Anion And Cation Binding In Proteins

James David Watson

Submitted for the degree of

Doctor of Philosophy

The University of Glasgow

Faculty of IBLS

September 2002

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

Abstract

Main-chain conformations where one amino acid is described as γR (or αR) and an adjacent one is γL (or αL) generally give rise to a structure where the three main-chain NH groups (one from each of the two residues and the one that follows) form a depression. Due to the electronegativities of the atoms involved, this depression is able to accommodate an atom with a whole or partial negative charge. The negatively charged atom, when present, is also stabilised through hydrogen bonds with the NH groups. We propose the name "nest" for this feature.

The nest is a common motif with 8% of residues in an average protein being involved in one. It is also a novel motif even though the situations it occurs in are well known. The negatively charged atom, or anion, that often occupies the nest varies and may be a main-chain carbonyl group such as in the paperclip (Schellman motif), it may also be a phosphate group as in the P-loop superfamily that binds ATP and GTP.

Overlapping (compound) nests and adjacent (tandem) nests are often observed and encourage the binding of larger more complex anionic groups. Examples of this include the compound nest of the P-loop which has 5 successive NH groups that bind the β phosphate group of nucleotide triphosphate. The longest compound nests are found surrounding cysteine bound [2Fe2S] and [4Fe4S] iron-sulphur centres, which are also anionic groups.

A totally novel motif occurs where the first residue of the nest is Asp, Asn, Ser or Thr. In many of these cases the sidechain acts as the negatively charged group so that the sidechain is effectively interacting with its own main-chain NH group and the two that follow. These motifs are relatively common and are named depending on the identity of the side chain involved: where the first amino acid is Asn or Asp the motif is known as an Asx-nest, when it is Ser or Thr the motif is an ST-nest. These are important motifs and are found in the calcium-binding loops of EF-hand proteins.

The discovery of the nest prompted an investigation into polypeptides where successive residues have main-chain ϕ,ψ conformations of opposite hand (alternating enantiomeric main-chain dihedral angles). Through the creation of model polypeptides on computer the full range of possible conformations was examined producing a plot not unlike the Ramachandran plot. All alternating enantiomeric structures are ring-shaped or extended, and some are found to occur in proteins, the commonest approximating to the nest.

Another conformation related to the nest but more extended is found in the four stretches of polypeptide that line the tetrameric K^+ channel, in this case the main-chain CO groups are the important feature and bind the K^+ ions in the channel.

A different ring-shaped structure where the main-chain CO groups point to the centre of the ring is employed for specific Ca^{2+} ion-binding in the annexin, phospholipase A2 and subtilisin loops, and the regularly arranged β-roll loops of the serralysin protease family. This feature we propose calling the catgrip.

Contents

Tables and illustrations

Abbreviations

αL, αR, γL, γR refer to conformational areas delimited by main-chain dihedral angles of individual residues. We use the definition of Efimov (Efimov, A. V., 1991; Efimov, A. V., 1993). Typical ϕ,ψ values are: αR, -70°, -40°; αL, 70°, 40°; γR, -90°, 0°; γL, 90°, 0°.

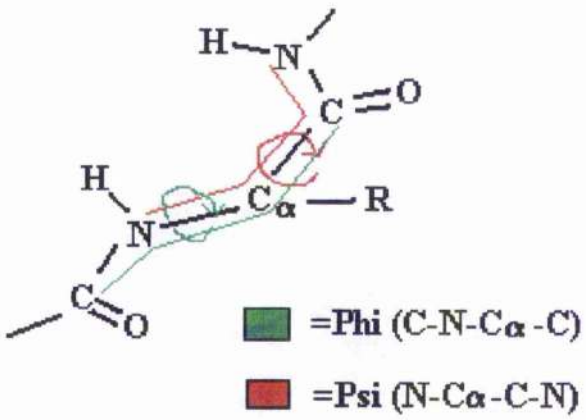
ADP	=	Adenosine diphosphate
Asx	=	Amino acid identity is Aspartate or Asparagine
ATP	=	Adenosine triphosphate
Cα	=	Alpha carbon
Cl^-	=	Chloride ion
CO	=	Mainchain carbonyl group
C-term	=	Carboxyl terminus
DSSP	=	Dictionary of secondary structures of proteins
GDP	=	Guanosine diphosphate
Glx	=	Amino acid residue is Glutamate or Glutamine
GTP	=	Guanosine triphosphate
H-bond	=	Hydrogen bond
HIPIPs	=	High potential Iron-sulphur proteins
K^+	=	Potassium cation
K-channel	=	Potassium channel
LR	=	A residue with mainchain dihedral angles in the αL or γL region followed by a residue with mainchain dihedral angles in the αR or γR region.
MMP	=	Matrix metalloprotease
NADP	=	Nicotinamide adenine diphosphate
NH	=	Mainchain amide group
NNN angle	=	Angle measured between 3 consecutive backbone nitrogen atoms
N-term	=	Amino terminus
PDB	=	Protein databank
P-loop	=	Phosphate-binding loop
RL	=	A residue with mainchain dihedral angles in the αR or γR region followed by a residue with mainchain dihedral angles in the αL or γL region.
ST	=	Amino acid identity is Serine or Threonine

Chapter 1: Introduction

1.1 Ramachandran Plots

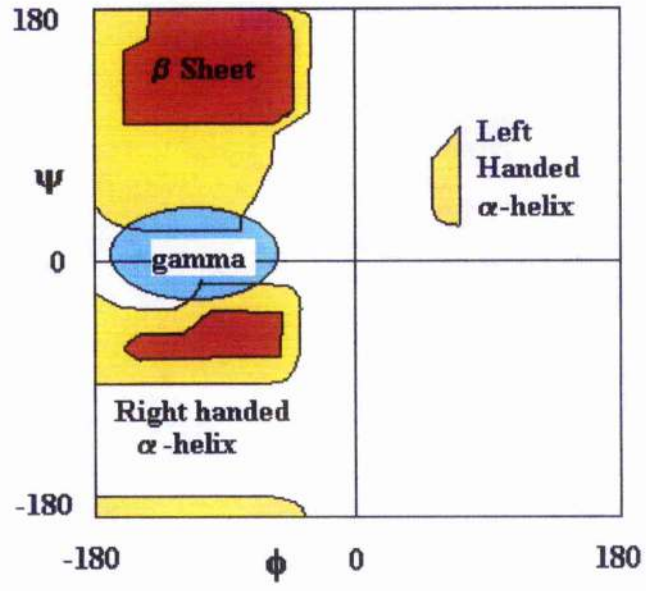
In 1968 G.N. Ramachandran and V. Sasisekharan published their work on the conformations of polypeptides and proteins, looking primarily at the steric hindrance when changing the polypeptide backbone torsion angles (Ramachandran, G. N. & Sasiekharen, V.,1968). Assuming that the peptide bond is planar, the main chain folding of a polypeptide chain is limited to only two degrees of freedom per residue. These are the angles of rotation around the Cα-N bond (Phi ϕ) and the Cα-C bond (Psi ψ) as shown in figure 1.1a.

Figure 1.1a : Diagram showing angles phi and psi



The work had two main parts to it, the first of which involved the energies of model "dipeptides". The energy of model "dipeptides" are plotted for every combination of phi and psi and gives rise to the familiar Ramachandran plot as illustrated in figure 1.1b below. The areas shaded on the plot are indicative of highly favourable energies and therefore are considered to be the 'allowed' regions. The regions considered to be disallowed are those where the energy of the model "dipeptide" is highly unfavourable, this is usually due to steric hindrance or electrostatic repulsion.

Figure 1.1b : Ramachandran plot.



The second part of the group's work involved moving from "dipeptides" to polypeptides. Here the interest lay in discovering what happens to the conformation of the polypeptide when successive residues have identical Phi and Psi values. The results of the modelling process showed that there is a tendency for these polypeptides to form helical structures (Dickerson & Geis, 1969). The many different types of helix are characterised by three measurements:

1. The number of residues per turn
2. The distance between residues measured parallel to the helix axis
3. The pitch (the parallel distance between two residues one turn apart)

A strand of Beta Sheet can be described as a flat helix with 2 residues per turn although this may not be obvious. As the number of residues per turn increases to 3 (for a 3_{10} helix), 3.6 (for an α helix) then 4.4 (for a π helix) structures appear more helical and are more easily recognised. Not all conformations are acceptable however, as there are many cases where the various atoms in the chain start to collide with one another (e.g. at $\phi = -60°$ and $\psi = -135°$ successive carbonyls clash severely). This interaction is dictated by the van der Waals radii minimum contact distances deduced from proteins and small structures.

1.2 Researching Enantiomeric Alternating Structures

This work started as an investigation into the formation of reverse turns at the C-termini of alpha helices. These occur when the regular pattern of right handed alpha helical residues is broken by a residue with a left handed combination of torsion angles (Richardson, J.S, 1981). The pattern of right and left handed residues was seen to occur in various other situations and the first extended example we found was that of the phosphate-binding loop of p21 Ras. This prompted an investigation into the conformations of polyglycine peptides where successive amino acids have enantiomeric (mirror image) alternating main chain dihedral angles, for example:

	Phi ϕ (degrees)	Psi ψ (degrees)
Residue 1	-50	-30
Residue 2	+50	+30
Residue 3	-50	-30
Residue 4	+50	+30

Searching a representative sample of the proteins from the Protein Data Bank (PDB) for examples of these enantiomeric structures demonstrates that they form anion- and cation-binding sites, many of which form recognised important structural and functional sites (e.g. P-loops, Iron-sulphur complexes, Potassium Channels). The vast majority, however, are structural small hydrogen-bonded motifs (e.g β bulge loops, reverse turns). In these examples the binding is to the partially negatively charged oxygen atom of carbonyls rather than to an obvious anion or cation such as calcium or phosphate.

1.3 Small Hydrogen Bonded Structural Motifs

Associated with secondary structure are many small motifs which can be identified by their patterns of hydrogen bonding. These are often structures of 4-6 amino acids forming stable loops which help direct the backbone to fold up correctly (Richardson, J. S., 1981; Baker, E.N. & Hubbard, R. E., 1984; Milner-White, E. J. & Poet, R., 1987; Ring, C. S. et. al., 1992; Donate, L. E. et. al., 1996; Doig, A. J. et. al., 1997; Oliva, B. et. al., 1997; Gunasekaran, K. et. al., 1997; Aurora, R. & Rose, G. D., 1998; Wan, W-Y. & Milner-White, E. J., 1999; Venkatachalam, C. M., 1968). Each motif is described

and its hydrogen bonding pattern is shown in figure 1.2b. The colour scheme used for the hydrogen bonds is described in figure 1.2a (Milner-White, E. J. & Poet, R. 1987).

Fig 1.2a:

Hydrogen Bond Colour Key

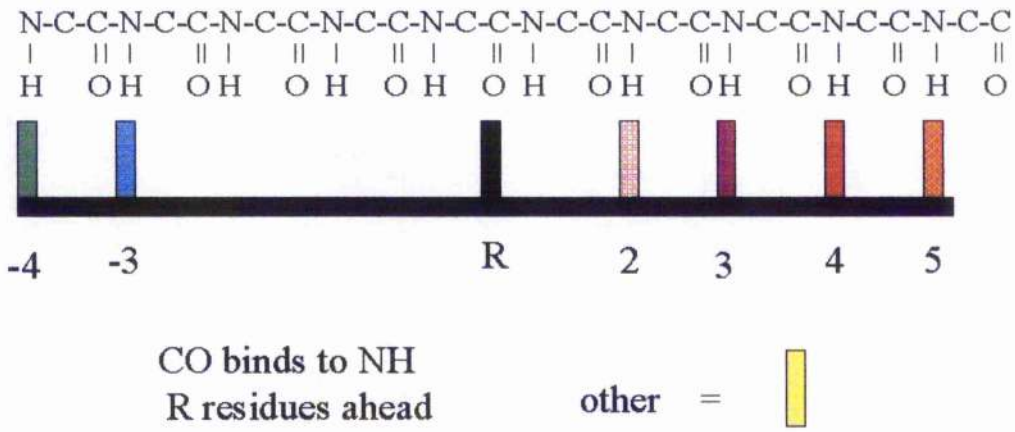
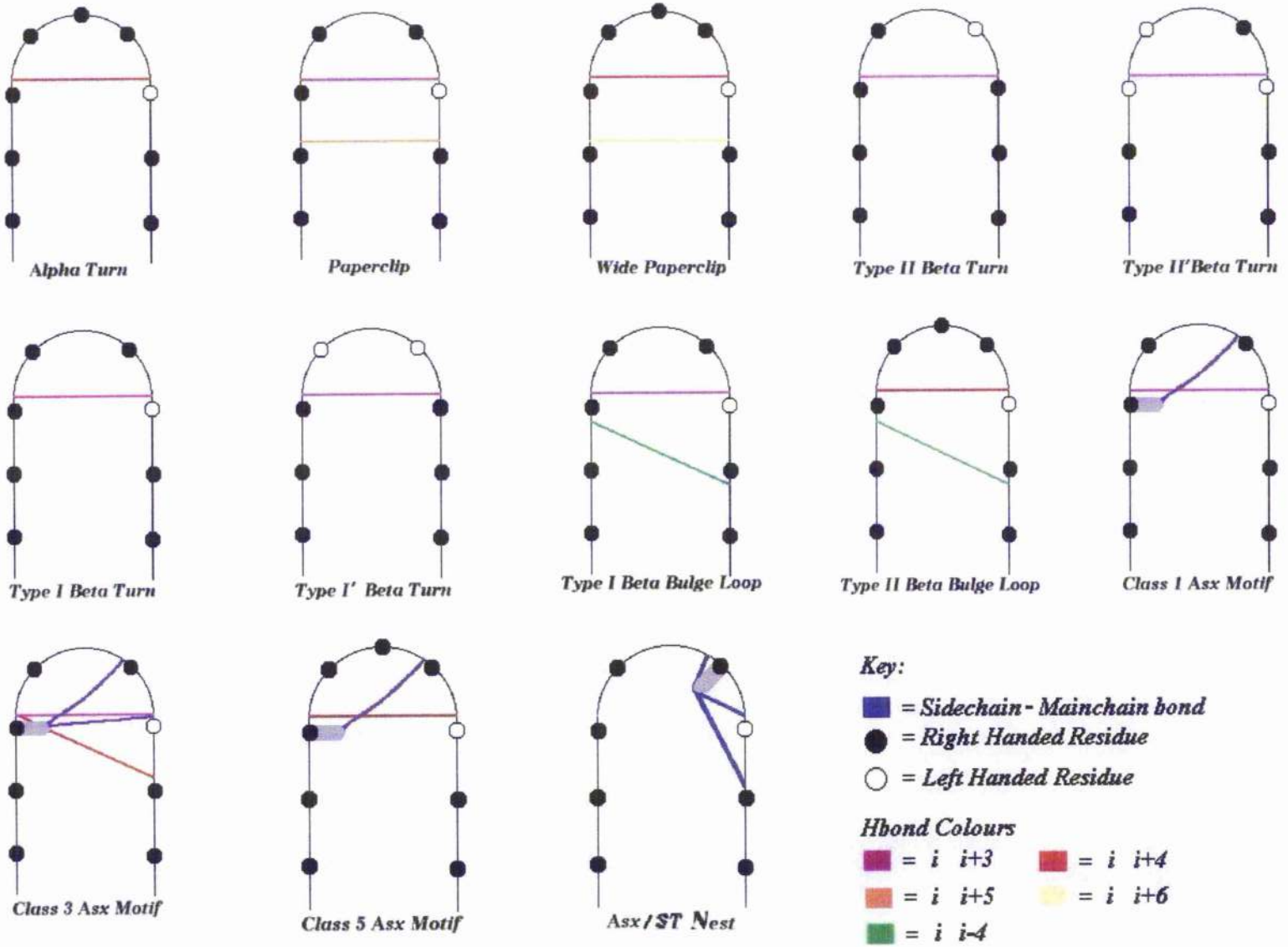


Fig 1.2b: Hydrogen bond diagrams.



1.3.1 Beta Turns

Beta turns were originally defined by the one hydrogen bond common to all (an i, i+3 hydrogen bond) (Venkatachalam, C. M., 1968) but some modern descriptions of beta turns do not require a hydrogen bond (Milner-White, E. J. & Poet, R., 1987). Almost all β-turns can be classified as one of four types as defined by the mainchain conformations of the two loop residues (arranged most common to least):

1. Type I

   Here the two loop residues are right handed in main-chain conformation and an i, i+3 bond is often found.

2. Type II

   In this type the first of the two loop residues has a right handed main-chain conformation and the second is left handed.

3. Type II'

   This is the enantiomeric form of the type II turn. The first of the two loop residues is left handed in main-chain conformation and the second one is the right handed one.

4. Type I'

   This is the enantiomeric form of the Type I turn: the two loop residues are left handed in main-chain conformation.

The prime symbol (') is used to indicate the enantiomeric (mirror image) mainchain conformation (similar to the difference between right and left handed helices).

1.3.2 Beta Bulge Loops

These are often associated with beta sheets and result from an additional residue being found in one strand (Chan, E. A. W. *et. al.*, 1993; Milner-White, E. J., 1987). This interrupts

the regular hydrogen bonding and causes a distinctive bulge. The bulge becomes particularly distorted when one of the residues is a left handed glycine residue. These loops can also be found as loops on their own (although it will be shown later that they may not be stable enough to form completely on their own). These motifs are classified as one of two types: type I and type II.

1.3.3 Alpha turns

The alpha turn (Pavone, V. et. al., 1996; Nataraj, D. V., 1996) is the simplest of all motifs and is characterised by one (i, i+4) hydrogen bond. It is found as part of the hydrogen bonding network of alpha helices as well as occurring on its own.

1.3.4 Paperclip/Schellman Motifs

A common motif found at the C-termini of alpha helices is that known as a Schellman motif (Schellman, C., 1980; Aurora, R. et. al., 1994; Viguera, A. R. & Serrano, L., 1995) or Paperclip (Milner-White, E. J., 1988). This is essentially a reverse turn that breaks the alpha helix out of its cycle and is characterised by two hydrogen bonds: an i, i+3 bond and an i, i+5 bond. The motif is also characterised by the presence of a left handed residue (commonly glycine). The motif is stable and can occur on its own independent of helix formation.

1.3.5 Gamma Turns

A rarer type of turn is the gamma turn (Milner-White, E. J. & Poet, R., 1987) (and inverse gamma turn). Here the characteristic hydrogen bond is an (i, i+2) bond, which is rather weak because of the bent geometry involved. The parameters chosen for the hydrogen bonding dictate the frequency with which these turns are detected.

Chapter 2: The Nest

2.1 Introduction

During an investigation into the C-termini of alpha helices it became apparent that the Schellman motif (Paperclip) shared features with the P-loop and the oxyanion hole of serine proteases. The main point of similarity was the way that three mainchain NH groups were used to bind negatively charged atoms or molecules. Looking at these motifs in more detail it became obvious that the mainchain was folding up in a similar way to form a positively charged concavity. Since the mainchain conformation is dictated by the dihedral angles of the backbone it was decided that dihedral angle patterns and their effect on the mainchain conformation would be the starting point of the project.

Work by Ramachandran *et. al.* (Ramachandran, G. N. & Sasisekharan, V., 1968; Dickerson & Geis, 1969) on the conformations of polypeptides had only taken into account situations where the polypeptide chain is made up of successive residues with the same conformation (dihedral angles), no studies we were aware of had investigated alternating enantiomeric dihedral angles as explained previously. Through analysis of a representative sample of the PDB we identified numerous examples where the mainchain dihedral angles alternated between enantiomeric (mirror image) right and left handed forms. This led to the identification of a novel anion-binding motif which we have termed the 'nest' (Watson, J. D. & Milner-White, E. J., 2001).

The 'nest' is a common motif which is shown to have both structural and functional properties. Although the situations the nest is found in are well described, this is the first time they have been related to one another as having a common core.

2.2 Geometry and Properties

There are two types of nest distinguished by the pattern of dihedral angles. There are those where the first residue of the nest has right handed dihedral angles followed by a left handed residue (RL), and there are those where the first residue is left handed followed by a right handed one (LR). The dihedral angles dictate whether a residue is right or left handed is as follows:

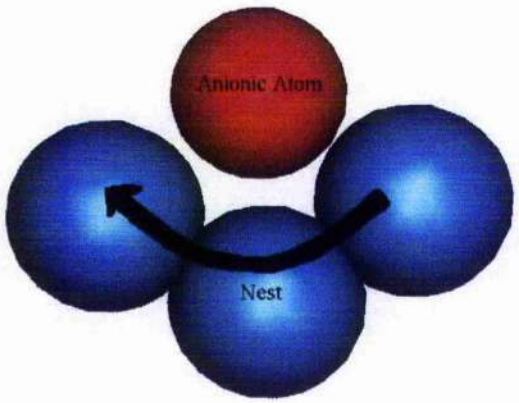
	Phi	Psi
Right handed (R)	-ve (-180° to 0°)	+ve or –ve (-180° to 180°)
Left handed (L)	+ve (0° to 180°)	+ve or –ve (-180° to 180°)

The two enantiomeric types have quite different properties because of the way the adjacent polypeptide chain has to fold to accommodate them. The RL type is more common than the LR type which is only associated with a couple of hydrogen bonded motifs.

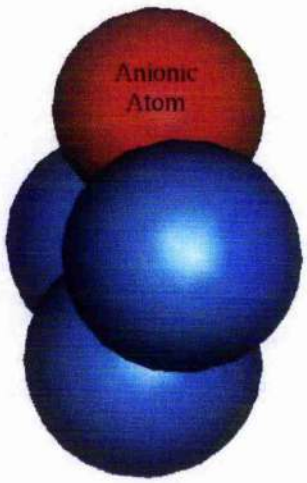
The nest is formed from three successive mainchain NH groups coming together to form a concave structure. The way the mainchain folds up is dictated by the backbone dihedral angles and therefore these control the formation of nests. Due to the nature of NH groups being slightly positively charged, the concavity therefore makes a good site for binding anionic atoms or molecules (Chakrabarti, P., 1993; Copley, R. R. & Barton, G. J., 1994). The way the three mainchain NH groups arrange themselves is that the first and third NH group point towards a common region where the primary anion would be bound. The middle NH group is not aligned with these two and points outwith the plane of the nest. Although it does not point in the correct direction it still makes some electrostatic interactions with the primary anion. More often than not a secondary anion is present, as in the paperclip motif (see figure 2.1d below), and it is this anion that is bound by the middle NH group.

Figure 2.1: Basic nest arrangement (figures a-c) and paperclip (figure d) from 1rhs (Gliubich, F. et. al., 1996)

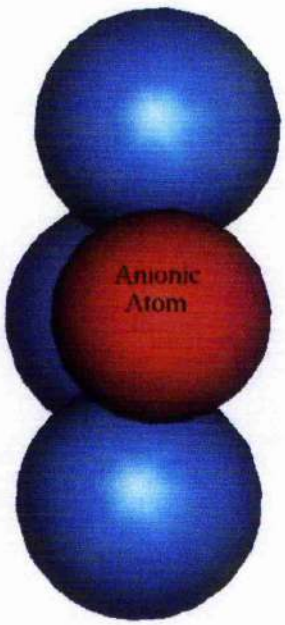
a.



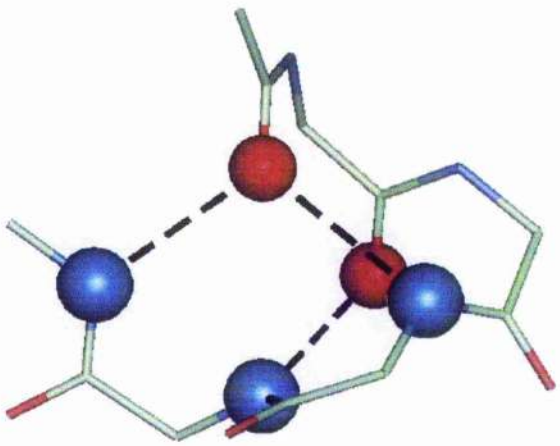
b.



c.

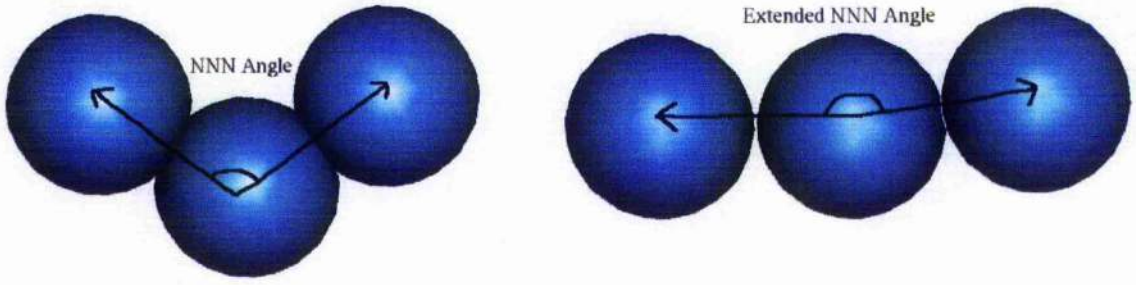


d.



The concavity of the nest is measured by taking the angle between the three nitrogen atoms of the mainchain NH groups (Figure 2.2 shows how this is measured).

Figure 2.2 : Average nest concavity and extended nest.



As the angle increases the nest loses its concavity becoming more linear and loses some of its ability to bind to anions. This is because the NH groups are more spaced out and do not point to a common region. There are also problems associated with very small angles. As the nest becomes more concave repulsive effects start to become more dominant and the NH groups clash with one another. The angle of the nest is therefore expected to be within a particular range and this is shown in figure 2.3 below. This is a histogram showing the distribution of the nest NNN angles and is split up into RL and LR types. This plot shows that the average NNN angle is different for the two types of nest with the LR motifs having greater NNN angles. This is probably a reflection of the differences between the various motifs that use each type.

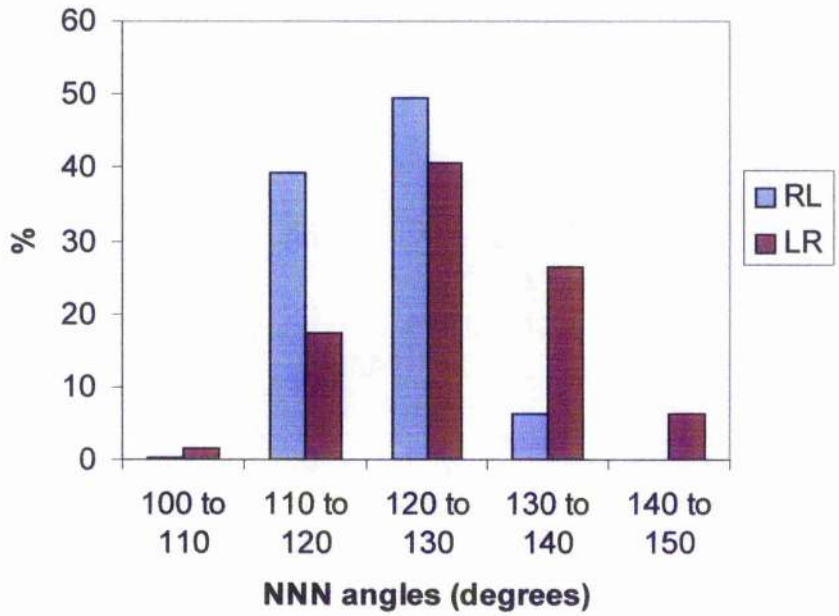
As discussed earlier, the nest is comprised of three mainchain NH groups each of which has a partial positive charge. The net effect allows the concavity to bind one or more atoms with a negative charge. In proteins the mainchain carbonyl oxygen has a partial negative charge dictated by its inherent electronegativity. There are two types of anions or anionic atoms that are commonly bound by the nest:

1. Structural – this involves other parts of the protein such as negative sidechains or mainchain carbonyl groups.

2. Functional – this involves external anions such as the phosphate group but it can involve other parts of the protein if they are involved in the mechanism.

Every nest binds to one anion but there is the potential to bind to a number of anions, the average number of anions bound being around two.

Fig 2.3: Distribution of NNN angles



2.3 Database and Methods

The primary database used in this study is a set of 67 protein structures which has been influenced by the 'PDB_SELECT' list (Hobohm, U. & Sander, C., 1994) at:

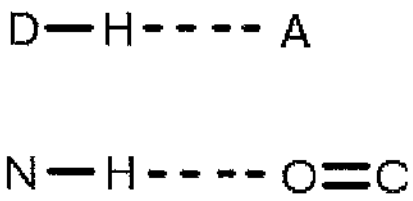
ftp://ftp.embl-heidelberg.de/pub/databases/pdb_select

All the chains in the database have less than 25% identity to each other, are refined by X-Ray diffraction at a resolution of less than 1.5Å and have an R-factor less than 20%.

The chains in the database were run through the DSSP program (Kabsch, W. & Sander, C., 1983) which is used to list the mainchain dihedral angles and other secondary structure features in an easily searchable output. The dssp files were searched for patterns of alternating enantiomeric dihedral angles as defined in section 1.2.

The new set of possible nests was subsequently analysed for hydrogen bonding patterns using the commercial molecular graphics package Quanta (QUANTA, 1997). The hydrogen bond parameters used are deliberately non-restrictive between relevant N and O or N and S atoms. This allows the detection of slightly weaker interactions and aids in the classification of motifs. Figure 2.4 below illustrates a basic hydrogen bond with donor and acceptor atoms and how it relates to the mainchain polypeptide atoms.

Figure 2.4 Basic Hydrogen Bond Arrangement (Jeffrey, G. A., 1997)



The hydrogen bond parameters we used are less restrictive and are as follows:

DHA angle =	No less than 90°
DA distance =	No greater than 3.8 angstroms for N---O
	No greater than 4.0 angstroms for N---S

When investigating functional motifs a different approach was used. For this purpose the PDB was searched using the following criteria: which ligands are bound, what the protein's function is, and homology with other members of the same family.

