Acta Cryst. (1997). D53, 316–320

C—H···O Hydrogen Bonds in β-sheets

G. FELCY FABIOLA,[a] S. KRISHNASWAMY,[b] V. NAGARAJAN[a] AND VASANTHA PATTABHI[a]*

[a]*Department of Crystallography and Biophysics, University of Madras, Madras 600 025, India, and* [b]*Bioinformatics Centre, Madurai Kamaraj University, Madurai 625 021, India. E-mail: crystal@giasmd01.vsnl.net.in*

(Received 25 July 1996; accepted 10 January 1997)

Abstract

A detailed analysis of the occurrence of the C—H···O hydrogen bonds in sheet regions of proteins has been presented. 11 unique protein structures with resolution 1.3 Å containing β-sheets show a widespread presence of C—H···O hydrogen bonds. These have average $C^{\alpha}\cdots O$, $CH\cdots O$ distances and a $C^{\alpha}—H\cdots O$ angle of 3.29, 2.38 Å and 143°, respectively. As in the case of N—H···O hydrogen bonds, parallel and antiparallel β-sheet regions show the same hydrogen-bond geometry. An inverse correlation is observed between the hydrogen-bond geometries involving the C^{α}_{i}—H···O=C and the N_{i+1}—H···O=C suggesting that C—H···O hydrogen bonds may act as an additional stabilizing factor. The propensity of different amino-acid residues to form such hydrogen bonds varies and shows a clear preference for valine and threonine. C—H···O hydrogen bonds involving side chains also occur extensively in β-sheet regions.

1. Introduction

C—H···O hydrogen bonds have now been widely accepted to contribute to the stability of interactions in small-molecule structures (Desiraju, 1991; Jeffrey & Saenger, 1991; Steiner & Saenger, 1993; Taylor & Kennard, 1982). A recent report based on small-molecule structures has indicated that water molecules, when not involved in other hydrogen bonds or coordination, show systematic involvement in C—H···O hydrogen bonds. The possibility of C—H···O hydrogen bonds contributing to the stability of nucleic acid base pairs and providing secondary hydrogen bonds in base pairs has been pointed out by Leonard, McAuley-Hecht, Brown, & Hunter (1995). However, even a recent analysis of stability of β-sheets in proteins has primarily considered only N—H···O main-chain hydrogen bonds in addition to side-chain interactions (Yang & Honig, 1995).

The question as to whether C—H···O hydrogen bonds are found in protein structures has been addressed recently (Nagarajan & Pattabhi, 1995; Derewenda, Lee & Derewenda, 1995). Our studies on heterochiral peptides have also shown that C—H···O hydrogen bonds can be formed in addition to N—H···O hydrogen bonds in β-sheets (Nagarajan, Pattabhi, Johnson, Durani & Bobde, 1997). In this context, the high-resolution structures (<1.3 Å) in the Protein Data Bank (Bernstein *et al.*, 1977), including those with hydrogen coordinates, have been analysed for C—H···O interactions that satisfy the C—H···O hydrogen-bonding criteria established in the studies referred to earlier. Derewenda *et al.* have, in general, dealt with C—H···O interactions in protein structures representing different categories of secondary and tertiary folds. Here, we focus on β-structures using only data from very high resolution structures and provide additional insights and information on the role of C—H···O hydrogen bonds in stabilizing them. The occurrence of such hydrogen bonds involving side chains in β-sheet regions and the propensity of different amino-acid residues to take part in C—H···O hydrogen bonds have also been explored.

2. Materials and methods

The protein structures with resolutions better than 1.3 Å were selected from the April 1995 release of the PDB. 11 of them had β-sheet regions and were non-identical structures. The relevant details on these structures are listed in Table 1. Three of the structures (1CBN, 5PTI, 7RSA) had H atoms included in the structure refinement. For 1CUS, N—H H atoms alone have been included in the PDB entry. Hence, C^{α} H atoms of 1CUS and all H atoms for the remaining structures were fixed at geometrically expected positions using the *BIOSYM* software, *Biopolymer* module (Biosym Technologies, 1993). In all instances, there was no disorder or problems associated with the residues identified as having C—H···O interactions. Based on the C—H···O hydrogen-bonding parameters suggested in the literature (Desiraju, 1991; Taylor & Kennard, 1982; Steiner, 1995), a stringent set of criteria namely, C···O <3.5 Å and C—H···O angle >130°, were used for selection of possible hydrogen bonds. The criteria suggested by Baker & Hubbard (1984) were used for delineating N—H···O hydrogen bonds.

