



<https://theses.gla.ac.uk/>

Theses Digitisation:

<https://www.gla.ac.uk/myglasgow/research/enlighten/theses/digitisation/>

This is a digitised version of the original print thesis.

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>  
[research-enlighten@glasgow.ac.uk](mailto:research-enlighten@glasgow.ac.uk)

**The Geometry of Hydrogen-Bonds and Carbonyl-Carbonyl Interactions  
Between *Trans*-Amides in Proteins and Small Molecules**

**William John Duddy**

**Submitted for the degree of  
Doctor of Philosophy**

**Division of Biochemistry and Molecular Biology  
Faculty of Biomedical and Life Sciences**



**UNIVERSITY  
of  
GLASGOW**

**April 2005**

ProQuest Number: 10391071

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10391071

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 – 1346

GLASGOW  
UNIVERSITY  
LIBRARY:

## Abstract

The geometries, at the level of atoms or groups of atoms, of interactions and motifs found in crystal structures in the Protein Data Bank (PDB) and the Cambridge Structural Database (CSD) are analysed. The bulk of the thesis examines electrostatic interactions that occur between *trans* secondary amide groups, or peptide groups as they are known in proteins. The two types considered are N-H...O=C hydrogen-bonds and carbonyl-carbonyl interactions. Additionally, a chapter investigates specific hydrogen-bonded motifs, known as *asx*- and *ST*-turns, that commonly occur within protein structures.

The geometry of hydrogen-bonds between *trans* secondary amide groups (i.e. peptide groups) has been studied extensively in proteins, and to a lesser extent in the CSD. They are a subset of the general case of N-H...O=C hydrogen-bonds. Previous analyses of the CSD have shown a tendency for N-H...O=C hydrogen-bonds to exhibit lone-pair directionality, where the hydrogen atom is near to the plane of the lone-pairs of the carbonyl oxygen atom, and the H...O=C angle approaches 120°. For *trans* secondary amides the in-plane preference is also observed, but the H...O=C angle is greater, averaging about 150°. Here, an examination of the CSD allows elucidation of four factors that together account for this difference in H...O=C: 1. A smaller proportion of *trans* secondary amide carbonyl oxygens accept more than one hydrogen-bond than do carbonyl oxygens in general. 2. N-H...O=C bonds often occur in 'ring' motifs with relatively constrained geometries and H...O=C values near 120°. These cannot be formed by *trans* secondary amides. 3. Chains of hydrogen-bonds between *trans* secondary amides, with large H...O=C values, often extend throughout the crystal lattice. 4. The steric accessibility of *trans* secondary amide carbonyl oxygens is less than for carbonyl oxygens in general.

It has been suggested that electrostatic interactions between carbonyl groups affect hydrogen-bond geometry in  $\alpha$ -helix and  $\beta$ -sheet, influence the twist of  $\beta$ -sheet, encourage polarization of carbonyl groups in  $\alpha$ -helix, and explain the propensity of asparagine and aspartate residues in unusual regions of the Ramachandran plot. The carbonyl groups of ketones in the CSD frequently interact with each other, and commonly occur in three geometric motifs, that can be described as antiparallel, parallel, or perpendicular. At

optimal geometry the antiparallel motif has energetic favourability approaching that of a medium-strength hydrogen-bond. Here it is shown that, after hydrogen-bonding has been taken into consideration, carbonyl-carbonyl interactions between *trans* secondary amides in the CSD also occur in these three motifs, and with a surprisingly similar propensity: 48% occur in the antiparallel arrangement (cf. 49% of those between ketones).

Furthermore, interactions between main-chain carbonyl groups in a 454-chain subset of the PDB are identified. For each carbonyl-carbonyl interaction, its geometry, local secondary structure, and local hydrogen-bonding, are considered. The three motifs present in the CSD are not found to be representative of the geometries present in the PDB. However, other favourable carbonyl-carbonyl interaction motifs are observed. These occur in a variety of situations with respect to secondary structure and hydrogen-bonding, and are prevalent at the C-termini of  $\alpha$ -helices. They are shown to contribute to the stability of common C-termini capping conformations, one being the Schellman loop, the other being the case where a proline terminates the helix.

The hydrogen-bonded  $\beta$ -turn is a small, well characterised, protein motif defined by a hydrogen-bond between the main-chain carbonyl group of one residue and the main-chain N-H group of another three residues ahead in the polypeptide chain. There are four common types, distinguished by geometry: I, I', II, and II'. In Asx- and ST-turns, the side-chain carbonyl of an asparagine, aspartate, serine, or threonine residue hydrogen-bonds with the main-chain N-H of a residue two ahead in the chain, such that they structurally mimic the  $\beta$ -turn. Asx-turns have previously been categorized into four classes and ST-turns into three categories, based on side-chain rotamer types and the conformation of the central residue of each turn. Here it is shown that the four classes of asx-turn are geometrically equivalent to the four types of hydrogen-bonded  $\beta$ -turn, and that the three categories of ST-turn are geometrically equivalent to three of the four types of hydrogen-bonded  $\beta$ -turn. It is proposed that asx- and ST-turns be named using the type I, II, I' and II'  $\beta$ -turn nomenclature. Using this nomenclature, the frequency of occurrence of both asx- and ST-turns is: type II' > type I > type II > type I', whereas for  $\beta$ -turns it is type I > type II > type I' > type II'. It is found that the type II Asx- or ST-turn is the same as a previously identified hydrogen-bonded motif called the Asx- or ST-nest.

*To the memory of my brother, James.*

## Acknowledgments

*Thanks to:*

*Prof. James Milner-White, my supervisor in Glasgow, for his constant guidance and encouragement throughout. Willem Nissink, my supervisor at the Cambridge Crystallographic Data Centre (CCDC), for his consistent support in all aspects of the work.*

*Frank Allen, executive director of the CCDC, for electing to spend so much valuable time on my project, for critical appraisals of the text, and for providing, at crucial stages, his carefully considered advice on the overall direction of the work. Also, the staff of the CCDC, particularly Sam Motherwell for all of his help with the steric accessibility section. He provided the RPluto data, along with much helpful advice. Robin Taylor for stimulating discussions regarding hydrogen-bond geometry. The BBSRC and CCDC for funding.*

*The protein crystallography group at the University of Glasgow, both for scientific and social support. Particularly: Alastair McEwen, Alan Riboldi-Tunncliffe, Neil Paterson, and Adrian Laphorn.*

*Zoe Gokhale for coding the identification and characterisation program for intraprotein carbonyl-carbonyl interactions. Gilleain Torrance for assistance with other aspects of the intraprotein carbonyl-carbonyl interaction section. Stuart Mackay for IT support and his friendship, but particularly for tolerating and responding to my endless wails of, "my computer's broken!" and, "why, why, why doesn't it work?".*

*My friends, whether back home in Northern Ireland, here in Glasgow, or elsewhere, for providing distraction and escape when most needed.*

*Lastly, and most importantly, my parents and my sister, for always being there when I need them, and for their persistent love and attention, even in my darkest moods.*



# Contents

	Page	
1	Abbreviations, Tables and Figures	1
2	Introduction	
2.1	Outline	5
2.2	General	7
2.3	<b>Hydrogen-bonding between <i>trans</i>-amide groups</b>	
2.3.1	Context	8
2.3.2	Small molecule studies	9
2.3.3	Protein studies	11
2.4	<b>Carbonyl-carbonyl interactions between <i>trans</i>-amide groups</b>	
2.4.1	Context	13
2.4.2	Small molecule studies	13
2.4.3	Protein studies	15
2.5	<b>Mimicry of <math>\beta</math>-turns by Asx- and ST-turns</b>	
2.5.1	Context	17
2.5.2	$\beta$ -turns	17
2.5.3	Asx-turns and ST-turns	18
2.5.4	Nests	20
3	<b>Hydrogen-bonding between <i>trans</i>-amide groups: Factors affecting hydrogen-bond geometry</b>	
3.1	Brief introduction and outline	21
3.2	Materials and methods	
3.2.1	Database searches	22
3.2.2	Statistical correction for angular data	23
3.3	Results & Discussion	
3.3.1	Hydrogen-bonds between <i>trans</i> -amide groups	24
3.3.2	Comparison with other hydrogen-bonded systems	26
3.3.3	Hydrogen-bonded chains	29
3.3.4	Multiple hydrogen-bonds	31
3.3.5	Ring motifs	33
3.3.6	Three-centered hydrogen-bonds	37
3.3.7	Summary of factors affecting H $\cdots$ O=C angle	39
3.3.8	Other Aspects of inter-amide geometry	42
3.3.9	Internal geometry of <i>trans</i> -amide group	48
4	<b>Hydrogen-bonding between <i>trans</i>-amide groups: steric accessibility</b>	
4.1	Brief introduction and outline	50
4.2	Materials and methods	
4.2.1	Identification and analysis of hydrogen-bonds	51
4.2.2	Calculation of steric accessibility	52
4.3	Results & Discussion	53

	Page	
5	<b>Carbonyl-carbonyl interactions between <i>trans</i>-amide groups: in the CSD</b>	
5.1	Brief introduction and outline	63
5.2	Materials and methods	
5.2.1	Database searches	64
5.2.2	Characterisation of carbonyl-carbonyl interactions	64
5.3	Results & Discussion	66
6	<b>Carbonyl-carbonyl interactions between <i>trans</i>-amide groups: in the PDB</b>	
6.1	Brief introduction and outline	75
6.2	Materials and methods	
6.2.1	Data collation and analysis	77
6.2.2	Dataset	77
6.2.3	Carbonyl-carbonyl interaction criteria and descriptors	78
6.2.4	Hydrogen-bond criteria and hydrogen-bonding patterns	79
6.2.5	Local secondary structure	80
6.2.6	Ramachandran plot regions	80
6.2.7	Energy calculations	81
6.2.8	Principal Component Analysis	81
6.3	Results & Discussion	
6.3.1	Overall data	82
6.3.2	Motif identification	85
6.3.3	Motif Description	97
6.3.4	Carbonyl-carbonyl interactions at the helical C-terminus	107
7	<b>Mimicry of <math>\beta</math>-turns by <i>asx</i>- and <i>ST</i>-turns</b>	
7.1	Brief introduction and outline	109
7.2	Materials and methods	110
7.3	Results & Discussion	
7.3.1	<i>Asx</i> -turn and <i>ST</i> -turn identification	112
7.3.2	Comparison with $\beta$ -turn geometry	113
8	<b>Conclusions</b>	
8.1	<b>Hydrogen-bonding between <i>trans</i>-amide groups</b>	
8.1.1	Factors affecting hydrogen bond geometry	118
8.1.2	An additional factor: steric accessibility	121
8.2	<b>Carbonyl-carbonyl interactions between <i>trans</i>-amide groups</b>	
8.2.1	<i>Trans</i> -amide carbonyl-carbonyl interactions in the CSD	123
8.2.2	<i>Trans</i> -amide carbonyl-carbonyl interactions in the PDB	125
8.3	<b>Mimicry of <math>\beta</math>-turns by <i>asx</i>- and <i>ST</i>-turns</b>	127
8.4	General conclusions	129
9	References	130

# 1 Abbreviations, Tables and Figures

## Abbreviations

$\alpha_L$  = Region of the Ramachandran plot where residues of left-handed  $\alpha$ -helix are found

(defined in chapter 6 as  $10 < \phi < 125$ ,  $-45 < \psi < 90$ ).

$\alpha_R$  = Region of the Ramachandran plot where residues of right-handed  $\alpha$ -helix are found

(defined in chapter 6 as  $-180 < \phi < 0$ , and  $-135 < \psi < 50$ ).

$\beta$  = Region of the Ramachandran plot where residues of  $\beta$ -sheets are found (defined in

chapter 6 as  $-180 < \phi < 0$ ,  $50 < \psi < 180$  or  $-180 < \psi < -135$ ).

$\gamma_L$  = Conformation describing a residue with  $\phi$ ,  $\psi$  values of approximately  $90^\circ$ ,  $0^\circ$ .

Asx = Asparagine or Aspartate residue.

BLAST = Basic Local Alignment Search Tools.

CSD = Cambridge Structural Database.

DSSP = Dictionary of Secondary Structure in Proteins.

IMPT = Intermolecular Perturbation Theory.

PCA = Principal Component Analysis.

PDB = Protein Data Bank.

PDB file = Coordinate file downloaded from the Protein Data Bank.

PDB id = Protein Data Bank identification tag.

SA = Steric accessibility.

sd = standard deviation.

ST = Serine or Threonine residue.

*Trans*-amide = *Trans* secondary amide.

## List of Tables

	Page	
3.1	Geometry of <i>trans</i> -amide... <i>trans</i> -amide hydrogen-bonds and hydrogen-bonds from other structural subsets.	27
3.2	Occurrence of <i>trans</i> -amide... <i>trans</i> -amide hydrogen-bonds within and outwith chains.	30
3.3	Acceptance of multiple hydrogen-bonds at oxygen atom.	32
3.4	The effect of ring motifs on hydrogen bond geometry.	34
3.5	Three-centered hydrogen-bonds.	38
3.6	Internal geometry of the <i>trans</i> amide group.	49
4.1	Average H...O=C angles and carbonyl oxygen atom steric accessibility values (SA) of N-II...O=C and <i>trans</i> -amide... <i>trans</i> -amide hydrogen-bonds.	55
5.1	Numbers and average angles of antiparallel, parallel, and perpendicular interactions between <i>trans</i> -amide carbonyl groups.	74
6.1	Constants ( <i>A</i> and <i>C</i> ) and partial charges ( <i>q</i> ) used in 9-6-1 Lennard-Jones potential.	81
6.2	Motif definition.	86
6.3	Geometry of sequentially proximate motifs.	99
6.4	Local secondary structure of sequentially proximate motifs.	100
6.5	Geometry of sequentially distant motifs.	104
6.6	Local secondary structure of sequentially distant motifs.	105
6.7	Ratio of C <sup>3</sup> ...O <sup>2</sup> distance to O <sup>2</sup> ...O <sup>4</sup> distance for motifs that are prevalent at helix C-termini.	108
7.1	Proportions of residues involved in $\beta$ -turn mimics.	112
7.2	Comparing torsion angles between $\beta$ -turns and their mimics.	115

## List of Figures

	Page	
2.1	Two types of electrostatic interaction between <i>trans</i> amide groups.	7
2.2	Hydrogen-bonded ring motif.	10
2.3	Defining residues and torsion angles in $\beta$ -turns and in <i>asx</i> - and <i>ST</i> -turns.	19
3.1	Geometric parameters of <i>trans</i> -amide... <i>trans</i> -amide hydrogen-bonds.	25
3.2	Distributions of H...O=C angle of N-H...O=C hydrogen-bonds.	36
3.3	H...O=C and N-H...O angles of various hydrogen bond subsets.	41
3.4	Parameters describing the orientation of the <i>trans</i> -amide... <i>trans</i> -amide hydrogen-bonded system.	42
3.5	Geometry of <i>trans</i> -amide... <i>trans</i> -amide hydrogen-bonded system.	43
3.6	Examples of chains of <i>trans</i> -amide... <i>trans</i> -amides with different orientations.	46
3.7	Distribution of C...O distance for <i>trans</i> -amide... <i>trans</i> -amide hydrogen-bonded systems.	47
3.8	Internal geometry of the <i>trans</i> -amide group.	48
3.9	Plot of C <sub>5</sub> -N-C <sub>6</sub> angle against N...C <sub>6</sub> distance of <i>trans</i> -amide groups.	49
4.1	Histograms of steric accessibility of carbonyl oxygen atoms of N-H...O=C hydrogen-bonds and of <i>trans</i> -amide... <i>trans</i> -amide hydrogen-bonds.	56
4.2	Histogram of average H...O=C angles of N-H...O=C and <i>trans</i> -amide... <i>trans</i> -amide hydrogen-bonds and different subsets of each.	59
4.3	H...O=C angle against steric accessibility (SA) of carbonyl oxygens of N-H...O=C hydrogen-bonds and <i>trans</i> -amide... <i>trans</i> -amide hydrogen-bonds.	62
5.1	Carbonyl...carbonyl interaction geometry.	65
5.2	Geometry of interactions between <i>trans</i> -amide carbonyl groups.	67
5.3	Geometry of antiparallel carbonyl...carbonyl interactions.	69
5.4	Examples of antiparallel and parallel interactions between <i>trans</i> -amide carbonyl groups.	70
5.5	Geometry of parallel carbonyl...carbonyl interactions.	71
5.6	Geometry of perpendicular carbonyl...carbonyl interactions.	73
5.7	Comparison of antiparallel, parallel, and perpendicular interactions of ketone carbonyl groups with those of <i>trans</i> -amide carbonyl groups.	74

	Page
6.1	Angle and torsion angle parameters of carbonyl...carbonyl interactions. 79
6.2	Regions defined in ramachandran plot annotation. 80
6.3	Overall energy distributions and geometry. 84
6.4	Definition of motifs 1 and 2. 87
6.5	Definition of motifs 3, 4, and 5. 89
6.6	Definition of motifs 6 and 7. 91
6.7	Definition of motifs 8 and 9. 92
6.8	Definition of motifs 10 to 15. 94
6.9	Carbonyl-carbonyl interactions in $\beta$ -sheet. 96
6.10	Examples of sequentially proximate carbonyl-carbonyl interaction motifs. 101
6.11	Examples of sequentially distant carbonyl-carbonyl interaction motifs. 106
7.1	Distributions of $\phi$ , $\psi$ torsion angles of residues $i+1$ and $i+2$ of $\beta$ -turns, compared with equivalent angles of asx- and ST-turns. 114
7.2	The four types of hydrogen-bonded $\beta$ -turns and equivalent asx-turns. 117

## 2 Introduction

### 2.1 Outline

This section describes the structure of the thesis, sections 2.2 to 2.5 introduce the various topics covered.

The central aim of the project is to investigate electrostatic interactions between carbonyl groups, in the context of protein structure. Recent research on small molecule and protein crystal structures indicates that they are more important than previously realised.

A search of the small molecule crystal structure database shows that many carbonyl-carbonyl interactions occur between *trans*-amides that are hydrogen-bonded to each other. Since the primary interest is in carbonyl-carbonyl interactions, an effort is made to separate out the hydrogen-bonded cases. The cases that are not hydrogen-bonded are considered in chapter 5. However, it is still of interest to examine the hydrogen-bonded cases, and particularly whether the carbonyl-carbonyl interaction inherent in the hydrogen-bonded system has a detectable effect on average hydrogen-bond geometry. Evidence relating to the effect of the carbonyl-carbonyl interaction on hydrogen-bond geometry is inconclusive, but the work does lead to new insights into the behaviour of *trans*-amides...*trans*-amide hydrogen-bonds compared with N-H...O=C hydrogen-bonds in general. This is covered in chapters 3 and 4.

The results of chapter 5 show that, in small molecules, the geometry of non-hydrogen-bonded carbonyl-carbonyl interactions between *trans*-amides is similar to energetically favourable interaction motifs that were identified previously between ketone carbonyl groups. This finding suggested that a similar analysis of carbonyl-carbonyl interactions in proteins might show the occurrence of the same motifs. Chapter 6 deals with the occurrence of carbonyl-carbonyl interactions in proteins, and although their geometry is not similar to the motifs identified in ketones, other favourable carbonyl-carbonyl interaction motifs are identified.

In the course of reviewing previous work on the interactions of amino acid side-chain carbonyl groups with nearby main-chain carbonyl groups, I became interested in the geometry of certain recurring small hydrogen-bonded motifs, called asx-turns and ST-turns. This led to the investigation of the geometric relationships between asx-turns, ST-turns, and the more widely known  $\beta$ -turns. This is presented in chapter 7, and has been published, in shortened form (Duddy *et al.*, 2004).

In summary, chapter 3 covers the factors that affect the geometry of hydrogen-bonds between *trans*-amide groups in small molecule crystal structures. Chapter 4 goes on to separately examine the factor of steric accessibility to the *trans*-amide carbonyl oxygen atom, and how this affects hydrogen-bond geometry. Chapters 5 and 6 cover the analysis of carbonyl-carbonyl interactions between *trans*-amides in small molecule and protein crystal structures, respectively. Chapter 7 is the geometric comparison among asx-turns, ST-turns and  $\beta$ -turns.

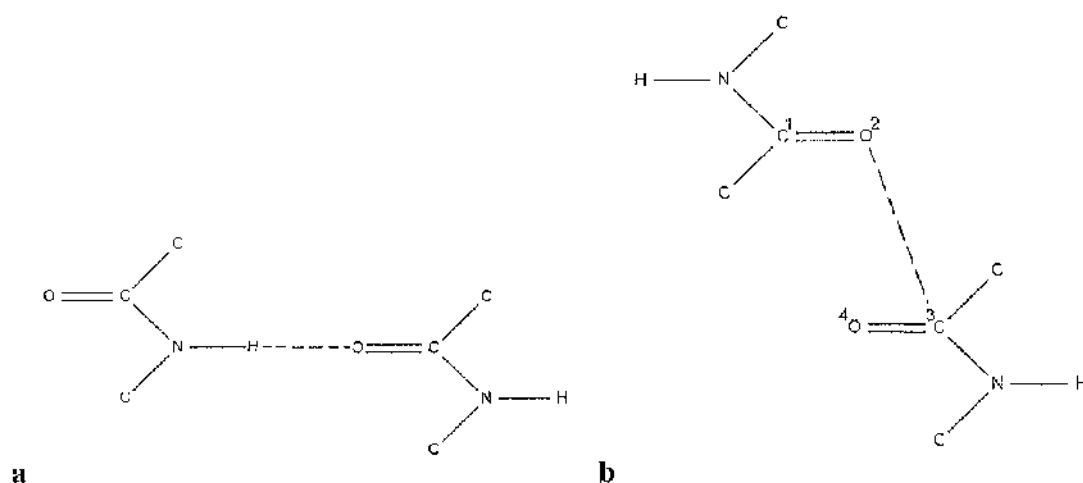
Section 2.2 introduces the general context. The following sections (2.3, 2.4, and 2.5) relate to the three topics: hydrogen-bonding between *trans*-amide groups, carbonyl-carbonyl interactions between *trans*-amide groups, and mimicry of  $\beta$ -turns by Asx- and ST-turns, respectively. Likewise, chapter 8 (Conclusions) has three sections (8.1, 8.2, and 8.3) relating to each of these three topics, leading onto a general conclusions section (8.4).



## 2.2 General

The *trans* secondary amide (*trans*-amide; as in Figure 2.1), or peptide unit, is formed by the joining of amino acids and is a central feature of protein backbone. The planarity of the *trans*-amide, caused by  $\pi$ -orbital overlap, provides rigidity within protein backbone (main-chain) structure such that, for each amino acid, there are only two torsion angles ( $\phi$  and  $\psi$ ) that possess significant rotational freedom. The formation of protein secondary structure can be regarded as the spatial re-orientation of backbone *trans*-amides, by rotations of  $\phi$  and  $\psi$ , such that the energetic favourability of their interactions with each other, and between them and their environment, is maximised.

Beyond steric considerations and hydrophobic effects, the greatest geometric influence is provided by the *trans*-amides...*trans*-amide hydrogen-bond (Figure 2.1a; reviewed by Rose *et al.*, 1993), although the effects of other types of hydrogen-bond (reviewed by Weiss *et al.*, 2001), particularly C-H...O (Derewenda *et al.*, 1995; Wahl & Sundaralingam, 1997) and C-H... $\pi$ -orbital (Brandl *et al.*, 2001), have recently been emphasised. Other electrostatic interactions, those between nearby carbonyl groups (Figure 2.1b), are also thought to be important (Maccallum *et al.*, 1995a,b; Allen *et al.*, 1998; Deane *et al.*, 1999; Lario & Vrielink, 2003).



**Figure 2.1.** Two types of electrostatic interaction between *trans*-amide groups. (a) Hydrogen-bonded. (b) Carbonyl...carbonyl.

## 2.3 Hydrogen-bonding between *trans*-amide groups

### 2.3.1 Context

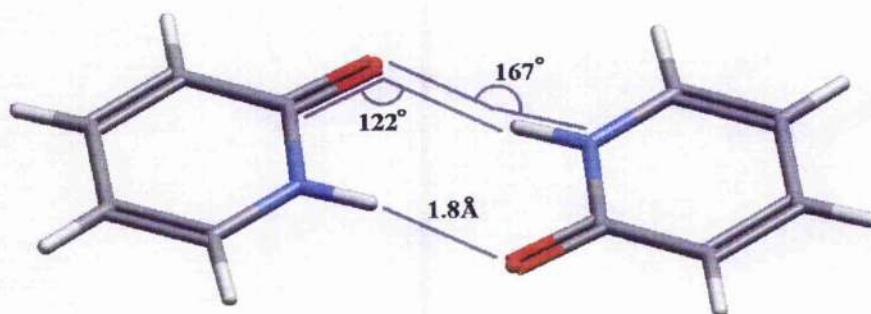
Chapters 3 and 4 present the results of an analysis of hydrogen-bonding between *trans*-amide groups in the CSD. The aim of this work was to resolve an apparent contradiction in current understanding of *trans*-amides...*trans*-amide hydrogen-bonds in proteins compared to the more general category of N-H...O=C hydrogen-bonds in small molecules. An influential body of work (Taylor *et al.*, 1983ab; Baker & Hubbard, 1984; Taylor & Kennard 1984; Vedani & Dunitz, 1985; Legon & Millen, 1988; Buckingham *et al.*, 1988; Scheiner, 1994; Aakeroy & Seddon, 1996; Jeffrey, 1997; Hay *et al.*, 2002; Sarkel & Desiraju, 2004), based largely on studies of small molecules with carbonyl groups, has led to emphasis on the idea that H...O=C angles tend to values of around 120°, consistent with the hydrogen atom bonding primarily with the oxygen's lone pair. However, the evidence from *trans*-amide...*trans*-amide hydrogen-bonds in proteins shows they mostly have much higher H...O=C angles, inconsistent with any lone pair effect. There is a need to examine *trans*-amide...*trans*-amide hydrogen-bonds in the CSD as a separate group and compare them with other N-H...O=C hydrogen-bonds.

### 2.3.2 Small molecule studies

Many examples of the *trans*-amide group can be found in small molecule crystal structures in the Cambridge Structural Database (CSD). Analysis of such crystal structures may reveal preferential intermolecular interactions between groups, and often detailed geometric information, such as interaction distances and angles, can be derived (Klebe, 1994; Mills & Dean, 1996). Such information about intermolecular interactions can be used, *inter alia*, in the design of drugs that occupy protein-ligand binding sites. Although information can be distilled from either small-molecule (CSD) or protein crystal structure (PDB), the use of small-molecule data does have some distinct advantages: generally, the resolution is much higher, and placement of hydrogen atoms is usually known.

Hydrogen-bonds between *trans*-amides (Figure 2.1a) are of the type N-H...O=C. Taylor *et al.* examined N-H...O=C hydrogen-bonds in the CSD, presenting their findings in a short series of papers (Taylor *et al.*, 1983a,b; Taylor *et al.*, 1984). In their analysis of lone-pair directionality (Taylor *et al.*, 1983a) they found that (a) there was a tendency for the hydrogen to sit in the plane described by the  $sp^2$  lone-pairs of the oxygen and (b) when the hydrogen position was projected onto this plane, the angle it made with the C=O group was on average 135°. They concluded that there is a distinct preference for N-H...O=C hydrogen-bonds to form in, or near to, the directions of the conventionally viewed  $sp^2$  lone-pairs. They found an increase in this preference for carbonyls that accepted two, rather than a single, hydrogen bond(s).

The results of Taylor *et al.* are in agreement with the empirically-derived Legon-Millen rules (Legon & Millen, 1987) for the lone-pair directionality of hydrogen-bonding of small molecules such as HF, HCCH, and HCN. Subsequent analyses of the CSD have also found agreement with these (Murray-Rust & Glusker, 1984; Vedani & Dunitz, 1985). However, it has been argued that maximisation of the number of favourable electrostatic interactions is more important than lone-pair directionality in determining the geometry of N-H...O=C hydrogen-bonds (Mitchell & Price, 1989). Specifically, the geometry of certain commonly occurring ring motifs containing more than one hydrogen-bond forces H...O=C angles to be near 120° (Figure 2.2). Ring motifs and their occurrence have been considered by Allen *et al.* (1999).



**Figure 2.2.** Hydrogen-bonded ring motif. From CSD refcode QIMHOJ. In this case the ring has 2-fold rotational symmetry. The H...O distance, N...H=O angle, and H...O=C angle are indicated.

### 2.3.3 Protein studies

The geometry of the hydrogen bond between two *trans*-amides in proteins has been studied most fully by Baker & Hubbard (1984), Stickle *et al.* (1992), and McDonald & Thornton (1994), and also by Fabiola *et al.*, (2002) and Kortemme *et al.*, (2003).

Baker & Hubbard (1984) analysed, *inter alia*, the H $\cdots$ O distances, the N-H $\cdots$ O angles, and the H $\cdots$ O=C angles of hydrogen-bonds found in  $\alpha$ -helices and  $\beta$ -sheets. The linearity of hydrogen-bonds at the hydrogen atom is fairly consistent, with N-H $\cdots$ O angles of around 160°. The *trans*-amide carbonyl group of each residue of an  $\alpha$ -helix tends to make two hydrogen-bonds, one with the N-H group of the residue three positions ahead (*i*, *i*+3) in the polypeptide chain, and one with the residue 4 ahead (*i*, *i*+4). The geometry of *i*, *i*+4 hydrogen-bonds and the hydrogen-bonds of  $\beta$ -sheets are similar, with H $\cdots$ O distances near 2Å (average of 2.04Å for *i*, *i*+4 and 1.94Å for  $\beta$ -sheet) and H $\cdots$ O=C angles near 150°. The *i*, *i*+3 bonds have long H $\cdots$ O distances, and H $\cdots$ O=C angles of about 116°. Although near to the value of 120° associated with lone-pair directionality, the hydrogen is well outside the *sp*<sup>2</sup> plane of the oxygen. Hydrogen-bonds of the *i*, *i*+3 type are likely to be much weaker than those of the *i*, *i*+4 type. Stickle *et al.* (1992) report similar conclusions regarding geometry, though they avoid the necessity of placing hydrogen atoms by measuring only the N $\cdots$ O distance and N $\cdots$ O=C angle. The H $\cdots$ O=C and N $\cdots$ O=C angles reported in these analyses describe a much more linear geometry at the oxygen atom than that observed by Taylor *et al.* (1983a) for the N-H $\cdots$ O=C hydrogen-bonds in the CSD.

McDonald & Thornton (1994) analysed buried hydrogen-bonds. This excludes those that are solvent-exposed and therefore prone to error or omission due to mobility and disorder at the protein surface. They found that hydrogen-bonds to buried main-chain carbonyl groups (from any N-H donor, including side-chains) have average H $\cdots$ O=C angles near 150°. They also found that a high proportion (about 80%) of buried main-chain carbonyls tend to make only one hydrogen bond, and that those making two had higher H $\cdots$ O distances. Baker & Hubbard (1984) had noted previously the tendency for main-chain carbonyls to form only one hydrogen bond -- 53% of their sample did, though this included solvent-exposed carbonyls.

A comprehensive analysis of the spatial distribution of biologically important functional groups within the CSD (Klebe, 1994) included the hydrogen-bonding of N-H and O-H groups to *trans*-amide carbonyl groups. As with the N-H...O=C hydrogen-bonds examined by Taylor *et al.* (1983a), a tendency was observed for the hydrogen to sit in the plane described by the  $sp^2$  lone-pairs. However, when the hydrogen position is projected onto this plane, the angle it makes with the C=O group is not in keeping with lone-pair directionality, and instead is much more like what is observed with *trans*-amide groups in proteins. Lone-pair directionality was, however, observed in this study when hydrogen-bonds to primary amides, rather than *trans*-amides, were examined.

Some early work (Leiserowitz & Tuval, 1977) on the hydrogen-bonded packing arrangements of a small number of amide-containing compounds noted the frequent occurrence of ribbons, or chains, of hydrogen-bonded symmetry-related *trans*-amides, running through the crystal lattice. It was suggested that, within these chains, lone-pair directionality is favoured. However, the small sample of observations therein presented show N...O=C angles in the range 139°-177°, higher than expected for lone-pair directionality.

## 2.4 Carbonyl-carbonyl interactions between *trans*-amide groups

### 2.4.1 Context

Chapters 5 and 6 present the results of an analysis of carbonyl-carbonyl interactions between *trans*-amide groups in the CSD and PDB, respectively. This work was inspired by the recent accumulation of evidence (Maccallum *et al.*, 1995a,b; Allen *et al.*, 1998; Deane *et al.*, 1999; Lario & Vrielink, 2003) for the importance of such interactions in protein and small molecule crystal structures.

### 2.4.2 Small molecule studies

Carbonyl groups are common organic building blocks, and their interactions have been studied extensively. The dipolar nature of carbonyl groups ( $>C^{\delta+}=O^{\delta-}$ ) allows favourable ( $\delta+$  to  $\delta-$ ) electrostatic interactions to take place between them. Early work on small, mainly non-biological molecules, in which dipolar carbonyl-carbonyl interactions were studied includes that of Bolton (1963, 1964, 1965), Bernstein *et al.* (1974), and Burgi *et al.* (1974). These studies showed that in crystal structures containing carbonyl groups, interactions between these groups occur frequently. Moreover, in structures containing both carbonyl groups and potential hydrogen donors, inter-molecular carbonyl-carbonyl interactions appear in some cases to be more important than inter-molecular hydrogen-bonds in determining the packing of molecules within the crystal structure.

More recently, a study of carbonyl-carbonyl interactions between ketone groups (carbon substituted  $>C=O$  groups) in the CSD, (Allen *et al.*, 1998) has shown that a majority (>70%) can be geometrically grouped into three common motifs (see Figure 5.1a): a slightly sheared antiparallel arrangement (in which each oxygen is near to the carbon of the other carbonyl); a perpendicular arrangement (in which the carbon of only one group can interact with the oxygen of the other group); and a highly sheared parallel arrangement (which also has only one close  $C\cdots O$  interaction). Intermolecular Perturbation Theory (IMPT) was used to calculate attractive interaction energies for the antiparallel and perpendicular types. For the antiparallel arrangement this was found to be comparable to

that of a medium-strength N-H...O=C hydrogen bond: A perfectly rectangular geometry had  $-22.3 \text{ kJ mol}^{-1}$  at a C...O distance of  $3.02 \text{ \AA}$ , the energy remained high ( $<-15 \text{ kJ mol}^{-1}$ ) out to a distance of  $3.6 \text{ \AA}$ , further than which calculations were not carried out. The energy of the perpendicular arrangement was comparable to that of a C-H...O=C hydrogen bond:  $-7.6 \text{ kJ mol}^{-1}$  at a distance of  $3.02 \text{ \AA}$ . The existence of these carbonyl-carbonyl interaction motifs in the crystal structures of acetone are thought to contribute to its unusual thermodynamic properties (Allan *et. al*, 1999).



### 2.4.3 Protein studies

Interactions between carbonyl groups in proteins have not attracted as much attention as those in the CSD. One exception to this, Maccallum *et al.* (1995a,b), used Lennard-Jones potentials to measure the coulombic energy of interactions between backbone *trans*-amides. The total energy was broken down by treating the N-H and carbonyl groups separately. It was found that the interaction between carbonyl groups was a principal source of stabilisation for a number of features of secondary structure. The right-handed twist of  $\beta$ -strands (and consequently the twist of  $\beta$ -sheet) and also the left-handedness of polyproline II helices and collagen helices (both of which are closely related to  $\beta$ -strands) are stabilised by favourable carbonyl-carbonyl interactions. Also, earlier observations (Artymiuk & Blake, 1981; Baker & Hubbard, 1984) of differences in the in-plane hydrogen-bond angles (i.e. the angle made with the C=O bond by the projection of the hydrogen atom position onto the  $sp^2$  plane of the oxygen) between parallel  $\beta$ -sheet, antiparallel  $\beta$ -sheet, and  $i, i+4$   $\alpha$ -helix hydrogen-bonds were explained in terms of carbonyl-carbonyl interactions (Maccallum *et al.*, 1995b).

Another example where carbonyl-carbonyl interactions were studied in proteins involves asparagine and aspartate (asx) residues (Deane *et al.*, 1999). These residues have a high propensity to adopt conformations in partially allowed regions of the Ramachandran plot, particularly the left-handed  $\alpha$ -helical region. In an effort to explain this, the distance between the carbonyl group of asx side-chains and their own main-chain carbonyl or the carbonyl of the preceding residue was measured. This distance was less than 4Å in more than 80% of cases. It was concluded that stacking of the carbonyls provides some explanation for the preponderance of asx residues in partially allowed regions of the Ramachandran plot.

Recently, Lario & Vrieling (2003) noted the difference in electron densities between the carbonyl groups of  $\alpha$ -helices and those of  $\beta$ -sheets, in an examination of the sub-Ångström resolution structure of cholesterol oxidase. In  $\alpha$ -helices the electron density is more polarised towards the oxygen atom. A number of possible explanations were considered for this including variations in hydrogen-bond geometry, differences in the peptide C-N

bond order associated with alternating resonance forms of the *trans*-amide, and carbonyl-carbonyl interactions. All but carbonyl-carbonyl interactions were ruled out.

Interactions have also been observed between ketone carbonyl groups of protein ligands with *trans*-amide carbonyl groups of the active site (Bergner *et al.*, 2001-2002). Of 821 ligands in the PDB comprising ketone moieties, 42 such examples were found, 13 of which occurred in the antiparallel arrangement identified by Allen *et al.* (1998).

## 2.5 Mimicry of $\beta$ -turns by Asx- and ST-turns

### 2.5.1 Context

Within proteins, asparagine and aspartate side-chain carbonyl groups can interact with nearby main-chain carbonyls. This stabilises conformations of these residues occurring within partially allowed regions of the Ramachandran plot (Deane *et al.*, 1999). Such interactions involve contact of the side-chain carbonyl with the main-chain carbonyl either of the asx residue itself, or of the previous residue in the chain. Review of the work of Deane *et al.* (1999) raised the possibility that many of the contacts between the asx side-chain carbonyl and its own main-chain carbonyl could be occurring within motifs called asx-turns. The geometric constraints of asx-turns dictate that the asx side-chain carbonyl is necessarily close to its own main-chain carbonyl. Based on this, I decided to investigate the geometry of asx-turns. The analysis did not produce new insights into carbonyl-carbonyl interactions, but did lead to an improved understanding of asx-, ST-, and  $\beta$ -turns, and the geometric relationships between them, as described in chapter 7, and as published, in shortened form (Duddy *et al.*, 2004).

