UNIVERSITY OF SOUTHAMPTON

Structure Led Drug Design for the Pentraxins

By Michelle Catherine Jenvey

A Thesis Submitted for the Degree of Doctor of Philosophy

Division of Biochemistry & Molecular Biology, Faculty of Medicine, Health & Life Sciences.

September 2006

University of Southampton

Abstract

Faculty of Medicine, Health and Life Sciences
School of Biological Sciences, Division of Biomolecular Sciences

Doctor of Philosophy

Structure Led Drug Design for the Pentraxins

By Michelle Catherine Jenvey

The pentraxin family of proteins have remained conserved throughout evolution and are thought to have important physiological roles. Serum Amyloid P component (SAP) and C-reactive protein (CRP) are members of the 'short' pentraxin sub-family and are composed of five identical subunits arranged symmetrically around a central, hollow pore. SAP is always found bound to amyloid deposits in organs; whilst CRP is the major acute phase protein in humans, and has been associated with the pathology of coronary heart disease. The three-dimensional structures of these proteins have already been solved using X-ray crystallographic methods and have enabled the structure-led design of ligands targeted specifically at these proteins. The X-ray crystal structures of SAP and CRP in complex with these ligands are presented.

CPHPC, the ligand targeted specifically at SAP, consists of two proline residues joined by a six carbon linker and was identified using the Roche compound library. Crystals of SAP in complex with CPHPC were grown in novel crystallisation conditions and diffracted X-rays to a resolution of 1.6 Å. The three-dimensional structure of the complex reveals the formation of an SAP decamer consisting of two molecules of SAP linked by five CPHPC molecules. The peptide bond in the proline residues of the CPHPC molecules can exist in either the *cis* or *trans* conformation. The identification of the preferred conformation in the CPHPC molecule could potentially increase the potency of the drug in its inhibition of SAP binding to amyloid deposits. The 1.6 Å structure of the SAP-CPHPC suggests that the *cis* conformation is the preferred state of the proline peptide bond in CPHPC.

PCHPC, the ligand targeted at CRP, was created using the rational drug-design approach and is composed of two phosphocholine residues (the natural ligand of CRP) joined by a six carbon linker. The structure of the CRP-PCHPC complex reveals an identical decamer to that formed by the SAP-PCHPC complex with five PCHPC molecules linking two CRP molecules.

Alternative approaches to drug design for both SAP and CRP are available that may improve the potency of CPHPC and PCHPC. Multivalent ligands have been designed for both the Shiga-like and cholera toxins and have been shown to improve binding affinities with respect to their monovalent equivalents. The design of multivalent compounds incorporating CPHPC and PCHPC may therefore provide an alternative method by which the design of inhibitory compounds for SAP and CRP may move forward in the future.

Contents

List of Figures

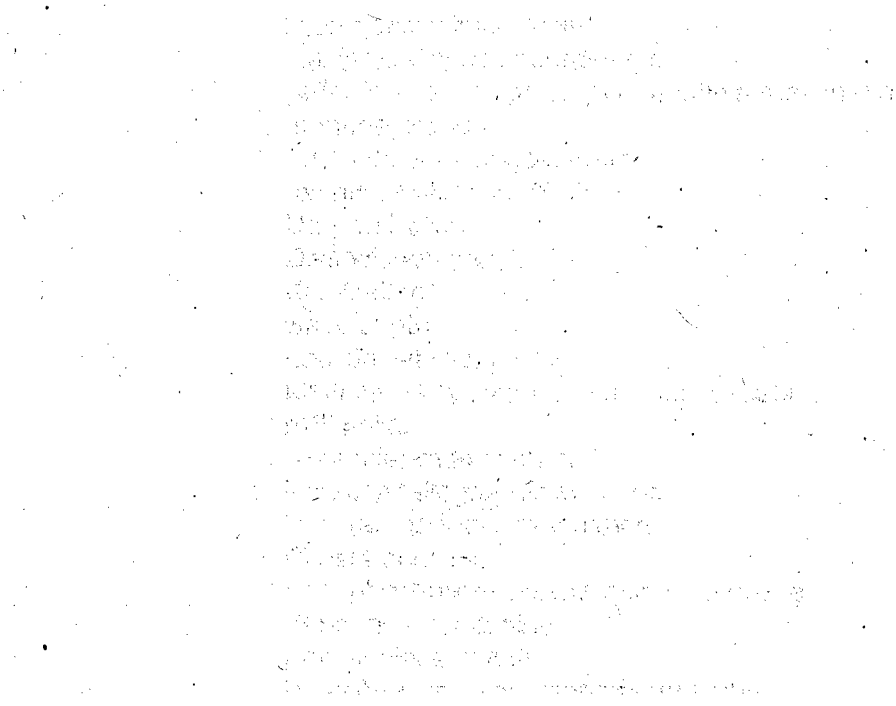
List of Tables

Acknowledgements

Mum, Dad and Sis, I couldn't have asked for any more love and support, lots of love as always. Nick, I'm so glad I came around in the end, and without you I would never have crossed the finish line, all my love.

Thanks to all my friends in Southampton, particularly Rob and Sah for putting up with me throughout the whole 3 year both at work and home. Also, to Andy, Hali and Vezza for drinks and moans.

I would like to thank the Wood/Cooper lab members past and present. A special thank you to Dazza and Fiyaz, for all their help in crystallographic matters and also for making the early days a real good laugh.



