

Short hydrogen bonds in proteins

Sathyapriya Rajagopal and Saraswathi Vishveshwara

Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India

Keywords

short hydrogen bonds; donor–acceptor of backbone/side chain; secondary structural and residue frequency; geometrical constraint; hydrogen bond strength

Correspondence

Molecular Biophysics Unit, Indian Institute of Science, Bangalore, Karnataka 560012, India

Fax: +91 80 23600535

Tel: +91 80 22932611

E-mail: sv@mbu.iisc.ernet.in

(Received 5 December 2004, revised 23 January 2005, accepted 8 February 2005)

doi:10.1111/j.1742-4658.2005.04604.x

The unique tertiary structures of proteins depend crucially on hydrogen bonds. Extensive investigations carried out on protein structures have shown that the bulk of hydrogen bonds in proteins belong to the normal type with neutral electronegative atoms as proton donors and acceptors [1–3]. The availability of a large number of high-resolution protein crystal structures in the last two decades, however, has revealed a variety of other types of hydrogen bonds, such as C–H–X hydrogen bonds [4–6], hydrogen bonds with π acceptors [7], protein–water hydrogen bonds [8], and other nonconventional hydrogen bonds [9,10].

The short-strong hydrogen bond (SSHB) is yet another class of hydrogen bond, which has been found in many chemical and biological systems. Particularly in proteins, the short hydrogen bonds have been observed at the active site of several enzymes [11–17]. These SSHBs are believed to play an important role in enzyme catalysis through low barrier hydrogen bonds (LBHB). The types of proton donor and acceptor, and the environment are the major determinants of the

Short hydrogen bonds are present in many chemical and biological systems. It is well known that these short hydrogen bonds are found in the active site of enzymes and aid enzyme catalysis. This study aims to systematically characterize all short hydrogen bonds from a nonredundant dataset of protein structures. The study has revealed that short hydrogen bonds are commonly found in proteins and are widely present in different regions of the protein chain, such as the backbone or side chain, and in different secondary structural regions such as helices, strands and turns. The frequency of occurrence of donors and acceptors from the charged side chains as well as from the neutral backbone atoms is equally high. This suggests that short hydrogen bonds in proteins occur either due to increased strength or due to geometrical constraints and this has been illustrated from several examples.

length and strength of these short hydrogen bonds [18]. Although there is a debate about the importance of LBHB in enzyme catalysis [19,20], the existence of SSHBs cannot be denied and can unambiguously be identified by experimental [21,22] methods and by *ab initio* calculations [23]. Because SSHBs are generally observed in systems with net charge, it is debated whether to consider such interactions as hydrogen bonds or as electrostatic interactions. Nevertheless, SSHBs have been shown to be present in nanotubes [calix (4)hydroquinone CHQ] even in a completely neutral environment [24,25]. The strengths of these short hydrogen bonds, however, have to be evaluated by *ab initio* quantum mechanical methods. Furthermore, the evidence for short hydrogen bonds in neutral systems has been provided from recent neutron diffraction studies and also from a search of the Cambridge Structural Database [26]. Such short hydrogen bonds are stabilized by charge, resonance and polarization effects and have been termed as synthon-assisted hydrogen bonds (SAHB).

Abbreviations

BB, backbone; LBHB, low barrier hydrogen bond; SC, side chain; SHB, short hydrogen bonds; SSHB, short-strong hydrogen bond.

Although short hydrogen bonds have been reported in several proteins and their implication in catalysis has been discussed in detail, we noticed that a systematic analysis in terms of the nature of the donors-acceptors and their possible roles in stabilizing the tertiary structure of proteins has not been undertaken so far. Therefore, we have performed a systematic analysis of short hydrogen bonds (SHBs) from a non-redundant dataset of 948 protein chains. As the bond distances –between the proton donor (D) and the acceptor (A) atoms – of normal hydrogen bonds in proteins vary from 2.7 Å to 3.2 Å, we have defined the hydrogen bonds with distance $d(D-A) < 2.7$ Å and angle $D-H-A \geq 150^\circ$ as SHBs, where the donor and the acceptor are nitrogen and oxygen atoms. Short hydrogen bonds involving sulfur atoms have also been investigated. The analysis has clearly shown that a large number of SHBs occur in proteins. The donor-acceptor specificities of SHBs and their role in stabilizing the tertiary structures of proteins are some of the highlights of this investigation.

Results and Discussion

Dataset validation

The positions of hydrogen atoms are not directly determined even in high resolution X-ray structures and their positions have to be fixed by modelling based on standard geometries. We have used AMBER software for fixing hydrogens and most of the results presented here are based on this method of hydrogen atom fixation. However, several checks have been made to assess the validity of this dataset. Firstly, the B-factors of the donor and the acceptor atoms involved in formation of SHBs are compared with those of the atoms forming normal hydrogen bonds. It was found that the B-factor distribution profile was very similar in the short and the normal hydrogen bonded cases (supplementary Fig. S1). The B-factors of the donor and acceptor atoms had values less than 50 Å² in short as well as normal hydrogen bonds, while the maximum value attained was in the region of 120–150 Å².

Secondly, different programs may vary slightly in assigning the positions of hydrogen atoms, which can give rise to varied results. To address this issue, we have compared the SHBs obtained by AMBER with those from HBPLUS [27] from the same dataset. It was seen that HBPLUS gave substantially larger numbers of SHBs compared with AMBER. However, it was seen that the list of SHBs given by AMBER is a subset of the HBPLUS results. A detailed analysis showed that the excess SHBs assigned by HBPLUS is mainly due to

hydroxyl group (OH) orientation (of Ser, Thr and Tyr) being optimized for formation of hydrogen bonds. It is not clear what percentage of these additional SHBs given by HBPLUS would be retained after energy minimization and how many more would be added to the AMBER list. However, it is likely that the prediction given by AMBER is an underestimation, while that of HBPLUS is an overestimation. A similar trend has also been seen for the SH group of Cys as donors of SHBs.

Finally, we have compared the SHBs from a set of 14 proteins, whose structures have been solved by neutron diffraction as well as X-ray crystallography (supplementary Table S1). The analysis showed only partial consistency of the identified SHB lists. Interestingly, the differences are not necessarily an artefact of the method of hydrogen fixation. Varied results are obtained even in the case of neutron diffraction studies on the same protein, which might be due to variations in experimental conditions. Whether the differences are due to the imposition of a fixed cut-off value ($d,0$) was examined for one case of sperm whale myoglobin solved by different groups (supplementary Table S2). In many instances, the SHB found in one neutron diffracted structure is seen as a normal hydrogen bond in the other neutron or X-ray structures. Thus in general, it is desirable to validate a specific hydrogen bond in a given protein by several methods. However, this analysis focuses on general features of SHBs in a large dataset.

Classification and statistics of SHBs in proteins

Using the criteria specified in the Experimental procedures section, SHBs have been identified from the dataset of 948 proteins, after fixing hydrogen atoms using both AMBER and HBPLUS. A total of 4087 and 7860 SHBs have been obtained for AMBER and HBPLUS, respectively. The statistics of the number of SHBs (from AMBER) in a given protein is presented as a histogram in Fig. 1. Of the 948 proteins in the dataset, the number of SHBs per protein chain varies widely from 0 to greater than 50. Approximately 800 of these proteins have at least one SHB. It is interesting to note that there are three enzymes [malate synthase G (1d8cA), carbamoyl phosphate synthetase (1a9xA) and glucose oxidase (1gpeA)], which have greater than 50 SHBs.

The SHBs are classified on the basis of several criteria. The first classification is based on the donor/acceptor atoms arising from the backbone (BB) or the side chain (SC) of the polypeptide chain. In this case, the SHBs are subclassified as: (a) BB-BB, where both the donor (N-H) and the acceptor (C=O) atoms come from the backbone; (b) BB-SC, in which the donor (N-H) is from the backbone and the acceptor

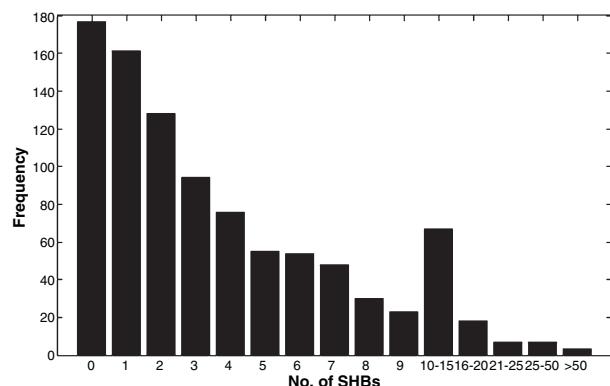


Fig. 1. A plot of the frequency of proteins containing varying numbers of short hydrogen bonds (SHBs) as determined by AMBER.

from the side chain; (c) SC-BB, side chain donor with backbone ($\text{C}=\text{O}$) acceptor; and (d) SC-SC bonds, where both the donor and the acceptor atoms are from the protein side chains. The second classification is based on the sequence separation between the donor and the acceptor, $|\text{D}_i-\text{A}_j|$. Here, the SHBs separated in sequence by less than four residues ($|\text{D}_i-\text{A}_j| \leq 4$) are termed local or short range SHBs, from five to nine residues ($5 \leq |\text{D}_i-\text{A}_j| \leq 9$) as medium range, and more than nine residues ($|\text{D}_i-\text{A}_j| > 9$) as long range SHBs. The third classification is based on the secondary structure to which the donors and the acceptors belong. The analyses based on these classifications are presented below.

