

HBexplore - A new tool for identifying and analyzing hydrogen bonding patterns in biological macromolecules

Klaus Lindauer, Cezar Bendic¹ and Jürgen Sühnel²

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

¹ *On leave from the Institute of Physical Chemistry, University of Bucharest, Romania*

² *To whom reprint requests should be sent (e-mail: jsuehnel@imb-jena.de).*

Running title:

H-bond analysis with HBexplore

Key words:

hydrogen bond, biological macromolecules, RNA, secondary structure, tertiary interaction

Abstract

The program HBexplore is a new tool for identifying and analyzing hydrogen bonding patterns in biological macromolecules. It selects all potential hydrogen bonds according to geometrical criteria. The hydrogen bond table can then be subjected to further automatic or interactive analysis tools. These tools include the calculation of mean values and distributions of geometrical hydrogen bond parameters for parts of a single structure, for complete single structures and for structure sets, the classification of each H-bond according to the participation of backbone, sidechain or base, ligand and water parts of nucleic acids or proteins, identification of Watson-Crick nucleotide pairs and of H-bonded pairs of equal nucleotides, the calculation of the mean number of H-bonds per residue and of the fraction of potential donor and acceptor atoms involved in H-bonds. HBexplore generates further automatically a H-bond residue interaction table. This table lists for all residues of the structure the other residues, ligands or water molecules directly connected via an H-bond. By means of a binary tree search algorithm this table is then converted into a H-bond cluster table. Clusters are understood here as an uninterrupted network of H-bonded residues. For nucleic acids secondary structures and tertiary interactions are automatically derived from the hydrogen bonding pattern. HBexplore is applied to two example RNA structures, a pseudoknot and a hairpin. It provides a comprehensive listing of individual hydrogen bonds and statistical information for larger structure sets. In addition, it can identify interesting new H-bond motifs. One example is a penta-nucleotide base-base H-bond interaction motif in the RNA pseudoknot. HBexplore is intended to contribute both to the elucidation of general principles of the architecture of biological macromolecules and to prediction and refinement of single structures.

Introduction

The hydrogen bond is one of the most important intra- and interatomic interactions in biological macromolecules (Jeffrey and Saenger, 1991). In the early days of structural biology hydrogen bonding considerations were of utmost importance both for extracting the model of a double-helical DNA from fiber diffraction data (Watson and Crick, 1953) and for proposing the α -helix and β -sheets as fundamental structural elements of proteins (Pauling et al., 1951). Further, the hydrogen bonding interaction between water and biological macromolecules can hardly be overestimated (Jeffrey and Saenger, 1991). In addition to numerous experimental and theoretical studies on individual structures a few attempts have been made to derive more general patterns and trends from a larger set of structures (Baker and Hubbard, 1984; Ippolito et al, 1990; Stickle et al., 1992; Bordo and Argos, 1994; McDonald and Thornton, 1994; Mandel-Gutfreund et al., 1995; Derewenda et al., 1995). For proteins, one of the results was that almost all of the potential donor and acceptor atoms are involved in H-bonds and that the major part of H-bonds is within the backbone. For nucleic acids there are no comparable

investigations. This is due to the fact that only a small number of RNA structures is known so far. Further, for DNA it is often assumed that hydrogen bonding is nothing more than standard Watson-Crick base pairing. This situation is currently changing. Due to the increased interest in RNA the still small number of known RNA 3D structures has already increased and will increase further in the near future. One of the most exciting new result in this field was the report of the first hammerhead ribozyme structure in 1994 (Tuschl et al., 1994; Pley et al., 1994). The RNA architecture is not simply a collection of helices but consists of further interesting structural elements like pseudoknots, tetraloops, non-canonical basepairs, coaxial stacking regions, parallel and trinucleotide interactions (Chastain and Tinoco, Jr., 1994). In the DNA field an increasing number of unusual structures like triplexes and quadruplexes becomes available (Joshua-Tor and Sussman, 1993). Further, the number of solved 3D structures of complexes between proteins and DNA or RNA is steadily increasing (Nelson, 1995; Moras and Poterszman, 1995). Thus, even though the number of known nucleic acid structures is still much smaller than for proteins we can now begin to search for general hydrogen bonding patterns in nucleic acids and their complexes with proteins and small molecule ligands.

Existing tools for H-bond analysis like the corresponding modules in structure determination or modeling packages do not have the flexibility required. They are often especially designed for proteins and will run into difficulties with nucleic acids. Further, their capabilities of statistical analysis and classification do not fulfil our requirements. Even the excellent program HBPLUS (McDonald and Thornton, 1994) is limited to a selection of potential H-bonds and has almost no options for further analysis. Therefore, we have written the C program HBExplore. This tool can be used for the detailed analysis of H-bonding patterns of both individual structures and of larger sets of structures. Even though the primary reason for the development of HBExplore was our interest in nucleic acid structures HBExplore can be applied to proteins as well. Even though we consider HBPLUS not as sufficient for the manner we would like to analyze H-bonding patterns in biological macromolecules, it should be stated that this program was extremely useful for our work on HBExplore in providing a lot of suggestions and examples for comparison.