2.4 Occurrence of nests and relationship with other motifs

2.4.1 Occurrence of nests in the database

In the database of 67 protein chains 323 motifs occur that can be described as nests and are shown in table 2.1a. The average geometric parameters for each subgroup are given in table 2.1b. The most common type is the RL form with 79% of the database the remaining 21% being the enantiomeric (with regards to the main chain) LR form.

Key to table 2.1a:

PC = Paperclip (Schellman motif)

WPC = Wide Paperclip

Ty1 BT = Type I Beta Turn

Ty2 BT = Type II Beta Turn

Ty1BBL = Type 1 Beta Bulge Loop

Ty2BBL = Type 2 Beta Bulge Loop

AsxM = Asx-motif

StM= ST-motif

Dnest, Nnest, Snest, Tnest, Cnest = A novel motif described in chapter 3.

Table 2.1a: data from all nests in the representative database

PDB code	Protein	Res Range	Motif	RL/LR	NNN angle	Phi 1	Psi 1	Phi 2	Psi 2	Phi 3	Psi 3	Comments
1a2p	Barnase (Ribonuclease)	17-19	Alpha Turn	RLB	128.9	-102.6	-14.2	60.3	34.3	-159.9	160.8	
1a2p	Barnase (Ribonuclease)	33-35	PC	RLB	120	-73.5	-10.3	114.9	5.7	-65.2	127.5	
1a2p	Barnase (Ribonuclease)	39-41	Ty1BT	RLR	127.1	-99.1	-6.3	61.1	35.5	-104.7	15.6	
1a2p	Barnase (Ribonuclease)	60-62	Ty1BT	RLR	125.7	-66.8	-10.6	64.6	28	-74.4	-36.5	
1a2p	Barnase (Ribonuclease)	93-95	Ty1BBL +StM	RLB	125.2	-110.1	15.6	73.4	25.7	-80.2	139.5	
1a2p	Barnase (Ribonuclease)	102-104	Ty 1 prime BT	LLR	106.4	59.8	47.6	74.2	14.4	-79.9	-41.8	
1a2y	Immunoglobulin/Hydrolase complex	54-56	Ty1BBL +Dnest	RLB	119.4	-81.1	5.4	91.3	-10.2	-69.4	150.3	
1a2y	Immunoglobulin/Hydrolase complex	75-77	PC + Ty2BBL +AsxM	RLB	121.6	-100.9	2.1	54.2	44.4	-126.7	143.7	
1aba	Glutaredoxin Mutant (Electron Trans.)	27-29	PC	RLB	119.9	-81.9	4.1	55.7	44.3	-110.9	119.7	
1aba	Glutaredoxin Mutant (Electron Trans.)	55-57	PC	RLB	121.8	-76.8	-19.9	71.8	33.7	-107.5	146.5	
1aba	Glutaredoxin Mutant (Electron Trans.)	72-74	Ty1BBL +Dnest	RLB	116.8	-82.3	3.6	92.9	-7	-72.6	151.8	
1ah7	Phospholipase C	75-77	Ty2BBL +Nnest	RLB	124.2	-120	-8.7	77.6	17	-120.2	159.5	
1ah7	Phospholipase C	103-105	PC	RLB	119.8	-87	4.7	66.3	26.4	-98.4	84.9	
1ah7	Phospholipase C	130-132	Alpha Turn	RLB	117.2	-128.3	11	54.7	41.7	-101.1	104.6	
1ah7	Phospholipase C	140-142	Ty 2 BT	LRR	115	98.1	-14.4	-66.6	-38.7	-56.8	-51.5	
1ah7	Phospholipase C	166-168	Ty 2 BT	LRx	142.9	50.4	45.3	-60.7	-47.5	-123.7	-167	
1amm	Gamma B-Crystallin	68-70	PC	RLB	111.2	-112.4	15.9	62	37.4	-74.9	154.8	
1amm	Gamma B-Crystallin	116-118	WPC	RLB	115.3	-123.3	1.5	58.2	36.5	-115.4	144.5	
1amm	Gamma B-Crystallin	157-159	PC	RLB	121.1	-101.2	8.8	89	6.6	-82.9	167.5	
1arb	Hydrolase (Ser Protease)	17-19	Ty 2 BT	LRR	107.6	86.7	-12.1	-120.3	19.1	-68	-14.5	
1arb	Hydrolase (Ser Protease)	45-47	Ty1BT +AsxM	RLR	129.5	-83.2	-9.4	49.2	46.1	-110.1	10.8	
1arb	Hydrolase (Ser Protease)	47-49	Dnest	RLB	117.7	-110.1	10.8	62.4	22	-53.7	146.5	
1arb	Hydrolase (Ser Protease)	58-60	Ty1BT	RLR	133.4	-72.8	-15.8	87.1	17.3	-92.1	-0.2	
1arb	Hydrolase (Ser Protease)	112-114	PC + Ty2BBL	RLB	128.8	-96.1	8.1	55.4	43.2	-115.6	154.5	
1arb	Hydrolase (Ser Protease)	127-129	Ty1BT	RLB	135	-84.6	-16.4	51	58.8	-79.1	168	
1arb	Hydrolase (Ser Protease)	192-194	Ty 2 BT	LRB	94.48	103	-32	-66.7	-20.5	-44.6	138.2	
1arb	Hydrolase (Ser Protease)	202-204	Ty1BBL +StM	RLB	122.6	-85.7	7.3	73.5	33.4	-113.4	140.1	
1arb	Hydrolase (Ser Protease)	237-239	Ty1BT	RLB	120.2	-81.1	1.6	99.5	6.8	-157.2	157.7	
1arb	Hydrolase (Ser Protease)	251-253	Ty1BT	RLR	123.8	-74.6	3.1	42.5	42.3	-69.6	-21.5	
1arb	Hydrolase (Ser Protease)	253-255	Tnest	RLB	125.5	-69.6	-21.5	83.8	8.1	-55.6	135	
1awd	Ferredoxin	29-31	PC	RLB	118.7	-88.9	5.5	87.5	11.1	-90.7	157.6	
1awd	Ferredoxin	37-39	Functional	RLR	129.7	-75.3	-27.7	51.1	26.6	-131.6	13.8	These two are part of an
1awd	Ferredoxin	39-42	Functional	RLR	115.5	-131.6	13.8	77.9	14.1	-116	21.2	iron sulphur binding loop.
1awd	Ferredoxin	69-71	PC	RLB	121.6	-85.7	-2.5	104.6	0.5	-66.1	144.9	
1awd	Ferredoxin	88-90	Unique	LRR	139.6	59	44.8	-104.4	14.8	-54.7	-41.7	
1bgf	Transcription factor	74-76	PC	RLB	124.7	-78.7	-9.6	57.2	40	-86.7	92.5	
1bkr	Actin binding	33-35	Ty1BT + Dnest	RLR	122.8	-94.7	1.3	79.4	-9.8	-65.5	-42.6	
1bkr	Actin binding	74-76	WPC	RLB	121.3	-106.9	-6.9	74.3	16.4	-91.5	118.6	
1brf	Iron-Sulphur Protein (rubredoxin)	8-10	Ty2BBL + Functional	RLB	124.1	-118.7	-11.6	87.7	-1.4	-55.3	139.6	
1brf	Iron-Sulphur Protein (rubredoxin)	21-23	PC + Ty2BBL	RLB	116	-117.7	27	89.2	16.8	-122	115.2	
1brf	Iron-Sulphur Protein (rubredoxin)	41-43	Ty2BBL +Functional	RLB	122.9	-111.2	-10	83.5	8.5	-65.8	142.4	
1brt	Bromoperoxidase A2	49-51	PC	RLB	125	-83.5	-2.1	98	24.3	-107.6	159.6	
1brt	Bromoperoxidase A2	87-89	PC	RLB	129.6	-86.1	-11.3	58.6	52.1	-80.7	139.5	
1brt	Bromoperoxidase A2	168-170	PC	RLR	117.2	-107.3	-0.5	49.9	59.7	-62.4	-25.5	
1bxo	Penicillopepsin (hydrolase)	14-16	Ty 1 prime BT	LRB	137.3	57.4	40.8	-65.9	-40.9	-147.5	167	
1bxo	Penicillopepsin (hydrolase)	77-79	Ty1BBL +Dnest	RLB	116	-81	8.2	85	-1.7	-73.2	162.7	
1bxo	Penicillopepsin (hydrolase)	188-190	Ty1BT	RLB	114.9	-112.8	15.2	81.2	-0.2	-100.7	162.2	

Table 2.1a: data from all nests in the representative database

PDB code	Protein	Res Range	Motif	RL/LR	NNN angle	Phi 1	Psi 1	Phi 2	Psi 2	Phi 3	Psi 3	Comments
1bxo	Penicillopepsin (hydrolase)	242-244	Ty2BBL +AsxM	RLE	127	-93	-29.8	113.6	19.9	144.5	-151	
1bxo	Penicillopepsin (hydrolase)	292-294	Ty 1 prime BT	LRL	133	82.4	18.9	-75.9	-10.4	88.8	10.3	
1byq	HSP 90 (chaperone)	40-42	Unique	LRR	133.2	59	38.6	-65.7	-21.6	-54.5	-19.6	
1byq	HSP 90 (chaperone)	86-88	PC +Ty2BBL	RLB	123.9	-87.9	-9.4	60.3	42.4	-122.3	148.7	
1byq	HSP 90 (chaperone)	124-126	PC	RLB	117.9	-106.5	8.3	102.1	6.5	-68	151.5	
1byq	HSP 90 (chaperone)	134-136	PC	RLR	121.6	-93.5	2.8	81	21.8	-136.3	33.3	
1c52	Cytochrome C552	15-17	Alpha Turn	RLR	158.6	-81.7	-11.1	76	151.7	-62	-31.7	
1c52	Cytochrome C552	18-20	Ty1BT + (Nnest)	RLR	119.7	-86.2	5.9	64.9	14.3	-90.2	-3.8	
1c52	Cytochrome C552	24-26	Ty 2 BT	LR		86.7	-5.8	-135.3	-65.1	126.8	124.5	
1c52	Cytochrome C552	41-43	Ty 2 BT	LRR	120.8	91.9	-8.9	-57.5	-47	-65.4	-37.2	
1c52	Cytochrome C552	91-93	Ty1 BT	RLR	114.9	-117.5	20.4	80.4	16.6	-65.9	-38	
1c52	Cytochrome C552	114-116	Unique	RLB	143.6	-58.1	-44.4	50.7	57.3	-95.3	160.6	
1c52	Cytochrome C552	128-130	Unique	RLR	133.3	-75.1	-26.3	71.6	49.6	-76.1	-79.9	
1cbn	Crambin	19-21	Ty1BT	RLB	118.8	-75.4	-6.8	103.4	9.3	-50.2	135.5	
1cbn	Crambin	30-32	Tnest + Alpha turn	RLB	115.7	-107.1	-18	91	-5.2	-66.6	168.5	
1cex	Cutinase (Serine esterase)	63-65	Alpha Turn	RLR	149.5	-104.7	-12.3	88.5	150.9	-61.9	-31.8	
1cex	Cutinase (Serine esterase)	78-80	Unique	RLB	128.8	-94.6	-4.1	55.2	46.4	-83.1	127.4	
1cex	Cutinase (Serine esterase)	88-90	Ty1BT	RLB	113.6	-92.1	5.8	82.5	9.9	-164.1	178	
1cex	Cutinase (Serine esterase)	155-157	PC	RLL	120.6	-99.5	9.8	55.3	37.4	80.5	10.6	
1cex	Cutinase (Serine esterase)	157-159	Ty 1 prime BT	LBB	159.1	80.5	10.6	-116.8	165.7	-102	122.9	
1cex	Cutinase (Serine esterase)	179-181	Unique	RLB	117	-117.1	24.6	97.1	4.4	-102.9	172.7	
1cka	Oncogene Protein/Peptide	176-178	Ty1BBL +AsxM	RLB	117.2	-98.5	0.9	87.6	4.2	-75.1	147.5	
1ctj	Cytochrome C6	66-68	Ty 2 BT	LRB	105.5	82.7	-17.8	-99.4	-24.9	-121.6	142.2	
1ctj	Cytochrome C6	86-88	PC	RLR	120.4	-92.7	-0.2	59.8	46.8	-97.1	-11.3	
1cyo	Cytochrome B5	61-63	Ty1BT	RLB	124.8	-112.8	29.8	54.1	47.6	-60.7	135	
1edm	Coagulation factor (factor IX)	53-55	Ty 2 BT	LLR	144.4	79.2	4.5	57.9	66.9	-69.5	-21.9	
1edm	Coagulation factor (factor IX)	81-83	Ty1BT +RevEnest	RLR	107.5	-118.7	24	60.7	30.6	-96.5	-5.1	
1ezm	Elastase (Zn Metalloprotease)	18-20	Ty 2 BT	LRR	121.5	94.2	4.8	-144.4	-71.7	-71.9	-34.9	
1ezm	Elastase (Zn Metalloprotease)	29-31	Ty1BBL +AsxM	RLB	114.1	-92.8	10.5	63.6	42.1	-105.8	148.1	
1ezm	Elastase (Zn Metalloprotease)	34-36	Dnest	RLR	113.2	-107.2	7.2	109.6	-36.6	-110.1	-36.8	
1ezm	Elastase (Zn Metalloprotease)	44-46	Ty 1 prime BT	LBR	159.8	66.5	14.5	-82	158.5	-102.4	1.7	
1ezm	Elastase (Zn Metalloprotease)	91-93	WPC	RLR	120.1	-127.3	-0.8	61.5	33.7	-140.9	-176.5	
1ezm	Elastase (Zn Metalloprotease)	116-118	Dnest	RLR	110.9	-120.4	4.1	110	-14.7	-135	-42.9	
1ezm	Elastase (Zn Metalloprotease)	151-153	Snest	RLB	133.5	-56.3	-34.3	71.5	32.8	-45.6	121.4	
1ezm	Elastase (Zn Metalloprotease)	179-181	Alpha Turn	RLB	119.6	-97.3	-13.7	84	-1.8	-155.7	153.7	
1ezm	Elastase (Zn Metalloprotease)	183-185	Dnest	RLB	118.7	-105.2	-1.7	62.3	28.7	-123.3	126.5	
1ezm	Elastase (Zn Metalloprotease)	187-189	Unique	LRR	125	77.5	11.7	-60.1	-37.6	-61.7	-30.7	
1ezm	Elastase (Zn Metalloprotease)	206-208	Ty1BT +Dnest	RLR	121	-88.9	2.1	85.7	-9.9	-133.3	-31.4	
1ezm	Elastase (Zn Metalloprotease)	279-281	PC	RLB	118.2	-86	0.7	68.7	35.5	-102.7	163.6	
1ezm	Elastase (Zn Metalloprotease)	293-295	PC	RLB	115.7	-96.4	14.9	73.3	26.6	-113.7	134.4	
1fus	Ribonuclease F1	29-31	PC	RLB	120.6	-98	1.2	55.1	32.5	-117.5	167.3	
1fus	Ribonuclease F1	46-48	Ty1BT +(AsxM)	RLB	132.4	-84.4	-17.5	88.6	36.5	-66.4	139.1	
1fus	Ribonuclease F1	83-85	Ty1BBL +AsxM	RLB	114.8	-104	7	67.9	23.7	-81.2	133.6	
1g3p	Minor Coat Protein	26-28	Ty2BBL + AsxM + Tnest	RLB	126.3	-93.2	-19.9	69.9	14.3	-76.1	160	
1g3p	Minor Coat Protein	49-51	Ty1BT + Dnest + StM	RLR	125.7	-76.5	-1.2	58.4	21.6	-89	-19.5	
1g3p	Minor Coat Protein	107-109	Dnest	RLR	128.3	-69.5	-6.5	76.4	10.7	-78.3	-19	
1g3p	Minor Coat Protein	182-184	PC	RLR	117.5	-81.5	10.7	73.7	15.7	-66.7	-20.1	

Table 2.1a: data from all nests in the representative database

PDB code	Protein	Res Range	Motif	RL/LR	NNN angle	Phi 1	Psi 1	Phi 2	Psi 2	Phi 3	Psi 3	Comments
1gci	Subtilisin	11-13	PC	RLR	126.9	-83	0.6	68.3	31.4	-61.1	-38.6	
1gci	Subtilisin	19-21	PC	RLB	114.2	-90.5	10.8	85.1	14.8	-107.6	120	
1gci	Subtilisin	22-24	Tnest	RLB	111.4	-133.4	-6	88.3	5.7	-53.2	137.4	
1gci	Subtilisin	99-101	Ty1BBL +Snest	RLB	122	-73.6	-12.3	96.6	2.3	-127.6	160.7	
1gci	Subtilisin	117-119	Unique	RLB	121	-126.4	17	81.8	18.2	-69.5	142.4	
1gci	Subtilisin	145-147	PC	RLB	118.7	-87.8	11.2	89.6	7.2	-81.9	132.9	
1gci	Subtilisin	156-158	Snest	RLB	119.7	-83.5	-11.2	82.2	-4.4	-75.9	158.7	
1gci	Subtilisin	183-185	Ty1BBL +AsxM	RLB	111.5	-91.5	2	61.3	34.9	-103.5	153.8	
1ifc	APO form2 (lipid binding)	13-15	Unique	LRR	135.5	59.1	23	-68.1	-39.3	-66.9	-30.6	
1ifc	APO form2 (lipid binding)	21-23	PC	RLB	121.5	-72.7	-15.1	101.1	17.6	-86.6	134.6	
1ifc	APO form2 (lipid binding)	34-36	Dnest	RLB	121.7	-84.3	-1.9	56.3	39.6	-64.5	135.6	
1ifc	APO form2 (lipid binding)	74-76	Ty1BBL +Dnest	RLB	124.3	-64.2	-20.7	112.9	-10.9	-63.8	126.6	
1ifc	APO form2 (lipid binding)	98-100	Ty2BBL +Nnest	RLB	122.6	-103.9	-0.9	67.4	7.7	-79.8	146.1	
1isu	HIPIP (High Pot FeS)	10-12	PC	RLB	118.8	-103.5	14.1	51.3	46.1	-60.1	142.1	
1ixh	Phosphate Binding Protein	27-29	Tnest + (Alpha turn)	RLB	124.3	-123.5	-19.6	86.4	-17.1	-62.9	140.5	
1ixh	Phosphate Binding Protein	47-49	PC	RLR	123.2	-72.9	-12.5	67.4	31	-95	-6.2	
1ixh	Phosphate Binding Protein	67-69	PC	RLB	118.1	-95.5	-2.1	69.2	28.2	-104.1	159.5	
1ixh	Phosphate Binding Protein	91-93	Ty 2 BT	LRB	122.3	84.1	-0.7	-80.6	-41.1	-79.2	138.5	
1ixh	Phosphate Binding Protein	105-107	PC	RLR	120.1	-80.6	-13.1	86.3	9.2	-77.2	-29.8	
1ixh	Phosphate Binding Protein	202-204	PC	RLB	119.3	-101	8.3	57.7	37.9	-92.1	140.9	
1ixh	Phosphate Binding Protein	213-215	Ty1BBL +AsxM	RLB	123.9	-84.1	0	92.7	19.1	-116	143.9	
1ixh	Phosphate Binding Protein	291-293	PC	RLB	120	-98.6	13.4	64.7	32.3	-97.6	129	
1ixh	Phosphate Binding Protein	316-318	Ty1BBL +AsxM	RLB	117.3	-98.6	2.4	93.6	2.2	-80.1	128.6	
1jhg	Trp Operon Repressor Complex	32-34	PC	RLR	116.5	-105.7	9.6	58	35.2	-111.8	6.4	
1jhg	Trp Operon Repressor Complex	63-65	Alpha Turn	RLR	143	-65.4	-43.9	84	14.3	-89.2	-42	
1jhg	Trp Operon Repressor Complex	75-77	Alpha Turn	RLB	122.1	-100.2	-15.1	78.1	9.9	-87.8	140	
1lkk	Human P56 Tyr Kinase	215-217	Ty 1 prime BT	LBR	141	104.9	-24.7	-66.3	164.5	-62.8	-29.4	
1luc	Bacterial Luciferase (Flavoprotein)	8-10	Unique	RLB	125.8	-89.9	-20.1	57.4	36	-75.6	142.7	
1luc	Bacterial Luciferase (Flavoprotein)	34-36	PC	RLB	116.4	-88.3	10.1	83.9	15.1	-89.5	133.7	
1luc	Bacterial Luciferase (Flavoprotein)	98-100	PC	LLR	123.7	60.6	50.8	80.6	19.8	-97.6	3	
1luc	Bacterial Luciferase (Flavoprotein)	117-119	PC	RLB	127.1	-81.5	-5.1	72.8	47	-121.2	157.9	
1luc	Bacterial Luciferase (Flavoprotein)	141-143	Alpha Turn	RLB	123.9	-106.2	-19.4	88.9	6.4	-148.9	158.1	
1luc	Bacterial Luciferase (Flavoprotein)	186-188	PC	RLB	120.1	-108.6	3.8	63.4	29.8	-90.7	136.7	
1luc	Bacterial Luciferase (Flavoprotein)	215-217	PC	RLB	120.3	-85.2	-7.2	97.8	4.4	-83.4	149.8	
1luc	Bacterial Luciferase (Flavoprotein)	318-320	Tnest +(Alpha Turn)	RLB	120.2	-112.3	-20	78.8	-0.2	-70.4	145.5	
1mro a	Methyl Coenzyme reductase (Methanogenesis)	26-28	Unique	LRR	113.7	72.2	-14.7	-114.3	29.4	-99.4	-8.3	
1mro a	Methyl Coenzyme reductase (Methanogenesis)	28-30	Unique	RLB	121.4	-99.4	-8.3	94.7	14.4	60.3	-130.1	
1mro a	Methyl Coenzyme reductase (Methanogenesis)	51-53	PC	RLB	126.8	-99.9	-14.3	77	23.4	-137.6	126.7	
1mro a	Methyl Coenzyme reductase (Methanogenesis)	65-67	Unique	LRB	120.5	87.6	30	-76.2	-20.7	-111	174.9	
1mro a	Methyl Coenzyme reductase (Methanogenesis)	120-122	WPC	RLB	130	-88.6	-24.6	73.9	26.9	-109.5	164	
1mro a	Methyl Coenzyme reductase (Methanogenesis)	141-143	Ty1BT	RLB	123.1	-69.2	-11.9	118	47.6	-127.6	158.5	
1mro a	Methyl Coenzyme reductase (Methanogenesis)	239-241	PC	RLB	128.5	-109.7	7.8	58.4	42.6	-68.8	166.1	
1mro a	Methyl Coenzyme reductase (Methanogenesis)	341-343	Ty 1 prime BT	LRR	138.5	55.2	28.2	-78.1	-53.8	-60.9	-32.2	
1mro a	Methyl Coenzyme reductase (Methanogenesis)	361-363	Ty 2 BT	LRB	125.7	66.9	19.5	-114.9	-2.1	-75.7	149	
1mro a	Methyl Coenzyme reductase (Methanogenesis)	417-419	Tnest +(Alpha Turn)	RLB	122.5	-88.8	-24.7	71	21.7	-138.5	127.9	
1mro a	Methyl Coenzyme reductase (Methanogenesis)	438-440	PC	RLB	118.9	-105.5	3.4	53.9	35.9	-164	165.6	
1mro a	Methyl Coenzyme reductase (Methanogenesis)	483-485	Unique	LRR	122.3	72.8	25.9	-138.5	-50.6	-61.4	-48	

Table 2.1a: data from all nests in the representative database

PDB code	Protein	Res Range	Motif	RL/LR	NNN angle	Phi 1	Psi 1	Phi 2	Psi 2	Phi 3	Psi 3	Comments
1mro a	Methyl Coenzyme reductase (Methanogenesis)	499-501	PC	RLB	123.2	-74.9	-21.6	84.8	16.3	-63.4	142.7	
1mro a	Methyl Coenzyme reductase (Methanogenesis)	534-536	PC	RLR	125.6	-86.8	-4.9	68.8	42.5	-120.6	-2.5	
1mro b	Methyl Coenzyme reductase (Methanogenesis)	14-16	Ty1BBL +AsxM	RLB	119.2	-105	3	78.5	24	-99.8	140.9	
1mro b	Methyl Coenzyme reductase (Methanogenesis)	59-61	PC + (Tnest)	RLB	118.8	-110.6	-7.6	57.7	26.2	-89.6	90.8	
1mro b	Methyl Coenzyme reductase (Methanogenesis)	71-73	Ty 2 BT	LRB	116.2	98.8	-6.9	-105	7.3	-58.1	136.8	
1mro b	Methyl Coenzyme reductase (Methanogenesis)	109-111	Ty 1 prime BT	LRB	117.3	96.5	4.1	-77.7	-23.9	-134.3	159.1	
1mro b	Methyl Coenzyme reductase (Methanogenesis)	146-148	PC	RLB	115.7	-106.2	6.1	49.4	50.9	-64.5	133	
1mro b	Methyl Coenzyme reductase (Methanogenesis)	189-191	Ty 2 BT	LRR	126.7	91.5	5.2	-95	1.4	-57.7	-44.7	
1mro b	Methyl Coenzyme reductase (Methanogenesis)	204-206	PC	RLL	124.3	-98.6	7	58.3	36.6	62.3	20.4	These two
1mro b	Methyl Coenzyme reductase (Methanogenesis)	206-208	Ty 1 prime BT	LBR		62.3	20.4	-77.5	110.6	-54.4	-50.2	are linked
1mro b	Methyl Coenzyme reductase (Methanogenesis)	226-228	PC	RLR	121.5	-78.8	-11.8	84	-0.6	-67.7	-34.2	
1mro b	Methyl Coenzyme reductase (Methanogenesis)	246-248	Unique	RLB	120	-107.4	6.9	54.8	40.4	-57.8	138.7	
1mro b	Methyl Coenzyme reductase (Methanogenesis)	250-252	Ty 1 prime BT	LRR	140.4	59.4	23	-66.9	-50.2	-57.9	-48.1	
1mro b	Methyl Coenzyme reductase (Methanogenesis)	279-281	PC	RLR	120.6	-92.3	2.7	75.7	16.9	-74.4	-39.2	
1mro b	Methyl Coenzyme reductase (Methanogenesis)	321-323	PC	RLB	119.6	-78.7	-17.5	44	40.5	-146.3	127.8	
1mro b	Methyl Coenzyme reductase (Methanogenesis)	342-344	Tnest	RLB	120.7	-121	-32	81.8	11.6	-108.6	155.9	
1mro b	Methyl Coenzyme reductase (Methanogenesis)	350-352	Ty 1 prime BT	LRR	119.2	102.3	-12.2	-75.7	-27	-70.4	-43.7	
1mro b	Methyl Coenzyme reductase (Methanogenesis)	387-389	StM	LRR	125.7	106	-6	-102.8	-2.3	-112.9	-48.9	
1mro c	Methyl Coenzyme reductase (Methanogenesis)	40-42	Ty1BT	RLR	143.3	-63.3	-46.1	77.2	16	-84.3	-29.2	
1mro c	Methyl Coenzyme reductase (Methanogenesis)	78-80	PC	RLB	126.8	-78	-21.5	83.3	28.9	-58.2	153.6	
1mro c	Methyl Coenzyme reductase (Methanogenesis)	165-167	Ty1BBL +AsxM	RLB	114.9	-85	-3.6	82.8	12.8	-85.9	141.6	
1mro c	Methyl Coenzyme reductase (Methanogenesis)	172-174	Ty1BT +AsxM	RLR	131.3	-105.1	-3.5	40.7	73.8	-85.4	-27.3	
1mro c	Methyl Coenzyme reductase (Methanogenesis)	181-183	Ty2BBL +Tnest + AsxM	RLB	127.1	-114	-21.4	75.9	12.2	-111.6	154.2	
1mro c	Methyl Coenzyme reductase (Methanogenesis)	191-193	Ty1BBL +AsxM	RLB	127.5	-91.2	-6.5	103.1	16.4	-95.1	136.3	
1nls	Concanavalin A (Agglutinin)	17-19	PC	RLB	116.4	-114.3	21.2	91	2.8	-62.6	145.2	
1nls	Concanavalin A (Agglutinin)	58-60	Alpha Turn	RLB	138.6	-83.5	-26.8	56.4	52.8	-137.9	117.5	
1nls	Concanavalin A (Agglutinin)	162-164	Ty1BBL +StM	RLB	114.2	-85.4	1.3	85.9	8.8	-83.3	131.1	
1nls	Concanavalin A (Agglutinin)	224-226	Ty 2 BT	LRB	133.4	69.7	22.8	-98.9	22.3	-63	173.3	
1nls	Concanavalin A (Agglutinin)	231-233	Ty 1 prime BT	LRB	126.5	70.6	10.1	-109	-36.6	-125.8	148.2	
1np4	Nitrophorin 4 (Nitric Oxide Transport)	19-21	Nnest	RLB	116.1	-101.1	10.7	83.2	-4.9	-82.6	174.4	
1np4	Nitrophorin 4 (Nitric Oxide Transport)	64-66	Ty2BBL +AsxM +Tnest	RLB	127.1	-81	-19.9	77.2	7.3	-66.3	129.1	
1np4	Nitrophorin 4 (Nitric Oxide Transport)	92-94	Ty1BBL +AsxM	RLB	119	-97.4	-5.3	97.7	-1	-70.8	128.8	
1np4	Nitrophorin 4 (Nitric Oxide Transport)	156-158	PC	RLB	115.4	-90.7	6	64.4	33.4	-112.3	153.1	
1np4	Nitrophorin 4 (Nitric Oxide Transport)	169-171	Unique	RLB	122.3	-85.6	5.6	48.3	63.4	-87.1	158.3	
1opd	Phosphotransferase	53-55	Ty1BT	RLB	140	-72.1	-22.4	53.8	54.6	-71.3	86	
1plc	Plastocyanin (Cu2+ Ph6)	9-11	Ty1BBL +Dnest	RLB	114.5	-101.8	8.9	87.9	0.6	-64.2	142.9	
1plc	Plastocyanin (Cu2+ Ph6)	90-92	PC	RLB	124.1	-73.4	-11.4	100	15.1	-83.8	80	
1rcf	Flavodoxin + FMN	12-14	Functional	RLR	121.5	-107.3	15	85.4	-7.4	-69.1	-38.2	
1rcf	Flavodoxin + FMN	99-101	Ty2BT +Functional	LRR	131.3	68.3	23.3	-58	-38.6	-52.2	-47.8	Binds FMN
1rcf	Flavodoxin + FMN	112-114	PC	RLB	119.8	-91.5	13.5	102.1	-0.9	-74	147.2	
1rge	Guanyloribonuclease	24-26	PC	RLB	115.9	-94.8	-1.2	61.7	40.8	91.6	-163.1	
1rge	Guanyloribonuclease	38-40	Functional	RLR		-74.9	-21	55.3	41.7	-64	-35.7	These are a loop which
1rge	Guanyloribonuclease	41-43	Ty1BT	RLR	123.3	-82.9	-2.9	60	32.1	-75.2	-38.8	binds Ribose PO4
1rge	Guanyloribonuclease	50-52	Ty 2 BT	LRB	112	97.3	-1.4	-66.5	-35.8	-108.4	142	
1rge	Guanyloribonuclease	86-88	Ty 1 prime BT	LRB	120.1	80.3	10.8	-77	-43	-135	152.8	
1rhs	Transferase	21-23	PC	RLR	122.9	-68.6	-10.5	84.7	15.9	-73.4	-19.8	
1rhs	Transferase	87-89	PC	RLB	117.3	-95.7	10.4	79.3	25.2	-106.4	123.7	

