Protein Engineering vol.15 no.5 pp.359–363, 2002

Geometric criteria of hydrogen bonds in proteins and identification of 'bifurcated' hydrogen bonds

Ivan Y.Torshin[1,2,3], Irene T.Weber[1] and Robert W.Harrison[1,4]

[1]Biology Department, and [4]Department of Computer Science, Georgia State University, Atlanta, GA 30303, USA, [2]Laboratory of Kinetics and Catalysis, Chair of Physical Chemistry, Chemistry Department, Moscow State University, Moscow 119899, Russia

[3]To whom correspondence should be addressed at: 410 NSC, 402 Kell Hall, Biology Department, Georgia State University, 24 Peachtree Ctr Avenue, Atlanta, GA 30303, USA.
E-mail: biotiy@suez.cs.gsu.edu; tiy1@mail.com

Empirical criteria for identification of hydrogen bonds were analyzed to produce a set of geometrically consistent criteria. For a data set of 30 structures, application of a set of purely geometrical criteria, along with exclusion of abnormal backbone conformations, also excluded a common interaction of Ser/Thr side chains with Asp/Glu side chains ([ST]/[DE] pairs). These interactions were termed 'bifurcated hydrogen bonds', which implies delocalization of a positively charged hydrogen of hydroxyl between the two acceptor atoms of the carboxylic group. These 'bifurcated' interactions are among the most common packing patterns for [ST]/[DE] pairs of side chains. Therefore, the identification of hydrogen bonds cannot be based on geometrical criteria only and requires introduction of some physico-chemical criteria.

Keywords: hydrogen-bonding criteria/protein tertiary structure/side chain–side chain packing

Introduction

The hydrogen bond is an interaction between two electronegative atoms, donor and acceptor, through an intermediate atom of hydrogen that is covalently connected with the donor (Latimer and Rodebush, 1920; Huggins, 1971; Baker and Hubbard, 1984; Ippolito *et al.*, 1990; Stickle *et al.*, 1992). The electron density of the bond donor atom-hydrogen is shifted to the donor atom, thus a positive charge is induced on the hydrogen atom. This partial charge interacts with the electronic cloud of the acceptor atom. Unlike the covalent bond, the hydrogen bond is a multipole interaction involving at least three atoms (D, H and A in Figure 1). A triangle DHA may be described by the lengths of the three sides, or, by the length of sides DH, HA and the angle DHA. The latter would be more compatible with the description of covalent bonds.

Hydrogen bonds are classified as ionic interactions (Baker and Hubbard, 1984; Ippolito *et al.*, 1990). Therefore, unlike a covalent bond, hydrogen bonds are characterized not by specific bond lengths and angles, but by using broader ranges of values. In order to determine these ranges for proteins, it is necessary to carry out measurements of these parameters in a number of protein structures. Such measurements have been periodically conducted for relatively large sets of the structures, recent results are presented in the papers (Ippolito *et al.*, 1990; Stickle *et al.*, 1992; McDonald and Thornton, 1994).

When such measurements are performed on a set of proteins, the results are assessed in terms of 'statistical significance': for example, 95% of all analyzed contacts in the work (McDonald and Thornton, 1994) could be identified using the rules 1–5 analyzed below. The remaining 5% or exceptional cases are not usually analyzed in detail. Therefore, it is not clear whether the remaining 5% of the cases were excepted due to errors in structural data or they represent chemical classes of the hydrogen bonds that are not easily identified by geometric-only criteria. Thus, in the present work, similar measurements were performed on a set of protein structures in order to consider in detail these exceptional cases: whether some specific 'chemical' type of bonds would be excluded from the list of hydrogen bonds using the proposed criteria which are analyzed below.

Materials and methods

Geometric criteria of hydrogen bond identification

In one of the recent studies on hydrogen bond criteria in proteins (McDonald and Thornton, 1994) results of the statistical analysis for approximately 50 protein structures were presented. The following empirical rules of hydrogen bond identification were proposed: rule 1, D–A <3.9 Å; rule 2, H–A <2.5 Å; rule 3, D–H–A $>90.0°$; rule 4, A′–A–D $>90.0°$; rule 5, A′–A–H $>90.0°$. These rules were analyzed in order to define a minimal geometrically consistent set of the criteria. Assuming that D and H coordinates are given, let us consider a set of all possible acceptor coordinates called the 'A-set' (comprised of 'A-dots' each of which represents potential coordinate of an acceptor atom). As a 'criterion' will be considered such a geometrical condition (rule), which lessens the volume of the 'A-set' (Figure 1). Let's consider the rules one by one.

