Theor Chem Acc (1999) 101:103–113
DOI 10.1007/s002149800m54

Theoretical
Chemistry Accounts

Regular article

C—H...O and C—H...N interactions in RNA structures*

Maria Brandl, Klaus Lindauer, Michael Meyer, Jürgen Sühnel

Biocomputing, Institut für Molekulare Biotechnologie, Postfach 100813, D-07708 Jena, Germany

Received: 17 May 1998 / Accepted: 4 August 1998 / Published online: 2 November 1998

Abstract. Small molecule studies indicate that C—H...X interactions (X: O,N) constitute weak H-bonds. We have performed a comprehensive analysis of their occurrence and geometry in RNA structures. Here, we report on statistical properties of the total set of interactions identified and discuss selected motifs. The distance/angle distribution of all interactions exhibits an excluded region where the allowed C—H...X angle range increases with an increasing H...X distance. The preferred short C—H...X interactions in RNA are backbone-backbone contacts between neighbour nucleotides. Distance/angle distributions generated for various interaction types can be used for error recognition and modelling. The axial C2′(H)...O4′ and C5′(H)...O2′ interactions connect two backbone segments and form a seven-membered ring that is specific for RNA. An AA base pair with one standard H-bond and one C—H...N interaction has been identified in various structures. Despite the occurrence of short C—H...X contacts their free energy contribution to RNA stability remains to be assessed.

Key words: RNA structure – C—H...X interactions – Hydrogen bonds – Statistical analysis

1 Introduction

With the discovery of catalytically active RNA molecules interest in RNA research has increased and a variety of new RNA structures has become available within the last few years. Therefore it has been claimed that the era of RNA structural biology has really arrived [1–3].

Hydrogen bonds belong to the most important interactions in biopolymer structures [4]. These bonds represent attractive interactions of the type Y-H...X where Y and X are electronegative atoms. In biological macromolecules N and O are usually considered as H-bond donors and acceptors. However, from small molecule studies there is evidence suggesting that C—H...O interactions should be viewed as weak H-bonds as well, given the less electronegative C—H group has appropriate electronic properties [5–7]. As a rule the 'acidity' of the donor C—H group and thus its potential H-bond donor strength increases in passing from sp^3 over sp^2 to *sp* hybridized C atoms and it is also increased when surrounded by electron withdrawing atoms [5, 7–9]. There is now growing evidence that interactions of this type may also be relevant to biopolymer structures [10]. Usually, however, for structure refinement or assessing the structural or functional role of interactions between biopolymer building blocks (amino acids, nucleotides) C—H...X contacts are not taken into account.

In RNA an axial C2′(H)(n)...O4′(n + 1) interaction was recently identified both in X-ray structures and in a molecular dynamics simulation (n: numbering of nucleotides in 5'-3' direction) [11]. Dynamically stable inter-residue C—H...O interactions were observed for the nucleotide pairs U(33)–C(36) and U(33)–U(35) in a molecular dynamics simulation of the anticodon loop of tRNA(Asp) [12]. Leonard et al. [13] and Starikov and Steiner [14] have discussed the possible stabilizing role of the C2(H)...O2 interaction in Watson-Crick AU base pairs. A short C—H...O contact was recently described in a UU base pair in the crystal structure of an RNA hexamer (Calcutta pair) [15]. Finally, intra-residue C-H...O contacts between the purine C8(H) or pyrimidine C6(H) and the backbone O5′ atoms of DNA can be assumed to play a similar role in RNA [16].

It is well known from protein studies that the free energy contribution of standard H-bonds to the protein folding process is still open to discussion [17]. Therefore, we cannot expect to obtain a simple answer for the much weaker C—H...X contacts. What we can do, however, in a first step is to identify possible candidates for an attractive interaction adopting geometrical criteria. These interactions have then to be studied in more detail by approaches taking into account energetic considerations. We have, therefore, performed a comprehensive geometrical analysis on RNA structures determined by X-ray dif-

*Contribution to the Proceedings of Computational Chemistry and the Living World, April 20–24, 1998, Chambery, France

Correspondence to: J. Sühnel

fraction and NMR spectroscopy and report here on the statistical information and discuss selected motifs.

2 Materials and methods

We have prepared five RNA structure sets from protein data bank (PDB) structures [18]:

set I (high-resolution X-ray structures with a resolution better than 2 Å, 6 structures) 157d, 1osu, 1rxb, 259d, 1urn, 255d; set II (X-ray, 23 structures) 157d, 1gid, 1mme, 1osu, 1rna, 1rxa, 1rxb, 1sdr, 1tn2, 1tra, 205d, 255d, 259d, 280d, 283d, 299d, 2tra, 300d, 301d, 3tra, 4tna, 4tra, 6tna; set III (X-ray RNA-protein complexes, 10 structures) 1asy, 1asz, 1gtr, 1gts, 1qrs, 1qrt, 1qru, 1ser, 1ttt, 1urn; set IV (NMR, 38 structures, 311 models) 1afx (13), 1ajf (1), 1ajl (1), 1ajt (1), 1am0 (8), 1anr (20), 1arj (20), 1elh (6), 1etf (1), 1fmn (5), 1guc (30), 1hlx (20), 1ikd (30), 1kaj (1), 1kis (1), 1koc (1), 1kod (1), 1kpd (1), 1mis (1), 1mwg (1), 1pbm (1), 1pbr (1), 1qes (30), 1qet (30), 1rau (1), 1raw (10), 1rht (1), 1rng (5), 1rnk (1), 1scl (6), 1slo (1), 1slp (16), 1tob (7), 1ull (7), 1yfv (1), 1zif (10), 1zig (10), 1zih (10); set V (all structures).

The NMR entries from the PDB may either contain minimized average structures or up to 30 different models. In the latter case we have included all models separately. However, each H-bond in one particular model was multiplied with a weighting factor (1/n), where n is the number of models. Within the NMR set the average structure of an RNA-peptide complex is included (PDB code: 1etf [19]). The related structure with 19 different models (PDB code: 1etg) was removed from the set because it contains unusual distance/angle combinations of C—H...X contacts in the light of the excluded region discussed in the text. The structure sets were analysed using HBexplore [20]. X-ray structures do not contain any H atom coordinates. Therefore HBexplore calculates the position of H atoms according to standard geometrical rules. The H atom coordinates given in the NMR PDB entries were not used. Rather, these coordinates were calculated in the same way as for the X-ray entries. There is no unique rule for naming the oxygen atoms O1P and O2P of the phosphate group. We have therefore scanned all structures with the program chirality kindly provided by P. Slickers and modified the PDB files if necessary. With regard to the plane defined by the atoms O5′, P and O3′ and looking from O5′ to O3′ the atom O1P is always pointing to the right and O2P to the left. Solvent accessible and buried atoms have been identified using an algorithm of Pearl and Honegger with a solvent radius of 1.4 Å and grid point distances of 0.5 Å [21].

