C pair observed in the crystal structure of several RNA oligomers (Holbrook et al., 1991; Cruse et al., 1994; Tanaka et al., 1999) where the two bases are linked by a O4(U)...N4(C) instead of an O2(U)...N4(C) hydrogen bond with a water molecule bridging the two N3 sites of both pyrimidines.

However U-U pairs, which account for approximately 7 % of the total number of pairs, cannot adopt a conformation isosteric to that observed for C-A pairs of family I. The arrangement shown in Figure 5 is the one observed in the crystal structures of the wild-type and transcripts of E. coli tRNA^Gln^ complexed with its cognate synthetase (see Figure 4). It is proposed to be a representative of family II. tRNAs with a U-U pair at position 32-38 may also belong to a separate family of tRNAs with specific and unknown features.

**Rare 32-38 base-pairs.** Besides the pairs in families I and II shown in Figure 5, several rare base oppositions are found at positions 32-38. They account for 7 % of the total number of tDNA (Table 2) and tRNA (Table 3) sequences and are gathered in family III. Table 2 indicates that only G-C pairs are completely avoided in the 2726 sequences reported in the tRNA database (Sprinzl et al., 1998). For several of these pairs, pairing arrangements approximately isosteric to family I or to family II can be suggested (Figure 6). In the following, only cis pairing types will be considered.

A-A (1.1 %, see Table 2), G-U (0.6 %), and G-A (0.1 %) pairs are roughly isosteric to C-A pairs (family I of Figure 5), while U-G (2.7 %), C-U (1.3 %), and A-U (0.4 %) pairs are approximately isosteric to U-U pairs (family II shown in Figure 5). For the first family, the A13-C22 pair, found in the crystal structure of E. coli tRNA^Gln^ (pte003), is close to the bifurcated C-A pair. For G-U pairs, no structures close to that of the C-A pair have been observed and, thus, the G-U pair shown in Figure 6 is suggested. The crystal structure of the 5 S rRNA loop E (ur1064; Correll et al., 1997) contains a sheared G98-A78 pair superposable to the singly hydrogen bonded C-A pair. For the second family, the G-U and A-U pairs shown in Figure 6 have
been modeled in order to be roughly isosteric to the U-U pair while interacting only through a single hydrogen bond. The C-U pair is extracted from the same crystal structure of an RNA duplex (ar0005; Tanaka et al., 1999). Note that NMR data obtained for E. coli tRNAPhe revealed signals associated to a rare A32/C38 interaction (Hyde & Reid, 1985a,b). For all these base-pairs, slight changes in the positions of the bases may “improve” their isosteric character. For A-G (0.3 %), A-C (0.2 %), C-G (0.2 %), G-C (0.04 %), and G-C (0 %) oppositions, no arrangement close to either the C-A or the U-U pair shown in Figure 5 can be proposed. Yet, for these rare base-pairs, the possibility of base editing (Price & Gray, 1998) and the probable existence of a noise level in the tRNA sequence database leading to a small number of incorrect sequences should be taken into account. Therefore, the sequences and model structures of family III should be considered with caution.

### Discussion

#### A conserved non-canonical 32-38 base-pair in the anticodon loop of tRNAs

The preceding structural and phylogenetic data lead to the conclusion that particular types of pairing occur systematically between the first (32) and last (38) residues in the anticodon loop of tRNAs. The largest family of base-pairs (family I) comprises the C-A pair, where the two bases are linked by a bifurcated hydrogen bond, and the isosteric U-A, U-C and C-C pairs (Figure 5). This category represents 86 % of the total number of sequences. A second family of base-pairs (family II) comprises the U-U pairs (7 %), which are not isosteric to the
pairs of family I. Finally, a third family (family III) comprises a set of rare base-pairs (7%) for which few structural data are available (Figure 6), but which could be roughly distributed among the first two families.

Importance of bifurcated hydrogen bonds

Up to now, pairs where the bases are linked by a single or a bifurcated hydrogen-bonding interaction have encountered limited interest, probably because of their apparent lability. Yet, such pairs were observed in the crystal structures of yeast tRNA\textsuperscript{The} (Quigley & Rich, 1976) and yeast tRNA\textsuperscript{Asp} (Westhof et al., 1985) at position 32-38 (where two chemical groups are involved) and at position G18-\Psi55 (where three chemical groups are involved). Furthermore, subsequent molecular dynamics simulations have shown that they are stable over nanosecond time scales in yeast tRNA\textsuperscript{Asp}, and the structure of the C-C pair is inferred from that of the U-C pair extracted from yeast tRNA\textsuperscript{Asp} (see Figure 2). In family II, the structure of the U-U pair is extracted from the E. coli tRNA\textsuperscript{Gln/RS} complex (see Figure 4). Family III gathers the rare oppositions. The glycosidic bonds of base 32 are similarly oriented in all the base-pair oppositions.