Methods

HBExplore is a collection of computer algorithms which identifies potential hydrogen bonding patterns in single structures or in structure sets in a three step procedure by means of geometrical criteria. In a first step the program generates the coordinates of hydrogens connected to potential hydrogen bonding donor atoms. In a second step HBExplore selects potential H-bonds according to a set of geometrical criteria and extracts statistical and other information from the set of H-bonds. Finally, it is possible to search in an interactive manner for H-bonds with particular properties.

HBExplore was written in ANSI C and was developed on a UNIX workstation IRIS INDIGO (Silicon Graphics) running the operating systems IRIX 5.2/5.3. HBExplore should be portable to any other computer with a standard ANSI/ISO C compiler. It can be obtained on

request from the authors (e-mail: jsuehnel@imb-jena.de). Work on a HBExplore web page is in progress. Readers should check out the IMB Jena WWW server (<http://www.imb-jena.de>).

In the following the program is described in more detail. HBExplore requires structural information in the Protein Data Bank (PDB) format (Bernstein et al., 1977). For structures determined by diffraction methods the positions of H atoms are usually not given. Therefore, HBExplore generates in a first step hydrogen coordinates according to standard geometrical rules (Cornell et al., 1995). The positions of the hydrogen atoms are dependent on the hybridisation of the donor atoms and on their atomic environment. All this information is stored in so-called connectivity files for donor and acceptor atoms. These files can easily be edited to include further residues or ligands with H-bonding capabilities. In the case of single bonded donor sp^3 atoms the hydrogen atom is rotated about this single bond and it is checked whether or not there is more than one possible acceptor atom. For NMR structures the H atom coordinates are given in the PDB files. Nevertheless, we generate H coordinates by HBExplore also in these cases. For most of the H atoms the coordinates generated by HBExplore and the coordinates given in the PDB file are practically identical. For freely rotatable donor groups, like the O_2^*H group in the RNA backbone, for example, we do not trust the PDB coordinates too much and therefore rotate these groups also for NMR structures. The geometrical criteria adopted are described in detail in Fig. 1. Criteria set I is identical to the criteria used by HBPLUS (McDonald and Thornton, 1994). If the acceptor atom A is part of a cycle, then there are two bonded neighbours of A, A_1 and A_2 . In this case A_m is used as bonded neighbour of A instead of A_1 for calculating the angles. A_m is located midway on the line connecting A_1 and A_2 . For sp^3 donor atoms first the minimum $H \cdots A$ distance is determined and then the remaining four geometrical parameters are calculated for this hydrogen position. The geometrical criteria can easily be changed by the user. It is well-known that hydrogen bonds display a marked directionality. This can easily be understood in terms of overlapping atomic hybrid orbitals. Whereas the s orbital of the donor hydrogen atoms has spherical symmetry, the hybrid orbitals of sp^2 or sp^3 acceptor atoms display angles of 120 deg and 109 deg. Therefore, a further criterion based on the directionality of hybrid orbitals can be applied. The orbitals are treated as vectors, whose direction is inferred from the surrounding atoms of A. This algorithm is similar to the approach for adding hydrogen atoms to the donor atoms. The corresponding criterion checks whether or not in addition to the distance constraints and the angle DHA , the angle between the $H \cdots A$ hydrogen bond and the direction of the hybrid orbitals on the acceptor atom A is smaller than 60 deg. As potential donor and acceptor atoms O, S and N are used. This means that H-bonds including C atoms are not taken into account.

HBExplore generates tables of all potential hydrogen bonds for both criteria sets and indicates automatically the differences. For the further analysis it is possible to select hydrogen-bonds which represent any logical combination of the two hydrogen bond tables. HBExplore does not consider H-bonds within one and the same residue but takes into account poly-center H-bonds and also, for sp^3 donor atoms, mutually exclusive H-bonds. Further analysis tools enable one to get both automatic information or to search in an interactive manner. The automatic analysis tools include the calculation of mean values and distributions of geometrical

parameters for parts of a single structure, for complete single structures and for structure sets, the classification of each H-bond according to the participation of backbone, sidechain (base), ligand and water parts of nucleic acids, proteins or the corresponding complexes, identification of Watson-Crick nucleotide pairs and of H-bonded pairs of equal nucleotides, and the calculation of the mean number of H-bonds per residue (hydrogen bond density) and of the fraction of potential donor and acceptor atoms involved in H-bonds. For structures with water there is an option to perform the analysis either without or with water. In the latter case potential water-water hydrogen bonds are not taken into account. They are, however, reported in a separate output file.