The distribution of the length $d(\text{D}-\text{A})$ of the SHBs classified on the basis of backbone and the side chain is presented in Fig. 2A (AMBER) and Fig. 2B (HBPLUS). A large number of BB-BB SHBs is found from

Fig. 2A, which indicates that SHBs are possible between neutral species. A greater proportion of these BB-BB SHBs occur in the distance range of about 2.6–2.7 Å. The distribution of the SC-SC SHBs is also high, and increases gradually from 2.3 Å to 2.65 Å. The distribution of the BB-SC and the SC-BB cases increases gradually from 2.45 Å to 2.65 Å though the numbers are significantly less when compared to the occurrence of the BB-BB and the SC-SC cases. The SHBs obtained from HBPLUS (Fig. 2B) also show a high frequency of BB-BB category. However large numbers were obtained for the SC-SC and SC-BB type. The origin of this difference is analysed in a later section. Nevertheless, a detailed analysis showed that the AMBER list is a subset of the HBPLUS list, with a negligible fraction (< 1%) identified only by AMBER.

The classification based on sequence separation helps in assessing the influence of SHBs at the local level or at the level of sequentially separated spatial interactions in the protein structure. The details of the number of SHBs observed as a function of the donor-acceptor sequence separation is presented in Table 1. A histogram from the AMBER list is also given in Fig. 3. It is evident from these that the contribution of SHBs (AMBER) from short, medium and long range are 30.5%, 10.8% and 58.7%, respectively. Very similar distribution (37.2%, 8.8% and 54%) has been obtained from HBPLUS. The long range SHBs contribute to more than half of the total SHBs observed in the dataset.

The BB-BB SHBs observed between the donor i and the acceptor $(i \pm 3)$ or $(i \pm 4)$ are the major components of the short-range SHBs. As expected, these bonds are found mainly in the helical regions of

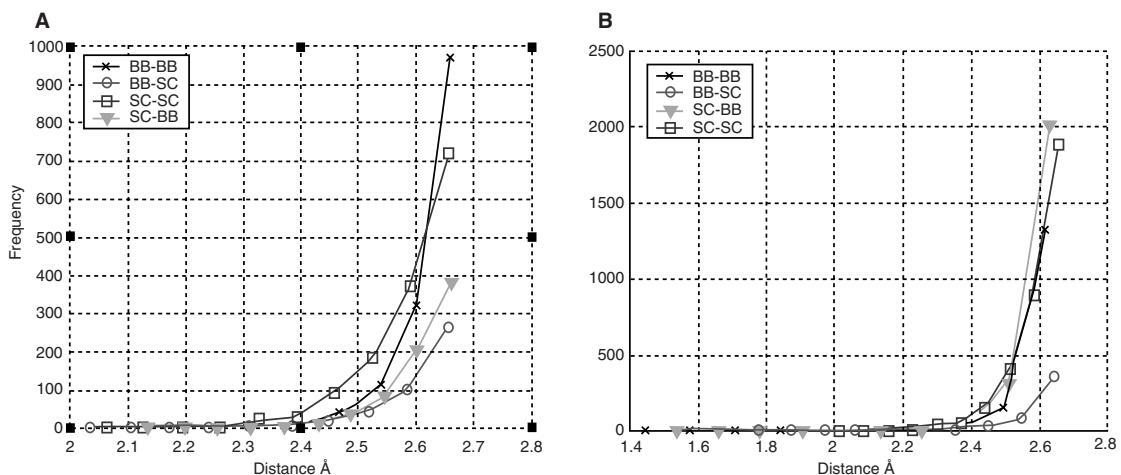


Fig. 2. (A) The frequency of SHBs as a function of donor-acceptor ($\text{D}-\text{A}$) distance (< 2.7 Å) in the AMBER list. (B) The frequency of SHBs as a function of donor-acceptor ($\text{D}-\text{A}$) distance (< 2.7 Å) in the HBPLUS list.

Table 1. Overview of SHB in proteins. Total number of proteins in the dataset = 948; total number of SHBs = 4087 (AMBER), 7860 (HBPLUS; values given in parentheses).

SHB	Short $ D_i-A_j \leq 4$	Medium $5 \leq D_i-A_j \leq 9$	Long range $ D_i-A_j > 9$	Total
BB-BB	600 (605)	191 (187)	692 (706)	1483 (1498)
BB-SC	275 (181)	36 (36)	140 (155)	451 (372)
SC-BB	142 (1173)	95 (197)	492 (2019)	729 (2389)
SC-SC	230 (963)	121 (274)	1073 (2364)	1424 (3481)
Total	1247 (2922)	443 (694)	2397 (4244)	4087 (7860)

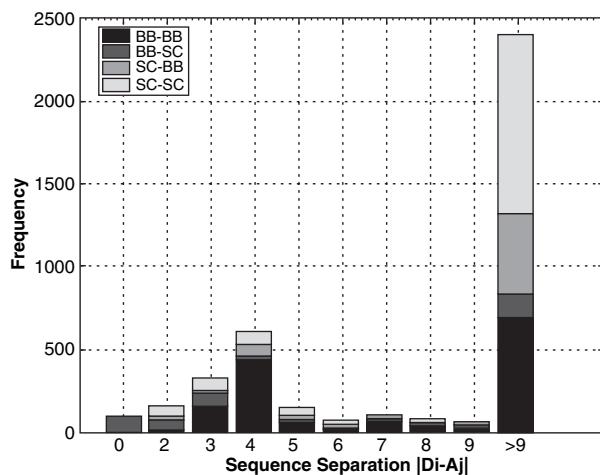


Fig. 3. A plot of the frequency of SHB (as determined by AMBER) as a function of donor-acceptor (D_i-A_j) sequence separation in the polypeptide chain. The four combinations of donor-acceptors from backbone (BB) and side chain (SC) are shown in different shades of grey.

proteins. In the case of the short range SHBs, apart from the intrahelical SHBs a considerable number of SHBs are observed in the D_i-A_i cases, as shown in Fig. 3. These SHBs are formed mainly from the backbone to the side chain of the same residue i . The residue side chains of Glu and Gln form these kinds of SHB in most of the cases where the backbone NH donates the proton to its own side chain carbonyl oxygen (BB_i-SC_i). An example of a D_i-A_i SHB from *p*-hydroxybenzoate hydroxylase (1pbe) is shown in Fig. 4. The side chains of Asn, Arg and Lys also participate in this type of D_i-A_i SHB wherein the side chains of the above mentioned residues donate to their own backbone oxygen atoms. The torsion angle, ϕ , of these residues seems to prefer narrow region in the Ramachandran map, around -50° to -100° . The occurrence of a similar preference for the ϕ in the (D_i-A_i) type of hydrogen bonds of normal distance range was reported

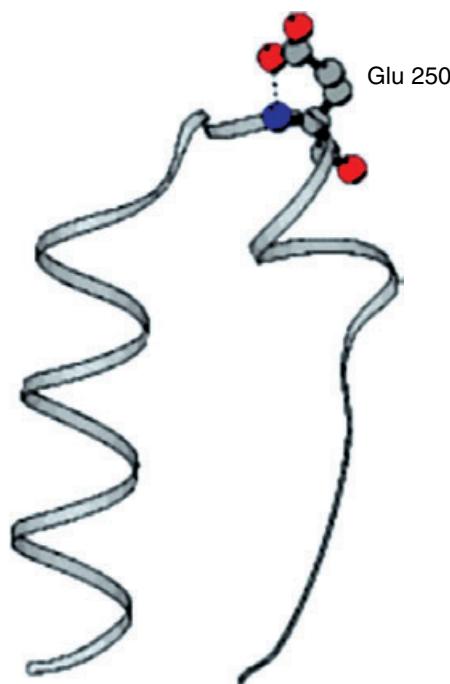


Fig. 4. The (D_i-A_i) SHB in *p*-hydroxybenzoate hydroxylase (PDB code 1pbe). The backbone N-H of Glu250 donates its hydrogen to OE1 of its own side chain.

by Eshwar *et al.* [27]. However, there is a difference in the residue preference as observed between the short and the normal hydrogen bonded cases. In contrast to the Glu and Gln residues in the SHB, Asn and Thr residue side chains donate to their backbone to form SC_i-BB_i in the normal hydrogen bonded cases.