Why a conserved pyrimidine at position 32 in tRNAs?

The non-canonical C32-A38 base-pair is characterized by a Watson-Crick-like distance between the C1' atoms and an asymmetric disposition of the angles at the C1' atoms (Figure 2). The latter geometrical characteristic leads to a pronounced non-isostericity upon reversal of the base-pair. Thus, C-A pairs are not isosteric to A-C pairs (Figure 7), U-A pairs are not isosteric to A-U pairs, and U-C pairs are not isosteric to C-U pairs. The C-C and U-U pairs shown in Figure 5 are obviously self-isosteric. Thus, the anticodon loop must begin with a pyrimidine at position 32 in order to adopt its characteristic functional shape. These structural considerations are in agreement with phylogenetic data indicating that, in tRNAs, a pyrimidine is present in 98% of the sequences at position 32 (Tables 2 and 3).
Is isostericity always required?

As demonstrated by the Watson-Crick pairs, isostericity is a central concept in nucleic acid structure. From crystal and phylogenetic data gathered from a set of sequences of the 5S rRNA loop E, isostericity has also been proposed for several non-canonical base-pairs (Leontis & Westhof, 1998a,b, 1999b). In the present study, a certain number of isosteric arrangements have been observed at position 32-38. Such isosteric pairs may result in a similar structure of the anticodon loop for free tRNAs. However, the occurrence of a non-negligible number of U·U and related pairs (7%) which are not isosteric to C·A pairs indicates that, depending on the context, isosteric base-pairs may not be systematically required. Thus, such sets of non-isosteric base-pairs may be characteristic of different tRNA families, for which adjustments for interactions with proteins, other RNA motifs or divalent ions are required. Significantly, when base-pairs are not strictly isosteric to the C·A or U·U pairs, their percentage of occurrence drops considerably (see Figure 6).

Several biochemical studies have evaluated the effects of changing the 32-38 base-pair type. Yarus and co-workers (1986b) have studied the suppression efficiency of a large number of mutants of the
amber suppressor Su7 and have shown that no mutation at position 32 or 38 was neutral. For example, Um·Ψ and Um·C pairs at position 32-38 were found to be less efficient than Um·A or Cm·A pairs (Yarus et al., 1986a). Interestingly, the measured suppression efficiency is Um·Ψ ≈ Um·C < Um·A < Cm·A and, thus, follows the proportion of 32-38 pairs derived from phylogenetic analysis (Figure 5). The substitution of the Cm32 nucleotide by any purine residue reduces the suppression efficiency more than tenfold (Smith & Yarus, 1989) in agreement with the fact that purines are strongly disfavored at position 32. Mutations of base 38 show less dramatic effects indicating that a larger mutational variability is allowed at position 38 than at position 32 (see Tables 2 and 3). In transcripts of E. coli tRNA\textsubscript{Gly}, with a U32·A38 pair the anticodon discriminates the glycine codons according to the wobble rules, but with a C·A pair it loses its discriminating power (Lustig et al., 1993). In transcripts of Mycoplasma mycoides tRNA\textsubscript{Gly}, the substitution of the wild-type C32·A38) by a U·A pair (Claesson et al., 1995) resulted in similar effects (but the transformation was the reverse of that performed in E. coli tRNA\textsubscript{Gly}). Since in both cases an isosteric substitution was performed (C·A↔U·A, see Figure 5), these data indicate that the nature of the contacts between the tRNA and the ribosome may involve the 32-38 base-pair and that such a substitution affects the recognition pattern of the anticodon hairpin more than its three-dimensional fold. It has also been demonstrated that substitution of U32·A38 by a C·A pair in M. mycoides tRNA\textsubscript{Gly} results in a substantial increase in its frameshifting efficiency (O’Connor, 1998). In E. coli tRNA\textsubscript{Gln} transcripts, substitution of the U32·C38 opposition by a C·C or C·A pair resulted in an increase of both aminoacylation and ribosome performance, but enhanced the latter function to a greater extent (McClain et al., 1998). It is possible that the replacement of a rare U32·U38 by a non-isosteric C·A pair has as a consequence the formation of a more “anticodon” like hairpin. Besides the isostericity of several of these non-canonical base-pair types, each of these base-pairs may interact specifically with proteins and ribosomal elements (see below). Iso- stericity, thus, is a generally valid concept indicating that it is possible to substitute, with minor structural changes, base-pairs displaying different shallow and deep groove recognition patterns. However, each of these isosteric base-pairs has specific functional roles, experimentally characterized by different translational efficiency. Furthermore, in some occurrences non-isosteric substitutions have been observed like the U·U pair of the rare base-pairs found in Tables 2 and 3. For example, in E. coli, a minor tRNA\textsuperscript{A\textsubscript{38}} isoacceptor differs from the other isoacceptors in possessing a rare purine instead of a pyrimidine at position 32. However, the structure and functional role of such base-pairs is not yet known.