Further, HBexplore generates automatically a H-bond residue interaction table. This table lists for all residues of the structure the other residues, ligands or water molecules directly bonded via an H-bond. By means of this table one can see at the first glance, whether or not there are residues in the structure not H-bonded and which residue is H-bonded to the largest number of other residues. By means of a binary tree search algorithm this table is then converted into a H-bond cluster table. Clusters are understood here as an uninterrupted network of H-bonded residues. To clarify the difference between the residue interaction table and the cluster table a short example is described. Let us assume that residue 10 exhibits H-bonds with residues 3 and 15. This information is stored in the residue interaction table. It may happen, however, that residue 3 has H-bonds with residue 22 and residue 15 with 12 and 13. If the residues 22, 12 and 13 have no further H-bonds, then there is a H-bond cluster in the structure which consists of residues 10, 3, 15, 22, 12, 13. For structures containing water both tables can be calculated with or without water. For nucleic acids, they are generated both for all H-bonds and for base-base H-bonds only. One interesting aspect is, for example, that clusters may consist of simple chains, branching points, and cyclic parts. Preliminary studies have shown that nucleic acid structures may display cluster patterns which range from H-bonded dinucleotide pairs, over structures with a combination of medium-sized clusters, separated dinucleotide pairs and single nucleotides not H-bonded at all, to extreme cases where the complete structure is almost one cluster. In addition, the secondary structure and tertiary interactions are automatically derived from the 3D structure for RNA. We assume that a RNA structure consists of helix-like and non-helix-like secondary structure elements. Helix-like elements are formed from at least three H-bonded nucleotide pairs of subsequent nucleotides in different sequence regions of one strand or in different strands. Non-helix-like elements are all other parts of a nucleic acid structure. All hydrogen bonds within one and the same secondary structure element are assumed to belong to this element. This includes, for example, H-bonds between non-neighbour nucleotides in loops and H-bonds between a nucleotide and more than one other nucleotides in helix-like regions. H-bond tertiary interactions are defined as H-bonded nucleotide pairs for which the nucleotides involved do not belong to one and the same secondary structure element.

In addition to these automatic analysis tools described HBexplore offers the possibility to select H-bonds according to various atomic, residue and geometrical search criteria which can be combined via logical operators within individual structures or larger structure sets.

HBExplore generates various output files: (a) a table of all H-bonds selected according to both criteria sets with all geometrical parameters, (b) a table of H-bonds sorted according to the residues involved and with all additional information on H-bond statistics and classification but with the D \cdots A distance only, (c) separate output files with the distributions of geometrical parameters for import into graphics programs, (d) coordinate files of the H atoms generated by HBExplore, (e) a file with all potential hydrogen bonds between water molecules.

The results obtained by HBExplore are dependent on the quality of the structure data and on the selection of appropriate geometrical criteria. It is important to be aware of the fact that HBExplore identifies H-bonds by means of geometrical selection criteria. This means it does not differentiate between strong or weak H-bonds. Fortunately, there are simple relationships between geometrical and energetic H-bonding properties. A linear H-bond with a short H \cdots A distance is certainly stronger than a nonlinear one with a larger value of this distance. Insofar it is possible to perform some crude energetic selection by simply varying the geometrical limits. The experimental structures may contain various errors, see the discussion by Kleywegt and Jones (1995). However, for crystal structures lower resolution does not mean automatically a more correct structure. For NMR and theoretical structures there are no resolution data at all. The determination of RNA structures by means of NMR techniques is still in its infancy. Therefore, it is not unlikely that published structures will have to be revised in the future. In spite of these arguments we do not exclude low resolution structures from the analysis. On the contrary, HBExplore can hopefully contribute to improved model building in the course of experimental structure determination from the H-bond point of view.

Results and Discussion

The application of HBExplore is described for two new structures both published in 1995, an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus (PDB code: 1rnk) (Shen and Tinoco, 1995) and a 24-nucleotide RNA hairpin (PDB code: 1rht) (Borer et al., 1995). These two examples serve to illustrate the capabilities of HBExplore. Even though this choice reflects our personal interest in RNA structures, HBExplore can be applied to DNA and proteins as well. More comprehensive studies on H-bonding patterns are in progress.

Note that images of various mono and stereo views and additional color coded distance plots of the structures discussed in this work are available in our IMAGE LIBRARY OF BIOLOGICAL MACROMOLECULES, which can be accessed via the world-wide web (URL: <http://www.imb-jena.de/IMAGE.html>) (Sühnel, in press).

RNA pseudoknot (1rnk)

Pseudoknots are well-known structural motifs in RNA (Pleij et al., 1985; Studnicka et al., 1978). A pseudoknot bridges two helical stems and could thus impose a higher barrier to the translating ribosome than the usual hairpin structure. The structure under study (PDB code: 1rnk) is the first report on a 3D structure of a pseudoknot that acts as frameshift stimulator (Shen and

Tinoco, Jr., 1995).