### 2.5.2 $\beta$ -turns

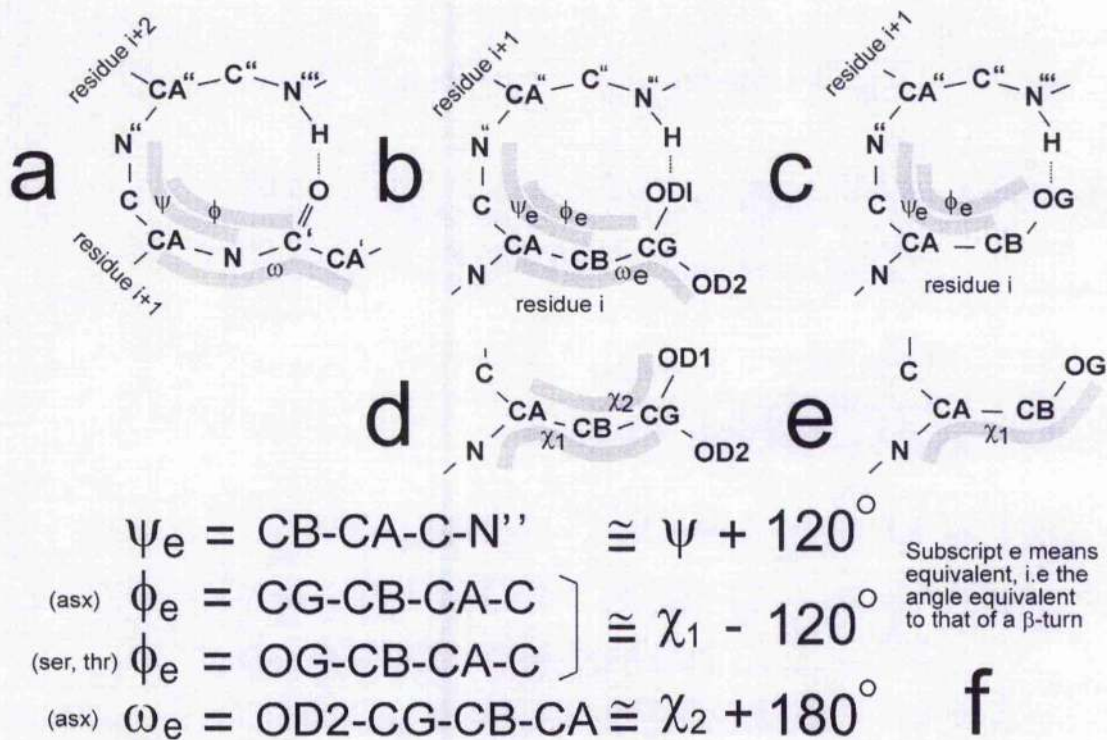
The four main types of  $\beta$ -turn, I, I', II and II', were first described by Venkatachalam (1968).  $\beta$ -turns are considered to have four residues,  $i$ ,  $i-1$ ,  $i+2$  and  $i+3$ . They are typically defined in two main ways, either by a distance of less than 7Å between the  $C_{\alpha}$  atoms of residues  $i$  and  $i+3$  (Lewis *et al.*, 1973) or by a hydrogen bond between the main-chain carbonyl group of residue  $i$  and the main-chain NH group of residue  $i+3$ . Here, like Venkatachalam, only the hydrogen-bonded turns are considered, eliminating those such as type VIII (Hutchinson & Thornton, 1996). Also excluded are those with *cis* peptide bonds. The validity of regarding  $\beta$ -turns of this sort as belonging to four main categories has been confirmed (Wilmot & Thornton, 1990) and later work has generally supported this classification (Richardson, 1981; Baker & Hubbard, 1984; Wilmot & Thornton, 1988, 1990; Hutchinson & Thornton, 1994, 1996; Gunasekaran *et al.*, 1998).

$\beta$ -turns are the most common small non-secondary structure motifs. They are often found at the loop ends of  $\beta$ -hairpins, and within several other motifs.

### 2.5.3 Asx-turns and ST-turns

Richardson (1981) pointed to the frequent occurrence of asx-turns and to their geometrical resemblance to hydrogen-bonded  $\beta$ -turns. The term asx is used to indicate either an aspartate or asparagine residue, both of which exhibit similar behaviour because both have side-chain carbonyl groups next to the  $\beta$ -carbon atom. Both asx-turns and  $\beta$ -turns form a 10-atom hydrogen-bonded ring. The difference, seen in Figures 2.3a and 2.3b, is that the main chain atoms in residue  $i$  and the NH group of residue  $i+1$  in the  $\beta$ -turn are replaced by the side-chain of an asx residue. Hence the asx side-chain atoms mimic the main-chain ones. Asx-turns have three residues,  $i$  (asx),  $i+1$  and  $i+2$ , so the residue numbering for homologous atoms differs from that of  $\beta$ -turns, such that  $i+1$  of asx-turns corresponds to  $i+2$  of  $\beta$ -turns. This is shown in Figure 2.3 (the 'equivalent angles' described in this figure are explained in Chapter 7). Several authors (Tainer *et al.*, 1982; Rees *et al.*, 1983; Richardson & Richardson, 1989; Eswar & Ramakrishnan, 1999; Wan & Milner-White, 1999a; Chakrabati & Pal, 2001) discuss the asx-turn and its similarity to the  $\beta$ -turn while others (Presta & Rose, 1988; Richardson & Richardson, 1988; Doig *et al.*, 1997; Aurora & Rose, 1998) show the asx-turn to be common at the N-termini of  $\alpha$ -helices.

The side-chain oxygen atoms of serine and threonine residues ( $i$ ) often form a hydrogen-bond with the main-chain NH groups of the residue two ahead ( $i+2$ ). Such features are like asx-turns except that they are 9-atom, instead of 10-atom, hydrogen-bonded rings. Data have been assembled (Baker & Hubbard, 1984; Eswar & Ramakrishnan, 1999, 2000) showing their common occurrence and pointing to their similarity with asx-turns. Here they are referred to collectively as ST-turns (as in Wan & Milner-White, 1999b). Sequence comparisons of homologous proteins (Vijayakumar *et al.*, 1999; Wan & Milner-White, 1999b) show that, within and between asx- and ST-turns, the four residues in question often substitute each other.



**Figure 2.3.** Defining residues and torsion angles in  $\beta$ -turns and in asx- and ST-turns. (a-c) are representations of the  $\beta$ -turn, asx-turn (with aspartate side chain) and ST-turn (with serine side chain), respectively. Residue positions are indicated, and it should be noted that position  $i+1$  of an asx- or ST-turn is equivalent to  $i+2$  of a  $\beta$ -turn. (d) and (e) show the  $\chi_1$  and  $\chi_2$  torsion angles of an aspartate and a serine. (b) and (c) also show the novel torsion angles  $\psi_e$ ,  $\phi_e$  and  $\omega_e$ , used in this work for comparison of asx- and ST-turns with  $\beta$ -turns (subscript e stands for equivalent). (f) shows the approximate relationships of  $\psi_e$ ,  $\phi_e$  and  $\omega_e$  with standard torsion angles, allowing short-hand comparisons to be made. The values of  $\psi_e$ ,  $\phi_e$  and  $\omega_e$  and of standard angles used in this work have been measured directly.

Previous work on ST-turns, in the general context of side-chain to main-chain hydrogen-bonding, has noted their prevalence and grouped them into three geometrically distinct categories, but their geometric similarity to  $\beta$ -turns has not been explicitly discussed. Asx-turns have been more extensively studied, and of particular relevance to the present work is that of Eswar and Ramakrishnan (1999). These authors categorised asx-turns into four geometrically distinct classes. They noted similarity with  $\beta$ -turns, and drew particular attention to the correspondence between one of their four classes and the type II'  $\beta$ -turn, and also to the observation that the second residue of asx-turns is often found in the bridge region of the Ramachandran plot, as is the equivalent residue of  $\beta$ -turns. Their findings are built upon in Chapter 7 through a direct comparison of asx-turns with  $\beta$ -turns, resulting in the observation that the four classes of asx-turn are geometrically equivalent to the four types of hydrogen-bonded  $\beta$ -turn. Additionally, it is found that the three categories of ST-turn are also equivalent to three (types I, II and II') of the four types of hydrogen-bonded  $\beta$ -turn. It is proposed that asx- and ST-turns be named using the type I, II, I' and II'  $\beta$ -turn nomenclature.

#### 2.5.4 Nests

A commonly occurring anion-binding motif has been described (Watson & Milner-White, 2002a,b; Pal *et al.*, 2002; Milner-White *et al.*, 2004) as a nest because it forms a concavity made from the main-chain NH groups of three successive residues. The anionic group bound is often a single oxygen atom with either a whole or a partial negative charge. A less common, though recurring, variant of this was called an asx-nest because the first nest residue is either aspartate or asparagine and its side chain oxygen is the anion in the nest. A similar distribution is found for ST-nests, which are like asx-nests except that they have either serine or threonine as the first residue and the S or T side-chain oxygen occupies the nest. They are relevant because it emerges that asx- and ST-turns mimicking type II  $\beta$ -turns are the same as asx- and ST-nests.

### 3 Hydrogen-bonding between *trans*-amide groups: Factors affecting hydrogen bond geometry

#### 3.1 Brief introduction and outline

The current evidence suggests that hydrogen-bonds made to *trans*-amide carbonyl groups are more linear at the oxygen acceptor atom than N-H...O=C hydrogen-bonds, despite the influence of lone-pair directionality, and not just within protein secondary structure.

Focussing on *trans*-amide...*trans*-amide hydrogen bonds in the CSD, a subgroup not previously examined, the possible reasons for this disparity are examined.

The geometry of the *trans*-amide...*trans*-amide hydrogen-bond is compared with the geometry of other hydrogen-bonded systems including the simple N-H...O=C hydrogen-bond, and hydrogen-bonds involving primary amides. The effects of three factors on hydrogen-bond geometry are analysed: the occurrence of hydrogen-bonded chains of *trans*-amides extending through the crystal lattice; the number of hydrogen-bonds made by the *trans*-amide carbonyl (including a comparison of the proportion of *trans*-amide carbonyls that accept multiple hydrogen-bonds compared with the proportion for C=O groups in general); and the occurrence of hydrogen-bonded ring motifs among the N-H...O=C hydrogen-bonds (as described by Mitchell & Price, 1989). The effects of these factors on the linearity at the oxygen acceptor atom are summarised. Lastly, the internal geometry of the *trans*-amide is considered.

## 3.2 Materials and methods

### 3.2.1 Database searches

A *trans*-amide is defined here as a secondary amide group with the C-CO-NH-C atomic configuration as in Figure 2.1 such that the CO and NH subgroups lie *trans* to one another.

The November 2003 release of the Cambridge Structural Database (version 5.25) contained a total of 298,097 crystal structures of small molecules. The ConQuest software package (Allen, 2002; Bruno *et al.*, 2002) was used to identify *trans* amide groups within a subset of this database. The subset consisted only of structures: (a) of organic compounds; (b) rated error-free by CSD evaluation procedures; (c) that were not polymeric; (d) with R-values  $< 0.075$ . Occurrences of intermolecular hydrogen-bonds between *trans*-amide groups were identified and characterised using ConQuest. The geometric search criteria were two amide units, constrained to the *trans* configuration by limiting the C-N-C-C torsion angle to  $180^\circ \pm 90^\circ$  (99% of the *trans*-amide groups identified in this way had CNCC =  $180^\circ \pm 20^\circ$ ).

A crystal structure may include more than one unique occurrence of an interaction, such that, in the database, the number of occurrences of an interaction exceed the number of crystal structures with the interaction. Also, there may be more than one independent molecule per asymmetric unit and each unique interaction is counted separately. A structural subset is defined as the set of crystal structures with one or more occurrences of the specified interaction. It is of importance to be aware of the situation when employing ConQuest to do database searches of the form "x **not** y" where x and y are specific interaction types. This specifies that the subset of structures containing occurrences of y have been excluded, and occurrences of x are identified only within the subset of structures not containing any instances of y.

The following parameters were calculated for each hydrogen bond occurrence (allowed limits for hydrogen bond definition are given in brackets): H $\cdots$ O distance ( $< 2.5 \text{ \AA}$ ); N-H $\cdots$ O angle ( $> 90^\circ$ ) and H $\cdots$ O=C angle ( $> 90^\circ$ ). Bond lengths to terminal hydrogen atoms are underestimated in X-ray studies due to hydrogen atom asphericity effects (Allen,



1986). In this study, hydrogen atom positions were normalised by ConQuest such that the N-H distance was fixed to 1.009 by moving the hydrogen atom out along the N-H vector. The 1.009 value is the mean bond length determined from neutron diffraction work.

Three-centered, or bifurcated, hydrogen bonds were defined as cases where the N-H, besides satisfying the hydrogen bond criteria above, was also close to another electronegative atom, X, where X is O, N, S, or halogen. The distance N...X was required to be within the sum of the Van der Waals radii of N and X plus an allowance of 0.2Å. Additionally the N-H...X angle was limited to  $> 90^\circ$ .

Occurrences of the different categories of hydrogen-bonded interactions were also identified using ConQuest. A *trans*-amide...*trans*-amide chain was defined as four *trans*-amide groups (a-d) arranged a...b...c...d such that there are three hydrogen bonds of the type shown (Figure 2.1a), between a...b, b...c, and c...d, the a-d labels being applied arbitrarily by the search procedure. Values describing hydrogen bond geometry are taken from the b...c interaction in each case. Two common types of repeating chain, involving symmetry related groups, were identified: the singly repeating  $\cdots(a)\cdots a\cdots a\cdots$  variety involving successive structurally identical *trans*-amide groups, and the doubly repeating  $(a\cdots b)\cdots a\cdots b\cdots$  variety involving two structurally distinct *trans*-amide groups. A ring motif was defined as a hydrogen-bonded structure of the type shown in Figure 2.2.

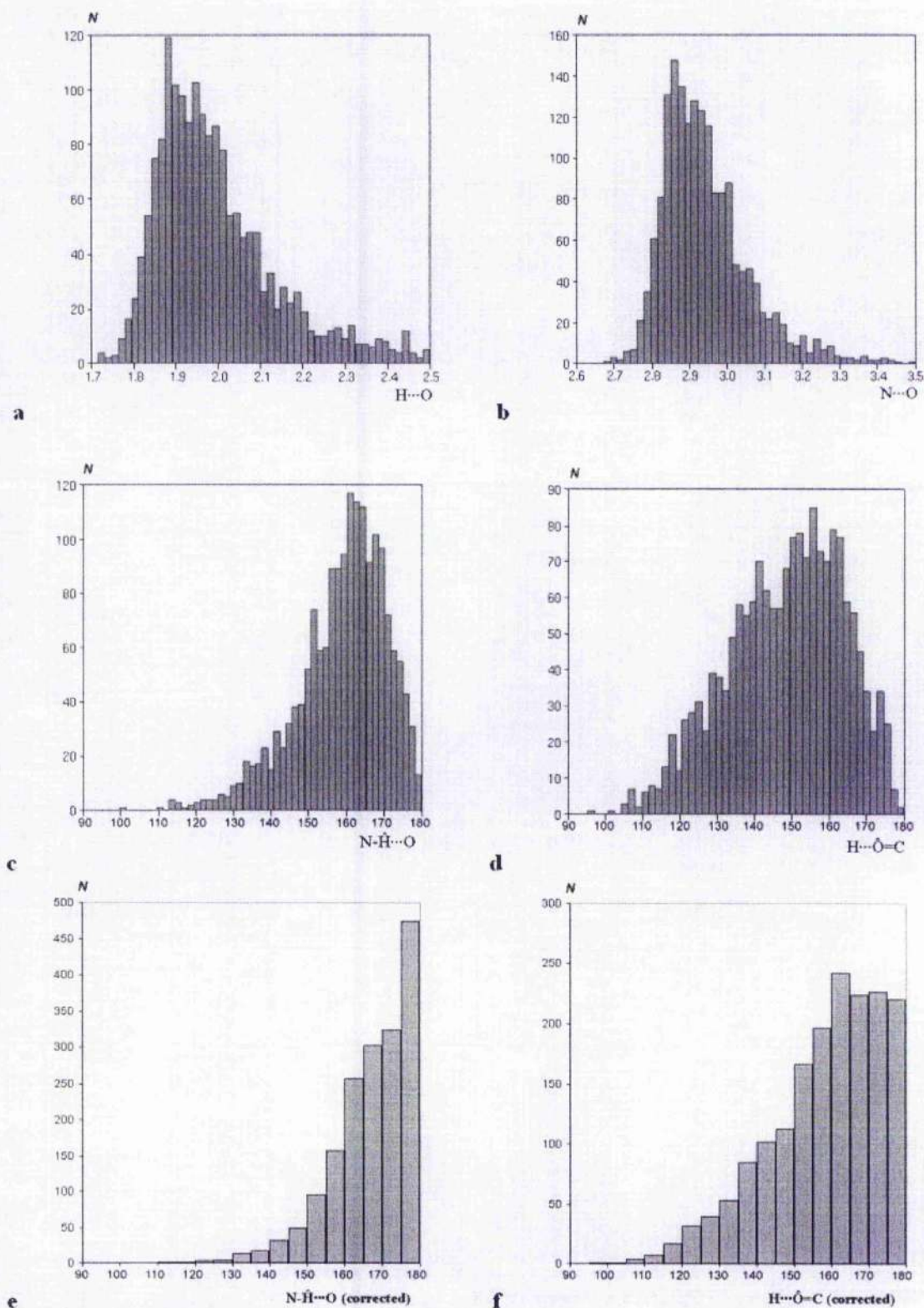
### 3.2.2 Statistical correction for angular data

The method used by Taylor & Kennard (1984) and Kroon *et al.*, (1975) was followed to correct statistical bias of angular data (Figures 3.1e and 3.1f). Occurrences were counted in bins with increments of  $5^\circ$  in the range  $90^\circ$  to  $180^\circ$ . The number of observations in each bin was divided by  $\sin x$ , where x is the average of the upper and lower boundaries of the bin, correcting the numbers of observations for each bin such that they were proportional to  $\sin x$ . The corrected numbers of observations were then multiplied by a normalisation value,  $N$ , derived by dividing the total number of observations before correction by the total number after correction.

### 3.3 Results & Discussion

#### 3.3.1 Hydrogen-bonds between *trans*-amide groups

The November 2003 release of the CSD contained 3077 structures containing at least one *trans*-amide group, and 5343 unique *trans*-amide groups. A proportion (2097; 39%) of *trans*-amide groups were found to make an inter-molecular N-H...O=C hydrogen bond (of the form shown in Figure 2.1a) to at least one other *trans*-amide. The following parameters were calculated for each hydrogen-bond: (a) H to O distance (H...O); (b) N to O distance (N...O); angles (c) N-H...O, (d) H...O=C and (e) N...O=C. The distributions of a-d are presented (Figure 3.1a-d) and the mean and standard deviation of each are given (Table 3.1). For three-dimensional angular data such as N-H...O and H...O=C angles there is a need to take account of the geometrical bias against values near 180°. This is because the available contact area on a sphere becomes proportionately less for a given change in angle as linearity is approached. A method is employed (Taylor & Kennard, 1984; Kroon *et al*, 1975) to correct for this bias. The corrected angle distributions are in Figures 3.1e and 3.1f. The peak at 175-180° in Figure 3.1e shows that the distribution of the N-H...O angle is consistent with a preference for linear geometry, while that for the H...O=C is less so.



**Figure 3.1.** Geometric parameters of *trans*-amide...*trans*-amide hydrogen-bonds. Histograms showing distributions of: (a) H to O distance ( $H\cdots O$ ); (b) N to O distance ( $N\cdots O$ ); (c) N-H...O angle; (d) H...O=C angle; (e) N-H...O after correction for geometrical bias; and (f) H...O=C angle after correction for geometrical bias.

### 3.3.2 Comparison with other hydrogen-bonded systems

The average geometry of the *trans*-amide...*trans*-amide hydrogen bond was compared with the geometry of several other hydrogen-bonded systems, ranging in complexity upwards from the simplest case of the N-H...O=C hydrogen bond (Table 3.1). Differences in mean values for all parameters were minimal, except those describing the angle at the oxygen acceptor atom: N...O=C and H...O=C. The *trans*-amide...*trans*-amide hydrogen bond had an average H...O=C of 148°, whereas for N-H...O=C the same angle had the lower value of 134°, dropping to 131° when structures containing *trans*-amide...*trans*-amide hydrogen-bonds were excluded.

The subsets N-H...*trans*-amide and *trans*-amide...O=C both (when structures containing *trans*-amide...*trans*-amide hydrogen-bonds were excluded) had average H...O=C values of around 140°, intermediate between the values for N-H...O=C and *trans*-amide...*trans*-amide (Table 3.1).

The hydrogen-bonds made by primary amides to each other or with other N-H or O=C groups, were also examined. Only occurrences where a hydrogen bond was made by the hydrogen *trans* to the carbonyl oxygen were considered. The primary amide...primary amide hydrogen bond had an average H...O=C of 139°, closer to the value for N-H...O=C than to the value for *trans*-amide...*trans*-amide. The average for primary amide...O=C was also 139°, but for N-H...primary amide it was 130° - much lower still, and very similar to that for the general N-H...O=C hydrogen bond after *trans*-amide...*trans*-amide occurrences are excluded (131°).

**Table 3.1.** Geometry of *trans*-amide...*trans*-amide hydrogen-bonds and hydrogen-bonds from other structural subsets. The mean and standard deviation are given for each geometric parameter. Subset size is the number of structures that contain one or more hydrogen bond of the specified type. The sample size, *N*, is the number of hydrogen-bonds of the specified type that occur within the subset. Where a subset is specified as 'not *trans*-amide...*trans*-amide' this indicates that structures containing *trans*-amide...*trans*-amide hydrogen-bonds have been excluded from that subset.

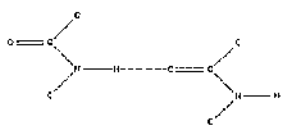
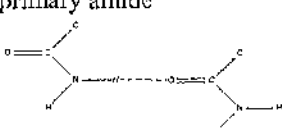
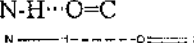
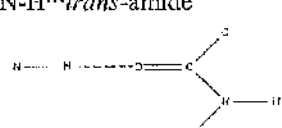
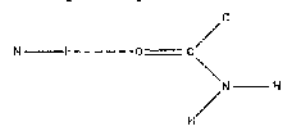
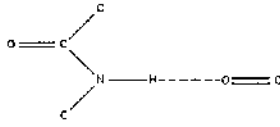
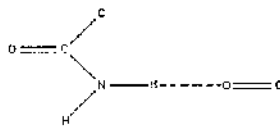
Type of hydrogen bond	Subset size	<i>N</i>	Parameter				
			H...O (sd)	N...O (sd)	N-H...O (sd)	N...O=C (sd)	H...O=C (sd)
<i>trans</i> -amide... <i>trans</i> -amide	1245	1731	2.00 (0.1)	2.94 (0.1)	159 (12)	152 (16)	148 (15)
							
primary amide ( <i>trans</i> H)... primary amide	202	259	2.05 (0.2)	2.96 (0.1)	153 (15)	143 (17)	139 (16)
							
N-H...O=C	8324	13825	1.98 (0.2)	2.91 (0.1)	158 (15)	135 (18)	134 (17)
							
N-H...O=C	7079	11657	1.97 (0.2)	2.91 (0.1)	158 (15)	133 (17)	131 (16)
not <i>trans</i> -amide... <i>trans</i> -amide							
N-H... <i>trans</i> -amide	1521	2190	2.00 (0.1)	2.94 (0.1)	158 (12)	149 (17)	147 (16)
							
N-H... <i>trans</i> -amide	276	335	1.99 (0.2)	2.92 (0.1)	156 (14)	144 (17)	141 (16)
not <i>trans</i> -amide... <i>trans</i> -amide							
N-H...primary amide	440	744	2.00 (0.1)	2.95 (0.1)	161 (14)	131 (16)	130 (15)
							
N-H...primary amide	238	306	1.97 (0.1)	2.94 (0.1)	165 (12)	128 (13)	128 (13)
not primary amide ( <i>trans</i> H)...primary amide							

Table continued...

<i>trans</i> -amide...O=C	1726	2378	2.01 (0.1)	2.95 (0.1)	159 (12)	149 (17)	146 (16)
							
<i>trans</i> -amide...O=C <b>not</b> <i>trans</i> -amide... <i>trans</i> -amide	481	568	2.02 (0.1)	2.97 (0.1)	160 (12)	141 (17)	139 (16)
primary amide ( <i>trans</i> H)... O=C	300	371	2.04 (0.1)	2.97 (0.1)	155 (14)	142 (18)	139 (16)
							
primary amide ( <i>trans</i> H)... O=C <b>not</b> primary amide ( <i>trans</i> H)...primary amide	98	110	2.03 (0.1)	2.98 (0.1)	159 (12)	141 (19)	139 (18)

### 3.3.3 Hydrogen-bonded chains

A frequent occurrence is the formation of hydrogen-bonded chains of *trans*-amides extending through the crystal lattice (examples are shown in Figure 3.6). These chains feature hydrogen-bonds between symmetry-related groups. Of the 2097 *trans*-amides involved in *trans*-amide...*trans*-amide hydrogen-bonds, 1336 (64%) were involved in chains (geometric data in Table 3.2). The number of chains totalled 1337. These were dominated by two repeating types,  $\cdots(\cdot a \cdot)_r \cdots$  and  $\cdots(\cdot a \cdots b \cdot)_r \cdots$  (see Materials and methods section 3.2.1), that together accounted for 91% of the total. There were 878 occurrences of  $\cdots(\cdot a \cdot)_r \cdots$ , and 340 occurrences of  $\cdots(\cdot a \cdots b \cdot)_r \cdots$ . Hydrogen bonds within chains have higher angles at the oxygen atom (average  $N \cdots O=C$  of  $153^\circ$  and  $H \cdots O=C$  of  $150^\circ$ ) than those occurring in structures that do not contain chains (average  $N \cdots O=C$  of  $146^\circ$  and  $H \cdots O=C$  of  $144^\circ$ ).

A large proportion (80%) of primary amide (*trans* H)...primary amide hydrogen-bonds were also found to occur within chains. In this case, the small number that did not occur within chains had a higher average  $H \cdots O=C$  angle, unlike the *trans*-amide...*trans*-amide case where it was lower.

**Table 3.2.** Occurrence of *trans*-amide...*trans*-amide hydrogen-bonds within and outwith chains. Subset size is the number of structures that contain one or more occurrence(s) of the specified type. The number of unique groups is the number of structurally distinct *trans*-amide groups involved in the occurrences. The sample size, *N*, is the number of occurrences of the specified type. The mean and standard deviation are given for each parameter. Values for all *trans*-amide...*trans*-amide, N-H...O=C (excluding *trans*-amide...*trans*-amide), and primary amide...primary amide hydrogen-bonds are given for ease of comparison.

Type	Subset size	Unique groups	<i>N</i>	Parameter				
				H...O (sd)	N...O (sd)	N-H...O (sd)	N...O=C (sd)	H...O=C (sd)
<i>trans</i> -amide... <i>trans</i> -amide	1245	2097	1731	2.00 (0.1)	2.94 (0.1)	159 (12)	151 (16)	148 (15)
primary amide ( <i>trans</i> H)...primary amide	202	284	259	2.05 (0.2)	2.96 (0.1)	153 (15)	143 (17)	139 (16)
N-H...O=C	7079	N/A	11657	1.97 (0.2)	2.91 (0.1)	158 (15)	133 (17)	131 (16)
<b>not</b> <i>trans</i> -amide... <i>trans</i> -amide								
<i>trans</i> -amide... <i>trans</i> -amide within chain	966	1336	1337	2.00 (0.1)	2.95 (0.1)	159 (12)	153 (15)	150 (15)
<i>trans</i> -amide... <i>trans</i> -amide <b>not</b> within chain	279	735	384	2.00 (0.1)	2.94 (0.1)	158 (13)	146 (16)	144 (16)
primary amide ( <i>trans</i> H)...primary amide within chain	181	227	233	2.05 (0.2)	2.96 (0.1)	153 (15)	143 (18)	138 (16)
primary amide ( <i>trans</i> H)...primary amide <b>not</b> within chain	21	54	28	2.05 (0.2)	2.96 (0.1)	154 (17)	143 (12)	145 (15)



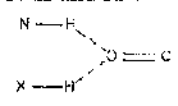
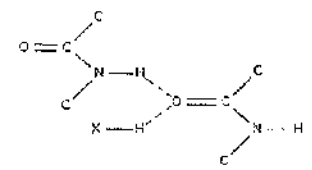
### 3.3.4 Multiple hydrogen-bonds

To investigate the influence of the number of hydrogen-bonds accepted on angles at the oxygen acceptor atom, and on general hydrogen bond geometry, the two systems, N-H $\cdots$ O=C and *trans*-amide $\cdots$ *trans*-amide, were tested for the presence or absence of at least one extra X-H $\cdots$ O hydrogen bond (Table 3.3). Where a second hydrogen-bond was present, the average H $\cdots$ O=C angles of both the original and second hydrogen-bonds were markedly lower than those observed for the singly bonded case.

In the case of the N-H $\cdots$ O=C system with a single hydrogen-bond, the average H $\cdots$ O=C angle was 138°, and with two hydrogen-bonds was 126° and 125° for the original and second bonds respectively. The *trans*-amide $\cdots$ *trans*-amide system with a single hydrogen-bond had an average H $\cdots$ O=C angle of 150°; this dropped to 136° and 130° for the original and second bonds respectively when a second bond was present. Other geometric parameters were not greatly affected by the influence of multiple hydrogen-bonds. The proportion of *trans*-amide $\cdots$ *trans*-amide occurrences with multiple hydrogen-bonds (9%) is much lower than the proportion of N-H $\cdots$ O=C occurrences with multiple hydrogen-bonds (26%). This relative weight will give multiple hydrogen-bonding in the N-H $\cdots$ O=C system a greater influence on the overall average H $\cdots$ O=C angle than in the *trans*-amide $\cdots$ *trans*-amide system.

The influence of multiple hydrogen-bonds on *trans*-amide $\cdots$ *trans*-amide occurrences within hydrogen-bonded chains was also considered, and compared with occurrences in the subset of structures that did not contain chains (Table 3.3). The two factors, multiple hydrogen-bonding and occurrence within a chain, are largely independent of one another as the proportion of within chain occurrences that had multiple hydrogen-bonds (9%) was approximately the same as the proportion of non-chain occurrences that had multiple hydrogen-bonds (9%).

**Table 3.3.** Acceptance of multiple hydrogen-bonds at oxygen atom. Occurrences of a second, X-H...O (X = N,O), hydrogen bond in the N-H...O=C and *trans*-amide...*trans*-amide hydrogen-bonded systems are considered. *Trans*-amide...*trans*-amide hydrogen-bonds occurring within hydrogen-bonded chains, and those occurring in the subset of structures that did not contain chains, are shown separately. Geometric data are presented for both the N-H...O and X-H...O hydrogen-bonds to O=C and *trans*-amide groups. The mean and standard deviation are given for each geometric parameter. Subset size is the number of structures that contain one or more occurrence(s) of the specified type. Values for each full subset (including both multiple and single hydrogen-bonds) are given for ease of comparison.

Type	Subset size	N	Parameter				
			H...O (sd)	N...O (sd)	N-H...O (sd)	N...O=C (sd)	H...O=C (sd)
N-H...O=C	8324	13825	1.98 (0.2)	2.91 (0.1)	158 (15)	135 (18)	134 (17)
N-H and X-H ...O=C	2288	3540					
			N-H...O=C 2.00 (0.2)	N...O 2.93 (0.1)	N-H...O 157 (16)	N...O=C 128 (16)	H...O=C 126 (15)
			X-H...O=C 1.94 (0.2)	N...O 2.86 (0.2)	N-H...O 158 (16)	N...O=C 127 (16)	H...O=C 125 (14)
N-H...O=C	6036	8094	1.97 (0.2)	2.91 (0.1)	159 (14)	139 (18)	138 (17)
not N-H and X-H ...O=C							
<i>trans</i> -amide... <i>trans</i> -amide	1245	1731	2.00 (0.1)	2.94 (0.1)	159 (12)	151 (16)	148 (15)
<i>trans</i> -amide and X-H ... <i>trans</i> -amide	129	153					
			N-H...O=C 2.08 (0.2)	N...O 3.02 (0.1)	N-H...O 156 (13)	N...O=C 138 (18)	H...O=C 136 (16)
			X-H...O=C 2.02 (0.2)	N...O 2.92 (0.2)	N-H...O 154 (16)	N...O=C 131 (14)	H...O=C 130 (13)
<i>trans</i> -amide... <i>trans</i> -amide not <i>trans</i> -amide and X-H ... <i>trans</i> -amide	1116	1501	1.99 (0.1)	2.94 (0.1)	159 (12)	153 (15)	150 (15)
<i>trans</i> -amide... <i>trans</i> -amide within chain	966	1337	2.00 (0.1)	2.95 (0.1)	159 (12)	153 (15)	150 (15)
<i>trans</i> -amide... <i>trans</i> -amide within chain and <i>trans</i> -amide and X-H ... <i>trans</i> -amide	103	117	2.07 (0.2)	3.01 (0.1)	156 (13)	140 (19)	137 (16)
<i>trans</i> -amide... <i>trans</i> -amide within chain not <i>trans</i> -amide and X-H ... <i>trans</i> -amide	863	1156	1.99 (0.1)	2.94 (0.1)	160 (11)	155 (15)	151 (14)
<i>trans</i> -amide... <i>trans</i> -amide not within chain	279	384	2.00 (0.1)	2.94 (0.1)	158 (13)	146 (16)	144 (16)
<i>trans</i> -amide... <i>trans</i> -amide not within chain and <i>trans</i> -amide and X-H ... <i>trans</i> -amide	26	36	2.10 (0.1)	3.04 (0.1)	158 (13)	134 (16)	133 (16)
<i>trans</i> -amide... <i>trans</i> -amide not within chain not <i>trans</i> -amide and X-H ... <i>trans</i> -amide	253	335	1.99 (0.1)	2.93 (0.1)	157 (14)	147 (15)	146 (15)

### 3.3.5 Ring motifs

N-H...O=C hydrogen-bonds can occur within ring motifs (Figure 2.2) of the type described by Mitchell & Price (1989) that involve two hydrogen-bonds. They require the N-H and C=O of an amide to be *cis*, so are formed by primary amides and *cis*- (secondary) amides, but not *trans*-amides. The number of structures in the CSD containing ring motifs was 1631. The geometry of ring motif hydrogen-bonds and of subsets excluding ring motifs was investigated (Table 3.4). A low average value (123°) of the H...O=C angle was observed in ring motifs. Their geometry was also characterised by the linearity of the N-H...O angle. This averaged 167°, which is 9° higher than the value for N-H...O=C hydrogen-bonds in general.

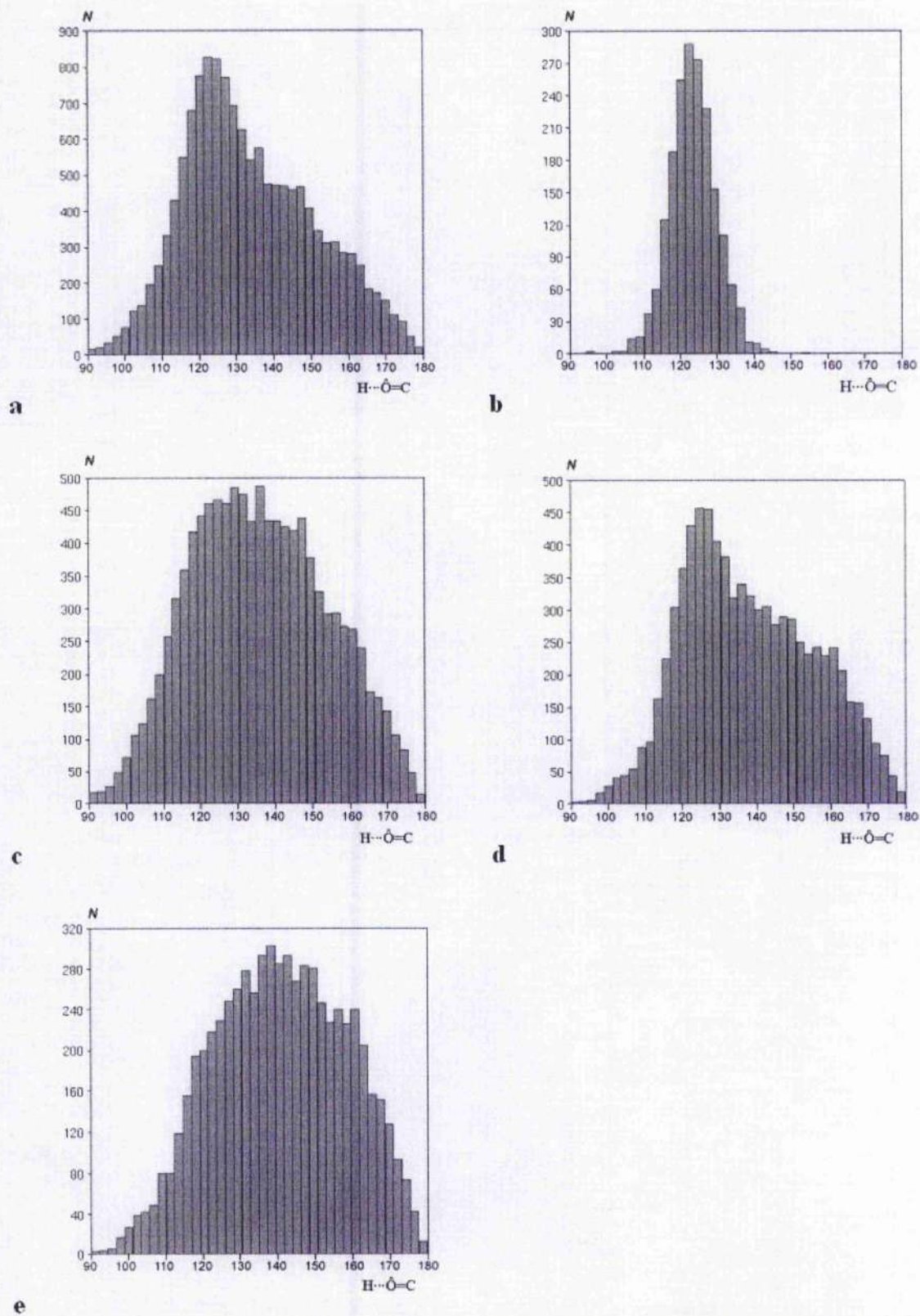
When structures containing ring motifs were excluded from the N-H...O=C hydrogen-bonds, the average H...O=C value increased from 134° to 136°. Further exclusion, of multiple hydrogen-bonding or of *trans*-amide...*trans*-amide hydrogen-bonds, produced values of 141° and 140° respectively.

To show the differential effects of exclusion of structures containing ring motifs on the geometry of hydrogen-bonds formed by either one or other of the two primary amide hydrogen atoms, values are given separately for hydrogen-bonds made by the primary amide hydrogen *trans* to O=C, and the primary amide hydrogen *cis* to O=C. As would be expected, exclusion of structures containing ring motifs causes little change in the case of the *trans* hydrogen, however for the *cis* hydrogen an increase in average H...O=C from 125° to 131° is observed. This increases further to 134° when multiple hydrogen-bonds are also excluded.

**Table 3.4.** The effect of ring motifs on hydrogen bond geometry. The mean and standard deviation are given for each geometric parameter. Subset size is the number of structures that contain one or more hydrogen bond of the specified type. The sample size,  $N$ , is the number of hydrogen-bonds of the specified type that occur within the subset.

Type	Subset size	$N$	Parameter				
			H...O (sd)	N...O (sd)	N-H...O (sd)	N...O=C (sd)	H...O=C (sd)
N-H...O=C of ring motif	1635	1892	1.91 (0.1)	2.90 (0.1)	167 (9)	122 (6)	123 (6)
N-H...O=C	8324	13825	1.98 (0.2)	2.91 (0.1)	158 (15)	135 (18)	134 (17)
N-H...O=C not ring motif	6689	10776	1.98 (0.2)	2.91 (0.1)	157 (15)	138 (19)	136 (18)
N-H...O=C neither ring motif nor N-H and X-H ...O=C	4913	6549	1.98 (0.2)	2.92 (0.1)	157 (14)	143 (18)	141 (18)
N-H...O=C neither ring motif nor N-H and X-H ...O=C nor <i>trans</i> -amide... <i>trans</i> -amide	3086	3860	2.00 (0.2)	2.94 (0.1)	157 (14)	142 (18)	140 (17)
primary amide...O=C ( <i>trans</i> hydrogen)	300	371	2.04 (0.1)	2.97 (0.1)	155 (14)	142 (18)	139 (16)
primary amide...O=C ( <i>trans</i> hydrogen) not ring motif	146	167	2.06 (0.2)	2.97 (0.1)	153 (16)	142 (18)	139 (18)
primary amide...O=C ( <i>trans</i> hydrogen) neither ring motif nor p. amide and X-H ...O=C	77	84	2.05 (0.1)	2.96 (0.1)	154 (16)	143 (19)	141 (19)
primary amide...O=C ( <i>cis</i> hydrogen)	410	520	1.97 (0.1)	2.95 (0.1)	166 (10)	124 (11)	125 (11)
primary amide...O=C ( <i>cis</i> hydrogen) not ring motif	143	160	2.00 (0.1)	2.96 (0.1)	161 (11)	131 (15)	131 (15)
primary amide...O=C ( <i>cis</i> hydrogen) neither ring motif nor p. amide and X-H ...O=C	69	76	1.98 (0.1)	2.94 (0.1)	161 (11)	134 (15)	134 (15)

Although the effect of excluding ring motifs on the average  $\text{H}\cdots\text{O}=\text{C}$  value of  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  hydrogen-bonds is relatively small (an increase of  $2^\circ$ ) compared with the effect of excluding multiple hydrogen-bonds (an increase of  $4^\circ$ ), its effect on the distribution of  $\text{H}\cdots\text{O}=\text{C}$  values is more dramatic (Figure 3.2). The large peak at approximately  $120^\circ$  in the  $\text{H}\cdots\text{O}=\text{C}$  distribution for all  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  hydrogen-bonds (Figure 3.2a), is caused by the ring motifs (Figure 3.2b), and disappears completely to leave a relatively smooth normal distribution when ring motifs are excluded (Figure 3.2c). It can also be seen that the effect of exclusion of multiple hydrogen-bonds on the distribution is less marked (Figures 3.2d and 3.2c).