3. Results and discussion

Statistical information pertaining to the C^{α}—H···O and N—H···O hydrogen bonds in the β-sheet regions in the 11 structures are given in Table 2 and Figs. 1 and 2.

Table 1. *Protein set used in C—H···O hydrogen bond calculations*

PDB code	Name	Resolution (Å)	*R*	Method of refinement	Remarks*	Reference
1CBN	Crambin	0.83	0.11	RLS	HPDB	1
5PTI	Trypsin inhibitor	1.0	0.20	RLS	HPDB	2
7RSA	RNase A	1.26	0.15	RLS	HPDB	3
2SN3	Scorpion neurotoxin	1.2	0.19	RLS	HF	4
8RXN	Rubredoxin	1.0	0.15	RLS	HF	5
1IGD	Protein G	1.1	0.19	RLS	HF	6
1IFC	Intestinal fatty acid binding protein	1.19	0.17	*X-PLOR*	HF	7
1ARB	*Anchromobacter* protease	1.2	0.15	RLS	HF	8
1YCC	Cytochrome *c*	1.23	0.19	RLS	HF	9
1CUS	Cutinase	1.25	0.16	*X-PLOR*	NHPDB HAF	10
1CSE	Subtilisin Carlsberg	1.2	0.18	EREF	HF	11

* HPDB, H atoms from PDB. HF, H atoms fixed. (Using *BIOSYM* software *Biopolymer* module. Biosym Technologies, 1993.) NHPDB, NH H atoms from PDB. HAF, C^{α} H atoms fixed. RLS, restrained least squares (Hendrickson & Konnert, 1980). EREF, energy refinement (Jack & Levitt, 1978). *X-PLOR*, (Brünger, 1992).

References: (1) Teeter, Roe & Heo (1993); (2) Wlodawer, Walter, Huber & Sjölin (1984); (3) Wlodawer, Svensson, Sjölin & Gilliland (1988); (4) Zhao, Carson, Ealick & Bugg (1992); (5) Dauter, Sieker & Wilson (1992); (6) Derrick & Wigley (1994); (7) Scapin, Gordon & Sacchettini (1992); (8) Tsunasawa, Masaki, Hirose, Soejima & Sakiyama (1989); (9) Louie & Brayer (1990); (10) Martinez, De Geus, Lauwereys, Matthyssens & Cambillau (1992); (11) Bode, Papamokos & Musil (1987).

Table 2. *Average values of the $C^{\alpha}\cdots O$, $C^{\alpha}H\cdots O$, $N\cdots O$, $NH\cdots O$ distances and the $C^{\alpha}—H\cdots O$, $N—H\cdots O$ angles in the data set*

N, number of structures. *n*, number of C—H···O as well as N—H···O hydrogen bonds.

			C—			N—	
		C···O	H···O	CH···O	N···O	H···O	NH···O
N	*n*	(Å)	(°)	(Å)	(Å)	(°)	(Å)
Antiparallel sheet							
9	49	3.27	141	2.37	2.89	158	1.94
Parallel sheet							
3	27	3.31	146	2.36	2.95	162	2.00
Summary							
Minimum		2.91	130	2.01	2.73	137	1.70
Maximum		3.50	160	2.82	3.23	175	2.84
Average		3.29	143	2.38	2.93	160	1.97
E.s.d.		0.11	6	0.13	0.10	9	0.17

The average hydrogen-bond parameters remain nearly the same irrespective of whether the sheets are parallel or antiparallel.