Table 2.1a: data from all nests in the representative database

PDB code	Protein	Res Range	Motif	RL/LR	NNN angle	Phi 1	Psi 1	Phi 2	Psi 2	Phi 3	Psi 3	Comments
1rhs	Transferase	118-120	PC	RLB	120.1	-98.1	18.3	88.3	8.4	-106.1	114.8	
1rhs	Transferase	136-138	PC	RLB	126.1	-85.8	-0.9	82	15.1	-70.8	160	
1rhs	Transferase	173-175	Snest +Alpha Turn	RLR	129.1	-80.2	-30.8	63.6	21	-65.9	-35.7	
1rhs	Transferase	188-190	PC	RLR	124.2	-78.1	-20.8	98.7	8.9	-92.3	-9.5	
1rhs	Transferase	219-221	Ty1BBL +StM	RLB	122.9	-84.5	-8.9	95.1	11.5	-91.1	152.7	
1rhs	Transferase	235-237	PC	RLB	114.8	-95.8	1.3	63.2	34.2	-106.1	112.4	
1rhs	Transferase	249-251	Functional	RLR	112.7	-143.7	21	81.2	12.6	-94.7	-68.3	
1rhs	Transferase	263-265	PC	RLB	115.1	-96.2	16	86.7	15.6	-120.8	90.7	
1rie	Rieske FeS Protein (Electron Transport)	142-144	Ty2BBL +Functional	RLB	118.8	-130.3	18.7	91.1	2.3	-69.4	178.7	
1rie	Rieske FeS Protein (Electron Transport)	153-155	Unique	RLR	113.3	-121.3	19.6	53.5	32.2	-115.2	-89.5	
1rie	Rieske FeS Protein (Electron Transport)	161-163	Ty2BBL +Functional	RLB	121.7	-137.9	14	76.8	41.5	-86.1	133.8	
1rie	Rieske FeS Protein (Electron Transport)	168-170	Ty1BBL +AsxM	RLB	117.6	-99.7	-6.3	83.8	11.4	-79.7	141.5	
1whi	Ribosomal Protein	14-16	Snest	RLB	124.1	-65.2	-26.5	121.3	-30.8	-70.8	140	
1whi	Ribosomal Protein	73-75	Ty1BBL +Dnest	RLB	115.1	-83	8.4	90.8	-4.7	-87.1	161	
1whi	Ribosomal Protein	109-111	PC	RLB	119	-96.3	13.9	68.7	29.1	-93.6	60.9	
1xnb	Xylanase	11-13	Dnest	RLB	119.2	-83.2	4.7	98.3	-11.4	-79.6	137.7	
1xnb	Xylanase	156-158	PC	RLB	110.1	-108	22.2	71.2	19.1	-101.3	87.8	
1xso	Cu,Zn superoxide dismutase	65-67	Ty1BT +(AsxM)	RLB	134.5	-98.8	7	59.4	55.6	-120.6	174.7	
1xso	Cu,Zn superoxide dismutase	125-127	Ty1priBT +AsxM	LRR	123.2	85.8	7.7	-111.1	21.3	-79	-9.3	
1xso	Cu,Zn superoxide dismutase	127-129	Unique	RLB	126.7	-79	-9.3	83.1	20	-148.6	166.5	
1xso	Cu,Zn superoxide dismutase	137-139	Ty1priBT +StM	LRE	135.5	56.6	34.2	-78	-7.8	71.5	-152	
1xyz	Xylanohydrolase	525-527	PC	RLB	119.2	-101.9	1.8	74.7	27.4	-116.5	143.1	
1xyz	Xylanohydrolase	539-541	PC + Nnest	RLR	122.3	-93.6	5.6	60.2	27.6	-76.2	-12.4	
1xyz	Xylanohydrolase	589-591	PC	RLB	117.8	-113.2	12.5	59.3	38.5	-110.3	151.1	
1xyz	Xylanohydrolase	635-637	Ty 2 BT	LRB	117.8	73.8	1.8	-107.6	-17.1	-115.6	117.6	
1xyz	Xylanohydrolase	645-647	Unique	LRB	138.6	48.3	50.5	-92.2	-11.2	-91.4	137.4	
1xyz	Xylanohydrolase	650-652	Ty1BT + AsxM + Snest	RLR	122.3	-75.1	-25.1	85.6	7.8	-79.3	-15.1	
1xyz	Xylanohydrolase	690-692	Unique	LRR	123.2	59.8	33.3	-124.7	13.7	-84.1	-28.4	
1xyz	Xylanohydrolase	711-713	PC	RLB	115.7	-104	13.7	74.7	14.6	-79.3	130.2	
1xyz	Xylanohydrolase	721-723	Unique	RLB	141	-62.4	-34.5	61.4	46	-91.2	88.6	
1xyz	Xylanohydrolase	746-748	PC	RLB	118.3	-104.4	12.7	70.8	28.5	-107.5	136.7	
1xyz	Xylanohydrolase	813-815	Unique	LRR	147.8	38.3	59.7	-92.2	-2	-114.3	-175.2	
1xyz	Xylanohydrolase	820-822	Ty1BBL +AsxM	RLB	114.2	-103.2	12.2	68.2	24.7	-90.5	132.4	
1ycc	Cytochrome C	34-36	Ty 2 BT	LRB	131.6	67.8	18	-61.1	-37.2	-65.5	126.5	
1ycc	Cytochrome C	41-43	Functional	LRB	133.8	77.8	25.2	-120.9	-1.8	-66.5	130	
1ycc	Cytochrome C	55-57	Ty1BT	RLB	142.2	-74.2	-28.3	41.5	46.2	-66.5	136.6	
256b	Cytochrome B562	81-83	PC	RLB	113.1	-93.8	8.5	78.6	9.8	-93	70.8	
2ctc	Carboxypeptidase A	43-45	Ty1BBL +StM	RLB	120	-86.5	15.1	80.1	11.3	-77.4	142.4	
2ctc	Carboxypeptidase A	67-69	Unique	LRB	122.7	73.7	18.4	-61.6	-33	-104.3	110.4	
2ctc	Carboxypeptidase A	90-92	Ty1BT	BLR	91.9	-69	120.2	91.8	-5.8	-110	-45.7	
2ctc	Carboxypeptidase A	171-173	Ty 2 BT	LRB	124.8	66.4	20.4	-65	-30.6	-81	126.8	
2ctc	Carboxypeptidase A	186-188	Ty1BT	RLR	140	-64.4	-48.4	64.4	2.9	-125.1	37	
2ctc	Carboxypeptidase A	234-236	Alpha Turn	RLB	118.5	-121.5	-0.5	81	14.9	-67.8	124.2	
2ctc	Carboxypeptidase A	261-263	PC	RLB	116.2	-86.1	-2.6	96.2	8.7	-90.7	95.8	
2ctc	Carboxypeptidase A	274-276	AsxTurn	RLR	113.1	-111.9	12.9	132.6	-21.8	-53.3	-48.3	
2end	Endonuclease V	37-39	PC	RLB	117.4	-88.1	7.7	80.8	13.8	-82.8	151.4	
2end	Endonuclease V	55-57	Ty 2 BT	LRR	124.2	97	-1.9	-60.1	-49.8	-66.7	-43	

PDB code	Protein	Res Range	Motif	RL/LR	NNN angle	Phi 1	Psi 1	Phi 2	Psi 2	Phi 3	Psi 3	Comments
2end	Endonuclease V	81-83	PC	RLB	119.8	-82.1	-2.2	83.2	18.9	-76.3	146.3	
2eng	Endoglucanase	10-12	Dnest	RLB	120.3	-110.5	0.1	67.7	2.4	-73.5	152.5	
2eng	Endoglucanase	20-22	Unique	LRB	124.9	69.8	19.8	-95.6	-31.6	-156.4	162.6	
2eng	Endoglucanase	34-36	Ty1BBL +AsxM	RLB	118.5	-92.1	16.8	69.5	22.8	-97.4	137.5	
2eng	Endoglucanase	126-128	Ty 2 BT	LRB	113.9	90.1	1.8	-75	-13.7	120.5	138.3	
2eng	Endoglucanase	139-141	WPC	RLL	113.8	-148.8	14.1	108.9	-6.1	107.5	154.6	
2eng	Endoglucanase	176-178	Ty 1 prime BT	LBB	157.7	56.6	34.3	-59.8	136.7	-96.5	105.2	
2eng	Endoglucanase	197-199	Tnest +(Alpha Turn)	RLB	125.7	-95	-23.3	77.5	1.5	-109.7	112.4	
2erl	Mating Pheromone ER1	9-11	PC	RLB	127	-79.1	-6.6	70.6	37.1	-113.2	91.5	
2fdn	Ferredoxin (Electron Transport)	9-11	Functional	RLR	125.2	-101.3	4.8	60.9	42.5	-77.4	-23.3	These two form an iron
2fdn	Ferredoxin (Electron Transport)	11-13	Cnest + Functional	RLR	131.5	-77.4	-23.3	84.7	-4.2	-63.8	-32.7	sulphur binding loop
2fdn	Ferredoxin (Electron Transport)	20-22	Ty1BT +Functional	RLR	120.4	-117.8	10.5	54.4	42	-91.3	-3.4	
2fdn	Ferredoxin (Electron Transport)	38-40	Functional	RLR	124.4	-119	18.3	68.6	27.5	-68.9	-25	These two form an iron
2fdn	Ferredoxin (Electron Transport)	40-42	Cnest + Functional	RLR	129	-68.9	-25	81	6.7	-66.1	-34.1	sulphur binding loop
2fdn	Ferredoxin (Electron Transport)	49-51	Ty1BT +Functional	RLR	123.3	-108.8	6.4	53.2	45.1	-77.1	-35.3	
2hbg	Deoxy Haemoglobin	23-25	Ty 1 prime BT	LRR	128	72.8	5	-53.7	-42.2	-64.1	-35.2	
2hbg	Deoxy Haemoglobin	45-47	PC	RLB	126.7	-72.9	-28.7	90	32.4	-107.6	155.2	
2igd	IGG Binding Protein	42-44	PC	RLB	120.8	-104.1	9.9	75.6	22	-92.4	126.3	
2igd	IGG Binding Protein	54-56	Ty2BBL +AsxM +Tnest	RLB	122.4	-109.5	-2.5	57.6	41.1	-125.1	129.4	
2izh	Biotin binding protein	25-27	Ty1BBL +AsxM	RLB	120.1	-90.7	-12.3	96.4	13.1	-80.3	156.1	
2izh	Biotin binding protein	36-38	Ty1BBL +Dnest	RLB	113	-86.5	5.8	93.3	8.4	-89.3	152.5	
2izh	Biotin binding protein	86-88	Unique	RLB	125.2	-99.8	-0.6	50.4	54.2	-154.5	167.8	
2mcm	Apoprotein(Macromomycin)	59-61	Ty1BBL +AsxM	RLB	118.7	-91.7	3.4	71.9	29	-118.9	158	
2mcm	Apoprotein(Macromomycin)	79-81	PC +Dnest	RLB	127.3	-91.1	-3.8	64.6	58.1	-82	131	
2mcm	Apoprotein(Macromomycin)	101-103	Ty1BBL +StM +Snest	RLB	116.7	-99.3	3.2	86	6.4	-74.9	129.6	
2phy	Photoactive Yellow Protein	36-38	Ty1BBL +AsxM	RLB	116.3	-93	-1.5	92.5	10.1	-79.9	139.9	
2phy	Photoactive Yellow Protein	50-52	Tnest +(Alpha Turn)	RLB	124.1	-102.8	-17.6	91	-12.6	-71.9	147.7	
2phy	Photoactive Yellow Protein	85-87	PC	RLB	125.3	-79.8	-21.9	81.6	14.8	-151.3	138	
2pth	Peptidyl tRNA hydrolase	34-36	PC	RLB	120.5	-94.9	-3.7	58.7	38.1	-116.2	83.9	
2pth	Peptidyl tRNA hydrolase	44-46	PC +Ty2BBL	RLL	123.6	-108.1	4.6	53.3	41.5	174.8	178.3	
2pth	Peptidyl tRNA hydrolase	81-83	PC	RLB	127.1	-91.8	0.2	54.8	61.5	-106.4	131.7	
2pth	Peptidyl tRNA hydrolase	112-114	Ty 1 prime BT	LBR	140	93.6	-5.9	-81.9	113.8	-68.2	-20	
2pth	Peptidyl tRNA hydrolase	123-125	PC	RLL	111.8	-89.9	13.4	85	43.1	54.8	47.1	
2pth	Peptidyl tRNA hydrolase	124-126	WPC	LLB	101.5	85	43.1	54.8	47.1	-120.6	108.7	
2pth	Peptidyl tRNA hydrolase	191-193	Ty 2 BT	LRx	109.1	78.4	-9.5	-94	-16.7	e.o.p	e.o.p	
2pvb	Parvalbumin (Ca binding)	22-24	Ty 2 BT	LRB	114.7	74.8	8.8	-94.8	-7	-68.4	135.8	
2pvb	Parvalbumin (Ca binding)	33-35	PC	RLR	125	-79.9	-9.9	102.6	-1.8	-68	-35.7	
2pvb	Parvalbumin (Ca binding)	53-55	Functional +Dnest	RLR	122.2	-90.5	4.9	61.3	34.6	-84.7	-11.7	All four of these
2pvb	Parvalbumin (Ca binding)	55-57	Functional +Snest	RLB	118.1	-84.7	-11.7	93	-5.2	-136.9	150.9	motifs bind to
2pvb	Parvalbumin (Ca binding)	92-94	Functional +Dnest	RLR	131.2	-104	-0.9	76.8	12.2	-84.9	4.1	the same
2pvb	Parvalbumin (Ca binding)	94-96	Functional +Dnest	RLB	119.4	-84.9	4.1	84.6	8.1	-143	169.3	Calcium cation
2rn2	Ribonuclease H	88-90	PC	RLL	119.6	-94.6	5.1	71.3	28.8	66.8	40	
2rn2	Ribonuclease H	90-92	Ty 1 prime BT	LBB	165.3	66.8	40	-126.4	155.9	-85.7	167.5	
2rn2	Ribonuclease H	94-96	Ty1BBL +StM +Dnest	RLB	118.8	-91.9	16.1	64.3	27.3	-104.5	159.4	
2rn2	Ribonuclease H	100-102	Ty 2 BT	LRR	134.8	60.9	31.2	-53.3	-30.8	-57.9	-46.7	
2sn3	Scorpion Neurotoxin	10-12	Ty2BBL +Dnest	RLB	120.3	-120.8	-5.8	77.8	8.7	-95.6	158.8	
2sn3	Scorpion Neurotoxin	34-36	Alpha Turn	RLB	147.6	-130	-52.4	84.9	16.4	-80.5	176.9	

Table 2.1a: data from all nests in the representative database

PDB code	Protein	Res Range	Motif	RL/LR	NNN angle	Phi 1	Psi 1	Phi 2	Psi 2	Phi 3	Psi 3	Comments
2sns	Staphylococcal Nuclease (Phosphoric Diester)	54-56	Ty 2 BT	LRR	139.2	89.8	25.9	-92.8	-61.1	-50.3	-56.9	
2sns	Staphylococcal Nuclease (Phosphoric Diester)	106-108	PC	RLR	117.8	-97.8	-7.7	81.4	17.1	-101	-2.4	
2sns	Staphylococcal Nuclease (Phosphoric Diester)	135-137	PC	RLB	138.5	-111.8	-35.5	100	77.7	-101.2	132.3	
3chb	Cholera Toxin	20-22	Unique	RLB	135	-98.7	-49.5	42.2	52.5	-156.1	158.5	
3chb	Cholera Toxin	34-36	Ty 2 BT	LRB	112.1	74.6	-0.3	-130.3	40.5	-82.5	88.9	
3chb	Cholera Toxin	44-46	Ty1BBL + Nnest	RLB	116.4	-68.6	-1.8	94.4	-12.3	-58.8	131.2	
3chb	Cholera Toxin	78-80	PC	RLB	119.2	-92.8	0	60.2	32.5	-66.4	136.4	
3lzt	Lysozyme	15-17	PC	RLR	123.7	-87.5	11.5	83.2	20.8	-82.5	-14.9	
3lzt	Lysozyme	36-38	WPC	RLL	111.8	-134.1	-0.9	54.3	37.7	66.5	17.8	
3lzt	Lysozyme	38-40	Ty 1 prime BT	LBR	154.4	66.5	17.8	-96.1	116.4	-64.9	-19.4	
3lzt	Lysozyme	48-50	Ty1BBL +Dnest	RLB	116.6	-81.4	0.8	90.2	-3.2	-94.8	166.8	
3lzt	Lysozyme	56-58	PC	RLB	127.6	-109.5	9.5	50.5	53.5	-83.4	137	
3lzt	Lysozyme	66-68	Dnest	RLR	122.9	-116.9	2.8	64.4	10.3	-125.9	-1	
3lzt	Lysozyme	73-75	Unique	RLR	135.2	-110.2	-24.3	42.9	55.5	-72.5	-21.7	
3lzt	Lysozyme	76-78	Ty1BT +AsxM	RLB	128.6	-82.4	-8.4	57.4	46.9	-142.6	153.2	
3lzt	Lysozyme	101-103	PC	RLB	122.1	-90	16.8	91.3	6.9	-91.7	138.2	
3seb	Staphylococcal Enterotoxin	58-60	Alpha Turn + AsxM	RLR	123.8	-134.8	-6	62	38.5	-86.9	-12	
3seb	Staphylococcal Enterotoxin	101-103	Ty 1 prime BT	LEB	174.5	74.3	-5.1	34.1	-177.7	-162.7	-171	
3seb	Staphylococcal Enterotoxin	105-107	Ty 1 prime BT	LR	138.4	46.6	64.1	-43.9	-42.4	12.3	174.8	
3seb	Staphylococcal Enterotoxin	123-125	Ty 2 BT	LRB	116.9	87.7	-13.9	-108.1	6.1	-92.6	132.5	
3seb	Staphylococcal Enterotoxin	172-174	WPC	RLR	118.5	-126.9	12	51	42.5	-53.7	-40.5	
3seb	Staphylococcal Enterotoxin	219-221	Ty1BT +Dnest	RLB	123.7	-81.3	-6.5	55.7	32.2	-60.7	135.4	
3sil	Glycosidase	20-22	Ty1BBL +AsxM	RLB	119.3	-87.6	-4.4	96.1	11	-91.9	144.9	
3sil	Glycosidase	46-48	Ty1BBL +StM	RLB	115.8	-86.6	-5.6	88.8	-0.4	-68.2	133.4	
3sil	Glycosidase	77-79	Ty1priBT +AsxM	LRB	123.9	93.5	10.7	-77.7	-43.3	-132.4	158	
3sil	Glycosidase	89-91	Unique	RLR	130.4	-73.3	-10.9	58.6	45.2	-86.2	-46	
3sil	Glycosidase	151-153	Ty 1 prime BT	LRB	122.9	89.5	6.6	-75.5	-43.9	-129.8	145.8	
3sil	Glycosidase	168-170	Ty1BT +Nnest	RLR	145	-59.5	-42.5	80.4	-2.7	-122.9	18.2	
3sil	Glycosidase	186-188	Ty1BBL +Dnest	RLB	119.5	-78.4	-2.7	104.2	-18.7	-60	138.4	
3sil	Glycosidase	214-216	Dnest	RLR	118	-103.9	-6.4	81.9	-2	-106	-55.1	
3sil	Glycosidase	229-231	Ty1BT	RLB	143.3	-89.2	10.9	61.4	-35.3	-58.1	134.3	
3sil	Glycosidase	260-262	Ty 1 prime BT	LRB	118.4	93.8	10.4	-80.2	-44.2	-139.8	153.3	
3sil	Glycosidase	271-273	Ty 2 BT	LRB	135.4	65.2	24.4	-89.1	-27.1	-120.7	133.2	
3sil	Glycosidase	301-303	Unique	RLR	114.2	-69.9	-29.7	59.3	58.3	-66.9	-22.1	
3sil	Glycosidase	304-306	Ty1BT +AsxM	RLR	128.9	-91.4	1.5	55.3	51.9	-162.9	-165.5	
3sil	Glycosidase	308-310	Ty1BT +AsxM	RLB	144.5	-77.4	-19.1	49.6	54	-65.6	163.7	
3sil	Glycosidase	321-323	Ty2BBL +AsxM +Snest	RLB	126	-100.7	-20.9	81.4	3.7	-63.9	149.3	
5p21	C-H-Ras P21	12-14	Functional	BLR	100.1	-59.7	136.7	80	5.3	-76.7	-14	These two form the
5p21	C-H-Ras P21	14-16	Functional	RLR	122.1	-76.7	-14	108.5	22.6	-57.2	-47.9	P-loop
5p21	C-H-Ras P21	25-27	Alpha turn	RLB	130.1	-114.9	-0.6	57.4	55.2	-165.9	152.1	
5p21	C-H-Ras P21	104-106	PC	RLB	120.5	-89.3	5.2	48.5	45.1	-152.1	144.4	
5p21	C-H-Ras P21	117-119	Unique	LRR	136.4	71.6	33.8	-70.6	-6.8	-70.7	-21.7	
5p21	C-H-Ras P21	137-139	PC	RLB	121.6	-86.3	-15.1	77.3	27.6	-126.7	159.7	
5p21	C-H-Ras P21	148-150	Ty2BBL +Tnest	RLB	117.1	-94.4	-11.1	78.5	6.2	-65.4	132.3	
5p21	C-H-Ras P21	151-153	Ty 2 BT	LRR	128.6	74.1	-5.5	-65.2	-52.3	-65.2	-37	
5pti	Trypsin Inhibitor	27-29	Ty2BBL +AsxM	RLB	126.2	-94.4	-17.8	65.8	32.8	-141.5	157	
5pti	Trypsin Inhibitor	36-38	Unique	RLB	128.3	-81	-8.6	104.9	-7.4	-146.3	158.2	

Table 2.1a: data from all nests in the representative database

PDB code	Protein	Res Range	Motif	RL/LR	NNN angle	Phi 1	Psi 1	Phi 2	Psi 2	Phi 3	Psi 3	Comments
5ptp	Beta Trypsin (Serine protease)	25-27	Ty 2 BT	LRB	123.4	63.3	22.2	-83.6	-7.4	-129.1	77.8	
5ptp	Beta Trypsin (Serine protease)	77-79	Functional	RLR	126.8	-99.6	-12.1	116.4	-15.8	-106.8	1.4	
5ptp	Beta Trypsin (Serine protease)	98-100	Ty2BBL +AsxM +Tnest	RLB	122.5	-98.3	-15.8	72	12.7	-68.4	134.8	
5ptp	Beta Trypsin (Serine protease)	174-176	Ty 2 BT	LRB	121.7	65.4	7.2	-116.3	-12.2	-93.8	126.9	
5ptp	Beta Trypsin (Serine protease)	193-195	Ty2BT +Functional	LRB	106.1	104.7	-14.8	-84.9	-16.9	-49.5	139.4	
7rsa	Ribonuclease A	33-35	PC	RLR	121.4	-85.3	4.6	67.7	29.2	-82.8	-15.8	
7rsa	Ribonuclease A	67-69	Ty1BBL +Nnest	RLB	119.5	-85.3	1.7	100.4	-15	-61.6	154.3	

Table 2.1b: Average parameters in nest subclasses

RL	NNN angle	Phi1	Psi1	Phi2	Psi2	No. Anions
α-Turn	128.5 (± 10.3)	-108 (± 21)	-16 (± 19)	72 (± 14)	26 (± 20)	1.6 (± 0.6)
Asx-nest	118.8 (± 4.7)	-101 (± 16)	2 (± 6)	79 (± 19)	4 (± 22)	1.5 (± 0.4)
ST-nest	122.8 (± 5.6)	-95 (± 24)	-22 (± 8)	84 (± 14)	3 (± 17)	2.0 (± 0.4)
Paperclip	120.8 (± 4.5)	-92 (± 11)	1 (± 12)	74 (± 16.0)	28 (± 16)	2.0 (± 0.3)
Type 1 β-Turn	127.7 (± 9.6)	-87 (± 17)	-6 (± 18)	67 (± 19)	28 (± 24)	1.8 (± 0.5)
Type 1 β-Bulge Loop	118.2 (± 3.6)	-89 (± 10)	2 (± 8)	86 (± 12)	10 (± 15)	2.1 (± 0.2)
Type 2 β-Bulge Loop	123.5 (± 3.1)	-108 (± 15)	-10 (± 13)	78 (± 12)	15 (± 13)	2.0 (± 0.5)
Unique	126.8 (± 8.6)	-92 (± 20)	-11 (± 21)	65 (± 19)	38 (± 20)	1.2 (± 0.6)
Totals	122.3 (± 6.9)	-94 (± 17)	-3 (± 15)	76 (± 18)	22 (± 20)	1.9 (± 0.5)

LR	NNN angle	Phi1	Psi1	Phi2	Psi2	No. Anions
Unique	129.5 (± 9.7)	65 (± 13)	29 (± 19)	-90 (± 25)	-15 (± 24)	1.0 (± 0.9)
Type 2 β-Turn	121.7 (± 11.7)	81 (± 14)	4 (± 17)	-84 (± 36)	-20 (± 31)	1.3 (± 0.4)
Type 1' β-Turn	126.8 (± 9.7)	75 (± 18)	19 (± 19)	-67 (± 41)	-30 (± 23)	2.0 (± 0.9)
Totals	124.5 (± 11.2)	74 (± 23)	16 (± 25)	-77 (± 45)	-20 (± 28)	1.5 (± 0.8)

The numbers in brackets indicate the standard error associated with each average value.

2.4.2 Hydrogen bonded motifs involving nests

Many of the nests listed in table 2.1 are part of previously defined hydrogen bonded motifs. How these motifs are described is clarified through the use of their characteristic hydrogen bond patterns. All of the relevant hydrogen bonded motifs are shown diagrammatically in figure 2.5b below. Figure 2.5a (Milner-White, E. J. & Poet, R., 1987) is a diagram explaining the hydrogen bond colour scheme used.