3 Results and discussion

The results of our analysis depend in a crucial manner on the accuracy of the experimental structures. Most of the X-ray structures, and in particular the large ones, were resolved at low resolution. Moreover, there is still no consensus on how reliable NMR-derived RNA structures are. Recently, it has been claimed, however, that NMR structures of RNA can be determined with an accuracy of 1.0–1.5 Å [22]. In order to check whether the geometrical parameters of C—H...X interactions differ between crystallographic and NMR structures and between X-ray structures of different resolution we have analysed these structure types separately. Figure 1 shows a comparison of distributions of the C—H...X angle versus the H...X distance for the NMR set, the total X-ray set and the high-resolution X-ray set. Except for the shape of the maximum the distributions of the NMR and total X-ray sets are similar, even though both sets consist of different RNA molecules. In the high-resolution X-ray set the location of the maxima is similar to the NMR and total X-ray sets, yet C—H...X interactions shorter than 2.0 Å are not found. Whether the shorter distances observed both for the less well resolved X-ray and for the NMR structures are due to experimental artefacts or indicate structural features not present in the high-resolution set is open to discussion. We have used the total set for a classification of interaction types, but checked the most frequent types for their occurrence and geometry in the high-resolution structures.

In the following we start out from a very general analysis of all C—H...X interactions and then take steps towards a more specific classification. In Fig. 2 contour plots of the distributions of the angles C—H...X versus the distances H...X of all inter-residue C—H...X contacts with an H...X distance smaller than 3.5 Å are shown. For comparison the corresponding distribution of standard H-bonds is given as well. In both cases excluded regions, which represent forbidden distance/angle combinations, have been observed and found to resemble each other in shape. A similar excluded region is also found for C—H...X interactions and standard H-bonds in proteins (Brandl, Meyer and Sühnel, unpublished results) and for standard H-bonds between small organic molecules and water in high-resolution neutron diffraction studies [23]. For both C—H...X and standard H-bond interactions the accessible angle range increases with an increasing H...X distance (Fig. 2) and the mean angle for a particular distance decreases. For all H...X distances the C—H...X mean angle is smaller than the angle for the standard H-bond interactions. An obvious explanation for this effect is the repulsion between the heavy atoms of the H-bond, which prevents too small H-bond angles. This repulsion can be assumed to be weaker for the C—H...X interaction. In addition, it is usually assumed that the directionality of H-bonds is more pronounced for stronger interactions [7]. The observed distance/angle correlation is relevant for the construction of H-bond potentials [24] and can be used for error recognition in experimental structures. For the NMR structure of the HIV-1 rev peptide-RRE RNA complex, 19 models (PDB code: 1etg) and an averaged structure (PDB code: 1etf) were reported [19]. Our analysis has shown that 23 C—H...X interactions of the NMR structures of 1etg are located far outside the allowed region. The most extreme example is an interaction between C(49):C6 and G(48):O2′ in model 2 with an H...O distance of only 1.3 Å and a C—H...O angle of 22°. This was the only structure for which pronounced deviations have been found. It was removed from the

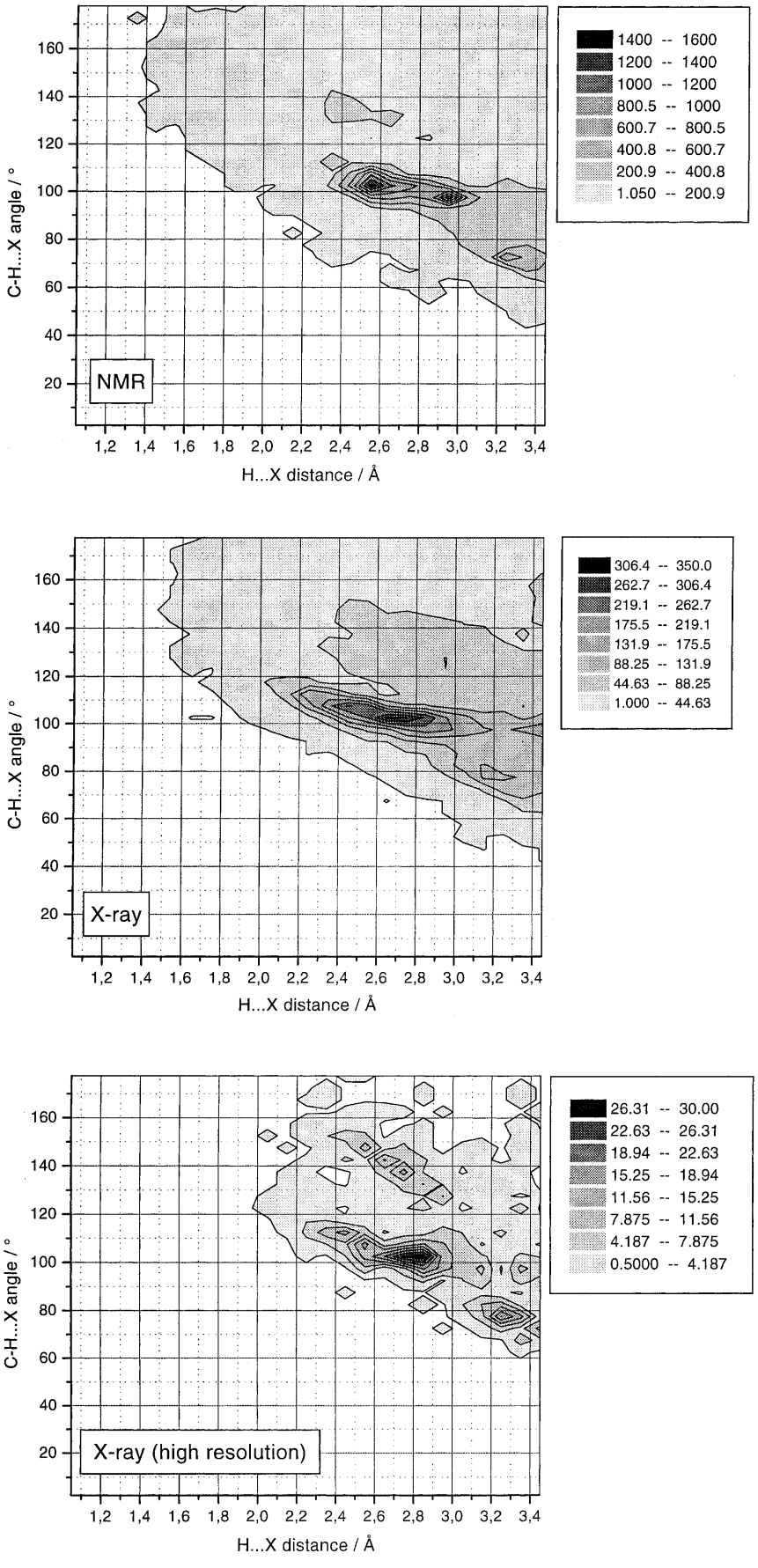


Fig. 1. Contour plots of C—H...X angles vs. H...X distances for inter-residue C—H...X interactions in the NMR, X-ray and high-resolution X-ray sets of RNA structures (X: O,N)