If one is interested in performing statistical evaluations of large datasets one has first to check for examples whether or not the algorithm works correctly. Therefore, we have performed a thorough comparison of the results obtained by HBExplore and HBPLUS for the pseudoknot structure using identical geometrical criteria. All of the following results were obtained using criteria set I. HBPLUS selects 67 H-bonds and HBExplore 62 H-bonds. A closer inspection shows, however, a more pronounced difference than could be expected from the simple comparison of the numbers 67 and 62. HBPLUS identifies 8 H-bonds not found by HBExplore and HBExplore finds 3 H-bonds not selected by HBPLUS. HBPLUS assumes erroneously for some reason that the atom O3* of the nucleic acid backbone is a H-bond donor atom. This leads to four wrongly assigned H-bonds. Further, HBPLUS finds H-bonds within a nucleotide in three cases, whereas HBExplore only identifies H-bonds between different residues. Finally, HBPLUS identifies one H-bond for which the distance H...A is 2.5 Å, whereas HBExplore gives a distance of 2.51 Å and therefore discards this bond due to the limit condition that the distance between the acceptor atom and the hydrogen should be smaller than 2.5 Å. The difference is due to slightly different assumptions in generating the H atom coordinates according to standard geometrical rules (distance OH: 0.96 Å (HBExplore), 1.0 Å (HBPLUS)). In summary, HBExplore finds all H-bonds detected by HBPLUS except for the wrong ones if precisely the same selection rules are adopted. However, it selects the following 3 additional H-bonds: C19:O2*-A20:O4*, A24:O2*-A25:O2*, C23:O2*-A24:N7). For all of them the backbone atom O2* acts as donor. A careful inspection of these H-bonds using the molecular graphics program SYBYL (Tripos Associates, Inc.) shows that all of them indeed fulfil the criteria adopted. In summary, we can state that both programs identify the overwhelming majority of H-bonds correctly, even though HBExplore is a little bit more precise. This may be due to the fact that HBPLUS was obviously developed with proteins in mind. The main advantage of HBExplore over HBPLUS is, however, not the correct identification of all H-bonds even though it is important to point out that HBExplore can, for example, cope with the modified nucleotides in tRNA. More interesting is the variety of additional analysis tools.

Both HBExplore and HBPLUS generate in a first step hydrogen atom coordinates and use angle criteria for identifying hydrogen bonds. On the other hand, there are approaches where simply the distance between the potential donor and acceptor atoms are used as selection criterion without taking into account any angle criteria (Ippolito et al., 1990; Mandel-Gutfreund et al., 1995). Finally, there are identification procedures which do not consider hydrogen atom coordinates but use angle criteria between donor and acceptor atoms (Stickle et al., 1992). The criterion with D...A distances only can easily be applied using HBExplore. For a maximum D...A distance of 3.5 Å HBExplore identifies 135 hydrogen bonds which should be compared to the 62 bonds selected with criteria set I. In other words, the two selection procedures lead to dramatic differences. This should be kept in mind if one compares result adopting criteria which are identical or similar to our approach (Baker and Hubbard, 1984, McDonald and Thornton, 1994), with a simple D...A distance selection (Ippolito et al., 1990; Mandel-Gutfreund et al., 1995). From Fig. 1 one can easily see why this approach yields a

substantial larger number of H-bonds. It does not exclude geometrical situations where the acceptor atom is located behind the DH group.

A few data on the H-bond statistics of the pseudoknot are presented in Table 1. The mean number of H-bonds per nucleotide, defined by two times the number of hydrogen bonds divided by the number of residues, is 3.64 (34 nucleotides). Note that for a standard B-DNA duplex there are 3.0 H-bonds per nucleotide for a GC homopolymer and 2.0 H-bonds per nucleotide for an AT homopolymer. The higher value found in the pseudoknot is due to a substantial number of H-bonds involving backbone atoms and to a minor part due to the interaction of bases with more than one other base. In proteins the mean number of H-bonds per amino acid is usually much smaller. One further difference is, that in proteins the major part of H-bonds occurs within the backbone, whereas standard nucleic acid structures have exclusively base-base H-bonds and only more unusual structures like the pseudoknot have a larger number of backbone H-bonds. The mean values of the distances $H\cdots A$ and $D\cdots A$ are 1.89 and 2.80 Å and the mean value of the angle $D-H\cdots A$ is 154.8 deg. The distributions of these parameters cannot be shown for the sake of brevity. It is further interesting to note that a substantial number of potential donor atoms and an even larger number of acceptor atoms is not involved in H-bonds. In Table 2 the H-bonding residue interaction table is shown. There is only one nucleotide not involved in any H-bond: C21. The secondary structure derived from the nucleotide interaction table is essentially identical to the structure given by Shen and Tinoco (1995) in their Fig. 9. There is, however, one borderline case. Table 2 shows that there is a H-bond between U34 and U8, the latter forming a Watson-Crick pair with A33. According to our definition of secondary structure this H-bond is part of a helix-like secondary structure element.