Abbreviations

Å	Ångstrom (1×10^{-10}m)
AA	Amyloid fibrils derived from serum amyloid A
a,b,c	Unit cell dimensions
α,β,γ	Unit cell angles, or Eulerian angles
Aβ	Alzheimer's diease β-protein peptide
AL	Amyloid fibrils derived from monoclonal immunoglobulin light chains
B-factor	Temperature factor
°C	Degree Celsius
CC	Correlation coefficient
CCD	Charge-coupled device
CCP4	Collaborative Computational Project Number 4
CLR	Collagen-like region
CNS	Crystallography & NMR system
CPHPC	R-1-[6-[R-2-carboxy-pyrrolidin-1-yl]-6-oxohexanoyl]pyrrolidine-2-carboxylic acid
CRP	C-reactive protein
Da	Dalton
dAMP	2'-deoxyadenosine-5'-monophosphate
DNA	Deoxyribose nucleic acid
EDTA	Ethylene-diamine-tetraacetic acid
EMBL	Europeran Molecular Biology Laboratory
ESRF	European Synchrotron Radiation Faciltity
GABA	γ-aminobutyric acid
HDL	High density lipoprotein
h,k,l	Reciprocal space coordinates
(h,k,l)	Miller indices for lattice planes within real space
Ig	Immunoglobulin
IC_{50}	50% inhibitory concentration
^{125}I-SAP	Iodine-125 labelled SAP
K	Degree Kelvin
K_d	Dissociation constant
KDa	Kilo-Dalton
λ	wavelength
LDL	Low density lipoprotein
MAD	Multiple wavelength anomalous scattering
mM	millimollar
MR	Molecular replacement
NCS	Non-crystallographic symmetry
NMR	Nuclear magnetic resonance
PC	Phosphocholine
PCHPC	Phosphocholine-hexane-phosphocholine
PE	Phoshoethanolamine
PEG	Polyethylene glycol
PEG MME	Polyethylene glycol monomethol ether
pH	$-\log_{10}$[hydrogen ion concentration]
ppm	Parts per million

ψ,φ,ω	(psi, phi, omega) main chain dihedral angles
φ,ω,χ	(phi, omega, chi) spherical polar angles
R factor	Residual index factor
rms	Root mean square
SAP	Serum amyloid protein
SLE	Systemic lupus erythematous
snRNPs	Small nuclear ribonucleoprotein particles
Tris	Tris (hydroxymethyl)aminomethane
x,y,z	Unit cell coordinates (real space)

Chapter 1
Introduction – The Pentraxins

1.1 Pentraxins

The pentraxin proteins were originally named for their distinct structural organisation of five identical subunits arranged non-covalently in pentameric radial symmetry. The name 'pentraxin' is derived from the Greek *penta* meaning five and *ragos* meaning berries. All members of the family contain the 'pentraxin family signature': a sequence of amino acids consisting of HxCxS/TWxS (x = any amino acid).

The family currently consists of seven members, which depending on their size, are divided into two distinct structural classes. The first of these structural groups are the novel 'long' pentraxins having molecular weights that range from 200-250 kDa. The long pentraxins are a fairly recent discovery, the first member pentraxin 3 (PTX3), was described in 1993 (Lee *et al.*, 1993). The second of these groups is the classical 'short' pentraxins, the two members having molecular weights of 115 kDa and 120 kDa. The first member of the short pentraxins was identified as early as the 1930's (Tillet & Francis, 1930). Homology between the long and short pentraxins is limited to the carboxy-terminus, suggesting that the long pentraxins resulted from the fusion of novel amino terminal domains to an ancestral pentraxin domain.

There are currently five members of the long pentraxin subgroup: XL-PXN1; the sperm acrosomal protein Apexin/p50; the neuronal proteins NPI and NPTX2; and pentraxin 3 (PTX3). Since these proteins have only recently been identified, their precise physiological roles have yet to be determined. However, they all function in a calcium-dependent manner, a characteristic that is common to all pentraxin family members.

This report will focus on the short pentraxins of which there are currently two members: C-reactive protein (CRP) and serum amyloid P component (SAP). CRP and SAP are evolutionary conserved in all vertebrates, and in some invertebrates including the phylogenetically ancient *Limulus polyphenus* (Horseshoe crab) (Shrive *et al.*, 1999). Human CRP and SAP share 51% sequence identity and 66% homology (Srinivasan *et al.*, 1994) (Figure 1.1). They are characterised by their different calcium-dependent binding specificities for phosphocholine (PC) and phosphoethanolamine (PE), a property that is utilised in the preparation of both proteins from blood plasma (Hawkins *et al.*, 1991; Pontet *et al.*, 1978; De Beer & Pepys, 1982).