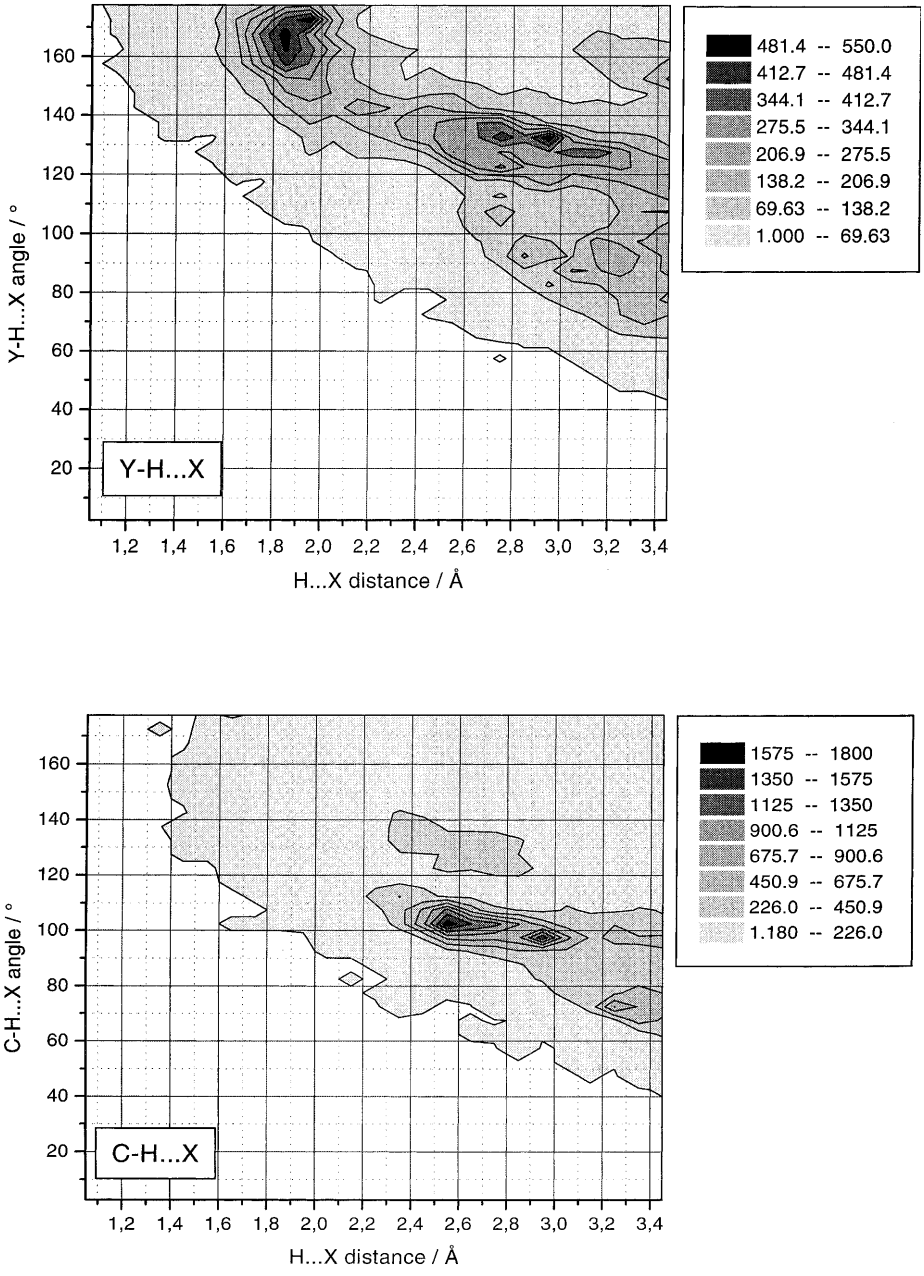


Fig. 2. Contour plots of C(Y)-H...X angles vs. H...X distances for inter-residue C—H...X and standard H-bond Y-H...X interactions in RNA structures (total set: Y,X; O,N)

dataset. On the other hand, C—H...X interactions of the average structure (1etf) fit well into the allowed region. The allowed area is slightly larger for the NMR set as compared to the X-ray data (Fig. 1). In addition to the allowed region for all C—H...X interactions the two-dimensional distributions of particular types of interactions shown below can be invoked for error recognition as well.

The contours describing the border between the forbidden and allowed regions are roughly similar for C—H...X and standard H-bond contacts, except for the fact that the standard H-bond distribution extends to shorter distances. By contrast the distribution maxima of standard H-bond and C—H...X interactions differ substantially. Whereas the maximum of the standard H-bond distribution is located at 1.8–2.0 Å and 160–180°, the C—H...X distribution displays a maximum within the distance range 2.4–2.8 Å and an angle range of 100–110°. The maximum of the standard H-bond distribution is mainly due to interactions between base donor and acceptor atoms and the C—H...X maximum primarily originates from contacts between backbone atoms.

We have selected all C—H...X interactions shorter than a certain cutoff distance. The usage of cutoffs for identifying H-bond interactions has been questioned [4, 7]. The basic argument was that it is not correct to assume that H-bonds become van der Waals interactions at longer distances. On the other hand, cutoff values are necessary for an automatic scanning of structures. We have therefore varied the cutoff values if possible in order to find out how this affects the results obtained. If

single cutoffs had to be used then a value of 3.0 Å was adopted. It is slightly larger than the van der Waals distance of 2.7 Å for an H...O interaction and is used by others as well [25]. In this regard it is interesting to point out that recently for an interaction of ethynyl and carbonyl groups with an H...O separation of 2.92 Å and a C—H...O angle of 130° a 25 cm^{-1} red shift of the ν_{CH} stretching wavenumber has been found [26]. Even though in this case the C—H group is more acidic than the ones found in biopolymers it clearly shows that there may be a weak H-bond interaction for distances larger than the van der Waals limit.

Angle cutoffs were not applied in the first step. For the discussion of the possible H-bond potential of a C—H...X interaction the angle geometry was taken into account, however. Table 1 shows the manner in which the number of contacts identified depends on the cutoff distance. The total number varies between approximately 23 000 contacts for a cutoff value of 3.5 Å and about 4500 for 2.5 Å. Our further classification is based on the location of the donor and acceptor atoms, which can be either in the sugar-phosphate backbone (B) or in the base (b) part of RNA. We have therefore distinguished between the categories BB, Bb, bB and bb, where the donor atom is represented by the first and the acceptor atom by the second letter. In addition, the fractions of interactions between nucleotides that are next neighbours in sequence (nn) have been calculated.

The great majority of C—H...X contacts identified occur between backbone donor and acceptor atoms (BB). The fractions for a cutoff distance of 3.0 Å are 86% (BB), 4% (Bb), 5% (bB) and 5% (bb). Contrary to the total number of interactions these values are only slightly dependent on the cutoff distance. In addition, it has been found that 96% of the BB interactions link next neighbours. For the other interaction types the fractions of next-neighbour contacts vary between 32 and 66%. In Table 2 a list of the most frequent interaction types and in Fig. 3 the corresponding two-dimensional distance/angle distributions for the next-neighbour backbone-backbone interactions are shown. For each interaction type separate distributions for X-ray and NMR data have been generated and found to be in good agreement. In addition, we have calculated the corresponding distributions for the high-resolution X-ray set, which includes five A-RNA helix structures and an RNA-protein complex with double helical and single-stranded regions in the RNA part. For the C2′(H)...O4′, C3′(H)...O5′, C2′(H)...O5′, C3′(H)...O1P, C5′(H)...O2′, C5′(H)...O1P and C5′(H)...O3′ interactions the data of the high-resolution set turned out to be within the range of the distribution obtained for the total set. C3′(H)...O2P interactions shorter than 3.5 Å were not found in the helical structures of the high-resolution set, yet rather short interactions of this type occur in single-stranded regions of the high-resolution RNA-protein complex.