The secondary structural preferences of the donors and acceptors as given by AMBER are represented as bars in Fig. 5 for the short and long range cases. Similar plots from the HBPLUS list are given in supplementary Fig. S2. In case of the short range SHBs (Fig. 5A), the predominant interactions seen within the BB-BB cases are the intrahelical regions (H-H) of the backbone as can be seen from Fig. 5A i. Interestingly, the intrahelical interactions from the side chains (SC-SC, Fig. 5A iv) are also considerably high. The effect of such short hydrogen bonds within the intrahelical region seems to have interesting structural consequences, which is discussed later. The SHBs from the BB-SC (Fig. 5A ii) and the SC-BB (Fig. 5A iii) cases show no predominance of a particular secondary structure (intrahelix or strand). This implies that many acceptors are from the turns, loops and other irregular secondary structures. The strand-strand (E-E) interaction is found to be minimal in the short range, as expected.

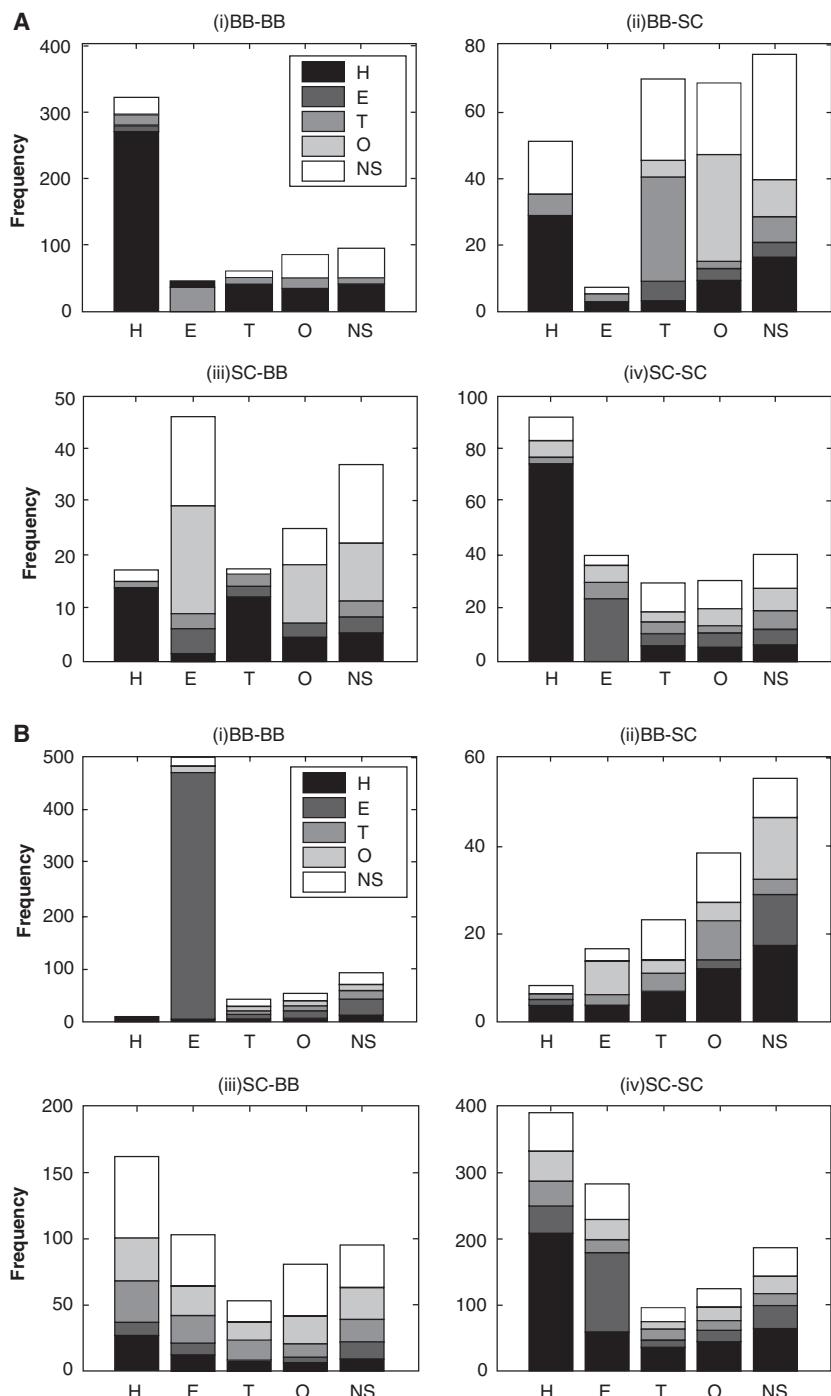


Fig. 5. The frequency bars of the secondary structural distribution (H, E, T, O, NS) in SHB donors (on the x-axis) and acceptors (on the y-axis) as given by AMBER. Different categories of SHBs from the backbone and the side chain are presented as follows: (i) BB-BB, (ii) BB-SC, (iii) SC-BB, (iv) SC-SC. (A) The short-range sequence separation ($D_i - A_j \leq 4$) and (B) the long-range sequence separation ($D_i - A_j > 9$).

In case of the long range SHBs (Fig. 5B), the BB-BB (Fig. 5B i) long range SHBs are essentially dominated by the hydrogen bonds within the extended regions of the strands (E-E). The interstrand interactions are also found in significant numbers in the SC-SC category (Fig. 5B iv). It can be seen that in the BB-BB and the SC-SC cases, SHBs from the regular

secondary structural conformations such as the extended and the helical conformations are observed in large numbers, whereas the BB-SC and the SC-BB cases host mainly irregular conformations akin to the short range SHBs.

The secondary structures of the short range SHBs as given by HBPLUS shows a substantial increase in the

SC-BB and the SC-SC cases. More numbers of helix-helix cases are seen in the former whereas there was no preference for a particular secondary structure in the latter. In case of long range SHBs, there was no difference in the patterns of secondary structure between HBPLUS and AMBER list; however, a general increase in the number of all four different categories are seen in HBPLUS.

The donor-acceptor types and their environment

Amino acid preferences

The SHBs between groups with net charge is well accepted in chemistry and biology [28–30], specifically when the proton acceptor is negatively charged. This is very well reflected in the amino acid preference (of donor and acceptor) of the side chains as shown in Fig. 6B,D (AMBER) and Fig. 7B,D (HBPLUS). On the other hand, no specific amino acid preference is seen for the backbone donors and acceptors (Figs 6A,C and 7A,C). In the case of SC-SC SHBs, the positively charged amino acids (Arg, His and Lys) frequently

pair with the negatively charged Asp and Glu residues. The side chains of polar residues also participate in SHBs. Here we see a major difference in the hydrogen position assignment between AMBER and HBPLUS. A large increase in the HBPLUS list is mainly due to the optimized orientation of the hydroxyl protons of Ser, Thr and Tyr towards the acceptors.

Strikingly a high proportion of Tyr is seen as a donor from both of the AMBER (Fig. 6B) and HBPLUS (Fig. 7B) lists. The negatively charged side chains of Asp and Glu (Figs 6B and 7D) are the acceptors commonly found for these Tyr side chain donors. The analysis on pairing frequencies of donor-acceptors has shown that more than 50% of the acceptors for Tyr side chain donors are from Asp and Glu residues and a small fraction of acceptors are from the side chain of other polar amino acids. Interestingly, more than 30% of the acceptors are the carbonyl oxygen atoms of the backbone. Thus, a significant number of SHBs are seen between pairs of neutral groups formed between some of the uncharged, polar amino acids and the carbonyl oxygen atom of the backbone. As mentioned