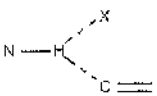
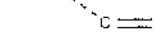
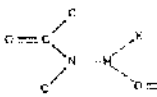
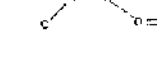


**Figure 3.2.** Distributions of  $\text{H}\cdots\text{O}=\text{C}$  angle of  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  hydrogen-bonds: (a) all  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  hydrogen-bonds; (b) those occurring in ring motifs; (c) those occurring in the subset of structures that do not contain ring motifs; (d) those occurring in the subset of structures that do not contain multiple hydrogen-bonds; (e) those occurring in the subset of structures that contain neither ring motifs nor multiple hydrogen-bonds.

### 3.3.6 Three-centered hydrogen-bonds

A further factor is the occurrence of 3-centered, or bifurcated, hydrogen-bonds (Taylor *et al.*, 1984); the frequency and geometry of these is shown in Table 3.5. Although rare among *trans*-amide...*trans*-amide hydrogen-bonds, they make up 14% of the N-H...O=C set. In the two cases, both N-H...O and H...O=C values are low. Also shown are average values after excluding 3-centered and other subsets (multiple hydrogen-bonds; ring motifs; and hydrogen-bonded chains). The values for the two sets of average N-H...O angles, compared to the multiple hydrogen bonds at the carbonyl oxygen, are more disparate, suggesting that, for 3-centered bonds, there are often one major and one minor hydrogen bond, rather than both being similar.

**Table 3.5.** Three-centered hydrogen-bonds. The mean and standard deviation are given for each geometric parameter. Subset size is the number of structures that contain one or more hydrogen bond of the specified type. The sample size,  $N$ , is the number of hydrogen-bonds of the specified type that occur within the subset.

Type	Subset size	$N$	Parameter				
			H... O/X(Å) (sd)	N... O/X(Å) (sd)	N-H... O/X(°) (sd)	N... O=C(°) (sd)	H... O=C(°) (sd)
N-H...O=C	8324	13825	1.98 (0.2)	2.91 (0.1)	158 (15)	135 (18)	134 (17)
3-centered	1222	1877					
 N-H...X			2.49 (0.3)	3.05 (0.2)	117 (18)	n/a	n/a
 N-H...O=C			2.03 (0.2)	2.89 (0.1)	147 (20)	128 (16)	125 (16)
N-H...O-C	2880	3553	2.00 (0.2)	2.90 (0.3)	158 (13)	143 (17)	141 (17)
<b>not</b> N-H and X-H ...O=C							
<b>not</b> ring motif							
<b>not</b> 3-centered							
<b>not</b> <i>trans</i> -amide... <i>trans</i> -amide							
<i>trans</i> -amide... <i>trans</i> -amide	1245	1731	2.00 (0.1)	2.94 (0.1)	159 (12)	151 (16)	148 (15)
3-centered	26	34					
 N-H...X			2.42 (0.2)	3.08 (0.1)	124 (15)	n/a	n/a
 N-H...O=C			2.18 (0.2)	3.04 (0.1)	145 (15)	140 (14)	134 (11)
<i>trans</i> -amide... <i>trans</i> -amide	242	316	1.99 (0.1)	2.93 (0.1)	158 (14)	148 (15)	146 (15)
<b>not</b> <i>trans</i> -amide and X-H ...O=C							
<b>not</b> chain							
<b>not</b> 3-centered							



### 3.3.7 Summary of factors affecting H...O=C angle

Differences in the H...O=C angle between different subsets are summarised in Figure 3.3. The effects of multiple hydrogen-bonding (where the O=C accepts two or more donors), ring motifs, and hydrogen-bonded chains on the H...O=C angle averages of N-H...O=C hydrogen-bonds, *trans*-amide...*trans*-amide hydrogen-bonds, and other structural subsets, are shown.

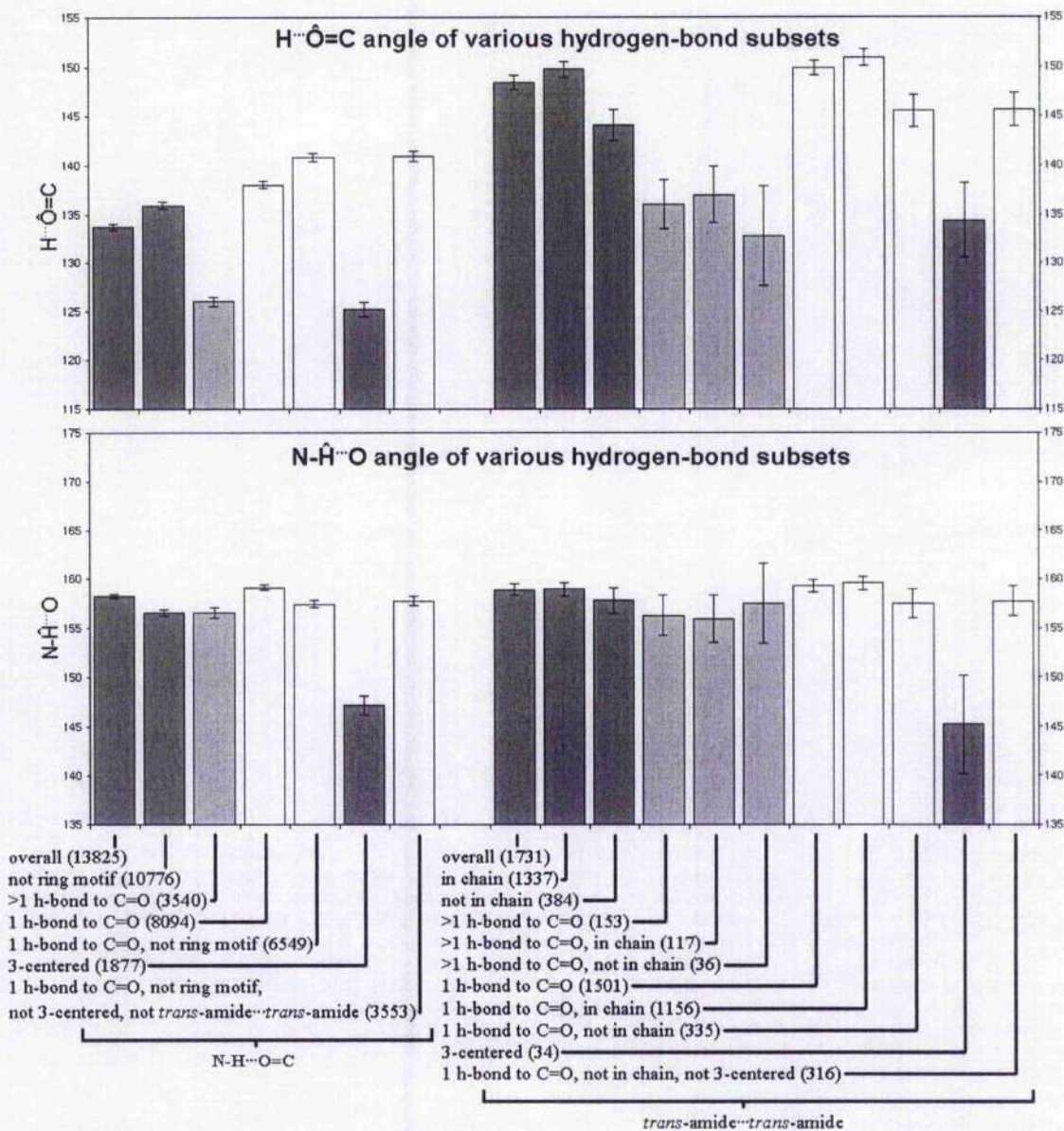
Multiple hydrogen-bonding and occurrence within (hydrogen-bonded) chains both have an effect on the average H...O=C angle of *trans*-amide...*trans*-amide hydrogen-bonds. This is most marked with multiple hydrogen-bonding, where a reduction of around 14° is observed when comparing single to multiple hydrogen-bonding. This is observed independently of whether the occurrence is within, or outside of, chains. Occurrences within chains show an increase of around 5° compared with those outside of chains, independently of multiple or single hydrogen-bonding. Highest values (H...O=C >150°) are observed when the *trans*-amide is both within a chain and accepts only a single hydrogen bond. In the small number of occurrences where the *trans*-amide is not within a chain and accepts multiple hydrogen-bonds, the average H...O=C value is 133°.

For N-H...O=C hydrogen-bonds the average H...O=C angle is affected by both multiple hydrogen-bonding and ring motifs. When multiple hydrogen-bonds are excluded, the average H...O=C angle rises to 138° from 126°; this 12° difference is similar to the 14° seen for *trans*-amide...*trans*-amide hydrogen-bonds. When ring motifs are excluded, this rises further, to 141°. Three-centered hydrogen-bonds, which occur in significant numbers in the N-H...O=C, but not the *trans*-amide...*trans*-amide, hydrogen bonds, are associated with low average H...O=C angles, although excluding them has little effect compared to the other factors.

The average value of H...O=C of the *trans*-amide...*trans*-amide hydrogen-bond (149°) is 15° higher than that of the more general N-H...O=C hydrogen-bond (134°). Some of this difference is accounted for by the observation that a greater proportion of N-H...O=C

(26%) than *trans*-amide...*trans*-amide (9%) hydrogen-bonds exist in multiply hydrogen-bonded situations. Independently of multiple hydrogen-bonding (i.e. in comparison both of cases with multiple hydrogen-bonds, and of cases without multiple hydrogen-bonds) the average value of  $H\cdots O=C$  in *trans*-amide...*trans*-amide hydrogen-bonds is about  $12^\circ$  higher than in  $N-H\cdots O=C$  hydrogen-bonds. Hydrogen-bonded chains in the *trans*-amide...*trans*-amide, and ring motifs in the  $N-H\cdots O=C$ , hydrogen-bonds both contribute to this difference. However, in the absence of multiple hydrogen-bonds, hydrogen-bonded chains, and ring motifs, a  $5^\circ$  difference in the average  $H\cdots O=C$  angle remains.

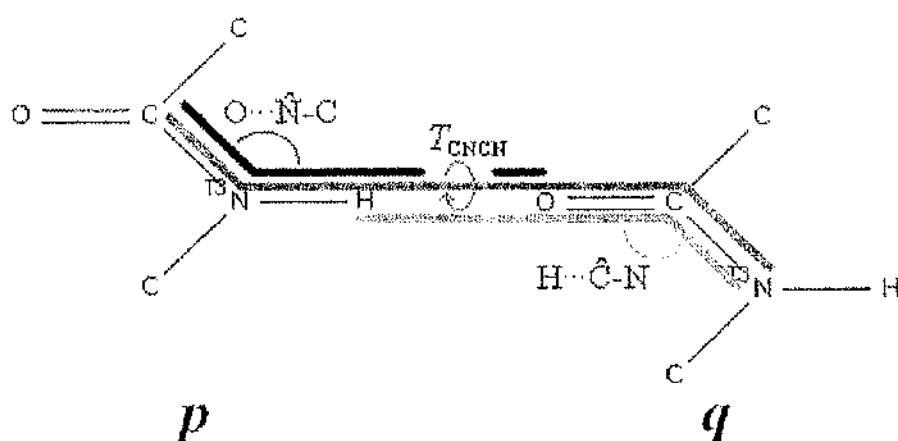
The average  $N-H\cdots O$  angle is consistently at values around  $158^\circ$  across subsets, with the only exception being the 3-centered hydrogen-bonds. Values as high as  $158^\circ$  are consistent with a preference for linearity, as discussed in relation to Figures 3.1c and 3.1f.



**Figure 3.3.** H...O=C and N-H...O angles of various hydrogen bond subsets. Average values are shown, with 95% confidence intervals. Light grey bars represent values for occurrences with multiple hydrogen bonds (where the O=C accepts more than 1 donor). Clear bars represent values for hydrogen-bonds from the subsets of structures containing only occurrences with single hydrogen-bonds. The annotations 'not ring motif', 'not *trans-amide*...*trans-amide*', 'not in chain', and 'not 3-centered' indicate hydrogen-bonds that occur within the subset of structures that do not contain these features; 'in chain' refers to hydrogen-bonded chains. The number of occurrences in each subset are shown in parentheses.

### 3.3.8 Other Aspects of inter-amide geometry

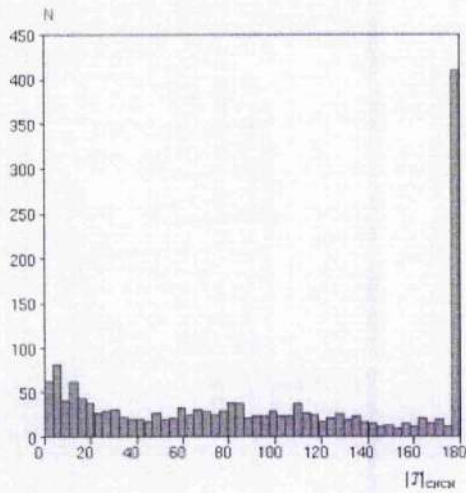
The geometry of the two *trans*-amide groups (*p* and *q* in Figure 3.4), relative to each other, in the hydrogen-bonded system (where the N-H of *p* is hydrogen-bonded to the O=C of *q*) was also considered. In addition to the hydrogen-bond parameters already described, three additional parameters were measured (indicated in Figure 3.4), the torsion angle  $T_{\text{CNCN}}$  describing the rotation of the C-N bond of *p* relative to the C-N bond of *q*, the angle  $\text{H}\cdots\text{C}-\text{N}$  made by the hydrogen atom of *p* with the C-N bond of *q*, and the angle  $\text{O}\cdots\text{N}-\text{C}$  made by the oxygen acceptor atom of *q* with the N-C bond of *p*.



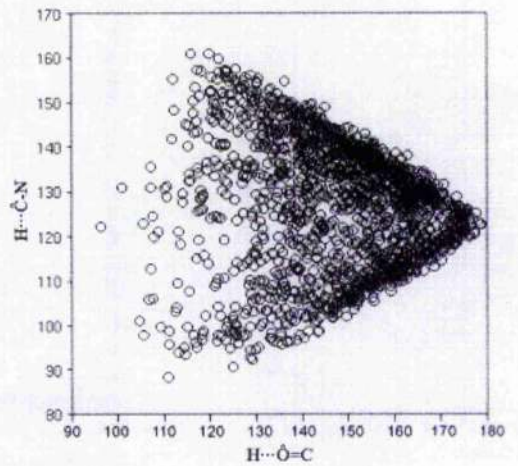
**Figure 3.4.** Parameters describing the orientation of the *trans*-amide...*trans*-amide hydrogen-bonded system. The two *trans*-amide groups are labelled *p* and *q*. The torsion angle  $T_{\text{CNCN}}$ , the angle  $\text{H}\cdots\text{C}-\text{N}$ , and the angle  $\text{O}\cdots\text{N}-\text{C}$  are indicated.

$T_{\text{CNCN}}$  provides an indication of the system's deviation from planarity, and also of whether the C-N bonds are considered *syn* or *anti* relative to each other about the hydrogen bond. The co-planar arrangement in Figure 3.4 has the *anti* configuration, with  $T_{\text{CNCN}}$  of  $180^\circ$ . The distribution of the modulus of  $T_{\text{CNCN}}$  is shown (Figure 3.5a). Allowing a  $20^\circ$  deviation from planarity, 28% of the *trans*-amide...*trans*-amide hydrogen-bonds had *anti*  $T_{\text{CNCN}}$ , and 18% were *syn*. The sharp peak at  $180^\circ$  occurs because many of these  $T_{\text{CNCN}}$  values are precisely  $180^\circ$  due to being interactions between symmetry-related molecules from different unit cells with an  $\cdots(\cdot a)\cdots$  arrangement; they make up 58% of the *anti* hydrogen-bonds.

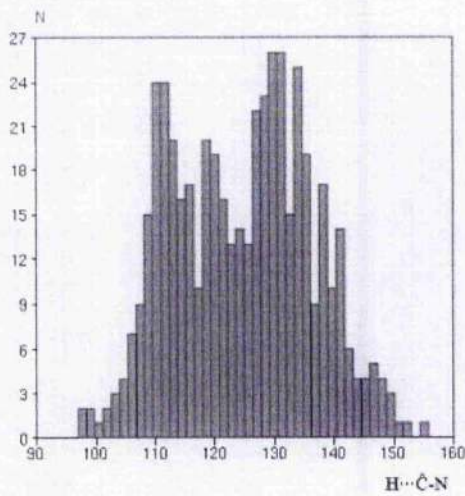




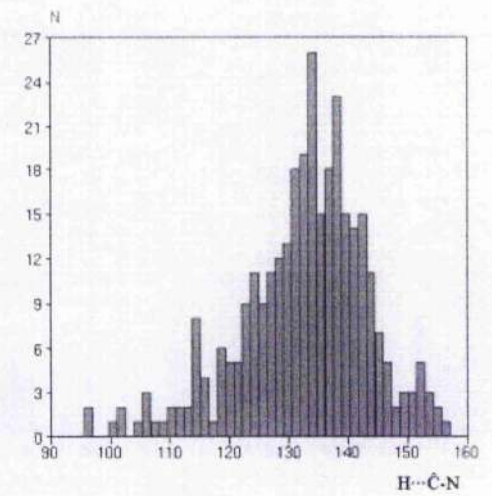
**a**



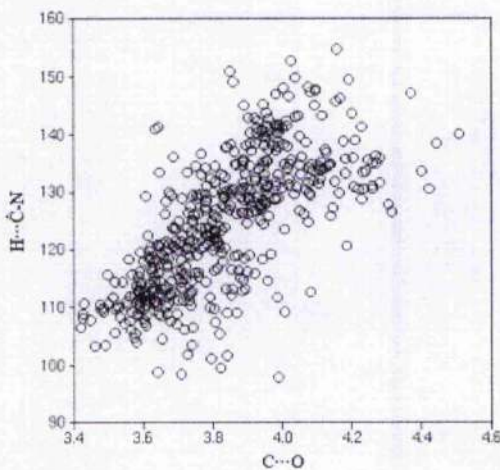
**b**



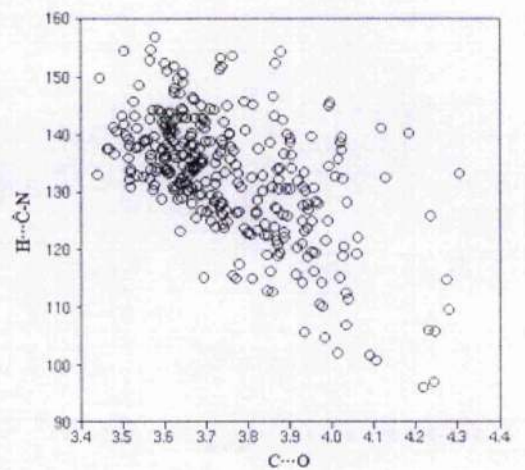
**c**



**d**

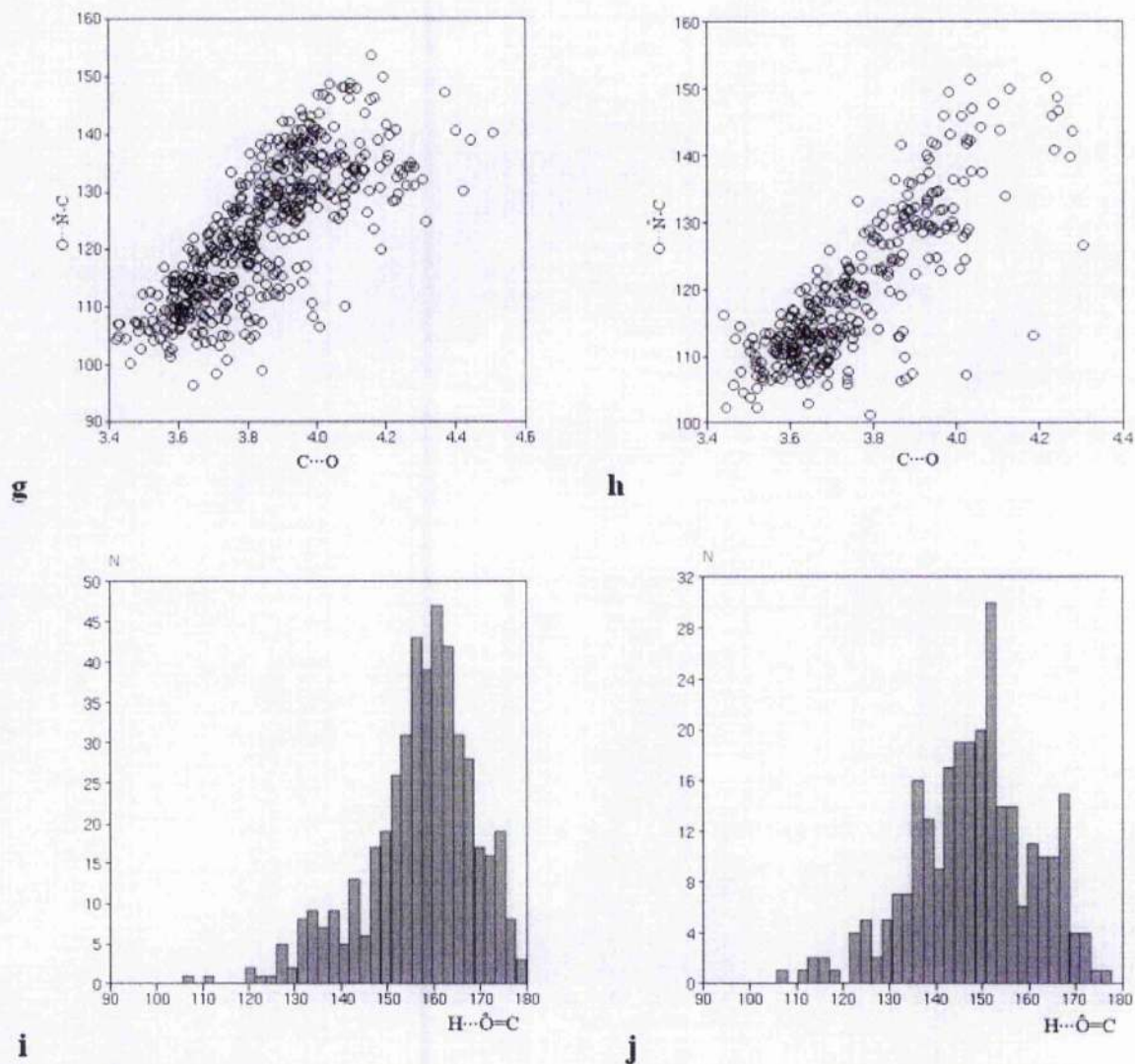


**e**



**f**

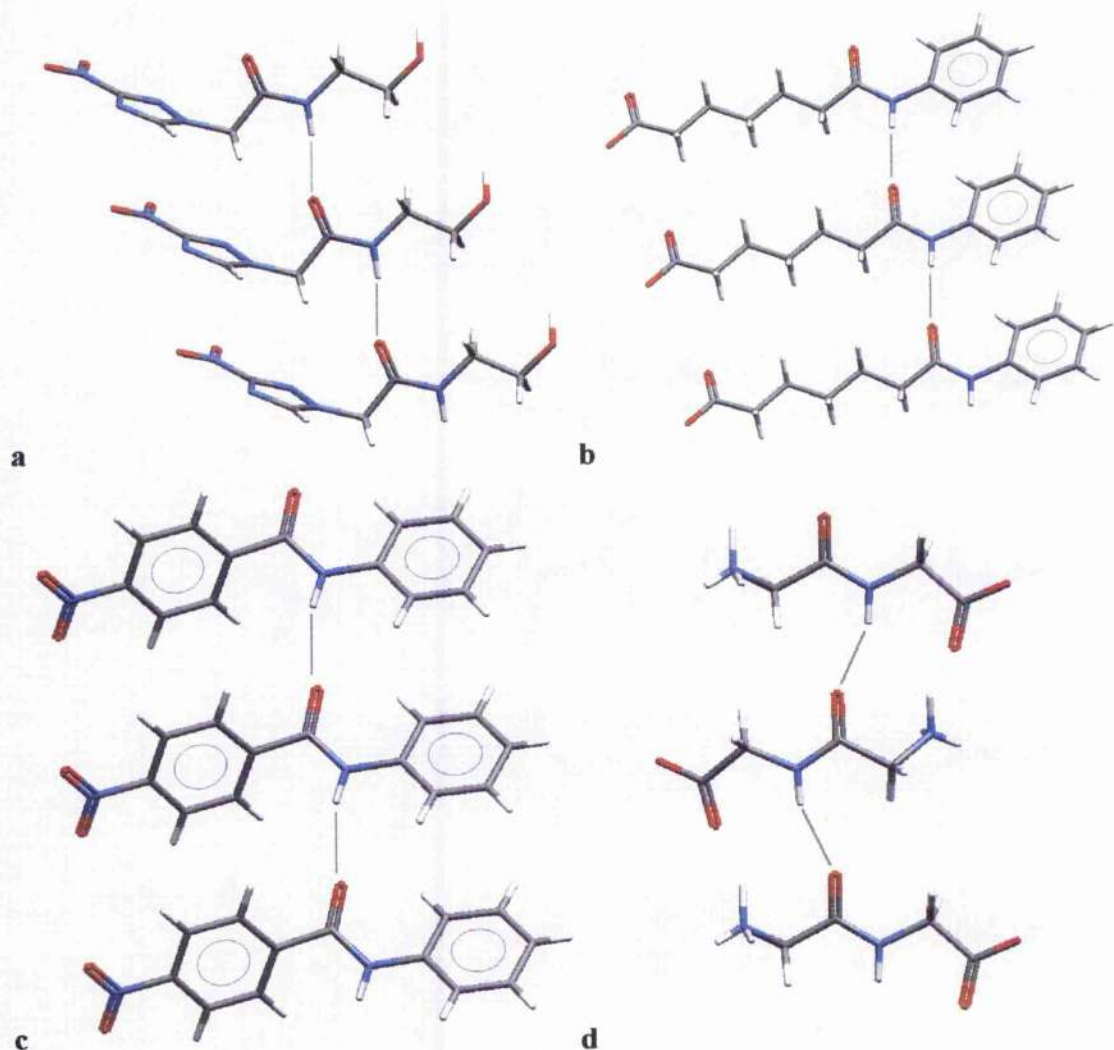
Figure continued...



**Figure 3.5.** Geometry of *trans*-amide...*trans*-amide hydrogen-bonded system. (a) distribution of the torsion angle  $T_{\text{CNCN}}$ , and (b) plot of  $\text{H}\cdots\text{C}-\text{N}$  against  $\text{H}\cdots\text{O}=\text{C}$ , for all occurrences. (c-j) various plots and distributions for *anti*  $T_{\text{CNCN}}$  on the left (c, e, g, and i) and *syn*  $T_{\text{CNCN}}$  on the right (d, f, h, and j). (c) and (d) distributions of  $\text{H}\cdots\text{C}-\text{N}$ . (e) and (f) plots of  $\text{H}\cdots\text{C}-\text{N}$  against  $\text{C}\cdots\text{O}$  distance. (g) and (h) plots of  $\text{O}\cdots\text{N}-\text{C}$  against  $\text{C}\cdots\text{O}$  distance. (i) and (j) distributions of  $\text{H}\cdots\text{O}=\text{C}$ .

The H...C-N (and also the O...N-C) angle indicates the deviation of the hydrogen-bond from linearity. In the planar, or near planar, arrangement, the H...C-N angle gives the direction of the deviation from linearity, as movement across the plane gives values either greater or less than 120°. In this context it is a more useful measure than H...O=C or N...O=C because, for these parameters, values of less than 180° are ambiguous in terms of the direction of deviation from linearity across the plane. It can be seen (Figure 3.5b) that, starting from a linear arrangement with H...C-N of 120° and H...O=C of 180°, a change of 10° in H...C-N approximately correlates with a 20° change in H...O=C. In the planar, or near planar, arrangement, rotation of *trans*-amide  $\phi$  about an axis perpendicular to the plane, relative to  $p$ , also affects H...C-N. The distribution of H...C-N for the *anti*  $T_{\text{CNCN}}$  occurrences was multimodal with one major peak above 120°, one below 120°, and a less pronounced peak at around 120° (Figure 3.5c). The *syn*  $T_{\text{CNCN}}$  occurrences had a normal distribution of H...C-N (Figure 3.7d) with an average of 133° (sd 10.8°). Typical examples of the three *anti*  $T_{\text{CNCN}}$  orientations, and of *syn*  $T_{\text{CNCN}}$  are seen in Figure 3.6.





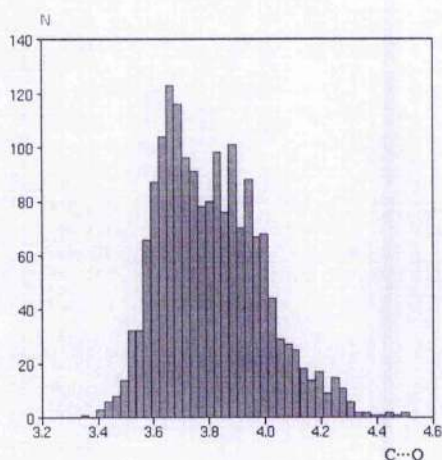
**Figure 3.6.** Examples of chains of *trans*-amide...*trans*-amides with different orientations. (a-c) examples of *anti*  $T_{CNCN}$ : (a)  $H \cdots C-N$  angle below  $120^\circ$  (CSD refcode: VACROG); (b)  $H \cdots C-N$  angle near  $120^\circ$  (refcode: HEVJUN); (c)  $H \cdots C-N$  angle above  $120^\circ$  (refcode: ACALAR). (d) example of *syn*  $T_{CNCN}$  (refcode: GLYGLY). Hydrogen bonds are indicated.



Both the H $\cdots$ C-N and O $\cdots$ N-C angles correlate with C $\cdots$ O, the distance from the carbonyl carbon of *p* to the oxygen of *q*. In the case of H $\cdots$ C-N, this is a positive correlation for *anti*  $T_{\text{CNCN}}$  (Figure 3.5e) and a negative correlation for *syn*  $T_{\text{CNCN}}$  (Figure 3.5f). Correlation with the O $\cdots$ N-C angle is tighter and is positive for both *anti*  $T_{\text{CNCN}}$  (Figure 3.5g) and *syn*  $T_{\text{CNCN}}$  (Figure 3.5h). The high values of H $\cdots$ C-N and low values of O $\cdots$ N-C observed for *syn*  $T_{\text{CNCN}}$  correlate with a preference for low C $\cdots$ O distances, giving an average C $\cdots$ O of 3.76 Å (sd 0.2), suggesting a preference for geometry that maximises the strength of the favourable electrostatic interaction between the two carbonyl groups.

The distributions of H $\cdots$ O=C for *anti*  $T_{\text{CNCN}}$  (Figure 3.5i) and for *syn*  $T_{\text{CNCN}}$  (Figure 3.5j) are also shown. Both cases have the high H $\cdots$ O=C angles observed for all *trans*-amide $\cdots$ *trans*-amide hydrogen-bonds, however the average for *anti*  $T_{\text{CNCN}}$  is higher (156.55°; sd 13.1) than that observed for *syn*  $T_{\text{CNCN}}$  (148.72°; sd 13.4).

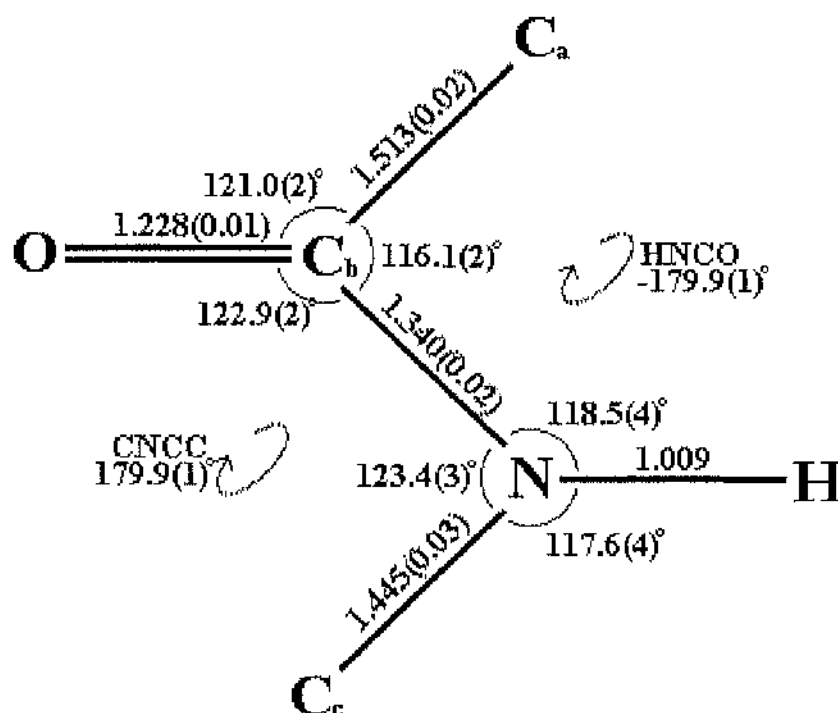
Overall C $\cdots$ O distances for the *trans*-amide $\cdots$ *trans*-amide hydrogen-bonds average 3.81 Å (sd 0.2), although the distribution is multimodal (Figure 3.7). The majority (1492; 86%) are below 4 Å, and so meet the criteria used in chapters 5 and 6 to identify carbonyl-carbonyl interactions between *trans*-amides.



**Figure 3.7.** Distribution of C $\cdots$ O distance for *trans*-amide $\cdots$ *trans*-amide hydrogen-bonded systems.

### 3.3.9 Internal geometry of the *trans*-amide group

The average internal geometries of *trans*-amides in the CSD are given in Table 3.6, along with values for the subset that make *trans*-amide...*trans*-amide hydrogen-bonds, and for those that do not. Values for the full set are shown schematically in Figure 3.8. The internal geometry of structures in the hydrogen-bonded subset was not found to differ significantly from the full set of *trans*-amides, or from *trans*-amides that occurred in the subset of structures that did not contain *trans*-amide...*trans*-amide hydrogen-bonds. The distributions of all internal parameters were uniform except for two: the N-C<sub>c</sub> distance, and the C<sub>b</sub>-N-C<sub>c</sub> angle. These had bimodal distributions that, when plotted against each other, show distribution of data points in two nearly distinct clusters (Figure 3.9). These clusters represent two distinct internal geometries, possibly resulting from differences in the effects of aromatic or aliphatic substituents on the carbon atom (C<sub>c</sub>).

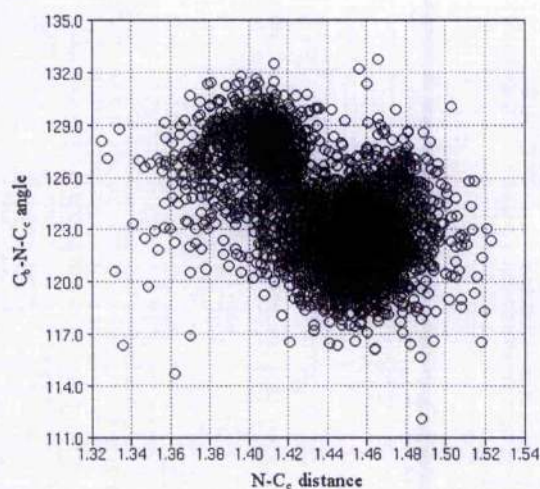


**Figure 3.8.** Internal geometry of the *trans*-amide group. Average values for bond lengths (Å), angles, and torsion angles are shown from 5343 structurally distinct occurrences within the CSD. The standard deviation, to one significant figure, of each parameter is given in brackets, except for the two torsion angles where the R bar value (more suitable for circular data; see Allen & Johnson, 1991) is given instead. The N-H bond length is a default normalisation value (see Materials and methods). The carbon atoms are labelled a, b, and c, for ease of reference.



**Table 3.6.** Internal geometry of the *trans* amide group. Mean values and standard deviations are given for each parameter for the full set of *trans*-amides, the hydrogen-bonded subset, and the non-hydrogen-bonded subset, except for torsion angles (where the R bar value is given instead). Subset size is the number of structures that contain one or more occurrence(s) from that subset. *N* is the total number of *trans*-amide groups that occur within the given subset.

Parameter	Full set of <i>trans</i> -amides		Hydrogen-bonded		Non-hydrogen-bonded	
	Mean	sd	Mean	sd	Mean	sd
bond lengths						
N-H	1.009	N/A	1.009	N/A	1.009	N/A
N-C <sub>b</sub>	1.340	0.02	1.337	0.02	1.343	0.02
N-C <sub>c</sub>	1.445	0.03	1.447	0.02	1.441	0.03
C <sub>b</sub> =O	1.228	0.01	1.229	0.01	1.227	0.01
C <sub>b</sub> -C <sub>c</sub>	1.513	0.02	1.510	0.02	1.512	0.02
angles						
H-N-C <sub>b</sub>	118.5	4	118.6	4	118.2	4
H-N-C <sub>c</sub>	117.6	4	117.8	4	117.3	4
C <sub>b</sub> -N-C <sub>c</sub>	123.4	3	123.1	2	123.8	3
N-C <sub>b</sub> -C <sub>a</sub>	116.1	2	115.9	1	116.0	2
N-C <sub>b</sub> =O	122.9	2	122.8	1	123.0	2
C <sub>a</sub> -C <sub>b</sub> =O	121.0	2	121.3	1	121.0	2
torsion angles						
C <sub>c</sub> NC <sub>b</sub> C <sub>a</sub>	179.9	1	179.7	1	179.8	1
HNC <sub>b</sub> O	-179.9	1	180.0	1	180.0	1
Subset size	3077		1245		1832	
<i>N</i>	5343		2097		2755	



**Figure 3.9.** Plot of C<sub>b</sub>-N-C<sub>c</sub> angle against N...C<sub>c</sub> distance of *trans*-amide groups.

## 4 Hydrogen-bonding between *trans*-amide groups: steric accessibility

### 4.1 Brief introduction and outline

Comparison of the findings of previous studies shows that the geometry of *trans*-amide...*trans*-amide hydrogen-bonds differs from that of the general N-H...O=C case in that the angle H...O=C made by the hydrogen with the accepting carbonyl group is more linear, averaging 148°, compared with an average of 134° in the general case. The previous chapter elucidated three factors that together account for most of the difference in average H...O=C: a smaller proportion of *trans*-amide carbonyl oxygens accept more than one hydrogen-bond; general N-H...O=C bonds often occur in ring motifs with relatively constrained geometry and low H...O=C values (these cannot be formed by *trans*-amides); and chains of hydrogen-bonds between *trans*-amides, with high H...O=C values, often extend throughout the crystal lattice. These factors were shown to be of similar importance. It is possible that the steric accessibility (SA) of the carbonyl oxygen atom may account for the remaining difference in the average H...O=C values.

This chapter is an investigation of the SA of the carbonyl oxygen atom of the N-H...O=C hydrogen-bond, and of the subset of N-H...O=C hydrogen-bonds of the type *trans*-amide...*trans*-amide. It centers on a comparison of the *trans*-amide...*trans*-amide subset with the general case, in an effort to determine the degree to which SA has influenced the average H...O=C angle. This work is dealt with separately because it involves the comparison only of certain selected subsets from chapter 3, and the creation of two new subsets, as explained in the materials and methods.