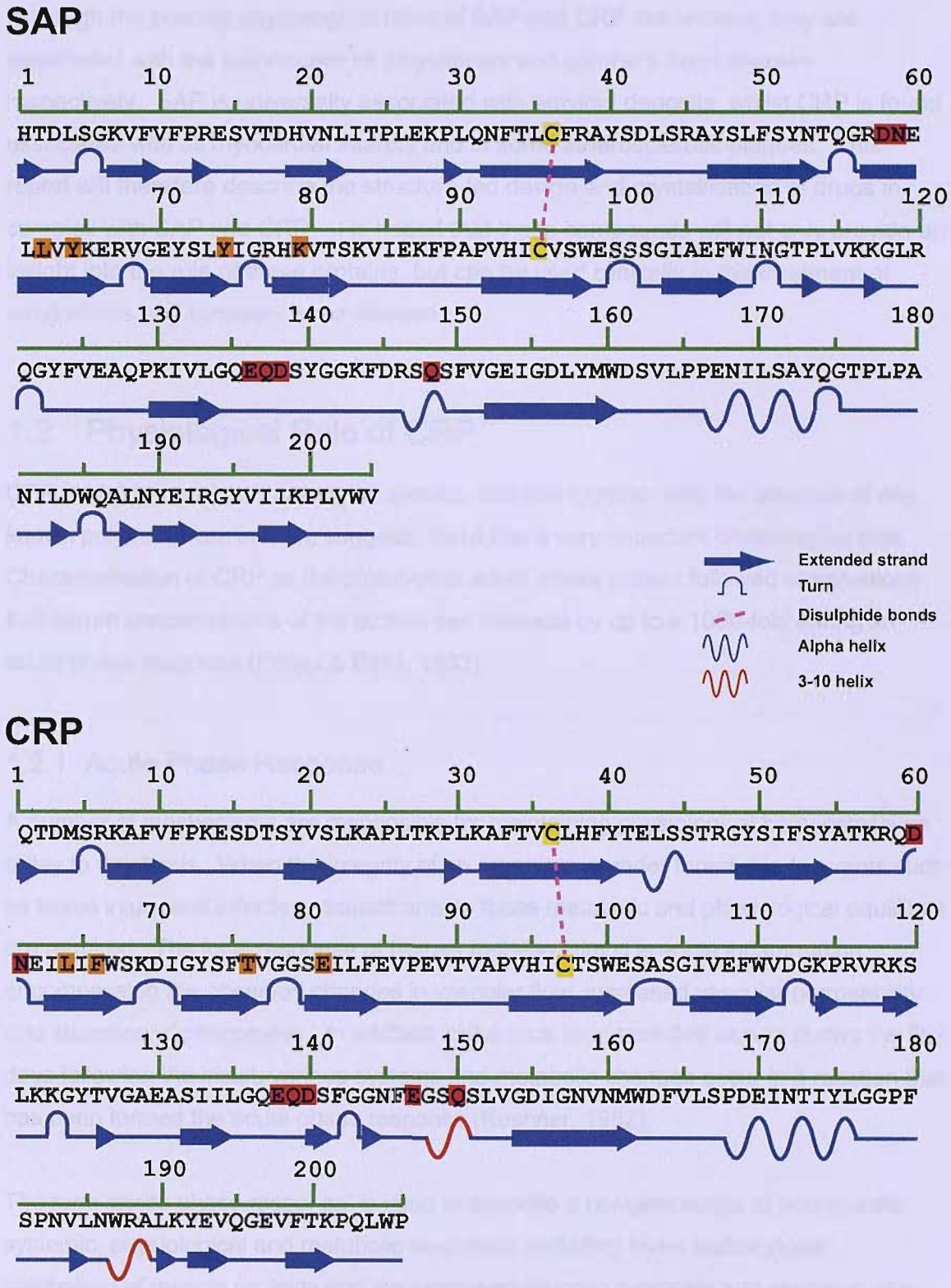


Figure 1.1 Primary sequence and secondary structure comparison of SAP and CRP. Residues involved in calcium and ligand binding are highlighted in red and orange respectively.

Although the precise physiological roles of SAP and CRP are unclear, they are associated with the pathologies of amyloidosis and coronary heart disease respectively. SAP is universally associated with amyloid deposits, whilst CRP is found associated with all myocardial infarcts and in some atherosclerotic plaques. This report will therefore describe the structure-led design and crystallisation of drugs in complex with SAP and CRP. It is hoped that these compounds will not only provide an insight into the role of these proteins, but can be used clinically in the treatment of amyloidosis and coronary heart disease.

1.2 Physiological Role of CRP

CRP is highly conserved among all species, and this together with the absence of any known polymorphism in man, suggests that it has a very important physiological role. Characterisation of CRP as the prototypical acute phase protein followed observations that serum concentrations of the protein can increase by up to a 1000-fold during an acute phase response (Pepys & Baltz, 1983).

1.2.1 Acute Phase Response

A number of mechanisms are responsible for maintaining physiological homeostasis on a day to day basis. When the integrity of an organism is under threat due to events such as tissue injury and infection, adjustments to these metabolic and physiological equillibria are required. The local response of tissues to these stimuli is acute inflammation, encompassing the observed changes in vascular flow, increased vascular permeability and attraction of phagocytes. In addition to the local response that occurs during the first days following the insult, various systems and metabolic changes occur in a reaction that has been termed the acute phase response (Kushner, 1982).

The term 'acute phase response' is used to describe a complex series of non-specific, systemic, physiological and metabolic responses including fever, leukocytosis, catabolism of muscle proteins and the increased *de novo* synthesis and secretion of a number of plasma proteins including CRP. A similar response is observed in all endothermic animals indicating that it may have survival value. The observed increase in proteinase inhibitors, complement, clotting and transport proteins observed during the response presumably enhances host resistance, minimises tissue injury and promotes tissue repair and regeneration (Pepys & Hirschfield, 2003).

1.2.2 Acute Phase Protein

CRP is a normal plasma protein, the circulating concentration of which increases dramatically during an acute phase response. The median concentration of circulating CRP in a healthy individual is approximately 0.8 mg/L, rising to peak values as high as 400-500 mg/L at the height of an acute phase response (Pepys & Baltz, 1983). These concentrations decrease rapidly following cessation of the stimulus for increased production. Due to this observation, CRP has been classified as one of the major acute phase proteins in humans.

CRP fails to undergo significant local sequestration or consumption, fragmentation, or complex formation, and the plasma half-life of CRP (19 hours) remains constant during all conditions of health and disease (Ridker, 2001). Therefore, the sole determinant of the dramatic increase in circulating concentrations observed during an acute phase response must largely reflect the increase in CRP synthesis.

Expression of the CRP gene is regulated predominantly at the transcriptional level by the cytokine interleukin-6 (IL-6). Synthesis of plasma CRP is performed predominantly in the liver; additional studies have, however, suggested that extra-hepatic expression may take place in alveolar macrophages and subsets of lymphocytes and monocytes (Black *et al.*, 2004).

Although the precise action of CRP in the acute phase response remains unclear, an insight has been provided by the identification of PC and the complement component C1q as major ligands of CRP.