Tertiary interactions are of particular importance for the kind the secondary structural elements are assembled to the overall 3D structure. From the H-bond nucleotide interaction table and from the secondary structure of the pseudoknot one can easily derive the following five tertiary interactions: G4:N2-A27:N1, A27:N6-G17:N3, G4:O2*-A26:N1, C5:O2*-A27:O2*, A24:N6-C19:O2P. The first two are tertiary interactions via base-base bonds only, whereas the remaining three include backbone atoms. Shen and Tinoco (1995) do not mention any tertiary interactions except for a possible H-bonding interaction between G7 and C12 in one of their calculated structures. However, for the structure given in the PDB file the distance G7:O6-C12:N4 is 16.75 Å. The cluster analysis shows that the complete structure consists of the non-H-bonded nucleotide C21 and only two clusters, a small one consisting of the 5 nucleotides 8,9,32,33,34 and a large one with all the other 28 nucleotides except for C21. Large clusters can be formed if the typical base-base hydrogen bonds in nucleic acids are supplemented by backbone-backbone or base-backbone H-bonds. These results indicate that for the pseudoknot structure the stabilization of the backbone by hydrogen-bonding is important. The authors point out that their calculated three-dimensional structure constrained by the NMR data does not exhibit any triple interactions. A triple interaction can be understood as either consisting of three H-bonded nucleotides where the H-bonds are not necessarily in the base parts of the nucleotides or as base triples for which only base-base H-bonds are taken into account. If all H-bonds are considered in the pseudoknot structure then the cluster analysis has shown that the large cluster

forms a network of 28 nucleotides. May be even more extreme is the finding that nucleotide G4 is directly H-bonded to five other nucleotides (C3, C5, C16, A26, A27). In this case, however, the H-bonds G4:O2*-C5:O4* and G4:O2*-A26:N1 are alternative. Further, there are various other nucleotides directly bonded to four, three or two nucleotides, see Table 2. Of course, these poly-nucleotide interactions include weak H-bonds. However, even if one takes into account only base-base H-bonds one penta-nucleotide base-base H-bond interaction and one base triple were identified by HBExplore in the pseudoknot structure: C3-G4-C16-G17-A27 and U8-A33-U34. For a stereo drawing of the penta-nucleotide motif, see Fig. 2. C3-G17 and G4-C16 form Watson-Crick base pairs. Further H-bonds are between A27:N1 and G4:N2 and between A27:N6 and G17:N3. One should note, that the A27 base plane is located almost perpendicular to the G4-C16 and C3-G17 planes. However, this is only of minor importance for the strength of the H-bond, which is formed by overlap of the H atom s orbital and of the lone pair sp^2 orbital of N. Further, the H...A distances of these hydrogen bonds are in the same distance range as the H-bonds within the Watson-Crick pairs and the D-H...A angles are between 160 and 180 deg. To the best of our knowledge, this seems to be the first case that a penta-nucleotide base-base hydrogen bond interaction was identified in an RNA structure. Work is in progress to screen all known RNA structures for similar unusual structure elements. In the base triple U8:N3 forms H-bonds with U34:O2 and A33:N1. A further H-bond is found between the donor atom U33:N6 and U8:O4. In this case the base plane of U8 is located between the planes of A33 and U34 which enables U8 to form H-bonds both with A33 and U34. This means that in this case the H-bonds cannot be linear. Rather the angles DHA are between 127 and 148 degrees, which, however, is within the limits set by our criteria.

One could argue, that hydrogen bonds including backbone atoms are less important. Therefore, we have reduced the D...A distance constraint from 3.9 Å to 3.0 Å and indeed the number of backbone-backbone H-bonds decreases from 24 to 18, whereas the number of base-base and base-backbone H-bonds remains almost constant. This means that almost all of the relatively weak H-bonds with D...A distances between 3.0 Å and 3.9 Å are within the backbone. However, if one decreases the D...A distance constraint further to 2.7 Å, then the total number of H-bonds found is 19, where 10 of them are backbone-backbone and 9 base-base bonds. This means that there is a substantial number of relatively strong backbone-backbone H-bonds as well. A search for the H-bond with the shortest D...A distance even yields the result that the backbone-backbone H-bond U22:O2*-C23:O5* is the H-bond within the pseudoknot with the shortest D...A distance (2.45 Å). Most of the other short backbone-backbone bonds are, however, between the O2* and O4* atoms.

Criteria set II (hybrid orbitals) does not identify three hydrogen bonds found by criteria set I, but finds three additional H-bonds. One example of the first category is the H-bond U8:N3-U34:O2. In this case the hydrogen is located on a line dissecting the angle between the two oxygen hybrid orbitals but in a plane above the plane of the hybrids. Therefore, the angle criterion of criteria set II is not fulfilled. One example of an H-bond interaction identified by criteria set II but not by I is G9:N1-C32:O2.

24-nucleotide RNA hairpin (1rht)

Hairpins represent important structure elements in RNA both from the point of view of the local RNA structure and of possible interaction sites with proteins or other nucleic acids (Varani, 1995). One of the hairpin structures known at atomic resolution is a 24-nucleotide variant of the RNA binding sequence for the coat protein of bacteriophage R17 (PDB code: 1rht) (Borer et al., 1995). The structure was determined using a combination of NMR, molecular dynamics, and energy minimization. Again, we have first performed a thorough comparison of the H-bonds identified by HBPLUS and HBExplore using criteria set I. The number of H-bonds found by HBPLUS and HBExplore are 41 and 36. HBExplore does not identify 8 H-bonds selected by HBPLUS. Five of them are due to the incorrect assignment of O3* as a H-bonding donor atom already mentioned in the pseudoknot case. The other three ones are due to a similar error. The atom N1 of cytosine C is also incorrectly treated as a H-bond donor. On the other hand, HBExplore identifies 3 additional H-bond all of which indeed fulfil all the criteria: C13:O2*-C14:O4*, G10:N1-C15:N3, A4:O2*-C5:O5*.