Importance of modified nucleotides at positions 32 and 38

Modified nucleotides are found in many RNA molecules and especially in tRNAs where they account for 12% of the total number of residues (Grosjean & Benne, 1998). A large number of modified nucleotides are located at the stem/loop junction of the anticodon hairpin in tRNAs (Auffinger & Westhof, 1998b). It has been proposed, in agreement with a large body of experimental evidence and molecular dynamics simulations, that some of these modifications, and especially pseudouridylation (Auffinger & Westhof, 1998a), are required to stabilize the three-dimensional structure of the functionally important anticodon loop. At position 32, the stabilization mechanism involves the formation of a pseudouridine/water complex where the water molecule mediates base to backbone interactions, and thus helps to increase the stability of the 32-38 interaction pairs (Davis & Poulter, 1991; Arnez & Steitz, 1994; Auffinger & Westhof, 1997, 1998a). The same nucleotide/water complexes can occur after pseudouridylation at positions 31, 38, and 39. Such modifications may thus help to strengthen the structure of the tRNA anticodon hairpin in the ribosome and prevent excessive unfolding of the anticodon loop when interacting with synthetase proteins.

Indeed, as noted above, in a few of the crystal structures of tRNAs complexed with their cognate synthetase, the 32-38 interaction is systematically maintained even after partial unfolding of the loop, and is stabilized by water mediated interactions involving modified nucleotides (Roud et al., 1989). Thus, one possible role of the 32·38 pair and of associated modified nucleotides is to limit the “unfolding” of the loop to the five bases, 33 to 37. The structure of base-pairs 31·39, 32·38, and the stacking of U33 below base 32 being at least partially preserved, as shown in Figure 8(a) for the
E. coli tRNA\textsuperscript{Gln}/RS complex, a subsequent "refolding" to the ribosomal active form, after decomplexation with the synthetase, may be facilitated. Besides a structural role, the question of a possible functional role of the non-canonical 32-38 base-pair can be raised. It has been noted that in some instances bases 32 and 38 can interact with the cognate synthetase and, thus, act as structural determinants in the recognition process. Biochemical data fully support this view (for a review, see Giege 8).

Figure 8. Turns in RNA motifs involving non-canonical base-pairs. (a) Views extracted from crystal structures of yeast tRNA\textsuperscript{Phe} (trna09) and tRNA\textsuperscript{Asp} (trna07) as well as from the crystal structures of the complex of E. coli tRNA\textsuperscript{Gln} with its cognate synthetase (pte003). For comparison, a fragment of a conventional Watson-Crick helix is drawn. The position of the glycosidic bond, for a hypothetical 32-38 Watson-Crick base-pair in the tRNA\textsuperscript{Phe} anticodon loop is marked by a green arrow. (b) Fragments of the crystal structures of the TΨC loop of yeast tRNA\textsuperscript{Phe} (trna09) and of the GNRA tetraloop of the hammerhead ribozyme (uhx026). (c) Fragments of the crystal structure of the leadzyme (ur0001) and the hammerhead ribozyme (uhx026) showing the C23-A45 and the C3-C17 base-pairs. In all these Figures, the glycosidic bonds of 32 have been superimposed.
et al., 1998). Nevertheless, the largest number of tRNA/synthetases do not use these bases as structural determinants and, thus, their structural role seems to prevail over their direct implications in a recognition process with a synthetase.

Is there a relationship between the 32-38 interaction and the U-turn?