Fig 2.5a:

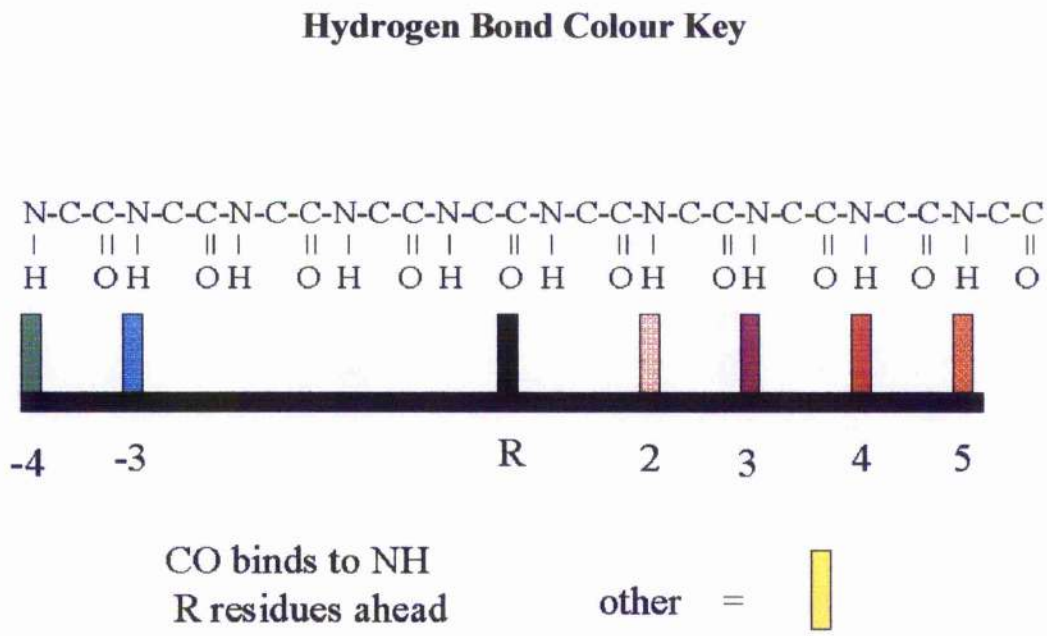
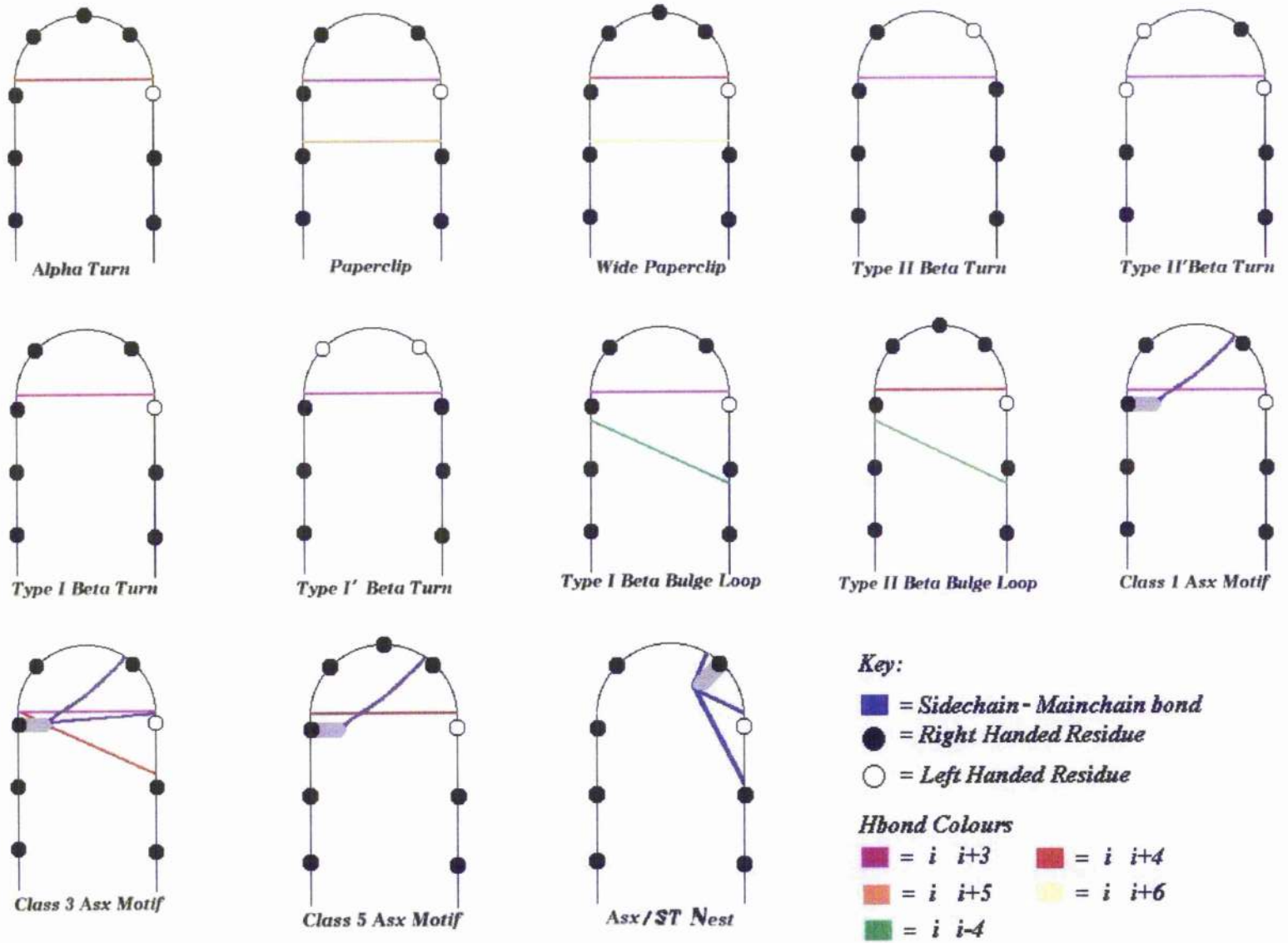


Figure 2.5b : Hydrogen bond diagrams



2.4.3 Association of nests with types of hydrogen bonded motifs

The nests that occur in the presence of other hydrogen bonded motifs can associate themselves with one, two or more motifs. Although this poses problems relating the number of nests with the number of motifs there is still valuable data in the distribution of the motifs. Tables 2.2 and 2.3 show the breakdown of the occurrence of nests associated with each type of motif.

The tables show that 27% of RL nests are part of Schellman loops (paperclips); 16% are within B-bulge loops; 12 % are part of Asx- or ST-motifs; and a further 17% form the novel Asx- and ST-nests that are the subject of the next chapter. The tables also show that even though LR nests are less common they do occur in particular characteristic situations i.e. the type I and type II β-turns.

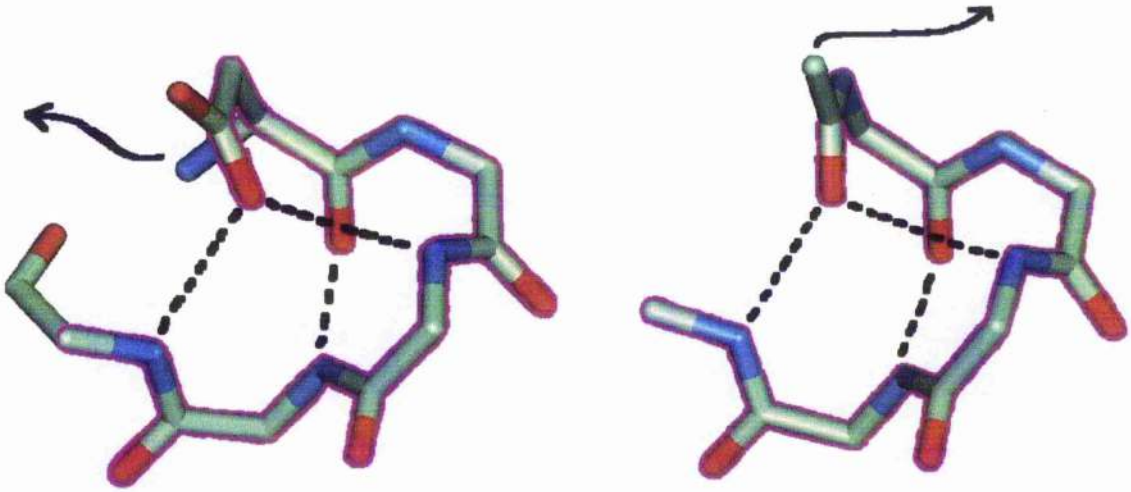
One of the most striking associations in the tables is that between Asx- ST-motifs (Wan, W-Y & Milner-White, E.J., 1999 - 1 and 2) and β-bulge loops. Almost all Asx – ST-motifs that are associated with a nest are also part of a β-bulge loop. This seems to suggest that these two motifs are in some way more stable when found together. There are no examples in our database of a β-bulge loop occurring on its own as a stable entity but this does not mean it does not occur in an isolated form at all.

A possible explanation for this association is structural mimicry (Eswar, N. & Ramakrisnan, C., 1999). The 3D structure of the Asx- ST-motif with β-bulge loop is very similar to that of the Paperclip or Schellman loop. Since the paperclip is a stable structure (occurring on its own in places) it might be expected to be a good shape to attempt to mimic. The combined Asx- ST-motif with β-bulge loop achieves this by replacing part of the mainchain loop of the paperclip with the sidechain of an Asx or ST residue as can be seen in figure 2.6 below.

Figure 2.6: Structural mimicry

a. Asx motif + β-bulge loop

b. Paperclip



The black arrow indicates the continuation of the mainchain. The structural mimicry is highlighted by a purple outline. The hydrogen bonds have also been included to emphasise the similarity.

Table 2.2: Distribution of RL nests by motif type.

Paperclip	86															
Wide Paperclip		7														
Type 1 β-Turn			19													
Type 1 β-Bulge Loops				0												
Type 2 β-Bulge Loops	5				0											
Asx-Motif			8*	17	7$	1										
ST-Motif				8#			0									
D-Nest	1		4	9	1			10								
N-Nest	1		2	2	2				1							
S-Nest				1						3						
T-Nest	1				1						3					
C-Nest												0				
α-Turn						1				1	6		10			
Functional			2		4			3		1		2		9		
Unique															18	
	Paperclip	Wide Paperclip	Type 1 β -Turn	Type 1 β -Bulge Loop	Type 2 β-Bulge Loop	Asx-Motif	ST-Motif	D-Nest	N-Nest	S-Nest	T-Nest	C-Nest	α-Turn	Functional	Unique	Totals
Numbers of nests within motif type	94	7	35	37	20	34	8	29	8	9	15	2	18	21	18	355

* = One of these also has an S-nest motif

\# = One of these has an D-nest and one has an S-nest

$ = Four of these have a T-nest and one has an S-nest

Table 2.3: Distribution of LR nests by motif type.

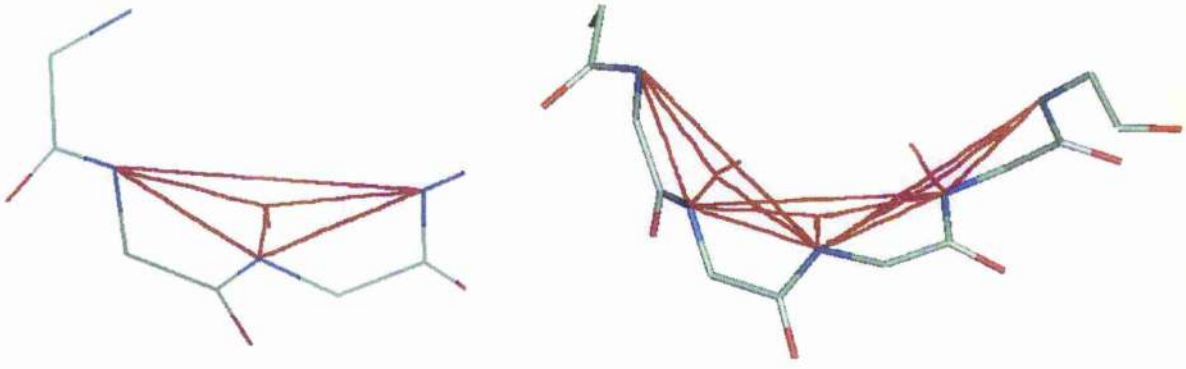
	Type 1' β-Turn	Type 2 β-Turn	Asx-Motif	ST-Motif	Paperclip	Functional	Unique	Totals
Type 1' β-Turn	13							
Type 2 β-Turn		30						
Asx-Motif	2		0					
ST-Motif	1			1				
Paperclip					1			
Functional		2				2		
Unique							13	
Numbers of nests within motif type	16	32	2	2	1	4	13	70

2.5 Compound and Tandem

Further examination of table 2.1a shows that two or more nests often follow each another closely. This prompted an investigation into the formation and function of compound and tandem nests. Compound nests occur where the nests overlap so that the residues alternate between R and L forms (e.g. RL, RLR, RLRL) whereas tandem nests are where two nests sit side by side (e.g. RLLR, LRRL, RLLRRL).

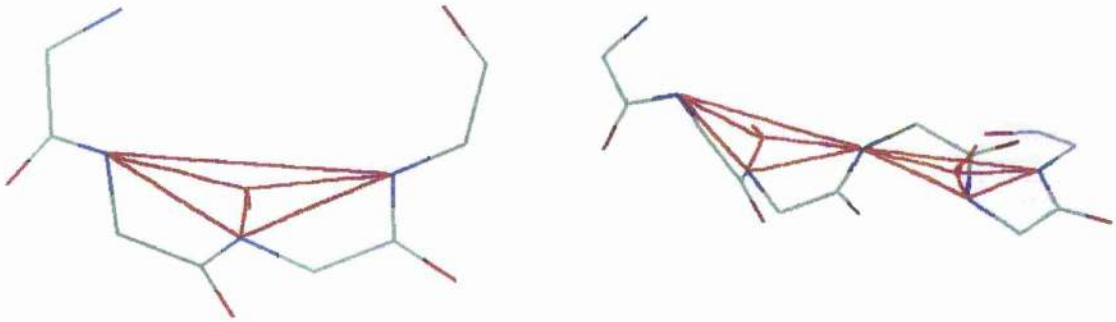
In compound nests the result is a long chain with all the overlapping nests facing a similar direction. This basically forms a much wider nest that is capable of binding a larger anionic group of atoms such as the phosphate ion.

Figure 2.7a : Diagram illustrating single nest going to a compound nest (planes of the nests are shown)



The tandem nest is not as common as the compound nest, which is probably due to the way subsequent nests align themselves. There is a greater change in the direction that adjacent nests face meaning that they no longer point to a common point or region. This makes it more difficult for the tandem nest to bind one large anion. Tandem nests tend to form structural features (e.g. holding distant parts of the protein together or simply allowing the polypeptide chain to make a turn) and only seem to perform functional roles when found in conjunction with one or more compound nests.

Figure 2.7b : Diagram illustrating single nest going to a tandem nest (planes of the nests are shown)



2.6 Functional Nests

Some of the examples found in the database turned out to have specific functional roles (e.g. P-loop of p21 Ras). Many proteins bind anions to perform a function such as to catalyse reactions or to activate other proteins and their structures have been solved with the ligands or substrate analogues bound. There is therefore considerable interest in whether nests are involved in the binding regions of these proteins and how they might work.

As mentioned in the Methods (section 2.3) investigating the anion-binding proteins involves a different strategy: simply knowledge-based browsing. The first to be detected in this way is the P-loop, a glycine rich sequence which binds triphosphates in the largest superfamily of ATP- and GTP-binding proteins. Other functional sites include oxyanion holes in serine proteases; Iron-sulphur proteins (Dauter, Z. et. al., 1996); Flavodoxin (Rao, S, et. al., 1992) active site; and, surprisingly, certain calcium-binding sites. These are shown along with other examples (Meador, W. E. et. al., 1993; McLaughlin, P. J. et. al., 1993; Bullough, P. A. et. al., 1994; Harrison, D. H. et. al., 1997; Bhatnagar, R. S. et. al., 1998; Scapin, G. et. al., 1995; Vassylyev, D. G. et. al., 1993) in table 2.4 below:

Table 2.4: Functionally important nests and compound nests

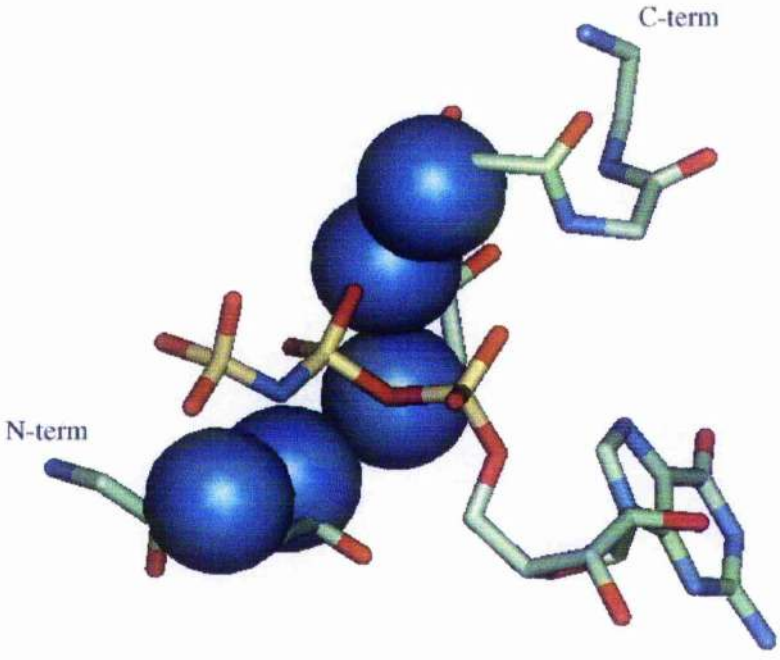
Nest type	Res. No.	PDB Code	Protein	Anion	Ligand
LR	193	2ptc	trypsin	O	peptide
LRLR	13	5p21	ras	P	GTP
LRLR	48	1aqu	oestrogen sulfotransferase	P	pyridoxal phosphate
RLR	12	1rcf	flavodoxin	P	FMN
RLR	88	1amo	NADPH cyt p450 reductase	P	FMN
RL	142	1rie	Rieske iron-sulfur protein	I	[2Fe2S]
RLRLR	39	1a70	spinach ferredoxin	I	[2Fe2S]
RLRLRRLR	57	1qla	fumarate reductase	I	[2Fe2S]
RLLRLR	46	1cje	adrenodoxin	I	[2Fe2S]
RLRLR	9	2fdn	Clostridium ferredoxin	I	[4Fe4S]
RLRLR	148	1feh	periplasmic hydrogenase	I	[4Fe4S]
RLRLR	346	2tmd	trimethylamine dehydrogenase	I	[4Fe4S]
RLR	387	1qj2	CO dehydrogenase		molybdo-pterin
RL	212	1az2	aldose reductase	P	NADP
RLR	37	1rge	ribonuclease	O	of guanine of rntd
LRLR	179	2nmt	N-myristoyl transferase	P	myristoyl-CoA
LRLR	119	1bo4	N-acetyl transferase	P	acetyl CoA
RLR	129	1opr	orotate P.R.transferase	P	PRPP
RL	741	1hqm	RNA polymerase β-chain	-	-
RLR*	3	1qfu	Haemagglutinin β-chain	O	D112
LR*	92	1cog	gelsolin	O	D87...Ca+
RLRL*	22	1cdm	calmodulin	O	D20 D22 D24...Ca+

Each line gives details of a nest, compound nest or tandem nest that is of functional importance. * indicates that the immediate anionic group bound in the nest is proteinaceous and not strictly a ligand. In the anion column a single letter indicates that an anionic atom or group binds in the nest, and shows whether it is a carbonyl or carboxylate oxygen atom (O), a phosphate group (P) or an iron-sulphur center (I).

2.6.1 P-loops

The P-loop (Phosphate-binding loop) is a well described ATP- or GTP-binding loop present in a large superfamily of important proteins which includes G-proteins and kinases (Via, A. et. al., 2000; Dreusicke, D. & Schulz, G. e., 1986; Pai, E. F. et. al., 1990). The main feature of the P-loop is a long compound LRLR nest that forms a binding site for the β-phosphate of ATP or GTP and is shown diagrammatically in figure 2.8. However, this is an example of a motif where the ligand also binds to the free main chain NH groups at the N-terminus of an alpha helix. On closer inspection of the hydrogen bonding in figure 2.8 it becomes evident that this interaction is in addition to the compound nest and does not interfere with it. Therefore the P-loop is actually more accurately described as a compound LRLR nest and an adjacent helical N-terminus that collectively bind to the α- and β-phosphates of the GDP substrate.

Figure 2.8: Diagram showing the arrangement of the P-loop



The P-loop, which is retained throughout the superfamily, has a highly conserved GxxxxGKS/T consensus sequence (where the underlined xxGK section forms the LRLR compound nest). A list of aligned P-loops from members of the superfamily is shown in table 2.5 with the LRLR compound nest highlighted. The most distantly

related member of this family is a sulphotransferase (Kakuta, Y. et. al., 1997) which transfers a sulphur to or from adenosine 3'-5'-diphosphate. In this protein there is a characteristic P-loop visible but here the compound nest is used to bind the 5'-phosphate of the substrate.

Table 2.5 List of aligned P-loop sequences from different proteins

PDB code	Sequence
1GIA	LGAG**ESGK**STIV
5P21	VGAG**GVGK**SALT
1HUR	VGLD**AAGK**TTIL
1MM	TGES**GAGK**TENT
1AQU	ATYP**KSGT**TWIS
conformation	XXXX**LRLR**RRRR

The P-loop is designed to bind to the α- and β-phosphates of ATP and GTP. Some proteins with P-loops, such as p21Ras, are GTPases and can be described as molecular switches. The activity of Ras is controlled by cycling between inactive GDP-bound and active GTP-bound forms. Studies have shown that GDP and GTP bind with equal affinity to p21Ras and that there are major structural changes between the two bound forms. All things being equal a basic equilibrium would be expected to be reached between the GDP and GTP bound forms. This equilibrium is never reached because of a number of other processes and proteins altering the balance (e.g. GTPase activating proteins [GAPs] tilt the balance in favour of the GDP bound inactive form by upregulating the GTP hydrolysis action of the Ras protein). It is the levels of these other factors changing the balance that determine whether Ras is active or inactive and therefore whether the signalling pathway is on or off.

If the P-loop was designed to bind not only to the α- and β- but also to the

γ-phosphate, there would be hydrolysis of GTP to GDP but no dissociation of product and essentially no change in the protein or its "signalling state". It is therefore important that the compound nest is situated at one end of the conserved sequence so it does not come into proximity with the third γ-phosphate of its substrate.

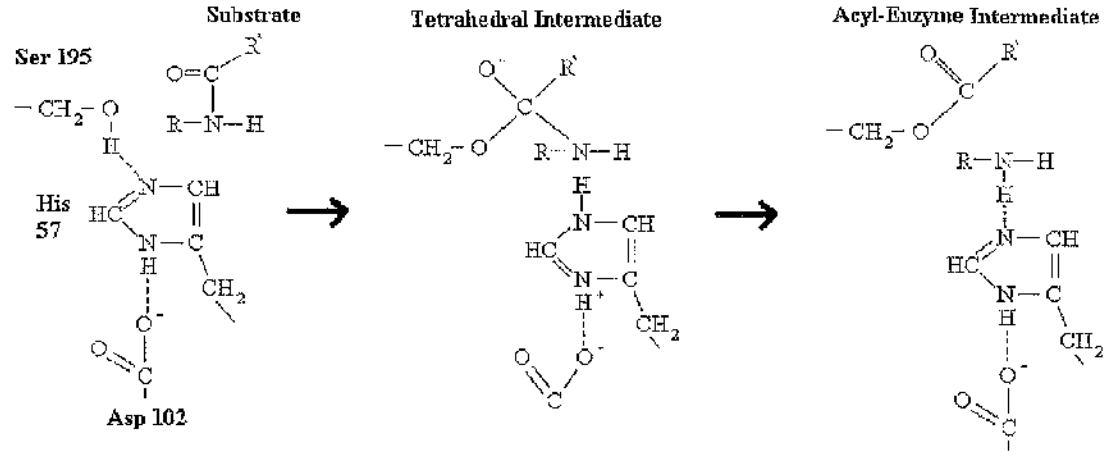
2.6.2 Oxyanion Holes Of Serine Proteases

Serine proteases (Marquart, M. et. al., 1983) catalyse the hydrolysis of peptide bonds between selected residues depending on the particular enzyme being used. Hydrolysis of the peptide bond occurs in two stages:

1. Binding of substrate and hydrolysis of the amine (or alcohol) component of the substrate peptide leaving a covalently bound acyl-enzyme intermediate
2. Hydrolysis of the acyl component from the acyl-enzyme intermediate to regenerate functional enzyme and release second product.

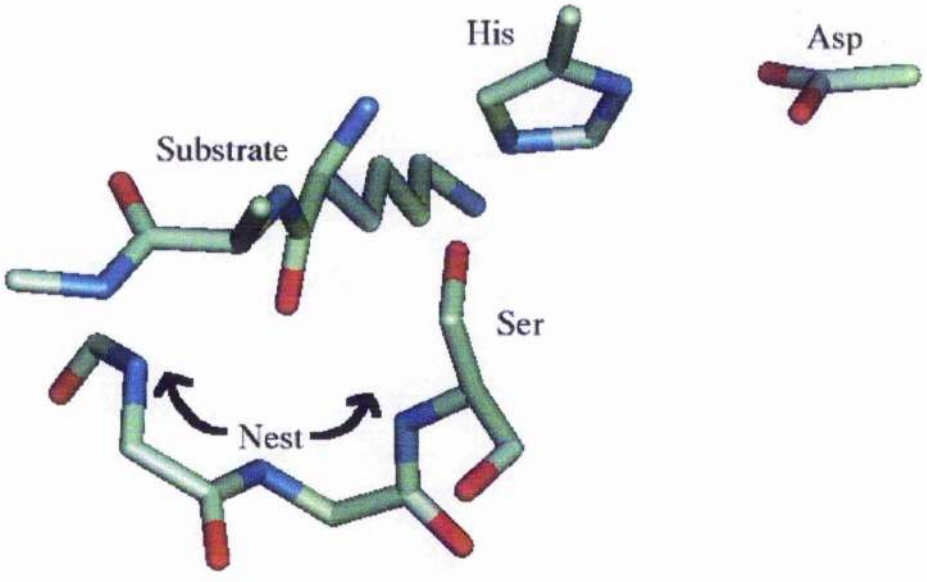
In the first step it is thought that a highly reactive serine (residue 195) nucleophillically attacks the carbonyl carbon atom of the peptide bond that is to be cleaved in the substrate. This forms a covalently bound tetrahedral intermediate shown in the reaction diagram below (figure 2.9a). The carbonyl carbon atom has changed from a trigonal arrangement to a tetrahedral one, therefore the carbonyl oxygen has only one bond and extra electron pairs. This oxygen is thus known as an oxyanion and is stabilised by hydrogen bonding to NH groups on the main chain of the protein at what is known as the oxyanion hole.

Figure 2.9a: Catalytic diagram: substrate to acyl-enzyme intermediate (Stryer, L. 1988).



The oxyanion hole found in all serine proteases has a characteristic GDS sequence which forms a simple LR nest visible whether or not substrate is bound (see figure 2.9b). This is an example of a nest with a vital function in an enzyme mechanism.

Figure 2.9b The oxyanion hole:



Serine proteases are not the only protein family which use oxyanion holes but, as far as can be seen, they are the only ones that have hydrogen bonding networks using nests.

2.6.3 Iron-sulphur Proteins

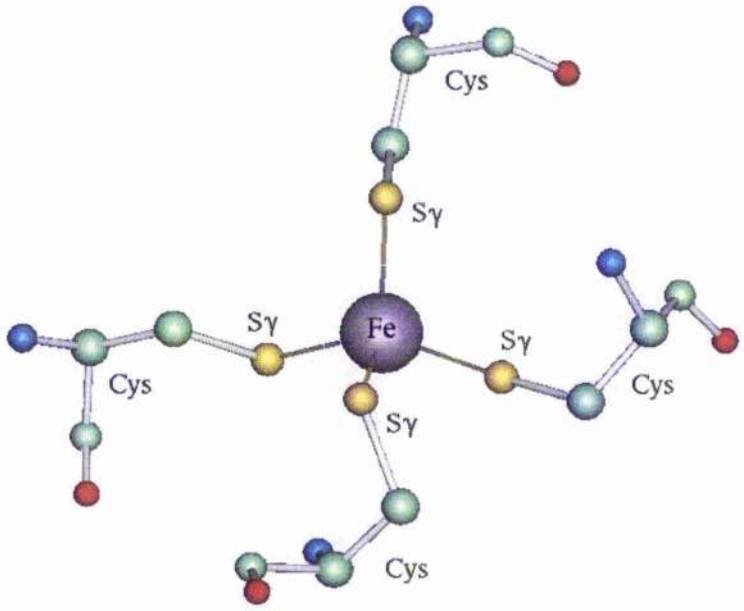
The Iron-sulphur proteins (Beinert, H. et. al., 1997; Sticht, H. & Rosch, P., 1998) are characterised by one or more non-haem iron ions ligated to inorganic sulphur and/or cysteine sulphur. These proteins are found in all organisms and have a wide variety of functions: catalytic centres, electron transfer, iron and oxygen sensors. All iron-sulphur proteins can be broadly classified into one of two groups: the ferredoxins & rubredoxins (which contain just iron-sulphur centres and generally show electron transfer activity rather than classical enzymatic activity); and the complex iron-sulphur proteins which have additional prosthetic groups such as flavin, haem or other metals (and show enzymatic activity). It is the ferredoxins that are of interest as they exhibit some of the longest compound and/or tandem nests found to date.

Within the ferredoxin/rubredoxin subgroup there are further divisions determined by the type of iron-sulphur centre found. It should be noted that there are variations to cysteine ligation such as histidine ligation in the Rieske iron-sulphur proteins (Iwata, S. et. al., 1996). There are four key types of ferredoxin defined by the number of iron ions and free sulphides associated with each.

1. [Fe]

Rubredoxin type. Here there is a single iron coordinated by four cysteines which can exist in either the oxidised Fe^{3+} or reduced Fe^{2+} forms. There are no nests associated with this type.

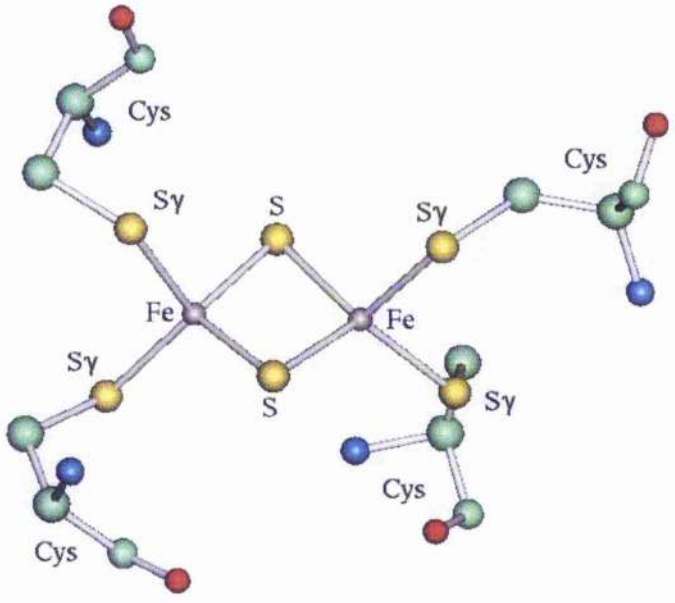
Figure 2.10 a: Rubredoxin type



2. [2Fe-2S]

In this cluster there are two high spin iron ions, two free sulphides, and four ligated cysteines (Binda, C. et. al, 1998). Compound and/or tandem nests are often associated with this group. The cluster can exist in an oxidised (Fe^{3+} Fe^{3+}) or reduced (Fe^{3+} Fe^{2+}) state.