The distributions of the C5′(H)...O2′ and C2′(H)...O4′ contacts are very similar in appearance. They cover an angle range between 90 and 180° and extend to smaller distances than the other interactions. The angle ranges are relatively large and thus indicate a certain flexibility of the interactions. By contrast, the C2′(H)...O5′, C3′(H)...O2P, C3′(H)...O5′ and C5′(H)...O3′ interactions are restricted to longer H...O distances and C—H...O angles below 120°. The accessible angle range is only about 20° and thus very small compared to C5′(H)...O2′ and C2′(H)...O4′. The interaction C3′(H)-O1P bears resemblance to the latter types even though it can adopt slightly smaller distances below 2.0 Å. With C—H...X angles between 100 and 150°, the interaction C5′(H)...O1P is somewhere in between the two types of distributions. These data constitute a kind of standard which can be used for modelling and refinement. In addition they provide information on RNA backbone flexibility. From the H-bonding perspective the interactions C5′(H)...O2′ and C2′(H)...O4′ are the most promising candidates. Their C—H groups and acceptor atoms are separated by six and seven covalent bonds and this can explain the flexibility observed. In the cases of the other contacts, donors and acceptors are only separated by three to five bonds. As mentioned in the introduction, the C2′(H)...O4′ interaction has already been identified by Auffinger and Westhof in molecular dynamics simulations and by a subsequent analysis of X-ray structures of RNA and A-DNA [11]. Whereas the C2′(H)...O4′ contact occurs between buried donor and acceptor atoms both C5′ and O2′ are in most cases located at the solvent accessible surface. Therefore interactions to the solvent may compete with the C5′(H)...O2′ interaction. Here, it is interesting to point out that C5′ has two H atoms which can both be involved in H-bond interactions. The orientation of the H in the 2′-OH group may affect a possible C5′(H)...O2′ interaction. The above-mentioned molecular dynamics

Table 1. Occurrence of inter-residue C—H...X interaction types in RNA structures (total set) for different H...X cutoff distances: BB backbone(donor)-backbone(acceptor); Bb backbone-base; bB base-backbone; bb base-base; the donor atom is listed first; nn fraction of next neighbour interactions for a particular interaction type; no angle cutoffs

Cutoff distance	3.5 Å		3.0 Å		2.75 Å		2.5 Å	
Total number	23057		13248		8824		4517	
Interaction type		nn		nn		nn		nn
BB	76%	0.95	86%	0.96	88%	0.96	88%	0.97
Bb	8%	0.57	4%	0.42	3%	0.34	3%	0.32
bB	8%	0.75	5%	0.66	5%	0.65	6%	0.70
bb	9%	0.52	5%	0.32	4%	0.26	3%	0.24

Table 2. The most frequent types of inter-residue C—H...X interactions in RNA structures (total set): H...X cutoff distance 3.0 Å; nn fraction of next-neighbour interactions for a particular interaction type; only interaction types with fractions larger than 10% are given; no angle cutoffs

Interaction type	Total	Fraction	nn
backbone-backbone			
C3′(H)...O1P	2628	23%	0.99
C5′(H)...O3′	1925	17%	1.00
C2′(H)...O4′	1892	17%	0.99
C5′(H)...O2′	1633	14%	0.98
backbone-backbone non-next-neighbour			
C1′(H)...O2′	54	12%	
C2′(H)...O1P	48	11%	
C4′(H)...O′	41	9%	
C1′(H)...O4′	33	8%	
base(donor)-backbone(acceptor)			
C6(H)...O2′	115	18%	0.72
C8(H)...O2′	103	16%	0.89
C6(H)...O3′	51	8%	0.94
C8(H)...O3′	48	7%	1.00
backbone(donor)-base(acceptor)			
C1′(H)...N3	87	16%	0.75
C2′(H)...N7	79	14%	0.69
C1′(H)...O2	69	12%	0.38
C1′(H)...N1	55	10%	0.38
base-base			
C2(H)...O2	239	36%	0.06
C5(H)...N7	67	10%	0.97
C2(H)...N3	41	6%	0.24
C8(H)...O4	41	6%	0.05

simulations have shown that the preferred orientations of the 2′-OH group in a standard C3′-endo conformation are towards the O3′ and O4′ atoms of the same residue and the shallow groove base atoms [11]. Thus, the H of the 2′-OH group should not prevent a short C5′(H)...O2′ interaction. In addition, molecular dynamics simulations on an RNA aptamer structure have shown that short C2′(H)...O4′ interactions can be accompanied by short C5′(H)...O2′ contacts (Schneider and Sühnel, unpublished results). They are characterized by backbone torsion angles $\alpha = 270$–$300°$ and γ around 60° characteristic for A-RNA. In addition, however, short C2′(H)...O4′ interactions can also occur with longer C5′(H)...O2′ interactions. In the latter case α varies continously between 60 and 360° and γ is approximately equal to 180°. This observation requires further investigations.

The C2′(H)...O4′ and the C5′(H)...O2′ interactions connect two segments of the sugar phosphate-backbone to a seven-membered ring. As DNA lacks the 2′-OH group, this structural motif is specific for RNA. Hence, it is tempting to speculate that it may contribute to the different structural features of DNA and RNA. In Fig. 4 an example is shown from a high-resolution RNA duplex [27].

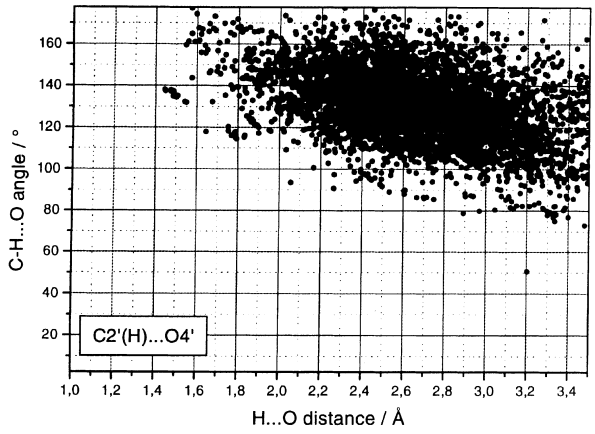
The shape of the total distribution for C—H...X contacts (Fig. 2) is dominated by backbone-backbone interactions between next-neighbour residues. Ninety percent of the interactions found within the main maximum at 2.4–2.8 Å and 90–110° belong to the types

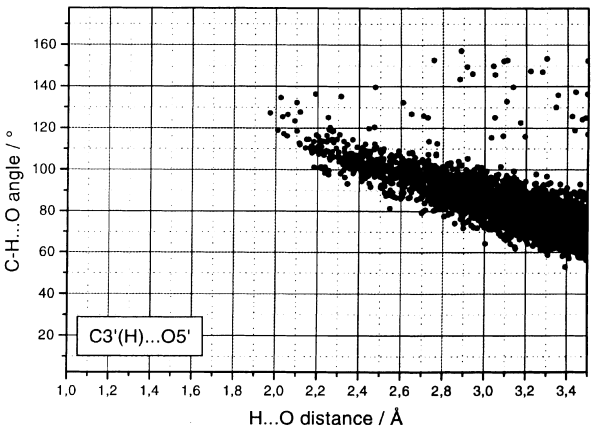
Fig. 3. Scatterplots of C—H...X angles vs. H...X distances for the eight most frequent next-neighbour backbone-backbone contacts in RNA structures (total set)

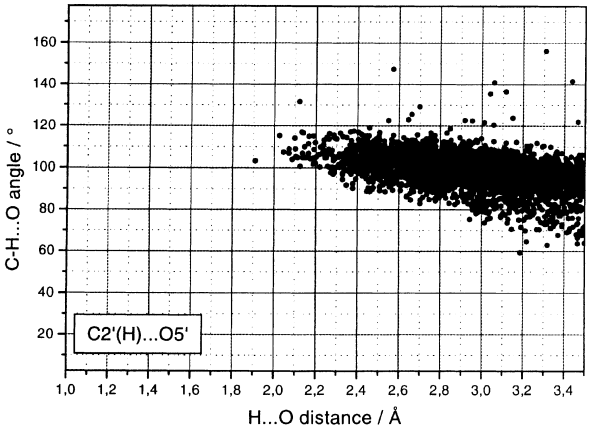
C2′(H)...O5′, C3′(H)...O2P, C3′(H)...O5′, C5′(H)...O3′ and C3′(H)...O1P, for which C—H groups and acceptor atoms are separated by less than five covalent bonds. The more flexible interactions C2′(H)...O4′ and C5′(H)...O2′ constitute about 65% of the second maximum at 2.4–2.8 Å and 120–140°.