The carbonyl O atoms are the acceptors and the C—H proton preferentially approaches the acceptor oxygen along one of its lone pair orbitals. In all cases the C^{α}—H of the *i*th residue of one strand and the N—H of the $(i+1)$th residue of the same strand are hydrogen bonded to the same carbonyl O atom of the opposite strand. This is true of antiparallel, parallel and mixed sheets (see, for example, Fig. 3). There is no unusual deviation in these regions from the normal β-strand conformations observed in proteins. Similarly there is no preference for any specific regions of sheet for the formation of C—H···O hydrogen bonds. The interstrand distance calculated between the carbonyl O atom to the centre of the peptide group is a constant with respect to φ as well as ψ (2.89 ± 0.05 Å) indicating that φ and ψ are correlated in such a way that the interstrand distance is maintained in β-sheets.

A cross distribution of the C—H···O and the N—H···O hydrogen-bond parameters (Fig. 4) reveals that there is an approximate inverse relation between them, especially the C···O *versus* N···O and CH···O *versus* NH···O distances. In general, for example, if the NH···O distance is large there is a tendency for the CH···O distance to be small. There are exceptions, probably dictated by the particular environment of interactions. Thus, there appears to be thus a compensatory mechanism to provide stability to the β-sheets.

In addition to the main-chain C^{α}—H···O hydrogen bonds, there is evidence for hydrogen bonds involving side-chain C atoms and main-chain carbonyl O atoms in β-sheet regions. This observation and the inverse correlation between C—H···O and N—H···O geometry clearly suggests that C^{α}—H···O interactions observed in β-sheets do not result from mere steric disposition. Fig. 5 shows the distribution in C—H···O hydrogen bonds involving side chains. Unlike the main-chain C—H···O interactions, C—H···O distances and C—H···O angles peak at 2.55 Å and 130°, respectively. Hence, these interactions are, in general, much weaker than the C^{α}—H···O interactions. All the same they are likely to contribute to the structural stability.

4. Amino-acid preferences

An analysis of the amino-acid preference for the formation of C—H···O hydrogen bonds in β-sheet regions was carried out. The calculations were performed for the entire data set without making any distinction between parallel and antiparallel β-sheets. The number of hydrogen bonds involving each amino acid was normalized with respect to the total number present in the protein as,

$$P_{ha} = \frac{h_a}{H_{aa}} \quad \frac{N_a}{N_{aa}},$$

where h_a is the number of C—H···O hydrogen bonds formed by a particular amino acid. H_{aa} is the total number of C—H···O hydrogen bonds in the data set. N_a is the number of a particular type of amino acid present in the data set. N_{aa} is the total number of amino acids in the data set.

The values of $P_{ha} \times 100$, given in Fig. 6, were evaluated separately for C^{α}—H···O hydrogen bonds and those involving side chains. In Fig 6(*a*), which shows the population distribution of various amino acids with respect to C^{α}—H···O hydrogen bonds, one can see a clear

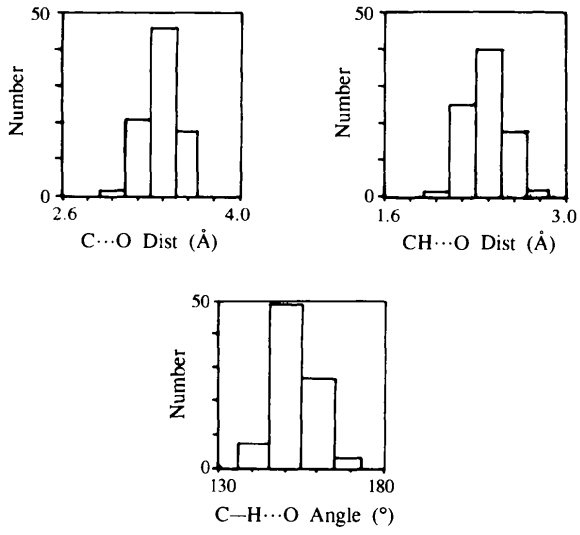

Fig. 1. The distribution of C^{α}—H···O hydrogen-bond parameters observed in β-sheet regions in the protein structures listed in Table 1.