1.2.3 Binding to PC Residues of Biological Membranes

The principal ligand of CRP was first identified in the 1930's by Tillet and Francis, who found that the addition of acute phase sera to a solution of C-polysaccharide from *pneumococci* caused precipitation of a then unknown protein, CRP. The substance in the cell wall C-polysaccharide of *pneumococci* to which CRP bound, termed 'fraction C', was shown to contain nitrogen and carbohydrate (Tillet *et al.*, 1930), and phosphorous , but remained unidentified until 1968. In 1968, Brundish and Baddiley successfully characterised fraction C as a ribitol techoic acid containing PC. The precise identity of the major reacting site for CRP in the cell wall C-polysaccharide however, continued to elude researchers. Three years later, in an attempt to elucidate

the precise nature of this reacting site, inhibition studies of CRP/C-polysaccharide precipitation were performed using PC and a series of organic phosphate monoesters. The results showed that PC was the most active inhibitor of CRP/C-polysaccharide precipitation (Volanakis & Kaplin, 1971), and identified PC as the major reacting site for CRP in C-polysaccharide. Subsequently it has been shown that CRP binds to PC residues with high affinity (K_d =1-2 x 10^{-5} M) (Anderson *et al.*, 1978).

CRP was shown to associate with the cell membranes of damaged and necrotic cells, but not normal intact cells, at sites of inflammation (Kushner and Kaplan, 1961). In response to these observations, the ability of CRP to bind to intrinsic phospholipids with PC-containing polar head groups (lecithin and sphingomyelin) was investigated. Indirect evidence of an interaction was shown when mixtures of CRP and an emulsion of these lipids resulted in activation of the classical pathway of complement in human serum (Kaplan & Volanakis, 1974; Volanakis & Kaplan, 1974); a reaction that was inhibited by the addition of PC. Volanakis and Wirtz (1979) showed that the binding of CRP to the PC head group of lecithin model membranes required the incorporation of lyso-lecithin. This provided support for earlier studies that had demonstrated CRP binding to membrane structures of altered or necrotic but not normal cells.

Further to this work, binding of CRP to red blood cell membranes was only observed following the treatment of the cells with phospholipase A_2 (Narkates & Volanakis, 1982). Human secretory phospholipase A_2 ($sPLA_2$) hydrolyses the fatty acid ester bond in position 2 of phospholipids generating lyso-phospholipids, as shown in Figure 1.2.

Due to the tightly packed nature of the outer leaflet of the membrane, the enzyme is not efficiently able to hydrolyse the outer leaflet phospholipids of normal intact cells. Phospholipids are distributed asymmetrically among the inner and outer leaflets of the membrane. The inner leaflet is composed mainly of phosphatidylserine (PS) and phosphatidylethanolamine (PE); sphingomyelin and phosphatidylcholine are the main constituents of the outer leaflet (see Figure 1.3a). This asymmetry is maintained by an energy-dependent process requiring adenosine triphosphate (ATP). Low intracellular concentrations of ATP due to cellular damage result in disruption of this asymmetry; phospholipids, in particular PS and PE, are now free to exchange between the inner and outer leaflets in a process referred to as `flip-flop'. As a consequence of this redistribution, outer leaflet-phospholipids are susceptible to hydrolysis by $sPLA_2$,