## 4.2 Materials and methods

### 4.2.1 Identification and analysis of hydrogen-bonds

This work is a continuation from chapter 3, and uses a number of the subsets of different hydrogen-bonded interactions identified in that study. The definition of a *trans*-amide, the hydrogen bond definition, and the methodology for geometric analysis of the hydrogen-bonds is also shared.

Conquest was used to create new subsets from the subset of N-H...O=C hydrogen-bonds that were not *trans*-amide...*trans*-amide hydrogen-bonds, and from the subset of *trans*-amide...*trans*-amide hydrogen-bonds. Certain subsets characterised in chapter 3 were chosen for this analysis. They were selected to reflect the three factors that were shown to affect the H...O=C angle: the number of hydrogen-bonds accepted by the carbonyl oxygen atom; the occurrence of 'ring' motifs among the general N-H...O=C set; the occurrence of hydrogen-bonded chains extending through the crystal lattice among the *trans*-amide...*trans*-amide subset. Five subsets were chosen for both the N-H...O=C and the *trans*-amide...*trans*-amide subsets: (1) those where the carbonyl oxygen accepted only a single hydrogen-bond; (2) those accepting multiple (two or more) hydrogen-bonds; (3) those occurring in 'ring motifs' (a subset of the general N-H...O=C subset), or in hydrogen-bonded chains extending through the crystal lattice (a subset of the *trans*-amide...*trans*-amide subset); (4) those not occurring in ring motifs or hydrogen-bonded chains; (5) those that had only a single hydrogen-bond and were neither in ring motifs nor hydrogen-bonded chains. Creation of the fifth subsets allowed comparison of N-H...O=C and *trans*-amide...*trans*-amide hydrogen-bonds, while accounting for the three previously identified factors that influence the H...O=C angle.

Throughout the text and Figures 'N-H...O=C' refers to the set of hydrogen-bonds of the type N-H...O=C (excluding *trans*-amide...*trans*-amide hydrogen-bonds), and '*trans*-amide...*trans*-amide' to the set of hydrogen-bonds that occur between *trans*-amide groups. These two subsets are further categorised into smaller subsets. The notation used

to refer to these is as follows: '>1 h-bond' denotes subsets where the carbonyl oxygen accepts multiple hydrogen-bonds; '1 h-bond' denotes subsets where the carbonyl oxygen accepts only a single hydrogen-bond; 'in ring' denotes N-H...O=C hydrogen-bonds within ring motifs; 'in chain' denotes *trans*-amide...*trans*-amide hydrogen-bonds within infinitely repeating hydrogen-bonded chains.

## 4.2.2 Calculation of steric accessibility (SA)

We wished to calculate the steric accessibility (SA) of the carbonyl oxygen atom from each hydrogen-bond, in the absence of the hydrogen-bonded N-H group and any other inter-molecular contact. The VATM feature of the program RPluto (Infantes & Motherwell, 2004) was used. This takes a probe of a given radius (in this case, that of a hydrogen atom, 1.2Å), and places it at random on the van der Waals surface of the atom of interest (in this case, the carbonyl oxygen). This random probe point placement is repeated a large number of times (in this case, 6000) to cover the entire van der Waals surface and ensure statistical significance. The SA is calculated as the fraction of probe points that do not intersect with another atom from that molecule. Atoms from neighbouring crystal-packed molecules are ignored, so the SA value is a measure of accessibility to the 'free' molecule, independent of packing. It is estimated that the standard deviation of SA for a sample of 6000 point placements is 0.005.

SA values from RPluto were integrated into data from Conquest describing hydrogen-bond geometry. This facilitated statistical analysis using the Vista component of the Conquest package. Geometric parameters such as the H...O=C angle could then be plotted directly against SA values, and comparisons made between the different subsets.

A software limitation of RPluto prevented SA calculation for oxygen atoms within crystal structures having more than 200 atoms per unit cell. This applied to a small number of structures (60 structures from the N-H...O=C subset, and 19 from the *trans*-amide...*trans*-amide subset), and explains why the numbers of structures and hydrogen-bond occurrences for each subset given here differ from those reported in Chapter 3.

### 4.3 Results & Discussion

The average  $\text{H}\cdots\text{O}=\text{C}$  angle for each subset is given in Table 4.1. The full set of  $\text{N-H}\cdots\text{O}=\text{C}$  hydrogen-bonds was separated into two subsets: those structures that contained hydrogen-bonds of the type *trans*-amide $\cdots$ *trans*-amide and those that did not. ' $\text{N-H}\cdots\text{O}=\text{C}$ ' is used throughout the text to refer to the non-*trans*-amide $\cdots$ *trans*-amide subset rather than the full set of  $\text{N-H}\cdots\text{O}=\text{C}$  hydrogen-bonds. *Trans*-amide $\cdots$ *trans*-amide hydrogen-bonds are  $17^\circ$  more linear at the oxygen atom than  $\text{N-H}\cdots\text{O}=\text{C}$  hydrogen-bonds. In cases where the carbonyl oxygen accepts only a single hydrogen-bond and where the hydrogen-bond does not occur in a ring motif or in a hydrogen-bonded chain, the difference between *trans*-amide $\cdots$ *trans*-amide and  $\text{N-H}\cdots\text{O}=\text{C}$  drops to  $7^\circ$ .

The steric accessibility (SA) of the carbonyl oxygen atom of each hydrogen-bond was measured. The distributions (Figure 4.1) and averages (Table 4.1) of SA values for the subsets are shown. A direct comparison of average  $\text{H}\cdots\text{O}=\text{C}$  angle with average SA values is provided in Figure 4.2. The average SA for the  $\text{N-H}\cdots\text{O}=\text{C}$  case was higher than that for *trans*-amide $\cdots$ *trans*-amide. This correlates with the difference in the average  $\text{H}\cdots\text{O}=\text{C}$  angle. The greater SA of the  $\text{N-H}\cdots\text{O}=\text{C}$  oxygen atom seems to allow geometries with lower  $\text{H}\cdots\text{O}=\text{C}$  values to be adopted. The distributions of SA for all subsets are unimodal (except for  $\text{N-H}\cdots\text{O}=\text{C}$  bonds that are not in ring motifs, where a bimodal distribution is observed) and tend to be skewed towards higher SA values.

There is a correlation of SA with the number of hydrogen-bonds accepted by the carbonyl oxygen. Carbonyl oxygens that accept more than one hydrogen-bond are more accessible than those accepting only one. This also correlates with the  $\text{H}\cdots\text{O}=\text{C}$  angle, where less linear geometry is observed in cases of multiple hydrogen-bonding. Greater accessibility favours multiple hydrogen-bonding.

Carbonyl oxygens of  $\text{N-H}\cdots\text{O}=\text{C}$  hydrogen-bonds that are involved in ring motifs are more sterically accessible than those that are not. Inflexible  $\text{H}\cdots\text{O}=\text{C}$  angles with lone-pair directionality are characteristic of the ring motif. The lack of flexibility is shown by the

low standard deviation of the average  $\text{H}\cdots\text{O}=\text{C}$  angle of ring motif  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  bonds (Table 4.1). The non-linear  $\text{H}\cdots\text{O}=\text{C}$  angle requires high steric accessibility, and this is indeed observed.

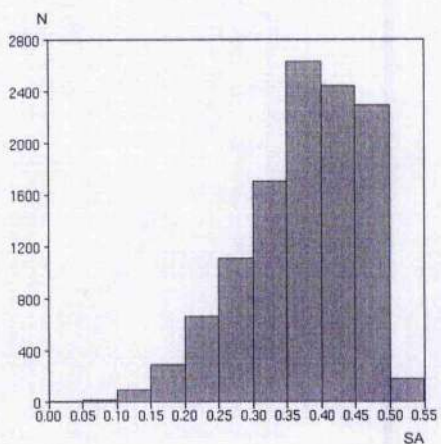
*Trans*-amide $\cdots$ *trans*-amide hydrogen-bonds within hydrogen-bonded chains have a more linear  $\text{H}\cdots\text{O}=\text{C}$  angle than those that do not, and for this reason we might have predicted lower average SA in those within chains, in keeping with the trend observed thus far in which the average  $\text{H}\cdots\text{O}=\text{C}$  correlates inversely with average SA. However, the average SA for those within hydrogen-bonded chains is higher rather than lower. Therefore, it may be supposed that some other factor, possibly packing related, acts against SA to produce more linear  $\text{H}\cdots\text{O}=\text{C}$  angles for *trans*-amide $\cdots$ *trans*-amide hydrogen-bonds within hydrogen-bonded chains.



**Table 4.1.** Average  $H\cdots O=C$  angles and carbonyl oxygen atom steric accessibility values (SA) of  $N-H\cdots O=C$  and *trans*-amide $\cdots$ *trans*-amide hydrogen-bonds. Overall values for the two sets are given, together with values for five subsets of each.

Type of hydrogen bond	Subset size	N	Parameters	
			$H\cdots O=C$ (sd)	SA (sd)
N-H $\cdots$ O=C	8264	13544	134 (17)	0.355 (0.09)
(>1 h-bond)	2268	3475	126 (15)	0.404 (0.07)
(1 h-bond)	5996	7969	138 (17)	0.329 (0.09)
(in ring)	1628	1870	124 (6)	0.398 (0.07)
(not ring)	6636	10549	136 (18)	0.343 (0.09)
(1 h-bond, not ring)	4875	6432	141 (18)	0.317 (0.09)
N-H $\cdots$ O=C (not <i>trans</i> -amide $\cdots$ <i>trans</i> -amide)	7038	11450	131 (16)	0.369 (0.09)
(>1 h-bond)	2095	3240	126 (14)	0.410 (0.06)
(1 h-bond)	4943	6360	135 (17)	0.344 (0.09)
(in ring)	1620	1863	124 (6)	0.399 (0.07)
(not ring)	5418	8486	133 (17)	0.359 (0.09)
(1 h-bond, not ring)	3827	4837	138 (17)	0.332 (0.09)
<i>trans</i> -amide $\cdots$ <i>trans</i> -amide	1226	1674	148 (15)	0.274 (0.06)
(>1 h-bond)	127	150	136 (16)	0.297 (0.05)
(1 h-bond)	1099	1453	150 (15)	0.272 (0.06)
(in chain)	964	1334	150 (15)	0.280 (0.05)
(not chain)	262	335	144 (16)	0.254 (0.07)
(1 h-bond, not chain)	238	295	145 (15)	0.252 (0.07)

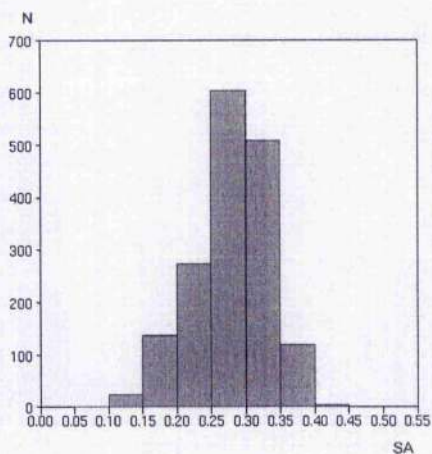
**N-H...O=C**



**a**

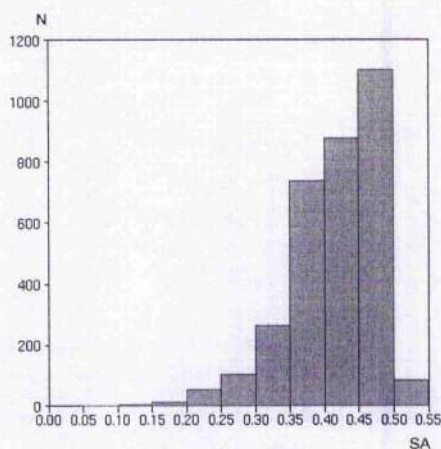
**>1 h-bond**

**trans-amide...trans-amide**



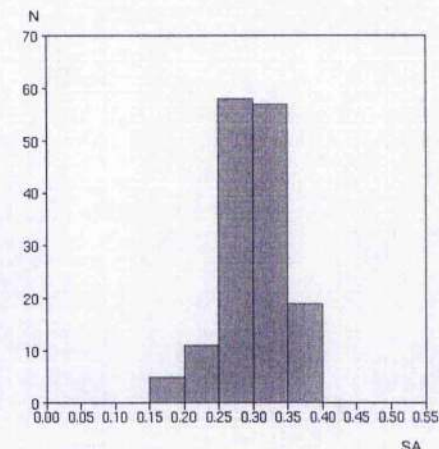
**b**

**>1 h-bond**



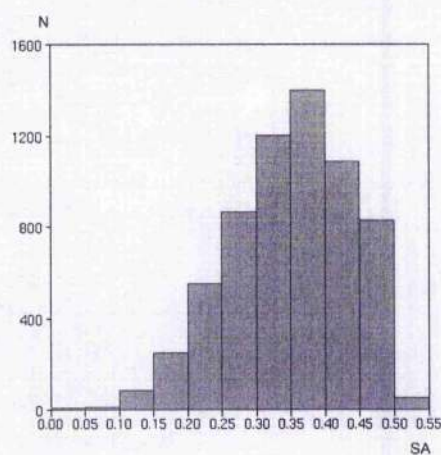
**c**

**1 h-bond**

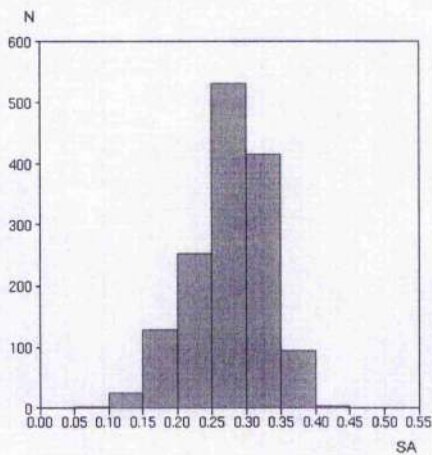


**d**

**1 h-bond**



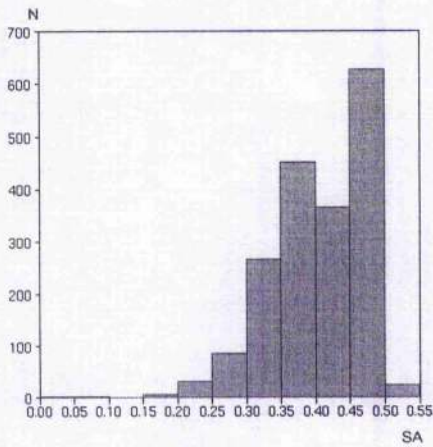
**e**



**f**

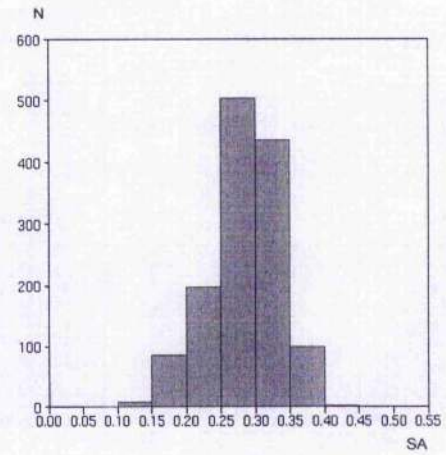
**Figure continued...**

**In ring**



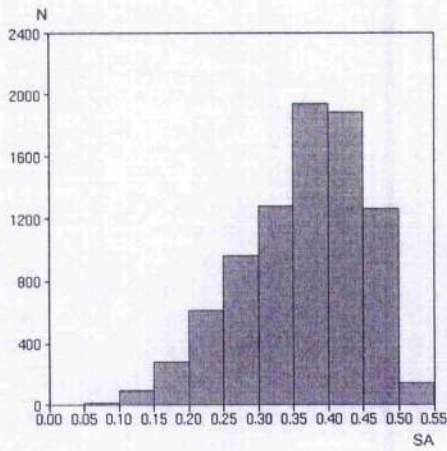
**g**

**In chain**



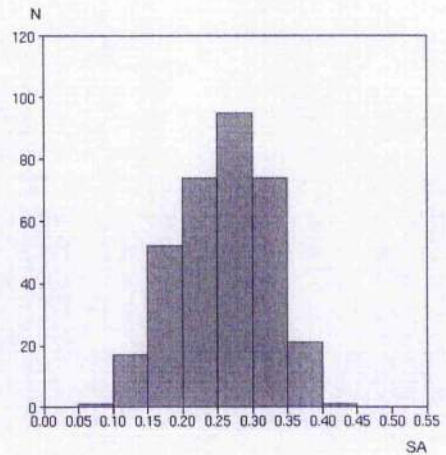
**h**

**Not ring**



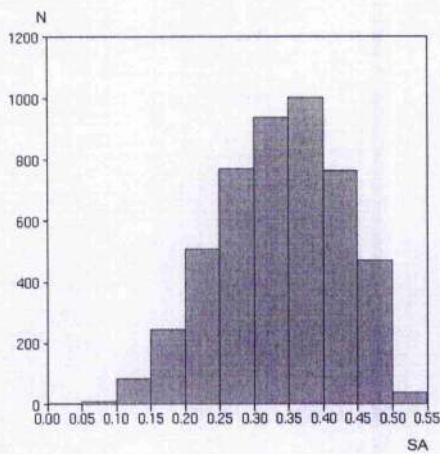
**i**

**Not chain**



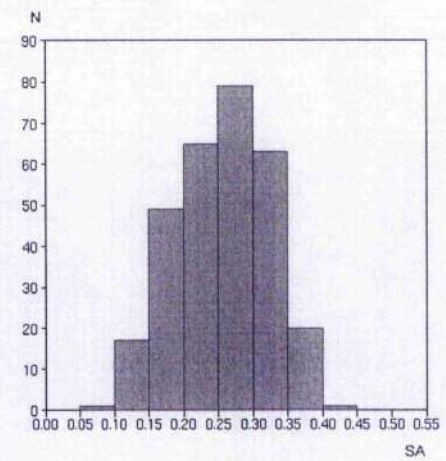
**j**

**1 h-bond, not ring**



**k**

**1 h-bond, not chain**

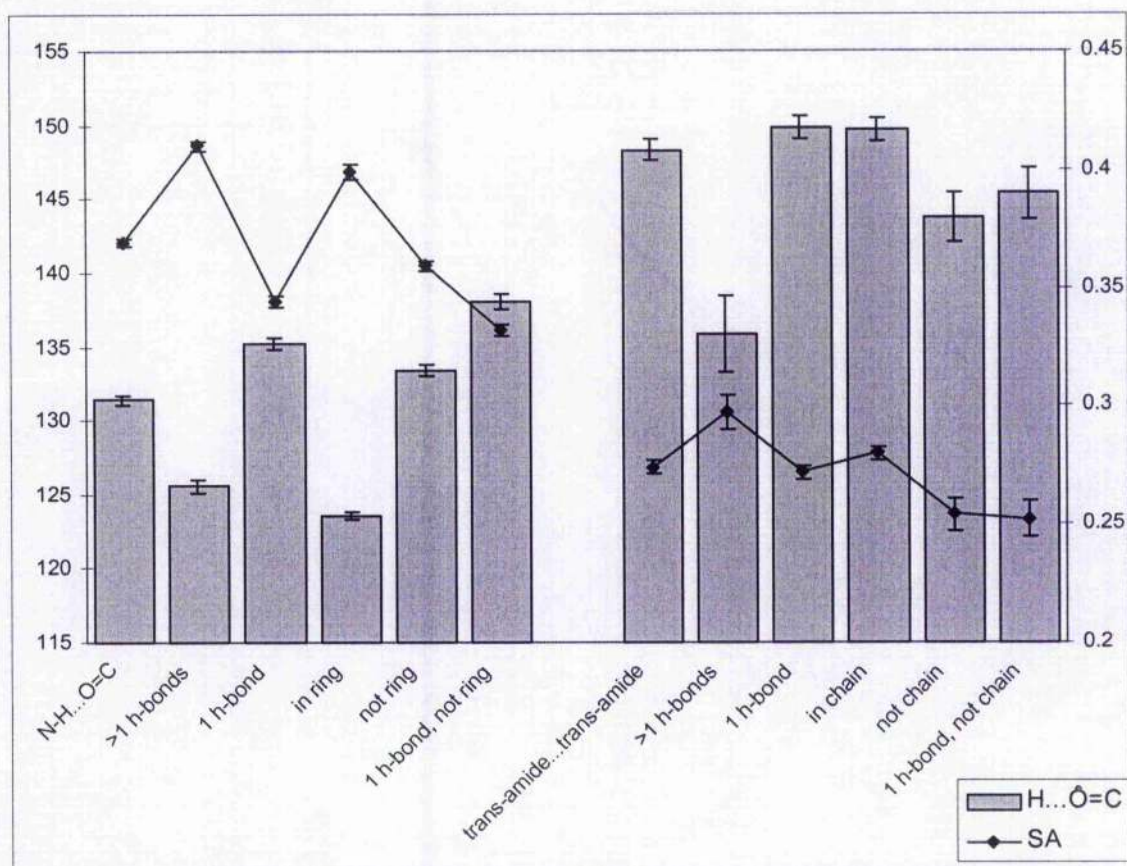


**l**

Figure continued...

**Figure 4.1.** Histograms of steric accessibility of carbonyl oxygen atoms of N-H...O=C hydrogen-bonds, excluding *trans*-amide...*trans*-amide hydrogen-bonds (left-hand column: **a, c, e, g, i, k**) and of *trans*-amide...*trans*-amide hydrogen-bonds (right-hand column: **b, d, f, h, j, l**). (**a & b**) Overall sets; (**c & d**) Cases where the carbonyl oxygen accepts multiple hydrogen-bonds; (**e & f**) Cases where the carbonyl oxygen accepts only a single hydrogen-bond; (**g & h**) Cases where the hydrogen-bond occurs within a ring or chain; (**i & j**) Cases where the hydrogen-bond does not occur in a ring or chain; (**k & l**) Cases where the carbonyl oxygen accepts only a single hydrogen-bond, and does not occur in a ring or chain.





**Figure 4.2.** Histogram of average  $H\cdots O=C$  angles of  $N-H\cdots O=C$  (excluding *trans*-amide...*trans*-amide) and *trans*-amide...*trans*-amide hydrogen-bonds and different subsets of each. Average steric accessibility values are overlaid in line plot form.  $H\cdots O=C$  values are given on the left axis, steric accessibility values on the right axis.  $N-H\cdots O=C$  hydrogen-bonds and associated subsets are grouped on the left, *trans*-amide...*trans*-amide hydrogen-bonds and associated subsets are grouped on the right. The 95% confidence intervals of the values are given.

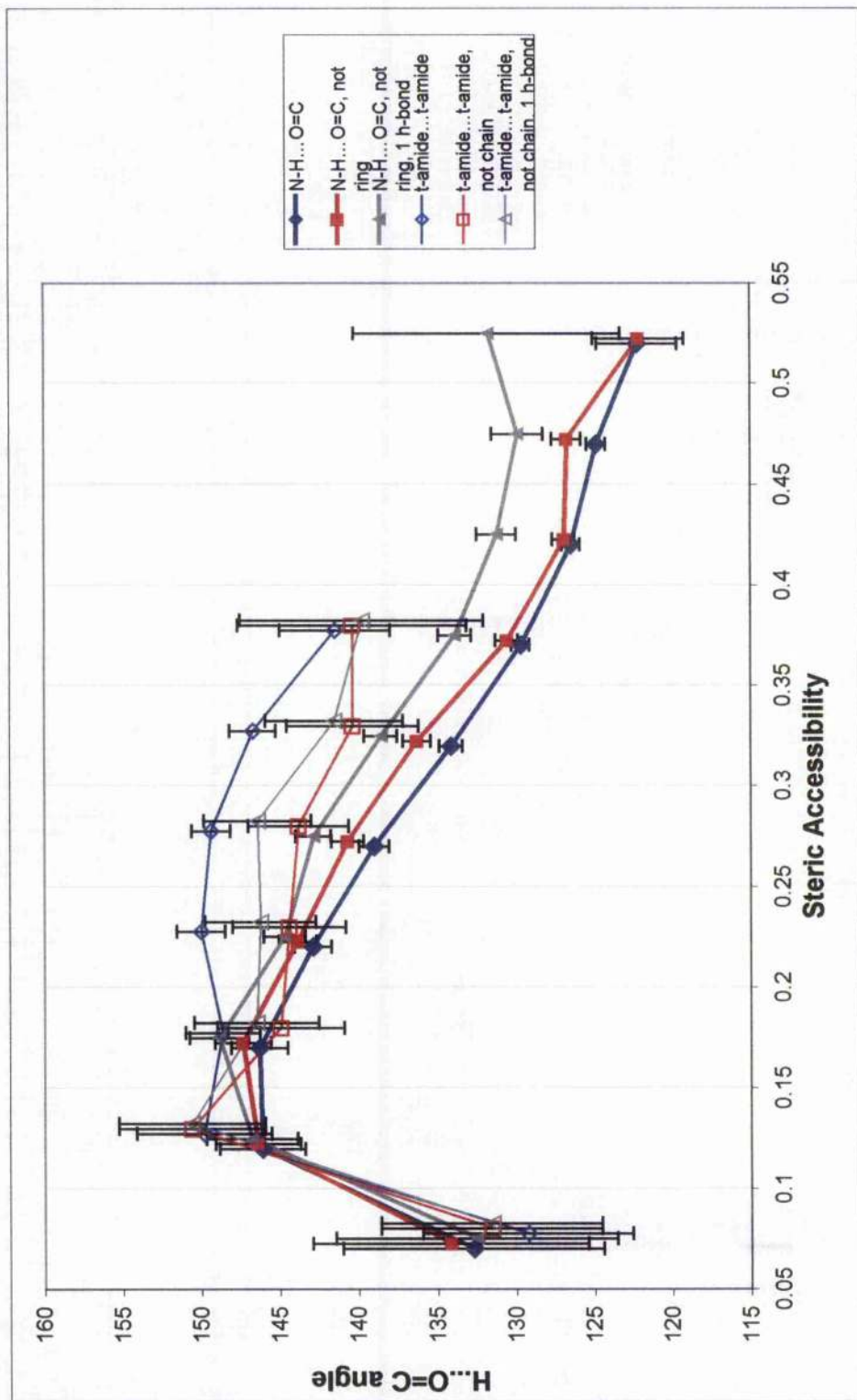
To further study the effect of SA on the  $\text{H}\cdots\text{O}=\text{C}$  angle, SA values were banded in increments of 0.05, and the average  $\text{H}\cdots\text{O}=\text{C}$  angle was calculated for each band. Comparison was made between different subsets for each SA band. This comparison is shown for the overall  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  and *trans*-amide $\cdots$ *trans*-amide cases, for the two subsets where multiple hydrogen-bonding and ring motifs have been excluded, and also for the two subsets where multiple hydrogen-bonding, ring motifs, and hydrogen-bonded chains have been excluded (Figure 4.3).

A small number of short polypeptides with intra-molecular hydrogen-bonds are present in the subsets examined. Carbonyl oxygens involved in these intra-molecular hydrogen-bonds tend to have very low accessibility for the formation of inter-molecular hydrogen-bonds. At very low values of SA (less than 0.1), a low average  $\text{H}\cdots\text{O}=\text{C}$  value is observed for all subsets. This is due to inter-molecular hydrogen-bonds to polypeptide carbonyl groups involved in intra-molecular hydrogen-bonds. These are not eliminated by the criteria for multiple hydrogen-bonding, as it specifies inter-molecular bonds. These occurrences are not considered further.

The plot (in the range of SA from 0.1 upwards) shows that for  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  hydrogen-bonds, there is a strong inverse correlation between SA and the  $\text{H}\cdots\text{O}=\text{C}$  angle. This suggests SA is an important factor in determining  $\text{H}\cdots\text{O}=\text{C}$  angles. For *trans*-amide $\cdots$ *trans*-amide hydrogen-bonds, there is also an inverse correlation, though the effect is weaker.

At the lower end of the range (from 0.1 to 0.2) the average  $\text{H}\cdots\text{O}=\text{C}$  angles are approximately the same for both  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  and *trans*-amide $\cdots$ *trans*-amide hydrogen-bonds. It seems likely that low SA is the dominant or limiting factor here, giving rise to consistently high  $\text{H}\cdots\text{O}=\text{C}$  angles. It appears that at low values SA is the only factor required to explain the difference in average  $\text{H}\cdots\text{O}=\text{C}$  between  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  and *trans*-amide $\cdots$ *trans*-amide hydrogen-bonds. In this range error bars from all the sets overlap.

At higher values of SA (greater than 0.2) the  $H\cdots O=C$  angles decrease for  $N-H\cdots O=C$ . For *trans*-amide $\cdots$ *trans*-amide the  $H\cdots O=C$  angles also decrease, but less markedly. At these higher values of SA, other factors are required to understand and explain the difference between the average  $H\cdots O=C$  angles of  $N-H\cdots O=C$  and *trans*-amide $\cdots$ *trans*-amide hydrogen-bonds. The plot suggests that the influence of SA overlaps to a large extent with the influence of ring motifs.  $H\cdots O=C$  angles for the set of  $N-H\cdots O=C$  that are not ring motifs are not very different from the full set of  $N-H\cdots O=C$ . The exclusion of hydrogen-bonded chains from the *trans*-amide $\cdots$ *trans*-amide set, and of ring motifs from the  $N-H\cdots O=C$  set, is sufficient to annul the difference in  $H\cdots O=C$  over almost the whole range of SA values. The exclusion of multiple hydrogen-bonding has the effect of a general increase in  $H\cdots O=C$  angles, both for *trans*-amide $\cdots$ *trans*-amide that are not within hydrogen-bonded chains, and for  $N-H\cdots O=C$  that are not within ring motifs.



**Figure 4.3.** H...O=C angle against steric accessibility (SA) of carbonyl oxygens of N-H...O=C (excluding *trans*-amide...*trans*-amide) hydrogen-bonds and *trans*-amide...*trans*-amide hydrogen-bonds. The SA is banded in increments of 0.05. The cut-offs of each band are indicated by vertical grey lines. The average H...O=C angle of each band (with 95% confidence intervals indicated) is given for the different subsets examined. The H...O=C angles for the different subsets at each band are horizontally staggered to allow confidence intervals to be seen.



## 5 Carbonyl-carbonyl interactions between *trans*-amide groups: in the CSD

### 5.1 Brief introduction and outline

An increased understanding of interactions between *trans*-amide carbonyls in proteins and in protein ligand-binding may be achieved by an analysis of the geometry of *trans*-amide carbonyl-carbonyl interactions in the CSD. The influence of such interactions on the crystal structures of small molecules in the CSD may to some extent be gauged by a comparison with what is observed in ketone carbonyl-carbonyl interactions in the CSD (Allen *et al.*, 1998).

In this chapter, occurrences of interactions between *trans*-amide carbonyl groups in the CSD are identified, and their geometry is characterised. The study examines the proportions of *trans*-amide interactions accounted for by each of the three motifs that were observed between ketones, and compares the geometry of ketone carbonyl-carbonyl interactions with *trans*-amide carbonyl-carbonyl interactions. It is found that, after exclusion of *trans*-amides that are hydrogen-bonded to other *trans*-amides, those remaining form carbonyl-carbonyl interactions with a propensity and geometric scatter strikingly similar to that observed for ketones.

## 5.2 Materials and methods

### 5.2.1 Database searches

Interactions between *trans*-amides were identified within the Cambridge Structural Database, as described in chapter 3 on hydrogen-bond geometry (see section 3.2.1). The definition of the *trans*-amide and the criteria for hydrogen-bonds was also as described in section 3.2.1.

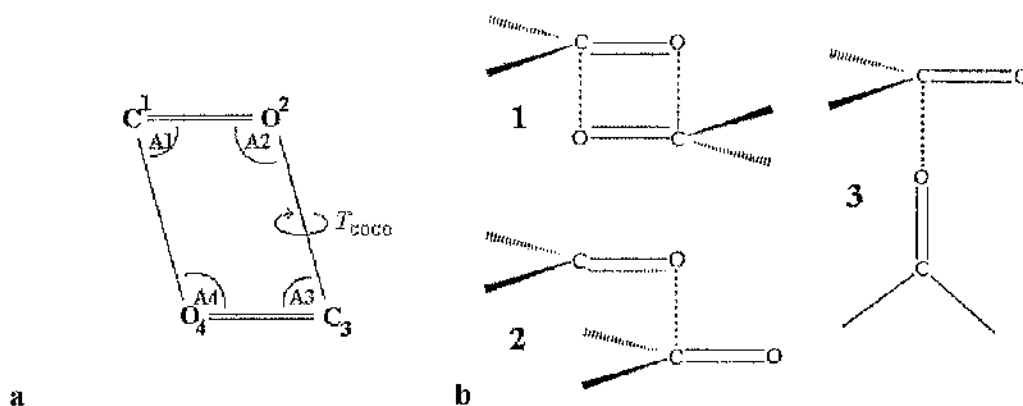
### 5.2.2 Characterisation of carbonyl-carbonyl interactions

Occurrences of inter-molecular carbonyl...carbonyl (Figure 2.1b) interactions between *trans*-amide groups were identified and characterised using Conquest. Interactions where a hydrogen-bond existed between the two *trans*-amide groups were identified initially, then excluded (Figure 2.1a).

The carbonyl...carbonyl interaction criteria required that the distance from the first to the second carbonyl group was constrained to within 4Å: at least one of C<sup>1</sup> and O<sup>2</sup> had to be within 4Å of at least one of C<sup>3</sup> and O<sup>4</sup>. For each occurrence of this, the search query specified a number of other parameters (Figure 5.1a) to be calculated by the software: (a) the torsion angle, referred to as  $T_{\text{COCO}}$ , made by the four carbonyl atoms in the order C<sup>1</sup>=O<sup>2</sup>...C<sup>3</sup>=O<sup>4</sup>; (b) the four angles, A1, A2, A3, and A4. Output data for all parameters was analysed using the Vista statistics program (Bruno *et al.*, 2002).

The carbonyl...carbonyl interactions were categorized by application of the same criteria used previously by Allen *et al.* (1998) to categorise contacts between ketone carbonyl groups into three motif types (Figure 5.1b). The criteria defining the three motif types was as follows: (**antiparallel**) the  $T_{\text{COCO}}$  torsion angle is equal to  $0^\circ \pm 20^\circ$ ; (**parallel**)  $T_{\text{COCO}}$  is equal to  $180^\circ \pm 20^\circ$ ; (**perpendicular**) the larger of the two values, A1 and A3, is greater than  $150^\circ$ .

The search query fragment is symmetrical such that angle A1 and A2 can be interchanged with A3 and A4 respectively, depending on the arbitrary atomic numeration of one carbonyl group as C<sup>1</sup> and O<sup>2</sup>, and the other as C<sup>3</sup> and O<sup>4</sup>, as applied by the Conquest search procedure. For this reason, measurements of angles A1-A4 for a given contact produce two possible sequences of values, where A1 is interchanged with A3 and A2 with A4. Where necessary, the process of categorisation and geometric analysis of interactions takes this permutational symmetry into account.

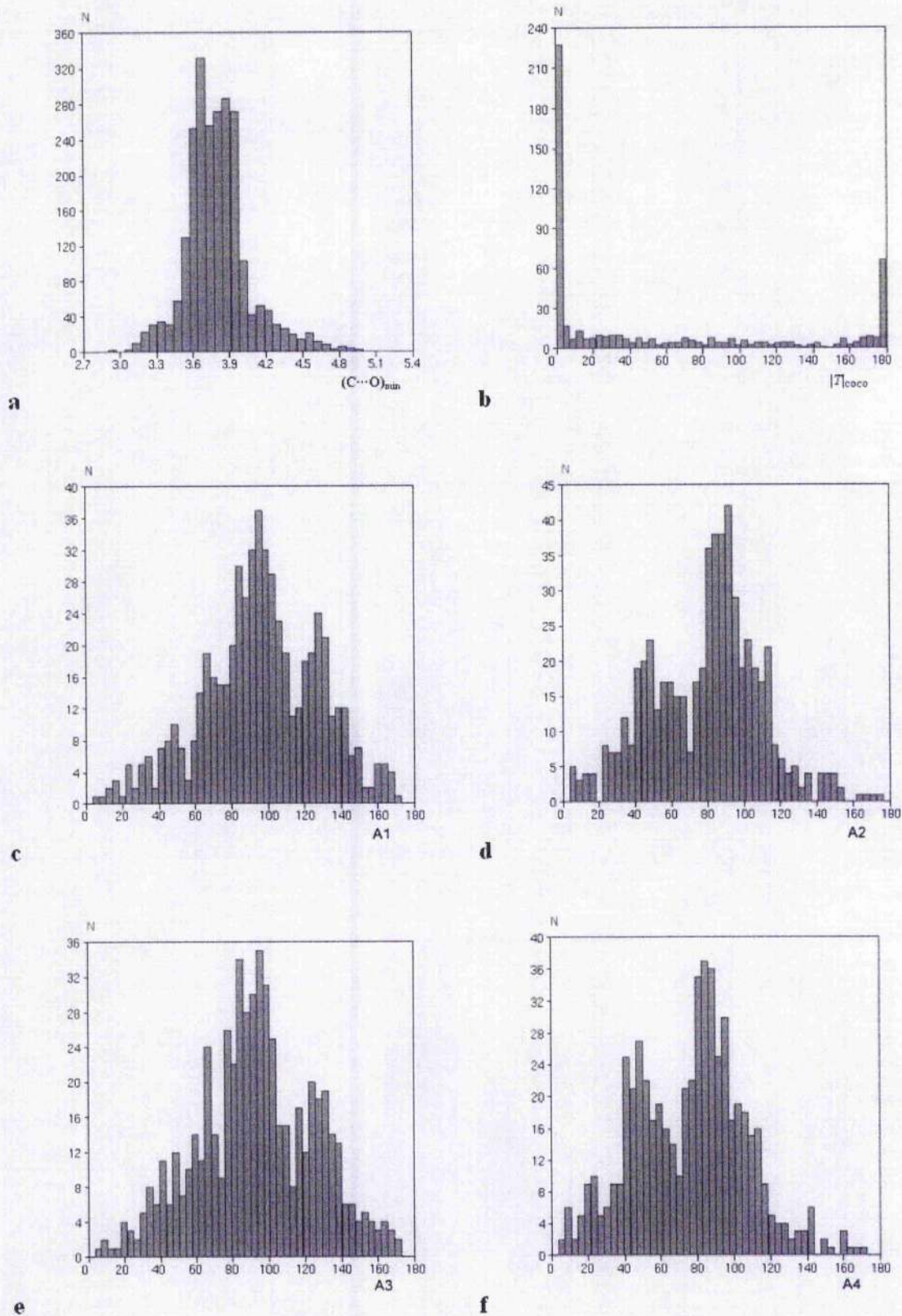


**Figure 5.1.** Carbonyl...carbonyl interaction geometry. (a) Angle and torsion angle parameters of carbonyl...carbonyl interactions. Angles A1-A4 are shown, and the torsion angle  $T_{COCO}$ . The four atoms are numbered 1-4. (b) Three motifs identified previously among contacts between ketone carbonyls: (1) Antiparallel; (2) Parallel; (3) Perpendicular.

### 5.3 Results & Discussion

Of 5343 unique *trans* amide groups in the CSD, 48% (2546) were involved in an inter-molecular carbonyl...carbonyl interaction with at least one other unit. This analysis is concerned with the characterisation of these interactions, with a view to discussion of their similarity to contacts between ketone carbonyl groups identified and characterised previously (Figure 5.1b; Allen *et al.*, 1998). Apart from prolines, all peptide bonds between the amino acids of proteins are *trans*-amides, and relevance of these results in that context will also be considered.