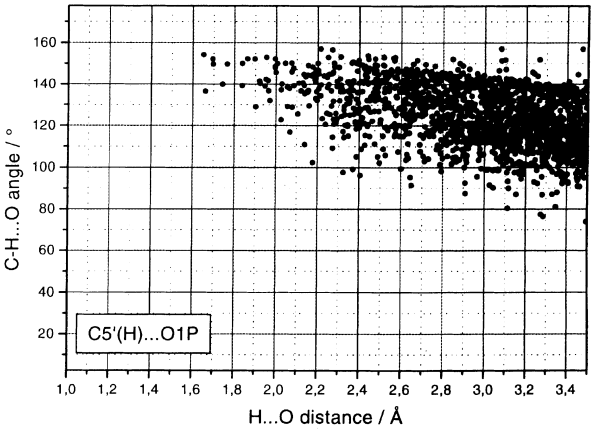
Besides the next neighbour backbone-backbone interactions described above, backbone-backbone interactions have also been found to connect nucleotides that are not next neighbours in sequence (Table 2). Five of the six most frequently occurring interaction types were found to connect predominantly nucleotides with a considerable difference in the nucleotide sequence position (tertiary interactions). The exception is the C2′(H)...O1P interaction, which has often been found in loops where it occurs between a nucleotide *n* which contains the C—H donor group and a nucleotide n + 2 which contains the O1P acceptor group. An interesting motif with a combination of next neighbour and non-neighbour backbone-backbone interactions is found in the structure of the complex between tRNA(Asp) and its cognate synthetase (PDB code: 1asy [28]). Upon protein binding the canonical U loop containing the anticodon sequence is disrupted, the bases of nucleotides 633–637 are unstacked and oriented towards the RNA-protein interface. For the nucleotides 636–638 the backbone folds back on itself and forms a small internal loop. Within the loop one standard H-bond between O2′ of U(235) and O2 of C(238) and a network of short C—H....O interactions mainly of the BB type is observed (Fig. 5).

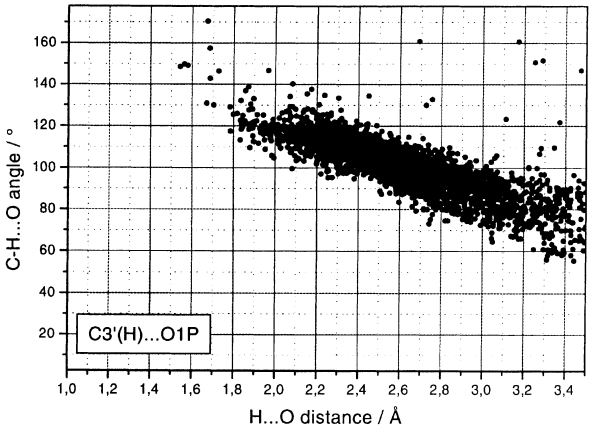
The occurrence of intra-residue interactions between the base atoms C8 in purines or C6 in pyrimidines and the backbone atom O5′ is well known [16]. According to our analysis short inter-residue interactions between the very same base atoms and O2′ are the most frequent interactions of the bB type (Table 2). In an ideal A-RNA conformation the O2′ atom of the previous nucleotide and O5′ are located on different sides of the base plane. With 2.35 and 2.55 Å the intra-residue C6(8)(H)...O5′ distances are much smaller than the values of 3.45 and 3.65 Å found for the C6(8)(H)...O2′ contacts. Yet, deviations from the ideal A-form can cause the C6(C8) hydrogen atoms to get closer to O2′ than to O5′. C6(C8)(H)...O2′ contacts shorter than 3.0 Å have been identified in both NMR and X-ray structures. However, the majority of interactions with especially short H...X distances and large C—H...X angles have been found in NMR structures, with C8(H) as the donor group. For example, in the NMR structure of a pseudoknot that causes frameshifting there are three short C—H...X contacts between C8(H)(n) and O2′(H)(n-1) in a loop region: A(27)–A(26), 1.78 Å; A(25)–A(24), 1.68 Å; A(24)–C(23), 2.02 Å (PDB code: 1rnk [29], Fig. 6). In all these cases the competing intra-residue contacts to O5′ are much longer. A further example is found in the

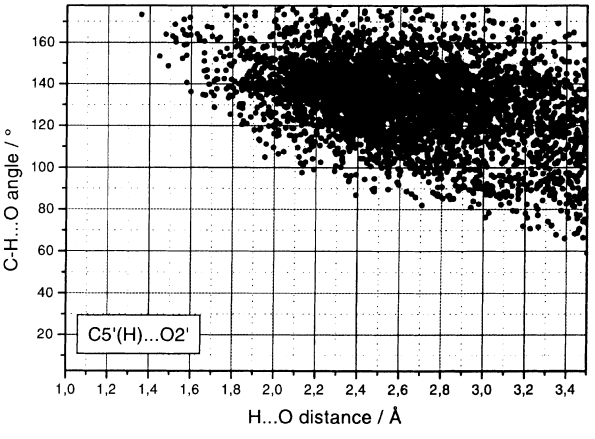

C-H...O angle / °
H...O distance / Å
C2'(H)...O4'

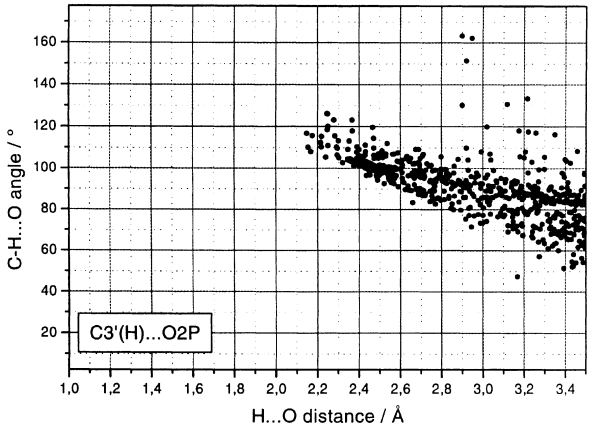

C-H...O angle / °
H...O distance / Å
C3'(H)...O5'

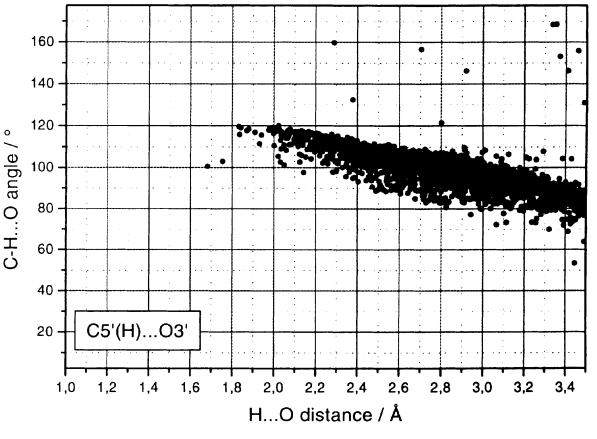

C-H...O angle / °
H...O distance / Å
C2'(H)...O5'