The mean values of the geometrical H-bond parameters are not very different to the pseudoknot case. However, the mean number of H-bonds per nucleotide is smaller for the hairpin than for the pseudoknot. On the other hand, the fraction of H-bonds where the backbone is involved is about 50% for both structures. This means that the pseudoknot has more H-bonds both for the backbone-backbone(base) and for the base-base subsets.

The H-bond nucleotide interaction table gives again a comprehensive overview of the network of H-bonds, which can be used to deduce the secondary structure and tertiary interactions. The authors present a secondary structure in Fig. 1 of their paper, with the base-pairing pattern G3-C21, A4-U20, C5-G19, U6-A18, G7-C17, C9-G16, G10-C15. It is obvious from Table 3, which contains the secondary structure deduced from the hydrogen-bonding patterns of the 3D structure, that the secondary structures derived by HBExplore from the three-dimensional structure is not identical to the secondary structure presented by the authors. Our analysis identifies the non-canonical pair G2-U22. On the other hand, the canonical pair A4-U20 is not H-bonded adopting our criteria. The reason is that the bases of A4 and U20 are not in one plane. This leads to H...A distances larger than 3 Å for the possible H-bonds and this is beyond the limit of 2.5 Å for this distance. Further, A8 assumed as not H-bonded by the authors has H-bonds with G16 and G16 is H-bonded both with C9 and A8. The appearance of non-canonical base-pairs is certainly not very surprising. However, that a formal Watson-Crick pair within a helix is not H-bonded is really interesting. This raises the question whether or not this is a real structural motif or an error in structure determination. Indeed, one of the possible applications of HBExplore is to identify unusual H-bonding patterns of this type during the process of structure refinement. One further interesting fact is that the nucleotide neighbours A18 and G19 form a base-base H-bond. It is quite clear that the DHA angle is rather small in this case, namely 97 deg. However, contrary to all other neighbour base-base interactions this is still within the limits of our geometrical criteria.

According to our definition of helix-like secondary structure element, which requires at

least three subsequent H-bonded nucleotide pairs, the H-bonds G2-U22 and G3-C21 have to be classified as tertiary interactions. The hairpin structure has only one nucleotide (U24), which is not H-bonded to other nucleotides. The complete structure consists of one small cluster (nucleotides: 1,2,3,21,22,23) and one very large cluster formed from all other nucleotides except for the loop nucleotides C13 and A14 which have mutually H-bonds but no H-bonds with other residues. Reducing the D \cdots A distance limit from 3.9 to 3.0 affects backbone-backbone and base-base H-bonds in almost the same manner as for the pseudoknot and the same is true for a further decrease of the D \cdots A distance limit to 2.7 Å, which yields still 6 backbone-backbone and 4 base-base H-bonds. And again like in the pseudoknot case are among the five H-bonds with the shortest D \cdots A distance four backbone-backbone and only one base-base bond, the H-bond with the very shortest D \cdots A distance being G2:O2*-G3:O5* with 2.37 Å.

In this report we have described the new program HBexplore for identifying and analyzing H-bonding patterns in biological macromolecules. It was applied to two RNA example structures, a pseudoknot (PDB code: 1rnk) and a hairpin (PDB code: 1rht). In both cases new H-bond interaction patterns could be identified. Work on a statistical analysis of larger structure sets and selected structure motifs for RNA, DNA and proteins is in progress.

Acknowledgements

We are grateful to I. McDonald for making available his excellent program HBPLUS.

References

- Baker, E.N. and Hubbard, R.E. (1984) Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* **44**, 97-179.
- Bernstein, F. C., Koetzle, T.F., Williams, G., Mayer, E. F., Bryce, M. D., Rodgers, J. R., Kennard, O., Simanouchi, T. and Tasumi, M. (1977), The Protein Data bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.
- Bordo, D. and Argos, P. (1994) The role of side-chain hydrogen bonds in the formation and stabilization of secondary structure in soluble proteins. *J. Mol. Biol.* **243**, 504-519.
- Borer, P.N., Lin, Y., Wang, S., Roggenbuck, M.W., Gott, J. M., Uhlenbeck, O.C. and Pelczar, I. (1995) Proton NMR and structural features of a 24-nucleotide RNA hairpin. *Biochemistry.* **34**, 6488-6503.
- Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179-5197.
- Chastain, M. and Tinoco, I., Jr. (1994) Structural elements in RNA. *Prog. Nucl. Acids Res.* **41**, 131-177.
- Derewenda, Z.S., Lee, L. and Derewenda, U. (1995) The occurrence of C-H \cdots O hydrogen bonds in proteins. *J. Mol. Biol.* **252**, 248-262.