The total number of carbonyl-carbonyl interactions located was 2393. These comprised all occurrences where at least one of the four considered atom to atom distances (two carbon to oxygen, one carbon to carbon, and one oxygen to oxygen), was less than 4Å. Occurrences where only the carbon to carbon and/or oxygen to oxygen distance, and neither of the carbon to oxygen distances, were less than 4Å, form a small proportion of the total (14%; 331), and are unlikely to be energetically favourable due to the proximity of atoms of like charge. They have been omitted from further analysis. Of those left, a further 1492 (62%) are excluded as they are *trans*-amide...*trans*-amide hydrogen-bonds (Figure 2.1a). This leaves 570 interactions that do not form *trans*-amide...*trans*-amide hydrogen-bonds and where  $(C\cdots O)_{\min}$  (defined as the smaller of the two carbon to oxygen distances) is less than 4Å. The distribution of  $(C\cdots O)_{\min}$ , for all 2393 interactions, is shown (Figure 5.2a). The average  $(C\cdots O)_{\min}$  was 3.81Å (sd of 0.3Å). This is longer than the average distance observed in contacts between ketone carbonyl groups, which was 3.52Å (Baalham, 1996).

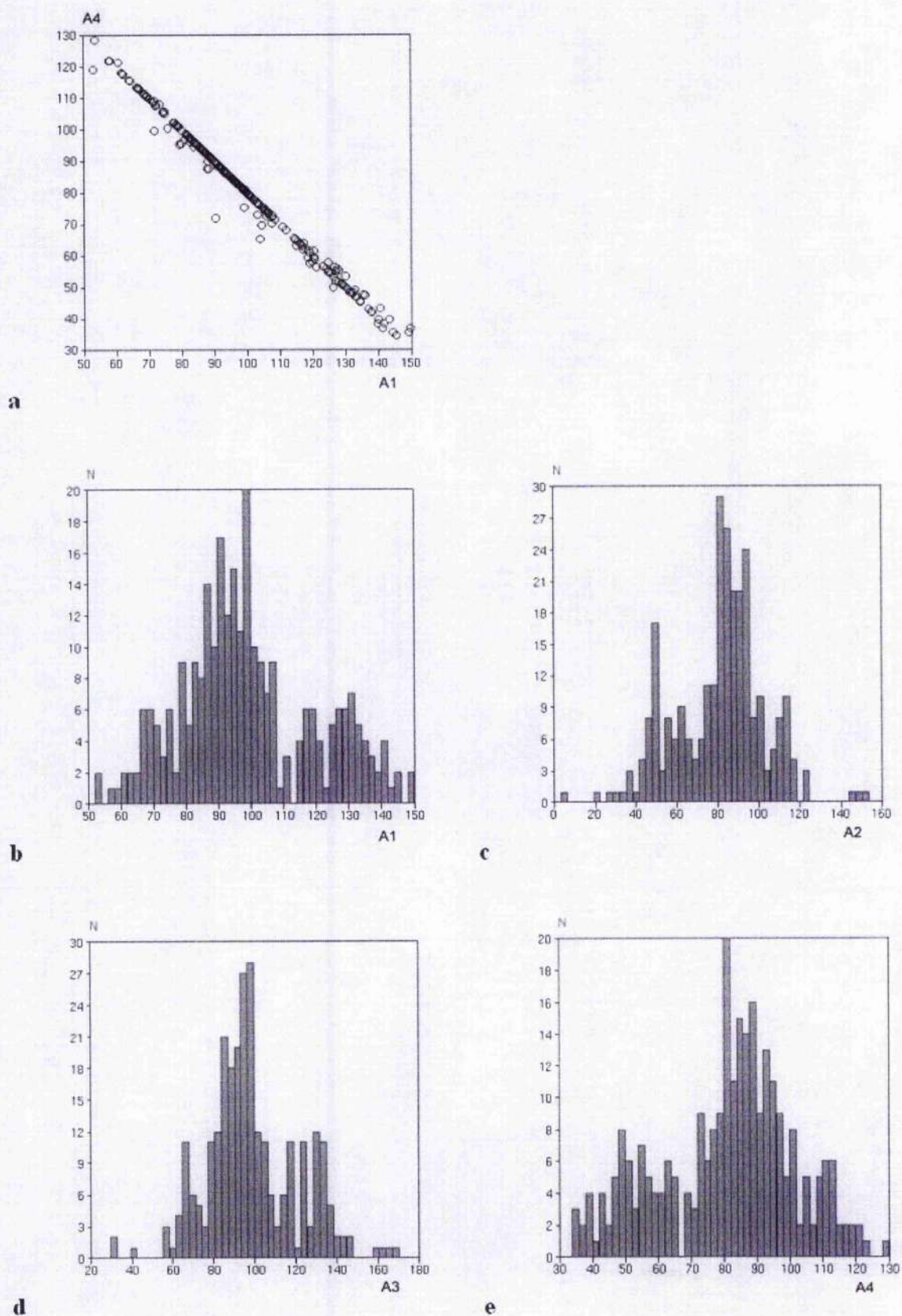


**Figure 5.2.** Geometry of interactions between *trans*-amide carbonyl groups. (a) Distribution of  $(C\cdots O)_{\min}$  for all carbonyl...carbonyl interactions. (b) Distribution of  $|T_{\text{COCO}}|$  for interactions with  $(C\cdots O)_{\min}$  less than 4 Å. (c-f) Distributions of angles A1-A4 for interactions with  $(C\cdots O)_{\min}$  less than 4 Å.

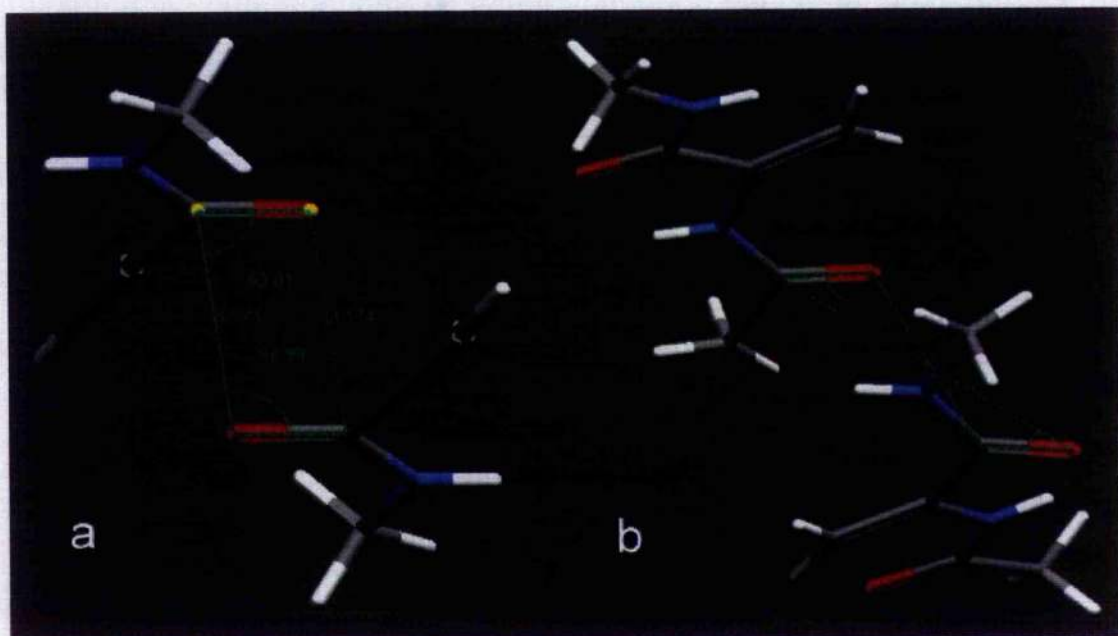
The remaining 570 interactions are now considered in detail. The distributions of the four angles, A1-A4, and of  $|T_{\text{COCO}}|$  (the modulus of the torsion angle  $T_{\text{COCO}}$ ) are shown for these interactions (Figure 5.2b-f). The distribution of  $|T_{\text{COCO}}|$  is similar to that observed for contacts between ketone carbonyl groups. It has two large peaks, one at  $0^\circ$  and one at  $180^\circ$ . The peak at  $0^\circ$  is particularly striking, and is due to occurrences where the two carbonyl groups are arranged antiparallel to each other. The total number of antiparallel occurrences, defined as  $|T_{\text{COCO}}| < 20^\circ$ , was 275. The majority (150; 55%) of these were interactions between carbonyl groups from symmetry related molecules and had  $T_{\text{COCO}}$  of precisely zero.

In a near-planar antiparallel arrangement with both C...O distances approximately equal, angles A1 and A3 have approximately equal values, as do angles A2 and A4. The degree of shear, or slippage, can be visualised by plotting one of each pair against one of the other pair, A1 against A4 is shown (Figure 5.3a). The distributions of angles A1-A4 for the antiparallel interactions are shown (Figure 5.3b-e). These distributions are multimodal, but all have one particularly pronounced peak. This is around  $97^\circ$  for both A1 and A3 and around  $84^\circ$  for A2 and A4. This corresponds to a geometry identical to that observed in antiparallel contacts between ketone carbonyls. A typical example is shown (Figure 5.4a). The other peaks observed in the distributions of A1-A4 are less pronounced, and no simple factors were identified that explained their occurrence. There are a substantial number of interactions with A1 and A3 near  $130^\circ$  and A2 and A4 near  $50^\circ$ . Visual inspection revealed many of these to occur in situations similar to what is found in protein  $\beta$ -sheets. In  $\beta$ -sheets, 'ladders' of hydrogen-bonds exist between polypeptide strands, with carbonyl-carbonyl interactions between carbonyl groups of adjacent 'rungs' and opposing strands. The interactions described here with A1 and A3 near  $130^\circ$  and A2 and A4 near  $50^\circ$  were often similar to a single, isolated, example of the type of carbonyl-carbonyl interaction found in  $\beta$ -sheet. They have only one or two hydrogen-bonds rather than the extended pattern of hydrogen-bonds definitive of protein  $\beta$ -sheet. The majority of these interactions do not occur in polypeptides, so could not form the extended pattern in any case.





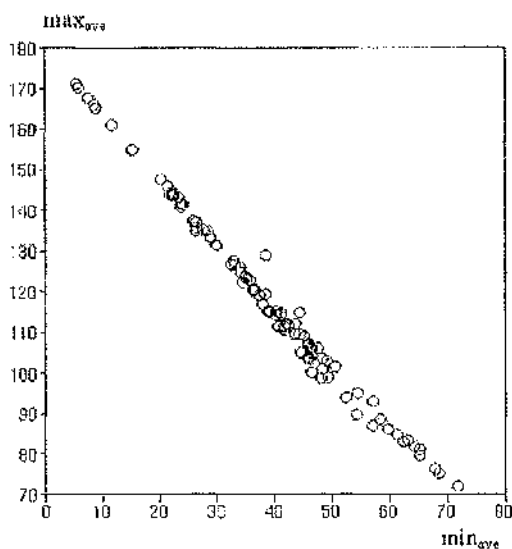
**Figure 5.3.** Geometry of antiparallel carbonyl...carbonyl interactions. (a) Plot of angle A1 against A4. (b-e) Distributions of angles A1-A4.



**Figure 5.4.** Examples of antiparallel and parallel interactions between *trans*-amide carbonyl groups. (a) Antiparallel interaction from CSD structure MPROLA10. The two groups are symmetry related ( $T_{\text{COCO}} = 0^\circ$ ). Both C...O distances are 3.17Å. Angles A1 and A3 are both  $83^\circ$ , A2 and A4 are both  $97^\circ$ . (b) Parallel interaction from CSD structure JUDZEN.  $T_{\text{COCO}}$  is  $176^\circ$ ,  $(\text{C}\cdots\text{O})_{\text{min}}$  is 3.70Å,  $\text{max}_{1,3}$  is  $111^\circ$ , and  $\text{min}_{1,3}$  is  $38^\circ$ .



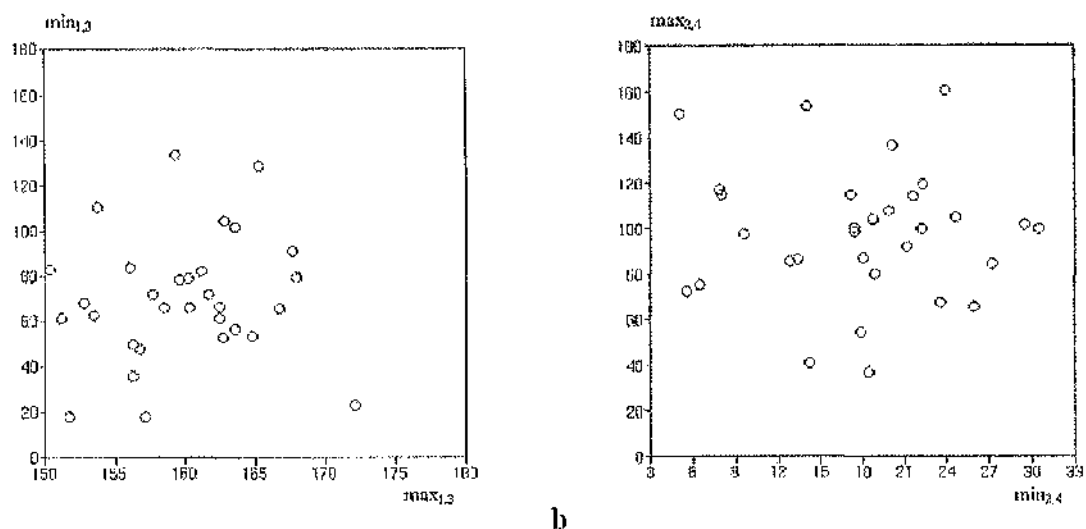
The second peak in the distribution of  $|T_{\text{COCO}}|$ , at  $180^\circ$ , is due to occurrences where the carbonyls are in a highly sheared parallel arrangement. Taking the parallel interactions to be those with  $|T_{\text{COCO}}| > 160$ , there were a total of 98. Due to the permutational symmetry issue described in the Materials and methods (section 5.2.2), direct analysis of angles A1-A4 is not useful for these interactions. Instead, permutational symmetry is dealt with by taking the two permutationally equivalent pairs of angles (A1 is equivalent to A3, as is A2 to A4), and re-assigning the lower value from each pair to the new parameters  $\text{min}_{1,3}$  and  $\text{min}_{2,4}$ , and the higher value from each pair to the new parameters  $\text{max}_{1,3}$  and  $\text{max}_{2,4}$ . The geometric descriptors can be simplified further by taking the average of the two lower values ( $\text{min}_{1,3}$  and  $\text{min}_{2,4}$ ), to produce  $\text{min}_{\text{ave}}$ , and likewise for the two higher values to produce  $\text{max}_{\text{ave}}$ . The plot of  $\text{max}_{\text{ave}}$  against  $\text{min}_{\text{ave}}$  (Figure 5.5) allows the geometry and degree of shear exhibited by the parallel interactions to be visualised. The average values of  $\text{min}_{\text{ave}}$  and  $\text{max}_{\text{ave}}$  were  $39^\circ$  and  $117^\circ$  respectively (Table 5.1). This corresponds to a more sheared arrangement than that found in parallel interactions between ketone carbonyls, where these two values are  $49^\circ$  and  $99^\circ$  respectively. A typical example is shown (Figure 5.4b).



**Figure 5.5.** Geometry of parallel carbonyl-carbonyl interactions. Plot of  $\text{max}_{\text{ave}}$  against  $\text{min}_{\text{ave}}$ .

A perpendicular motif, defined previously in an analysis of ketone carbonyl...carbonyl contacts (Allen *et al.*, 1998), was used here to group together contacts where the approach of the oxygen of one ketone carbonyl group to the carbon of another approximately described a Bürgi-Dunitz trajectory with  $O\cdots C=O$  of  $100-110^\circ$ . There are 42 *trans*-amide carbonyl...carbonyl interactions that meet the criterion for this perpendicular arrangement. The same correction for permutational symmetry as was applied for the parallel arrangement, creating min and max values from the A1-A4 parameters, is also applied here. The criterion for the perpendicular arrangement requires that  $\max_{1,3} > 150^\circ$ . This allows overlap with the antiparallel and parallel arrangements as they are defined using the  $T_{\text{COCO}}$  torsion angle solely. Therefore, antiparallel and parallel arrangements with  $\max_{1,3} > 150^\circ$  will also meet the criteria for the perpendicular arrangement. In the ketone analysis, overlap was minimal and precedence was given to the perpendicular classification. However,  $\max_{1,3}$  for parallel interactions between *trans*-amides is distributed towards higher values than for ketones. As a result, many of the interactions with  $\max_{1,3} > 150^\circ$  have low values of  $\min_{1,3}$ , more in keeping with the parallel arrangement than the perpendicular. For this reason, the parallel classification is given precedence here. The parallel arrangement accounts for 11 of the 42 interactions with  $\max_{1,3} > 150^\circ$ , the remaining 31 are considered to be of the perpendicular type.

The perpendicular interactions have average geometry similar to that observed for perpendicular contacts between ketones (Table 5.1). There is considerable spread in the distributions of the geometric descriptors, particularly  $\min_{1,3}$  and  $\max_{2,4}$ . This is visualised by plotting  $\min_{1,3}$  against  $\max_{1,3}$  (Figure 5.6a) and  $\max_{2,4}$  against  $\min_{2,4}$  (Figure 5.6b). The spread of these distributions may be due to the lack of a  $T_{\text{COCO}}$  torsion angle constraint on the perpendicular arrangement, so the carbonyl groups do not have to be in the same plane.

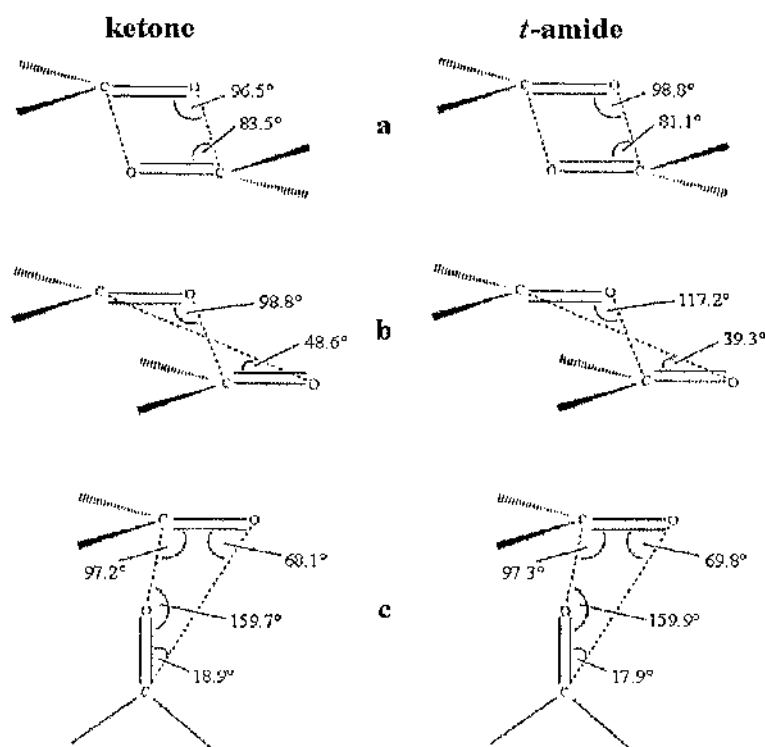


**Figure 5.6.** Geometry of perpendicular carbonyl...carbonyl interactions. (a)  $\min_{1,3}$  against  $\max_{1,3}$ . (b)  $\max_{2,4}$  against  $\min_{2,4}$ .

The proportions of *trans*-amide carbonyl...carbonyl interactions accounted for by the antiparallel, parallel, and perpendicular arrangements, together with the average geometry of these arrangements, are given in Table 5.1. The majority (70%) of the *trans*-amide carbonyl...carbonyl interactions that are not integral to a hydrogen-bond, and have  $(C\cdots O)_{\min} < 4\text{\AA}$ , are accounted for by these three arrangements. Of the 2062 *trans*-amide carbonyl...carbonyl interactions that have  $(C\cdots O)_{\min} < 4\text{\AA}$ , only 166 (8%) neither occur within a *trans*-amide...*trans*-amide hydrogen-bond, nor are in an antiparallel, parallel, or perpendicular arrangement. It should be noted that each *trans*-amide carbonyl group may be making one or more hydrogen-bonds with other non-*trans*-amide groups, but such occurrences are not considered in this analysis. The geometry of the antiparallel, perpendicular, and parallel interactions, as found for contacts between *trans*-amide carbonyls, is shown, compared with that found for contacts between ketone carbonyls (Figure 5.7).

**Table 5.1.** Numbers ( $N$ ) and average angles (A1-4) of antiparallel, parallel, and perpendicular interactions between *trans*-amide carbonyl groups. The total number of non-hydrogen-bonded interactions with  $(C...O)_{\min} < 4\text{\AA}$  was 570. The percentage of this number accounted for by each of the categories is shown. Permutational corrections are applied to the angle parameters of the parallel and perpendicular interactions such that: average values for the pair of angles A1 and A2, and the pair A3 and A4, of the parallel interactions, are identical as they are averages of  $\max_{\text{ave}}$  and  $\min_{\text{ave}}$  respectively; average values given for A1-A4 for the perpendicular interactions are averages of  $\min_{1,3}$ ,  $\min_{2,4}$ ,  $\max_{1,3}$ , and  $\max_{2,4}$ , respectively.

Description of interaction	$N$	Percentage of total 570 (%)	Average Angles( $^{\circ}$ )			
			A1 (sd)	A2 (sd)	A3 (sd)	A4 (sd)
Antiparallel	275	48	99 (21)	81 (22)	99 (22)	81 (20)
Parallel	98	17	117 (23)	117 (23)	39 (15)	39 (15)
Perpendicular	31	5	70 (28)	18 (7)	160 (5)	97 (30)



**Figure 5.7.** Comparison of antiparallel, parallel, and perpendicular interactions of ketone carbonyl groups with those of *trans*-amide carbonyl groups. Average geometries of ketone interactions (left) and *trans*-amide interactions (right) are shown. (a) Antiparallel; (b) Parallel; (c) Perpendicular.

## 6 Carbonyl-carbonyl interactions between *trans*-amide group: in the PDB

### 6.1 Brief introduction and outline

Protein studies suggest the importance of carbonyl-carbonyl interactions, particularly in repetitive structure such as  $\alpha$ -helix and  $\beta$ -sheet, but also in side-chain to main-chain interactions. The previous chapter showed that, in the CSD, carbonyl-carbonyl interactions are important between *trans*-amides. Compared to those found between ketones, they have similar geometry, and the antiparallel motif is still prevalent. Taking this evidence together, it seems worthwhile to analyse the frequency and geometry of carbonyl-carbonyl interactions in proteins, and determine whether their geometry is similar to that found in the CSD. To place them in the context of other protein structure features, information was gathered about their secondary structure and hydrogen-bonding environment.

Interactions between main-chain carbonyl groups in a subset of 454 chains from the PDB are identified. For each carbonyl-carbonyl interaction, its geometry, secondary structure, and nearby hydrogen-bonding, are considered. The interactions are categorised by the separation (in sequence) of the two interacting residues. For each separation, geometry is used to distinguish different motifs. The three motifs present in ketone and *trans*-amide carbonyl-carbonyl interactions in the CSD are not found to be representative of the geometries present in the PDB. However, other carbonyl-carbonyl interaction motifs are observed. They occur in a variety of situations with respect to secondary structure and hydrogen-bonding, and are prevalent at the C-termini of  $\alpha$ -helices.

The C-terminal end of the helix has been studied extensively in the context of helix 'capping' (reviewed by Aurora & Rose, 1998). The C-terminus is defined as where the helix hydrogen-bonding pattern ends. This leaves exposed carbonyl oxygens that must be hydrogen-bonded by other, non-repetitive, structural conformations. A common type of C-terminal 'cap' is the Schellman loop (Schellman, 1980; Milner-White, 1988). Another cap involves a proline residue at the end of the helix. This was first documented by Aurora & Rose (1998), but prolines were already known as "helix breakers" (Richardson, 1981). Here it is found that carbonyl-carbonyl interactions contribute to the stabilisation

of both of these capping conformations. Furthermore, a hydrophobic interaction within many Schellman loops (identified by Aurora & Rose; 1998) is often stabilised by a favourable carbonyl-carbonyl interaction.

## 6.2 Materials and methods

### 6.2.1 Data collation and analysis

In collaboration with Zoe Gokhale, a Bioinformatics masters student at the University of Glasgow (2003), a Perl program was created to identify carbonyl-carbonyl interactions among the dataset described in section 6.2.2, applying the criteria described in section 6.2.3. It also calculated distance and angle measurements, collated secondary structure information, and carried out Lennard-Jones energy calculations, all of which are described in the following sections. The program generated a single output file suitable for the Vista statistics program (CCDC, 1994; Bruno *et al.*, 2002). Vista was used for statistical analysis, plot preparation, and Principal Component Analysis (PCA).

### 6.2.2 Dataset

A dataset of 500 PDB files was obtained from the kinemage website (Lovell *et al.*, 2003; <http://kinemage.biochem.duke.edu>). The resolution of chains in this set is 1.8Å or better, clashscore (for atoms with B-factor < 40) is less than 22 per 1000 atoms (see Word *et al.*, 1999, for clashscore definition), fewer than 10 per 1000 atoms have main-chain bond angles (including those formed by C<sub>β</sub>) > 5 standard deviations from Engh & Huber geometry (Engh & Huber, 1991), and hydrogen atoms have been added using the program Reduce (Word *et al.*, 1999).

Blastp (Altschul *et al.*, 1990 & 1997) was set up locally and used with default settings to identify cases where two structures had more than 30% sequence identity. In such cases, the structure with lower resolution was removed. A further 16 were removed due to irregularities in PDB file helix and sheet annotation, leaving a final set of 454 protein chains.

### 6.2.3 Carbonyl-carbonyl interaction criteria and geometric descriptors

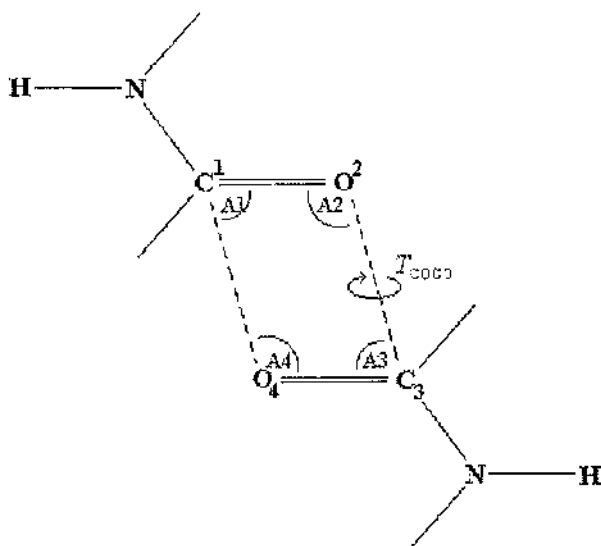
Carbonyl-carbonyl interactions were defined as occurrences where an atom of a carbonyl group was within 4Å of an atom of another carbonyl group. The residues of each of the two carbonyl groups were labelled *i* and *j*, with *i* being the first by sequence number. The sequence separation was defined as *j-i*.

The dataset was comprised of single chains, so no inter-chain interactions were present. If the interaction occurred within a helix or sheet it was ignored. Such cases were identified by checking whether residues *i* and *j* were both included in the same element in the HELIX or SHEET annotation of the PDB file. Additionally, if there was a mainchain-mainchain hydrogen-bond between *i* and *j*, the interaction was ignored (see hydrogen-bonding section 6.2.4).

Much of the analysis concerns data on local residues, specifically residues *i-1* to *i+2* and *j-1* to *j+2*. Interactions were ignored in the small number of cases where a nonstandard amino acid type occurred within either of these ranges. Nonstandard amino acid types were defined as those with a PDB annotation that was not one of the twenty standard three-letter codes.

For each carbonyl-carbonyl interaction the carbonyl atoms of residues *i* and *j* were labelled  $C^1$  and  $O^2$  and  $C^3$  and  $O^4$  respectively. A number of geometric parameters were measured: (a) the distances  $C^1 \cdots O^4$ ,  $C^3 \cdots O^2$ ,  $C^1 \cdots C^3$ ,  $O^2 \cdots O^4$ ,  $O^2 \cdots H$ ,  $O^4 \cdots H$ ,  $O^2 \cdots N$ , and  $O^4 \cdots N$  (where H and N are from the peptide unit of the opposing residue); (b) the torsion angle,  $T_{COCO}$ , made by the four carbonyl atoms in the order  $C^1=O^2 \cdots C^3=O^4$ ; (c) the four angles,  $A1$ ,  $A2$ ,  $A3$ , and  $A4$ .  $T_{COCO}$  and the angles  $A1$ - $A4$  are shown (Figure 6.1).





**Figure 6.1.** Angle and torsion angle parameters of carbonyl-carbonyl interactions. Angles A1-A4 are shown, and the torsion angle  $T_{\text{COCO}}$ . The four carbonyl atoms are numbered 1-4, for ease of reference.

## 6.2.4 Hydrogen-bond criteria and local hydrogen-bonding patterns

The program HBPLUS (v 3.0; McDonald & Thornton, 1994) was used to generate files containing information on all hydrogen-bonds present in each chain. Hydrogen-bonds were defined by the criteria described in section 3.2.1.

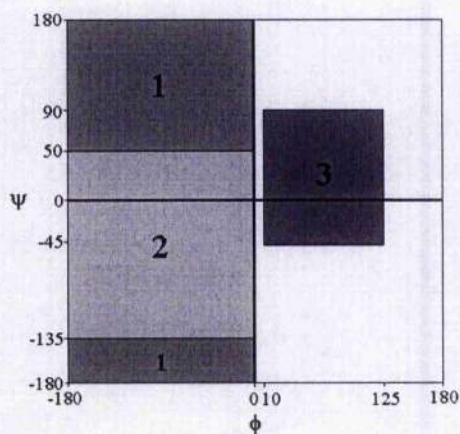
For the two residues (*i* and *j*) of each carbonyl-carbonyl interaction, data on local hydrogen-bonding was obtained as follows: the presence of hydrogen-bonds between the carbonyl of *i* with the N-H of residues *j*-1, *j*, *j*+1, and *j*+2 was checked for (and vice versa for *j* with *i*-1, *i*, *i*+1, and *i*+2). Two bit-strings of length three (the first bit-string for bonds made by the carbonyl of *i*, and the second for those of *j*) were returned such that, for example, 001-100 represents the presence of *i* to *j*+2, and *j* to *i*-1 bonds, but no others. Thus, the top-left interaction in Figure 6.9 would return the bit-strings 001-010. These bit-strings are used in Tables 6.4 and 6.6. Interactions with hydrogen-bonds of type *i* to *j*+1 or *j* to *i*+1 were eliminated from further analysis as these are cases such as that in Figure 2.1a where the carbonyl-carbonyl interaction is likely to be secondary to the hydrogen-bond.

## 6.2.5 Local secondary structure

The program DSSP (Kabsch & Sander, 1983) was used to generate secondary structure information on each chain. For each interaction, a single letter secondary structure annotation was retrieved for residues  $i-1$  to  $i+2$  and  $j-1$  to  $j+2$  to create two four-digit codes describing the local secondary structure. Two-digit codes, referring only to  $i$ ,  $i+1$  and  $j$ ,  $j+1$ , were also created.

## 6.2.6 Ramachandran plot regions

For each interaction, residues  $i-1$  to  $i+2$  and  $j-1$  to  $j+2$  were ascribed a value from 1 to 4 depending on the region of the Ramachandran plot they occupied. The regions 1 to 3 are indicated in Figure 6.2, region 4 being the remaining area. The boundaries are as follows: region 1 ( $\beta$ ),  $-180 < \phi < 0$ ,  $50 < \psi < 180$  or  $-180 < \psi < -135$ ; region 2 ( $\alpha_R$ ),  $-180 < \phi < 0$ ,  $-135 < \psi < 50$ ; region 3 ( $\alpha_L$ ),  $10 < \phi < 125$ ,  $-45 < \psi < 90$ . Four- and two-digit codes were created for each interaction, as for the secondary structure data.



**Figure 6.2.** Regions defined in Ramachandran plot annotation. Regions 1 to 3 are sometimes referred to in the text as the  $\beta$ ,  $\alpha_R$ , and  $\alpha_L$ , regions, respectively.

## 6.2.7 Energy calculations (Lennard-Jones)

The Lennard-Jones potential was used to calculate a potential energy for each carbonyl-carbonyl interaction. This method is used in most energy calculations in proteins (Maccallum *et al.*, 1995a, b). The calculation was limited to the four atoms ( $C^1$ ,  $O^2$ ,  $C^3$  and  $O^4$ ) of the two carbonyl groups. A potential is calculated for each pair ( $i$  and  $j$ ) of atoms (in this case the carbon···carbon, oxygen···oxygen, and two carbon···oxygen pairs) and the final energy is the addition of each potential. The potential is proportional to the distance between interacting atoms ( $r$ ). It is broken into three components: a strong repulsive interaction (a function of either  $1/r^{12}$  or  $1/r^9$ , in this case  $1/r^{12}$  is used, this is justified by Maccallum *et al.*, 1995b) between the van der Waals atomic cores; a weak attraction (a function of  $1/r^6$ ) representing dispersion forces; and the electrostatic interaction (a function of  $1/r$  and of the point charges ( $q$ ) of the interacting atoms). Partial charges were assigned from empirically derived values, as described by Lifson *et al.* (1979), and listed in Table 6.1.

For each pair, the energy ( $E$ ) =  $A_{ij}/r^9 - C_{ij}/r^6 + 332qiqj/r$

where the constants  $A$ ,  $C$ , and  $q$ , have the values in Table 6.1,  $r$  is the distance between interacting atoms, and where  $A_{ij} = (A_i A_j)^{1/2}$  and  $C_{ij} = (C_i C_j)^{1/2}$ .

**Table 6.1.** Constants ( $A$  and  $C$ ) and partial charges ( $q$ ) used in 9-6-1 Lennard-Jones potential, as derived by Lifson *et al.* (1979) using empirical refinement.

	C	O
$A$	12500	45800
$C$	355	1410
$q$	+0.46	-0.46

## 6.2.8 Principal Component Analysis

The four angle parameters, A1-A4, were subjected to Principal Component Analysis (PCA) to aid the process of categorizing interactions into distinct motifs. This was carried out using the Vista statistics program (CCDC, 1994; Bruno *et al.*, 2002).

## 6.3 Results & Discussion

### 6.3.1 Overall data

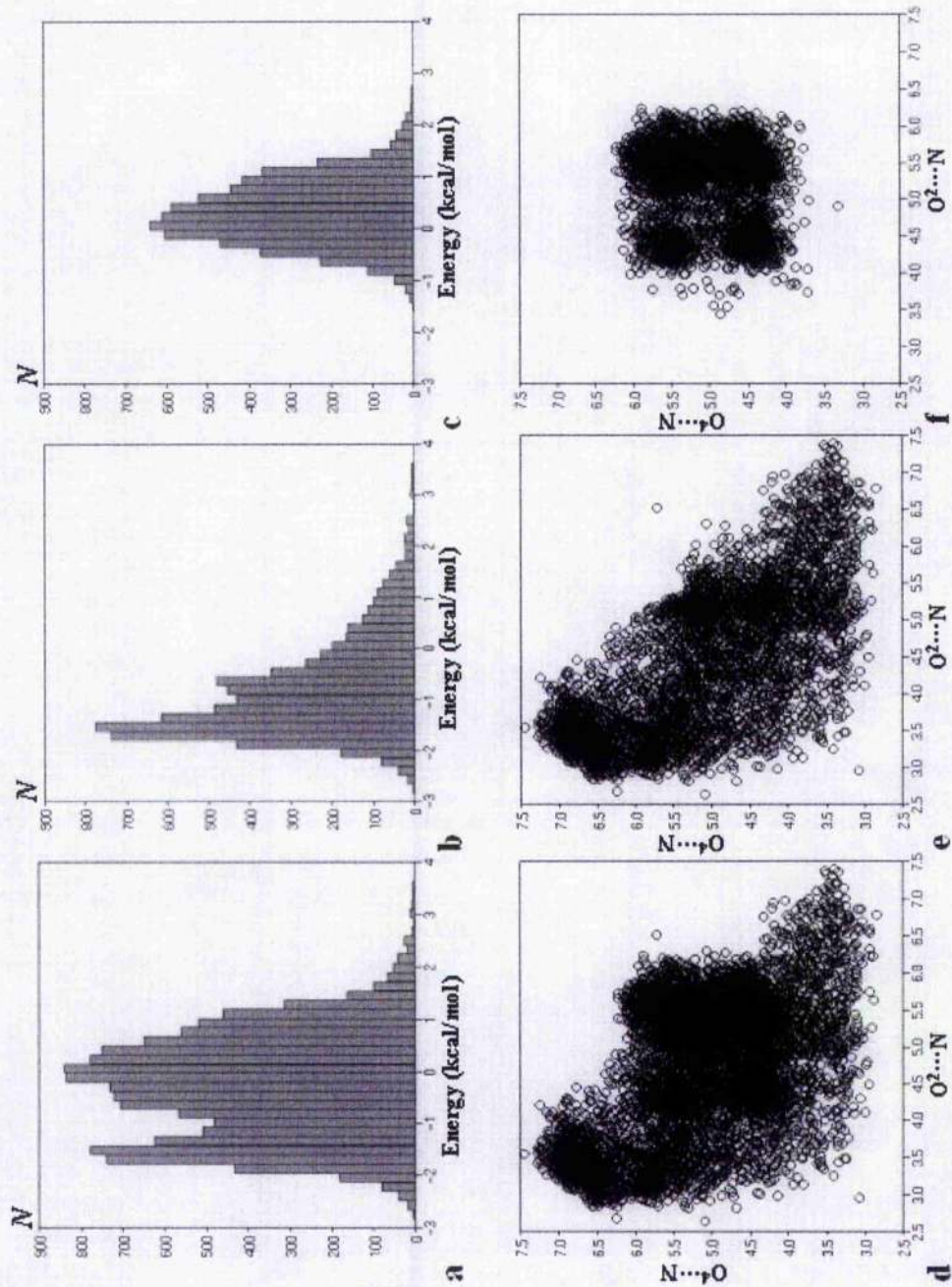
Interactions between main-chain carbonyl groups at least two residues apart were identified in a database of 454 proteins. Those that occurred between residues in the same secondary structure element ( $\alpha$ -helix or  $\beta$ -sheet) were ignored. Cases where the N-H of the peptide group of one of the interacting carbonyls was hydrogen-bonded to the oxygen of the other interacting carbonyl (as in Figure 2.1a) were eliminated, as the carbonyl-carbonyl interaction in such cases is likely to be secondary to the hydrogen-bond. Such occurrences are also eliminated in chapter 5, so this allows comparison of the results from proteins with those from the small molecule database.

A carbonyl-carbonyl interaction was defined as any case where an atom of one carbonyl group is within  $4\text{\AA}$  of an atom of another. There were 12,850 such interactions in the database. A measure of electrostatic favourability was obtained using a Lennard-Jones potential, limited just to the four atoms of the two interacting carbonyl groups. The geometry and local environment, in terms of hydrogen-bonding and secondary structure, of each interaction was examined. The parameters, and the terms used to describe them, are explained in the Materials and methods section 6.2.

The distribution of the energy for all interactions is shown (Figure 6.3.a). It has two peaks, at approximately 0 and  $-1.5$  kcal/mol. Of the 12,850 interactions, there are 54% where one of the two carbon to oxygen distances is less than  $4\text{\AA}$ . The peak at  $-1.5$  kcal/mol is populated almost exclusively by these, as shown in Figure 6.3.b. Those cases with carbon to oxygen distances greater than  $4\text{\AA}$  are responsible for the peak at 0 kcal/mol, as shown in Figure 6.3.c. This would be expected, as the favourable component of the electrostatic interaction comes from the attraction between the carbon and oxygen atoms. It is also observed that the neutral interactions, that form the peak at 0 kcal/mol, fall into four clusters when the  $O^4 \cdots N$  and  $O^2 \cdots N$  distances are plotted against each other (Figure 6.3.f.). The plot of  $O^4 \cdots N$  against  $O^2 \cdots N$  for all occurrences (Figure 6.3.d.) shows these four clusters together with two other clusters, one with a relatively long  $O^4 \cdots N$  and short  $O^2 \cdots N$ , and the other with short  $O^4 \cdots N$  and long  $O^2 \cdots N$ . The two other clusters are

prevalent in the plot of  $O^4 \cdots N$  against  $O^2 \cdots N$  for the favourable interactions (Figure 6.3.e.).