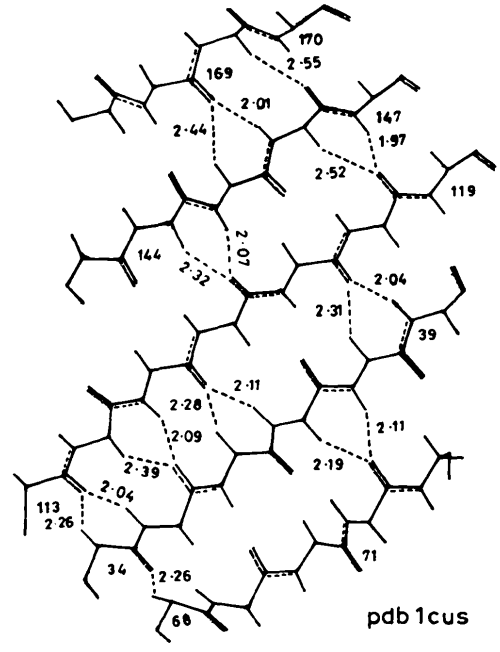

Fig. 3. The β-sheet in cutinase. N—H···O and C^{α}—H···O hydrogen bonds are indicated.

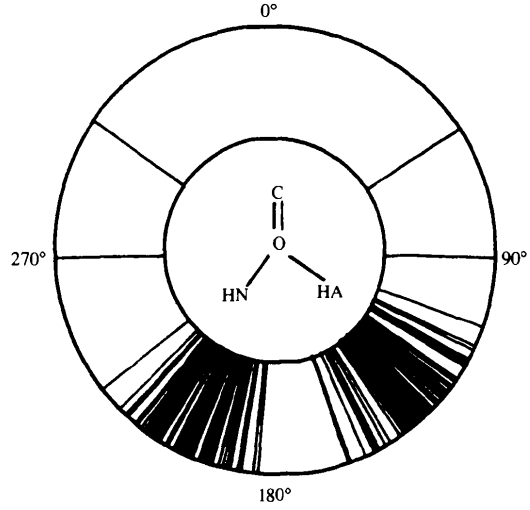

Fig. 2. The distribution of the angles C=O···H—C and C=O···H—N in the respective hydrogen bonds.

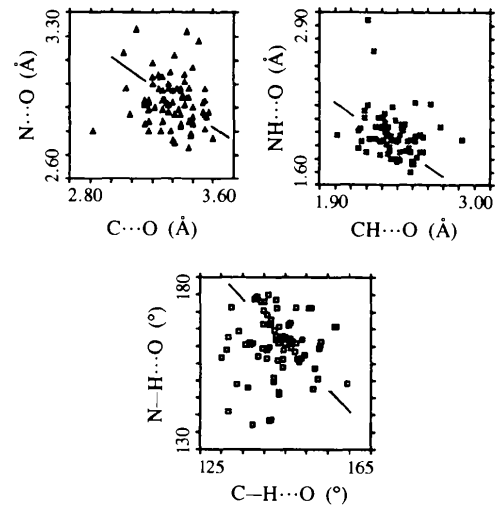

Fig. 4. A cross distribution of observed C—H···O and N—H···O parameters.

preference for threonine and valine. Though glycine occurs very frequently in sheets, surprisingly C^{α} atoms of glycine do not take part in C—H···O interactions at all. This may be because of the high flexibility of glycyl residues which may cause the C^{α} protons to deviate from the hydrogen-bonding positions. Fig. 6(*b*) gives the residue-wise distribution of C—H···O hydrogen bonds involving side-chain atoms and shows a preference for valine and tyrosine.

5. Conclusions

The present analysis clearly shows that C—H···O hydrogen bonds can offer additional stability to sheet regions and act as a compensating factor if there is a weak N—H···O interaction. In this context, the observation of Wishart, Sykes & Richards (1991) on NMR chemical shifts is of interest. They have shown that there is a linear correlation between hydrogen-bond strength and amide proton chemical shifts. Their analysis shows that for helical regions the difference in chemical shift *viz* $\Delta\delta$ ($= \delta_{obs} - \delta_{rc}$, δ_{rc} being the chemical shift for random coil) is negative and has the lowest hydrogen-bond energy, whereas for β-sheet regions $\Delta\delta$ is positive and has higher hydrogen-bond energy. The NMR database on some of the proteins used in this analysis (NMR Database, University of Wisconsin, Madison) show downfield shift for C^{α} protons in the β-sheet region ($\Delta\delta$ is positive) and upfield shift in the helix region ($\Delta\delta$ is negative). This, along with the observation of Wishart *et al.*, appears to indicate that C^{α}—H protons form stronger hydrogen bonds in β-sheet regions than in the α-helical regions. This inference agrees with that of Derewenda *et al.* (1995) that 'short C—H···O contacts in helices appear to be rather esoteric in character and relatively few in number'. The propensity for C—H···O interactions in β-sheets is such that the contribution from these hydrogen bonds to structural stability is more substantial in sheets than in any other region of