- Ippolito, J.A., Alexander, R.S. and Christianson, D.W. (1990) Hydrogen bond stereochemistry in protein structure and function. *J. Mol. Biol.* **215**, 457-471.
- Jeffrey, G.A. and Saenger, W. (1991) *Hydrogen bonding in biological structures*. Springer-Verlag, New York.
- Joshua-Tor, L. and Sussman, J.J. (1993) The coming of age of DNA crystallography. *Curr. Opin. Struct. Biol.* **3**, 323-335.
- Kleywegt, G.J. and Jones, T.A. (1995) Where freedom is given, liberties are taken. *Structure*, **3**, 535-540.
- Mandel-Gutfreund, Y., Schueler, O. and Margalit, H. (1995) Comprehensive analysis of hydrogen bonds in regulatory protein DNA-complexes: In search of common principles. *J. Mol. Biol.* **253**, 370-382.
- McDonald, I.K. and Thornton, J.M. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777-793.
- Moras, D. and Poterszman, A. (1995) Diverse modes of recognition. *Curr. Biol.* **5**, 249-251.
- Nelson, H.C.M. (1995) Structure and function of DNA-binding proteins. *Curr. Opin. Genetics Develop.* **5**, 180-189.
- Pauling, L., Corey, R.B. and Branson, H.R. (1951) The structure of proteins. Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA* **37**, 205-211.
- Pleij, C.W.A., Rietveld, K. and Bosch, L. (1985) A new principle of RNA folding based on pseudoknotting. *Nucl. Acids Res.* **13**, 1717-1731.
- Pley, H.W., Flaherty, K.M. and McKay, D.B. (1994) Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**, 68-74.
- Shen, L.X. and Tinoco, Jr., I.J. (1995) The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. *J. Mol. Biol.* **247**, 963-978.
- Stickley, D.F., Presta, L.G., Dill, K.A. and Rose, G.D. (1992) Hydrogen bonding in globular proteins. *J. Mol. Biol.*, **226**, 1143-1159.
- Studnicka, G.M., Rahn, G.M., Cummings, I.W. and Salser, W.A. (1978) Computer method for predicting the secondary structure of single-stranded RNA. *Nucl. Acids Res.* **5**, 3365-3387.
- Sühnel, J. (in press) Image Library of Biological Macromolecules. *Comput. Appl. Biosci.*
- Tuschl, T., Gohlke, C., Jovin, T.M., Westhof, E. and Eckstein, F. (1994) A three-dimensional model for the hammerhead ribozyme based on fluorescence measurements. *Science* **266**, 785-798.
- Varani, G. (1995) Exceptionally stable nucleic acid hairpins. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 379-404.
- Watson, J.D. and Crick, F.H.C. (1953) A structure for deoxyribose nucleic acid. *Nature* **171**, 737-738.

Received on

;

Table 1

Hydrogen bond statistics of an RNA pseudoknot (1rnk) and of a hairpin (1rht)¹

	1rnk	1rht
Total number of H-bonds:	62	36
backbone-backbone	24	18
base - base	34	17
backbone - base	4	1
Mean number of H-bonds per nucleotide:	3.64	3.0
Mean value of H-bond distance D [⋯] A:	2.80 Å	2.86 Å
Mean value of H-bond distance H [⋯] A:	1.89 Å	2.04 Å
Mean value of H-bond angle D-H [⋯] A:	154.8 deg	144.7 deg
Number of potential donor atoms (H-bonded)	79 (55)	55(29)
backbone	34 (21)	24(13)
base	45 (34)	31(16)
Number of potential acceptor atoms (H-bonded)	289 (62)	205(36)
backbone	202 (26)	144(19)
base	87 (36)	61(17)

¹ 1rnk: Shen and Tinoco, Jr. (1995); 1rht: Borer et al. (1995).

Table 2

Hydrogen-bond nucleotide interaction table for an RNA pseudoknot (1rnk) and a hairpin (1rht)

1rnk						1rht			
n	nucleotides H-bonded with n					n	nucleotides H-bonded with n		
1:	19					1:	2		
2:	18					2:	1	3	22
3:	4	17				3:	2	21	
4:	3	5	26	16	27	4:	5		
5:	4	27	15	6		5:	4	19	
6:	7	5				6:	18		
7:	6					7:	8	17	
8:	9	33	34			8:	7	16	
9:	8	32				9:	10	16	
10:	31					10:	9	11	15
11:	30					11:	10	12	
12:	29					12:	11		
13:	14	28				13:	14		
14:	13	15				14:	13		
15:	5	14	16			15:	10		
16:	4	15	17			16:	8	9	
17:	3	16	27			17:	7	18	
18:	2	19				18:	6	17	19
19:	1	18	20	24		19:	18	20	5
20:	19					20:	19		
21:	--					21:	3	22	
22:	23					22:	2	21	23
23:	22	24				23:	22		
24:	23	25	19			24:	--		
25:	24	26							
26:	4	25	27						
27:	4	5	26	17					
28:	13	29							
29:	12	28	30						
30:	11	29	31						
31:	10	30							
32:	9	33							
33:	8	32	34						
34:	8	33							

Table 3

Hydrogen bond secondary structure of the pseudoknot (1rnk) and of the hairpin (1rht)