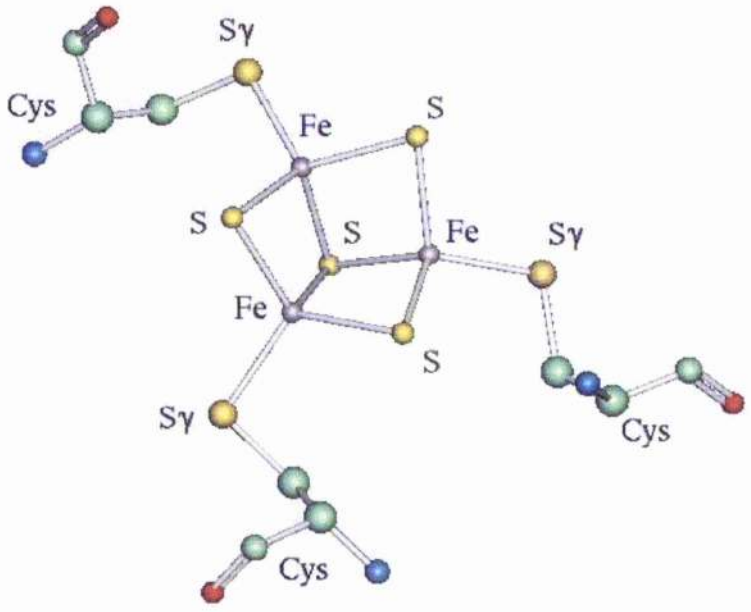
Figure 2.10 b: [2Fe-2S] type.



3. [3Fe-4S]

Here three iron ions are bound to four free sulphides and three cysteines. There are no nests associated with this group as far as can be determined. Once again the centre can be found in either an oxidised (3 Fe^{3+}) or reduced (2 Fe^{3+} Fe^{2+}) state.

Figure 2.10 c: [3Fe-4S] type.



4. [4Fe-4S]

This type of centre is more complicated and is another group that compound and/or tandem nests are found (Dauter, Z. et. al., 1996; Barber, M. J., 1992; Peters, J. W., 1998). These centres have four iron ions surrounded by four free sulphides and four cysteine groups. There are three possible oxidation states possible with this subgroup (Sticht, H. & Rosch, P., 1998) but only two are possible as the oxidised and reduced forms. The levels chosen depends on the type of [4Fe-4S] protein involved and is explained in table 2.6 below.

Figure 2.10 d: [4Fe-4S] type.

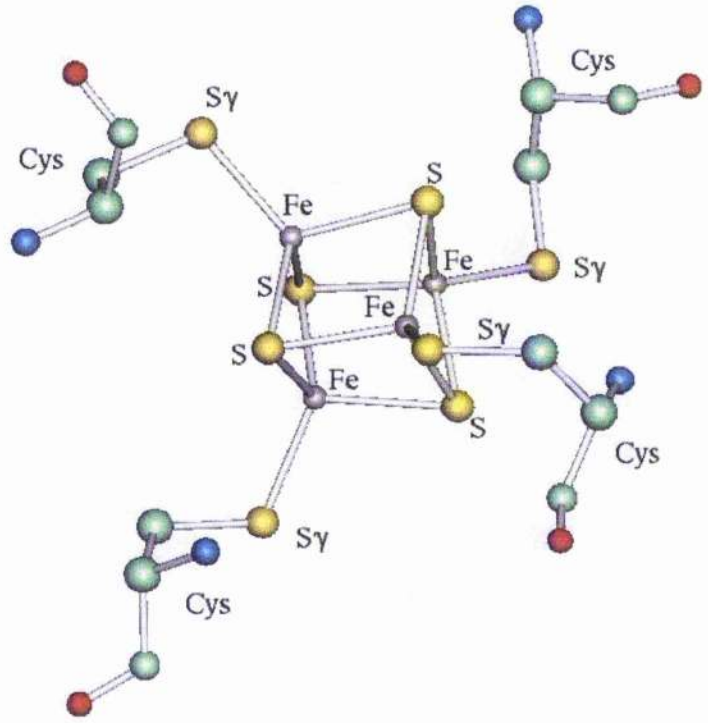


Table 2.6 Oxidation states for [4Fe-4S] proteins

Redox State	Symbol	Iron Charges	[4Fe4S]Ferredoxins	HIPIPS*
Oxidised	C+	3 Fe^{3+} 1 Fe^{2+}		Oxidised Form
Intermediate	C°	2 Fe^{3+} 2 Fe^{2+}	Oxidised form	Reduced Form
Reduced	C-	1 Fe^{1+} 3 Fe^{2+}	Reduced Form	

*HIPIP = HIgh Potential Iron-sulphur ProteinS (Cowan, J. A. & Lui, S. M., 1998)

The main question is why are nests associated with some types of Iron-sulphur complex but not with others, or even why they are associated at all. Since we are regarding nests as anion-binding sites the question to be asked is whether iron sulphur centres can be regarded as anionic. This requires some consideration. If the cysteine, as well as the non-cysteine, sulphur atoms are taken into account the net charge on the entire centre is negative (Stephens, P. J. et. al., 1996; Cowan, J. A. & Lui, S. M., 1998) and is in essence a giant anion. This is in line with the nest being described as an anion-binding feature. Examination of the way the nests interact with the [2Fe-2S] and [4Fe-4S] clusters shows that there is some hydrogen bonding with the free sulphides, but this is not always seen. Since the function of these proteins is to alternate between reduced and oxidised forms it is interesting to note that the formation of hydrogen bonds would be likely to favour the formation of the reduced form of the protein (Carter, C. W. Jr., 1977; Stephens, P. J. et. al., 1996).

2.6.4 Calcium-binding Sites

A more unusual function for a nest is seen in proteins with EF-hand (Allouche, D. et. al., 1999) domains such as calmodulin (Meador, W. E. et. al., 1993; Milner-White, E. J., 1999). Here the nest is indirectly involved in binding calcium cations. EF-hand domains are well studied and it is known the calcium-binding region is the most conserved section with a characteristic DxDxDGxxxxE consensus sequence. The calcium is coordinated by seven oxygen atoms: one from each of the three conserved Aspartates; one from a mainchain carbonyl group; two from the conserved Glutamate; and one from a water molecule.

The particular arrangement of all these sidechains brings three negatively charged aspartates and one glutamate in close proximity to one another. This is highly unfavourable electrostatically and is alleviated by a compound RLRL nest (DxDG in the consensus sequence). The compound nest binds to the free oxygen in each of the three Aspartate sidechains and helps hold the whole structure together. This is shown in figures 2.11a and 2.11b.

It would seem strange to bind a cation in this way with a calcium surrounded by negative charge which is subsequently stabilised by a compound nest, however it seems to be the best way to stabilise three aspartate side chains next to each other. Another use for the nest may be to collectively reduce the negative charge surrounding the calcium. Calcium is a Ca^{2+} cation yet is surrounded by four sidechains with a net negative charge. Even allowing for pKa and microenvironmental differences this is an excessive concentration of electrons, therefore it would make sense to mitigate the negativity of the three Aspartate sidechains. This is achieved by hydrogen bonding one of the oxygens of each of the carboxylate sidechains to the compound nest.

It may be that the EF-hand motif evolved from an ancestral cation-binding loop that was used to bind larger more cationic metals, and the use of the compound nest may have arisen after the loop was used for calcium-binding.

Figure 2.11a: Hydrogen bonding diagram of EF hand.

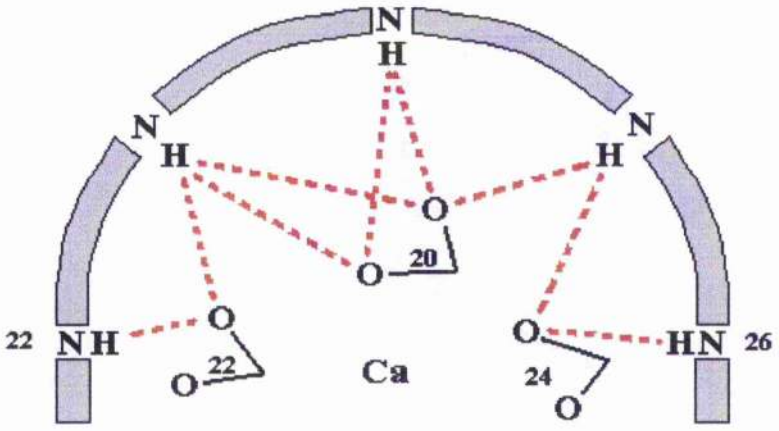
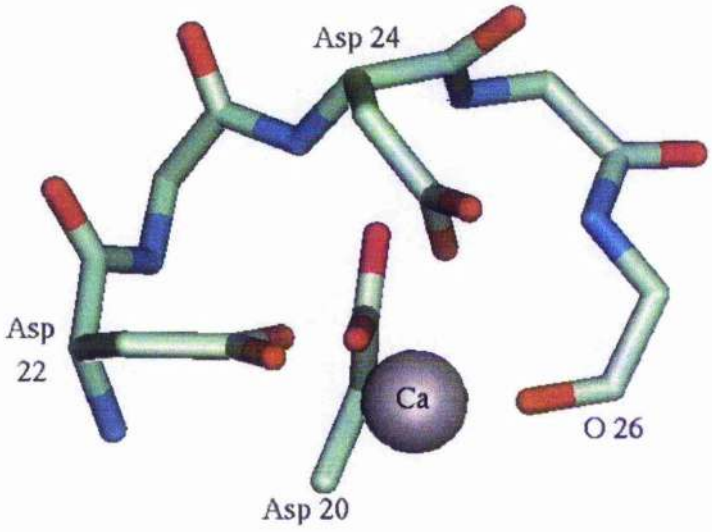


Figure 2.11b: Illustration of calcium-binding to EF-hand.



2.7 Summary

The nest is a conformation common to a number of previously identified hydrogen bonded motifs. It is found to have important structural roles such as in the Paperclip (Schellman loop) which is used as the most common alpha helical terminator at the C-terminus. The nest is also useful in functional situations such as in the serine protease where it helps to stabilise the enzyme's tetrahedral intermediate. The situations the nest occurs in are familiar and well studied but the fact that they all have a common core shape has been missed.

The nest may have some evolutionary significance due to its presence in what would be considered "old" proteins such as those involved in methanogenesis. It is a very basic unit as it could be very easily formed from a simple mixture of L- and D-amino acids. This is another reason why it may be of some evolutionary significance.

The work on the nest has led to the identification of a totally novel motif termed the Asx- or ST- nest (Watson, J. D. & Milner-White, E. J., 2001) that is discussed in detail in the next chapter. The Asx-nest is important to some calcium-binding proteins and may have implications for protein folding as will be discussed later.

Chapter 3: A Novel Nest-containing Motif: The Asx-nest and ST-nest

3.1 Introduction

During the investigation into the RL nests a novel nest-containing motif was observed. In this motif the major anionic atom in the nest concavity is the side chain oxygen of the first residue whose mainchain NH group forms the nest. In other words, the first residue that forms the nest binds to its own main chain NH group and the two that follow. In all of these motifs the important first residue whose side chain acts as the anion is always an Asx (Asp or Asn) or ST (Ser or Thr), therefore these motifs are called Asx-nests (D-nests or N-nests) and ST-nests (S-nests and T-nests) respectively. They are reasonably common motifs with 0.6 occurrences per protein in the 67-protein database (Watson, J. D. & Milner-White, E. J., 2001). There is a significantly higher proportion of Aspartate (D-nests) which suggests these motifs are good at stabilising negative charges.

These motifs are not just structural features but functional, as is shown by their discovery in EF-hand calcium-binding motifs (Chapter 2.6.4).

3.2 Geometry and Properties

The key point about these motifs is that the sidechain of the first nest residue is binding to its own mainchain NH group (figures 3.1a and b). As a result the sidechain of the first residue of the nest is acting as the main anion bound by the nest.

Figure 3.1a : The Asx-nest from 1a2y (Bhat, T. N. et. al., 1994)

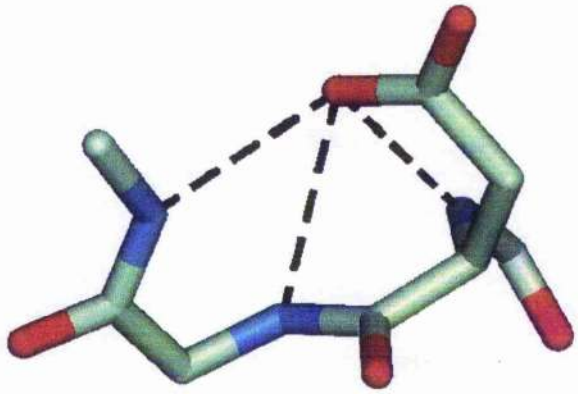
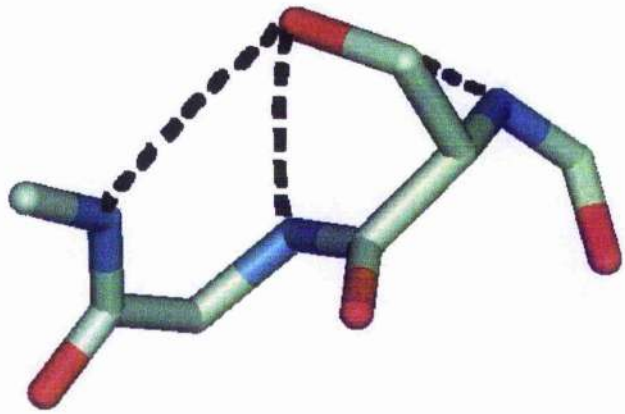


Figure 3.1b: The ST-nest



There may be some concern that the (i,i) bond is not a legitimate hydrogen bond due to distance and angular constraints. It is generally agreed that the distance for a hydrogen bond can be subject to variation depending on the circumstance such as the types of donor and acceptor atoms but the angular constraints are more important.
Even if the interaction is not a true hydrogen bond there may still be considerable electrostatic interactions between the NH and the sidechain. In order to address this, the motifs were analysed using a force field program that calculates energetic interactions using Lennard-Jones energy potentials. The program takes each CONH peptide group

and compares it with every other CONH group in the protein as well as with all the charged side chains, giving the energy of interaction and an indication of whether or not the geometry is favourable. This technique has been successful in investigating the occurrence of bifurcated hydrogen bonds (Preissener, R. et. al., 1991) where the geometry can be significantly different; it has also been successful in identifying carbonyl-carbonyl (Maccallum, P. H. et. al., 1995) interactions that can often be of greater importance than the hydrogen bond.

The main focus of this analysis is to determine whether or not the (i,i) bond is a legitimate interaction or not. The analysis showed that the interaction does not have the geometry to be considered a true hydrogen bond however, it is a significant electrostatic interaction which gives the motif increased stability. The Asx- and ST-nests should therefore be considered to be legitimate stable motifs.

3.3 Occurrence

In the 67-protein database there are 61 examples of Asx- or ST-nests as can be seen in Table 3.1 below. The breakdown of each type is as follows: 29 D-nests; 8 N-nests; 9 S-nests; and 15 T-nests. There are 2 examples where the sidechain is a Cysteine but this is rare.

It should be noted that the Asx- and ST-nests often occur in conjunction with other types of motif, especially the two types of β-bulge loop. They may therefore help provide additional stability to other types of structural hydrogen-bonded motifs.

Table 3.1: Distribution of RL nests by motif type.

	Paperclip	Paperclin Wide	Type 1 β-Turn	Type 1 β-Bulge Loop	Type 2 β-Bulge Loop	Asx-Motif	ST-Motif	D-Nest	N-Nest	S-Nest	T-Nest	C-Nest	α-Turn	Functional	Unique	Totals
Paperclip	86															
Wide Paperclip		7														
Type 1 β-Turn			19													
Type 1 β-Bulge Loops				0												
Type 2 β-Bulge Loops	5				0											
Asx-Motif			8*	17	7$	1										
ST-Motif				8#			0									
D-Nest	1		4	9	1			10								
N-Nest	1		2	2	2				1							
S-Nest				1						3						
T-Nest	1				1						3					
C-Nest												0				
α-Turn						1				1	6		10			
Functional			2		4			3		1		2		9		
Unique															18	
Numbers of nests within motif type	94	7	35	37	20	34	8	29	8	9	15	2	18	21	18	355

* = One of these also has an S-nest motif

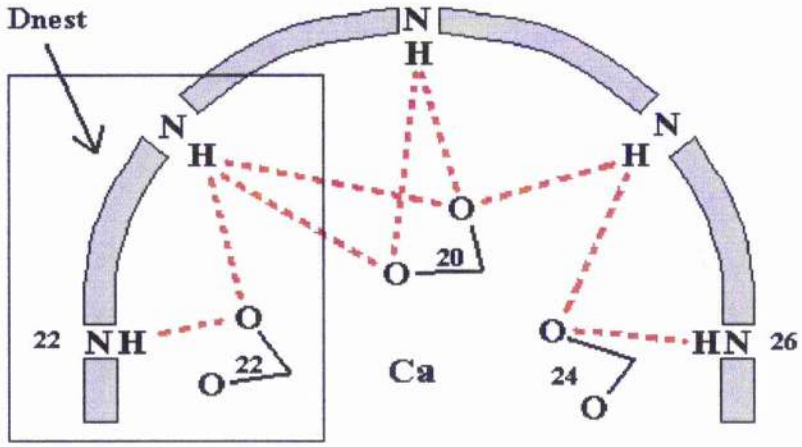
\# = One of these has an D-nest and one has an S-nest

$ = Four of these have a T-nest and one has an S-nest

3.4 Discussion and Importance

The Asx- and ST-nests are novel motifs in that they have not been described in the literature before, although some of the situations in which they occur are well known. They are shown to have significant stability and can be used in a functional capacity such as in the EF-hand motif (Meador, W. E. et. al., 1993; Milner-White, E. J., 1999; Allouche, D. et. al., 1999). In the EF-hand motif a calcium cation is bound by three aspartates (Chakrabarti, P., 1994) that come in close proximity with one another. In order to stabilise this unfavourable proximity the side chains are also bound by a compound nest which includes a D-nest (see figure 3.2).

Figure 3.2: Hbond diagram of EF-hand motif with highlighted D-nest.



In the database used there are no examples of Glutamate or Glutamine forming these type of motifs (Glx-nests). This is probably due to the increased length of the sidechain preventing the correct geometry to form, so the carboxylate cannot approach close enough to the nest concavity.

The Asx- and ST-nests may have importance in protein folding and may even act as nucleation points for folding. In a newly synthesised polypeptide in an aqueous environment it would be more favourable for main chain NH groups to be hydrogen bonded to other parts of the emerging peptide rather than to water; main chain CO groups on the other hand are easily solvated by water and this is due to the properties of water when solvating different anions and cations.

It is known that small ions (+ve and –ve) of high charge density (known as kosmotropes) bind water molecules much more strongly than those of a low charge density (known as chaotropes) and are therefore more easily solvated (Collins, K. D., 1997). When the properties of main chain NH and CO groups are examined it is seen that NH groups behave more like chaotropes and do not bind water well (are poorly solvated) whereas CO groups are more kosmotropic and are readily solvated.

Proteins are made in the cell in a generally aqueous environment therefore a linear polypeptide as it is synthesised would attempt to fold into the most stable conformation. For the main-chain carbonyl groups solvation would not pose a problem, however the main-chain NH groups are not so easily solvated. It would therefore be more preferable to use other parts of the protein to bind these main-chain NH groups. One quick and simple way to achieve this is to allow the Asx and ST residues to bind to nests forming temporary Asx-nests and ST-nests. Once the rest of the protein has been synthesised, these small stable structures would be able to be displaced in later folding stages by other hydrogen bonded motifs, such as the paperclip (Schellman loop) or by sidechain-sidechain interactions from more distant parts of the protein.

If the Asx-nest and ST-nest are used to stabilise an emerging polypeptide until other interactions can be formed then the evidence may lie in the sequences surrounding the secondary structure.

The first place to look is at the paperclip (Milner-White, E. J., 1988) or Schellman (Schellman, C., 1980) motif that is commonly used to terminate alpha helices at their C-termini. The majority of these motifs have a left-handed glycine as the C-cap residue (Schellman, C., 1980; Gunasekaran, K. et. al., 1998). The definition we use for the C-cap residue is as follows: if the residues of a helix are labelled as shown below the C-cap residue is the one with dihedral angles that do not belong to an alpha helix (non αR conformation).

N-term.....C4...C3...C2...C1...Ccap...C'...C''...C'''...C''''.....C-term

If the Asx-nest is important to paperclip (Schellman loop) formation then when the sequence is examined there should be a higher propensity for Asx residues at the C1 and C2 positions since these are close enough to the C-cap residue to interact. It should also be noted that although the Glx-nest is rarely seen in protein structures (due to the extra length of the sidechain) there is no reason to assume that it cannot form on a temporary basis to be displaced at a later date. If Glx-nests are also important to paperclip formation then they should be expected to have a higher than normal propensity at the C2 and C3 positions (since their sidechains are longer).

If the propensities for amino acid types are looked at in alpha helices it is found that Asp and Asn are found at higher levels than expected at the C-cap, C1 and C2 positions and Glu and Gln are found in higher levels at C2 and C3. This would seem to add some backing evidence to support the theory but this is by no means conclusive as there may be other reasons these residues are more common at these positions.

Although an appealing theory, the work involved to investigate this properly would include mutagenesis and protein folding studies that are beyond the scope of this thesis.

3.5 Summary

The Asx- and ST-nests are novel motifs that are found to be reasonably common. They are stable structures used to provide additional stability to other hydrogen-bonded motifs as well as having an important structural role in EF-hand proteins where they are used to bind calcium.

The geometry of the bonding may not be perfect but the electrostatic interactions are significant enough to allow the motif to spontaneously fold up in solution. This stability means the motif may have significant influences early on in protein folding pathways and could possibly act as a nucleation point for the development of more stable hydrogen-bonded motifs, however this would have to be investigated further.

Chapter 4: The Potassium Channel

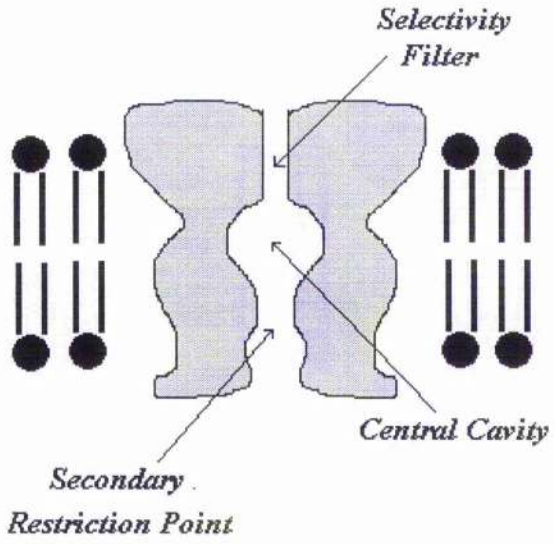
4.1 Introduction

During the investigation into the nest an unusual compound αLαR nest was noticed in the selectivity filter of the potassium channel KscA from *Streptomyces lividans*. Instead of the regular concave nest the structure showed a more extended linear backbone with the carbonyl groups arranged to point to the centre of the channel. The extended nest is found to be part of the selectivity filter: the part of the channel responsible for allowing only potassium cations to cross the channel. This is the first time a nest has been used where the carbonyl group (Carell, C. J. et. al., 1988; Chakrabarti, P., 1990, Chakrabarti, P., 1990) arrangement is the most important feature and it is also the first time an extended version of the nest has been observed in a functional situation.

4.2 Potassium Channel Structure

Potassium channels are integral membrane proteins common to both prokaryotes and eukaryotes (Minor, D. L. Jr., 2001). They are highly selective for the rapid transport of potassium across the cell membrane. Recently the high resolution crystal structures of bacterial potassium channels (Doyle, D. A. et. al., 1998; Zhou, M. et. al., 2001; Choe, S., 2002) have been determined and shed light on how this selectivity is achieved. All the channels sequenced have been found to contain a characteristic GYG sequence near the extracellular surface known as the selectivity filter (Doyle, D. A. et. al., 1998; Morais-Cabral, J. H. et. al., 2001). This is the narrowest region of the pore and is formed from a linear row of backbone carbonyl groups of each of the four subunits coming together. Just below this region is the water-filled central cavity with another restriction point at the bottom (Yellen, G., 2001).

Figure 4.1 Basic channel structure (Choe, S., 2001) illustrating the various regions and membrane.



The GYG sequence contains the most linear part of the row of carbonyls and is formed by the αLαR backbone conformation. The conformation is similar to that of the regular compound nest previously discussed, but the concavity is lost. This is due to the difference in dihedral angles between the regular γRγL concave nest and the extended αLαR conformation.

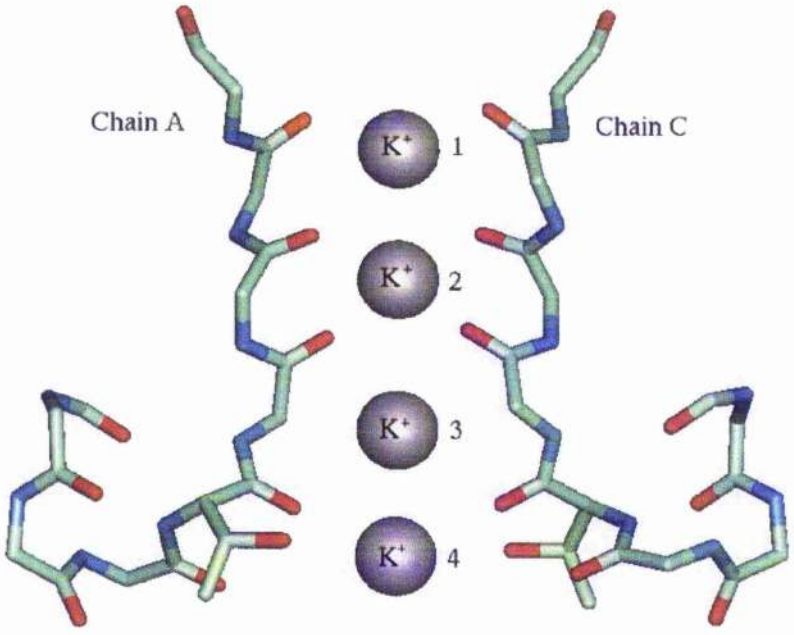
The αLαR extended nest is part of a larger RLRLRL compound nest made up of residues 74-79 as shown in figure 4.2 below. Details of the residues are shown in table 4.1.

Table 4.1 Conformation of residues at the selectivity filter

Residue No.	Residue Type	Phi (°)	Psi (°)	Conformation
74	T	-79	-13	γR
75	T	53	74	αL
76	V	-123	-26	γR
77	G	42	77	αL
78	Y	-63	-37	αR
79	G	98	-13	γL

The first part of the compound nest is at the C-terminus of a helix that points into the water-filled cavity below. This section forms a structure similar to the Schellman (Schellman, C., 1980) motif and is a common reverse turn at the C-termini of helices. The compound nest is then continued into the linear section of the selectivity filter. The result of this is that the entire top region of the water-filled cavity and the selectivity filter is lined with backbone carbonyl groups. There are also mainchain carbonyl groups that point into solution at the extracellular side of the selectivity filter (Gly79) that are important to the dehydration of solvated K^+ cations (Zhou, Y. et. al., 2001).

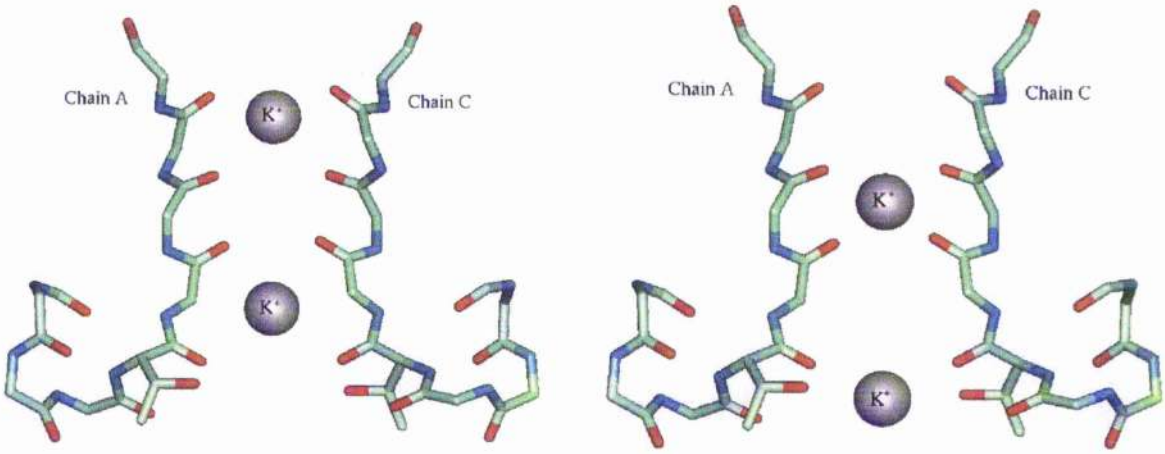
Figure 4.2 : Selectivity filter with bound potassium ions



4.3 Mechanisms Of Action

The potassium channel selectivity filter has the ability to bind up to four potassium cations (four 'sites' numbered 1-4 from extracellular to intracellular) within the channel plus one ion on the extracellular surface and one in the central cavity. The selectivity filter is found to exist in two ion bound configurations (Zhou, M. et. al., 2001; Yellen, G., 2001): one where sites 1 and 3 are bound; the other with sites 2 and 4 occupied (see figure 4.3 below).