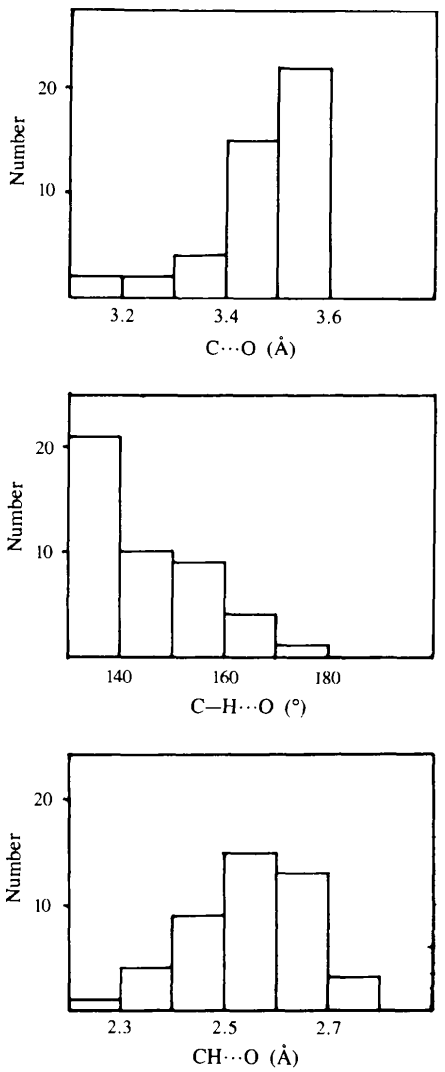

Fig. 5. Distribution of the parameters of the C—H···O hydrogen bonds involving side chains in β-sheets.

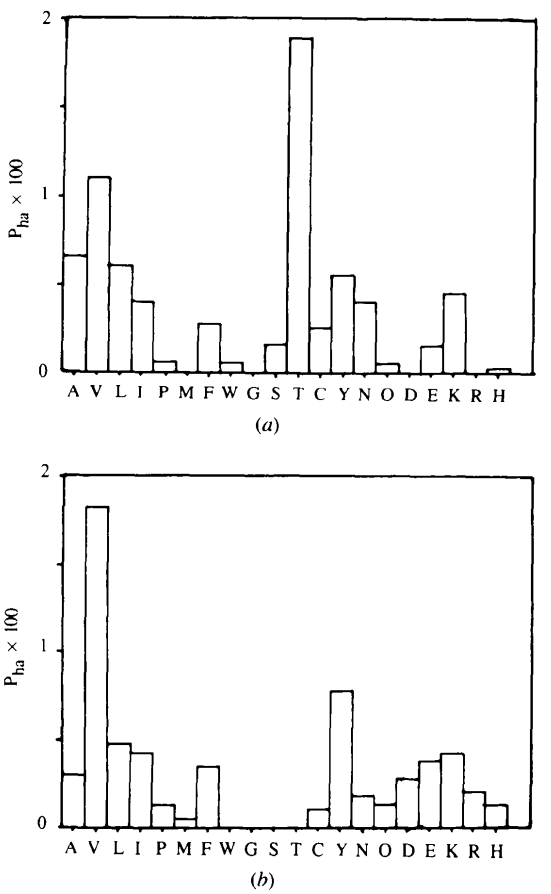

Fig. 6. The observed amino-acid preference of (*a*) C^{α} and (*b*) side-chain C atoms for forming C—H···O hydrogen bonds (see text for details).

proteins. Thus, the present study emphasizes the need for taking into account C—H···O interactions in proteins especially in the modelling of β-sheet regions.