Rules 1 and 2

A-dots, satisfying the rule 1, are placed inside sphere D with radius 3.9 Å, circumscribed around point D, whereas dots, satisfying the rule 2 are all inside sphere H with radius of 2.5 Å, circumscribed around H (Figure 1b). The atom D is placed inside sphere H. As distance DH is ~1 Å, sphere H is always placed inside the sphere D with minimum distance between surfaces of these spheres equal to $3.9 - 2.5 - 1 = 0.4$ Å. Therefore, the dots that are placed inside the sphere H in any case belong to the sphere D, in other words, the dots that satisfy rule 2, in any case satisfy rule 1. Thus, rule 1 is redundant for hydrogen bond identification using the given boundary distances and hydrogen coordinates. Rule 1, although it may be applied for rough estimates, could not be called a 'criterion', since its application does not narrow the A-set (sphere H). At the same time, rule 2 is a criterion.

Rule 3

Applying restrictions for the angle DHA narrows the sphere H into a spherical cone with the axis DH (cone A, Figure 1c) and with a cone angle equal to 2(180 – angle_DHA) for the

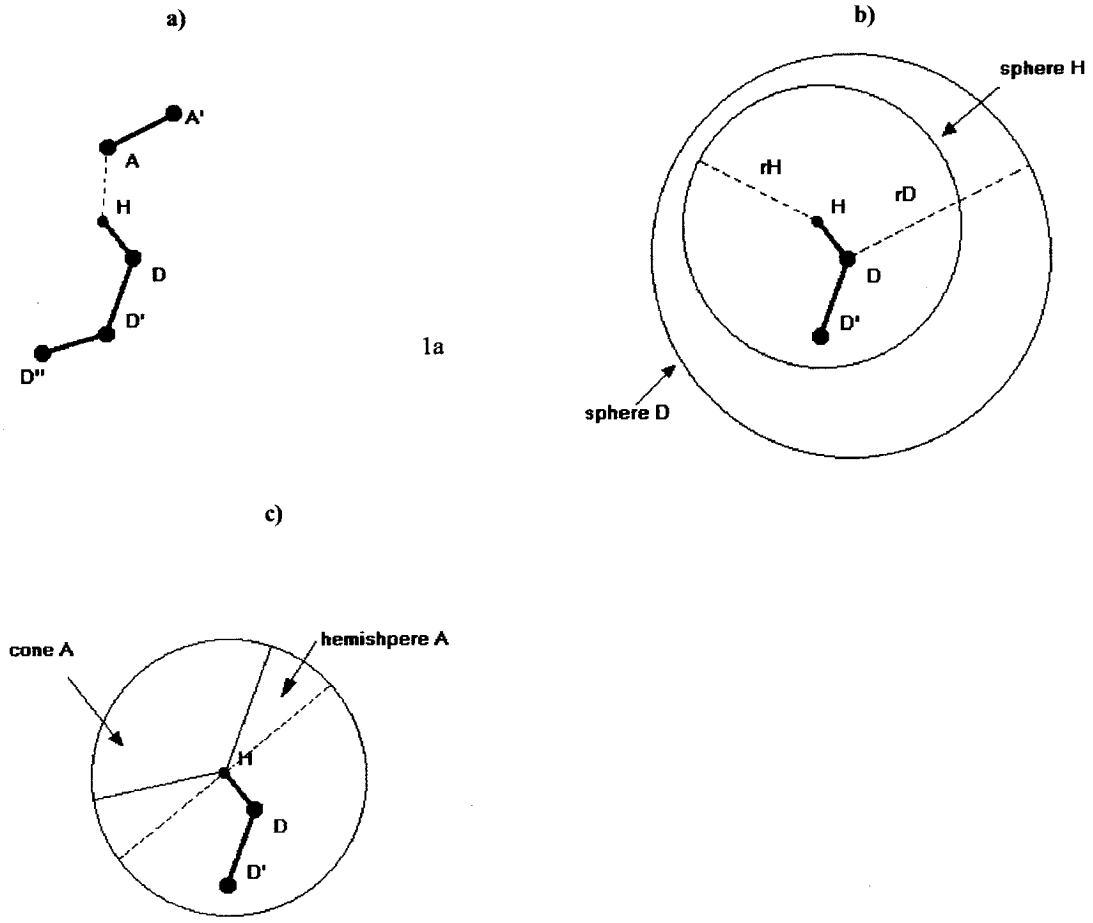

Fig. 1. (**a**) Spatial model of a hydrogen-bonding interaction. Atoms participating in formation of the hydrogen bond and their nearest neighbors are shown. D, a hydrogen donor; H, an atom of hydrogen; A, a hydrogen acceptor; D′, an atom that has a covalent bond with the donor; D″, an atom that is covalently bound to D′; A′, an atom that has a covalent bond with the acceptor; (**b**) Applying restrictions on distances D–A and H–A (rD = 3.9 Å; rH = 2.5 Å). (**c**) Applying restrictions on angle D–H–A. Explanations are provided in the text.

angles >90°. At DHA = 90°, the cone becomes a hemisphere (hemisphere A, Figure 1c). Therefore, this rule is a criterion.

Rules 4 and 5

In the case when the structure of a protein does not contain violations of the molecular geometry, restrictions on the A′–A–D and A′–A–H angles would not change the A-set, since the A′ atom is not directly involved in the hydrogen-bonding interaction.