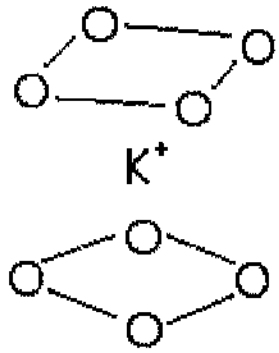
Figure 4.3 : Two possible bound 'states'



Conduction of ions across the channel is therefore achieved by alternating between the two bound forms. This can be achieved by either simply moving the ions back and forth, or as one ion enters at one side another is expelled at the other side. Regardless of how the ions are moved the mechanism will work at its optimum rate when the two stable intermediates (1,3) and (2,4) are at equal energies. It is found that, when only potassium is available, all four sites are occupied i.e. both occupied states are of almost identical energies. If other permeable ions are used such as rubidium, one state is usually preferred over another therefore the channel conducts more poorly (Zhou, M. et. al., 2001; Morais-Cabral, J. H. et. al., 2001; Zhou, Y. et. al., 2001). The same situation occurs when there is a mixture of potassium and rubidium, the rubidium prefers one type of occupied state and prevents potassium transport from working as effectively.

The water-filled cavity is important for stabilising the potassium ion mid transfer (Doyle, D. A. et. al., 1998; Yellen, G., 2001; Zhou, Y. et. al., 2001) and addresses how the channel manages to achieve a very high effective K^+ concentration at the centre of the membrane. The K^+ ion is at its most stable when fully hydrated therefore being in the centre of a membrane is energetically unfavourable. The cavity holds the potassium ion in the centre surrounding it with ordered water molecules. This forms a structure with the K^+ ion at the centre of a square antiprism (see figure 4.4) that is mimicked in the selectivity filter by mainchain carbonyls (Doyle, D. A. et. al., 1998; Carrell, C. J. et. al., 1988) at each of the bound 'sites'.

Figure 4.4: Illustration of the square antiprism



The ordering of water molecules does not impede the movement of potassium through the pore and allows the smooth transfer of the ions in and out of the selectivity filter. The dehydration and rehydration of the ion is achieved by firstly replacing half the hydration shell with four mainchain carbonyl groups then replacing the remainder of the hydration shell with another four carbonyls (Morais-Cabral, J. H. et. al., 2001; Zhou, Y. et. al., 2001). There are tiny structural adjustments between the (1,3) and (2,4) conformations but the polypeptide conformations are largely unchanged.

Crystallography studies show that the selectivity filter exists in two states dependant on the concentration of potassium:

1. High $[K^+]$

This structure shows four K^+ ions each surrounded by 8 oxygens from the mainchain carbonyl groups. The selectivity filter is linear and this would be considered to be the conductive state of the channel.

2. Low [K^+]

When there is a low concentration of potassium the structure becomes significantly altered. There are no ions at all to be found at positions 2 & 3 in the filter and Val76 and Gly77 have very different conformations. The selectivity filter backbone rotates and moves the carbonyl groups away from the ions in the centre disrupting the ion-binding sites. This gives the filter an hourglass appearance and essentially pinches it shut. This conformation is the non-conductive state (Choe, S., 2001).

There are large conformational differences between conducting and non-conducting states. The amount of changes involved means the time taken for interconversion between the two states is on the gating timescale (milliseconds) rather than the conduction timescale (nanoseconds) (Yellen, G., 2001; Morais-Cabral, J. H. et. al., 2001; Choe, S., 2002). This helps to explain how the selectivity filter manages to maintain its stability in low K^+ concentrations.

4.4 Summary

The potassium channel is the first example where the function of the nest is dictated by the arrangement of its mainchain carbonyl groups (Carell, C. J. et. al., 1988; Chakrabarti, P., 1990, Chakrabarti, P., 1990) rather than the mainchain NH groups. This is due to its more unusual extended conformation formed by alternating $\alpha L \alpha R$, rather than $\gamma L \gamma R$, dihedral angles. The arrangement of the carbonyl groups of four subunits collectively results in a channel specifically designed for cations of a particular size, namely potassium and rubidium. Other cations are excluded by being either too large to enter the pore or, in the case of sodium, too small to interact with the carbonyl groups lining the channel. An interesting aspect to the formation of the selectivity filter is that it is unstable without cations present in it: the K^+ ion is the subject of transport but is also essential to the formation of the channel in the first place (Choe, S., 2001). Other regions of the channel are designed to stabilise the transport of the potassium ion and help overcome the unfavourable energetics of moving it from a favourable solvated state through a hydrophobic membrane and back into solution again.

Work into anion transport has involved investigation of the ClC chloride channel (Minor, D. L. Jr., 2001; Dutzler, R. et. al., 2002; Choe, S., 2002). The structure and mechanism of the Cl^- channel is fundamentally different from that of the cationic potassium channel. The Cl^- channel is a dimer and, unlike the tetrameric potassium channel, the pore is not made up from an interface between two subunits; each subunit has its own pore and associated selectivity filter. The selectivity filter in the Cl^- channel does not involve carbonyls as these would repel chloride ions. The Cl^- ion is bound by partial charges made up from the mainchain amide groups of the N-terminus of an alpha helix (Hol, G. W. J. et. al., 1978) and serine and tyrosine sidechains. The only similarity between the chloride and potassium channels is that both use partial charges to hold their ions (mainchain amide and carbonyl groups respectively). Presumably this negates the need for fully charged groups that may bind the ions too tightly and prevent their efficient transport.

Recent work on Aquaporins (AQP1) (Sui, H. et. al., 2001), the water channels used by cells for the rapid transport of water in response to osmotic gradients, has shown the involvement of selectivity filters in their structures. Although the structures greatly differ they also seem to have carbonyl groups lining the constriction point and may be another example of the extended nest as seen in the potassium channel. This needs to be confirmed by examination of the crystal structure when it becomes available. The main difference in the aquaporins lies in the greater use of sidechains and hydrophobic regions to restrict movement of water. The reasons for the differences between types are probably down to the requirement for the water channel to be optimised for water transport while some other aquaporins are selective for the transport of other solutes with the water.

Chapter 5: Molecular Modelling

5.1 Introduction

The nest is characterised by alternating enantiomeric dihedral angles and is shown to have anion-binding properties that are functional as well as structural. The identification of the extended nest conformation and its cation-binding function led to a consideration that polypeptides with different alternating enantiomeric dihedral angles may give rise to other types of motif. I investigated this further by creating artifical model polypeptides on computer with different sets of dihedral angles to indicate the range of possible structures that can be formed.

5.2 Creating Model Polyglycines

The first step in the modelling process was to create a polyglycine short polypeptide. A regular PDB coordinate file from a protein was selected for a region with 10 amino acids in a beta strand. Using a simple text editor (jot, emacs, word, etc) the lines corresponding to the 10 amino acid section were isolated. All amino acid identities were changed to GLY and lines containing coordinates for sidechains deleted. This gave a file containing a 10-residue polyglycine strand.

Fig 5.1 Examples of lines during the editing process

a) Original File

```
HEADER    ONCOGENE PROTEIN                          30-APR-90   5P21      5P21   2
COMPND    $C-*H-RAS $P21 PROTEIN (AMINO ACIDS 1 - 166) COMPLEX WITH      5P21   3
COMPND   2 GUANOSINE-5'-(BETA,GAMMA-IMIDO) TRIPHOSPHATE (GPP*NP)        5P21   4
SOURCE    HUMAN (HOMO $SAPIENS) CELLULAR HARVEY-RAS GENE (TRUNCATED)    5P21   5
SOURCE   2 EXPRESSED IN (ESCHERICHIA $COLI)                             5P21   6
AUTHOR    E.F.PAI,A.WITTINGHOFER,W.KABSCH                               5P21   7
REVDAT   1   15-JAN-92 5P21    0                                        5P21   8

ATOM    112  N   SER    17       6.913  32.942  16.135  1.00  7.74      5P21 231
ATOM    113  CA  SER    17       6.661  34.295  15.649  1.00  8.78      5P21 232
ATOM    114  C   SER    17       7.854  34.859  14.919  1.00  9.73      5P21 233
ATOM    115  O   SER    17       7.653  35.446  13.868  1.00  9.70      5P21 234
```

```
ATOM    116  CB  SER    17       6.360  35.264  16.769  1.00  8.22      5P21 235
ATOM    117  OG  SER    17       5.083  34.853  17.204  1.00  8.79      5P21 236
ATOM    118  N   ALA    18       9.072  34.687  15.444  1.00  8.95      5P21 237
ATOM    119  CA  ALA    18      10.253  35.199  14.769  1.00  8.95      5P21 238
ATOM    120  C   ALA    18      10.438  34.576  13.374  1.00  8.54      5P21 239
ATOM    121  O   ALA    18      10.832  35.254  12.430  1.00  8.75      5P21 240
ATOM    122  CB  ALA    18      11.521  34.931  15.640  1.00 10.11      5P21 241
ATOM    123  N   LEU    19      10.136  33.284  13.176  1.00  8.31      5P21 242
ATOM    124  CA  LEU    19      10.225  32.667  11.837  1.00  9.30      5P21 243
ATOM    125  C   LEU    19       9.204  33.281  10.884  1.00 10.98      5P21 244
ATOM    126  O   LEU    19       9.534  33.652   9.751  1.00 10.47      5P21 245
ATOM    127  CB  LEU    19       9.938  31.165  11.885  1.00  9.41      5P21 246
ATOM    128  CG  LEU    19      11.035  30.275  12.487  1.00 11.80      5P21 247
ATOM    129  CD1 LEU    19      10.491  28.866  12.723  1.00 10.43      5P21 248
ATOM    130  CD2 LEU    19      12.256  30.288  11.554  1.00 13.42      5P21 249
ATOM    131  N   THR    20       7.950  33.419  11.335  1.00 10.72      5P21 250
ATOM    132  CA  THR    20       6.888  34.015  10.529  1.00 11.87      5P21 251
ATOM    133  C   THR    20       7.173  35.492  10.161  1.00 12.65      5P21 252
ATOM    134  O   THR    20       7.012  35.917   9.009  1.00 12.60      5P21 253
ATOM    135  CB  THR    20       5.547  33.888  11.312  1.00 13.43      5P21 254
ATOM    136  OG1 THR    20       5.299  32.509  11.544  1.00 12.67      5P21 255
ATOM    137  CG2 THR    20       4.346  34.410  10.518  1.00 14.64      5P21 256
```

b) Removal Of Sidechain Atoms

```
ATOM    112  N   SER    17       6.913  32.942  16.135  1.00  7.74      5P21 231
ATOM    113  CA  SER    17       6.661  34.295  15.649  1.00  8.78      5P21 232
ATOM    114  C   SER    17       7.854  34.859  14.919  1.00  9.73      5P21 233
ATOM    115  O   SER    17       7.653  35.446  13.868  1.00  9.70      5P21 234
ATOM    118  N   ALA    18       9.072  34.687  15.444  1.00  8.95      5P21 237
ATOM    119  CA  ALA    18      10.253  35.199  14.769  1.00  8.95      5P21 238
ATOM    120  C   ALA    18      10.438  34.576  13.374  1.00  8.54      5P21 239
ATOM    121  O   ALA    18      10.832  35.254  12.430  1.00  8.75      5P21 240
ATOM    123  N   LEU    19      10.136  33.284  13.176  1.00  8.31      5P21 242
ATOM    124  CA  LEU    19      10.225  32.667  11.837  1.00  9.30      5P21 243
ATOM    125  C   LEU    19       9.204  33.281  10.884  1.00 10.98      5P21 244
ATOM    126  O   LEU    19       9.534  33.652   9.751  1.00 10.47      5P21 245
ATOM    131  N   THR    20       7.950  33.419  11.335  1.00 10.72      5P21 250
ATOM    132  CA  THR    20       6.888  34.015  10.529  1.00 11.87      5P21 251
ATOM    133  C   THR    20       7.173  35.492  10.161  1.00 12.65      5P21 252
ATOM    134  O   THR    20       7.012  35.917   9.009  1.00 12.60      5P21 253
```

c) Changing All Residues To Glycine (Final Model File)

```
ATOM    113  CA  GLY    17       6.661  34.295  15.649  1.00  8.78      5P21 232
ATOM    114  C   GLY    17       7.854  34.859  14.919  1.00  9.73      5P21 233
ATOM    115  O   GLY    17       7.653  35.446  13.868  1.00  9.70      5P21 234
ATOM    118  N   GLY    18       9.072  34.687  15.444  1.00  8.95      5P21 237
ATOM    119  CA  GLY    18      10.253  35.199  14.769  1.00  8.95      5P21 238
ATOM    120  C   GLY    18      10.438  34.576  13.374  1.00  8.54      5P21 239
```

```
ATOM    121  O   GLY    18      10.832  35.254  12.430  1.00  8.75      5P21 240
ATOM    123  N   GLY    19      10.136  33.284  13.176  1.00  8.31      5P21 242
ATOM    124  CA  GLY    19      10.225  32.567  11.837  1.00  9.30      5P21 243
ATOM    125  C   GLY    19       9.204  33.281  10.884  1.00 10.98      5P21 244
ATOM    126  O   GLY    19       9.534  33.652   9.751  1.00 10.47      5P21 245
ATOM    131  N   GLY    20       7.950  33.419  11.335  1.00 10.72      5P21 250
ATOM    132  CA  GLY    20       6.888  34.015  10.529  1.00 11.87      5P21 251
ATOM    133  C   GLY    20       7.173  35.492  10.161  1.00 12.65      5P21 252
ATOM    134  O   GLY    20       7.012  35.917   9.009  1.00 12.60      5P21 253
```

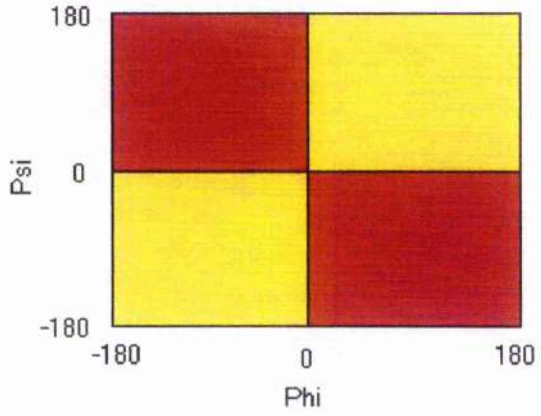
To perform a manipulation on this model, its file was imported into Quanta (QUANTA, 1997). Within Quanta are various modelling options that allow the user to alter features of an imported file. The user can change factors such as bond angles, bond lengths, hydrogen bond parameters, add or remove bonds with other sections of a protein, set dihedral angles, etc.

For the model peptide all peptide bonds were set to trans bonds with ω angles of 180°. The main point of all the modelling is to investigate alternating enantiomeric dihedral angles, therefore a starting point is needed. I decided to start at (0°,0°) and go from there. The phi and psi dihedral angles were all altered to 0° and the resulting peptide was saved to a file named [0.0].

For the next model phi was kept at 0° but psi was changed to 10° for the first amino acid, -10° for the second, 10° for the third and so on. This was saved as a new file called [0.10]. Psi was changed by 10° at a time saving the new model files until reaching 180°. Returning to the [0.0] file, phi was then altered by 10° so that the first residue was -10°, the second 10°,and so on to form file [10.0]. The phi angles for this [10.0] model were adjusted as with the [0.0] model.

It was soon discovered that this process was only solving half the problem. This is explained by figure 5.2 below. In each of the files created the polyglycine model has dihedral angles (-x,-y) (x,y) (-x,-y) ... etc. and only covers the two quadrants of the Ramachandran plot as shown in fig 5.2 as the yellow region.
In order to cover the rest of the plot the modelling process had to be repeated except that each model created had dihedral angles of (-x,y) (x,-y) (-x,y)... etc. This is indicated in figure 5.2 as the red region.

Figure 5.2: Ramachandran plot indicating regions covered by models.

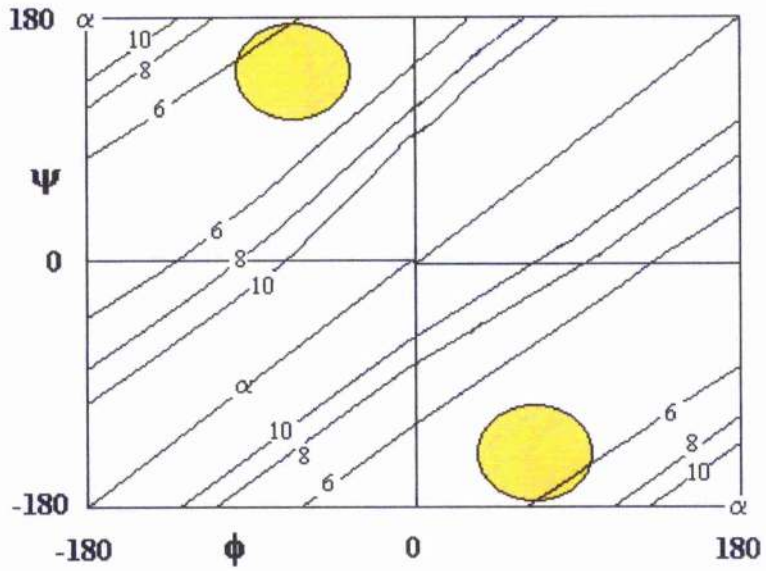


5.3 Ramachandran-like Plots

The modelling process resulted in 703 different model files and for each several measurements were taken. The resultant mass of data is difficult to analyse in its raw form therefore it was determined the best way to display and interpret the data was to plot it.

The plots produced are similar to the Ramachandran plot except that in the original work model polypeptides with identical conformations for successive residues were considered, whereas this modelling is for polypeptides with successive alternating enantiomeric dihedral angles. Another key difference is that in the current work all the polypeptides are, in principle, ring shaped rather than the mainly helical models designed by Ramachandran *et al.* (Ramachandran, G. N. & Sasisekharan, V., 1968; Dickerson & Geis, 1969) It should be noted that only half of each plot needs to be shown due to the symmetry across the plot (see figure 5.3).

Figure 5.3: Full Ramachandran plot indicating symmetry between halves.

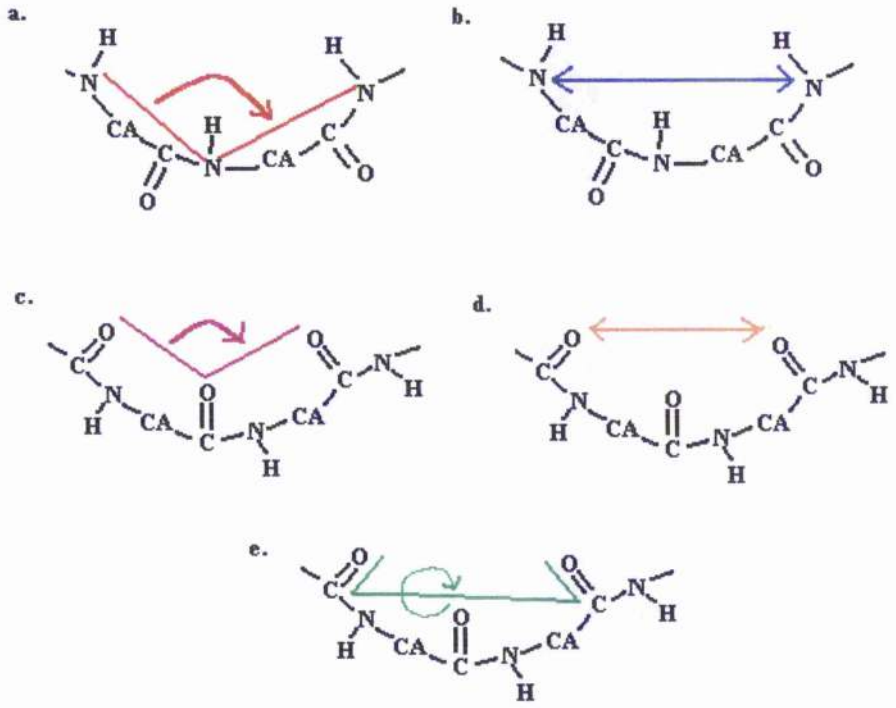


The yellow circled region corresponds to the dihedral angles which give rise to the conformation we have termed the catgrip.

The measurements taken for each model polypeptide are as follows (for diagrams illustrating how the measurements were taken see figure 5.4):

1. The number of residues for a complete ring.
2. The distance between alternate NH groups (N-N distance) – this gives an estimate of the 'width' of any nest in a model.
3. The angle between three successive NH groups (NNN angle) – this gives a measure of how extended or concave any model structure is.
4. The distance between alternate carbonyl oxygens (O-O distance) – this gives an estimate of how close carbonyls come with each other.
5. The angle between successive carbonyl oxygens (OOO angle) – this gives another estimate of how extended the polypeptide model is.
6. A dihedral angle measured between successive carbonyl groups (OCCO angle) – this indicates whether the carbonyls point towards the centre of the ring or tangential to the ring.

Figure 5.4: Diagrams illustrating how the different measurements are taken



Legend:

a. = NNN angle

b. = N-N distance

c. = OOO angle

d = O-O distance

e. = OCCO dihedral angle

Other plots were made including information such as regions highlighting the dihedral angles for nests, catgrips and K channel, whether the NH groups or CO groups point to the ring centre or away. All of the plots are shown in figure 5.5 a-h.

Fig 5.5 (a-d): Ramachandran-like plots

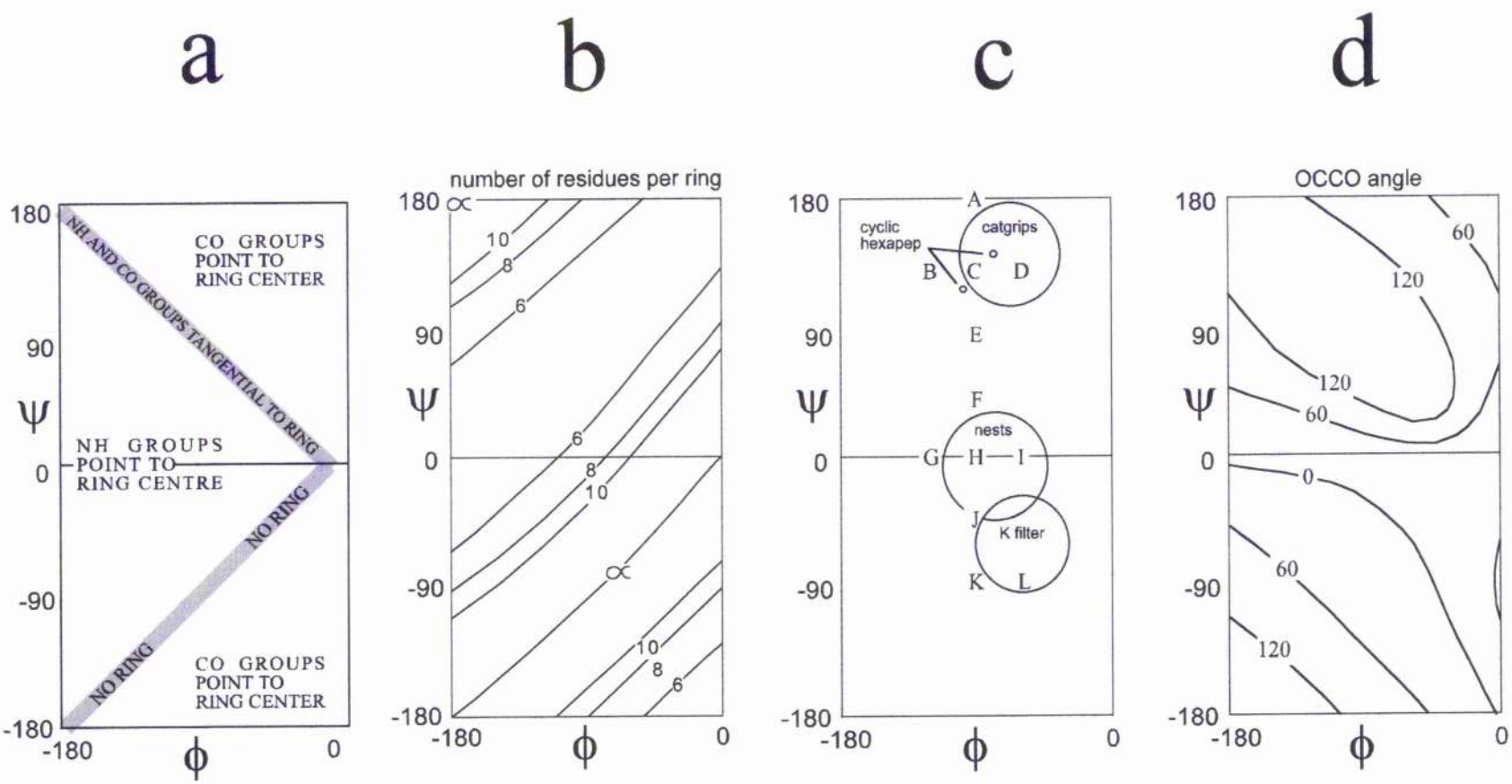
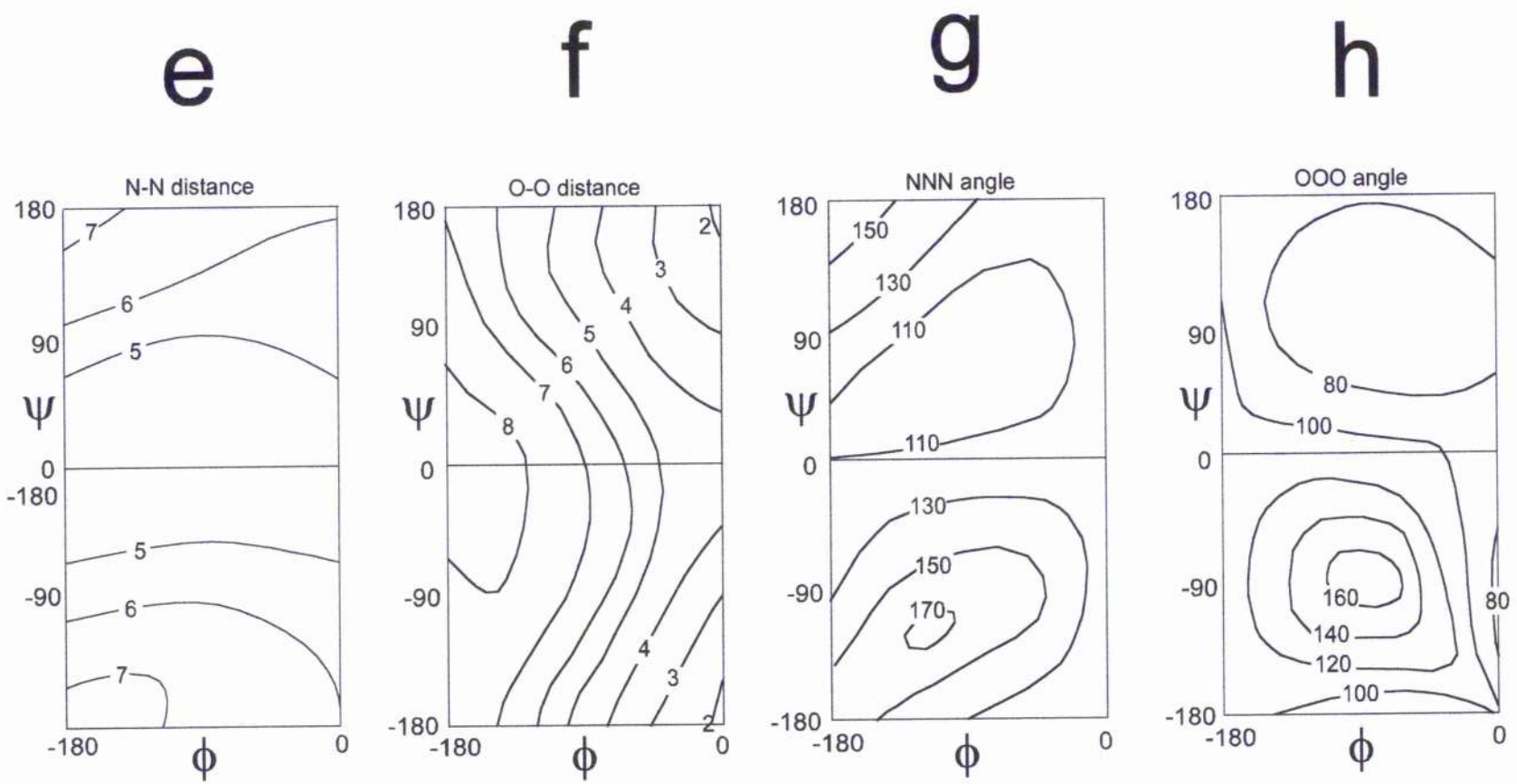


Fig 5.5 (e-h): Ramachandran-like plots



5.4 Results And Discussion

The theoretical minimum ring size is 5.08 residues in length. The reason the minimum ring size is not 5 residues is due to the difference in geometry between the pentagon and the tetrahedron (the angle between sides of a pentagon is 108° whereas the tetrahedral angle is 109°)

If these types of structures are to be found in real situations outwith that of the computer model two points become apparent:

.

1. Ring conformations with non-integral or odd numbers can exist only as partial rings (therefore the smallest sterically feasible cyclic polypeptides of this type are hexapeptides and octapeptides).
2. If pairs of alternate atoms are to bind to another atom then they should be about 4-5 angstroms apart. Therefore it would be here that structures binding other substrates would be found.

Figure 5.6 shows pictures of some of the polyglycine models and figure 5.7 shows where they lie on the Ramachandran-like plots in relation to the locations of different motifs.