The use of the High Resolution Graphics facility and the Bioinformatics Center facilities at School of Biotechnology, Madurai Kamaraj University is gratefully acknowledged. We thank Dr G. G. Dodson, M. Sundaralingam and T. C. Pochapsky for useful discussions. Our thanks are due to the NMR group at University of Wisconsin, Madison, for providing the NMR data from their database and the referees for very useful suggestions.

References

Baker, E. K. & Hubbard, R. E. (1984). *Progr. Biophys. Mol. Biol.* **44**, 97–179.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rogers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Biosym Technologies (1993). *Insight*II, version 2.3.0, Biosym Technologies, San Diego, CA, USA.

Bode, W., Papamokos, E. & Musil, D. (1987). *Eur. J. Biochem.* **166**, 673–692.

Brünger, A. T. (1992). *X-PLOR: A system for X-ray crystallography and NMR*, Yale University Press, New Haven, CT, USA.

Dauter, Z., Sieker, L. & Wilson, K. S. (1992). *Acta Cryst.* B**48**, 42–59.

Derewenda, Z. S., Lee, L. & Derewenda, V. (1995). *J. Mol. Biol.* **252**, 248–262.

Derrick, J. P. & Wigley, D. B. (1994). *J. Mol. Biol.* **243**, 906–918.

Desiraju, G. R. (1991). *Acc. Chem. Res.* **24**, 290–296.

Hendrickson, W. A. & Konnert, J. H. (1980). *Biomolecular Structure, Function, Conformation and Evolution*, edited by R. Srinivasan, pp. 43–57. Oxford: Pergamon Press.

Jack, A. & Levitt, M. (1978). *Acta Cryst.* A**34**, 931–934.

Jeffrey, G. A. & Saenger, W. (1991). *Hydrogen Bonding in Macromolecular Structures*. Berlin: Springer-Verlag.

Leonard, G. A., McAuley-Hecht, K, Brown, T. & Hunter, W. N. (1995). *Acta Cryst.* D**51**, 136–139.

Louie, G. V. & Brayer, G. D. (1990). *J. Mol. Biol.* **214**, 527–555.

Martinez, C., De Geus, P., Lauwereys, M., Matthyssens, G. & Cambillau, C. (1992). *Nature (London)*, **356**, 615–618.

Nagarajan, V. & Pattabhi, V. (1995). International seminar cum school on Macromolecular Crystallographic data, Calcutta, India, Abstract I40.

Nagarajan, V., Pattabhi, V., Johnson, A., Durani, S. & Bobde, V. (1997). *Int. J. Peptide Protein Res.* In the press.

Scapin, G., Gordon, J. I. & Sacchettini, J. C. (1992). *J. Biol. Chem.* **267**, 4253–4269.

Steiner, Th. (1995). *Acta Cryst.* D**51**, 93–97.

Steiner, Th. & Saenger, W. (1993). *J. Am. Chem. Soc.* **115**, 4540–4547.

Taylor, R. & Kennard, O. (1982). *J. Am. Chem. Soc.* **104**, 5063–5070.

Teeter, M. M., Roe, S. M. & Heo, N. H. (1993). *J. Mol. Biol.* **230**, 292–311.

Tsunasawa, S., Masaki, T., Hirose, M., Soejima. M. & Sakiyama, F. (1989). *J. Biol. Chem.* **264**, 3832–3839.

Wishart, D. S., Sykes, B. D. & Richards, F. M. (1991). *J. Mol. Biol.* **222**, 311–333.

Wlodawer, A., Svensson, L. A., Sjölin, L. & Gilliland, G. L. (1988). *Biochemistry*, **27**, 2705–2717.

Wlodawer, A., Walter, J., Huber, R. & Sjölin, L. (1984). *J. Mol. Biol.* **180**, 301–329.

Yang, A.-S. & Honig, B. (1995). *J. Mol. Biol.* **252**, 366–376.

Zhao, B., Carson, M., Ealick, S. E. & Bugg, C. E. (1992). *J. Mol. Biol.* **227**, 239–252.