Thus, out of the five rules above, only rules 2 and 3 may be definitely called 'criteria'. However, rules 4 and 5 may be necessary for sieving out some interactions, corresponding to abnormal conformations of the backbone. On other hand, theoretically, the energy of the hydrogen bond is dependent on configurations of neighboring atoms around the sphere H. In this case, extremely low values of angles A′–A–D and A′–A–H will correspond to the introduction of an additional atom (atom A′) in the region of the hydrogen-bonding interaction, thus influencing, to some extent, the stability of the bonding interaction. Therefore, these two rules should be analyzed on a set of protein structures.

Calculation of hydrogen atom positions in proteins

In X-ray crystallography, the atomic coordinates are determined by fitting a model to the experimental electron density map. The structures used for this study were determined from crystals that diffracted between 2.5 and 1.5 Å. In this range of resolutions, positions of the hydrogen atoms cannot be determined. Thus, positions of hydrogen atoms were calculated using coordinates of the donor, angle D′–D–H and distance D–H. For sp3 donors (Ser and Thr with 'rotating' hydrogen atoms), Lys and Tyr, possible coordinates of hydrogen atoms were analyzed during identification of hydrogen-bonding interactions (as described later). Hydrogen donor protein atoms analyzed in this study were: NH of the main chain, ARG NE, ASN ND2, HIS NE2, SER OG, TYR OH, ARG NH1, CYS SG, HIS ND1, THR OG1, ARG NH2, GLN NE2, LYS NZ and TRP NE1; acceptor atoms were: carboxyl oxygen of the main chain, ASN OD1, GLN OE1, MET SD, ASP OD1, GLU OE1, SER OG, ASP OD2, GLU OE2, THR OG1,CYH SG, HIS ND1 and TYR OH.

Identification of hydrogen bonds

Each structure was analyzed in five steps.

Step 1. Positions of the hydrogen atoms were calculated, with geometry for individual residues as in Stickle *et al.* (Stickle *et al.*, 1992) and McDonald and Thornton (McDonald and Thornton, 1994). For 'rotating' hydrogen atoms in Ser, Thr and Lys, the hydrogen positions, corresponding to all the energy minima, were taken (and then analyzed as described further). For Tyr, only the two possible positions were taken

Table I. Hydrogen bonds with angles A′–A–H and A′–A–D <90° ('non-4,5') in a number of protein structures