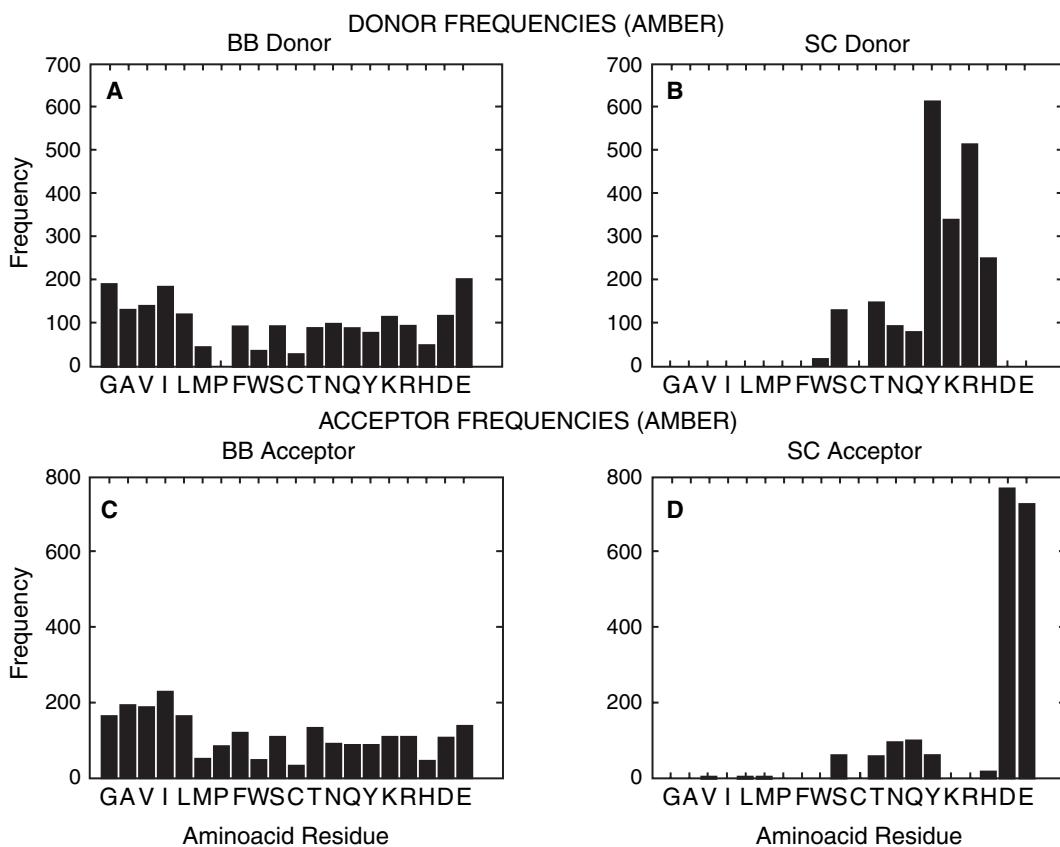


Fig. 6. Amino acid residue-wise frequencies of donors and acceptors in SHBs determined by AMBER from the backbone (A and C) and from the side chains (B and D).

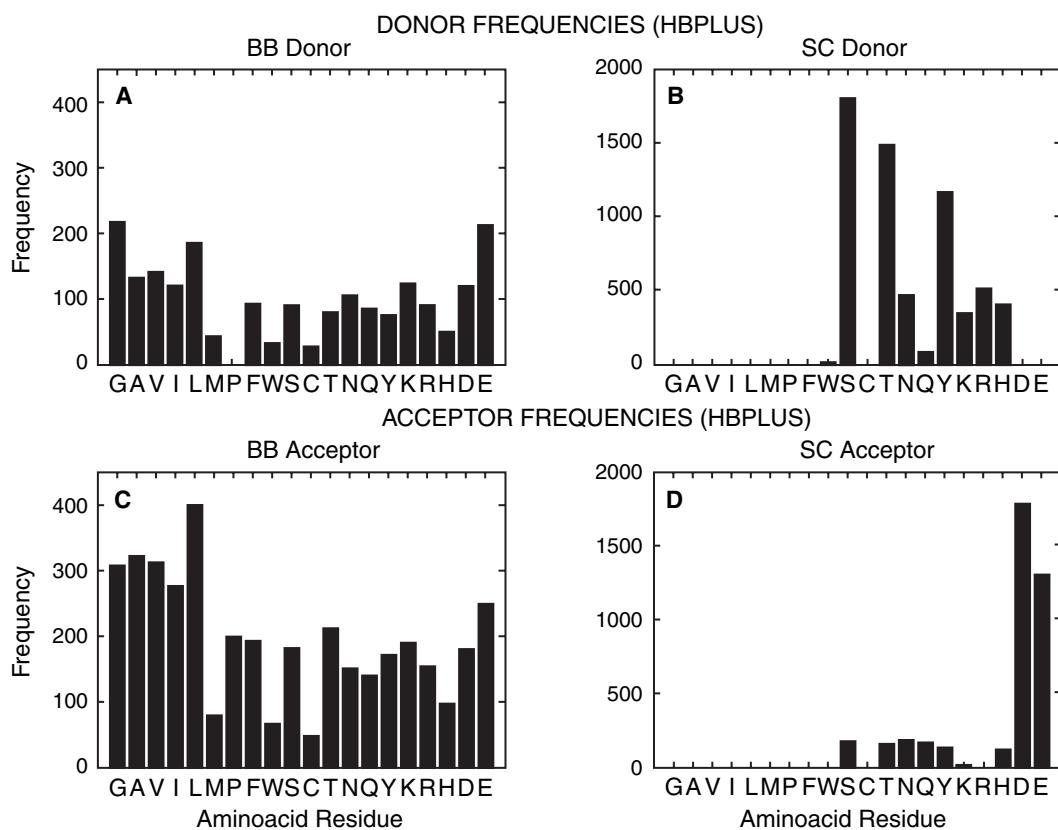


Fig. 7. Amino acid residue-wise frequencies of donors and acceptors in SHBs determined by HBPLUS from the backbone (A and C) from the side chains (B and D).

earlier, such short hydrogen bonds between neutral atoms have also been reported from neutron diffraction studies and the Cambridge data search on small molecules, which have been termed as synthon assisted hydrogen bonds (SAHB) [26].

Thus, a significant number of SHBs has been observed between pairs of neutral groups, contributed to by some of the uncharged, polar amino acids. The participation of the large number of Tyr side chains in SHBs may influence protein structure at the global level and may also influence the function of the protein. For example, a ‘tyrosine corner’ (with an ‘LxPGxY’ sequence motif) is found in Greek key proteins [31] and has been shown to be involved in stabilizing the Greek key connections of the strands. This is characterized by a highly conserved Tyr residue hydrogen bonding to the i-4 backbone N-H as well as to the C=O atoms. In our dataset we could find about 14 well-defined Tyr corner SHBs and additionally nine examples with minor variations to this motif. The presence of an SHB in the tyrosine corner motif emphasizes the requirement of a strong and specific interaction that is necessary to contribute to the stability of the

tertiary structure. The role of these Tyr corners in imparting stability to the tertiary structure has also been experimentally verified [32].

Environment of the backbone donors and acceptors

As mentioned in the introduction, the strength of SHBs between neutral species is highly debated. However, we have encountered a significant number of SHBs in protein structures and a large number of them are observed in the protein BB-BB category. To investigate the reason for these occurrences, we have examined the environment of the hydrogen bonds. Environment induced SHBs have been reported in the active site of several enzymes [12–14] and in enzyme–ligand complexes [18]. We analysed the types of residue side chains that are present in the vicinity of the BB-BB donor–acceptor pairs in the SHBs. All of the side chain atoms that occur within a distance of 4.5 Å from the donor or the acceptor atom are said to form the environment of the residue involved in the formation of SHB. The results presented in Fig. 8 (only AMBER results are presented in this section and in the

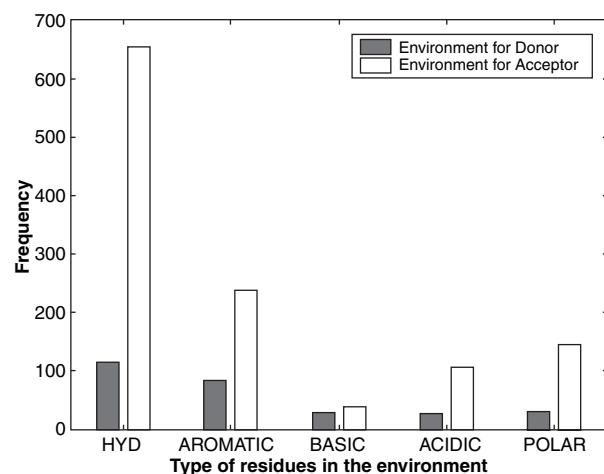


Fig. 8. A plot of the frequencies of the environment (hydrophobic, aromatic, basic, acidic and polar) of the backbone donors (N-H) and the acceptors (C=O) of the AMBER list.

next section) indicate a high number of hydrophobic residues in the neighbourhood of both the donor and acceptors involved in SHB formation. Thus a large number of SHBs are found in the hydrophobic environment of the protein. Furthermore, the presence of a greater number of aromatic side chains in the environment is also quite interesting. Both the charged and the polar environment around the BB-BB SHBs are considerably less, with the polar side chain environment being relatively higher. These results indicate clearly that SHBs could not only exist between neutral donor-acceptor pairs but also could exist in the absence of a charged environment. Perhaps the hydrogen bond between neutral groups in the distance range of 2.5–2.7 Å is energetically reasonable, although it may not be optimal. The energy cost involved in the shortening of the hydrogen bond is probably compensated by the overall optimized geometry of the protein. The correlation of such SHBs with fine-tuned interactions of secondary structures is presented in a later section.