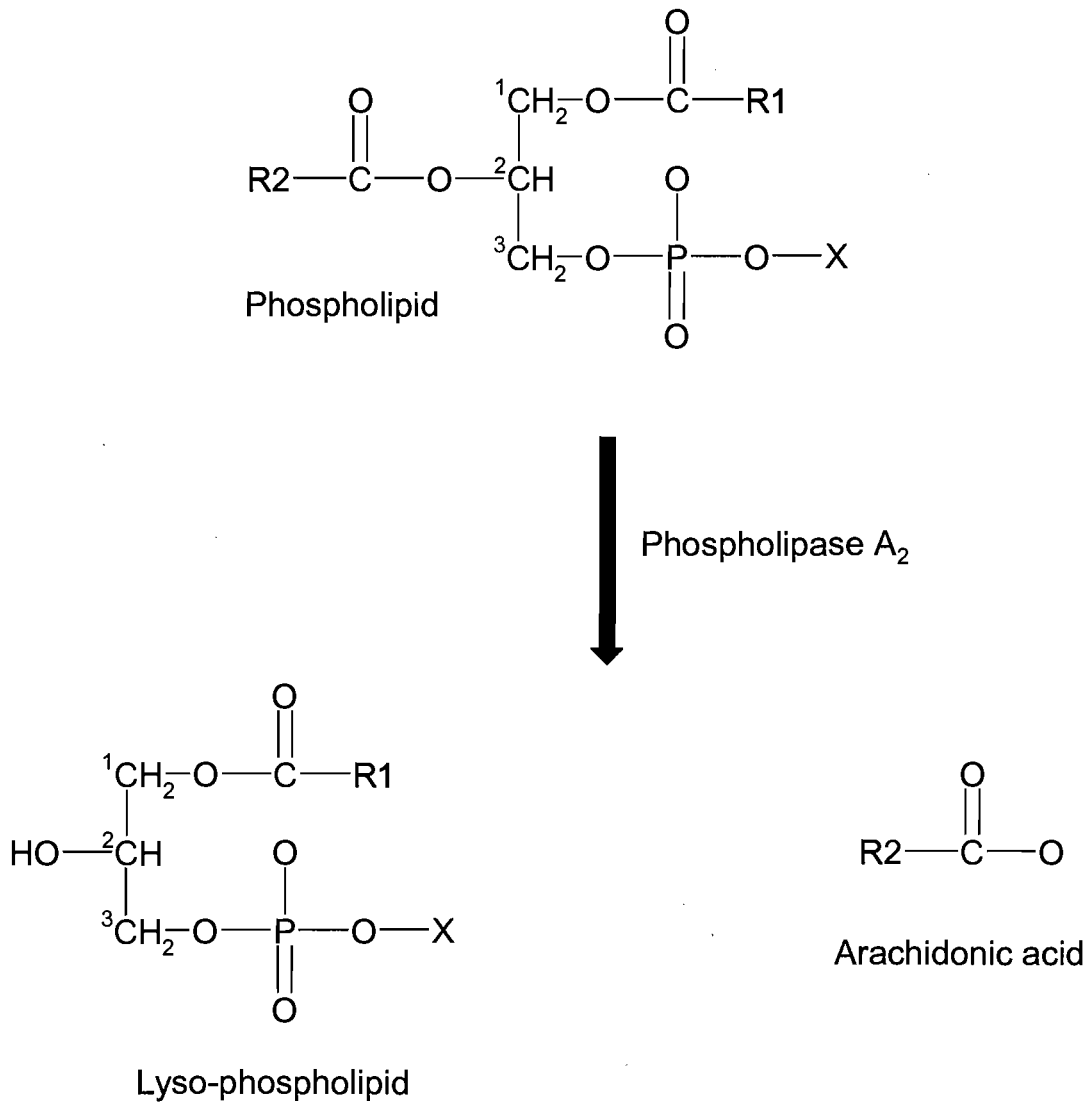


Figure 1.2 Mechanism of $sPLA_2$ showing the hydrolysis of the fatty acid ester bond in position 2 (labelled) of PC to generate lyso-PC.

thereby generating lyso-phospholipids including lyso-PC (Hack *et al.*, 1997) (see Figure 1.3b).

The increase in the amount of lyso-phospholipid, together with the presence of PE and PS, exposes PC head groups in the outer leaflet of the membrane and generates potential binding sites for CRP. Following its binding to PC, CRP is thought to bind Clq and activate the classical pathway of the complement system, as shown in Figure 1.3c (Thompson *et al.*, 1999).

1.2.4 CRP and Complement System

The complement system was first identified in the late 1800's when it was observed that antibody-mediated lysis of bacteria and red blood cells required an additional, non-specific, heat-sensitive factor present in all normal sera (Volanakis, 1982). It was proposed that this factor 'complemented' the effects of specific antibody. Since these initial observations over a century ago, a plethora of research has been carried out on the complement system. This research has identified, in comparison to antibodies, an active role for complement in the lysis of invading cells via the formation of lethal cell lesions. It is also now accepted, however, that the role of complement in humoral immunity extends much further than the killing of cells. Once activated, the complement system functions not only directly to dispose of target cells, but also indirectly to mediate a whole host of other mechanisms essential to host defence. These include the release of histamine from mast cells, chemotactic attraction of phagocytic cells, and opsonisation of cells and particles (Volanakis, 1982). It is now appreciated that the complement system is a complex system that encompasses approximately 30 serum proteins and cell surface receptors involved in a range of functions including direct cell lysis and the enhancement of B and T cell responses (Walport, 2001b).

The contributory effects of complement to host defence are mediated primarily by fragments derived from the early complement components C3 and C5 (Volanakis, 1982; Reid & Porter, 1981). Cleavage of C3 generates two fragments: C3a and C3b. C3a is an anaphylatoxin causing an increase in capillary permeability through the release of histamine from mast cells. C3b, a powerful opsonin enhancing the phagocytosis of particles to which it binds, has been highlighted as the most essential complement-derived biologically active fragment in terms of host defence against