**Figure 6.3.** Overall energy distributions and geometry. (a-c) Histograms showing energy distributions. (d-f) Plots of  $O^4 \dots N$  against  $O^2 \dots N$ . (a and d) for all carbonyl-carbonyl interactions; (b and e) for those with a carbon to oxygen distance less than 4Å; (c and f) for those with both carbon to oxygen distances greater than 4Å. Energy is in kcal/mol.

### 6.3.2 Motif identification

This section describes the identification and definition of the different motifs. Initially, Principal Component Analysis (PCA) was applied to the four angles A1-A4, and the parameters generated were used to identify distinct motifs. However, it was found that the two nitrogen to oxygen distances ( $O^4 \cdots N$  and  $O^2 \cdots N$ ) provided sufficient discrimination between clusters to allow categorisation based on these alone, with the added benefit that the distances could be more easily translated into meaningful descriptions of the geometry than their PCA counterparts. Motifs are identified based on the occurrence of clusters in plots of  $O^4 \cdots N$  against  $O^2 \cdots N$ . The cut-off chosen for each cluster was based on visual inspection of the plots. Although crude, this method was effective in discriminating between interactions occurring in a range of distinct protein structural features, as evidenced in the motif description section: 6.3.3. Images of typical motifs, and parameters describing each motif are also presented in section 6.3.3. The range of values of  $O^4 \cdots N$  and  $O^2 \cdots N$  chosen to define each motif are shown in Table 6.2, and boxes are marked on Figures 6.4 to 6.8 showing each motif's region of the  $O^4 \cdots N$  against  $O^2 \cdots N$  plot.

A distinction was made between sequentially proximate interactions, where the two residues were near to each other in sequence, and sequentially distant interactions between residues far apart in sequence. Separations of 6 residues or less were considered sequentially proximate (sequence separation is defined as  $j-i$  where  $i$  and  $j$  are the residue numbers, and  $i$  is first in the sequence). This value was chosen because those with separations greater than 6 have similar geometry to each other, as is shown later in this section.

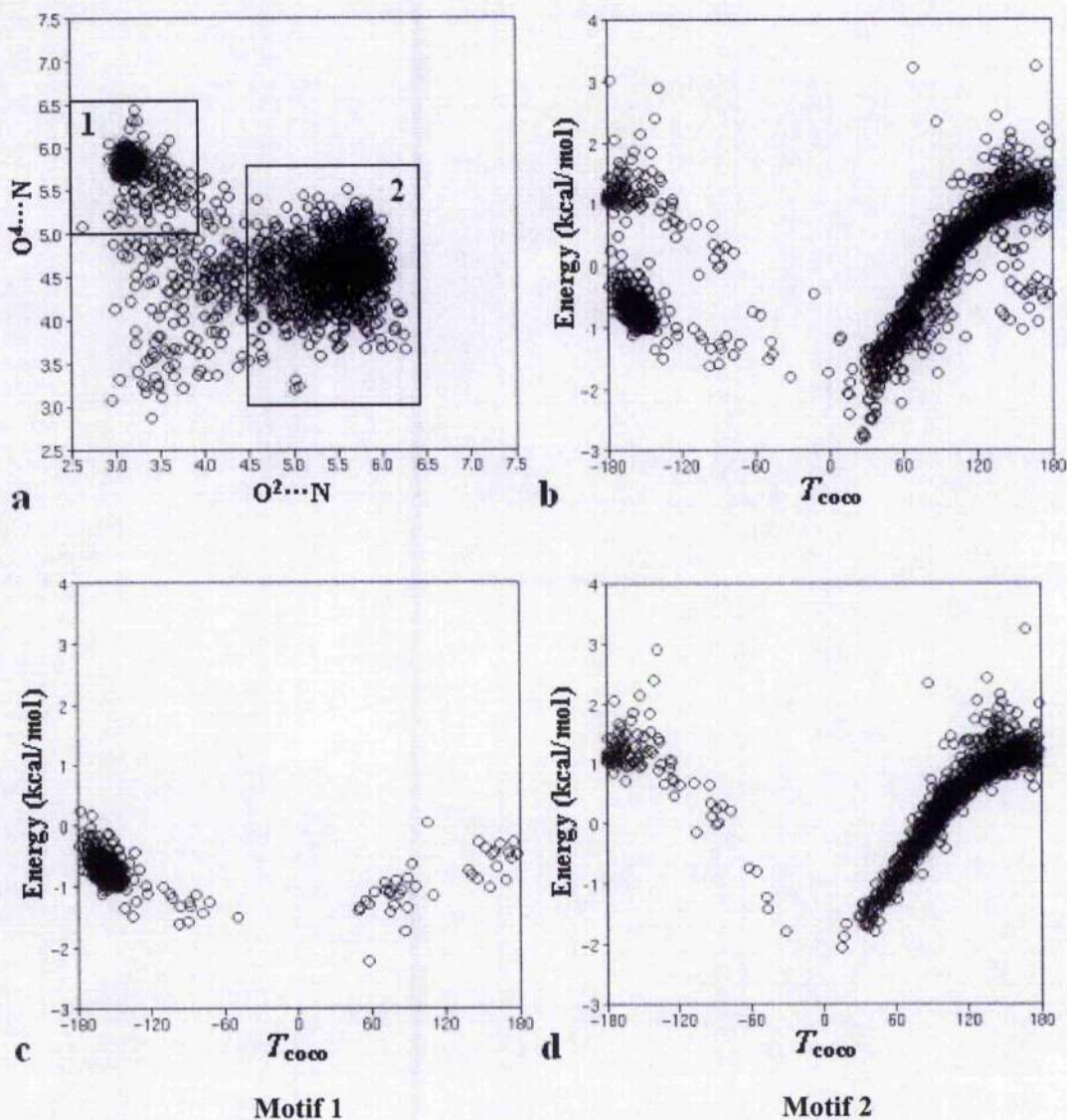
**Table 6.2.** Motif definition. For each motif, its sequence separation, and the range of  $O^4 \cdots N$  and of  $O^2 \cdots N$  used to define it is shown.

Motif	Sequence separation	$O^4 \cdots N$ range (Å)		$O^2 \cdots N$ range (Å)	
1	2	5.0	6.6	2.5	4.0
2	2	3.0	5.8	4.5	6.5
3	3	6.1	7.5	2.5	4.2
4	3	4.3	6.3	4.5	6.5
5	3	4.5	6.1	2.5	4.2
6	4	6.1	7.5	2.5	4.2
7	4	3.2	6.1	4.5	6.3
8	5	5.0	7.5	2.5	4.0
9	5	2.8	6.5	6.3	4.8
10	$\geq 6$	5.0	6.25	3.9	5.0
11	$\geq 6$	5.0	6.25	5.0	6.25
12	$\geq 6$	3.9	5.0	5.0	6.25
13	$\geq 6$	3.9	5.0	3.9	5.0
14	$\geq 6$	5.0	7.5	2.5	3.9
15	$\geq 6$	2.5	3.9	5.0	7.5

*Sequentially proximate interactions (residue difference < 6):*

The number of sequentially proximate interactions was 6,579, 51% of the total. Of these, 1815 had a separation of 2 residues. The plot of  $O^4 \cdots N$  against  $O^2 \cdots N$  for these (Figure 6.4.a) shows two clusters. Each cluster was considered a separate motif, defined by the ranges of  $O^4 \cdots N$  and of  $O^2 \cdots N$  shown in Table 6.2. Also shown is the carbonyl-carbonyl interaction energy plotted against the torsion angle,  $T_{\text{CCCO}}$  (Figure 6.4.b). Again, this shows two clusters. Figures 6.4.c and d show the carbonyl-carbonyl interaction energy plotted against the torsion angle,  $T_{\text{CCCO}}$ , for motifs 1 and 2, respectively. It is apparent from Figures 6.4.c and d that the two clusters in Figure 6.4b correspond to motifs 1 and 2.

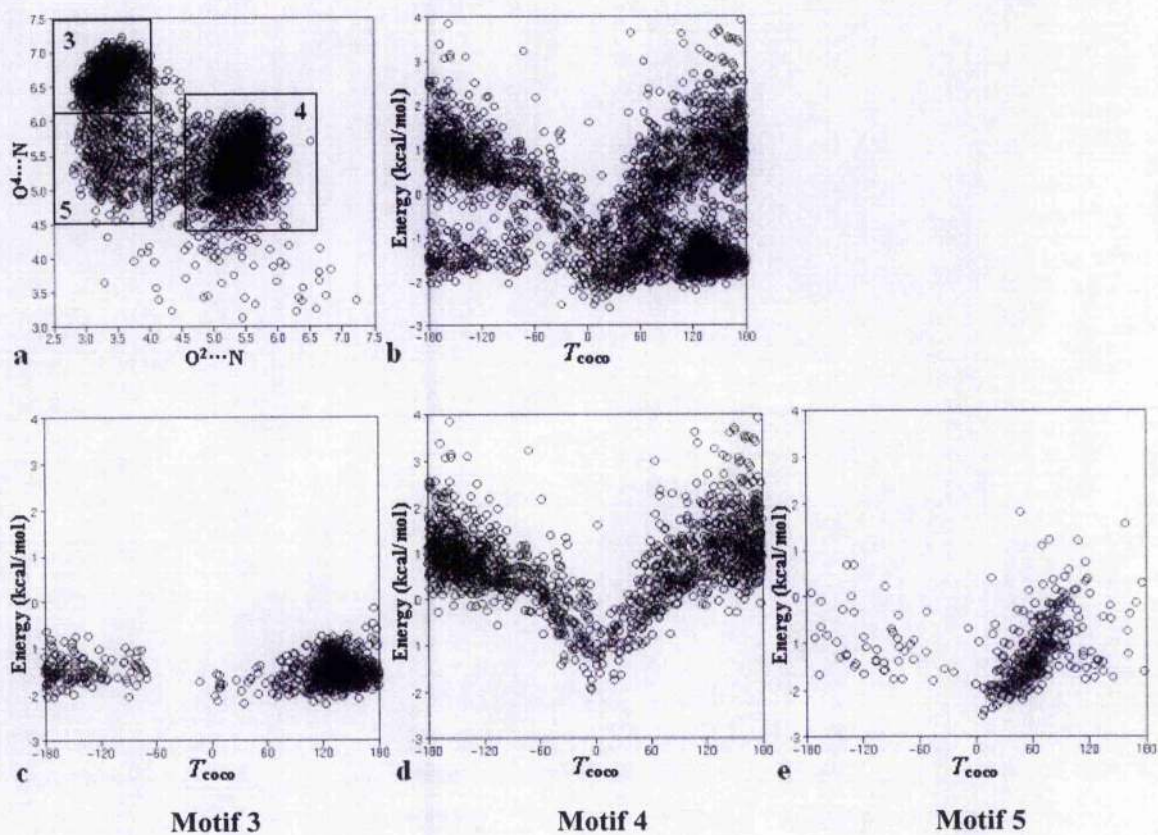




**Figure 6.4.** Definition of motifs 1 and 2. (a) Plot of  $O^4 \cdots N$  against  $O^2 \cdots N$  for interactions with a separation of two residues. Boxes indicating defining regions for each motif are labelled with motif number. (b) Plot of energy against  $T_{coco}$  for interactions with a separation of two residues. (c) and (d) Plots of energy against  $T_{coco}$  for motifs 1 and 2 respectively. Energy is in kcal/mol.

A subset of 3160 interactions featured a separation of 3 residues. In this case three clusters were observed on the plot of  $O^4 \cdots N$  against  $O^2 \cdots N$  (Figure 6.5.a.). The clusters were categorised as motifs 3 to 5, defined by the ranges of  $O^4 \cdots N$  and of  $O^2 \cdots N$  shown in Table 6.2. These three motifs also represent different clusters on the plot of energy against  $T_{\text{coco}}$  (Figure 6.5.b-c.).

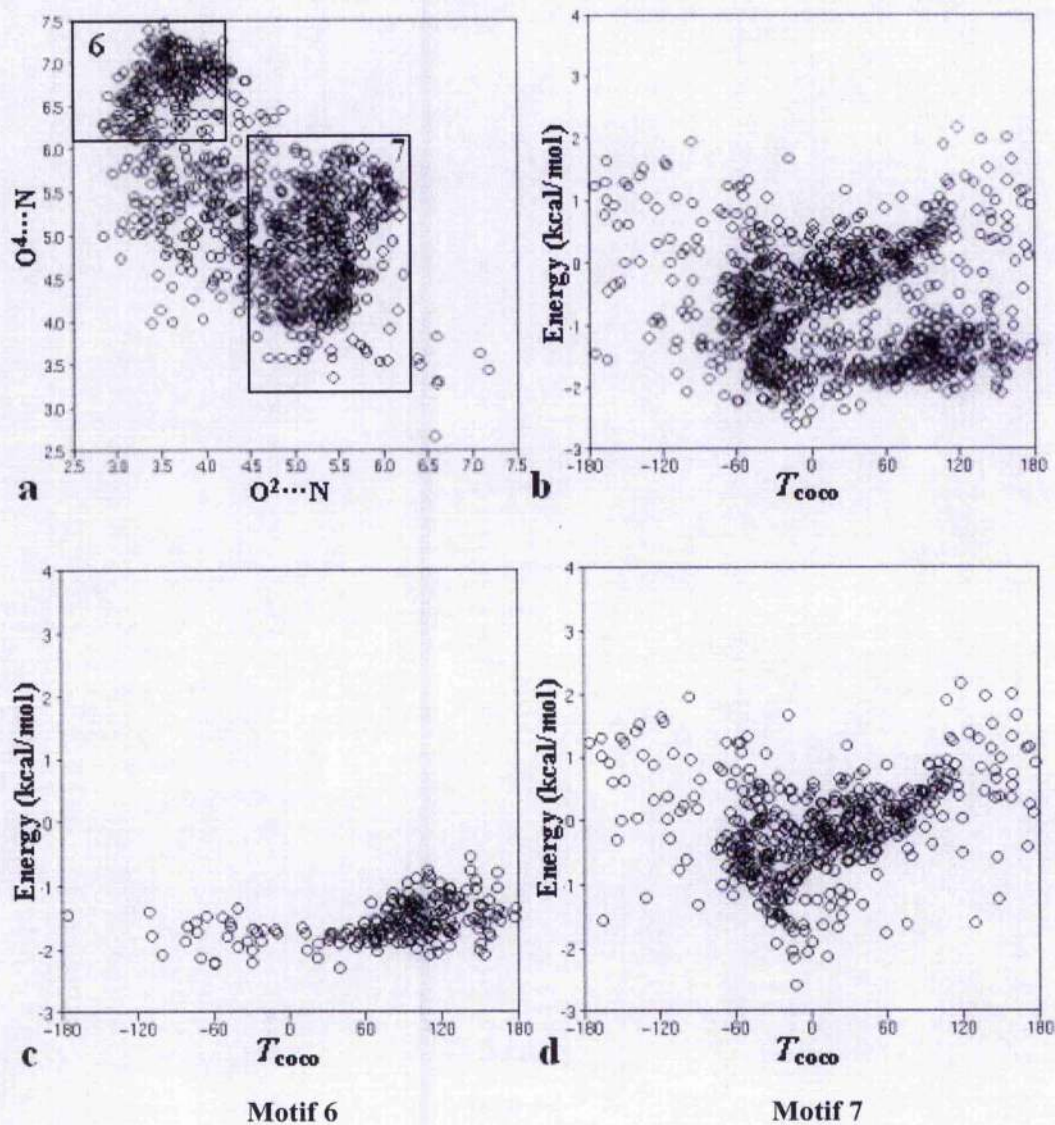




**Figure 6.5.** Definition of motifs 3, 4, and 5. (a) Plot of  $O^4 \cdots N$  against  $O^2 \cdots N$  for interactions with a separation of three residues. Boxes indicating defining regions for each motif are labelled with motif number. (b) Plot of energy against  $T_{coco}$  for interactions with a separation of three residues. (c-e) Plots of energy against  $T_{coco}$  for motifs 3, 4, and 5 respectively. Energy is in kcal/mol.

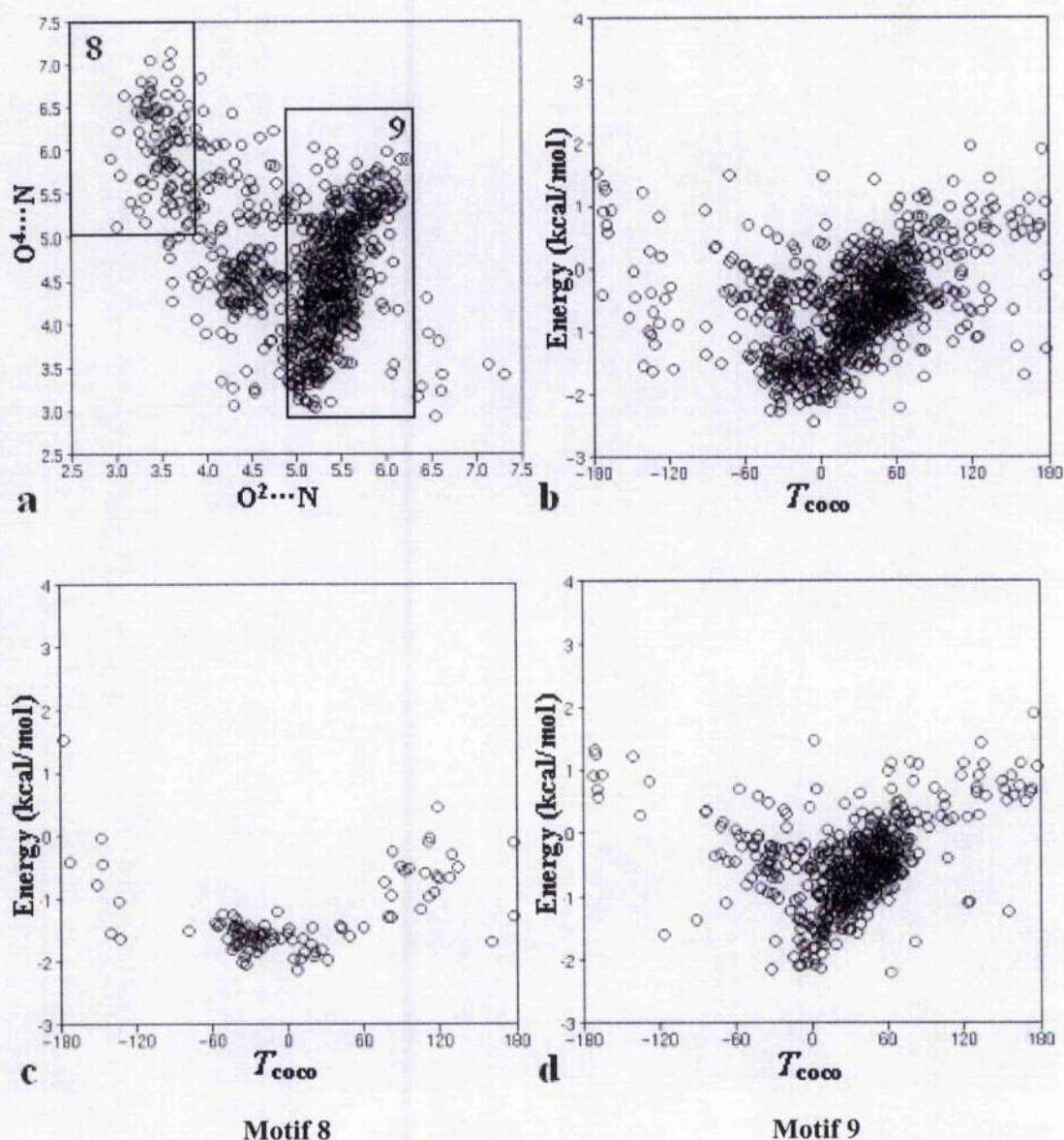
The same process was followed for categorisation at sequence separations of four and five residues, as described for two and three residues, where motifs are defined by clusters observed in the plot of  $O^4 \cdots N$  against  $O^2 \cdots N$ . At a separation of four residues, 845 interactions are found, and at five residues, 758 are found. Plots of  $O^4 \cdots N$  against  $O^2 \cdots N$  are shown (Figure 6.6.a. and 6.7.a.). Motifs 6 and 7, were identified among sequence separations of four residues, and motifs 8 and 9 among sequence separations of five residues. Ranges of  $O^4 \cdots N$  and of  $O^2 \cdots N$  for each motif are shown in Table 6.2. Again, these motifs represent different clusters on the plots of energy against  $T_{\text{coco}}$  (Figure 6.6.b-d. and 6.7.b-d.). A third cluster can be seen in Figure 6.7.a., with  $O^4 \cdots N$  and  $O^2 \cdots N$  both at approximately 4.5 Å. The interactions populating this cluster are similar in character to those identified as motif 12 among the sequentially distant interactions, and are analysed along with those (see below).





**Figure 6.6.** Definition of motifs 6 and 7. (a) Plot of  $O^4 \cdots N$  against  $O^2 \cdots N$  for interactions with a separation of four residues. Boxes indicating defining regions for each motif are labelled with motif number. (b) Plot of energy against  $T_{coco}$  for interactions with a separation of four residues. (c-d) Plots of energy against  $T_{coco}$  for motifs 6 and 7 respectively. Energy is in kcal/mol.



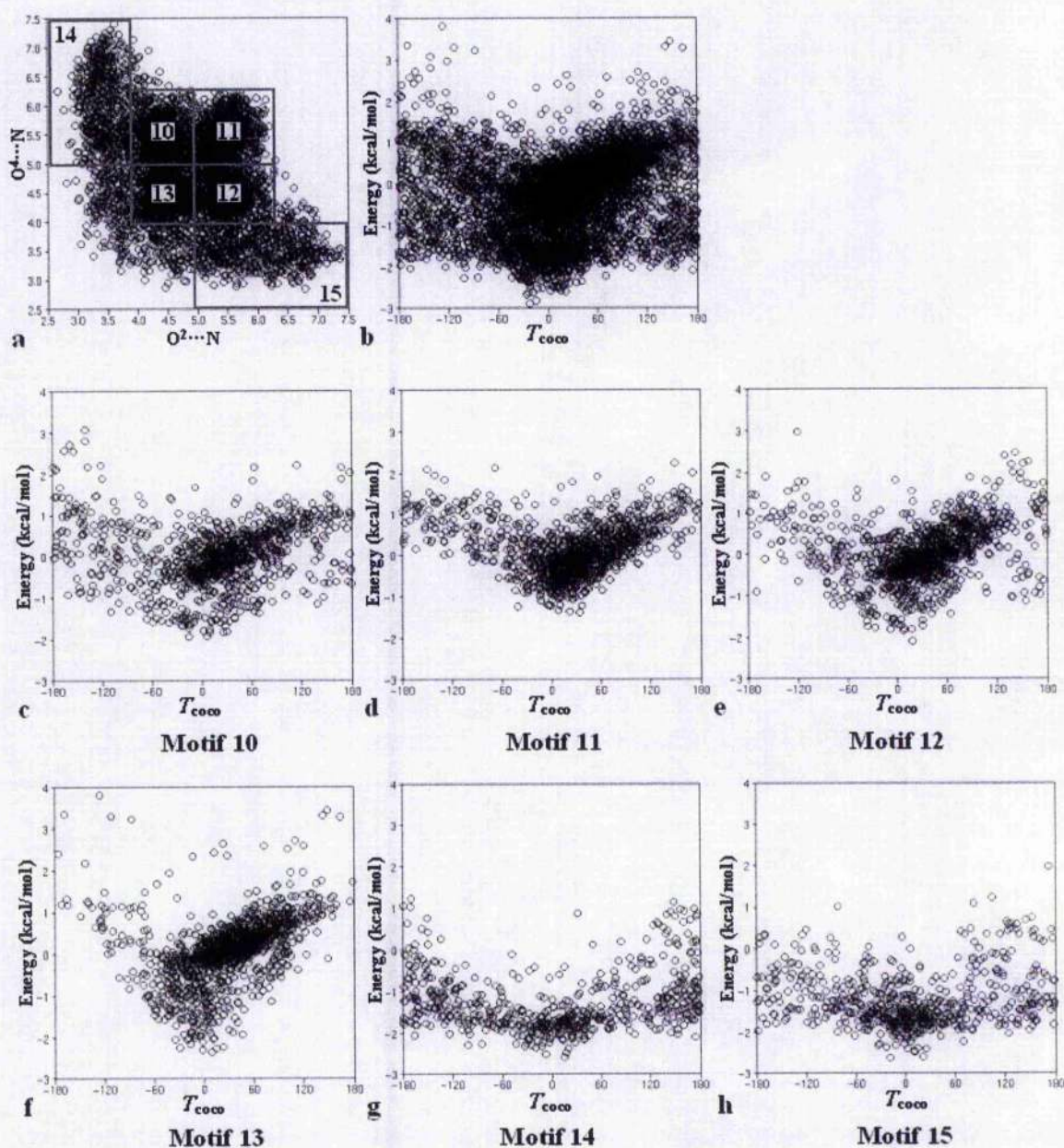


**Figure 6.7.** Definition of motifs 8 and 9. (a) Plot of  $O^4 \cdots N$  against  $O^2 \cdots N$  for interactions with a separation of five residues. Boxes indicating defining regions for each motif are labelled with motif number. (b) Plot of energy against  $T_{coco}$  for interactions with a separation of five residues. (c-d) Plots of energy against  $T_{coco}$  for motifs 8 and 9 respectively. Energy is in kcal/mol.

*Sequentially distant interactions (residue difference  $\geq 6$ ):*

Beyond a sequence separation of five, this factor becomes less important, and as a result the distributions of  $O^4 \cdots N$  against  $O^2 \cdots N$  (and of any other geometric parameters examined) are similar at different separations. Because of this all the sequentially distant interactions were grouped together, and categorised into motifs 10 to 15, based on the six main clusters observed on the plot of  $O^4 \cdots N$  against  $O^2 \cdots N$  (Figure 6.8.a.). Motifs were defined by the ranges of  $O^4 \cdots N$  and of  $O^2 \cdots N$  shown in Table 6.2. Plots of energy against  $T_{\text{coco}}$  are similar for motifs 10 to 13 (Figure 6.8.c-f.), and for motifs 14 and 15 (Figure 6.8.g and h.). The cluster identified previously at a residue separation of five is included among motif 13.

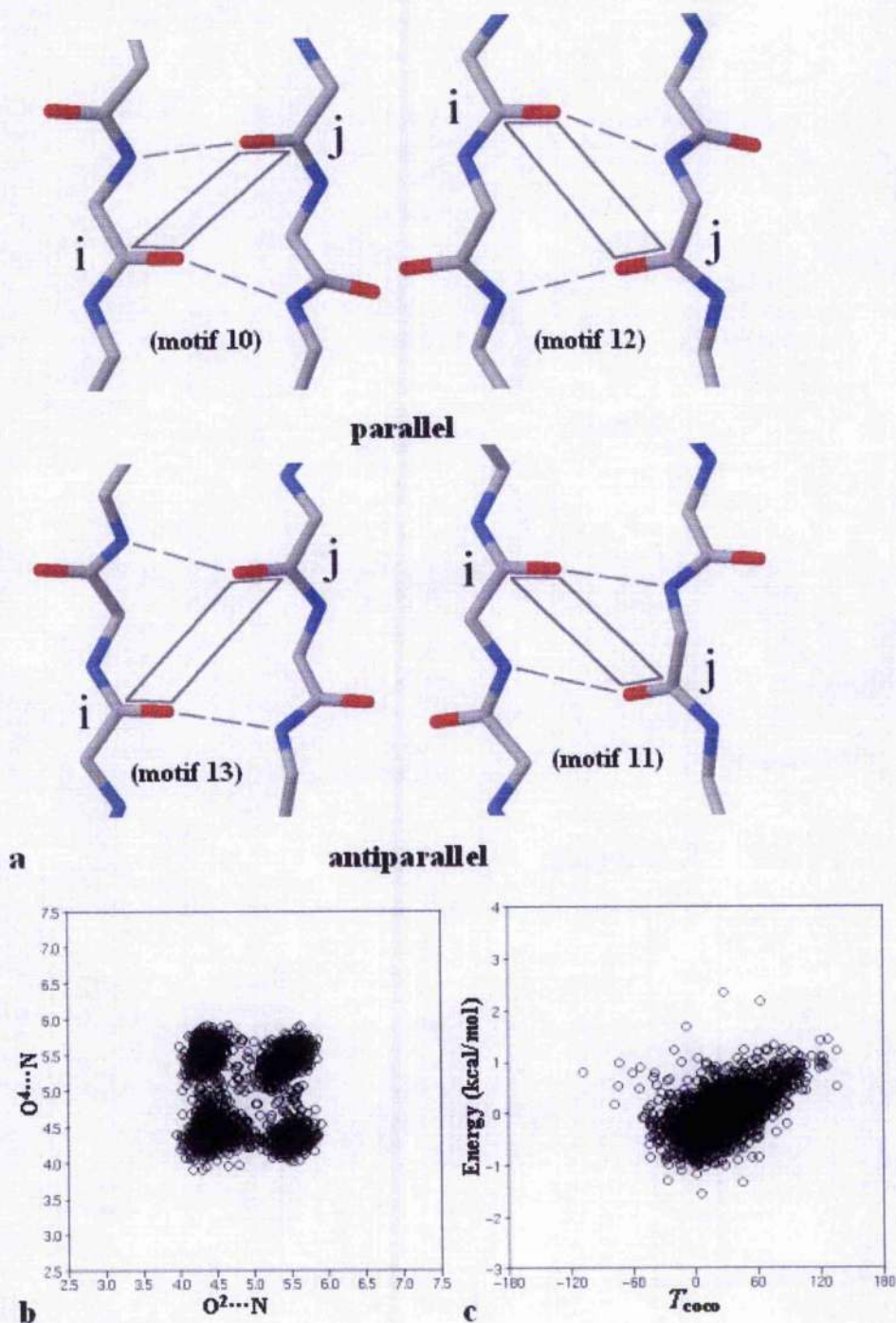




**Figure 6.8.** Definition of motifs 10 to 15. (a) Plot of  $O^4 \cdots N$  against  $O^2 \cdots N$  for interactions with a separation of six or more residues. Boxes indicating defining regions for each motif are labelled with motif number. (b) Plot of energy against  $T_{coco}$  for interactions with a separation of six or more residues. (c-h) Plots of energy against  $T_{coco}$  for motifs 10 to 15 respectively. Energy is in kcal/mol.



A large proportion (1785; 37%) of motifs 10-13 occur between strands that are hydrogen-bonded in the fashion found in  $\beta$ -sheets, but are not annotated as  $\beta$ -sheet in the PDB annotation (and are therefore not eliminated before analysis), either because they are at the end of the sheet, or the length of hydrogen-bonded contact between the strands is too short to be considered  $\beta$ -sheet. Local hydrogen-bonding patterns were used to identify these. They occurred in four types, each of which were defined by two hydrogen-bonds, each made by one of the two interacting carbonyls to an NH group on the opposing strand, as shown (Figure 6.9a). The plots of  $O^4 \cdots N$  against  $O^2 \cdots N$  and of energy against  $T_{\text{coco}}$  for these interactions are shown (Figure 6.9 b and c.). The similarity of Figure 6.9. c. with the relevant plots for motifs 10-13 (Figure 6.8 c-f.) can be observed.



**Figure 6.9.** Carbonyl-carbonyl interactions in  $\beta$ -sheet. (a) The four types of carbonyl-carbonyl interaction that occur within parallel (top two) and antiparallel (bottom two)  $\beta$ -sheet. Continuous grey lines indicate each carbonyl-carbonyl interaction, dashed lines show the defining hydrogen-bonds. Residues  $i$  and  $j$  for each interaction are marked (interacting residues are labelled  $i$  and  $j$  as described in the Materials and methods section 6.2.3). Similar geometry and hydrogen-bonding is found among the carbonyl-carbonyl interaction motifs indicated in brackets. (b) and (c) Plots of  $O^4 \cdots N$  against  $O^2 \cdots N$ , and of energy against  $T_{coco}$ , respectively, for interactions with hydrogen-bonding of the type found in  $\beta$ -sheet. Energy is in kcal/mol.

### 6.3.3 Motif Description

Each motif was examined in terms of its average geometry, energy and local environment. The parameters measured are described in the Materials and methods section. The numbers of each motif, together with the average energy, the torsion angle ( $T_{\text{coco}}$ ), and the four angles (A1-A4), are shown in Tables 6.3 and 6.5. The positions of the two interacting residues on the Ramachandran plot, their secondary structure annotation, and a description of local hydrogen-bonding, are shown in Tables 6.4 and 6.6. Examples from each motif are shown in Figures 6.10 and 6.11. The following text describes sequentially proximate and sequentially distant interactions.

#### Sequentially proximate interactions

The following text describes motifs 1 to 9 (examples shown in Figure 6.10), summarising Tables 6.3 and 6.4.

Motif 1 is a favourable interaction (in terms of the energy of the carbonyl-carbonyl interaction) that often occurs at the N-terminal end of  $\alpha$ -helices. In 78% of cases residue  $j$  (sequentially the second of the two interacting residues; interacting residues are labelled  $i$  and  $j$  as described in the Materials and methods section 6.2.3) is within an  $\alpha$ -helix. In a smaller number of cases, this motif occurs at the beginning of a hydrogen-bonded turn. This is essentially the same interaction as that occurring internally within many  $\alpha$ -helices, between residues two apart in the sequence.

Motif 2 is usually unfavourable, and often occurs between the carbonyl of a central residue of a  $\beta$ -turn and the carbonyl of the fourth residue of the turn. The large proportion of occurrences (43%) with both residues  $i$  and  $i+1$  in region 2 of the Ramachandran plot (an example is shown in Figure 6.10. Motif 2a) suggests that these  $\beta$ -turns are almost all type I. The rest of the occurrences of this motif represent a varied range of bends in the polypeptide chain. An example from the large number with residue  $i$  in the  $\beta$  region (region 1 in Figure 6.2.) of the Ramachandran plot is shown (Figure 6.10. Motif 2b).

Motif 3 is favourable and often occurs at the C-terminal end of an  $\alpha$ -helix. The hydrogen-bonding patterns found in these motifs reflects the different ways in which helices end. In 49% of cases residue  $i$  hydrogen-bonds to the N-H of residue  $j$ , but not to the N-H of residue  $j+2$ . A structure of this type is shown in motif 3a in Figure 6.10. In some cases a hydrogen-bond is formed with the N-H of  $j+2$ , as in motif 3b in Figure 6.10. In these cases the  $i$  to  $j$  hydrogen-bond may or may not be present, in motif 3b it is not. Motif 3b is also known as a paperclip or Schellman loop. The favourable carbonyl-carbonyl interaction contributes to the stability of the paperclip/Schellman loop. Such loops are frequently observed at the C-termini of  $\alpha$ -helices.

Motif 4, like motif 2, most often occurs in association with a  $\beta$ -turn (67% have a hydrogen-bond from the N-H of  $i$  to the carbonyl of  $j$ ), except that the carbonyl-carbonyl interaction in this case is between the carbonyl groups of the two residues that form the  $\beta$ -turn hydrogen-bond. The interaction is unfavourable. A large proportion (52%) occur at the C-termini of  $\alpha$ -helices or 3/10-helices, and a proportion (22%) occur at the N-terminal start of  $\beta$ -sheets.

Motif 5, like motif 3, is a favourable interaction that often occurs at the C-termini of  $\alpha$ -helices. The majority of these motifs (185; 52%) have a proline at  $j+1$ . An example of this is shown in Figure 6.10. 5a. In 36% of cases there is a  $\beta$ -turn defined by a hydrogen-bond from the N-H of  $j$  to the carbonyl of  $i$ . As in the motif 2 interactions, nearly all these  $\beta$ -turns are type I. An example is shown in Figure 6.10. motif 5b.

Motifs 6 and 7 both occur primarily at the C-termini of  $\alpha$ -helices. It is interesting that no other significant data clusters were observed for carbonyls of residues four apart in the sequence. These represent another way in which the helix can end. In the case of motif 6 (shown in Figure 6.10. motif 6), the carbonyl-carbonyl interaction is favourable, and residue  $j$  tends to lie in the  $\alpha_L$  region of the Ramachandran plot (region 3 in Figure 6.2). For motif 7 the energy of the carbonyl-carbonyl interaction is close to 0 kcal/mol and residue  $j$  lies in the  $\beta$  region (region 1 in Figure 6.2).

Motifs 8 and 9 have structural similarities to motifs 6 and 7, respectively. The carbonyl-carbonyl interaction geometry is also similar, and again the motifs are found primarily at the ends of helices. The difference is that residue  $i+4$  in motifs 8 and 9 (which is residue  $j$  in motifs 6 and 7) now forms a hydrogen-bond with the carbonyl of residue  $i$ , and the carbonyl-carbonyl interaction is now with residue  $i+5$  rather than  $i+4$ .

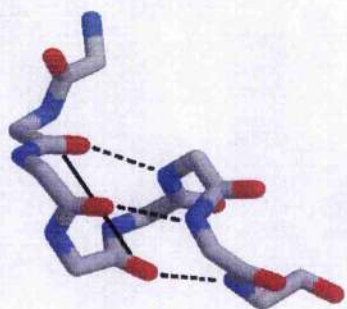
**Table 6.3.** Geometry of sequentially proximate motifs. The Motif id, Sequence separation, and numbers (N), of each motif are shown, along with the averages and standard deviations of the carbonyl-carbonyl interaction energy (in kcal/mol), the  $T_{\text{coco}}$  torsion angle, and the angles A1-A4 in degrees. \*The Outliers in  $T_{\text{coco}}$  were eliminated so that the distribution was restricted to a range of no more than  $180^\circ$ , to allow calculation of the mean and R bar values for circular data (Allen & Johnson, 1991).

Motif id	Seq. sep.	N	Energy (sd)	$T_{\text{coco}}$ * (R bar)	A1 (sd)	A2 (sd)	A3 (sd)	A4 (sd)
1	2	503	-0.69 (0.3)	-156.4 (1.0)	113.0 (7)	130.8 (9)	29.6 (9)	47.1 (6)
2	2	1142	0.46 (0.8)	117.2 (0.8)	99.8 (12)	51.2 (9)	109.3 (8)	59.8 (16)
3	3	1115	-1.49 (0.3)	139.6 (0.9)	151.6 (13)	130.5 (10)	36.6 (9)	17.9 (11)
4	3	1492	0.79 (0.8)	-169.3 (0.8)	148.0 (16)	52.3 (13)	109.6 (13)	17.4 (12)
5	3	359	-1.16 (0.7)	63.3 (0.9)	128.7 (20)	88.7 (12)	80.7 (12)	47.7 (17)
6	4	209	-1.56 (0.3)	100.1 (0.8)	144.7 (15)	135.4 (14)	35.6 (12)	28.4 (12)
7	4	458	-0.13 (0.8)	4.4 (0.7)	126.0 (24)	56.9 (16)	117.8 (18)	51.8 (20)
8	5	100	-1.32 (0.6)	-13.0 (0.8)	119.1 (15)	126.1 (22)	52.1 (17)	56.8 (12)
9	5	522	-0.51 (0.7)	33.1 (0.9)	108.2 (14)	53.1 (11)	124.1 (14)	65.6 (14)

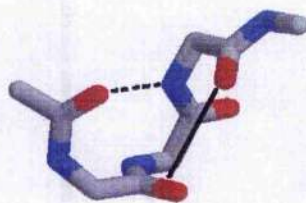
**Table 6.4.** Local secondary structure of sequentially proximate motifs. For all interactions, and separately for each motif type, the percentage of interactions where residue *i* and *j* are in a given region of the Ramachandran plot are shown, listed in order of frequency. In brackets, the most common *i, i+1* and *j, j+1* combinations are given. The preferences of *i* and *j* for different DSSP annotations are shown. 'na' indicates that DSSP has made no assignment; 'h-b turn' refers to hydrogen-bonded turn. The percentages of different patterns of local hydrogen-bonding for *i* and *j* are also shown. Each preference is only shown if it makes up greater than 10% of total for each motif type.