C-H...O angle / °
H...O distance / Å
C5'(H)...O1P


C-H...O angle / °
H...O distance / Å
C3'(H)...O1P


C-H...O angle / °
H...O distance / Å
C5'(H)...O2'


C-H...O angle / °
H...O distance / Å
C3'(H)...O2P


C-H...O angle / °
H...O distance / Å
C5'(H)...O3'

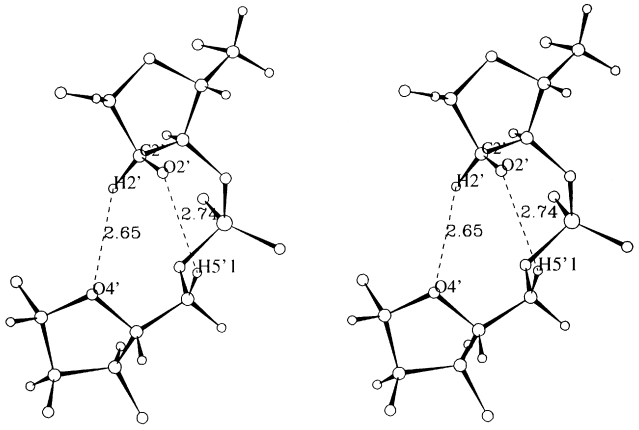


Fig. 4. The C2′(H)...O4′/C5′(H)...O2′ backbone motif for neighbour nucleotides in RNA. The example is taken from the crystal structure of $R(CCCCGGGG)_2$ (PDB code: 259d [27]). The Figures displaying RNA structure elements were generated with InsightII from Molecular Simulations, Inc. If the motifs reported are checked by other graphics programs slightly different distances and angles may be obtained. The hydrogen atom of the 2′-OH group was omitted because its orientation cannot be reliably inferred from the experimental structure

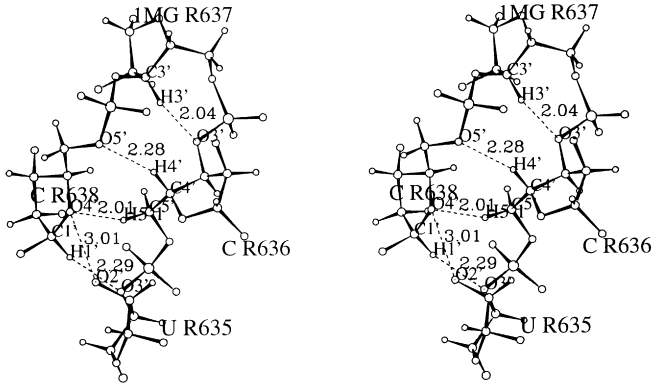


Fig. 5. Backbone-backbone C—H..O contacts between next-neighbour and non-next neighbour in an RNA loop motif of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp) (PDB code: 1asy [28]). The PDB numbering U(635), C(636), 1MG(637), C(638) corresponds to U(35), C(36), 1MG(37), C(38) in the local RNA numbering. R is the chain identifier. The hydrogen atom of the 2′-OH group was omitted because its orientation cannot be reliably inferred from the experimental structure

sarcin/ricin loop NMR structure (PDB code: 1scl) [30]. All models exhibit a short C8-O2′ contact at the 5′-end, which links the base part of the second nucleotide to the backbone of the first. Common to these two examples is the fact that in comparison to the A-helical conformation the bases are moved towards the RNA backbone. Finally, in all NMR models of an unusually stable parallel-stranded RNA tetraplex (PDB-code: 1rau [31]) short C8(H)...O2′ contacts are found between G(3) and G(2) in all four chains. In contrast to the examples mentioned above the bases are oriented perpendicular to the helix axis. In this case the interaction is due to the small spacing between base planes 2 and 3, the non-A-

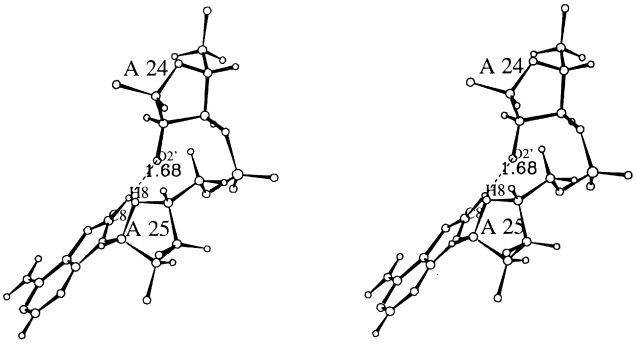


Fig. 6. Base-backbone C—H...O interaction in an RNA pseudoknot structure (PDB: 1rnk [29]). The hydrogen atom of the 2′-OH group was omitted because its orientation cannot be reliably inferred from the experimental structure

like backbone torsional angles and the unusual ribose conformation of G(2).

For backbone-base C—H...X interactions the dominating donor atom is C1′ (Table 2). The four most frequently occurring interaction types shorter than 3.0 Å are C1′(H)...N3, C2′(H)...N7, C1′(H)...O2 (cytosine) and C1′(H)...N1 (adenine). The first two interactions are primarily observed between neighbour nucleotides, the latter two between non-neighbours. An example, which mediates a tertiary contact is shown in Fig. 7. In the hammerhead structure the nucleotides A(60) in chain A and U(161) in chain B are connected by a backbone-backbone standard H-bond between O2′H and O2′ (PDB code: 1mme [32]). In addition, there are two short C—H...O contacts between the donor groups C1′H and C4′H, and O2. In the group I intron structure a very short C1′(H)...N3 interaction between neighbour nucleotides is found (PDB code: 1gid [33]). Here, the hydrogen of C1′ in A(122) makes contact with N3 in A(123) (H...N3 distance: 1.83 Å; C1′-H...N3 angle: 129°). However, such short contacts seem to be the exception rather than the rule. The shortest backbone-base C—H...X interaction in the high-resolution X-ray set has a length of 3 Å and has been found in an RNA helix with two G(anti).A(anti) base pairs (PDB code: 157d [13]). The interaction occurs between C15:C2′ and G(16):N7 within the B chain, where G is part of one of the GA base pairs.

The largest fraction of base-base C—H...X contacts is due to C2(H)...O2 interactions (Table 2), most of which occur in the Watson-Crick AU base pair. In Fig. 8 the corresponding distance/angle distribution is shown. The preferred angle range is between 110 and 140° and the distances exhibit a maximum between 2.6 and 2.8 Å. The geometrical parameters obtained from the few C2(H)...O2 contacts in the high-resolution data-set are in line with the results obtained from the total structure set. Whether this interaction contributes to the stability of AU base pairs remains to be clarified. Recent ab initio calculations on model systems suggest that the C2(H)...O2 interaction in AU base pairs can be considered as weakly attractive [14]. In any case this C2(H)...O2 interaction is expected to play a minor role because of the additional two standard H-bonds.