The active three-dimensional structure of tRNA anticodon loops is characterized by a U-turn associated with the conserved base U33 and the A-form helical stack of loop residues 34 to 38 in continuity with the helical stem residues 39 to 44 (as shown in Figure 8(a) for yeast tRNAAsp and yeast tRNAPhe). While both residues 32 and 38 are located in a helical track, the rotation of the glycosidic bonds on the side of residue 32 is much more pronounced than in regular helical A-form stacks (Figure 8(a)). Furthermore, between residues 34 and 38, the rotation angles are rather regular and not far from the helical case. But, clearly, residue 38 has not rotated sufficiently to form a Watson-Crick pair with residue 32. Hence, a substitution of the non-canonical 32-38 pair by a regular Watson-Crick pair would hinder the formation of the U-turn and the regularity of the loop. This explains why no G·C and very few C·G pairs are reported at these positions (Table 2). Therefore, the isosteric C·A, U·A, U·C, and C·C pairs at position 32-38, and possibly also U·U pairs, appear as a crucial transition element linking the stem to the loop of the anticodon hairpin. Residues 32 and 38 should therefore be considered, along with the conserved base U33 and likely a conserved purine at position 37, as a signature for tRNA anticodon loops (Figure 9). The view that only specific base combinations occur at position 32-38 is in agreement with the extended anticodon concept, indicating that the coding performance of the triplet anticodon is enhanced by the appropriate loop and stem sequence (Yarus, 1982; Yarus et al., 1986b; Yarus & Smith, 1995).

In the seven base T-loops of tRNAs, a non-canonical two hydrogen-bonded trans Hoogsteen T54·A58 pair precedes the U-turn. This base-pair appears to play a similar role in T-loops as that played by the 32-38 pair in anticodon loops (Figure 8(b)). Note that a T54·A58 pair is reported in 78% of tRNA gene sequences (Sprinzl et al., 1998). Similarly, in the recent crystal structure of the L11 binding region in LSU RNA, a trans Watson-Crick A·U pair flanks the U-turn of a five nucleotide loop (Conn et al., 1999). GNRA tetra-loops are known to form U-turns as well (Westhof et al., 1989; Jucker & Pardi, 1995). In such motifs, G·A sheared pairs, which involve two hydrogen bonds, are used to close the loop (Figure 8(b)). Thus, sheared G·A pairs, almost isosteric to C·A pairs (Figure 6), are used in short loops, presumably because of their intrinsic higher stability and because they induce a sharper turn, while the apparently more labile singly hydrogen bonded base-pairs found in anticodon motifs are used in longer loops.

![Figure 9. tRNA anticodon loop signature showing the structure/sequence relationship at the 32-38 base-pair level. The distribution of the four A, G, C, and U nucleotides is extracted from the set of 2726 tDNA gene sequences present in the tRNA database (Auffinger & Westhof, 1998b; Sprinzl et al., 1998).](image-url)
Other examples of interfacial non-canonical base-pairs

In hairpins. A recent NMR structure of the tRNA1ys3 anticodon hairpin obtained at low pH revealed the formation of a C32·A 38 base-pair (Durant & Davis, 1999). The pKₐ of the adenine base was estimated to be close to 6 for A38. Likewise, evidence for the formation of a non-canonical bifurcated U·C pair have been obtained by NMR for the central hairpin of the HDV ribozyme containing a seven nucleotide loop (Kolk et al., 1997). In this structure, the sharp turn which is associated with the change of direction of the backbone is shifted in the 3′ direction by one nucleotide, and characterized as a reversed U-turn. Interestingly, an independent NMR determination of the structure of the same hepatitis delta virus hairpin proposed a different hydrogen-bonding scheme for the U·C pair of the nucleotides on the 5′ and 3′-ends of the loop (Lynch & Tinoco, 1998). In this latter structure, the cytosine amino group interacts with O4 instead of O2 of U resulting in a water structure, the cytosine amino group interacts ends of the loop (Lynch & Tinoco, 1998). In this later structure, the cytosine amino group interacts with O4 instead of O2 of U resulting in a water inserted U·C pair similar to that observed in the crystal structure of RNA oligonucleotides (Holbrook et al., 1991; Cruse et al., 1994; Tanaka et al., 1999). In any case, in each example of structurally determined hairpins containing loops with seven bases, a non-canonical base-pair is observed at the junction between the stem and the loop.

