More hydrogen bonds for the (structural) biologist

Manfred S. Weiss, Maria Brandl, Jürgen Süehnel, Debnath Pal and Rolf Hilgenfeld

Why does a given protein structure form and why is this structure stable? These fundamental biochemical questions remain fascinating and challenging problems because the physical bases of the forces that govern protein structure, stability and folding are still not well understood. Now, a general concept of hydrogen bonding in proteins is emerging. This concept involves not only N–H and O–H donor groups, but also C–H and nitrogen atoms as acceptor groups, but also \( \pi \)-systems. We postulate that the incorporation of the entirety of these interactions leads to a more complete description of the problem, and that this could provide new perspectives and possibly new answers.

The world of proteins used to be simple. Hydrogen atoms bound to nitrogen and oxygen atoms formed hydrogen bonds with lone electron pairs on other oxygen and nitrogen atoms. These ‘classical’ hydrogen bonds\(^1,2\) (Fig. 1a) have been held responsible for the formation of secondary structural elements such as \( \alpha \)-helices and \( \beta \) sheets and, along with van der Waals and hydrophobic forces, they constitute one of the main pillars of overall protein stability and a principal determinant of protein conformation. However, this seemingly simple picture is not able to provide more than a qualitative explanation of protein structure, folding and stability. A quantitative description that would allow the calculation and prediction of the energetics of these phenomena needs to be more elaborate and, indeed, the ‘classical’ view on hydrogen bonds has evolved considerably over the years. In 1982, Taylor and Kelland presented unequivocal evidence for the existence of hydrogen bonds between C–H donor groups and oxygen acceptors\(^3\), and in the 1990s these hydrogen bonds were discovered for proteins at first\(^4\) and then for other biological macromolecules\(^5\) (Fig. 1b). Early this year, two papers dealing with even more exotic hydrogen bonds were published: Steiner and Koellner described hydrogen bonds involving aromatic acceptors\(^6\) (Fig. 1c), and Brandl et al. exhaustively surveyed the occurrence of interactions involving all possible C–H groups\(^7\) (Fig. 1d). Burley and Petsko have termed these interactions ‘weakly polar interactions’\(^8\) but they can also be classified as hydrogen bonds, although they are considerably weaker than ‘classical’ hydrogen bonds. Because of their frequent occurrence in proteins, these interactions can be expected to contribute significantly to the overall stabilization energy of a protein, which is often not more than a few kcal mol\(^{-1}\). Of course there is no reason to restrict their potential importance to just the intramolecular protein context itself. As the respective donor and acceptor groups also occur on protein surfaces, these interactions can be used to recognize and selectively bind other proteins, ligands, substrates, inhibitors and so forth.

The nature of these various types of hydrogen bonds can be described qualitatively by Pearson’s hard/soft–acid/base (HSAB) concept\(^9\). If N–H and O–H groups are considered hard acids, S–H groups intermediate acids and C–H groups soft acids, and oxygen and nitrogen atoms the hard bases, sulfur atoms intermediate bases and \( \pi \)-systems soft bases, then soft-acid–soft-base hydrogen bonds should form as do hard-acid–hard-base hydrogen bonds. A more quantitative description can be provided by quantum-chemical \( ab\) initio calculations, which suggest that at least four different attractive energy terms ought to be considered: (1) electrostatic energy arising from interactions between charges, partial charges and dipoles; (2) charge-transfer or delocalization energy; (3) polarization energy from interactions between permanent dipoles and induced dipoles; and (4) dispersion energy originating from interacting temporary dipoles and induced dipoles. These stabilizing energies are counterbalanced with a destabilizing, repulsive energy, termed exchange repulsion energy. The various types of hydrogen bonds can now be distinguished by how much the different energy terms contribute to the overall energy of a hydrogen bond. The hard-acid–hard-base hydrogen bonds are dominated by electrostatic and charge-transfer energy, and the soft-acid–soft-base hydrogen bonds by...
dispersion and polarization energy. As a consequence, soft-acid–soft-base hydrogen bonds are persistent in both highly polar solvents such as H\textsubscript{2}O and apolar environments such as those found in the interior of proteins. The different types of hydrogen bonds also appear to exhibit a different behaviour when studied by vibrational spectroscopy. Observed frequency shifts of D–H (where D represents any atom) bonds are usually a good indication for the formation of a hydrogen bond and they can also be used to distinguish between different hydrogen bond types. The classical hydrogen bonds always show a red shift upon hydrogen bond formation. This implies that the D–H bond is weakened and concomitantly the D–H bond length increased. By contrast, the C–H stretching frequency can be either red- or blue-shifted and accordingly the C–H bond shortened, respectively.