PDB	RES	nres	nhb	Non-4,5	MC–MC	SC–MC	SC–SC	s.b.	NH2..OOC	OH..OOC	NH..OH
1bz0 B	1.50	146	126	6 (4.8%)	2 (1.6%)	3 (2.4%)	1 (0.8%)	1	0	0	
1hpg A	1.50	187	136	10 (7.4%)	7 (5.1%)	2 (1.5%)	1 (0.7%)	1	0	0	
1cqw A	1.50	295	246	9 (3.7%)	4 (1.6%)	3 (1.2%)	2 (0.8%)	2	0	0	
1qhv A	1.51	195	162	10 (6.2%)	4 (2.5%)	4 (2.5%)	2 (1.2%)	1	0	0	1
1qtp A	1.60	247	207	5 (2.4%)	1 (0.5%)	3 (1.4%)	1 (0.5%)	1			
1a8u A	1.60	277	249	5 (2.0%)	1 (0.4%)	2 (0.8%)	2 (0.8%)	1	0	1	
2tmn E	1.60	316	275	12 (4.4%)	6 (2.2%)	5 (1.8%)	1 (0.4%)	0	0	1	
1ajs A	1.60	412	338	9 (2.7%)	2 (0.6%)	3 (0.9%)	4 (1.2%)	0	0	1	2
1bu7 A	1.65	455	405	15 (3.7%)	2 (0.5%)	4 (1.0%)	8 (2.2%)	7	1	0	
1qsa A	1.65	618	569	14 (2.5%)	7 (1.2%)	3 (0.5%)	4 (0.7%)	2	1	1	
1bkp B	1.70	278	227	4 (1.8%)	2 (0.9%)	0 (0.0%)	2 (0.9%)	0	0	2	
1bf6 B	1.70	291	257	14 (5.4%)	3 (1.2%)	5 (1.9%)	6 (2.3%)	0	0	6	
1qfz A	1.70	308	236	7 (3.0%)	1 (0.4%)	2 (0.8%)	3 (1.7%)	3	0	0	
1mty B	1.70	384	362	6 (1.7%)	0 (0.0%)	3 (0.8%)	3 (0.8%)	3			
1xgs A	1.75	295	224	5 (2.2%)	3 (1.3%)	2 (0.9%)	0 (0.0%)				
1b32 A	1.75	517	486	21 (4.3%)	11 (2.3%)	4 (0.8%)	5 (1.2%)	1	0	3[a]	
1ppk E	1.80	323	262	8 (3.1%)	6 (2.3%)	2 (0.8%)	0 (0.0%)	2	0	1	
1gd1 P	1.80	334	281	14 (5.0%)	5 (1.8%)	7 (2.5%)	2 (0.7%)	2	0	0	
2udp A	1.80	338	313	13 (4.2%)	7 (2.2%)	6 (1.9%)	0 (0.0%)				
1b6h A	1.80	517	485	19 (3.9%)	7 (1.4%)	3 (0.6%)	9 (1.9%)	3	1	3	1
1a9x A	1.80	1058	981	34 (3.5%)	10 (1.0%)	8 (0.8%)	16 (1.6%)	8	1	5	2
1ai9 B	1.85	192	137	10 (7.3%)	5 (3.6%)	3 (2.2%)	2 (1.5%)	2	0	0	
1jkm B	1.85	361	299	7 (2.3%)	3 (1.0%)	1 (0.3%)	3 (1.0%)	0	0	2	1
7ahl B	1.90	293	218	18 (8.3%)	5 (2.3%)	8 (3.7%)	4 (2.3%)	2	1	1	
1qqo A	1.90	378	315	17 (5.4%)	9 (2.9%)	5 (1.6%)	3 (1.0%)	2	0	1	
1tf4 B	1.90	605	518	15 (2.9%)	8 (1.5%)	2 (0.4%)	5 (1.0%)	2	1	2	
1a1x _	2.00	108	68	1 (1.5%)	1 (1.5%)	0 (0.0%)	0 (0.0%)				
1ebg A	2.10	436	410	25 (6.1%)	15 (3.7%)	2 (0.5%)	8 (2.0%)	4	0	2[b]	
5gpb _	2.30	842	806	47 (5.8%)	11 (1.4%)	7 (0.9%)	29 (3.6%)	21	4	0	4
1hdx A	2.50	374	282	23 (8.2%)	11 (3.9%)	5 (1.8%)	7 (2.5%)	5	1	1	
1alh A	2.50	449	417	17 (4.1%)	5 (1.2%)	5 (1.2%)	7 (1.7%)	1	2	3[c]	
1pky A	2.50	470	385	20 (5.2%)	9 (2.3%)	5 (1.3%)	6 (1.6%)	1	0	3	2
Total			10 806	427 (4%)	173 (1.6%)	117 (1.1%)	137 (1.3%)	74	12	37	11