Motif id (N)	<i>i</i> Rama (region: %)	<i>j</i> Rama (region: %)	<i>i</i> DSSP (%)	<i>j</i> DSSP (%)	H-bonds to <i>i</i> (%)	H-bonds to <i>j</i> (%)
All (12850)	1: 48% (11: 27%) (12: 19%) 2: 44% (22: 34%)	1: 64% (11: 45%) (12: 16%) 2: 26% (22: 11%) (21: 10%)	na: 26% α-helix: 22% β-sheet: 22% h-b turn: 11% bend: 10%	na: 28% β-sheet: 23% bend: 21% h-b turn: 12%	000: 43% 010: 32% 001: 18%	000: 74% 010: 13% 001: 12%
1 (503)	1: 70% (12: 65%) 2: 26% (22: 23%)	2: 94% (22: 91%)	na: 57% bend: 13% h-b turn: 11%	α-helix: 78% h-b turn: 16%	001: 91%	000: 100%
2 (1142)	2: 49% (22: 43%) 1: 42% (12: 29%) (11: 10%)	1: 92% (11: 57%) (12: 29%)	h-b turn: 33% na: 29% bend: 13%	bend: 62% na: 21%	000: 92%	000: 98%
3 (1115)	2: 65% (22: 64%) 1: 32% (12: 28%)	2: 96% (23: 37%) (21: 31%) (22: 26%)	α-helix: 53% na: 25%	h-b turn: 56% 3/10 helix: 32%	010: 49% 001: 21% 011: 18% 000: 12%	000: 100%
4 (1492)	2: 51% (22: 47%) 1: 44% (12: 29%)	1: 97% (11: 67%) (12: 28%)	na: 30% α-helix: 22% 3/10 helix: 20%	na: 37% bend: 36% β-sheet: 20%	010: 67% 000: 32%	000: 89% 010: 10%
5 (359)	2: 66% (22: 64%) 1: 32% (12: 26%)	1: 83% (12: 54%) (11: 28%) 2: 16%	α-helix: 42% na: 19% h-b turn: 14%	na: 41% bend: 33% h-b turn: 10%	000: 57% 010: 36%	000: 94%
6 (209)	2: 70% (22: 68%) 1: 29% (12: 26%)	3: 67% (31: 59%) 2: 30% (23: 15%) (22: 11%)	α-helix: 65% β-sheet: 15% na: 12%	na: 35% h-b turn: 31% bend: 25%	010: 65% 000: 13% 001: 12%	000: 100%
7 (458)	2: 66% (22: 62%) 1: 30% (12: 18%)	1: 88% (11: 55%) (12: 29%) 4: 11%	α-helix: 52% na: 21%	na: 53% bend: 27% β-sheet: 10%	110: 39% 010: 30% 100: 16% 000: 15%	000: 87% 010: 13%
8 (100)	2: 86% (22: 80%) 1: 10%	2: 78% (22: 42%) (21: 33%) 3: 10%	α-helix: 57% h-b turn: 16% 3/10 helix: 10% na: 10%	na: 42% h-b turn: 24% α-helix: 15% bend: 16%	010: 57% 000: 42%	000: 99%
9 (522)	2: 88% (22: 87%) 1: 11%	1: 96% (11: 78%) (12: 17%)	α-helix: 81%	na: 74% bend: 12%	010: 55% 110: 22% 000: 14%	000: 91%

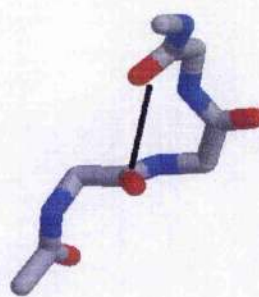




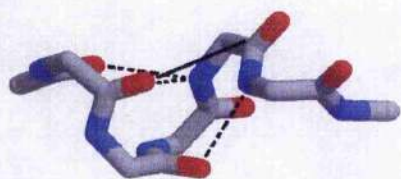
Motif 1



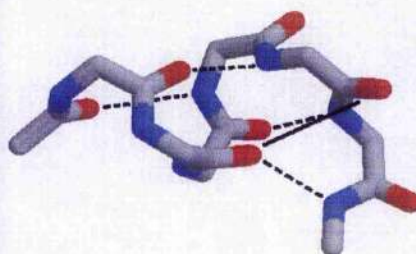
Motif 2a



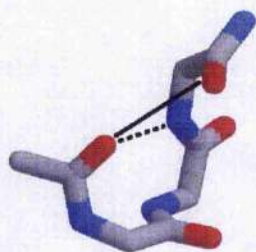
Motif 2b



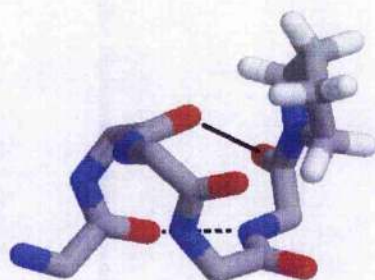
Motif 3a



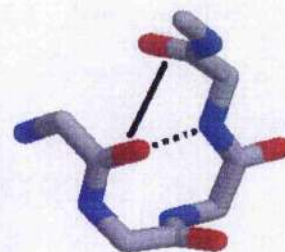
Motif 3b



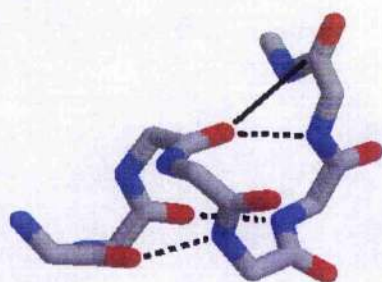
Motif 4



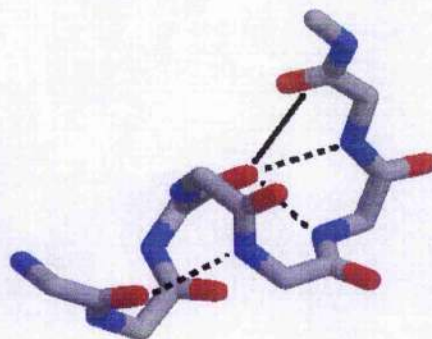
Motif 5a



Motif 5b

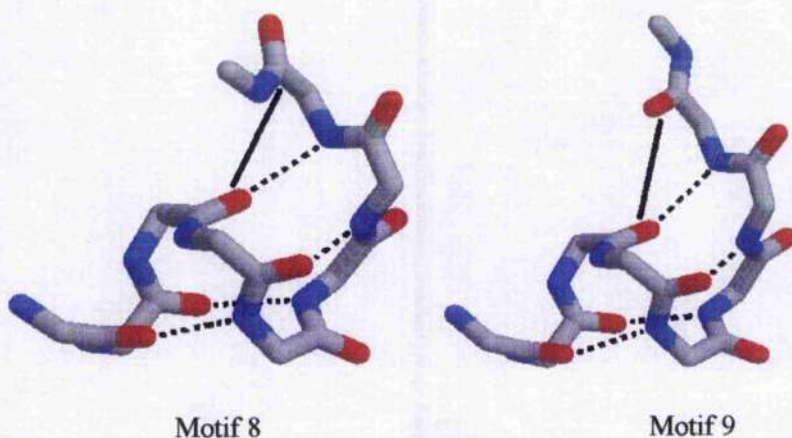


Motif 6



Motif 7

Figure continued...



**Figure 6.10.** Examples of sequentially proximate carbonyl-carbonyl interaction motifs. Each interaction is indicated by a continuous black line. Dashed lines indicate hydrogen-bonds. Only main-chain atoms are shown (except for the proline side-chain in Motif 5a). The motif type, PDB id, and residue numbers for the interacting carbonyl groups are as follows: Motif 1, from 1d2n, between carbonyl groups of residues 585 and 587. Motif 2a, 1luc chain B, 194 and 196. Motif 2b, 1hpm, 90 and 92. Motif 3a, 153l, 38 and 41. Motif 3b, 1hxn, 427 and 430. Motif 4, 1gai, 296 and 299. Motif 5a, 1toa chain A, 153 and 156. Motif 5b, 19hc chain A, 285 and 288. Motif 6, 4xis, 79 and 83. Motif 7, 1qnf, 24 and 28. Motif 8, 2gar, 183 and 188. Motif 9, 1cb0, 156 and 161.



## Sequentially distant interactions

The following text describes motifs 10 to 15 (examples shown in Figure 6.11), summarising Tables 6.5 and 6.6.

Motifs 10 to 13 are dominated by interactions that occur near the ends of  $\beta$ -sheets. It should be recalled that interactions between residues in the same secondary structure element are eliminated at an earlier stage of the analysis. Secondary structure elements are defined by hydrogen-bonding patterns. The  $\beta$ -sheet interactions that dominate motifs 10-13 are cases where one of the two interacting residues is not considered part of the  $\beta$ -sheet, because the hydrogen-bonding pattern ends before or at that residue. The examples shown in Figure 6.11 all occur at the ends of  $\beta$ -sheets. It can be seen in each case that the  $\beta$ -sheet hydrogen-bonding pattern ends just above or below the carbonyl-carbonyl interaction. They can be compared with the examples of  $\beta$ -sheet hydrogen-bonding shown in Figure 6.9a.

More than 50% of occurrences of motifs 10-13 do not match the  $\beta$ -sheet hydrogen-bonding criteria shown in Figure 6.9a, yet they share the geometry of occurrences that do. These are often cases where the  $\beta$ -sheet hydrogen-bonding pattern has been disrupted at the carbonyl-carbonyl interaction. There are also occurrences unrelated to  $\beta$ -sheets. These inhabit a varied range of secondary structures, and no distinct types could be isolated for separate examination.

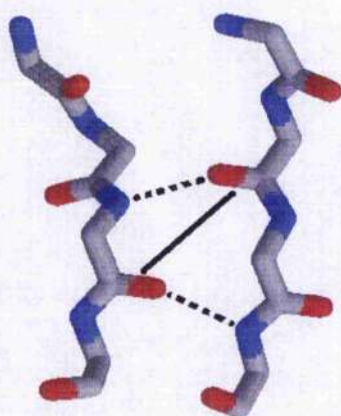
Motifs 14 and 15 represent a range of interactions where one of the nitrogen-to-oxygen distances is relatively short. These cases are often very weak hydrogen bonds, which have fallen outside the criteria used for hydrogen-bond definition. They inhabit a large range of different Ramachandran plot geometries and secondary structure descriptors, with no distinct correlations emerging from the data. For this reason no examples of motifs 14 and 15 are shown in Figure 6.11.

**Table 6.5.** Geometry of sequentially distant motifs. The Motif id, Sequence separation, and numbers (N), of each motif are shown, along with the averages and standard deviations of the carbonyl-carbonyl interaction energy (in kcal/mol), the  $T_{\text{cccc}}$  torsion angle, and the angles A1-A4 in degrees. \*The Outliers in  $T_{\text{cccc}}$  were eliminated so that the distribution was restricted to a range of no more than  $180^\circ$ , to allow calculation of the mean and R bar values for circular data (Allen & Johnson, 1991). <sup>§</sup>Distributed across the full range of  $T_{\text{cccc}}$ , with peaks at  $0^\circ$  and  $180^\circ$ , so that the range could not be restricted to within  $180^\circ$ .

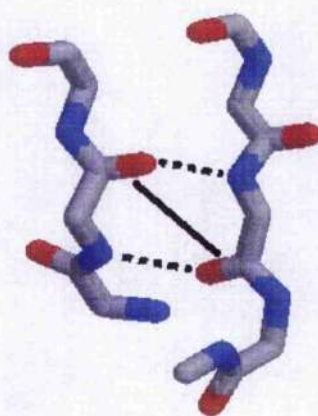
Motif id	Seq. sep.	N	Energy (sd)	$T_{\text{cccc}}^*$ (R bar)	A1 (sd)	A2 (sd)	A3 (sd)	A4 (sd)
10	$\geq 6$	1060	0.02 (0.7)	28.7 (0.8)	124.2 (14)	52.5 (16)	124.0 (18)	51.7 (11)
11	$\geq 6$	1253	0.09 (0.6)	32.6 (0.8)	125.3 (14)	51.2 (12)	124.6 (14)	50.7 (11)
12	$\geq 6$	1080	0.02 (0.7)	35.9 (0.8)	121.9 (19)	50.5 (13)	124.5 (15)	53.3 (16)
13	$\geq 6$	1369	0.08 (0.8)	28.2 (0.8)	125.2 (14)	51.0 (14)	124.7 (15)	50.6 (12)
14	$\geq 6$	598	-1.28 (0.7)	§	139.6 (18)	117.9 (26)	51.2 (23)	32.3 (18)
15	$\geq 6$	580	-1.24 (0.7)	§	55.7 (26)	32.9 (17)	140.2 (18)	113.6 (29)

**Table 6.6.** Local secondary structure of sequentially distant motifs. For all interactions, and separately for each motif type, the percentage of interactions where residue *i* and *j* are in a given region of the Ramachandran plot are shown, listed in order of frequency. In brackets, the most common *i*, *i*+1 and *j*, *j*+1 combinations are given. The preferences of *i* and *j* for different DSSP annotations are shown. 'na' indicates that DSSP has made no assignment; 'h-b turn' refers to hydrogen-bonded turn. The percentages of different patterns of local hydrogen-bonding for *i* and *j* are also shown. Each preference is only shown if it makes up greater than 10% of total for each motif type. The percentages of motifs 10-13 with the same local hydrogen-bonding as found in  $\beta$ -sheet are indicated.

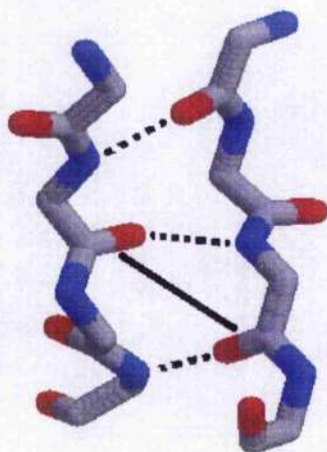
Motif id (N)	<i>i</i> Rama (region: %)	<i>j</i> Rama (region: %)	<i>i</i> DSSP (%)	<i>j</i> DSSP (%)	H-bonds to <i>i</i> (%)	H-bonds to <i>j</i> (%)
10 (1060)	1: 66% (11: 55%) 2: 25% (22: 13%) (21: 10%)	1: 51% (11: 42%) 2: 35% (21: 23%) 3: 11%	$\beta$ -sheet: 47% na: 20%	na: 29% $\beta$ -sheet: 26% h-b turn: 15% bend: 15%	001: 48% 000: 45%	000: 49% 010: 48% $\beta$ -sheet-like hydrogen-bonding (001/010): 32%
11 (1253)	1: 71% (11: 67%) (12: 10%) 2: 22% (21: 10%) (22: 10%)	1: 76% (11: 59%) (12: 13%) 2: 22%	$\beta$ -sheet: 49% na: 18%	$\beta$ -sheet: 52% na: 21%	010: 58% 000: 38%	010: 52% 000: 43% $\beta$ -sheet-like hydrogen-bonding (001/010): 37%
12 (1080)	1: 55% (11: 46%) 2: 35% (21: 19%) (22: 14%)	1: 75% (11: 60%) (12: 10%) 2: 16%	na: 33% $\beta$ -sheet: 25% bend: 16% h-b turn: 11%	$\beta$ -sheet: 50% na: 21% bend: 10%	000: 47% 010: 44%	000: 44% 001: 42% $\beta$ -sheet-like hydrogen-bonding (001/010): 28%
13 (1369)	1: 63% (11: 55%) 2: 28% (21: 19%)	1: 72% (11: 64%) 2: 19% (21: 13%)	$\beta$ -sheet: 41% na: 25% bend: 11%	$\beta$ -sheet: 44% na: 33% bend: 10%	001: 60% 000: 33%	001: 61% 000: 36% $\beta$ -sheet-like hydrogen-bonding (001/010): 46%
14 (598)	1: 46% (11: 30%) (12: 12%) 2: 43% (22: 20%) (21: 17%)	1: 47% (12: 23%) (11: 18%) 2: 38% (22: 19%) (21: 16%)	na: 25% bend: 22% $\beta$ -sheet: 22% h-b turn: 16%	na: 39% bend: 18% $\beta$ -sheet: 14% h-b turn: 13% $\alpha$ -helix: 10%	000: 61% 001: 23% 010: 16%	000: 97%
15 (580)	1: 50% (11: 22%) (12: 20%) 2: 36% (12: 20%) (21: 11%)	1: 57% (11: 40%) (12: 10%) 2: 27% (21: 13%) 3: 12%	na: 35% $\beta$ -sheet: 21% bend: 16% $\alpha$ -helix: 13%	$\beta$ -sheet: 30% na: 28% bend: 18% h-b turn: 14%	000: 90%	000: 69% 001: 19% 010: 12%



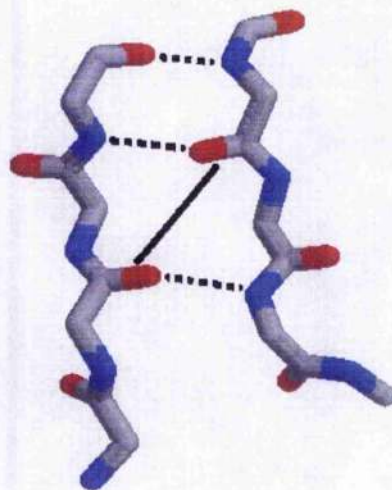
Motif 10



Motif 11



Motif 12



Motif 13

**Figure 6.11.** Examples of sequentially distant carbonyl-carbonyl interaction motifs. Each interaction is indicated by a continuous black line. Dashed lines indicate hydrogen-bonds. Only main-chain atoms are shown. The motif type, PDB id, and residue numbers for the interacting carbonyl groups are as follows: Motif 10, 1din, 30 and 116. Motif 11, 1aay, 135 and 144. Motif 12, 1mgt, 110 and 145. Motif 13, 1koe, 142 and 304.

### 6.3.4 Carbonyl-carbonyl interactions at the helical C-terminus

A surprising finding is that seven of the fifteen motifs identified are mainly found at the C-termini of  $\alpha$ -helices. These motifs (3 through to 9) occur at separations of three, four, and five, residues. Closer inspection of the geometry shows similarity between some of them. Motif 5 is different from the rest, but motifs 3, 6, and 8 are similar to each other, as are motifs 4, 7, and 9. This similarity is best demonstrated by comparison of the ratio of the  $C^3 \cdots O^2$  distance to the  $O^2 \cdots O^4$  distance, shown in Table 6.7. The low ratio observed for motifs 3, 6, and 8, shows a relative proximity of the carbon to the oxygen, whereas with motifs 4, 7, and 9, the oxygen atoms are relatively close. As shown in Table 6.3., motifs 3, 6, and 8, are, on average, favourable, and 4, 7, and 9, are slightly favourable or unfavourable.

Ignoring motif 5, it appears that, for each residue separation from three to five, there are two main conformations that can be adopted by carbonyl-carbonyl interactions at this helical C-terminus location. An informative way to consider this is to look at residues  $i$  and  $j$  in motif 8 (Figure 6.10. motif 8.). The structure in which this motif occurs also includes examples of motifs 6 and 3. The interaction between the carbonyl group of residue  $i$  with that of residue  $j-1$  in motif 8 is an example of motif 6 and, likewise, the interaction between the carbonyl of  $i$  with that of  $j-2$  is an example of motif 3. As an approximation, a  $180^\circ$  rotation of the  $\psi$  angle of residue  $j$  (taking it from the  $\alpha_r$  region of the Ramachandran plot to the  $\beta$  region) transforms motif 8 into motif 9. Likewise, a 'flip' of residue  $j-1$  transforms motif 6 into motif 7, and of residue  $j-2$  transforms motif 3 into motif 4.

Motifs 8 and 9 are both examples of the Schellman loop (Schellman, 1980; Milner-White, 1988). This common C-terminal capping conformation is defined by the presence of two hydrogen-bonds (an  $i, i-5$ , and an  $i, i-3$ ) at the end of the helix. Aurora & Rose (1998) studied hydrophobic interactions at C-terminal caps. They defined the last residue of the helix (the final residue to make an  $i, i-4$  hydrogen-bond) as Ccap, subsequent residues as  $C'$ ,  $C''$ ,  $C'''$ , etc., and preceding residues as  $C1$ ,  $C2$ ,  $C3$ , etc., in the order  $\cdots C3-C2-C1-Ccap-C'-C''-C''' \cdots$ . They identified a common hydrophobic interaction in the Schellman loop that occurs between  $C''$  and  $C3$ . Using their nomenclature, the

carbonyl-carbonyl interaction shown in motifs 8 and 9 occurs between residues C'' and C3. Therefore, it is likely the hydrophobic interaction described by Aurora & Rose (1998) often incorporates a favourable carbonyl-carbonyl interaction, and that this contributes to the stability of Schellman loops.

Aurora & Rose (1998) also identified cases where the helix terminates with a proline at position C'. The favourable carbonyl-carbonyl interaction in motif 5a occurs between residues Ccap and C3, and contributes to the stability of proline-terminated helices.

**Table 6.7.** Ratio of C<sup>3</sup>...O<sup>2</sup> distance to O<sup>2</sup>...O<sup>4</sup> distance for motifs prevalent at helix C-termini. Sequence separation, average ratio, and standard deviation of the ratio are shown for each. The residue separation is also shown.

Motif	Sequence separation	C <sup>3</sup> ...O <sup>2</sup> / O <sup>2</sup> ...O <sup>4</sup> (sd)
3	3	0.81 (0.02)
4	3	1.18 (0.00)
5	3	0.95 (0.07)
6	4	0.80 (0.03)
7	4	1.15 (0.00)
8	5	0.85 (0.08)
9	5	1.17 (0.07)

## 7 Mimicry of $\beta$ -turns by asx- and ST-turns

### 7.1 Brief introduction and outline

Hydrogen-bonded  $\beta$ -turns in proteins occur in four categories: type I (the commonest), type II, type II' and type I'. Asx-turns resemble these  $\beta$ -turns, in that both have an N-H...O=C hydrogen bond forming a ring of 10 atoms. Serine and threonine side chains also commonly form hydrogen-bonded turns, herein referred to as ST-turns. Asx-turns have previously been categorized into four classes and ST-turns into three categories, based on side chain rotamer types and the conformation of the central residue of each turn. Here it is shown that the four classes of asx-turn are geometrically equivalent to the four types of hydrogen-bonded  $\beta$ -turn, and that the three categories of ST-turn are geometrically equivalent to three of the four types of hydrogen-bonded  $\beta$ -turn. Almost all type II asx-turns occur in the form of a recently-described 3-residue feature named an asx-nest.

## 7.2 Materials and methods

The Relibase<sup>†</sup> knowledge base (Bergner *et al.*, 2001-2002; Gunther *et al.*, 2003; Hendlich *et al.*, 2003) was used for identification and geometric analysis of asx- and ST-turns within the dataset of 500 proteins described in section 6.2.2. Statistical analysis of the turns was carried out using the Vista program (CCDC, 1994; Bruno *et al.*, 2002).

Data regarding dihedral angle distributions of hydrogen-bonded  $\beta$ -turns (Figures 7.1a and 7.2b) was obtained from a dataset of 135 PDB files (resolution  $< 1.5\text{\AA}$ , R-factor  $< 20\%$ , chain identity  $< 25\%$ ) based on that used in a previous study (Watson & Milner-White, 2002a). The PDB codes, with the letter of the selected subunit appended when appropriate, are: 1a2yb, 1ah7, 1aie, 1amm, 1benb, 1bgf, 1brt, 1byqa, 1c52, 1ckaa, 1cse, 1ezm, 1g3p, 1luca, 1msi, 1opd, 1plc, 1poa, 1rie, 1rpo, 1whi, 1xyza, 2arob, 3vub, 1c5ea, 1czpa, 1dbfa, 1dj0a, 1dqza, 1ezga, 1fm0d, 1fm0e, 1fsga, 1g7aa, 1g7a, 1gk8a, 1gk8i, 1gk9a, 1gk9b, 1heta, 1i0ha, 1i4fa, 1i4ua, 1ijva, 1isua, 1jz8a, 1k0ma, 1k20a, 1k55a, 1lkka, 1psra, 1qh4a, 1qopa, 1qopb, 1qtna, 1qtnb, 1rgea, 1sgpi, 1swua, 3chbd, 1a2pa, 1a6m, 1aho, 1atg, 1b0y, 1bkr, 1byi, 1c0p, 1c7k, 1cc8, 1cex, 1ctj, 1cxq, 1cy5, 1cyo, 1d4o, 1d5t, 1dcs, 1ds1, 1e29, 1e58, 1e6u, 1eu1, 1euw, 1fle, 1f94, 1fcy, 1fo8, 1fw9, 1fye, 1g2b, 1g6x, 1hbz, 1hlr, 1hnj, 1i27, 1i2t, 1i40, 1i71, 1i8o, 1ifc, 1iqq, 1j98, 1j9b, 1jbe, 1jfb, 1jg1, 1jhg, 1jk3, 1kb0, 1koi, 1mun, 1nls, 1pa2, 1qg8, 1qj4, 1qqf, 1rb9, 1rhs, 2end, 2eng, 2erl, 2lis, 2mcm, 2nlr, 3ebx, 3eug, 3lzt, 3pyp, 3seb, 3sil and 9rnt.

Because of their preference for staggered rather than eclipsed conformations,  $\chi_1$  angles have a tendency to cluster in the region of either  $180^\circ$ ,  $60^\circ$  or  $-60^\circ$ . Lovell *et al.* (2000) have pointed out that different authors have used the gauche+/gauche- angle descriptors to refer to either  $60^\circ/-60^\circ$  or  $-60^\circ/60^\circ$ , which has resulted in considerable confusion. Here, as suggested by these authors, the descriptors are avoided by simply referring to the values.

Hydrogen-bonds were defined by the criteria described in section 3.2.1. Asx- and ST-turns are identified by a single hydrogen-bond, that between the side chain oxygen of an aspartate, asparagine, serine or threonine residue (i) and the main chain NH group of a residue two ahead (i+2). Beta-turns are identified by a single hydrogen bond between the



main chain CO of residue  $i$  and the main chain NH of residue  $i+3$ . Dihedral angles used are defined in Figure 2.3. In asx- and ST-nests the first residue is D, N, S or T and its side chain oxygen atom is within 4.0 Å of at least two of the main chain nitrogen atoms of residues  $i$ ,  $i+1$  or  $i+2$ .

## 7.3 Results & Discussion

### 7.3.1 Asx-turn and ST-turn identification

Both asx-turns and  $\beta$ -turns consist of a 10-atom hydrogen-bonded ring closed by a N-H $\cdots$ O=C hydrogen-bond. ST-turns are similar but the ring is 9-membered. Figure 2.3 provides a comparison of the relevant torsion angles of the two types of turn. To facilitate this, a new set of angles called  $\omega_e$ ,  $\phi_e$  and  $\psi_e$  are defined. These are angles in the asx- or ST-turn equivalent to corresponding angles in the  $\beta$ -turn. Their approximate relationship to the standard angles is given in Figure 2.3f, though values given throughout this work have been measured directly. If there is mimicry, the angles  $\phi_e$  and  $\psi_e$  in asx- and ST-turns should resemble  $\phi$  and  $\psi$  of residue  $i+1$  in  $\beta$ -turns, while  $\phi$  and  $\psi$  of residue  $i+1$  of asx- and ST-turns should resemble  $\phi$  and  $\psi$  of residue  $i+2$  of  $\beta$ -turns.

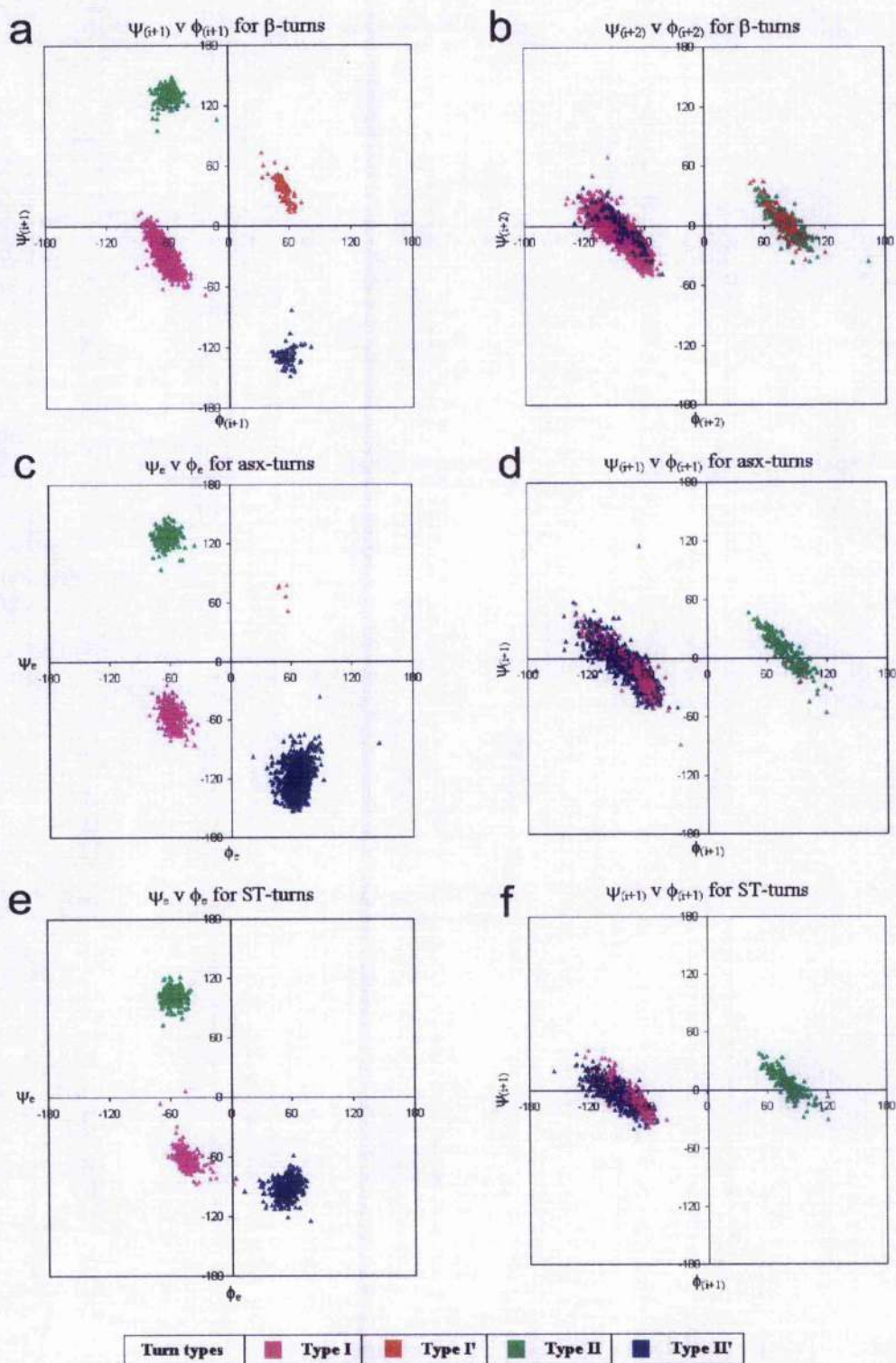
The database of 500 proteins contains 24,786 asparagine, aspartate, serine or threonine residues. Table 7.1 shows the numbers of each and of those involved in asx- or ST-turns.

**Table 7.1.** Proportions of residues involved in  $\beta$ -turn mimics. The total numbers of aspartate, asparagine, serine, and threonine residues in the dataset are given, along with the numbers and percentages of each that form asx- and ST-turns (i.e. occupy position  $i$  of an asx- or ST-turn).

Residue type	Asn	Asp	Ser	Thr	N/D/S/T
Number in database	5096	6617	6728	6345	24786
Number forming asx- or ST- turns	569	1071	491	367	2498
Percentage forming asx- or ST- turns	11.2	16.2	7.3	5.8	10.1

### 7.3.2 Comparison with $\beta$ -turn geometry

Figure 7.1 (c and e) shows that the  $\phi_e, \psi_e$  values for asx- and ST- turns cluster into four groups placed symmetrically on the Ramachandran plot, making allowance for one of the groups being sparsely populated. In Figure 7.1 (d and f) the  $\phi, \psi$  values for residue  $i+1$  are plotted; here two clusters are observed, which is consistent with the four groupings. Finding four groups leads to a consideration of the extent to which they are related to the four known groups of  $\beta$ -turns. The equivalent  $\beta$ -turn torsion angle distributions are shown in Figure 7.1 (a and b), and the similarity with asx- and ST-turn distributions can be observed. Torsion angle values were used to group the asx- and ST-turns into types equivalent to the four types of hydrogen-bonded  $\beta$ -turn. The data points in Figure 7.1 have been coloured accordingly, to allow comparison of distributions of different types across the asx-, ST- and  $\beta$ -turns.



**Figure 7.1.** Distributions of  $\phi$ ,  $\psi$  torsion angles of residues  $i+1$  and  $i+2$  of  $\beta$ -turns, compared with equivalent angles of asx- and ST-turns. Turns of different types are represented by data points of different colours, as indicated by the legend. (a) and (b) show  $\psi$  versus  $\phi$  for residues  $i+1$  and  $i+2$ , respectively, of  $\beta$ -turns. (c) and (d) are the asx-turn equivalent angles, and (e) and (f) are the ST-turn equivalent angles, to the  $\beta$ -turn angles shown in (a) and (b). (c) shows  $\psi_e$  versus  $\phi_e$  for asx-turns, (d) shows  $\psi$  versus  $\phi$  for residue  $i+1$  of asx-turns, (e) shows  $\psi_e$  versus  $\phi_e$  for ST-turns, and (f) shows  $\psi$  versus  $\phi$  for residue  $i+1$  of ST-turns.

Table 7.2 provides a comparison of the average values of the  $\phi$ ,  $\psi$  angles of residues  $i+1$  and  $i+2$  of types I, I', I and II'  $\beta$ -turns with those of the corresponding  $\phi_e$ ,  $\psi_e$  angles, and residue  $i+1$   $\phi$ ,  $\psi$  angles, of asx- and ST-turns. It is evident that the four groups of asx- and ST-turns have average torsion angle values that resemble those of  $\beta$ -turns. The average values of asx-turns show some divergence from those of  $\beta$ -turns, but are close enough that the similarity is striking. Slight differences are observed between the ST-turns and  $\beta$ -turns. Most notably, the average  $\psi_e$  angles of the type II and II' ST-turns differ from  $\psi$  of the type II and II'  $\beta$ -turns by  $18^\circ$  and  $27^\circ$  respectively. The apparent similarity suggests that it is appropriate that the nomenclature for the four  $\beta$ -turn types be used for asx- and ST-turns.

**Table 7.2.** Comparing torsion angles between  $\beta$ -turns and their mimics. Numbers ( $N$ ) and average torsion angles of four different types of asx- and ST-turns are given, along with ideal  $\beta$ -turn torsion angle values (Lewis *et al.*, 1973; Wilmot & Thornton, 1990). Ideal torsion angle values of  $\beta$ -turns and values of equivalent torsion angles of asx- and ST-turns share the same columns. The shorter side-chains of ST-turns do not have  $\omega_e$  angles.

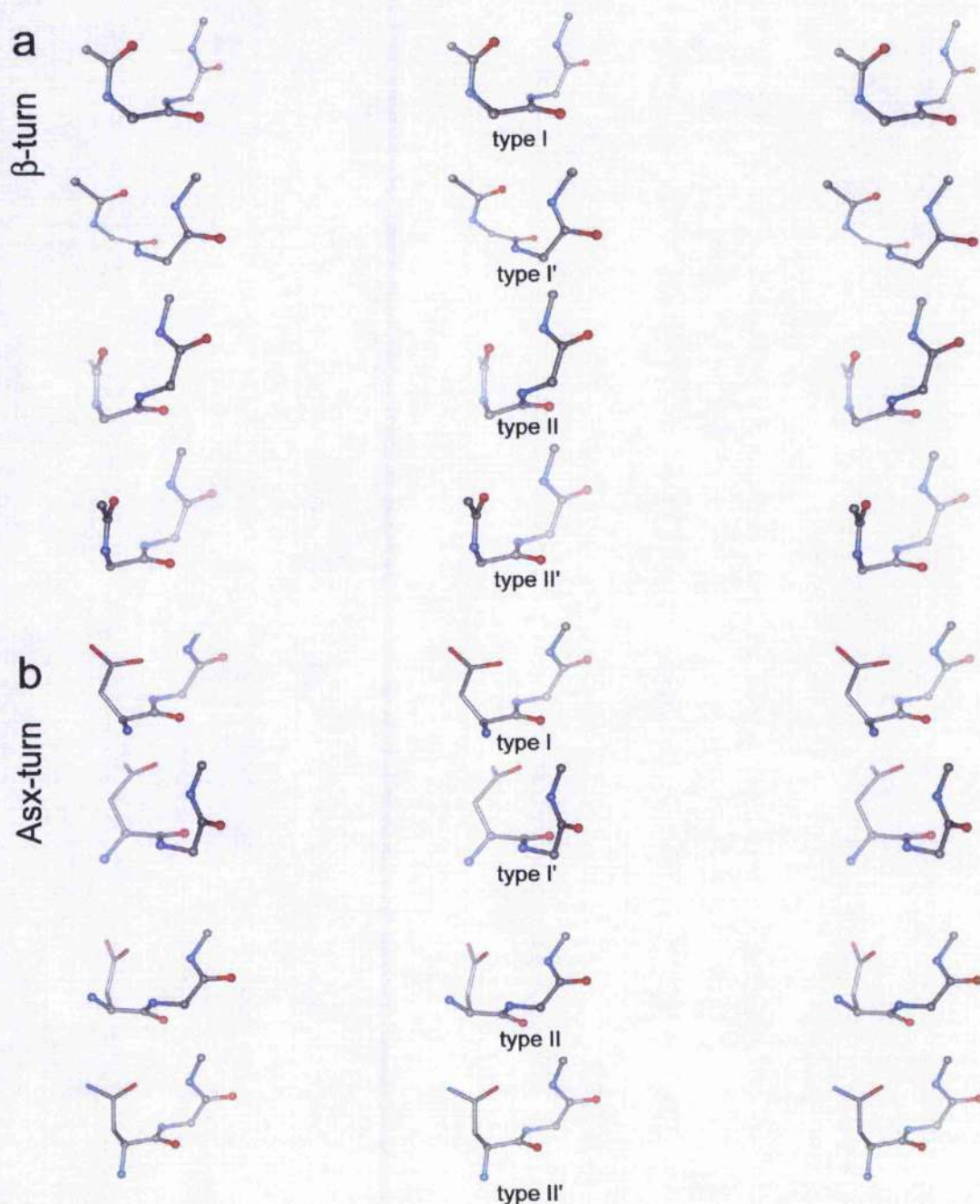
Turn type	$N$	Torsion angle values									
		$\omega$	$\omega_e$	$\phi$	$\phi_e$	$\psi$	$\psi_e$	$\phi_{(i+2)}$	$\phi_{(i+1)}$	$\psi_{(i+2)}$	$\psi_{(i+1)}$
Type I $\beta$ -turn		180		-60		-30		-90		0	
Type I asx-turn	302		-143		-58		-55		-72		-13
Type I ST-turn	205		N/A		-44		-62		-75		-9
Type I' $\beta$ -turn		180		60		30		90		0	
Type I' asx-turn	5		102		53		70		98		-18
Type II $\beta$ -turn		180		-60		120		80		0	
Type II asx-turn	203		-176		-62		128		80		5
Type II ST-turn	207		N/A		-57		102		80		7
Type II' $\beta$ -turn		180		60		-120		-80		0	
Type II' asx-turn	1130		155		65		-123		-73		-14
Type II' ST-turn	445		N/A		53		-93		-89		-5

The numbers of different types of asx- and ST-turns in the database are shown in Table 7.2. In  $\beta$ -turns the commonest occurrence is of type I. This turn resembles a 3/10-helix with only one hydrogen-bond. In asx- and ST-turns the commonest is type II' while the next most common is type I. The structures of type II asx- and ST-turns are the same as those previously identified as asx-nest or ST-nest motifs (Watson & Milner-White, 2002a). The tight clustering of  $\phi_e$  and  $\psi_e$  angles in type II asx- and ST-turns, as observed in Figure 7.1, may result from the nest being conformationally constrained. One factor allowing asx-turns (but not ST-turns) to exhibit minor conformational differences from  $\beta$ -turns is the fact that  $\omega_e$ , corresponding to  $\omega$  of residue  $i+1$  of  $\beta$ -turns, is a side chain angle ( $\chi_2$ ), so is not restrained to the *trans* conformation of  $\omega$ . The flexibility of  $\omega_e$  provides extra conformational possibilities for asx-turns compared to  $\beta$ -turns.