Multiple hydrogen bonds

Hydrogen bonds with multiple donors (acceptor furcation) and multiple acceptors (donor furcation) are known to be common in protein structures [33,34]. The role of SSHBs in protein-ligand complexes has been studied and the donor and acceptor furcation has been analysed in detail [10], from the point of view of recognition of the ligand by the protein. In this analysis, we have investigated the cases of donor (approach of many acceptors towards a donor) and acceptor

(approach of many donors towards an acceptor) furcated multiple hydrogen bonds, where one of them is an SHB and the other is a normal hydrogen bond with regular geometry ($2.7 \text{ \AA} \leq d \leq 3.2 \text{ \AA}$, $\theta \geq 150^\circ$). The analysis shows that about 11% of the total SHBs in the dataset are involved in the formation of multiple hydrogen bonds. The details of donor and acceptor furcation are given in Table 2. The acceptor furcation is more predominantly seen as compared to the donor furcation. This could mainly be due to the fact that the geometrical constraints for acceptor furcation are lower, as the two donor atoms are separated in space without sacrificing the hydrogen bond geometry. On the other hand, the geometrical constraints are greater for donor furcation. A majority of multiple hydrogen bonds are found in the long-range between the side chain donors and acceptors. This is true for both the acceptor and the donor furcated systems.

In the cases where both the donor and the acceptor are from the backbone, a particular pattern of acceptor furcation is observed. The two donors (backbone N-H) are from sequential neighbours (i and $i + 1$). About 26 of these bonds are observed from the dataset. Furthermore, in many of these cases, the acceptor is from the $(i - 3/i - 4)$ residue, either from the backbone C=O group or from a side chain. The evaluation of the ϕ , ψ of the i^{th} and the $(i + 1)$ residues revealed that it belongs to a specific type of beta-turn (type VIII) [35]. Furthermore, the angle $(N_i)-(O)-(N_{i+1})$ showed a consistent value of 57.8° ($\pm 3.3^\circ$). This geometrical pattern can be visualized from Fig. 9. No such specific pattern was seen in the donor furcated cases.

Sulfur-containing SHBs

Sulfur atoms have been known to participate in hydrogen bonds. Gregoret *et al.* [36] have examined the

Table 2. SHB with multiple hydrogen bonds (from the AMBER list).

SHB	SHORT	MEDIUM	LONG	TOTAL
Acceptor furcation				(297)
BB acceptors				
BB(NH) Donors	17	4	9	(30)
SC donors	9	5	16	(30)
SC Acceptors				
BB(NH) Donors	25	6	12	(43)
SC Donors	25	16	163	(194)
Donor furcation				(156)
SC Donors				
BB Acceptors	9	8	41	(58)
SC Acceptors	12	3	83	(98)

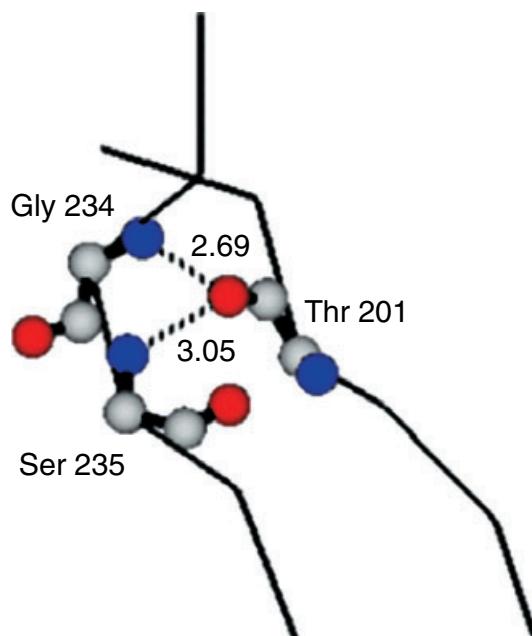


Fig. 9. An example of acceptor furcated multiple hydrogen bond in chlorella virus DNA ligase-adenylate (PDB code 1fviA). The backbone C=O of Thr201 accepts hydrogen from the backbone donors of both Glu234 (i) and Ser235 (i +1). Here, the backbone N-H from Glu234 forms a SHB (2.69 Å) with Thr201 C=O and Ser235 N-H forms a normal hydrogen bond (3.05 Å). Glu234, Ser235 and Thr201 are shown in 'ball and stick' representation and the protein backbone in 'trace'.

prevalence and the geometry of sulfur containing hydrogen bonds in proteins. Their study indicated a substantial number of cysteine-SH-O=C hydrogen bonds at S-O distance around 3.5 Å. In this study, we investigated the possible occurrence of short hydrogen bonds involving sulfur atoms by analysing our dataset of 948 proteins with the distance (D-A) < 3.5 Å and D-H-A angle $\geq 150^\circ$. The distance profile is presented in Fig. 10 and the details of donor-acceptor types are presented in Table 3. We indeed see a significant number of sulfur-containing short hydrogen bonds. As in the case of hydroxyl groups, SH groups of cysteine as donors have been identified by HBPLUS in large numbers, which is not so from the AMBER list. The acceptors for the SH donors are mainly from carbonyl oxygen atoms as seen earlier in the case of the normal hydrogen bonds. We also see a significant number of examples where the sulfur atom of cysteine and methionine act as acceptors.

The SHBs with sulfur atoms were specifically investigated for their location in the three dimensional structure in several proteins. Two examples are presented in Fig. 11. In the case of adenylate cyclase from *Trypanosoma brucei* (PDB code 1fx2A), the sulfur-SHB is

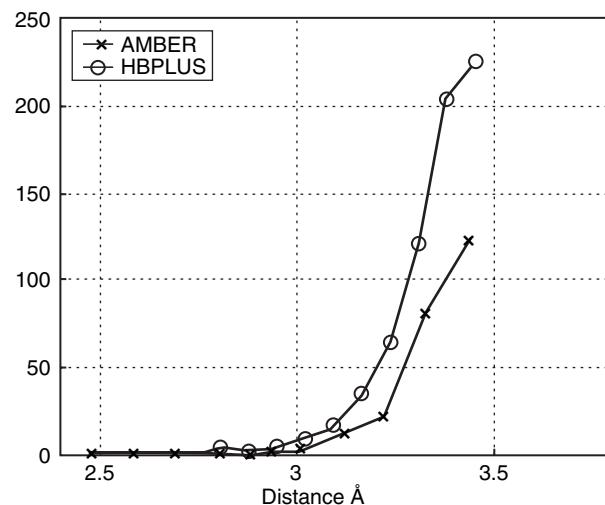


Fig. 10. Distance distributions of sulfur SHBs from AMBER and HBPLUS lists.

Table 3. A list of Sulfur-SHB, the donors, acceptors and the frequencies. BB, backbone; SC, side chain; SH, sulfur-containing group of cysteine; SD, sulfur-containing group of methionine.

Amino acid D-A	AMBER	HBPLUS
CYS(SH)		
SH-O(BB)	4	396
SH-O(SC)	5	71
SH-N(SC)	–	6
N(BB)-SH	119	94
N(SC)-SH	30	19
O(SC)-SH	10	32
MET		
N(BB)-SD	49	30
N(SC)-SD	20	8
O(SC)-SD	8	13
Total	245	680

found to add additional stability to a helix through the formation of a sidechain-backbone sulfur-SHB whereas in *Escherichia coli* cytotoxic necrotizing factor (PDB code 1hq0A), the sulfur-SHB is involved in clamping the ends of two strands in a sheet.