Figure 5.6 Pictures of various models

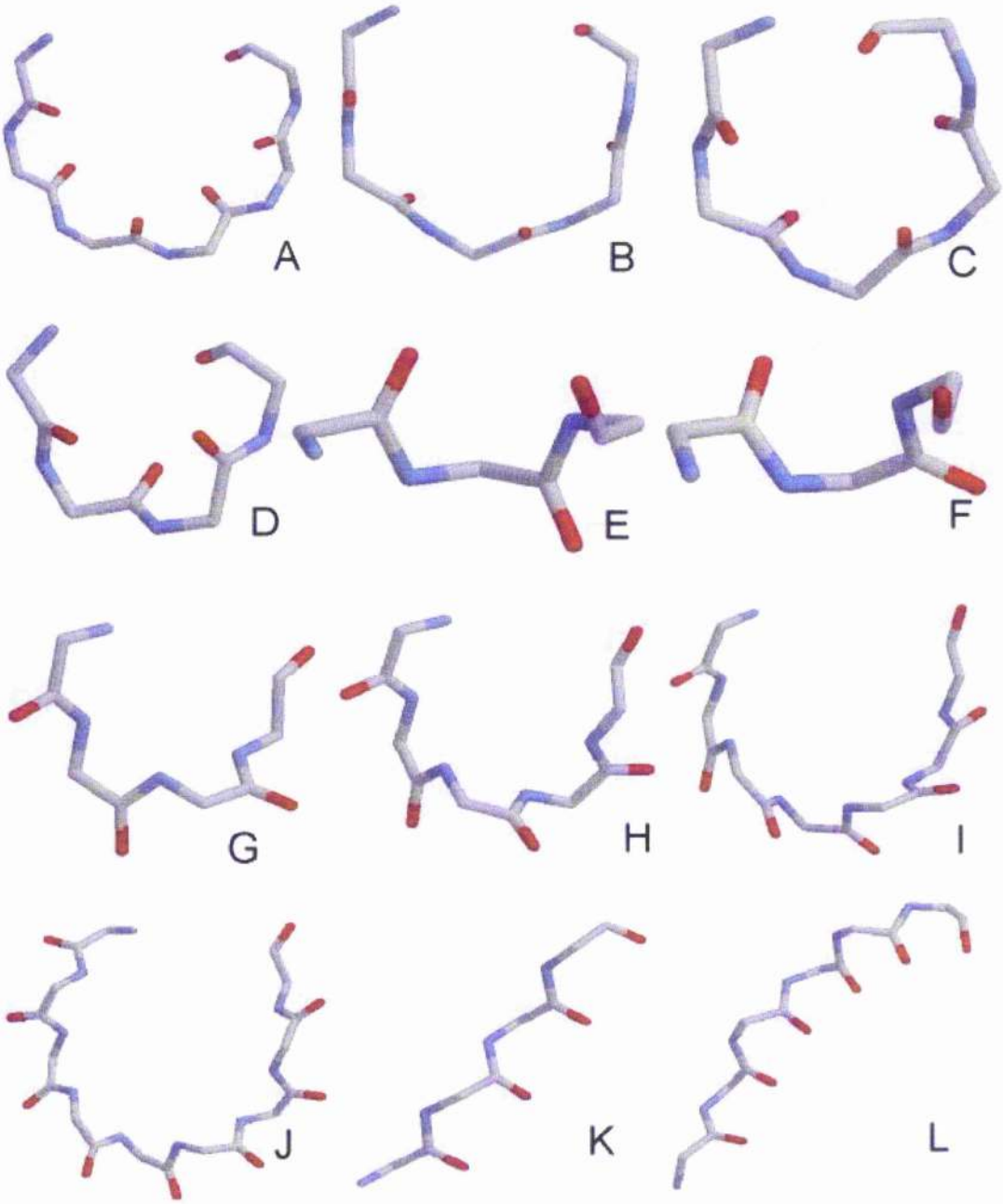
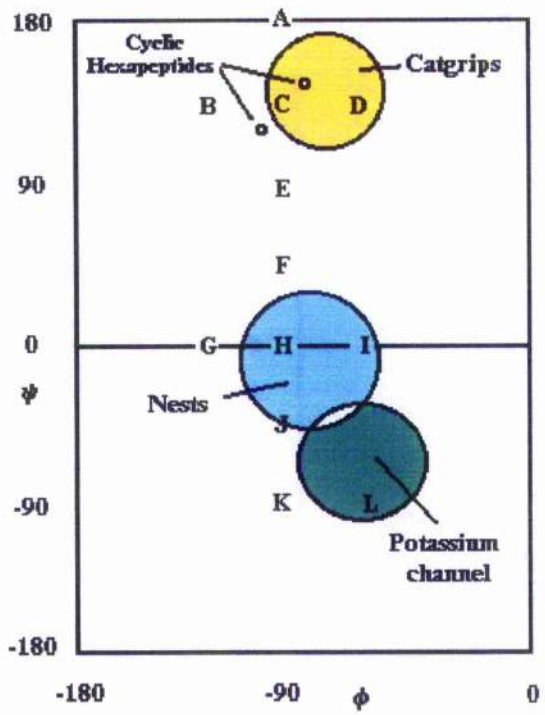


Figure 5.7: Ramachandran-like plot where particular conformations are highlighted (this also shows the locations of the example polypeptide models from figure 5.6).



In Figures 5.7 and 5.6c-d there is a conformation which could give rise to cation-binding by the backbone carbonyl groups. It was decided to investigate this structure further by looking for examples in the PDB where a cation is bound by mainchain carbonyl groups. Several examples were found where calcium ligation is mediated by the mainchain and led to the identification of a novel motif termed the "catgrip" (cation grip) (Watson, J. D. & Milner-White, E. J., 2001) as discussed in the next chapter.

Recent work on cyclic polypeptides (Pavone, V. et. al., 1989) involved synthesis of hexa- or octa-peptides with alternate L and D amino acids. These form rings capable of assembling into tubes through antiparallel hydrogen bonding (Ghadhiri, M. R. et. al., 1993). They are found to have particular dihedral angles that are highlighted in figure 5.7.

5.5 Summary

Through the creation of the model polyglycines it has been possible to investigate the range of possible structures where the backbone dihedral angles are alternating and enantiomeric. The structures formed are ring-shaped or extended and data collected on them has been used to create Ramachandran-like plots for easy analysis. The investigation of the shapes generated by the modelling process has shown several conformations that occur regularly in protein structures. Three key types of structure are found:

1. The nest – this has alternating enantiomeric dihedral angles in the γR and γL regions of the Ramachandran plot and gives rise to a concavity with anion-binding potential. This is because the conformation arranges three successive backbone NH groups to come together in a small pocket, since NH groups have a partial positive charge they are able collectively to bind to anions. Multiple nests can come together to form larger structures with the abiltiy to bind larger anions such as the P-loop of p21 Ras.
2. The extended nest – this takes dihedral angles in the αR and αL region of the Ramachandran plot and give rise to a more flattened extended conformation. This conformation has been used by the potassium channel to form the selectivity filter (it lines the channel with mainchain carbonyl groups in order to facilitate selective K+ transport).
3. The catgrip – here successive amino acids have dihedral angles in the βR and epsilon regions of the Ramachandran plot. This conformation aligns its carbonyl groups so as to bind to its cation e.g. Calcium. Like the nest it is possible to get multiple catgrips to come together to form compound or tandem ring-shaped structures. These rings are put to use to bind Calcium in a number of matrix metalloproteases as well as other proteins.

Recent work on cyclic polypeptides has involved the synthesis of alternating cyclic L,D-hexapeptides (Pavone, V. et. al., 1989) which have been shown by crystallography to form ring-shaped structures similar to some of the model polyglycines.

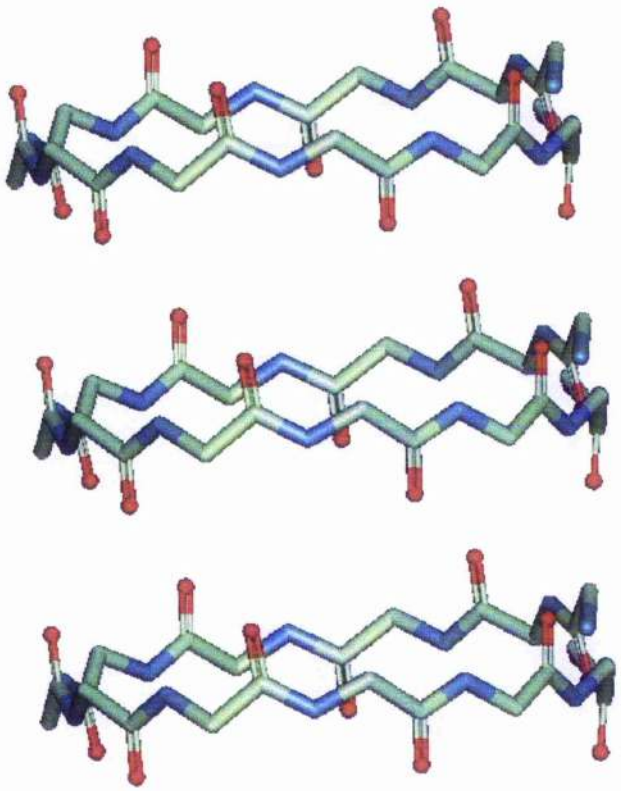
There are two main synthesised cyclopeptides:

1. cyclo(L-phe, D-phe)3 : $\phi_1\psi_1$= (89°, -152°) $\phi_2\psi_2$= (-89°, 152°)
2. cyclo(L-val, D-val)3 : $\phi_1\psi_1$= (98°, -130°) $\phi_2\psi_2$= (-98°, 130°)

These have a conformation similar, but not identical, to that of the βR βL model polyglycine that has been previously identified as the catgrip.

Other work has been conducted with synthesised cyclic octapeptides with alternating L and D amino acids. These have been shown to self-assemble by antiparallel hydrogen bonding between adjacent octapeptides to form nanotubules (Ghadhiri, M. R. et. al., 1993). This would be expected to happen in more extended rings where the CO and NH groups are arranged tangential to the plane of the ring (see figure 5.8)

Figure 5.8 Possible nanotubule formation



Chapter 6: The Catgrip

6.1 Introduction

The identification of the extended nest in the potassium channel showed that even a small change in the mainchain dihedral angles can change the conformation of a polypeptide chain. This also indicated that there are functional motifs in existence where the mainchain carbonyl groups are the most important atoms. Since carbonyl groups have a partial negative character the motifs should make good cation-binding motifs. Subsequent molecular modelling on computer showed there are certain combinations of dihedral angles where the model polypeptide forms structures capable of forming in real proteins. It was therefore decided to investigate the cation-binding in proteins to determine whether or not these new motifs that use alternating enantiomeric conformations are common. This resulted in the identification of another novel motif, which we have named the "catgrip" (Cation-grip) (Watson, J. D. & Milner-White, E. J., 2001).

6.2 Geometry

Catgrips are ring-shaped cation-binding motifs. They are formed from residues with alternating mainchain dihedral angles and, like the nest, can occur in two enantiomeric forms RL and LR. Their particular conformation arranges the mainchain so that the backbone CO groups point to a common region. Since the carbonyl groups have a partial negative charge the catgrip motif makes an excellent cation-binding motif (Carrell, C. J. et. al., 1988; Chakrabarti, P., 1990; Chakrabarti, P., 1990; Holm, R. H. et al, 1996; Harding, M. M., 1999; Harding, M. M., 2001).

Multiple catgrips can form larger tandem or compound catgrips in the same way nests can but there is a much greater difference in conformation between compound and tandem forms of the catgrip. This can be seen in the diagrams below.

Another major difference between catgrips and nests involves how they bind to their ligand. Where a nest results from the enantiomeric dihedral angles of residues i and

i+1, the nest concavity is formed from the NH groups of i, i+1, and i+2. In the equivalent catgrip structure it is the main chain CO groups of residues i-1 and i+1 that bind the cation.

There are three main types of catgrip motif found and are distinguished by the numbers of calciums bound as well as the number of carbonyls used (Watson, J. D. & Milner-White, E. J., 2001):

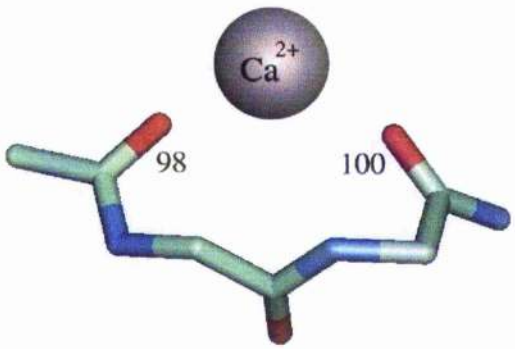
1. Two alternating main chain carbonyl oxygens bound to one Ca^{2+}
   This is the simplest type of motif and can be LR or RL (see figure 6.1a)

2. Three alternating main chain carbonyl oxygens bound to one Ca^{2+}
   This can occur in either a tandem or a compound form which greatly differ in structure (see figures 6.1b for tandem and 6.1c for compound)

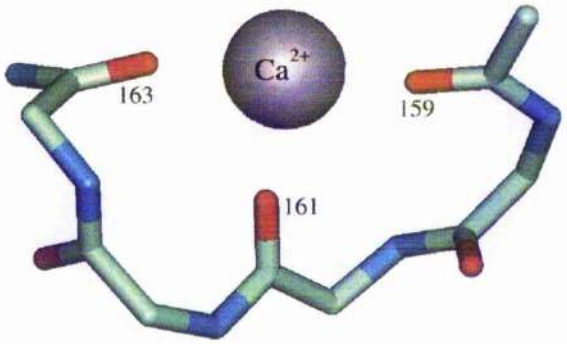
3. Four successive main chain carbonyl oxygens bound to two Ca^{2+}
   These are the most complicated type of motif and also the most interesting. They are observed in the serralysin proteins (a matrix metalloprotease [MMP] (Karlin, K. D., 1993; Bode, W. et. al., 1993; Baumann, U., 1994; Browner, M. F. et. al., 1995; Borkakoti, N., 1998; Fernandez-Catalan, C. et. al., 1998; Spurlino, J. C. et. al., 1994) found in Gram-negative bacteria) at a domain described as a β-roll (Baumann, U. et. al., 1993). This region consists of β-helices with six or seven calcium ions at the loops between strands. There are sets of loops with calciums between them arranged to form a row. Three of the loops examined exhibit the regular βLβRβL conformation that allows the carbonyl groups to contact the calcium ions above and below each loop (see figure 6.1d)

Figure 6.1: Various types of catgrip

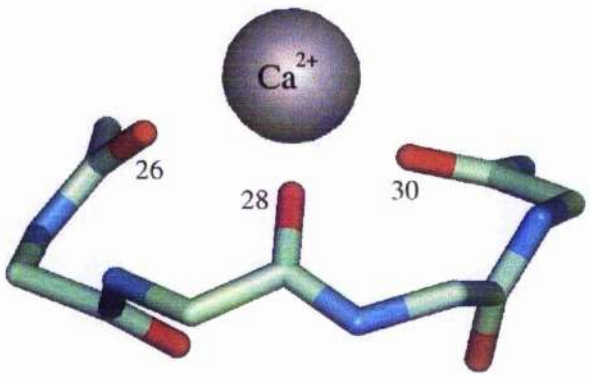
A. 2ran (Huber, R. et. al.,1992)
[2 CO groups 1 calcium]



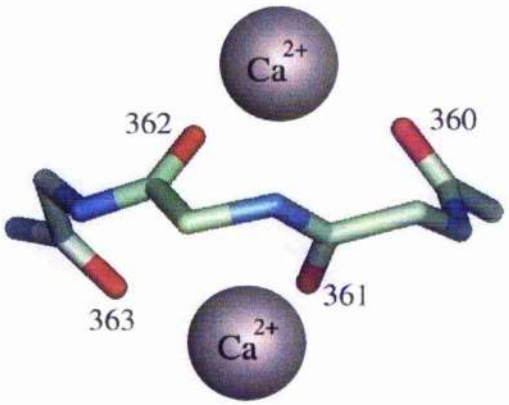
B. 2usn (Becker, J. W. et. al, 1995)
[3 CO groups 1 calcium]



C. 2ran2 (Huber, R. et. al.,1992)
[3 CO groups 1 calcium]



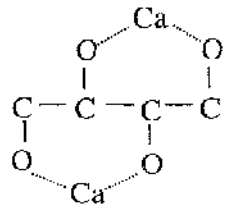
D. 1kap (Baumann, U. et. al., 1993)
[4 CO groups 2 calcium]



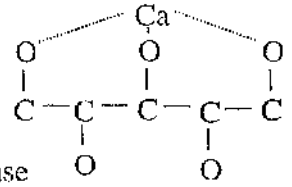
6.3 Methods And Results

To investigate cation-binding the PDB was searched for all proteins that ligate metal ions using their mainchain carbonyl groups (Laskowski, R. A. et. al., 1997; Degtyarenko, K. N. et. al., 1999). This set was then searched for those proteins where the backbone carbonyl groups used are from alternating amino acid residues. The dihedral angles for these proteins were then looked at for alternating enantiomeric angles (similar to the nest situation). The majority of 'hits' are members of the matrix metalloproteases (Karlin, K. D., 1993; Bode, W. et. al., 1993; Baumann, U., 1994; Browner, M. F. et. al., 1995; Borkakoti, N., 1998; Fernandez-Catalan, C. et. al., 1998; Spurlino, J. C. et. al., 1994) but other examples exist such as annexin (Huber, R. et. al., 1992; Koncha, N. O. et. al., 1993), phospholipase A2 (Dijkstra, B. W. et. al., 1981; Scott, D. L. et. al., 1990), fibrinogen (Spraggon, G. et. al., 1997), thermitase (Gros, P. et. al., 1989) and subtilisin (Schnittke, J. L. et. al., 1996; Dauter, Z. et. al., 1991). The results of this work are shown in table 6.1 and are split up into each of the three types of motif found.

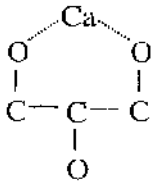
Table 6.1: Observed Catgrips



Catgrip type	Resno	PDBcode	protein
Four successive mainchain carbonyl oxygens bound to 2 Calciums			
°L°R°L°	352,361,370	1af0	MMP: Serralysin
°L°R°L°	352,361,370	1kap	MMP: alkaline protease



Three alternating mainchain carbonyl oxygens bound to 1 calcium			
°R L°R L°	27	2ran	annexin
°R L°R L°	29	1bp2	phospholipase A2
°R L°L R°	160	2usn	MMP: stromelysin
°R L°L R°	177	1mmp	MMP: matrilysin
°R L°L R°	177	1nfc	MMP: fibroblast collagenase
°R L°L R°	156	1bzs	MMP: neutrophil collagenase
°R L°L R°	195	1buv	MMP: membrane-type collagenase

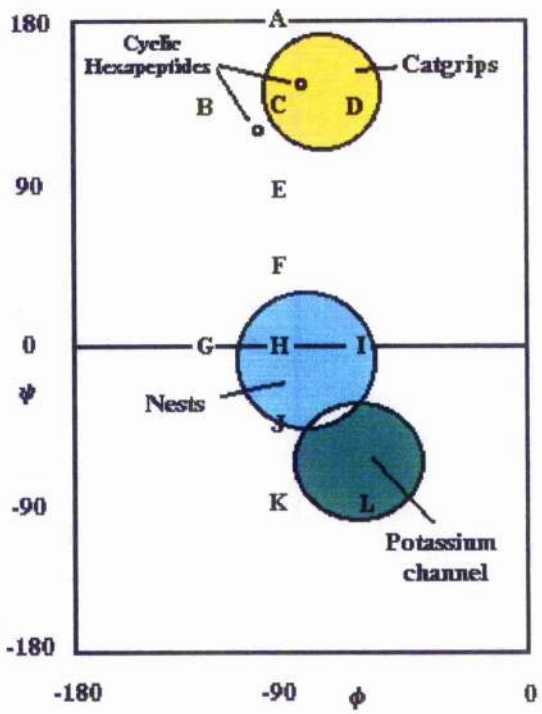


Two alternating mainchain carbonyl oxygens bound to 1 calcium			
°R L°	99,185,259	2ran	annexin
°R L°	323	1fza	fibrinogen
°R L°	160	1qib	MMP: gelatinase
°L R°	87	2tec	thermitase
°L R°	79	1af4	subtilisin Carlsberg
°L R°	80	1mpt	M-protease
°L R°	80	1scj	subtilisin E
°L R°	80	1mee	mesentericopeptidase

(Note that the location of the bound carbonyl oxygens are indicated by °)

These motifs have alternating enantiomeric dihedral angles and it is of interest to know where on the Ramachandran plot they are found. The dihedral angles of each type of motif are plotted on the Ramachandran plot as shown in figure 6.2 below. The results show that the different motifs cluster at different regions although all are found to have βR or βL conformations. The K channel has been included to show that, although it is using carbonyl groups to bind cations, it is a different, more extended, form of the nest and different in conformation from the catgrip.

Figure 6.2: Ramachandran plot showing locations of catgrips.



6.4 Discussion

The catgrip is a novel cation-binding motif used by some proteins to ligate metal ions. So far we have only found examples where the cation bound is calcium and a search of magnesium-binding proteins failed to detect any, but this does not mean other proteins may use the motif to ligate other metals. All of the catgrips found so far fall into one of three groups as determined by the number of calciums bound and the number of carbonyls used by the motif.

Catgrips are more specific motifs than nests which may explain their rarity. They are related to the nest because they are formed from alternating mainchain dihedral angles and have the ability to form compound and tandem forms, however, it is the differences between the two that are important. The catgrip is formed from dihedral angles in the β_R and β_L regions of the Ramachandran plot whereas the nest is formed from alternating γ_R and γ_L region residues. This is shown by molecular modelling to be important to the shape of the motif. The β-region where the catgrips are found is where the model polypeptides all have their CO groups more towards the centre of the ring. It is also the area where the CO groups are spaced apart enough to work together yet not interfere with each other, making the motif ideal for binding cations (see figure 6.3 below)

Figure 6.3 Four Ramachandran-like plots:

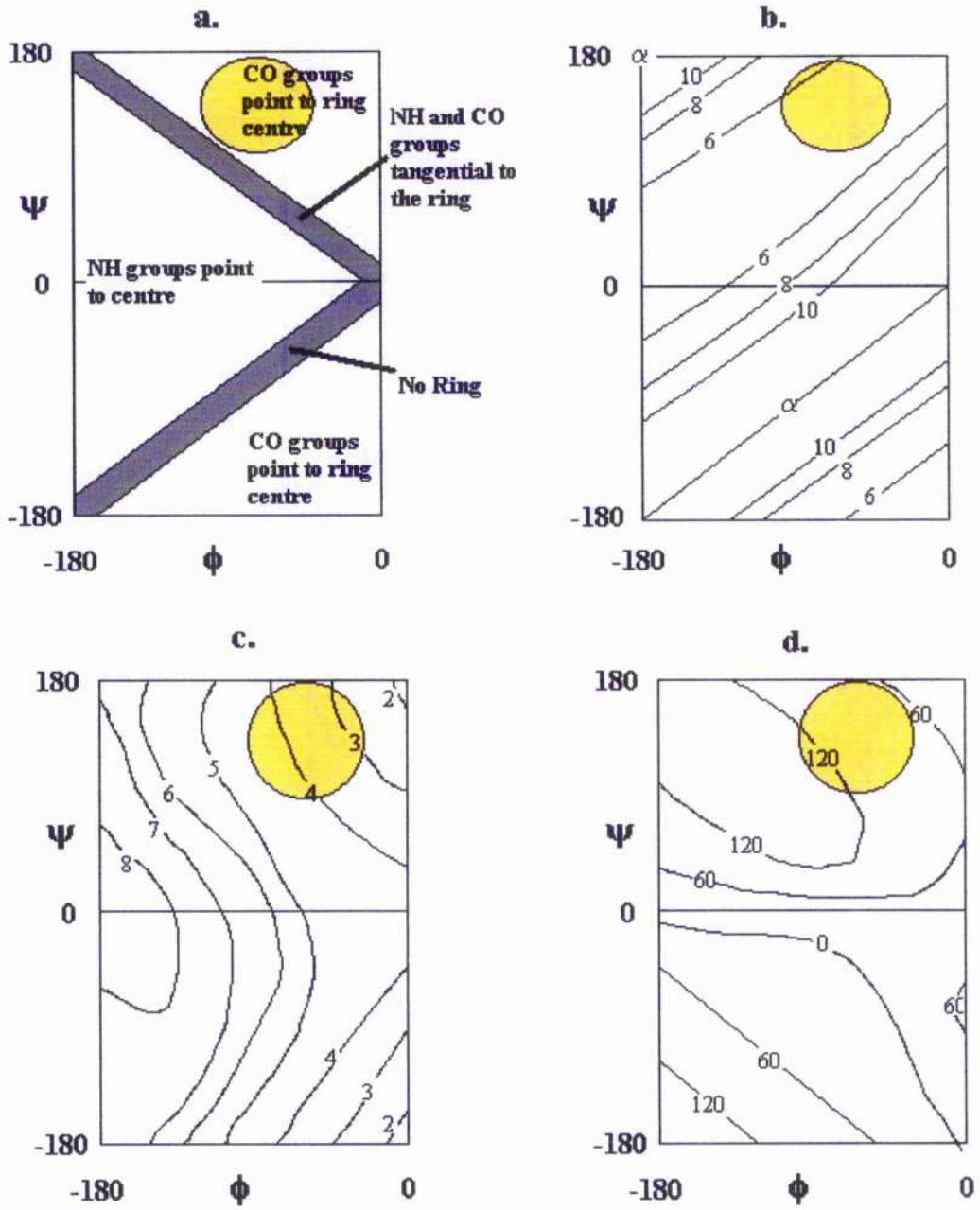
The catgrip region is highlighted in each diagram by a yellow circle.

6.3a= general geometry

6.3b= number of residues per ring

6.3c= O-O distance (angstroms)

6.3d= OCCO dihedral angle (degrees)



The other main difference between nests and catgrips is how the two motifs bind their ligand. In the nest the residues whose dihedral angles are important are labelled i and i+1, therefore the three NH groups that come together are from residues i, i+1 and i+2. In the equivalent catgrip, where the residues with enantiomeric mainchain dihedral angles are also labelled i and i+1, the mainchain CO groups that are involved in ligand-binding are from residues i-1 and i+1.

In conclusion the catgrip is a specific calcium-binding motif in a few proteins but has the potential to bind to any cationic ligand. Given more time the next step would be to conduct a more thorough search of the PDB using the dihedral angle definitions to find out if this conformation is a common one and if other catgrip motifs as yet unknown to us exist.

Chapter 7: Summary

Understanding how proteins fold up to form their three dimensional shape is of importance to the continued study of protein structure and function, yet our understanding of the processes involved is still limited. The investigation of protein structure continues to prove informative and much remains to be understood. It is for this reason we decided to investigate the conformations of polypeptides native in protein, which has led to the identification of three important novel motifs that occur widely.

When considering the conformation of polypeptides or proteins it is important to study the properties of the backbone or mainchain atoms. The peptide bond is unusual in that, due to resonance, it has a partial double bond character. This means that, instead of having free rotation around a single bond, it has a relatively fixed planar state similar to that of a double bond, but the bond lengths are midway between those of single and double bonds. There are therefore only two types of bond along the polypeptide backbone capable of rotation: phi(ϕ) and psi(ψ). The conformation of any polypeptide is therefore dictated by the combination of angles measured for ϕ and ψ. This was investigated by Ramachandran *et. al.* (Ramachandran, G. N. & Sasisekharan, V., 1968; Dickerson & Geis, 1969) in two ways. The first part of the study involved creating model "dipeptides" with every possible combination of ϕ and ψ then measuring the energy of each "dipeptide" giving an indication of how stable the conformation pertaining to each pair of dihedral angles is. The energy for each model "dipeptide" is then plotted and gives the familiar Ramachandran plot. This was solely a measure of which combinations of dihedral angles are favourable and which, through steric hindrance, are unfavourable.
The second part of Ramachandran's study involved creating model polypeptides to investigate what happens when successive amino acids have the same ϕ, ψ angles. Many of the polypeptides are helical in nature, and among them are the familiar secondary structural elements such as alpha helices, polyproline helices and beta strands.

While investigating the polypeptide conformations at the C-termini of alpha helices it became apparent that a pattern of dihedral angles occurs where successive amino acids

had mirror image mainchain conformations. It was decided to repeat the analysis by Ramachadran *et al* using polypeptides where successive amino acids have alternating enantiomeric (mirror image) mainchain conformations. The polypeptides formed are ring shaped or extended linear structures some of which are sterically impossible. A number, however, are shown to be common motifs in proteins and to have important structural and functional roles.

The first stable conformation to be found is that we have termed the 'nest' (Watson, J. D. & Milner-White, E. J., 2001). It is characterised by three mainchain NH groups that align to form a concavity. The partial charges on each of the atoms is determined by their electronegativity and gives the concave surface of the nest a partial positive charge making it a good anion-binding motif. There are two main types of nest motif distinguished by their pattern of dihedral angles. The most common type is the RL form, defined by the first residue having dihedral angles in the right handed region of the Ramachandran plot, while the second residue is left handed in conformation and has the mirror image range of dihedral angles. The other type of nest is the LR type and is less common, association occurring in only a couple of motifs.

The concavity of the nest can be defined by the angle between the three successive nitrogen atoms. In a bar-chart showing the distribution of this angle there is seen to be a distinct difference between the RL and LR forms. Although both follow a normal distribution the average angle is wider for the LR type of nest. This may reflect the differences between the motifs that use the LR and RL conformations or it may be due to the smaller sample size for the LR type (it is a less common conformation). The concavity of the nest has an average value of 125° and appears to provide a big enough opening to locate an anion in it. As the angle decreases the nest becomes more concave to the point where steric hindrance makes the conformation impossible. Increasing the angle of the nest forms a linear extended structure where the successive NH groups of the mainchain no longer point to a common point and the conformation is less likely to bind anions.