To be fair, it should be mentioned that the concept of hydrogen bonds involving C–H donor groups and/or π-acceptor groups is not new. For many years, it has been well known and widely accepted in the field of small organic and inorganic molecules, where inclusion compounds or host-guest complexes often form solely on the basis of these weak hydrogen bonds. The recently published books by Nishio et al. and Desiraju and Steiner provide excellent reviews on this topic. However, the protein community has been slow in recognizing the importance of these weak hydrogen bonds for proteins, although some papers of the past decade have pointed in this direction (see, for example, Ref. 14). It can only be hoped that the recent findings will initiate a re-evaluation of this scepticism.

To shed some light on the relative importance of the various types of hydrogen bonds in protein structures, a complete hydrogen-bond analysis of two high-resolution protein structures taken from the Protein Data Bank was carried out. Phospholipase C (PLC; PDB entry 1AH7, 1.5 Å resolution) is a protein of mainly α-helical structure consisting of 245 amino acid residues (Fig. 1, central panel), and human heparin-binding protein (HBP; PDB entry 1A7S, 1.12 Å resolution) is a β-sheet protein comprising 221 residues. The total numbers of identified hydrogen bonds in each class are given in Table 1; the criteria for their identification are outlined in the table legend. The four classes of hydrogen bonds described are also illustrated in Fig. 1 using examples identified in the PLC structure.

From the frequency of occurrence of the various interactions presented in Table 1, it can be inferred that for both all-α- and all-β-proteins, the ‘classical’ hydrogen bonds dominate the picture by their sheer number. The absolute numbers presented should certainly be taken with a grain of salt as they represent the numbers for individual proteins rather than those for whole groups of proteins. Nevertheless, they are able to reveal some general trends. Forming ~20–25% of the total number of hydrogen bonds, C–H...O interactions constitute the second most important group. It has already been described for proteins containing a large fraction of β-structure, that C–H...O hydrogen bonds occur almost ubiquitously in β sheets, but they also appear to occur frequently in α-helical proteins. Hydrogen bonds with π-acceptors constitute yet another considerable fraction. Even though π-systems are relatively infrequently observed to interact with O–H, N–H and S–H donors, they seem to exhibit a distinct preference to team up with C-H donors, thus forming C–H...π hydrogen bonds. The formation of these bonds is consistent with the HSAB concept. Again, we would like to emphasize that the observation that more C–H...π hydrogen bonds occur in PLC than in HBP, does not mean that these hydrogen bonds are more abundant in all-α proteins than in all-β proteins; it merely reflects the possible spread of the numbers. Brandl et al. had shown previously that proteins can contain anywhere between 0.0 and 22.6 C–H...π hydrogen bonds per 100 amino acids.

It has also been noted in some cases that these weaker hydrogen bonds occur at molecular interfaces and that they are important for molecular recognition.

In terms of hydrogen bonding energies, many battles have been fought over the years. The commonly accepted numbers now range from ~10 kcal mol\textsuperscript{-1} for the strongest (e.g. O–H...O) bonds to ~0.5–1.0 kcal mol\textsuperscript{-1} for C–H...O hydrogen bonds. The ‘classical’ N–H...O hydrogen bonds commonly observed in biological macromolecules are considered to be of intermediate strength, ~5–6 kcal mol\textsuperscript{-1}.
interaction energies was one of the reasons why the weaker hydrogen bonds have been neglected for such a long time. Interestingly however, Scheiner et al. 19 recently published a quantum-mechanical study on C–H–O hydrogen bonds between C–H atoms of amino acids and H2O molecules. They found that such a hydrogen bond is about half as strong as a ‘classical’ one between two H2O molecules and, in the case of the positively charged lysine, the C–H–O bond is thought to be even stronger.

The numbers presented unequivocally show that the weaker interactions cannot and must not be neglected. Albeit weak, they are numerous and therefore might help explain the well-known problem that protein stabilities, interaction energies and folding energies cannot be calculated very accurately. The consideration of these important interactions might enhance the usefulness of these calculations in general, and further our understanding of protein structures and their functions.