SHBs mediating structural constraints in protein structures

In the above sections we have examined the frequency of occurrence, the residue and environment preferences of SHBs in proteins. In this section, we investigate the role of such hydrogen bonds in the context of protein tertiary structure and its function. These short hydrogen bonds are found to occur in specific regions of the proteins, which seem to contribute to the rigidity of

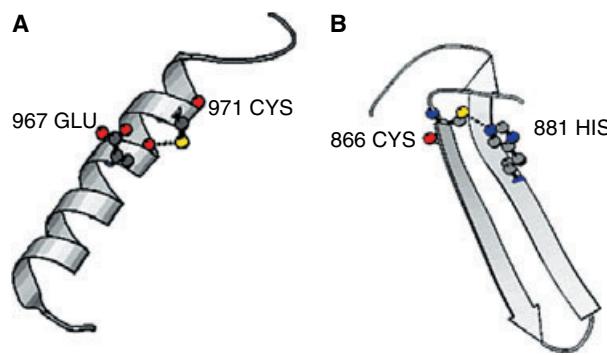


Fig. 11. (a) Sulfur-SHB in adenylate cyclase from *T. brucei* (PDB code 1fx2A). A sulfur-SHB is observed between the S-Gamma atom (SG) of Cys961 with backbone O of Glu967 in a helix. (b) Sulfur-SHB in *E. coli* cytotoxic necrotizing factor (PDB code 1hq0A). The sulfur-SHB is observed between SG of Cys866 with sidechain of His881. It occurs at the edge of the sheet holding the strands in registry.

the local structural region or to the supersecondary structures. By supersecondary structures, we refer to the situation where two or more secondary structures in the vicinity are connected and do not refer to the ideal supersecondary structural motifs as they are often referred to in the literature. The structural constraints mediated through SHBs can be found both at the short range or the local regions and at the long-range interactions of the protein structure.

In case of the short-range ($|D_i - A_j| \leq 4$) hydrogen bonds formed between the backbone donors and acceptors, a majority of cases are found in the intra-helical regions of the protein structure. A turn can be introduced in a helix by one or two residues adopting nonhelical (ϕ, ψ) values. We have seen examples of such turns being stabilized by SHBs. For example, from Fig. 12A it can be seen that in carbamoyl phosphate synthetase (PDB code 1a9xA), the residues Arg675(O) and Gln679(N) form backbone hydrogen bonds in the helix. But Asp674 has a nonhelical ϕ, ψ value of $-88.9^\circ, 112.3^\circ$, and forms a sharp turn. This turn is stabilized by the backbone SHBs that Asp674 forms with both Asp670 and Phe678 backbone atoms, which exist in the helical region (details shown in Fig. 12). Thus, SHBs could in fact stabilize the mutual orientation of two fragments of a helix, which adopts a change of direction at the turn residue. Such short range BB-BB SHBs are also found in β -hairpins as can be seen in copper amine oxidase from *E. coli* (PDB code 1oacA) from Fig. 13.

Long-range SHBs between backbone atoms are commonly found in long β -strands. This can lead to a variety of geometrical consequences. We have presented several examples of SHB occurrence in the protein

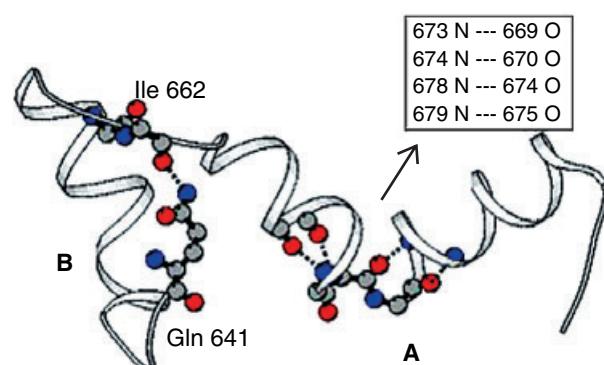


Fig. 12. SHBs in supersecondary structures of proteins (helix-helix, helix-loop) is shown in the example carbamoyl phosphate synthetase (PDB code 1a9xA). A sharp turn stabilizing the flanking helical regions is shown in (A). Four backbone SHBs (NH-CO) from residues 673, 674, 678 and 679 (inset) are involved in stabilizing the sharp turn. From these four residues, residue 674 takes up a non-helical ϕ, ψ and forms the turn, causing a change in the direction of propagation of the helix. (B) An SHB is formed between Gln641 side chain (NE2) and the backbone of Ile662 (C=O) (BB-SC SHB). Both the residues Gln641 and Ile662 are found in loop regions on either side of a helix. The residues involved in the formation of SHB in this figure and all the subsequent figures are shown in 'ball and stick' representation and the secondary structures (helices and strands) in 'cartoon' representation.

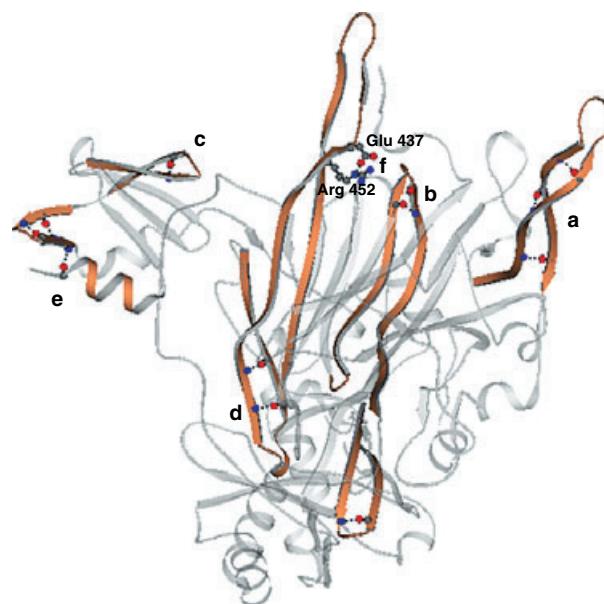


Fig. 13. SHBs in extended strands from copper amine oxidase. The SHB-containing strands in different regions (a-f) of the protein are coloured orange. The regions correspond to: (a) the strands bending together; (b) strands moving away in different directions; (c) beginning of the sheet; (d) the end of the sheet; (e) between a strand and helix; and (f) between the side chains Arg452 and Glu437 of strands.

copperamine oxidase, like simultaneous bend of strands (Fig. 13, a), strands moving away in different directions (Fig. 13, b), SHBs in the beginning of the sheet (Fig. 13, c), at the end of the sheet (Fig. 13, d) and between a strand and helix (Fig. 13, e). Two or more backbone hydrogen bonds are also involved in stabilizing long strands as shown in Fig. 13. There is also an instance of SC-SC SHB observed between residues Arg452 and Glu437 in the long strand, which might provide additional rigidity to the region of the strand that bends subsequently (Fig. 13, f).

While the extended β -sheets formed from sequentially well-separated regions of the protein chain are stabilized by backbone hydrogen bonds, there is no such well-defined force existing between interhelical interactions found in large helical bundles. The stability of helix-helix interacting motifs is due to a variety of reasons [37], such as the charge-charge, hydrophobic or hydrogen bond interactions of the side chains. From this analysis, it is interesting to see a large number of interhelical stabilizations through SC-SC SHBs. Examples of helix-helix interaction mediated by long-range SC-SC SHB are given in Fig. 14. A simple helix-helix interaction through Tyr76 and Asp32 side chains in haemolysin from *E. coli* (PDB code 1qoyA) is shown in Fig. 14A. The stabilization of the mutual orientation of three helices in the diphtheria toxin repressor (PDB code 2dtr-) through two SHBs is shown in Fig. 14B. The SHBs between Arg69-Glu19 and Arg77-Glu113 (both arginines are from the same helix) have aided in packing of the three helices. In another instance, a

cluster of SHBs between Tyr208-Asp29 and Tyr208-Thr94 and between residues Asn177-Glu44 (in ribonucleotide reductase protein R2F; 1kgnA) has brought the helices, which are far away in sequence, to spatial proximity (Fig. 14C). Thus SHBs are found in helix-helix orientation and stabilization, and might contribute to helix packing and stability of these supersecondary structures.

Apart from interhelical stabilization, SHBs involving side chains are also observed in intrahelical interactions. An example is shown in Fig. 15A where short-range SC-SC SHB is formed by Arg537-Glu533 and Lys592-Glu589 in leukotriene A4 hydrolase (PDB code 1hs6A). Furthermore, the involvement of side chain SHBs in stabilizing other types of secondary structures namely the loops, bends and turns, are also seen. For example, an SHB between side chains of Tyr214-Glu217 is found to occur in a region where a β -strand bends as shown in Fig. 15B. An SHB between Ile662 in the backbone and Gln641 in the side chain is found in carbamoyl phosphate synthetase (Fig. 12B). This SHB actually bridges the two loops from either side of the helix and might restrict the conformational flexibility of the loop in the protein structure.