Several examples of loops closed by non-canonical base-pairs can be found in crystal structures (Pley et al., 1994; Scott et al., 1995; Cate et al., 1996; Correll et al., 1998; Perbandt et al., 1998). A hairpin closing wobble U·U pair is found in the crystal structure of the complex formed between a fragment of the U2 snRNA and a spliceosomal protein (Price et al., 1998). Interestingly, in the crystal structure of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin, no loop closing C·A base-pair is observed (Oubridge et al., 1994) indicating that, in this other structural context, a possible non-canonical base-pair is disrupted after complex formation.

In ribozymes. Interestingly, the crystal structure of the leadzyme (Wedekind & McKay, 1999), a C23·A45 pair, identical with the C·A pair observed in tRNAs, is present (Figure 8(c)). It has been proposed, on the basis of crystallographic data, that this non-canonical pair is involved in the binding of a Pb²⁺ in the RNA deep groove which may be important for catalysis. In one motif of the crystal structure of a Ba²⁺ occupies this site (in tRNA crystal structures, a related Mg²⁺ binding site is located in the deep groove of the 32-38 pair). Interestingly, the formation of a C·A pair is not mandatory for catalysis, since substitution experiments have shown that the ribozyme is still active after replacement of the adenine residue by an abasic ribose (Chartrand et al., 1997).

In the hammerhead ribozyme, a U·A non-canonical base-pair involving a single hydrogen bond closes stem III (Pley et al., 1994; Scott et al., 1995). This hydrogen bond, which is not a bifurcated one as in tRNAs, is well maintained during an MD simulation of the ribozyme (Hermann et al., 1998). Also, and most interestingly, the base preceding the U-turn in the hammerhead ribozyme is normally a C and it forms a wobble base-pair with the cleavable C residue (C3-C17: see Figure 8(c)). This base-pairs differ from the singly hydrogen-bonded base-pair shown in Figure 5. Nevertheless, it seems essential for the ribozyme as a substitution of this C·C pair by a G·C pair results in a considerable loss of activity (McKay, 1996).

Could the 32·38 base-pair participate in RNA/RNA recognition in the ribosome?

As emphasized above, several biochemical studies have demonstrated that modifications of the 32-38 base-pair can alter both aminoacylation and ribosome performance suggesting that this base-pair is involved in direct or indirect recognition features with their cognate synthetase and with the ribosome (Lustig et al., 1993; Claesson et al., 1995; Yarus & Smith, 1995; Giegé et al., 1998; McClain et al., 1998; O’Connor, 1998) or, as proposed by Yarus and co-workers (Smith & Yarus, 1989; Yarus & Smith, 1995), with a second tRNA inside the ribosome. For RNA/RNA interactions within a given organism, the Watson-Crick binding sites of residue 38, located in the shallow groove of the anticodon stem, are not blocked and could be involved in the formation of a triple interaction with a ribosomal residue (Figure 10). Thus, the adenine base of C·A and U·A pairs and the cytosine base of U·C and C·C pairs could be involved in Watson-Crick interactions with U or G residues, respectively. Yet, other interaction types cannot be excluded. For example, the formation of a trans A38·U or C38·U pair could be proposed. This interaction scheme presents the advantage of involving a uridine base in all four cases (Figure 10). Further, a uridine base could also be placed in the shallow groove of the U32·U38 pair. Interestingly, GNRA tetraloops in which the Watson-Crick sites of the guanine or the adenine residue of the closing G·A pair are available, are involved in RNA/RNA recognition via the Watson-Crick sites of the A in the shallow groove side of G·C pairs in helices (Michel & Westhof, 1990; Jaeger et al., 1994). Crystal structures have demonstrated that the G·A pair of a GNRA tetraloop can indeed form of a base triple with a helix (Pley et al., 1994).