In the regular nest not all the NH groups point in the same direction. The first and the third NH groups point to a common region whereas the middle NH group points to the side, away from the plane of the nest. This negates the unfavourable electrostatic clashes that would be expected from bringing three NH groups together. The unusual

orientation of the three NH groups also allows the nest to bind to multiple anions. The primary anion is usually bound by the first and third NH group and the middle NH group is often used to bind a second anion. The types of anions bound vary and include the carbonyl oxygen atoms that have a partial negative charge (Baker, E. N. & Hubbard, R. E., 1984; Jeffrey, G. A., 1997), negatively charged Asp and partially negative Asn side-chains (Le Questel, J-Y et. al., 1993; Doig, A. J. et. al., 1997; Aurora, R. & Rose, G. D., 1998; Wan, W-Y & Milner-White, E. J., 1999; Chakrabarti, P. & Pal, D., 2001), and ligands such as phosphate groups (Chakrabarti, P., 1993; Copley, R. R. & Barton, G. J., 1994).

The majority of nests are associated with paperclip (Milner-White, E. J., 1988) (Schellman (Schellman, C., 1980)) motifs although other hydrogen bonded motifs also incorporate nests. These include Beta bulge loops (Milner-White, E. J. & Poet, R., 1987; Ring, C. S. et. al., 1992; Chan, E. A. W., 1993; Milner-White, E. J., 1987), Beta turns (four types) (Venkatachalam, C. M.,1968; Baker, E. N. & Hubbard, R. E., 1984; Milner-White, E. J. & Poet, R., 1987; Stickle, D. F. et. al., 1992) and Asx-turns (Le Questel, J-Y et. al., 1993; Shimoni, L. & Glusker, J. P., 1995; Wan, W-Y & Milner-White, E. J., 1999). Nests are not isolated structures and examples commonly occur where multiple nests associate to form larger structures capable of binding more complex anionic groups. These extended structures can be made from either overlapping (Compound) or adjacent (Tandem) nests. The compound nests are more common and generally have functional roles such as in the P-loop containing superfamily of proteins (Dreusicke, D. & Schulz, G. E., 1986; Pai, E. F. et. al., 1990). The tandem nests are usually structural motifs allowing the polypeptide chain to turn and accommodate other motifs. However, when found in conjunction with a compound motif, they can have functional roles.

It is evident there are a number of situations in proteins where the nest has a function. Those identified in our database are the P-loop (Dreusicke, D. & Schulz, G. E., 1986; Pai, E. F. et. al., 1990) (the phosphate-binding loop important to many kinases and G-proteins); the oxyanion hole of some serine proteases (Marquart, M. et. al., 1983); loops surrounding the Iron-Sulphur (Sticht, H. & Rosch, P., 1998) centres of a number of proteins; and the calcium-binding (Meador, W. E. et. al., 1993; Finn, B. E. & Drakenberg, T., 1993) loops of various proteins.

The way certain EF hand (Allouche, D. et. al., 1999) calcium-binding loops utilise nests requires further explanation. Here the cationic calcium ion is bound by Asx residues which are in turn stabilised through hydrogen bonding to nests (Watson, J. D. & Milner-White, E. J., 2001; Watson, J. D. & Milner-White, E. J., 2001; Chakrabarti, P., 1994). This arrangement seems to be stable and under closer examination contains another novel motif we have termed the Asx-nest (there is another related motif called the ST-nest). This is a nest where the first residue of the three is an Asp or Asn (Ser or Thr for the ST-nest) whose side chain oxygen occupies the nest. Therefore the sidechain binds to its own NH group and the two that follow. This interaction is not particularly strong but a program that calculates the energies of interaction between peptide groups shows that although the geometry is slightly unfavourable, electrostatic interactions do stabilise the motif.

The Asx- and ST-nests are useful motifs that occur in certain functional situations such as the EF hand calcium-binding loops. They may also have significance for protein folding. As the protein is synthesised it folds up in an aqueous environment. This encourages the protein to attempt to solvate those parts that are hydrophilic and bury the hydrophobic sections. It has been shown that the mainchain CO groups are more easily solvated than the NH groups (Collins, K. D., 1997). Therefore the polypeptide is more likely to use the NH groups to bind other parts of protein and the CO groups bind to the solvent. Since Asx- and ST-nests are relatively stable and require only that the residue after the Asx or ST in question be in a left handed conformation, it is likely they spontaneously form. These motifs could be nucleation points for the folding of the emerging polypeptide. Development of other secondary structural features would follow shortly afterwards to displace these nucleation points, often replacing the Asx- and ST- nests with more favourable interactions.

As mentioned earlier, the nest can also adopt a more extended form with a linear arrangement, having NH groups on one side and CO groups on the other. This is found in the potassium channel KscA (Zhou, M. et. al., 2001) from *Streptomyces lividans* in a stretch of polypeptide known as the selectivity filter that gives the channel its specificity for potassium ions over other cellular ions.
The potassium channel is a tetrameric protein and each chain has a conserved GYG signature sequence at the selectivity filter. Investigation of this polypeptide shows the residues involved have alternating enantiomeric dihedral angles. The dihedral angles

are found to be in the αR and αL regions of the Ramachandran plot. The mainchain conformation is more extended than the regular nest which usually has dihedral angles in the γR and γL regions. The four extended nests one from each subunit align next to one another around the four-fold axis with their mainchain carbonyl groups pointing to the centre. This forms a channel capable of allowing cations through. The high specificity for potassium ions is achieved in two ways: the first being the fact that the selectivity filter is lined with carbonyl groups making it impossible for anions to traverse the channel. The second way selectivity is achieved is through limiting the width of the filter. This only allows cations the size of potassium through since those that are too large cannot get into the channel and smaller cations do not pass because they try to drag through solvating water molecules.

This interesting situation is the only nest like structure observed where the mainchain carbonyl groups are the significant ligand-binding factors. It has been suggested that the selectivity filter may also be of significance to a group of proteins known as Aquaporins (Sui, H. et. al., 2001). These proteins are only just being characterised and transport water and/or solutes across membranes, and as such are important as they allow the cells to cope with osmotic potentials. Papers published on the subject suggest that the Aquaporins have an extended nest not unlike that of the potassium channel selectivity filter, but that the Aquaporins are more complex as they use sidechains in the channel to aid in their selectivity.

The third novel motif we have found is the catgrip (Watson, J. D. & Milner-White, E. J., 2001). This is a specific cation-binding motif used by a few types of protein to bind calcium. Although all the examples found so far bind calcium there is no reason to assume that other proteins might not use a form of the catgrip to bind other cations. The catgrip is formed by alternating mainchain dihedral angles in the βR and βL regions of the Ramachandran plot. This gives a ring shaped structure where the mainchain carbonyl groups of alternate residues point towards the centre of the ring. Due to the partial negative charge on the mainchain carbonyl group the motif forms a cation-binding motif.

All catgrips found to date can be classed in one of three groups determined by how many calciums are bound and how many mainchain carbonyl groups are used in the process. There may be other types yet to be determined but this is unlikely as the motif has a specific function and is uncommon apart from in those proteins that use it for a specific purpose.

The catgrip is a relative of the nest in the sense that it is formed from polypeptides with alternating enantiomeric mainchain dihedral angles, and multiple catgrips can come together to form larger compound and tandem structures. This is where the similarities end. The catgrip binds cations not anions, the mainchain carbonyls involved come from the residues before and after the first residue of the conformation (i.e. i-1 and i+1), and there is a much greater difference in conformation between compound and tandem catgrips than in nests.

The discovery of these novel motifs shows that continued analysis of protein structure reveals new features even in areas considered well understood. Although the situations these motifs are found in are well known the underlying processes involved are still to be fully elucidated. We find that these motifs do not seem to have a specific consensus sequence except for a propensity for Glycine at the γL residue, and as such do not contribute greatly to structure prediction. One significant aspect of these motifs is that they can be formed from a simple mixture of L- and D- amino acids. This means they may prove to be important in an evolutionary sense as well as being of interest to those studying how proteins fold up. In an evolutionary sense some of these motifs, namely the nest and Asx-/ST-nests, may be ancient. Their occurrence in proteins that are considered to have evolved early on (such as the methanogenesis and iron-sulphur proteins) seems to add weight to this opinion. Their importance to protein folding is more relevant to the Asx- and ST-nests which, as discussed earlier, may act as nucleation points for folding.

Overall there is still much work to be done. The first point I would address would be to try to automate the identification of the nests and catgrips using their dihedral angle ranges and patterns. This could then be used to identify the full range of motifs in the entire PDB which would help classify more motifs as well as aid in the analysis of those we are already aware of. I would also like to perform mutagenesis and protein folding studies to see if the Asx- and ST-nests can act as initiation points for folding up of secondary structure. Further analysis of the Aquaporin family of proteins would also be beneficial. If the extended nest conformation is found, it could aid in the understanding of how this conformation forms and what functions it has.

Chapter 7: Acknowledgements

Many thanks to:

BBSRC for funding the research.

Dr. E. J. Milner-White for all his help and support through the years.

Prof. G. Lindsay for his guidance and patience.

Prof. J. Coggins for his advice.

The X-ray crystallography department in particular:
Paul Emsley
Adrian Lapthorn

Stuart McKay for all his technical support.

All the staff in the Chemistry department.

I would also like to thank my friends for making sure I continued with the work even when things were not going my way, you are fantastic.

Most of all I would like to thank my family for their support and help through the years, I wouldn't have made it this far without you.

References

Allouche, D., Parello, J. & Sanejouand, Y-H. (1999). Ca^{2+}/Mg^{2+} exchange in parvalbumin and other EF-hand proteins. A theoretical study. *J. Mol. Biol.* **285,** 857-873

Aurora, R., Srinivasan, R. &Rose, G. (1994). Rules for α-helix termination by glycine. *Science*, **264**, 1126-1130

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

Barber, M. J., Neame, P. J., Lim, L. W., White, S. & Matthews, F. S. (1992). Correlation of X-ray and deduced amino acid sequence of trimethylamine dehydrogenase. *J. Biol. Chem.* **267**, 6611-6629

Baumann, U., Wu, K. M., Flaherty, K. M. & McKay, D.B. (1993). Three dimensional structure of alkaline protease of *Ps. aeroginosa*, a two-domain protein with a calcium-binding parallel β-roll. *EMBO J.* **12**, 3357-3364

Baumann, U. (1994). Crystal structure of the 50kDa metalloprotease from Serratia marcescens. *J. Mol. Biol.* **242**, 244-251

Becker, J. W., Marcy, A. I., Rokosz, L. L., Axel, M. G., Burbaum, J. J., Fitzgerald, P. M. D. *et al.* (1995). Stromelysin-1: three-dimensional structure of the inhibited catalytic domain. *Protein Sci.* **4**, 1966-1976

Beinert, H., Holm, R. H. & Munck, E. (1997). Iron-sulphur clusters: nature's modular multipurpose structures. *Science*, **277**, 653-659

Bhat, T. N., Bentley, G. A., Boulot, G., Greene, M. I., Tello, D., Dall'Acqua, W. et al. (1994). Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl. Acad. Sci. USA*, **91**, 1089-1098

Bhatnagar, R. S., Futterer, K., Farazi, T. A., Korolev, S., Murray, C. L., Jackson-Machelski, E. *et al.* (1998). Structure of N-myristoyltransferase with bound myristoyl CoA and peptide substrate analogs. *Nature Struct. Biol.* **5**, 1091-1097

Binda, C., Coda, A., Aliverti, A., Zanetti, G. & Mattevi, A. (1998). Structure of the mutant E92K of [2Fe-2S] ferredoxin from *Spinacia oleracia* at 1.7 Å resolution. *Acta Crystallog. sect. D*, **54**, 1353-1367

Bode, W., Gomis-Ruth, F. X. & Stocker, W. (1993). Astacins, serralysins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments. *FEBS Letters*, **331**, 134-140

Borkakoti, N. (1998). Matrix metalloproteinases: variations on a theme. *Prog. Biophys. Mol Biol.* **70**, 73-74

Browner, M. F., Smith, W. W. & Castelhano, A. L. (1995). Matrilysin – inhibitor complexes: common themes among metalloproteinases. *Biochemistry*, **34**, 6602-6610

Bullough, P. A., Hughson, F. M., Skehel, J. J. & Wilet, D. C. (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature*, **371**, 37-43

Carrell, C. J., Carrell, H. L., Erlebacher, J. & Glusker, J. P. (1988). Structural aspects of metal ion – carboxylate interactions. *J. Am. Chem. Soc.* **110**, 8651-8656

Carter, C. W., Jr (1977). New stereochemical analogies between iron sulphur electron transfer proteins. *J. Biol. Chem.* **252**, 7802-7811

Chakrabarti, P. (1990). Systematics in the interaction of metal ions with the main-chain carbonyl group in protein structure. *Biochemistry*, **290**, 651-658

Chakrabarti, P. (1990). Interactions of metal ions with carboxylic and carboxamide groups in protein structures. *Protein Eng.* **4**, 49-56

Chakrabarti, P. (1993). Anion-binding sites in protein structures. *J. Mol. Biol.* **234**, 463-482

Chakrabarti, P. (1994). Conformational analysis of carboxylate and carboxamide side-chains bound to cations. *J. Mol. Biol.* **239**, 306-314

Chakrabarti, P. & Pal, D. (2001). Interrelationships between side-chains and main-chain atoms in proteins. *Prog. Biophys. Mol. Biol.* **76**, 1-102

Chan, E. A. W., Hutchinson, E. G., Harris, D. & Thornton, J. M. (1993). Identification, classification and analysis of β-bulges in proteins. *Protein Sci.* **2**, 1574-1590

Choe, S. (2002). Potassium channel structures. *Nature Rev. Neurosci.* **3**, 115-121

Collins, K. D. (1997). Charge density-dependent strength of hydration and biological structure. *Biophys. J.* **72**, 65-76

Copley, R. R. & Barton, G. J. (1994). A structural analysis of phosphate- and sulphate-binding sites in proteins. Estimation of propensities for binding and conservation of phosphate binding sites. *J. Mol. Biol.* **242**, 321-329

Cowan, J. A. & Lui, S. M. (1998). Structure-function correlations in high potential iron proteins. *Advan. Inorg. Chem.* **45**, 313-350

Dauter, Z., Betzel, C., Genov, N. Pipon, N. & Wilson, K. S. (1991). Complex between the subtilisin from mesophilic bacteria and the leech inhibitor eglin C. *Acta Crystallog. sect. B*, **47**, 707-720

Dauter, Z., Wilson, K. S., Sieker, L. C., Meyer, J. & Moulis, K. M. (1996). Atomic resolution structure of *Clostridium acidurici* ferredoxin. *Biochemistry*, **36**, 16065-16073

Degtyarenko, K. N., North, A. C. T. & Findlay, J. B. C. (1999). PROMISE: a database of bioinorganic motifs. *Nucl. Acids. Res.* **27**, 233-236

Dickerson & Geis. *The structure and action of proteins (1969) Chapter 2: Bricks and mortar – the structural proteins.* Harper & Row, New York, USA.

Dijkstra, B. W., Kalk, K. H., Hol, W. G. J. & Drenth, J. (1981). Structure of bovine pancreatic phospholipase A2. *J. Mol. Biol.* **147**, 97-111

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

Donate, L. E., Rufino, S. D., Canard, L. H. J. & Blundell, T. L. (1996). Conformational analysis and clustering of short and medium-sized loops connecting regular secondary structures. *Protein Sci.* **5**, 2600-2616

Doyle, D. A., Cabral, J. M., Pfuetzner, R. A., Kuo, A., Gulbis, J. M. & Cohel, S.L. (1998). The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science*, **280**, 69-77

Dreusicke, D. & Schulz, G. E. (1986). The glycine-rich loop of adenylate kinase forms a giant anion hole. *FEBS Letters,* **208**, 301-304

Dutzler, R., Campbell, E. B., Cadene, M., Chalt, B. T. & MacKinnon, R. (2002). X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature*, **415**, 287-294

Efimov, A. V. (1991). Structure of α–α-hairpins with short connections. *Protein Eng.* **4**, 245-250

Efimov, A. V. (1993). Standard structures in proteins. *Prog. Biophys. Mol. Biol.* **60**, 201-239

Eswar, N. & Ramakrisnan, C. (1999). Secondary structures without backbone: an analysis of backbone mimicry by polar side-chains in proteins. *Protein Eng.* **12**, 447-455

Fernandez-Catalan, C., Bode, W., Huber, R., Turk, D., Calvete, J. J., Lichte, A. *et al.* (1998). Crystal structure of a complex formed by the membrane type 1 matrix metalloproteinase and the soluble gelatinase A receptor. *EMBO J.* **17**, 5238-5248

Finn, B. E. & Drakenberg, T. (1993). Calcium-binding proteins. *Advan. Inorg. Chem.* **46**, 441-494

Ghadhiri, M. R., Granja, J. R., Milligan, R. A., McRee, D. E. & Khazanovich, N. (1993). Self-assembling organic nanotubules based on a cyclic peptide architecture. *Nature*, **366**, 324-327

Gandini, D., Gogioso, L., Bolognesi, M. & Bordo, D. (1996). Patterns in ionisable side-chain interactions in protein structures. *Proteins: Struct. Funct. Genet.* **24**, 439-449

Gliubich, F., Gazzero, M., Zanotti, G., Delbono, S.,Bombieri, G. & Berni, R. (1996). Active site structural features of chemically modified forms of rhodanese. *J. Biol. Chem.* **271**, 21054-21061

Gros, P., Betzel, Ch., Dauter, Z., Wilson, K. S. & Hol, W. G. L. (1989). Molecular dynamics refinement of A thermitase-eglin complex at 1.9 Å. *J. Mol. Biol.* **210**, 347-367

Gunasekaran, K., Nagarajaram, H. A., Ramakrishnan, C. & Balaram, P. (1998). Stereochemical punctuation marks in protein structures: glycine and proline containing helix stop signals. *J. Mol. Biol.* **275**, 917-932

Gunasekaran, K., Ramakrishnan, C. & Balaram, P. (1997). β-Hairpins in proteins revisited: lessons for de novo design. *Protein Eng.* **10**, 1131-1141

Harding, M. M. (1999). The geometry of metal-ligand interactions relevant to proteins. *Acta Crystallog. sect. D*, **55**, 1432-1443

Harding, M. M. (2001). The geometry of metal-ligand interactions in proteins. *Acta Crystallog. sect. D*, **57**, 401-411

Harrison, D. H., Bohren, K. M., Ringe, D., Petsko, G. A. & Gabbay, K. H. (1997). The alrestatin double decker: binding of inhibitors to human aldose reductase. *Biochemistry*, **36**, 16134-16140

Hemmingsen, J. M., Gennert, K. M., Richardson, J. S. & Richardson, D. C. (1994). The tyrosine corner: a feature of most greek key β-barrel proteins. *Protein Sci.* **3**, 1927-1937

Hobohm, U. & Sander, C. (1994). Enlarged representative set of protein structures. *Protein Sci.* **3**, 522-541

Hol, G. W. J., Van Duijnen, P. T. & Berendsen, H. J. C. (1978). the alpha-helix dipole and the properties of proteins. *Nature*, **273**, 443-446

Holm, R. H., Kennepohl, P. & Solomon, E. I. (1996). Structural and functional aspects of metal site binding in biology. *Chem. Rev.* **96**, 2239-2314

Huber, R., Berendes, R., Burger, A., Schneider, M., Karshikov, A. & Luecke, H. (1992). Crystal and molecular structures of human annexin V. *J. Mol. Biol.* **223**, 683-704

Ippolito, J. A., Alexander, R. S. & Christianson, D. W. (1990). Hydrogen bond stereochemistry in protein structure and function. *J. Mol. Biol.* **215**, 457-471

Iwata, S., Saynovits, M., Link, T. A. & Michel, H. (1996). Structure of a water-soluble fragment of the 'Rieske' iron-sulphur protein of the bovine heart mitochondrial cytochrome bc1 complex. *Structure*, **4**, 567-579

Jeffrey, G. A. (1997) *An introduction to hydrogen bonding*. Oxford University 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.

Kakuta, Y., Pedersen, L. G., Carter, C. W., Negishi, M. & Pedersen, L. C. (1997). Structure of estrogen sulphotransferase. *Nature Struct. Biol.* **4**, 904-908

Karlin, K. D. (1993). Metalloenzymes: structural motifs and inorganic models. *Science*, **261**, 701-708

Koncha, N. O., Head, J. F., Kaetzel, M. A., Dedman, J. R. & Seaton, B. A. (1993). Rat annexin V crystal structure: calcium-induced conformational changes. *Science*, **261**, 1321-1329

Laskowski, R. A., Jones, M. L. & Thornton, J. M. (1997). PDBsum: a web-based database of summaries and analyses of all PDB structures. *Trends Biochem. Sci.* **22**, 488-490

Le Questel, J-Y., Morris, D. G., Maccallum, P. H., Poet, R. & Milner-White, E. J. (1993). Common ring motifs in proteins involving asparagine or glutamine amide groups hydrogen-bonded to main chain atoms. *J. Mol. Biol.* **231**, 888-896

Maccallum, P. H., Poet, R. & Milner-White, E. J. (1995). Coulombic attractions between partially charged main-chain atoms stabilise the right-handed twist found in most β-strands. *J. Mol. Biol.* **248**, 374-384

Maccallum, P. H., Poet, R. & Milner-White, E. J. (1995). Coulombic attractions between partially charged main-chain atoms stabilise the right-handed twist found in most β-strands. *J. Mol. Biol.* **248**, 374-384

Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983). The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and complexes with inhibitors. *Acta Crystallog. sect. B*, **39**, 480-498

McDonald, I. K. & Thornton, J.M. (1994). Hydrogen bonding in proteins. *J. Mol. Biol.* **238**, 777-793

McLaughlin, P. J., Gooch, J. T., Mannherz, H. G. & Weeds, A. G. (1993). Structure of gelsolin segment 1-actin complex and the mechanism of filament severing. *Nature*, **364**, 685-689

Meador, W. E., Means, A. R. & Quiocho, F. A. (1993). Modulation of calmodulin plasticity in molecular recognition on the basis of X-ray structures. *Science*, **262**, 1718-1721

Milner-White, E. J. (1987). Beta-bulges within loops as recurring motifs of protein structure. Biochim. *Biophys. Acta*, **911**, 261-265

Milner-White, E. J. & Poet, R. (1987). Loops, bulges, turns and hairpins in proteins. *Trends Biochem. Sci.* **12**, 189-192

Milner-White, E. J. (1988). Recurring loop-motif in proteins that occurs in right-handed and left-handed forms. *J. Mol. Biol.* **199**, 503-511

Milner-White, E. J. (1999). The N-terminal domain of MDM2 resembles calmodulin and its relatives. *J. Mol. Biol.* **292**, 957-963

Minor, D. L., Jr (2001). Potassium channels: life in the post-structural world. *Curr. Opin. Struct. Biol.* **11**, 408-414

Morais-Cabral, J. H., Zhou, Y., & MacKinnon, R. (2001). Energetic optimization of ion conduction rate by the K^+ selectivity filter. *Nature*, **414**, 37-42

Nataraj, D. V., Srinivasan, N., Sowdhami, R. & Ramakrishnan, C. (1996). Alpha turns in protein structures. *Curr. Sci.* **69**, 434-447

Oliva, B., Bates, P. A., Querol, E., Aviles, F. X. & Sternberg, M. J. E. (1997). An automated classification of the structure of protein loops. *J. Mol. Biol.* **266**, 814-830

Pai, E. F., Krengel, U., Petsko, G. A., Goody, R. S., Kabsch, W. & Wittinghofer, A. (1990). Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.3 Å. *EMBO J.* **9**, 2351-2359

Pavone, V., Benedetti, E., DiBlasio, B., Lombardi, A., Pedone, C., Tomisch, L. & Lorenzi, G. P. (1989). Regularly alternating L,D-peptides III. Hexacyclic peptides from valine or phenylalanine. *Biopolymers*, **28**, 215-223

Pavone, V., Maglio, O., Isernia, C. & Saviano, M. (1996). Discovering protein secondary structures. Classification and description of isolated α-turns. *Biopolymers*, **38**, 705-721

Peters, J. W., Lanzilotta, W. N., Lemon, B. J. & Seefeldt, L. C. (1998). X-ray crystal structure of the Fe-only hydrogenase from *C. pasteurianum*. *Science*, **282**, 1853-1861

Preissener, R., Egner, U. & Saenger, W. (1991). Occurrence of bifurcated three-centre hydrogen bonds in proteins. *FEBS Letters*, **288**, 192-196

QUANTA Generating and Displaying Molecules, July 1997. San Diego: Molecular Simulations Inc., 1997.

Ramachandran, G. N. & Sasisekharan, V. (1968). Conformations of polypeptides and proteins. *Advan. Protein Chem.* **23**, 283-437

Rao, S., Schaffie, F., Yu, C., Satyshur, K., Stockman, B., Markley, J. & Sundaralingam, M. (1992). Structure of the oxidised long chain flavodoxin from anabena 7170. *Protein Sci.* **1**, 1413-1431

Richardson, J. S. (1981). Protein anatomy. *Advan. Protein Chem.* **34**, 167-339

Ring, C. S., Kneller, D. G., Langridge, R. & Cohen, F. E. (1992). Taxonomy and conformational analysis of loops in proteins. *J. Mol. Biol.* **22**, 685-699

Scapin, G., Grubmeyer, C. & Sacchettini, J. C. (1995). The crystal structure of orotate phosphoribosyltransferase complexed with orotate α-D-5-phosphoribosyl-1-pyrophosphate. *Biochemistry*, **34**, 10744-10754

Schellman, C. (1980). The αL conformation at the ends of helices. In *Protein Folding* (Jaenicke, R., ed.), pp. 53-61, Elsevier, Amsterdam, North Holland

Schnittke, J. L., Stern, L. J. & Klibanov, A. M. (1996). Crystal structure of subtilisin Carlsberg in anhydrous dioxane. *Proc. Natl. Acad. Sci. USA*, **94**, 4250-4264

Scott, D. L., White, S. P., Otwinowski, Z., Yuan, W., Gelb, M.H. & Sigler, P. B. (1990). Interfacial catalysis: the mechanism of phospholipase A2. *Science*, **250**, 1541-1546

Shimoni, L & Glusker, J. P. (1995). Hydrogen bonding motifs of protein side-chains. *Protein Sci.* **4**, 65-74

Spraggon, G., Everse, S. J. & Doolittle, R. F. (1997). Crystal structures of fragment D from human fibrinogen and its crosslinked counterpart from fibrin. *Nature*, **389**, 455-462

Spurlino, J. C., Smallwood, A. M., Carlton, D. D., Banks, T. M., Vavra, K. J., Cook, E. R. *et al.* (1994). Structure of mature truncated human fibroblast collagenase. *Proteins: Struct. Funct. Genet.* **19**, 98-109

Stephens, P. J., Jollie, D. R. & Warshel, A. (1996). Protein control of redox potentials of iron-sulphur proteins. *Chem. Rev.* **96**, 2491-2513

Sticht, H. & Rosch, P. (1998). The structure of iron-sulphur proteins. *Prog. In Biophys. Mol. Biol.* **70**, 95-136

Stickle, D. F., Presta, L. G., Dill, K. A. & Rose, G. D. (1992). Hydrogen bonding in globular proteins. *J. Mol. Biol.* **226**, 1143-1159

Stryer, L. *Biochemistry - Third Edition*. 1988. W. H. Freeman and Company, New York, USA.

Sui, H., Han, B-G., Lee, J. K., Wallan, P. & Jap, B. K. (2001). Structural basis of water-specific transport through the AQP1 water channel. *Nature*, **414**, 872-878

Vassylyev, D. G., Katayanagi, K., Ishikawa, K., Tsujimoto-Hirano, M., Danno, M., Pahler, A. *et al.* (1993). Crystal structures of ribonuclease F1 from *Fusarium moniliforme* in its free form and complexed with 2'GMP. *J. Mol. Biol.* **230**, 979-996

Venkatachalam, C. M. (1968). Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. *Biopolymers* **6(10)**, 1425-1436

Via, A., Ferre, F., Brannetti, B., Valencia, A. & Helmer-Citterich, M. (2000). Three-dimensional view of the surface motif associated with the P-loop structure. *J. Mol. Biol.* **303**,45-465

Viguera, A. R. & Serrano, L. (1995). Experimental analysis of the schellman motif. *J. Mol. Biol.* **251**, 150-160

Wan, W-Y. & Milner-White, E. J. (1999). A natural grouping of motifs with an aspartate or asparagine forming two hydrogen bonds to residues ahead in sequence: their occurrence at α-helical N termini and in other situations. *J. Mol. Biol.* **286**, 1633-1649

Wan, W-Y. & Milner-White, E. J. (1999). A recurring two-hydrogen-bond motif incorporating a serine or threonine residue is found both at α-helical N termini and in other situations. *J. Mol. Biol.* **286**, 1650-1666

Watson, J. D. & Milner-White, E. J. (2001). A novel main-chain anion binding site in proteins: the nest. A particular combination of ϕ, ψ values in successive residues gives rise to anion binding sites that occur commonly and are found at functionally important regions. *J. Mol. Biol.* **315**, 171-182

Watson, J. D. & Milner-White, E. J. (2001). The conformations of polypeptide chains where the ϕ, ψ values of alternating residues are enantiomeric. Their occurrence in cation and anion binding regions of proteins. *J. Mol. Biol.* **315**, 183-191

Yellen, G. (2001). Keeping K^+ completely comfortable. *Nature Struct. Biol.* **8(12)**, 1011-1013

Zhou, Y., Morais-Cabral, J. H., Kaufmann, A. & MacKinnon, R. (2001). Chemistry of ion coordination and hydration revealed by a K^+ channel-Fab complex at 2.0 Å resolution. *Nature*, **414**, 43-48

Zhou, M., Morais-Cabral, J. H., Mann, S. & Mackinnon, R. (2001). Crystal structure of the potassium channel KscA. *Nature*, **411**, 657-661