References
15 Hough, E. et al. (1998) High resolution (1.5 Å) crystal structure of phospholipase-C from Bacillus cereus. Nature 338, 357–360

Manfred S. Weiss*
Maria Brandl
Jürgen Sünnel
Deborah Pal
Rolf Hilgenfeld
Institute of Molecular Biotechnology, Beutenbergstr. 11, D-07745 Jena, Germany.
*e-mail: msweiss@imb-jena.de

**Table 1. Complete hydrogen-bond analysis for phospholipase C (PLC) and human heparin-binding protein (HBP)**

<table>
<thead>
<tr>
<th>Hydrogen-bond type</th>
<th>PLC (all-α; 245 amino acids) (number of hydrogen bonds)</th>
<th>HBP (all-β; 221 amino acids) (number of hydrogen bonds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D–H–A (D = N,O; A = N,O,S)</td>
<td>333 (135.9)</td>
<td>192 (86.9)</td>
</tr>
<tr>
<td>S–H–A (A = N,O,S)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>C–H–A (A = N,O,S)</td>
<td>82 (33.5)</td>
<td>81 (36.7)</td>
</tr>
<tr>
<td>D–H–X (D = N,O)</td>
<td>2 (0.8)</td>
<td>–</td>
</tr>
<tr>
<td>S–H–π^A</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>C–H–π^A</td>
<td>50 (20.4)</td>
<td>20 (9.0)</td>
</tr>
<tr>
<td>Total</td>
<td>467 (190.6)</td>
<td>293 (132.6)</td>
</tr>
</tbody>
</table>

*Numbers represent the total number of hydrogen bonds found (numbers in parentheses represent the number of hydrogen bonds per 100 amino acid residues).

1 Hydrogen bonds were selected when the distance between the donor atom D and the acceptor atom A (d_D,A) was ≤ 3.5 Å, the angle at the hydrogen bond (θ_D,A,B) ≥ 90°, and the angle at the acceptor atom also ≥ 90°. These are the standard Baker-and-Hubbard criteria for hydrogen bonds.

2 Hydrogen bonds with S–H donor groups could not occur in the two cases studied because PLC does not contain cysteine residues and all eight cysteine residues in PLC are involved in disulfide bridges.

3 Hydrogen bonds were selected when the distance between the donor atom D and the acceptor atom A (d_D,A) was ≤ 3.5 Å, and the angles at the hydrogen bond (θ_D,A,B) < 90°, the angle at the donor atom also < 90°. These are the standard Baker-and-Hubbard criteria for hydrogen bonds.

4 Hydrogen bonds with S–H donor groups could not occur in the two cases studied because PLC does not contain cysteine residues and all eight cysteine residues in PLC are involved in disulfide bridges.

5 Hydrogen bonds were selected when the distance between the donor atom D and the acceptor atom A (d_D,A) was ≤ 2.0 Å to the plane of the π-system containing the acceptor atom.

6 Hydrogen bonds were selected when the distance between the donor atom D and the acceptor atom A (d_D,A) was ≤ 2.0 Å depending on the π-system in question.

7 Hydrogen bonds were selected when the distance between the donor atom D and the acceptor atom A (d_D,A) was ≤ 3.5 Å and the angle at the acceptor atom also ≤ 90°.

8 Hydrogen bonds were selected when the distance between the donor atom D and the acceptor atom A (d_D,A) was ≤ 4.5 Å, where X is the center of the π-system, the selection criteria had to be expanded by an additional geometric constraint. For a C–H–O interaction to occur, the donor C atom had to be within a distance of 2.0 Å to the plane of the π-system containing the acceptor O atom.

9 Hydrogen bonds were selected when the distance between the donor atom D and the acceptor atom A (d_D,A) was ≤ 4.5 Å, where X is the center of the π-system, the selection criteria had to be expanded by an additional geometric constraint. For a C–H–O interaction to occur, the donor C atom had to be within a distance of 2.0 Å to the plane of the π-system containing the acceptor O atom.

10 Hydrogen bonds with S–H donor groups could not occur in the two cases studied because PLC does not contain cysteine residues and all eight cysteine residues in PLC are involved in disulfide bridges.

11 Hydrogen bonds were selected when the distance between the donor atom D and the acceptor atom A (d_D,A) was ≤ 2.0 Å to the plane of the π-system containing the acceptor atom.

12 Hydrogen bonds were selected when the distance between the donor atom D and the acceptor atom A (d_D,A) was ≤ 2.0 Å depending on the π-system in question.