R-factors of the experiments are in the 0.11–0.2 range, average bond length errors are in the 0.007–0.02 Å range. RES, resolution; nres, number of residues; nhb, number of hydrogen bonds in the globule; MC, main chain; SC, side chain; s.b., salt bridge.
[a]SER111A–SER435A OG-OG.
[b]CYS247A–TYR282A SG-OH.
[c]TYR64A–CYS286A OH-SG.

(as the hydroxyl in tyrosine is conjugated with the aromatic ring, thus the rotation is impeded). The energy function used for the analysis was the electrostatic term (Coulomb's law) complemented by hard-sphere repulsion. This potential includes two fundamental types of interactions and is successfully used in molecular mechanics calculations (Weber and Harrison, 1999).

Step 2. Geometric rules 2 and 3 (above) were then used to compile a list of potential hydrogen bonds (nhb in Table I).

Step 3. For each potential bond in this list, the angles A′–A–D and A′–A–H (used in rules 4 and 5 above) were measured. If, for a potential bond, either of the angles was <90° (as in rules 4 and 5), the bond was moved to the 'non-4,5' (Table I) list.

Step 4. The bonds in these 'non-4,5' lists were further classified into main chain–main chain, side chain–main chain, side chain–side chain types (columns MC–MC, SC–MC and SC–SC in Table I).

Step 5. The side chain–side chain interactions of the non-4,5 lists were then classified into 'salt bridges' and other chemical types of the pairs of interacting atoms (s.b., NH2..OOC, OH..OOC and NH..OH in Table I).

Structures of the proteins obtained by X-ray crystallography were taken from the PDB (Berman *et al.*, 2000). Hydrogen bond analysis was performed using a program by I.Torshin; the excluded interactions were analyzed using the bond lists produced by the program and also in Rasmol (Sayle and Milner-White, 1995).

Results and discussion

In order to assess the plausibility of using the geometrical rules 4 and 5 for selection of hydrogen bonds, it was necessary to conduct measurement of the angles A′–A–D and A′–A–H in a number of solved structures. Such calculations were conducted using structures from PDB (Berman *et al.*, 2000). Approximately 30 structures, selected on the base of resolution, *R*-factor, bond length error, chain size and completeness of the atomic coordinates were analyzed, bonds formed by hetero atoms were not considered. The results are presented in the Table I.

The interactions with the angles A′–A–H and A′–A–D <90° comprise 2–8% of the whole set of hydrogen-bonding interactions for a protein (Table I). Approximately one-third of them are main chain–main chain interactions between O and N atoms of (i) and (i + 2) residues in a chain (MC–MC Table I). This kind of interaction corresponds to 'strained' backbone conformations, therefore, rules 4 and 5 are particularly useful to exclude them. Another one-third of the non-4,5 interactions are side chain–main chain interactions