It is well known that non-canonical base pairs are important elements of RNA structure. Usually it is assumed that they all contain two direct inter-base H-bonds [4]. Recently, however, examples with only one direct standard H-bond interaction have been found. They include water-mediated GU, AG and GG pairs in loop E of 5S *r*RNA, CU pairs and the already mentioned UU base pair with a short C—H...O contact in an RNA hexamer with a UU overhang [15, 34, 35]. In the latter base pair a standard H-bond N3(H)...O4 is accompanied by a short C5(H)...O4 interaction with an H...O length of about 2.25 Å. We have observed a non-canonical AA base pair that resembles the Calcutta base pair in the respect that the two bases are linked by one standard H-bond and one short C—H...X contact. The AA base pair occurs in the group I ribozyme domain (PDB code: 1gid [33]), in tRNA(Glu) complexed with the cognate synthetase (PDB codes: 1qrs, 1qrt, 1qru, 1gtr, 1gts [36, 37]) and in one model of the RNA-FMN aptamer complex (PDB code: 1fmn [38]). The standard H-bond is formed between the N6(H2) group of the one adenine and the N3 nitrogen of the second adenine; C2(H) of the second adenine and N7 of the first adenine participate in a C—H...N contact, which covers an H...N distance range between 1.9 and 2.8 Å. In the group I intron structure (1gid) the interaction occurs twice. The two AA pairs directly follow each other in the nucleotide sequence [A(206)–A(114), A(207)–A(113)], (Fig. 9). It is interesting to point out that these AA pairs together with a GU wobble pair represent a motif that is essential for 5′-splice site selection [39]. In glutaminyl transfer RNA the AA base pair is involved in a base triple formed by a symmetric AA pair [A(45)–A(13)] linked by two standard H-bonds of the type N6(H)...N1 and the additional base pair with the short C2(H)...N7 contact [A(22)–A(13); 1qru, 1qrt, 1qrs, 1gts, 1gtr]. A further interesting base-base C—H..X motif is found in an unusually stable RNA tetraplex structure (PDB code: 1rau [31]). The tetraplex consists of four central G quartets and two U quartets at both ends. Whereas one of the U quartets is linked by standard H-bonds, the other one shows an interesting C—H...X motif. The four U bases are connected by eight C6(H)...O4 and C5(H)...O4 interactions with H...X distances below 2.5 Å. In addition, there are four standard H-bonds of the type N3(H)...O2′.

In 24 of the 33 structures of the crystallographic dataset water molecules are included. We have found 615 contacts with an H...O distance of less than 3.0 Å. For water the H atom coordinates cannot be calculated from

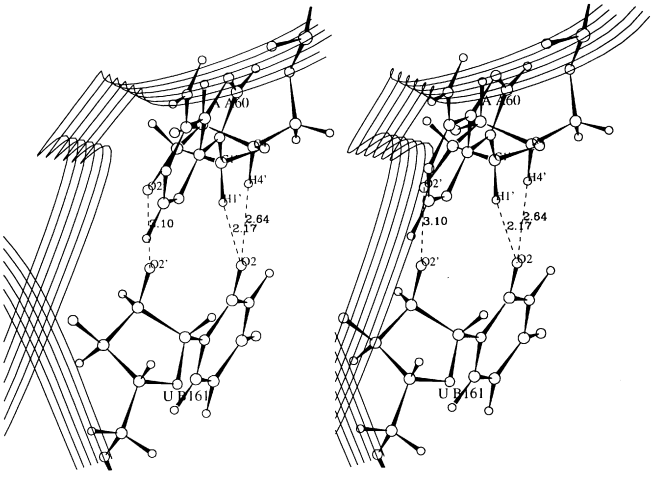


Fig. 7. Tertiary C—H...X interactions between the donor groups C4'-H and C1'-H of nucleotide A(60) and the acceptor O2 in nucleotide U(161) of a hammerhead structure (PDB-code: 1mme [32]). The hydrogen atom of the 2′-OH group was omitted because its orientation cannot be reliably inferred from the experimental structure

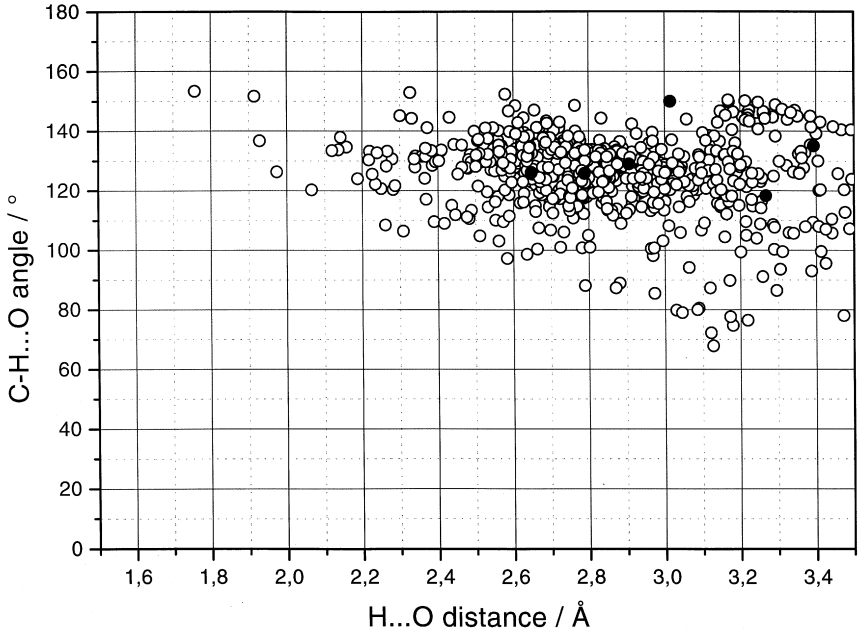


Fig. 8. Scatterplot of C—H...X angles vs. H...X distances for C2(H)...O2 interactions in Watson-Crick AU base pairs (empty circles: total set; filled circles: high-resolution X-ray set)

the positions of the oxygen atoms. Therefore, short C—H...O contacts may be found, for which the water oxygen lone pairs are pointing away from the C—H group. The following data should be seen under this limitation. The most frequently occurring donor atoms are pyrimidine C5, C1′ and C5′ (Table 3). For the base C5 donor atom cytosine water contacts are more frequent than interactions between uracil and water. For about one-third of those C—H...O contacts between C5 of cytosine and water that are shorter than 3 Å, there is also a standard H-bond between N4(H) and water. The preference of C1′ for contacts to water is interesting because this donor is only rarely involved in short C—H...X contacts within RNA. By contrast C5′ has been identified as a candidate for next-neighbour interactions with O2′. Molecular dynamics simulations on the hydration of C—H groups in RNA have shown that pyrimidine bases C5(H) water contacts should be considered as bona fide H-bonds [40]. This observation can now be supplemented by the fact that in experimental structures C5 indeed belongs to the preferred donor groups for C—H...X water interactions. On the other hand, it is likely that in cytosine the neighbour N4(H) donor group may affect the C—H...O interaction.