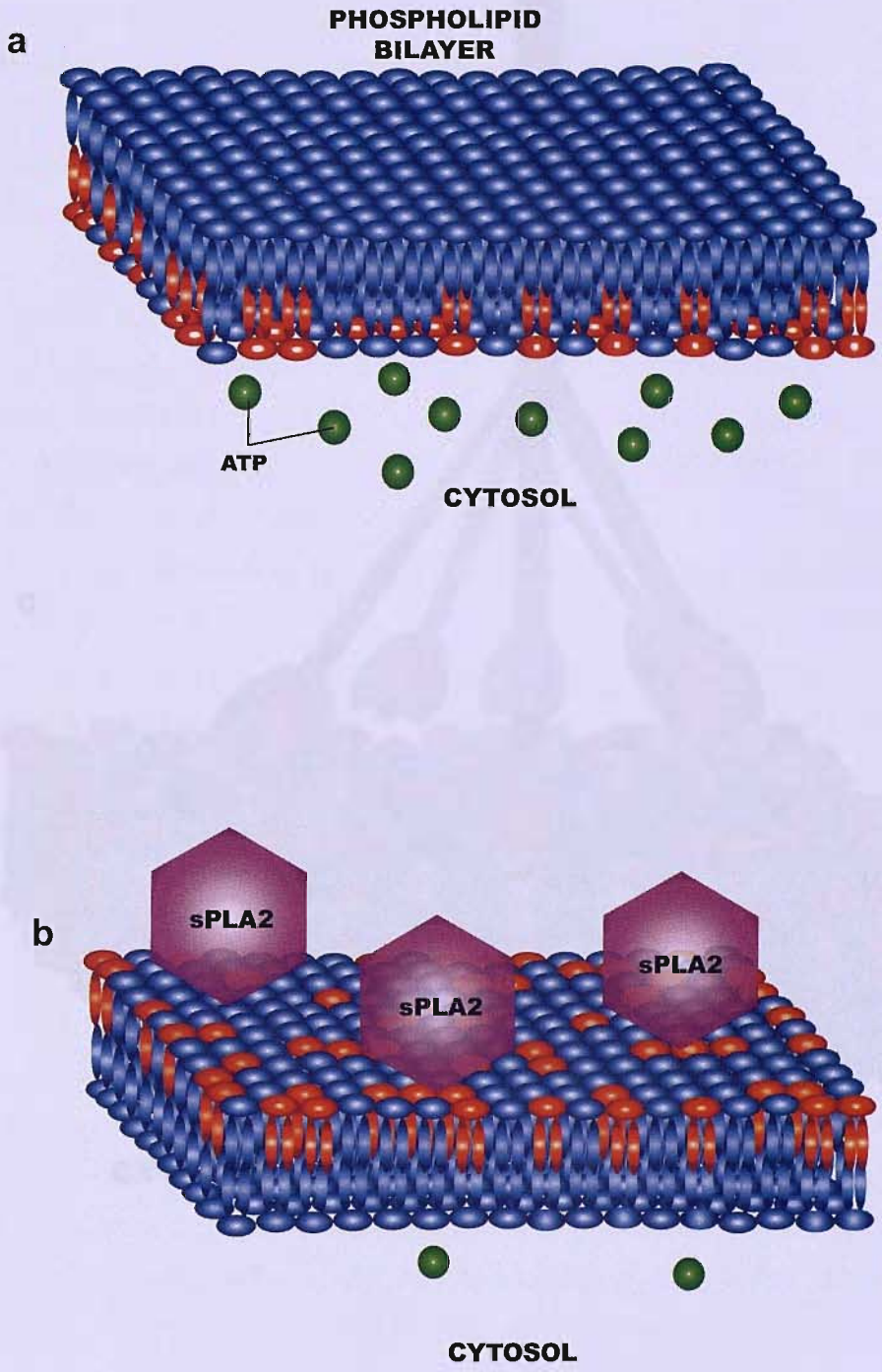


Figure 1.3 Model for the binding of CRP to biological membranes. **a** Asymmetric distribution of phospholipids among the inner and outer leaflets of the membrane (PS & PE are shown in red; sphingomyelin & PC shown in blue). **b** Low ATP levels result in 'flip-flop' of the membrane and it is now susceptible to hydrolysis by $sPLA_2$.

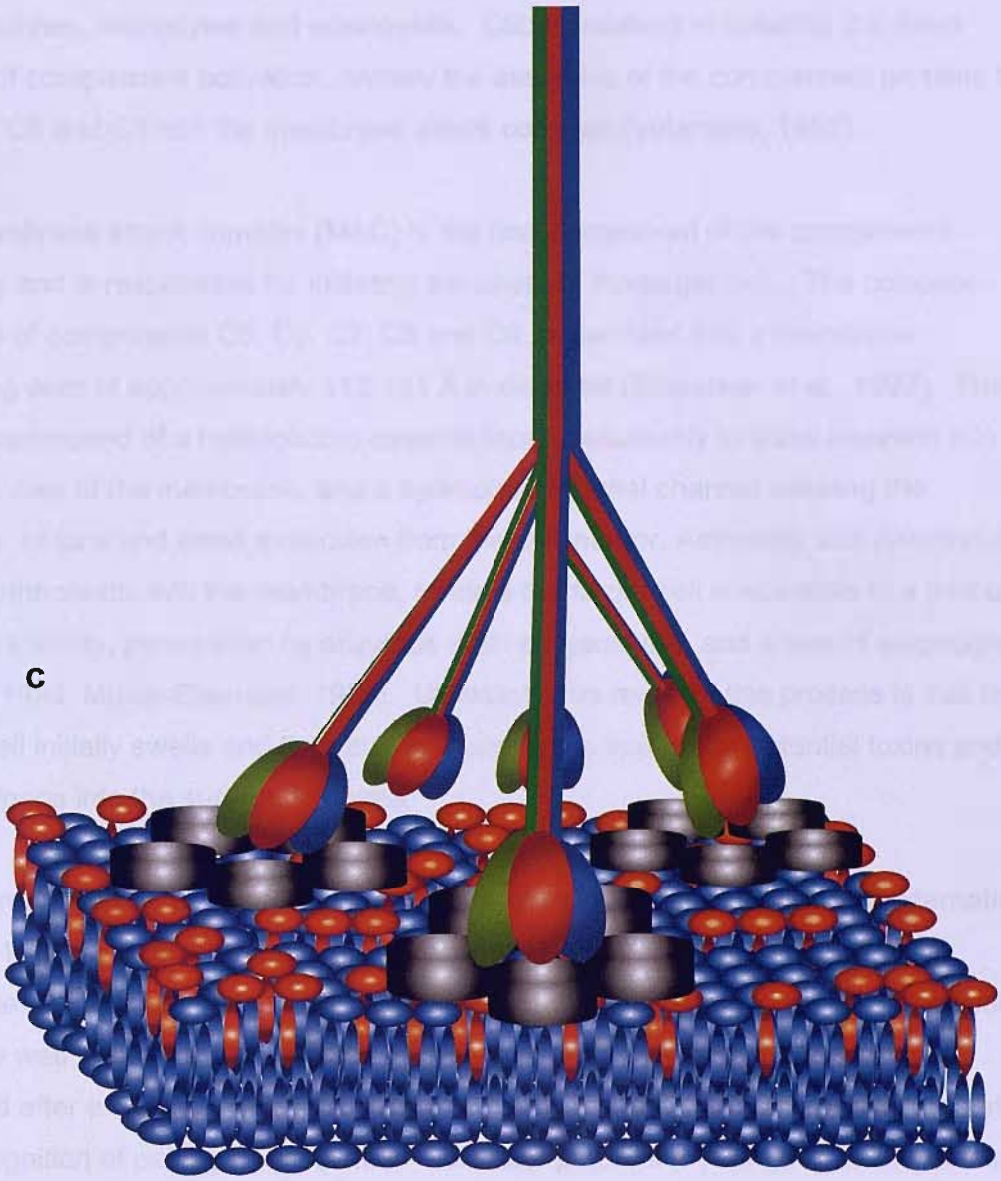


Figure 1.3 c The presence of PE and PS and the increase in the amount of lyso-phospholipid in the outer leaflet of the membrane exposes potential binding sites for CRP (shown in grey). C1q (shown in green/red/blue) can then bind to CRP and activate the classical pathway of complement.

infection. C5a, generated by enzymatic cleavage of C5, is a potent chemotactic factor for leukocytes, monocytes and eosinophils. C5b is involved in initiating the direct effects of complement activation, namely the assembly of the complement proteins C6, C5, C7, C8 and C9 into the membrane attack complex (Volanakis, 1982).