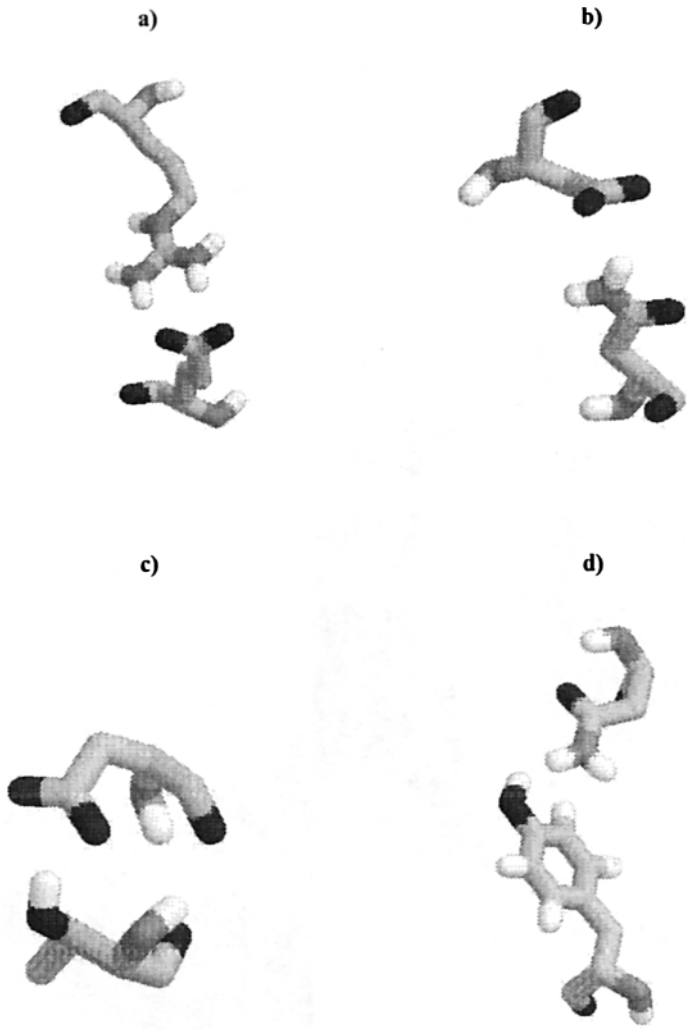

Fig. 2. Examples of the excluded side chain–side chain interactions in the proteins of the set. The examples were taken from oligo-peptide binding protein (PDB 1b6h_a). In the whole set, interactions of the 'salt bridge' (a) and OH..OOC types (c) comprise >80% of all the excluded side chain–side chain interactions. (**a**) 'Salt bridge' type (Arg237 NH2, Glu259 OE1); (**b**) NH2..OOC type (Asn487 ND2, Asp505 OD1); (**c**) OH..OOC type (Asp140 OD2, Thr143 OG1); (**d**) NH..OH type (Asn246 ND2, Tyr269 OH). The Figure was prepared using Rasmol (Sayle and Milner-White, 1995).

between (i) and (i + 1) residues (MC–SC, Table I). These bonds are not analyzed here, however, the related backbone conformations are regular. Most of the side chain–main chain bonds are formed by residues placed on the borders of the β-strands and/or helices. The last one-third of the excepted interactions are between side chain atoms (SC–SC, Table I). Although side chain–side chain interactions are rather rare (~1% of all hydrogen bonds), it was found that they fall mostly into four chemically distinct types (Table I and Figure 2): roughly, half of the bonds are of the 'salt bridge' ([RK]/[DE]) type (Figure 2a), one-third are OH..OOC ([STY]/[DE] type, Figure 2c) and one-sixth are of NH2..OOC ([QN]/[DE], Figure 2b) and NH..OH ([RKQN]/[STY], Figure 2d) types.

When both excluded (non-4,5, Table I) and non-excluded interactions of the same residue pairs in each structure were analyzed visually and using the lists of the bonds, ~90% of the OH..OOC interactions were found to be 'bifurcated' (that is, with hydrogen bonding being possible for both carboxyl atoms). Approximately 90% of the 'bifurcated' bonds (thus ~81% of all OH..OO analyzed) were of [ST]/[DE] type. The term 'bifurcated hydrogen bonds' implies that hydrogen of the 'rotating' hydroxyl of Ser or Thr may interact with the two oxygen atoms of a carboxyl group. In a crystal structure, which is an average over all molecules in the lattice, this also may correspond to ~50% of molecules existing in one configuration and the other 50% in the other configuration. However, as the positions of the 'rotating' hydrogen atoms were calculated using the energy minima, the rotation rather corresponds to a delocalization of the positively charged hydrogen between the two oxygen atoms. This conclusion is based on the physico-chemical principles of the hydrogen bonding (modeled using the electrostatic and the repulsion terms, as above) and not only on some geometric criteria.