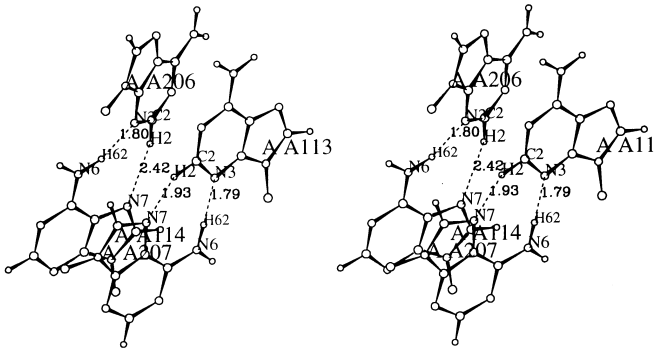


Fig. 9. Non-canonical AA base pairs with one standard H-bond and one C2(H)...N7 interaction in the group I ribozyme structure (PDB code: 1gid [33])

Table 3. Occurrence of C—H...O(water) contacts in X-ray RNA structures

H...O cutoff distance	3.5 Å	3.0 Å	2.5 Å
donor			
C1′	278	145	38
C5′	269	109	40
C5	243	131	31
C3′	168	76	13
C4′	153	53	14
C2′	122	33	12

McDonald and Thornton have demonstrated that in proteins almost all N-H and O-H groups not exposed to the solvent accessible surface are involved in hydrogen bonding within the protein [41]. For C—H...X interactions it has been claimed that a possible role might be to satisfy acceptor groups if no standard donor groups are available [4]. Therefore, it is interesting to study to what extent the C—H donor groups are satisfied by H-bond acceptors. In a first step we have calculated the fractions of buried and exposed C—H donors according to a standard algorithm (Table 4). It is found that only the ribose C2′ and C3′ donors are mainly buried. For C6 and C8 base donors the fractions of exposed cases are 58% and 71% and all other donor groups are more than 84% exposed. We have identified all donor groups not exposed to the surface and then calculated the percentage of these groups involved in at least one inter-residue C—H...X interaction with an H...X distance shorter than 3.0 Å. For the mainly exposed donors the number of groups available for the analysis is rather small, however. We have found three types of donor groups. For C2′, C3′, C5′ and C2 more than 88% of the buried C—H groups are involved in C—H...O or C—H...N interactions. For C1′, C4′ and C5 the fraction of such interactions ranges from 48–72%, and for C6 and C8 the corresponding values are below 30%. The latter donors are, however, known to be involved in intra-residue contacts. The fractions of C6 or C8 donors involved either in intra-residue interactions with O5′ or in inter-residue contacts are 89% and 88%, respectively. These percentage values are corrected for the fact that a fraction of donors may form both intra- and inter-residue contacts. So we can conclude that for six out of the nine C—H donor types the fraction of buried C—H donors with an O or N acceptor atom within a distance of 3 Å is larger than 88%. For the remaining three donor types the corresponding value is between 48 and 72%.

Table 4. Exposed, buried and satisfied C—H donor groups in RNA structures

Donor	Total number of donors	Fraction of donors satisfied	Fraction of exposed donors	Total number of buried donors	Fraction of buried donors satisfied	Total number of exposed donors	Fraction of exposed donors satisfied
C1′	3183	12%	84%	511	48%	2672	5%
C2′	3180	75%	21%	2516	88%	664	24%
C3′	3179	90%	22%	2470	97%	709	65%
C4′	3185	11%	93%	208	72%	2977	7%
C5′	3186	78%	98%	74	90%	3112	78%
C2	661	50%	86%	95	95%	566	42%
C5	1443	18%	93%	100	67%	1343	15%
C6	1498	12%	58%	623	15%	875	11%
C8	1682	15%	71%	488	27%	1194	10%

4 Conclusion

We have identified a large number of C—H...X interactions shorter than 3 Å in RNA structures. If one adopts the widely used assumption that interactions of this type have to be viewed as weak H-bonds they should be taken into account in RNA structure refinement, analysis and modelling. However, it has to be added that the free energy contribution of C—H...X contacts to the stability of RNA remains to be assessed.

Acknowledgements. We are grateful to the Thüringer Ministerium für Wissenschaft, Forschung und Kultur for financial support.

References

1. Ramos A, Gubser CC, Varani G (1997) Curr Opin Struct Biol 7: 317
2. Uhlenbeck OC, Pardi A, Feigon J (1997) Cell 90: 833
3. Doudna JA, Cate JH (1997) Curr Opin Struct Biol 7: 310
4. Jeffrey GA, Saenger W (1991) Hydrogen bonding in biological structures. Springer, Berlin
5. Desiraju GR (1991) Acc Chem Res 24: 290
6. Taylor R, Kennard O (1982) J Am Chem Soc 104: 5063
7. Steiner T (1997) Chem Commun 727
8. Koch U, Popelier PLA (1995) J Phys Chem 99: 9747
9. Rovira MC, Novoa JJ, Whangbo MH, Williams JM (1995) J Chem Phys 200: 319
10. Wahl MC, Sundaralingam M (1997) Trends Biochem Sci 22: 97
11. Auffinger P, Westhof E (1997) J Mol Biol 274: 54
12. Auffinger P, Louise-May S, Westhof E (1996) J Am Chem Soc 118: 1181
13. Leonard GA, McAuley-Hecht K, Brown T, Hunter WN (1995) Acta Cryst D51: 136
14. Starikov EB, Steiner T (1997) Acta Cryst D53: 345
15. Wahl MC, Rao ST, Sundaralingam M (1996) Nature Struct Biol 3: 24
16. Rubin J, Brennan T, Sundaralingam M (1972) Biochemistry 11: 3112
17. Sippl MJ (1996) J Mol Biol 260: 644
18. Bernstein FC, Koetzle TF, Williams GJ, Meyer EE Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) J Mol Biol 112: 535
19. Battiste JL, Mao H, Rao NS, Tan R, Muhandiram DR, Kay EL, Frankel AD, Williamson JR (1996) Science 273: 1547
20. Lindauer K, Bendic C, Sühnel J (1996) Comput Appl Biosci 12: 281
21. Pearl CH, Honegger A (1983) J Mol Graphics 1: 9
22. Allain FH, Varani G (1997) J Mol Biol 267: 338
23. Savage HFJ, Finney JL (1986) Nature 322: 717
24. Hooft RWW, Sander C, Vriend G (1996) Proteins 26: 363
25. Berger I, Egli M, Rich A (1996) Proc Natl Acad Sci USA 93: 12116
26. Steiner T, Lutz B, Van der Maas J, Schreurs AMM, Kroon J, Tamm M (1998) Chem Commun 171
27. Egli M, Portmann S, Usman N (1996) Biochemistry 35: 8489
28. Ruff M, Krishnaswamy S, Boeglin M, Poterszman A, Mitschler A, Podjarny A, Rees B, Thierry JC, Moras D (1991) Science 252: 1682
29. Shen LX, Tinoco I Jr (1995) J Mol Biol 247: 963
30. Szewczak AA, Moore PB (1995) J Mol Biol 247: 81
31. Cheong C, Moore PB (1992) Biochemistry 31: 8406
32. Scott WG, Finch JT, Klug A (1995) Cell 81: 991
33. Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA (1996) Science 273: 1678
34. Corell CC, Freeborn B, Moore PB, Steitz TA (1997) Cell 91:705
35. Holbrook S, Cheong C, Tinoco I Jr, Kim SH (1991) Nature 353: 579
36. Rould MA, Perona JJ, Steitz TA (1991) Nature 352: 213
37. Perona JJ, Rould MA, Steitz TA (1993) Biochemistry 32: 8758
38. Fan P, Suri AK, Fiala R, Live D, Patel DJ (1996) J Mol Biol 258: 480
39. Strobel SA, Ortoleva-Donnelly L, Ryder SP, Cate JH, Moncoeur E (1998) Nature Struct Biol 5: 60
40. Auffinger P, Louise-May S, Westhof E (1996) Faraday Discuss Chem Soc 103: 151
41. McDonald IK, Thornton JM (1994) J Mol Biol 238: 777