The membrane attack complex (MAC) is the final component of the complement pathway and is responsible for initiating the death of the target cell. The complex consists of components C5, C6, C7, C8 and C9, assembled into a membrane-spanning pore of approximately 113-181 Å in diameter (Biesecker *et al.*, 1993). The pore is composed of a hydrophobic external face, presumably to allow insertion into the lipid core of the membrane, and a hydrophilic internal channel allowing the 'leakage' of ions and small molecules from the cell interior. Assembly and insertion of these components into the membrane, renders the target cell susceptible to a loss of osmotic stability, penetration by enzymes such as lysozyme, and a loss of electrolytes (Esser, 1994; Muller-Eberhard, 1986). Ultimately, the result of this process is that the target cell initially swells and then bursts, causing the spillage of potential toxins and autoantigens into the surrounding area.

Complement can be activated by one of three pathways: classical, lectin or alternative (Figure 1.4). All three of these pathways converge onto a common complement component, C3; these however, all differ in their nature of recognition. The classical pathway was the first to be identified and is activated by antibody (IgG and IgM) released after a humoral response (Reid, 1983a). The lectin pathway is activated after the recognition of pathogen-associated molecular patterns (PAMPs) by lectin proteins (Carroll, 2004). One such protein is mannan-binding lectin (MBL), a member of the collectin family of proteins that also includes the classical component C1q. Members of this family all contain characteristic collagen stalk and globular regions (Carroll, 2004). The six globular head domains of MBL bind N-acetyl glucosamine and mannose residues common among microbes. The alternative pathway, in contrast to the other two pathways, is unspecific, spontaneously activating C3 in low levels. Once the activated form of C3 (C3b) is attached to the surface of a pathogen, an amplification loop consisting of the alternative pathway components is activated, thereby generating many more active C3 molecules (Fujita *et al.*, 2004).

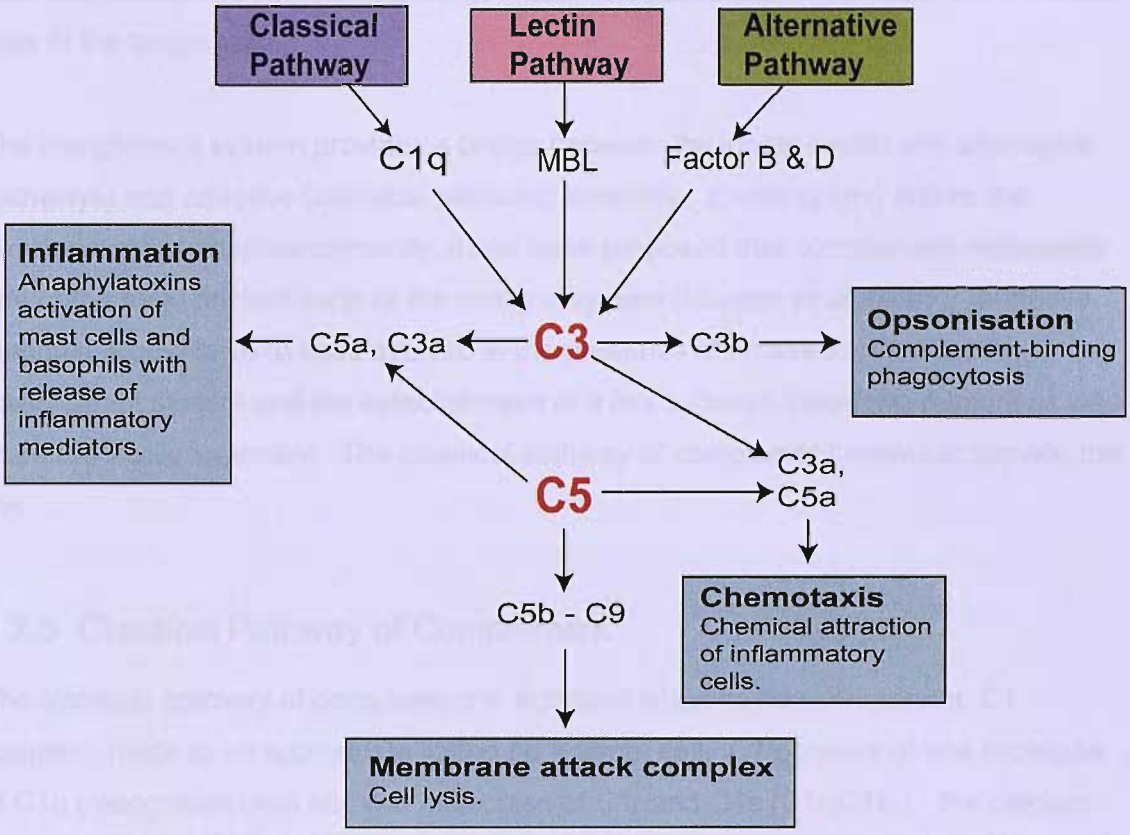


Figure 1.4 The three pathways of complement.

Following ligand recognition and subsequent activation, each one of these pathways involves a complex enzymatic cascade. The enzymes of these cascades exist and circulate as inactive pro-enzymes, or zymogens, activated following cleavage of their active site inhibitory fragments. Activation of the first enzyme in a pathway initiates a cascade whereby the pro-enzyme product of one step is the enzyme catalyst for the next step (Reid, 1981). The end result of this cascade is the formation of the MAC and lysis of the target cell.

The complement system provides a bridge between the innate (lectin and alternative pathways) and adaptive (classical pathway) immunity. Evolving long before the development of adaptive immunity, it has been proposed that complement represents one of the most ancient parts of the immune system (Morgan *et al.*, 2005). Adaptive immunity would have to have evolved in the presence of an already functioning complement system and the establishment of a link between these two responses was therefore vitally important. The classical pathway of complement seems to provide this link.