In addition, the spatial structures of the [ST]/[DE] fragments (bulk of the excluded side chain–side chain interactions), were analyzed using an atlas of protein side chain–side chain interactions (Singh and Thornton, 1992). This atlas represents all 400 possible pairwise interactions between the 20 side chains and is based on 533 protein chains with sequence identity no greater than 20% and includes only proteins solved by X-ray crystallography to 2.0 Å resolution or better. Spatial configurations of the most of the excluded 'bifurcated' hydrogen bonds of all side chain–side chain pairs (Ser–Asp, Ser–Glu, Thr–Asp and Thr–Glu) were found to belong to the most statistically significant clusters [cluster numbers 1 and/or 2, according to the atlas (Singh and Thornton, 1992)]. This confirms the conclusion that application of the restrictions on the angles A'–A–H and A'–A–D (rules 4 and 5 above), along with exclusion of unfavorable main chain–main chain hydrogen-bonding interactions (~30%), also excludes from the list of potential hydrogen bonds ~70% of valid side chain interactions (particularly some of the 'salt bridges' and many [ST]/[DE] fragments, termed here as 'bifurcated hydrogen bonds').

Recently, the higher stability and special functional value in protein structures of 'salt bridges' has been questioned (Sindelar *et al.*, 1998), and the lack of apparent stability benefit for many salt bridges requires an alternative explanation for their occurrences in protein structures (Sindelar *et al.*, 1998). Their identification as a kind of 'bifurcated' or a 'double' hydrogen bond would allow one to tackle the problem from a chemical viewpoint. At the same time, the stability of the [ST]/[DE] fragments may be even higher than that of salt bridges, particularly due to delocalization of the partial positive charge of 'rotating' hydrogen between the oxygen atoms of a carboxyl. As the van der Waals radius of oxygen is ~1.4 Å, in most cases the destabilizing interaction due to overlap of the van der Waals atomic spheres would be less than in the case of salt bridges. In this protein data set, ~20% of all excluded 'bifurcated' bonds are involved in the formation of three to four residue turns, whereas the bulk of them were formed between sequentially distant residues. This suggests these bonds are important for stabilizing the tertiary structure.

These 'bifurcated' hydrogen bonds between acid and hydroxyl residues and salt bridges (considered as 'pairs' of hydrogen bonds) may be important for protein structure. Therefore, usage of some physico-chemical rules, in addition to the valid geometric criteria (H–A <2.5 Å; D–H–A> 90.0°), is likely to be more accurate than the purely geometric rules 4 and 5, which merely exclude such interactions. These physico-chemical rules are: (1) analysis of the energy minima for 'rotating' hydrogen atoms; (2) direct exclusion of the

'strained' main chain–main chain interactions between the O atom of *each* (*i*) residue and the N atom of *each* (*i* + 2) residue (instead of restricting values of the A′–A–H and A′–A–D angles for *any* donor–acceptor pair).

References

Baker,E.N. and Hubbard,R.E. (1984) *Prog. Biophys. Mol. Biol.*, **44**, 97–179.
Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) *Nucleic Acids Res.*, **28**, 235–242.
Huggins,M.L. (1971) *Angew. Chem.*, **10**, 147–155.
Ippolito,J.A., Alexander,R.S. and Christianson,D.W. (1990) *J. Mol. Biol.*, **215**, 457–471.
Latimer,W.M. and Rodebush,W.H. (1920) *J. Am. Chem. Soc*., **42**, 1419–1423.
McDonald,I.K. and Thornton,J.M. (1994) *J. Mol. Biol.*, **238**, 777–793.
Sayle,R.A. and Milner-White,E.J. (1995) *Trends Biochem. Sci.*, **2**, 374–378.
Singh,J. and Thornton,J.M. (1992) *Atlas of Protein Side-Chain Interactions*, Vols I and II. IRL Press, Oxford.
Sindelar,C.V., Hendsch,Z.S. and Tidor,B. (1998) *Protein Sci.*, **7**, 1898–1914.
Stickle,D.F., Presta,L.G., Dill,K.A. and Rose,G.D. (1992) *J. Mol. Biol.*, **226**, 1143–1159.
Weber,I.T. and Harrison,R.W. (1999) *Protein Eng.,* **12**, 469–474.

Received May 22, 2001; revised September 24, 2001; accepted February 8, 2002