Figure 7.2 shows examples of the four different types of hydrogen-bonded  $\beta$ -turn structures (Figure 7.2a), along with examples of the equivalent asx-turn structures (Figure 7.2b). The geometric similarity of each type of  $\beta$ -turn with the equivalent asx-turn can be observed easily from visual inspection.

The average O $\cdots$ H distances in the defining hydrogen bonds of  $\beta$ -turns and their mimics differ significantly. The shortest are those of the asx-turns (2.09Å; sd 0.17), then  $\beta$ -turns (2.16Å; sd 0.16), and the longest are the ST-turns (2.2Å; sd 0.15). It is likely asx-turn hydrogen bonds are shorter than those of  $\beta$ -turns because of the extra flexibility of the  $\omega_e$  angle. ST-turns have 9 instead of 10 atoms per hydrogen-bonded ring. This may confer a limitation on the allowed proximity of the hydrogen-bonding atoms of these turns.





**Figure 7.2.** The four types of hydrogen-bonded  $\beta$ -turns and equivalent asx-turns. Structures are represented in relaxed-eye (first and second images) and crossed-eye (second and third images) stereo. Main chain atoms are in ball and stick, with side chain atoms in stick. (a) Examples of type I, I', II and II'  $\beta$ -turns. (b) Examples of equivalent asx-turns. Details of the examples chosen are listed, in order, as residue number, PDB code and chain (if named): asx-turn type I: 12-14, 1mro b (Ermler *et al.*, 1997);  $\beta$ -turn, type I: 7-10 1g3p; asx-turn type I': 62-64 1cse e (Bode *et al.*, 1987);  $\beta$ -turn type I': 44-47 1sgp i; asx-turn type II: 93-95 1k55 a (Golemi *et al.*, 2001);  $\beta$ -turn type II: 345-348 1d5t; asx-turn type II': 29-31 1f94 (Kuhn *et al.*, 2000);  $\beta$ -turn type II': 598-601 1jz8 a. Although not illustrated, it may be useful to list some typical ST-turns: type I: 14-16 1fw9 a; type II: 234-236 1gk9 b; type II': 296-298 1mro b. More examples of asx- and ST-turns of different types have been shown previously (Aurora & Rose, 1998; Wan & Milner-White, 1999a,b; Eswar & Ramakrishnan, 1999).



## 8 Conclusions

### 8.1 Hydrogen-bonding between *trans*-amide groups

#### 8.1.1 Factors affecting hydrogen bond geometry

The average value of the  $\text{H}\cdots\text{O}=\text{C}$  angle in *trans*-amide $\cdots$ *trans*-amide hydrogen-bonds is  $149^\circ$ . This value resembles that between *trans*-amides in proteins, but is  $15^\circ$  higher than that of the more general  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  hydrogen-bond ( $134^\circ$ ) in small molecules. The results of chapter 3 show that three factors contribute to this difference: multiple hydrogen-bonding (where the carbonyl oxygen accepts more than one hydrogen-bond); hydrogen-bonded ring motifs (where an  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  hydrogen-bond forms part of a ring containing seven or eight atoms and two hydrogen-bonds); and hydrogen-bonded chains (where hydrogen-bonding between *trans*-amide groups adjacent to one another in the crystal structure extends repetitively across the lattice). When comparing only those  $\text{C}=\text{O}$  groups that make a single hydrogen-bond, and with ring motifs excluded from the  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  hydrogen-bonds, and hydrogen-bonded chains excluded from the *trans*-amide $\cdots$ *trans*-amide hydrogen-bonds, the difference in average  $\text{H}\cdots\text{O}=\text{C}$  becomes just  $5^\circ$ .

Figure 3.3 summarises many of the key points of the work. It shows that, in spite of the  $\text{H}\cdots\text{O}=\text{C}$  variations,  $\text{N}-\text{H}\cdots\text{O}$  varies comparatively little and its average value remains high at around  $160^\circ$ . This is consistent with a tendency for these three atoms to approach linearity. This preference emerges especially in the case of *trans*-amides compared to most other hydrogen bond systems because of their single conformationally fixed NH group (e.g. in hydroxyls the hydrogen can rotate;  $\text{NH}_2$  groups have two hydrogens).

Some explanation for the effects of these three factors can be offered. There are obvious steric reasons why multiple hydrogen-bonds will cause reduction in  $\text{H}\cdots\text{O}=\text{C}$ , since each additional hydrogen must find space to approach the oxygen acceptor atom. Examination of the ring motifs seen with *cis*-amides suggests that in this case the linear  $\text{N}-\text{H}\cdots\text{O}$  angles constrain the  $\text{H}\cdots\text{O}=\text{C}$  angles to be low although there may be an extra favourable effect

due to the hydrogen being attracted to the oxygen's lone pair. In hydrogen-bonded chains of *trans*-amides we expect to see high  $\text{H}\cdots\text{O}=\text{C}$  values because most of them only involve single, rather than multiple, hydrogen-bonds. However, when the set occurring in hydrogen-bonded chains is divided into those with multiple hydrogen-bonding and those with only a single hydrogen-bond, the average  $\text{H}\cdots\text{O}=\text{C}$  is higher in both cases than that for single or multiply hydrogen-bonded occurrences that are not in hydrogen-bonded chains. This observation is poignant as the majority of hydrogen-bonds in proteins occur within hydrogen-bonded chains that run along helices and across  $\beta$ -sheets. Small molecule *trans*-amides might be supposed to have more freedom to adopt preferred hydrogen-bond geometries than *trans*-amides in protein secondary structure, so it is interesting that the small molecule geometry resembles that in proteins.

Uncertainty persists as to the nature of hydrogen bonding. One idea is that the hydrogen atom bonds to a lone pair of the oxygen atom acceptor, giving rise to  $\text{H}\cdots\text{O}=\text{C} = 120^\circ$ , while the other (Mitchell & Price, 1989, 2000; Apaya *et al.*, 1997) emphasizes electrostatic interactions between partially charged atoms. This approach gives rise to a broad electrostatic potential around carbonyl oxygen atoms such that there is little energetic preference for bent (near  $120^\circ$ ) or linear ( $>150^\circ$ )  $\text{H}\cdots\text{O}=\text{C}$  interactions. The conclusion from *trans*-amide $\cdots$ *trans*-amides in the CSD is that, while  $\text{N}-\text{H}\cdots\text{O}$  tends to linearity in all situations,  $\text{H}\cdots\text{O}=\text{C}$  varies widely, ranging from  $120^\circ$  to  $180^\circ$  depending on the environment. For  $\alpha$ -helix and  $\beta$ -sheet in proteins, as in small molecule crystals, the factors just described such as hydrogen-bonded chains encourage the more linear angles.

The analysis of the geometry of the *trans*-amide $\cdots$ *trans*-amide hydrogen-bonded system shows that almost half of *trans*-amide $\cdots$ *trans*-amide hydrogen-bonds exist in a near planar arrangement (with a  $T_{\text{CNCN}}$  value within  $20^\circ$  of either  $0^\circ$  or  $180^\circ$ ). This propensity for planarity may be a consequence of crystal packing factors. About 60% of the planar systems are in the *anti* configuration ( $T_{\text{CNCN}}$  near  $180^\circ$ ) and the rest are in the *syn* configuration ( $T_{\text{CNCN}}$  near  $0^\circ$ ). It is interesting that, within the plane, those in the *anti* configuration have no distinct preference in terms of their 'shear' relative to each other, whereas those in the *syn* configuration show a clear preference for geometry that maximises the proximity of the two carbonyl groups, and thereby the strength of the

favourable electrostatic interaction between them. The *syn* configuration in proteins occurs within antiparallel  $\beta$ -sheet, and the carbonyl-carbonyl interaction present there is maximised in the same way.

Another factor, steric accessibility of the carbonyl oxygen, was addressed in chapter 4 and is described in the following section (8.1.2). If the *trans*-amide carbonyl oxygen is, on average, less accessible than other carbonyl oxygen atoms, then this could be responsible for some, or all, of the unexplained difference in average  $\text{H}\cdots\text{O}=\text{C}$ , after accounting for the factors listed above.

### 8.1.2 An additional factor: steric accessibility

The results of Figure 4.2 show that steric accessibility of the carbonyl oxygen atom has a substantial influence on the geometry of *trans*-amide...*trans*-amide hydrogen-bonds, and on N-H...O=C hydrogen-bonds in general. Figure 4.3 leads to the conclusion that, when steric accessibility is limited to low values, H...O=C angles of all N-H...O=C hydrogen-bonds (including *trans*-amide...*trans*-amide) are restricted to relatively linear values averaging near 150°, while greater steric accessibility allows lower values of H...O=C. When steric accessibility is no longer the limiting factor (i.e. at greater values of SA), there is a difference between the average H...O=C angle of *trans*-amide...*trans*-amide hydrogen-bonds and N-H...O=C hydrogen-bonds in general and the additional factors examined in chapter 3 and listed in section 8.1.1 are required to explain it. These are: the number of hydrogen-bonds accepted by the carbonyl oxygen; the formation of ring motifs among the N-H...O=C hydrogen-bonds; and the formation of hydrogen-bonded chains among the *trans*-amide...*trans*-amide hydrogen-bonds. Hydrogen-bonded chains have relatively very linear H...O=C angles, and the steric accessibility of the carbonyl oxygen is unexpectedly high. Hence, at higher values of steric accessibility, the occurrence of hydrogen-bonded chains is the most important factor in explaining the relatively linear geometry of *trans*-amide...*trans*-amide hydrogen-bonds.

Within proteins, the ratio of donors to acceptors is approximately 1:1 and as a result most carbonyl oxygens accept only one hydrogen-bond. Hydrogen-bonded chains are also common, in the form of  $\alpha$ -helices and  $\beta$ -sheets. These are not infinite such as those found in small molecule structures, but may share some of the same characteristics. It would be interesting to analyse the steric accessibility of the carbonyl oxygen atoms of the polypeptide chain, in the absence of distant residues, to establish whether there is also similarity with the *trans*-amide...*trans*-amide hydrogen-bonds in the CSD.

In the wider context, this work shows how the large quantity of geometric data in the CSD can be used effectively to answer questions regarding the relative importance of different geometric factors on the average geometry of a specific interaction. As the quantity of data

in the CSD increases, approaches of this type can be widened to analyse ever more detailed aspects of these interactions.

## 8.2 Carbonyl-carbonyl interactions between *trans*-amide groups

### 8.2.1 *Trans*-amide carbonyl-carbonyl interactions in the CSD

The most striking observation of chapter 5 is that, after exclusion of *trans*-amide...*trans*-amide hydrogen-bonds (Figure 2.1a), a similar overall proportion of *trans*-amide...*trans*-amide carbonyl-carbonyl interactions are found to occur in the three geometric motifs observed among interactions between ketone carbonyl groups. The proportions that occur in each of these motifs are also strikingly similar. The proportion of interactions featuring the antiparallel motif was 48% (cf. 49% of contacts between ketone carbonyls), and the parallel and perpendicular motifs accounted for 17% and 5%, respectively (cf. 9% and 10%, respectively, in ketones). This similarity occurs despite the replacement of one of the substituent groups of the ketone with an amine (NH) moiety in the *trans*-amide. NH groups of one *trans*-amide can form hydrogen-bonds with any other acceptors present in the structure. This would be expected to affect the proportional distribution of motif types even after elimination of contacts in which the two *trans*-amides are hydrogen-bonded to each other. The lack of a large change in the distribution suggests that the influence, on the crystal structure, of interactions between carbonyl groups, is manifested despite competition with hydrogen-bond formation by the NH substituent.

The average angles of the three motifs are similar to those found for ketone carbonyl groups. The average geometry of the antiparallel motif is almost identical (A1 and A3 are 99°, cf. 97° in ketones; A2 and A4 are 81°, cf. 84°), as is that of the perpendicular motif. Average geometry of the parallel motif shows an arrangement that is slightly more sheared (the angle made by C=O...C is 117°), than that found for ketone carbonyls (in which C=O...C is 99°). Reasons for this are speculative, but may include differences in steric hindrance or in packing arrangements. Differences in the substituents on the nitrogen of the *trans*-amide with those on the carbon of the ketone may affect steric accessibility. There may also be differences in the preferred packing arrangements of molecules containing *trans*-amides with those containing ketones.

Interactions between *trans*-amide carbonyls are, on average, longer (3.8Å) than those observed between ketone carbonyls (3.52Å). In part this is due to the preponderance of hydrogen-bonded interactions, which have a relatively fixed carbon-to-oxygen distance averaging 3.76Å. However, after exclusion of the hydrogen-bonded interactions, those remaining still have a substantially higher average of 3.67Å. This may, again, be due to differences in steric hindrance or in packing arrangements.

In cases where the *trans*-amide units are hydrogen-bonded, the interaction between their carbonyl groups is always favourable. The peptide bond between amino acid residues is a *trans* amide unit, and interactions of the type described, involving both a hydrogen-bond and carbonyl-carbonyl interaction, are ubiquitous in proteins. Their occurrence in  $\alpha$ -helices and  $\beta$ -sheets is accompanied by a quantifiable influence on the geometry of these types of secondary structure elements (Maccallum & Milner-White, 1995a,b). From the perspective of protein structure, the prevalence of these hydrogen-bonded contacts is therefore unsurprising. However, it is of interest from the same perspective to note the high proportion of non-hydrogen-bonded carbonyl-carbonyl interactions that can be classified as antiparallel, perpendicular and parallel motifs. The question arises whether, and in what abundance, these motifs occur in protein structures. This was addressed in chapter 6, and the conclusions are described in the following section (8.2.2).



## 8.2.2 *Trans*-amide carbonyl-carbonyl interactions in the PDB

Chapter 6 shows that carbonyl-carbonyl interactions, both favourable and unfavourable, are common between main-chain peptide groups in non-repetitive protein secondary structure. Using the separation in sequence between the two residues, combined with a simple geometric measure, the distance between the carbonyl oxygen of one peptide unit and the nitrogen of the other, the interactions can be categorised into distinct motifs.

Chapter 5 investigated the occurrence of carbonyl-carbonyl interactions between *trans*-amide groups in the CSD. These were found to occur with a frequency and geometry similar to that observed between ketone carbonyls in the CSD. In particular, about 50% of interactions, between both ketones and *trans*-amides (after exclusion of hydrogen-bonded cases), occurred in the antiparallel arrangement. This arrangement is noted for having energetic favourability comparable to that of a medium-strength hydrogen-bond. Main-chain protein carbonyl-carbonyl interactions (examined in chapter 6) did not often meet the criteria for the antiparallel arrangement, as described in section 5.2.2, but several of the motifs (7, 9, and all the  $\beta$ -sheet-like motifs 10 through to 13) do have similar geometric character to it. Motifs 7 and 9 are favourable and occur at the ends of  $\alpha$ -helices, therefore stabilising the recurring conformations found there. Motifs 10 to 13 are energetically neutral, therefore neither stabilising nor destabilising the  $\beta$ -sheet ends and other  $\beta$ -sheet-like conformations in which they tend to occur.

Interactions are found in many local secondary structure situations, representing the full range of conformations adopted by non-repetitive regions of protein architecture. However, most of the motifs exhibit a tendency to occur within particular hydrogen-bonded features, or within regions resembling repetitive secondary structure. Motif 1 is favourable and occurs at the N-termini of  $\alpha$ -helices and in regions resembling  $\alpha$ -helix. Motifs 2 and 4 often occur within  $\beta$ -turns. They are unfavourable and therefore will act to destabilise  $\beta$ -turns. Motifs 3 through to 9 are all commonly found at the C-termini of helices, and are all favourable, so may be an important factor in providing stability. Motifs 10 through to 13 commonly occur either at the ends of  $\beta$ -sheets or in regions resembling  $\beta$ -sheet.

The preponderance of carbonyl-carbonyl interactions at the C-termini of  $\alpha$ -helices suggests an important role for them in stabilising the conformations required to cap the helix.

Motifs 8 and 9 stabilise the common capping conformation known as the Schellman loop. These motifs are likely to represent the Schellman loop hydrophobic interaction identified by Aurora & Rose (1998). Motif 5a stabilises a common capping conformation where the helix terminates with a proline residue.

It is possible that many of the unfavourable motifs result from both carbonyl oxygen atoms hydrogen-bonding to the same hydrogen-bond donor atom, or group of atoms, such as a water molecule.

### 8.3 Mimicry of $\beta$ -turns by asx- and ST-turns

As described in chapter 7, the four types of asx-turn and three types of ST-turn were identified in a large and diverse database. Measurement of those asx- and ST-turn torsion angles that are equivalent to the standard dihedral angles of  $\beta$ -turns shows (in Figure 7.1 and Table 7.2) that the four types of asx-turns are geometrically equivalent to the four main types of  $\beta$ -turn: I, I', II and II', and that the three types of ST-turn are equivalent to three of those types: I, II and II'. It is proposed here that asx- and ST-turns be named using the same nomenclature as for  $\beta$ -turns. The most common of the asx- and ST-turn conformations is type II', characterized by a side chain  $\chi_1$  value of  $\sim 180^\circ$  (angle inter-conversion is discussed in the Legend to Figure 2.3). Types I and II are less common. Type I has  $\chi_1$  of  $\sim 60^\circ$  and mimics the main chain part of a 3/10-helix. Type II also has  $\chi_1$  of  $\sim 60^\circ$ , and has a central residue with the  $\alpha_L$  or  $\gamma_L$  conformation. Type I' is very rare and examples are found only among the asx-turns.

Given that ST-turns feature a hydrogen-bonded ring of 9, rather than 10, atoms, the similarity of their structural classification to that of asx-turns is unexpected. However, chemical properties of the four residues (D, N, S and T) are not dissimilar and it has been observed that they exhibit a tendency in such situations to substitute each other over evolutionary time (Vijayakumar *et al.*, 1999; Wan & Milner-White, 1999b).

The shorter average O...H distance of the hydrogen-bond observed in asx-turns compared with  $\beta$ -turns indicates that the asx-turn is able to adopt a conformation which is more favourable. The asx-turn has extra flexibility over the  $\beta$ -turn, conferred by the rotational freedom of the side-chain  $\chi_2$  angle. However,  $\phi_c$  of the asx-turn ( $\chi_1$  of the side chain) is more constrained than  $\phi$  of the  $\beta$ -turn, as it must adopt a rotamer conformation.

Considering this, it is interesting to reverse our perspective on how  $\beta$ -turns are mimicked by asx- and ST-turns, and instead ponder the mimicry of the asx side chain rotamer angle by  $\phi$  of the  $\beta$ -turn. It would seem that the ideal  $\phi$  angle of residue  $i+1$  of the  $\beta$ -turn mimics the rotameric conformation of the asx-turn.

Common motifs incorporating asx- or ST-turns are the asx- or ST-nests. These are recently identified, yet by no means uncommon, protein motifs (Watson & Milner-White,

2002a; Pal *et al.*, 2002). The observation that almost all type II asx- and ST-turns occur as asx- or ST-nest motifs is an unexpected finding of the present work. Conversely, most asx- or ST-nests are also type II asx- or ST-turns so they are nearly synonymous.

## 8.4 General conclusions

The findings presented here provide new insights into protein structure at the atomic level. They have relevance to protein folding and protein-ligand binding, and could potentially be applied to protein fold prediction methods and drug design.

The analysis of hydrogen-bonding of *trans*-amides in small molecule structures, and its comparison with the situation in proteins, shows the importance of examining many factors when attempting to explain the geometry of specific inter-molecular interactions. Packing considerations, competition with other interactions, steric accessibility, and the recurrence of certain motifs, can all play a part. Similarly, the examination of main-chain carbonyl-carbonyl interactions in proteins shows the complexity of secondary structure, the analysis of which requires attention not only to main-chain geometry and local hydrogen-bonding, but also to combinations of the two, and other factors besides. In the worlds of both small molecules and proteins the factors that contribute to the geometry of an interaction are likely to be many, and care is needed to separate and evaluate their influences.

In clarifying the geometric relationship of *asx*- and *ST*-turns to  $\beta$ -turns, a deeper understanding of these small hydrogen-bonded motifs is gained, aiding future efforts in this area of research.

The work on intra-protein carbonyl-carbonyl interactions was limited to main-chain carbonyl groups. It would be interesting to apply the same search criteria and analysis to side-chains as well, by considering the interactions of asparagine and glutamine (and also aspartate and glutamate) carbonyl groups with each other and with main-chain carbonyls. The *asx*-turn represents a single example of a commonly recurring side-chain-main-chain carbonyl-carbonyl interaction motif, but it is likely that others exist among the recurring features of protein secondary structure. It is possible that some of these will mimic the main-chain carbonyl-carbonyl interaction motifs, just as the *asx*-turn carbonyl-carbonyl interaction mimics the  $\beta$ -turn carbonyl-carbonyl interactions identified in chapter 6 (motifs 2 and 4).

## 9 References

- Aakeroy, C.B. & Seddon, K.R. (1993) The hydrogen bond and crystal engineering. *Chem. Soc. Revs.* **22**, 397-407.
- Allan, D.R., Clark, S.J., Ibberson, R.M., Parsons, S., Pulham, C.R. & Sawyer, L. (1999) The influence of pressure and temperature on the crystal structure of acetone. *Chem. Commun.* **8**, 751-752.
- Allen, F.H. & Johnson, O. (1991) Automated conformational analysis from crystallographic data. 4. Statistical descriptors for a distribution of torsion angles. *Acta Cryst.* **B47**, 62-67.
- Allen, F.H. (1986) A systematic pairwise comparison of geometric parameters obtained by X-ray and neutron diffraction. *Acta Cryst.* **B42**, 515-522.
- Allen, F.H. (2002) The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Cryst.* **B58**, 380-388.
- Allen, F.H., Baalham, C.A., Lommerse, J.P.M. & Raithby, P.R. (1998) Carbonyl-carbonyl interactions can be competitive with hydrogen bonds. *Acta Cryst.* **B54**, 320-329.
- Allen, F.H., Motherwell, W.D.S., Raithby, P.R., Shields, G.P. & Taylor, R. (1999) Systematic analysis of the probabilities of formation of bimolecular hydrogen-bonded ring motifs in organic crystal structures. *New J. Chem.* 25-34.
- Altschul, S.F., Gish W., Miller W., Myers E.W. & Lipman D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215** (3), 403-10.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25** 3389-3402.

- Apaya, R.P, Bondi, M. & Price, S.L. (1997) The orientation of N-H...O=C and N-H...N hydrogen bonds in biological systems: How good is a point charge as a model for a hydrogen bonding atom? *J. Comp-aided molecular design* **11**, 479-490.
- Artymiuk, P.J. & Blake, C.C.F. (1981) Refinement of human lysozyme at 1.5 Å resolution. Analysis of non-bonded and hydrogen-bond interactions. *Journal of Molecular Biology* **152**, 737-762.
- Aurora, R. & Rose, G.D. (1998) Helix capping. *Protein Sci.* **7**, 21-38.
- Baker, E.N. & Hubbard, R.E. (1984) Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* **44**, 97-179.
- Brandl, M., Weiss, M.S., Jabs, A., Suhnel, J. & Hilgenfeld, R. (2001) C-H... $\pi$ -interactions in proteins. *J. Mol. Biol.* **307**, 357-377.
- Bergner, A., Günther, J., Hendlich, M., Klebe, G. & Verdonk, M. (2001-2002). Use of Relibase for retrieving complex three-dimensional interaction patterns. *Biopolymers* **61**, 99-110.
- Bernstein, J., Cohen, M.D. & Leiserowitz, L. (1974) The chemistry of quinonoid compounds (Patai, S., Ed.), Wiley, London.
- Bode, W., Papamokos, E. & Musil, D. (1987) The high-resolution X-ray crystal structure of the complex formed between subtilisin Carlsberg and eglin c, an elastase inhibitor from the leech *Hirudo medicinalis*. *Eur. J. Biochem.* **166**, 673-692.
- Bolton, W. (1963) The crystal structure of anhydrous barbituric acid. *Acta Cryst.* **16**, 166-173.
- Bolton, W. (1964) The crystal structure of alloxan. *Acta Cryst.* **17**, 147-152.
- Bolton, W. (1965) The crystal structure of triketoindane (anhydrous ninhydrin). A structure showing close C=O...C interactions. *Acta Cryst.* **18**, 5-10.



- Bruno, I.J., Cole, J.C., Edgington, P.R., Kessler, M., Macrae, C.F., McCabe, P., Pearson, J., & Taylor, R. (2002) New software for searching the Cambridge Structural Database and visualising crystal structures. *Acta Cryst.* **B58**, 389-397.
- Buckingham, A.D., Fowler, P.W. & Hutson, J.M. (1988) Theoretical studies of van der Waals molecules and intermolecular forces. *Chem. Rev.* **88**, 963-988.
- Burgi, H.-B., Dunitz, J. D. & Shefter, E. (1974) Chemical reaction paths. IV. Aspects of O...C=O interactions in crystals. *Acta Cryst.* **B30**, 1517-1523
- CCDC. (1994) Vista - a program for the analysis and display of data retrieved from the CSD. Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, England, U.K.
- Chakrabati, P. & Pal, D. (2001) The inter-relationships of side-chain and main-chain conformations in proteins. *Prog. Biophys. Mol. Biol.* **76**, 1-102.
- Deane, C. M., Allen, F. H., Taylor, R. & Blundell, T. L. (1999) Carbonyl-carbonyl interactions stabilize the partially allowed Ramachandran conformations of asparagine and aspartic acid. *Protein Engineering* **12**, 1025-1028.
- Derewenda, Z.S., Lee, L. & Derewenda, U. (1995) The occurrence of C-H...O hydrogen bonds in proteins. *J. Mol. Biol.* **252**, 248-262.
- Doig, A.J., Macarthur, M.W., Stapley, B.J. & Thornton, J.M. (1997) Structures of N-termini of helices in proteins. *Protein Sci.* **6**, 147-155.
- Duddy, W.J., Nissink, J.W.M., Allen, F.H. & Milner-White E.J. (2004) Mimicry by asx- and ST-turns of the four main types of beta-turn in proteins. *Protein Sci.* **13**, 3051-3055.

- Engh, R.A. & Huber, R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst.* **A47**, 392-400.
- Ermiler, U., Grabase, W., Shima, S., Goubeaud, W.M. & Thauer, R.K., (1997) Crystal structure of methyl-coenzyme M reductase: the key enzyme of biological methane formation. *Science* **278**, 1457-1462.
- Eswar, N. & Ramakrishnan, C. (1999) Secondary structures without backbone: an analysis of backbone mimicry by polar side chains in proteins. *Prot. Eng.* **12**, 447-455.
- Eswar, N. & Ramakrishnan, C. (2000) Deterministic features of side-chain hydrogen bonds in globular protein structures. *Prot. Eng.* **13**, 27-238.
- Fabiola, F., Bertram, R., Korostelev, A. & Chapman, M.S. (2002) An improved hydrogen bond potential: impact on medium resolution protein structures. *Prot. Sci.* **11**, 1415-1423.
- Gavezzotti, A. & Filippini, G. (1994) Geometry of the intermolecular X-H...Y (X, Y = N,O) hydrogen bond and the calibration of empirical hydrogen bond potentials. *J. Phys. Chem.* **98**, 4831-4837.
- Golemi, D., Maveyraud, L., Vakeulenko, S., Samama, J.P. & Mobashery, S., (2001). Critical involvement of a carbamylated lysine in catalytic function of class D beta-lactamases. *Proc. Nat. Acad. Sci. USA* **98**, 14280-14285.
- Gunasekaran, K., Gomathi, L., Ramakrishnan, C., Chandrasekhar, J. & Balaram, P. (1998) Conformational interconversions in peptide  $\beta$ -turns: analysis of turns in proteins and computational estimates of barriers. *J. Mol. Biol.* **284**, 1505-1516.
- Günther, J., Bergner, A., Hendlich, M. & Klebe, G. (2003) Utilising structural knowledge in drug design strategies: applications using relibase. *J. Mol. Biol.* **326** (2), 621-636.

- Hay, B.P., Dixin, D.A., Bryan, J.C. & Moyer, B.A. (2002) Crystallographic evidence for oxygen transfer directionality in oxyanion hydrogen bonds. *J. Am. Chem. Soc.* **124**, 182-183.
- Hendlich, M., Bergner, A., Günther, J. & Klebe, G. (2003) Relibase: design and development of a database for comprehensive analysis of protein-ligand interactions. *J. Mol. Biol.* **326** (2), 607-620.
- Hutchinson, E.G. & Thornton, J.M. (1994) A revised set of potentials for beta-turn formation in proteins. *Prot. Sci.* **3**, 2207-2216.
- Hutchinson, E.G. & Thornton, J.M. (1996) PROMOTIF - a program to identify and analyze structural motifs in proteins. *Prot. Sci.* **5**, 212-220.
- Infantes, L. & Motherwell, S. (2004) Prediction of H-Bonding Motifs for Pyrazoles and Oximes Using the Cambridge Structural Database. *Struct. Chem.* **15** (3), 173-184.
- Jeffrey, G.A. (1997) *An Introduction to Hydrogen Bonding*. Oxford Univ. Press, Oxford, UK.
- Kabsch, W. & Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen bonded and geometrical features. *Biopolymers* **22**, 2577-2637
- Klebe, G. (1994) The use of composite crystal field environments in molecular recognition and the de novo design of protein ligands. *J. Mol. Biol.* **237**, 212-235.
- Kortemme, T., Morozov, A.V. & Baker, D. (2003) An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J. Mol. Biol.* **326**, 1239-1259.
- Kroon, J., Kanters, J.A., Van Duijneveldt-Van de Rijdt, J.G.C.M., Van Duijneveldt, F.B. & Vliegthart, J.A. (1975) O-H...O Hydrogen bonds in molecular crystals. A statistical and quantum chemical analysis. *J. Mol. Struct.* **24**, 109-129.

- Kuhn, P., Deacon, A.M., Comoso, S., Rajaseger, G., Kini, R.M., Uson, I. & Kolatakur, P.R. (2000) The atomic resolution structure of Bucandin, a toxin isolated from the Malayan Krait determined by direct methods. *Acta Cryst.* **D56**, 1401-1407.
- Lario, P.I. & Vriclink, A. (2003) Atomic resolution density maps reveal secondary structure dependent differences in electronic distribution. *J. Am. Chem. Soc.* **125** (42), 12787-12794.
- Legon, A.C. & Millen, D.J. (1987) Directional character, strength and nature of the hydrogen bond in gas-phase dimers. *Acc. Chem. Res.* **20**, 39-45.
- Leiserowitz, L. & Tuval, M. (1978). Molecular-packing modes. N-Methylamides. *Acta Cryst.* **B34**, 1230-1247..
- Lewis, P.N., Momany, F.A. & Scheraga, H.A. (1973) Chain reversals in proteins. *Biochem. Biophys. Acta* **303**, 211-229.
- Lifson, S., Hagler, A. T. & Dauber, P. (1979) Consistent force field studies of intermolecular forces in hydrogen bonded crystals. 1. *J. Amer. Chem. Soc.* **101**, 5111-5121.
- Llamas-Saiz, A.L., Foces-Foces, C., Mpo, O., Yanez, M. & Elguero, J. (1992) Nature of the hydrogen bond: Crystallographic versus theoretical description of the O-H...N(sp<sup>2</sup>) hydrogen bond. *Acta Cryst.* **B48**, 700-713.
- Lovell, S., Word, J.M., Richardson, J.S. & Richardson, D.C. (2000) The penultimate rotamer library. *Proteins* **40**, 389-408.
- Lovell, S.C., Davis, I.W., Arendall, W.B., de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S. & Richardson, D.C. (2003) Structure validation by C-alpha geometry: phi, psi and C-beta deviation. *Proteins* **50** (3), 437-450.

- Maccallum, P.H., Poet, R. & Milner-White, E.J. (1995a) Coulombic interactions between partially charged main chain atoms not hydrogen-bonded to each other influence the conformations of  $\alpha$ -helices and antiparallel  $\beta$ -sheet. A method for analysing forces between hydrogen bonding groups in proteins includes all coulombic interactions. *J. Mol. Biol.* **248**, 361-373.
- Maccallum, P.H., Poet, R. & Milner-White, E.J. (1995b) Coulombic attractions between partially charged main-chain atoms stabilise the right-handed twist found in most beta-strands. *J. Mol. Biol.* **248**, 374-384.
- McDonald, I.K. & Thornton, J.M. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777-793.
- Mills, J.E.J. & Dean, P.M. (1996) Three-dimensional hydrogen bond geometry and probability information from a crystal survey. *J. Comp.-aided Molec. Design* **10**, 607-622.
- Milner-White, E.J. (1988) Recurring loop motif in proteins that occurs in righthanded and left-handed forms. Its relationship with alpha-helices and betabulge loops. *J. Mol. Biol.* **199**, 503-511.
- Milner-White, E.J., Nissink, J.W.M., Allen, F.H. & Duddy, W.J. (2004) Recurring main-chain anion-binding motifs in short polypeptides: nests. *Acta Cryst.* **D60**, 1935-1942.
- Mitchell, J.B.O. & Price, S.L. (1989) On the electronic directionality of N-H...C=O hydrogen bonding. *Chem Phys Letts.* **154**, 267-272.
- Mitchell, J.B.O. & Price, S.L. (2000) A systematic nonempirical method of deriving model intermolecular potentials for organic molecules: application to amides. *J. Phys. Chem.* **104**, 10958-10971.

- Murray-Rust, P. & Glusker, J.P. (1984) Directional hydrogen bonding to  $sp^2$  and  $sp^3$  hybridised oxygen atoms and its relevance to ligand-macromolecule interactions. *J. Am. Chem. Soc.* **106**, 1018-1025.
- Pal, D., Suhnel, J. & Weiss, M.S. (2002) New principles of protein structure: Nests, Eggs-and what next? *Angew. Chem. Int. Ed.* **41**, 4663-4665.
- Presta, L.G. & Rose, G.D. (1988) Helix signals in proteins. *Science* **240**, 1632-1641.
- Rees, D.C., Lewis, M. & Lipscomb, W.N. (1983) Crystal structure of Carboxypeptidase A. *J. Mol. Biol.* **168**, 167-387.
- Richardson, J.S. (1981) The anatomy and taxonomy of protein structure. *Adv. Prot. Chem.* **34**, 168-340.
- Richardson, J.S. & Richardson, D.C. (1988) Amino acid preferences for specific locations at the ends of alpha helices. *Science* **240**, 1648-1652.
- Richardson, J.S. & Richardson, D.C. (1989) in Fasman GD. (Ed.) Prediction of protein structure and the principles of protein conformation. Plenum Press, N.Y. pp 1-98.
- Rose, G.D. & Wolfenden, R. (1993) Hydrogen bonding, hydrophobicity, packing, and protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **22**, 381-415.
- Sarkel, S. & Desiraju, G.R. (2004) N-H...O, O-H...O and C-H...O hydrogen bonds in protein-ligand complexes. *Proteins* **54**, 247-259.
- Scheiner, S. (1994) Bent hydrogen bonds and proton transfer. *Acc. Chem. Res.* **10**, 294-303.
- Schellman, C. (1980) The alpha-L conformation at the ends of helices. In: Jaenicke R, ed. Protein folding. Elsevier/North Holland Biomedical Press, Amsterdam. pp 53-61.

- Stickle, D.F., Presta, L.G., Dill, K.A. & Rose, G.D. (1992) Hydrogen bonding in globular proteins. *J. Mol. Biol.* **226**, 1143-1159.
- Tainer, J.A., Getzoff, E.D., Beem, K.M., Richardson, J.S. & Richardson, D.C. (1982) Determination of the structure of copper zinc superoxide dismutase. *J. Mol. Biol.* **160**, 181-217.
- Taylor, R. & Kennard, O. (1984) Hydrogen-bond geometry in organic crystals. *Acc. Chem. Res.* **17**, 320-326.
- Taylor, R., Kennard, O., & Versichel, W. (1983) Geometry of the N-H...O=C hydrogen bond. 1. Lone-pair directionality. *J. Am. Chem. Soc.* **105**, 5761-5766.
- Taylor, R., Kennard, O., & Versichel, W. (1984) The geometry of the N-H...O=C hydrogen bond. 3. Hydrogen bond distances and angles. *Acta. Cryst.* **B40**, 280-288.
- Thornton, J.M., McArthur, M.W., McDonald, I.K., Jones, D.T., Mitchell, J.B.O., Nandi, C.L., Price, S.L. & Zwelebil, J.J.M. (1993) Protein structures and complexes: What they reveal about the interactions which stabilize them. *Phil Trans Roy Soc ser A* **345**, 113-129.
- Torshin, I.Y, Weber, I.T. & Harrison, R.W. (2002) Geometric criteria of hydrogen bonds in proteins and identification of bifurcated hydrogen bonds. *Prot. Eng.* **15**, 359-363.
- Vedani, A. & Dunitz, J.D. (1985) Lone-pair directionality in hydrogen bond potential functions for molecular mechanics calculations. *J. Am. Chem. Soc.* **107**, 7653-7658.
- Venkatachalam, C.M. (1968) Stereochemical criteria for polypeptides and proteins V. *Biopolymers* **6**, 1425-1436.
- Vijayyakumar, M., Quian, H. & Zhou, H.X. (1999) Hydrogen bonds between short polar side chains and peptide backbone: prevalence in proteins and effects on helix-forming propensities. *Proteins* **34**, 497-507.



- Wan, W-Y. & Milner-White, E.J. (1999a) A natural grouping of motifs with an aspartate or asparagine forming two hydrogen bonds to residues ahead in sequence: their occurrence at  $\alpha$ -helical N termini and in other situations. *J. Mol. Biol.* **286**, 1633-1649.
- Wan, W-Y. & Milner-White, E.J. (1999b) A recurring two-hydrogen-bond motif incorporating a serine or threonine residue is found both at  $\alpha$ -helical N-termini and in other situations. *J. Mol. Biol.* **286**, 1650-1666.
- Wahl, M.C. & Sundaralingam, M. (1997) C-H...O hydrogen bonding in biology. *Trends Biochem. Sci.* **22**, 97-102.
- Watson, J.D. & Milner-White, E.J. (2002a) A novel main-chain anion binding site in proteins: a particular combination of  $\phi, \psi$  angles in successive residues gives rise to anion binding sites that occur commonly and are often found at functionally important regions. *J. Mol. Biol.* **315**, 171-182.
- Watson, J.D. & Milner-White, E.J. (2002b) The conformations of polypeptide chains where the  $\phi, \psi$  values of alternating residues are enantiomeric. Their occurrence in cation and anion binding regions of proteins. *J. Mol. Biol.* **315**, 183-191.
- Weiss, M.S., Brandl, M., Suhnel, J., Pal, D. & Hilgenfeld, R. (2001) More hydrogen bonds for the (structural) biologist. *Trends Biochem. Sci.* **26**, 521-523.
- Wilmot, C.M. & Thornton, J.M. (1988) Analysis and prediction of the different types of beta-turn in proteins. *J. Mol. Biol.* **203**, 221-232.
- Wilmot, C.M. & Thornton, J.M. (1990)  $\beta$ -turns and their distortions: a proposed new nomenclature. *Prot. Eng.* **3**, 479-493.



Word, J.M., Lovell, S.C., LaBean, T.H., Taylor, H.C., Zalis, M.E., Presley, B.K.,  
Richardson, J.S. & Richardson D.C. (1999a) Visualizing and quantifying molecular  
goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J. Mol.  
Biol.* **285**, 1711-1733.

Word, J.M., Lovell, S.C., Richardson, J.S. & Richardson, D.C. (1999b) Asparagine and  
glutamine: using hydrogen atom contacts in the choice of side-chain amide  
orientation. *J. Mol. Biol.* **285**, 1735-1747.