However, RNA/RNA interactions could also take place on the less accessible deep groove side, especially at the border of a helix. Biochemical data indicate that the substitution of a C·A by an isosteric U·A pair affects the discriminating ability of tRNA transcripts (Lustig et al., 1993; Claesson et al., 1995) and their frameshifting efficiency (O'Connor, 1998). This seems only possible if recognition phenomena involve the deep groove side which presents different shapes. However, interaction
patterns involving a third residue are less obvious. Further, the deep groove side of the non-canonical C·A pair in yeast tRNA^Phe (Ψ·C in yeast tRNA^Asp) has been described as forming an ion binding site in tRNAs and in the leadzyme (Wedekind & McKay, 1999). Several rare modifications were also found to block the access to the deep groove side. These modifications are m3C‡ at position 32 and m5C at position 38. Note that m3C‡ has only been detected at position 32 in tRNAs (Auffinger & Westhof, 1998b). In order to test the preceding hypothesis, several biochemical assays can be proposed. For example, a methylation of the N1 site of A38 or of the N3 site of C38 should block access to the shallow groove, while a methylation of the N7 site of A38, the C5 site of C38, the N3 site of C32, or the N3 site of U32 would at least partially hinder access to the deep groove. Interestingly, in the recent crystal structure of a conserved ribosomal RNA/protein domain, a U·A pair adopts the conformation of the C·A pair observed in yeast tRNA^Phe (Conn et al., 1999). Here, the hydroxyl group of a distant adenine residue seems hydrogen bonded to the uridine residue on the deep groove side N3(U)…O2'(A) ≈ 2.6 Å. Although such a hydrogen bond is not compatible with a methylation at the N3 site of the uridine base, it does point to another potential binding site in the deep groove. Furthermore, binding of ribosomal protein residues cannot be excluded.

Conclusions

Biomolecular motifs with specific structure and functions can be recognized by a set of sequence characteristics constituting their molecular signatures. The knowledge of these signatures has many implications such as (i) the ability to recognize motifs with specific structure in sequence databases; and (ii) the possibility to derive a 3D model from the sequence. Up to now, the signature of the tRNA anticodon hairpins comprised the phylogenetically conserved U33 residue associated with a conserved pyrimidine at position 32 and a conserved purine at position 37. A strong preference for a purine at position 38 had also been noted while an almost regular distribution of the four bases is observed at the anticodon positions 35 to 36. At position 34 adenine bases are poorly represented (Grosjean et al., 1982; Auffinger & Westhof, 1998b). An analysis of available data led to the refinement of the signature of the tRNA anticodon loop. tRNA anticodon loops are generally closed by a set of non-canonical isosteric base-pairs involving a single inter-residue hydrogen bond at position 32-38, which can be distributed in three families. Family I (86% of the sequences) comprises the four C·A, U·A, U-C, and C·C pairs, i.e. Y·C/A, base-pairs (interestingly, the U·A pair is not of the Watson-Crick type). Family II (7%) gathers the set of U·U base-pairs, not isosteric to the preceding ones, also observed at position 32-38. Thus, sequences with U32-U38 pair may comprise a tRNA family with distinct properties. Family III (7%) comprises a set of rare base oppositions for which only a limited number of approximately isosteric base-pairs could be tentatively proposed. Besides the structural aspects, it is proposed that conserved patterns associated with the set of the four isosteric 32-38 base-pairs of family I,
could be involved in the formation of base triples mediating tertiary interactions of the tRNA with a ribosomal RNA residue.

Acknowledgements

The authors thank Stephen Cusack for providing coordinate sets for the tRNA anticodon hairpin of structures pte002 and ptr014.

References


Biou, V., Yaremchuk, A., Tukalo, M. & Cusack, S. (1994). The 2.9 \( \hat{A} \) crystal structure of T. thermophilus seryl-tRNA synthetase complexed with tRNA{\textsuperscript{Ser}}. Science, 263, 1404-1406.

Brown, R. S., Dewan, J. C. & Klug, A. (1985). Crystallographic and biochemical investigation of the lea-

d(II)-catalyzed hydrolysis of yeast phenylalanine tRNA. Biochemistry, 24, 4785-4801.


Cusack, S., Yaremchuk, A. & Tukalo, M. (1996a). The crystal structure of T. thermophilus lysyl-tRNA synthetase complexed with E. coli tRNA{\textsuperscript{Glu}} and a T. thermophilus tRNA{\textsuperscript{Aas}} transcript: anticodon recognition and conformational changes upon binding of a lysyl-adenylate analogue. EMBO J. 15, 6321-6334.

Cusack, S., Yaremchuk, A. & Tukalo, M. (1996b). The crystal structure of the ternary complex of T. thermophilus seryl-tRNA synthetase with tRNA{\textsuperscript{Ser}} and a seryl-adenylate analogue reveals a conformational switch in the active site. EMBO J. 15, 2834-2842.


Edited by D. E. Draper

(Received 7 June 1999; received in revised form 26 July 1999; accepted 30 July 1999)