Thus the SHBs from the dataset are observed in different secondary (or supersecondary) structural regions. These SHBs in general are found to be involved in stabilizing the three dimensional structures of proteins at specific structural locations as discussed above. A single protein such as copper amine oxidase

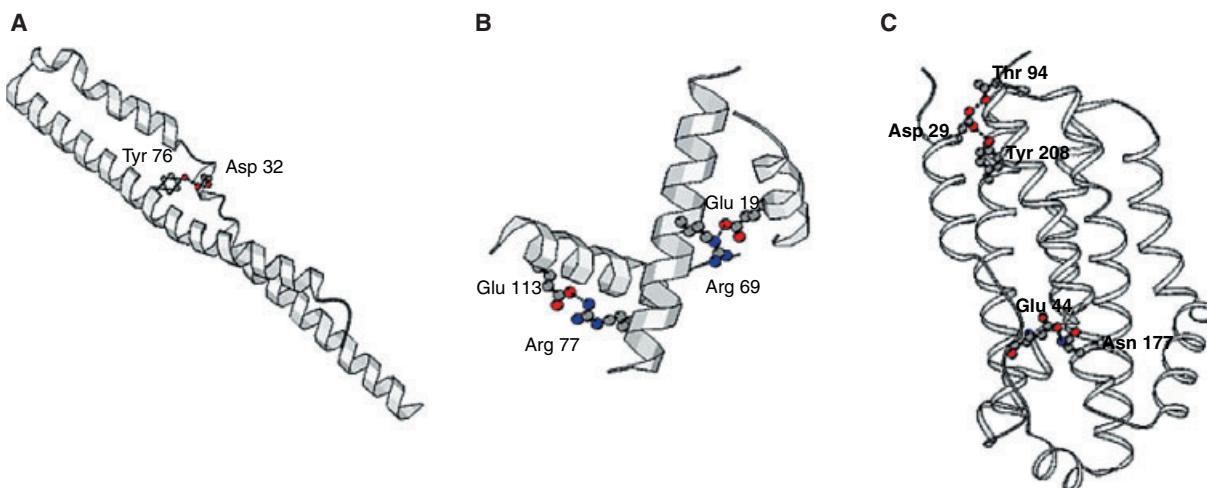


Fig. 14. Examples of long-range SHBs between helices. (A) SHB between side chains of Tyr76(OH) and Asp32(OD1) in haemolysin from *E. coli*. (B) In diphtheria toxin repressor, mutual orientation of three helices is controlled by two SHBs. These are formed between the side chains of Arg77(NE) and Glu113(OE1), and Arg69(NH1) and Glu19(OE1), where both arginines are from the same helix. (C) A cluster of SHBs from side chains of Asp29, Thr94 and Tyr208 from different helices in spatial proximity in ribonucleotide reductase. Another side chain SHB [Asn177(ND2)-Glu44(OE2)] occurring in middle of the helical bundle is also shown.

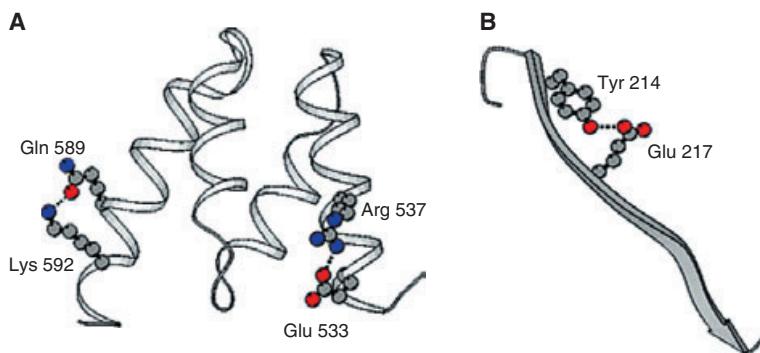


Fig. 15. Short range SHBs in secondary structures. (A) Intrahelical SHBs between the side chain atoms of residues Lys592(NZ) and Gln589(OE1), and Arg537(NH1) and Glu533(OE1) in leukotriene A4 hydrolase. (B) Intrastrand SHB between side chain atoms of residues Tyr214(OH) and Glu217(OE1) in *Pseudomonas* serine carboxyl proteinase.

has 31 numbers (42 from HBPLUS) of SHBs and several of them in crucial locations are shown in Fig. 13. Furthermore, it is amazing to see a large number (> 50) of such interactions in single proteins as in case of enzymes carbamoyl phosphate synthetase, malate synthase G (PDB code 1d8cA) and glucose oxidase (PDB code 1gpeA). SHBs are generally associated with higher interaction strength, as evidenced in many chemical and biological systems. Most of the strong SHBs are due to either direct or indirect participation of charged groups. We have seen a large number of such examples in protein structures, which perhaps contribute to the increased interaction strength. Interestingly, we have also observed a large number of SHBs between neutral backbone atoms in the hydrophobic environment. We believe that such SHBs have formed not necessarily because of increased strength of specific interaction, but also because they fine-tune the overall tertiary structure of proteins.

Conclusions

An analysis of a large number of proteins from a non-redundant dataset has shown that short hydrogen bonds (SHB) frequently occur in protein structures. As expected, many of them are found between charged groups of Arg, Lys, His, Asp and Glu amino acids. However, the polar groups also substantially contribute to SHBs. In particular, it is interesting to note the participation of a large number of tyrosine residues in the formation of SHBs. Surprisingly, many SHBs are also seen between the neutral groups of backbone (N-H and C=O) atoms. They occur in all types of environment, with hydrophobic and aromatic residues being substantially larger than the polar and charged residues. SHBs involving sulfur atoms have also been identified. Our analysis has shown that SHBs are present in a variety of secondary structural regions such as α -helix, β -strands, turns and loops, and contribute to the structural constraints required for the tertiary

structural integrity of proteins. Thus, the occurrence of SHBs in proteins could either be due to the need for a strong interaction or due to structural necessity with implied structural constraints dictated by the three dimensional structure of proteins. These short hydrogen bonds should be taken into account in protein structure modelling and site-directed mutagenesis experiments.

Experimental procedures

A nonredundant dataset of 948 protein chains with resolution better than 2.0 Å and R factor ≤ 0.20 was obtained from [38]. The tLEaP and CARNAL modules of the AMBER7 package [39] was used to fix the hydrogen atoms and to determine the hydrogen bonds from the protein coordinate files. Hydrogen atoms were also fixed by HBPLUS [40] and the results are compared. The SHBs involving nitrogen and oxygen atoms have been defined according to the distance and the angle criteria, $d(D-A) < 2.7$ Å and $\theta(D-H-A) \geq 150^\circ$. A distance of $d(D-A) < 3.5$ Å was used when a sulfur atom (as donor or acceptor) was involved. The secondary structural assignments (H, helix; E, extended; T, turn; O, other conformations eg S-bend, G-3₁₀ helix, B residue in isolated β -bridge, etc.) of the donor as well as the acceptor residues in the protein structures were obtained from the DSSP [41] program. The amino acid residues for which DSSP could not assign any structural annotation were labelled as nonstructured (NS). The donor-acceptor pairs are represented in terms of secondary structures; for example (H-NS) represents the donor and acceptor atoms, respectively, from the α -helix and the non-structured conformations. The proteins were visualized using programs VMD [42] and the figures were prepared using MOLSCRIPT [43].

Acknowledgements

The motivation for this work came from the Discussion Meeting on 'Intermolecular Interactions' sponsored by the Indian Academy of Sciences in Coorg

(December 2003). We thank Sanjeev B.S. and Anandhi S. for useful discussions on technical details and acknowledge the support from the Computational Genomics Initiative at the Indian Institute of Science, funded by the Department of Biotechnology (DBT), India.