1rnk	1rht
G 1 - C 19	G 1
G 2 - C 18	G 2
C 3 - G 17	G 3
G 4 - C 16	A 4
C 5 - G 15	C 5 - G 19
A 6	U 6 - A 18
G 7	G 7 - C 17
U 8 - A 33 , U 8 - U 34	A 8 - G 16
G 9 - C 32	C 9 - G 16
G 10 - C 31	G 10 - C 15
G 11 - C 30	A 11
C 12 - G 29	U 12
U 13 - G 28	C 13
A 14	A 14
G 15 - C 5	C 15 - G 10
C 16 - G 4	G 16 - C 9 , G 16 - A 8
G 17 - C 3	C 17 - G 7
C 18 - G 2	A 18 - U 6
C 19 - G 1	G 19 - C 5
A 20	U 20
C 21	C 21
U 22	U 22
C 23	A 23
A 24	U 24
A 25	
A 26	
A 27	
G 28 - U 13	
G 29 - C 12	
C 30 - G 11	
C 31 - G 10	
C 32 - G 9	
A 33 - U 8	
U 34 - U 8	

Figure Captions.

Fig. 1.

Geometrical criteria for selecting hydrogen bonds

(D - donor; A - acceptor; H - hydrogen; A_1, A_2, A_m - bonded neighbours of A;

D, A - O, N, S; d_1 - distance $D \cdots A$; d_2 - distance $H \cdots A$; α - angle DHA; β - angle HAA₁(A_m);

γ - angle DAA₁(A_m); ϵ - angle between the lines defined by H and A and the vector of the hybrid orbitals - hyb).

I. criteria set I (default values):

$$d_1 < 3.9 \text{ \AA}, d_2 < 2.5 \text{ \AA}, \alpha > 90 \text{ deg}, \beta > 90 \text{ deg}, \gamma > 90 \text{ deg}.$$

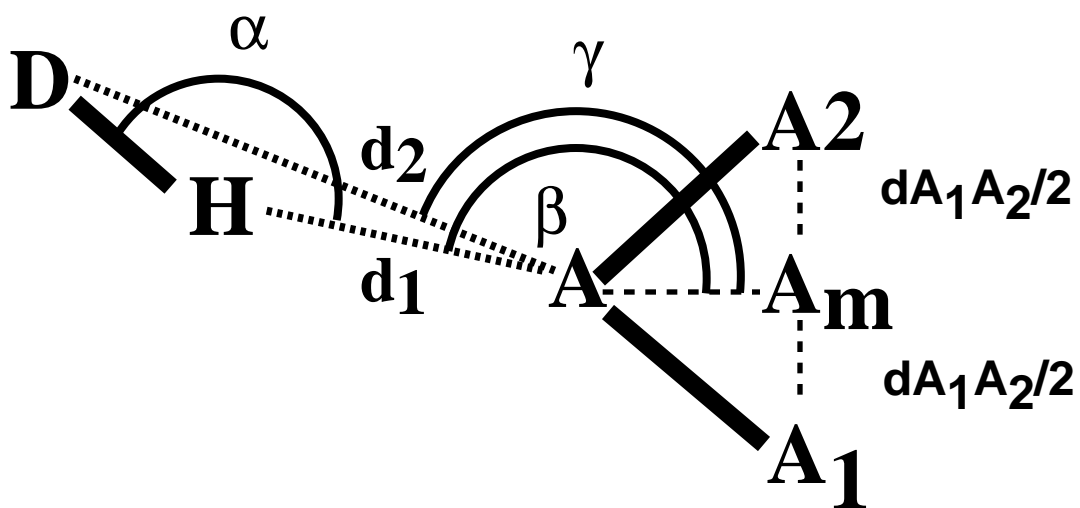
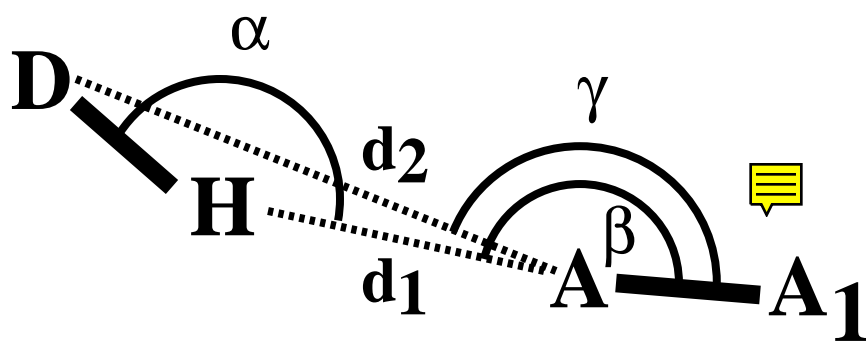
II. criteria set II (default values):

$$d_1 < 3.9 \text{ \AA}, d_2 < 2.5 \text{ \AA}, \alpha > 90 \text{ deg}, \epsilon < 60 \text{ deg}.$$

Fig. 2.

Stereo drawing of the penta-nucleotide hydrogen bond base-base interaction motif between C3, G4, C16, G17 and A27 in an RNA pseudoknot (PDB code: 1rnk). The numbers indicate the $H \cdots A$ distances in \AA .

I.



II.