References

- 1 Baker EN & Hubbard RE (1984) Hydrogen bonding in globular proteins. *Prog Biophys Mol Biol* **44**, 97–179.
- 2 McDonald IK & Thornton JM (1994) Satisfying hydrogen bonding potential in proteins. *J Mol Biol* **238**, 777–793.
- 3 Stickle DF, Presta LG, Dill KA & Rose GD (1992) Hydrogen bonding in globular proteins. *J Mol Biol* **226**, 1143–1159.
- 4 Chakrabarti P & Chakrabarti S (1998) C–H–O hydrogen bond involving proline residues in α helices. *J Mol Biol* **284**, 867–873.
- 5 Madan Babu M, Kumar Singh S & Balaram P (2002) A C–H triplebond O hydrogen bond stabilized polypeptide chain reversal motif at the C terminus of helices in proteins. *J Mol Biol* **322**, 871–880.
- 6 Manikandan K & Ramakumar S (2004) The occurrence of C–H...O hydrogen bonds in alpha-helices and helix termini in globular proteins. *Proteins* **56**, 768–781.
- 7 Steiner T (1995) Water molecules which apparently accept no hydrogen bonds are systematically involved in C–H–O interactions. *Acta Crystallogr D Biol Crystallogr* **51**, 93–97.
- 8 Jeffrey G, A & Saenger W (1991) *Hydrogen Bonding in Biological Structure*. Springer, Berlin.
- 9 Steiner T & Koellner G (2001) Hydrogen bonds with π acceptors in proteins: Frequencies and role in stabilising local 3D structures. *J Mol Biol* **305**, 535–557.
- 10 Sarkhel S & Desraju GR (2004) N–H–O, O–H–O, and C–H–O hydrogen bonds in protein-ligand complexes: strong and weak interactions in molecular recognition. *Proteins* **54**, 247–259.
- 11 Shuou S, Loh S & Herschlag D (1996) The energetics of hydrogen bonds in model system: Implications for enzyme catalysis. *Science* **272**, 97–101.
- 12 Stranzl GR, Gruber K, Steinkellner G, Zanger K, Schwab H & Kratky C (2004) Observation of a short, strong hydrogen bond in the active site of hydroxynitrile lyase from *Hevea brasiliensis* explains a large pKa shift of the catalytic base induced by the reaction intermediate. *J Biol Chem* **279**, 3699–3707.
- 13 Nishina Y, Sato K, Tamaoki H, Tanaka T, Setoyama C, Miura R & Shiga K (2003) Molecular mechanism of the drop in the pKa of a substrate analog bound to medium-chain acyl-CoA dehydrogenase: implications for substrate activation. *J Biochem* **134**, 835–842.
- 14 Huang X, Zeng R, Ding X, Mao X, Ding Y, Rao Z, Xie Y, Jiang W & Zhao G (2002) Affinity alkylation of the Trp-B4 residue of the beta-subunit of the glutaryl 7-aminocephalosporanic acid acylase of *Pseudomonas* sp. 130. *J Biol Chem* **277**, 10256–10264.
- 15 Kim KS, Oh KS & Lee JY (2000) Catalytic role of enzymes: Short strong H-bond-induced partial proton shuttles and charge redistributions. *Proc Natl Acad Sci USA* **97**, 6373–6378.
- 16 Lauble H, Kennedy MC, Emptage MH, Beinert H & Stout CD (1996) The reaction of fluorocitrate with aconitase and the crystal structure of the enzyme-inhibitor complex. *Proc Natl Acad Sci USA* **93**, 13699–13703.
- 17 Katz BA, Spencer JR, Elrod K, Luong C, Mackman RL, Rice M, Sprengeler PA, Allen D & Janc J (2002) Contribution of Multicentered Short Hydrogen Bond Arrays to Potency of Active Site-Directed Serine Protease Inhibitors. *J Am Chem Soc* **124**, 11657–11668.
- 18 Vishveshwara S, Madhusudhan MS, Maizel JV Jr (2001) Short-strong hydrogen bonds and a low barrier transition state for the proton transfer reaction in RNase A catalysis: a quantum chemical study. *Biophys Chem* **89**, 105–117.
- 19 Schutz CN & Warshel A (2004) The low barrier hydrogen bond (LBHB) proposal revisited: the case of the Asp...His pair in serine proteases. *Proteins* **55**, 711–723.
- 20 Perrin CL & Nielson JB (1997) ‘Strong’ hydrogen bonds in chemistry and biology. *Annu Rev Phys Chem* **48**, 511–544.
- 21 Ash EL, Sudmeier JL, Day RM, Vincent M, Torchilin EV, Haddad KC, Bradshaw EM, Sanford D, G, Bachovchin W & W (2000) Unusual 1H NMR chemical shifts support (His) C (epsilon) 1...O==C H-bond: proposal for reaction-driven ring flip mechanism in serine protease catalysis. *Proc Natl Acad Sci USA* **97**, 10371–10376.
- 22 Anderson S, Crosson S & Moffat K (2004) Short hydrogen bonds in photoactive yellow protein. *Acta Cryst D* **60**, 1008–1016.
- 23 Schiott B (2004) The influence of solvation on short strong hydrogen bonds: a density functional theory study of the Asp–His interaction in subtilisins. *Chem Commun (Camb)* **5**, 498–499.
- 24 Kim KS, Suh SB, Kim JC, Hong BH, Lee EC, Yun S, Tarakeshwari P, Lee JY, Kim Y, Ihm H, Kim HG, Lee JW, Kim JK, Lee HM, Kim D, Cui C, Youn SJ, Chung HY, Choi HS, Lee CW, Cho SJ, Jeong S & Cho JH (2002) Assembling phenomena of calix[4]hydroquinone nanotube bundles by one-dimensional short hydrogen bonding and displaced pi-pi stacking. *J Am Chem Soc* **124**, 14268–14279.
- 25 Suh SB, Kim JC, Choi YC, Yun S & Kim KS (2004) Nature of one-dimensional short hydrogen bonding: bond distances, bond energies, and solvent effects. *J Am Chem Soc* **126**, 2186–2193.

- 26 Vishveshwara P, Babu NJ, Nangia A, Mason SA, Puschmann H, Mondal R & Howard JAK (2004) Variable temperature neutron diffraction analysis of a very short O-H—O hydrogen bond in 2,3,5,6-pyazine-tetracarboxylic acid dihydrate: Synthon-assisted short O_{acid}-H – O_{water} hydrogen bonds in a multicellular array. *J Phys Chem A* **108**, 9406–9416.
- 27 Eshwar N & Ramakrishnan C (2000) Deterministic features of side chain main chain hydrogen bonds in globular structures. *Protein Engineering* **15**, 227–238.
- 28 Kim Y, Lim S & Kim Y (1999) The role of a Short Strong Hydrogen bond on the double proton transfer in the formamide-formic acid complex: Theoretical studies in gas phase and in solution. *J Phys Chem* **103**, 6632–6637.
- 29 Cho HS, Ha NC, Choi G, Kim HJ, Lee D, Oh KS, Kim KS, Lee W, Choi KY & Oh BH (1999) Crystal structure of δ (5)-3-ketosteroid isomerase from *Pseudomonas testosteroni* in complex with equilenin settles the correct hydrogen bonding scheme for transition state stabilization. *J Biol Chem* **274**, 32863–32868.
- 30 Wu ZR, Ebrahimian S, Zawrotny ME, Thornburg LD, Perez-Alvarado GC, Brothers P, Pollack RM & Summers MF (1997) Solution structure of 3-oxo-delta5-steroid isomerase. *Science* **276**, 415–418.
- 31 Hemmingsen JM, Kim M, Gernert Richardson JS & Richardson DS (1994) The tyrosine corner: a feature of most Greek key beta barrel proteins. *Protein Sci* **3**, 1927–1937.
- 32 Hamill SJ, Cota E, Chothia C & Clarke J (2000) Conservation of folding and stability within a protein family: the tyrosine corner as an evolutionary cul-de-sac. *J Mol Biol* **295**, 641–649.
- 33 Preissner R, Enger U & Saenger W (1991) Occurrence of bifurcated three-center hydrogen bonds in proteins. *FEBS Lett* **288**, 192–196.
- 34 Fain AV, Berezovsky IN, Chekhov VO, Ukrainskii DL & Esipova NG (2001) Double and bifurcated Hydrogen bonds in the alpha helices of globular proteins. *Biophysics* **46**, 969–977.
- 35 Wilmot CM & Thornton JM (1988) Analysis and prediction of the different types of beta-turn in proteins. *J Mol Biol* **203**, 221–232.
- 36 Gregoret LM, Rader SD, Fletterick RJ & Cohen FE (1991) Hydrogen bonds involving sulphur atoms in proteins. *Proteins* **9**, 99–107.
- 37 Crick FHC (1953) The packing of alpha helices: simple coiled coils. *Acta Cryst* **6**, 689–697.
- 38 Wang G & Dunbrack RL Jr (2003) PISCES: a protein sequence culling server. *Bioinformatics* **19**, 1589–1591.
- 39 Amber, Version 7 (2002) University of California, San Francisco.
- 40 McDonald IK & Thornton JM (1994) Satisfying hydrogen bonding potential in proteins. *J Mol Biol* **238**, 777–793.
- 41 Kabsch W & Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **12**, 2577–2637.
- 42 Humphrey W, Dalke A & Schulten K (1996) 'VMD – visual molecular dynamics'. *J Mol Graphics* **14**, 33–38.
- 43 Kraulis PJ (1991) MOLSCRIPT - a program to produce both detailed and schematic plots of protein structures. *J Appl Crys* **24**, 946–950.

Supplementary Material

The following material is available from <http://www.blackwellpublishing.com/products/journals/suppmat/EJB/EJB4604/EJB4604sm.htm>

Figure S1. Histogram of B factors of residues participating in normal and SHBs.

Figure S2. (A) Secondary structures of donor–acceptors in short range SHBs given by HBPLUS. (B) Secondary structures of donor–acceptors in long range SHBs given by HBPLUS.

Table S1. List of neutron diffracted structures considered for this study.

Table S2. Comparison of neutron diffraction and X-ray structures. A case study (myoglobin).