1.2.5 Classical Pathway of Complement

The classical pathway of complement is activated when its first component, C1 complex, binds to an appropriate ligand on a target cell. C1 consists of one molecule of C1q (recognition unit) and two molecules of C1r and C1s ($C1r_2C1s_2$) - the calcium-dependent catalytic unit (Cooper, 1985). Once C1q has recognised and bound ligand, a conformational change is thought to occur in the recognition unit, leading to the activation of the catalytic unit. The activated catalytic unit of C1 is able to mediate the proteolytic cleavage of C4 and C2 leading to the formation of a C3 convertase and subsequent cleavage of C3 into C3a and C3b (Sim & Tsiftsoglou, 2004).

1.2.6 C1q - Structure and Function

Human C1q has a molecular weight of 460 kDa (Reid & Thompson 1983b) and consists of 18 polypeptide chains of three different types: six A (A=223 residues), six B (B=226 residues) and six C (C=217 residues) (Kishore & Reid, 2000). Each chain consists of an N-terminal region of approximately 3-9 residues involved in the formation of A-B and C-C inter-chain disulphide bonds. This is followed by a stretch of approximately 81 residues containing repetition of the characteristic collagen sequence Gly-X-Y (X=Proline, Y=4-hydroxyproline or 5-hydroxylysine) (Reid & Porter, 1976).

This collagen-like region (CLR) gives rise to the formation of six collagen-like triple helices each containing one A, one B and one C chain (Cooper, 1985). Due to C-terminal interruptions in the Gly-X-Y sequence motif, the stalk region diverges into six arms each ending in a C-terminal globular head (GLR) of approximately 135 residues (Sellar *et al.*, 1991). Current models for the molecular architecture of C1q are based on electron microscopy images (Knobel *et al.*, 1975). These studies reveal a distinctive structure resembling a bunch of tulips, as shown in Figure 1.5. The 'stem' is formed by the CLR of the C1q molecule, whilst the six C-terminal globular heads form the 'flowers'.

In the proposed model of C1 activation, the six globular heads of C1q recognise and bind to Fc regions of IgM and IgG, causing a conformational change in its CLR and activation of the catalytic components C1r and C1s (Kishore *et al.*, 2004). Although the precise details of this mechanism remain unclear, it has been proposed that movement of the collagenous stalk regions of C1q relative to each other causes a conformational change in one or both of the C1r molecules. Once both C1r molecules are activated, cleavage of the C1s pro-enzyme proceeds and the catalytic unit becomes fully functional and able to initiate the complex cascade of complement (Arlaud *et al.*, 2001). This represents the traditional view of classical pathway activation. It is now widely accepted that the classical pathway can also be activated by various non-immunoglobulin substances binding to the gC1q domain. These include lipopolysaccharides, porins of gram negative bacteria, phospholipids, apoptotic cells, CRP and PTX-3 (Kishore *et al.*, 2004).

Although the crystal structure of complete, intact C1q has still to be solved, the globular domain of C1q has been elucidated via crystallographic methods (Gaboriaud *et al.*, 2003). The structure reveals a tight (buried surface 5490 $Å^2$) heterotrimeric assembly of subunits formed by the C-termini of chains A, B and C arranged in clockwise fashion when viewed from above (see Figure 1.6). As anticipated, the overall morphology of the assembly exhibits a globular, almost spherical structure of approximately 30 Å in diameter. When viewed from the side, the N and C termini of all three chains are located at the base of the trimer. A single calcium ion is positioned opposite the base of the trimer, well exposed to solvent and defining the upper entrance of a discontinuous channel that runs through the centre of the assembly. The overall structure of the heterotrimer consists of a 10- stranded β sandwich with jelly-roll topology, consisting of two antiparallel stranded β-sheets (Gaboriaud *et al.*, 2003).

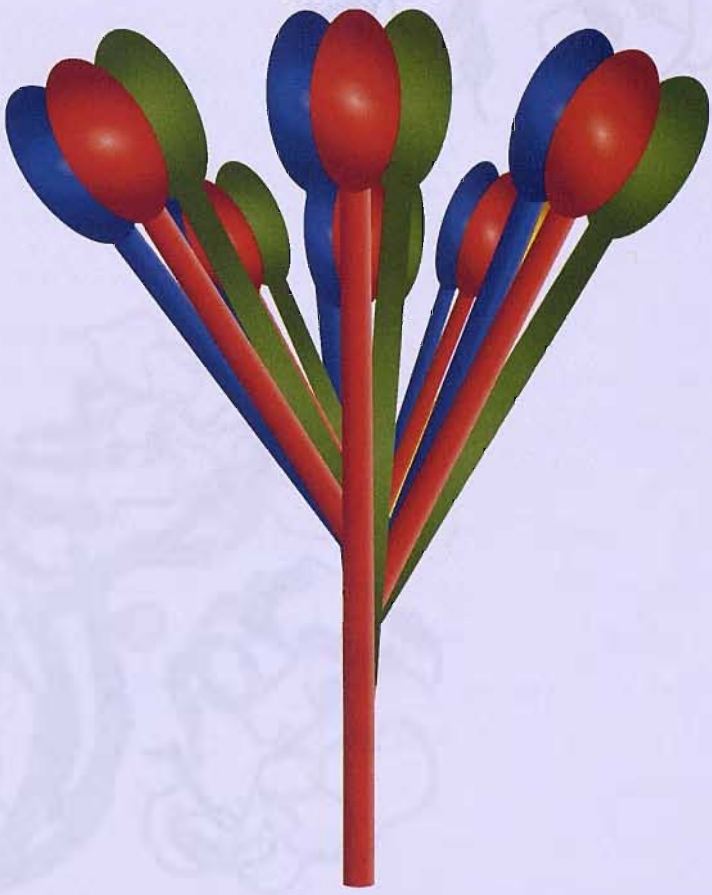


Figure 1.5 The structure of C1q resembling a bunch of tulips. The 'stalk' is composed of the collagenous region (CLR); the 'flowers' represent the globular head regions (GLR). Individual chains are coloured separately (A=red, B=blue, C=green).

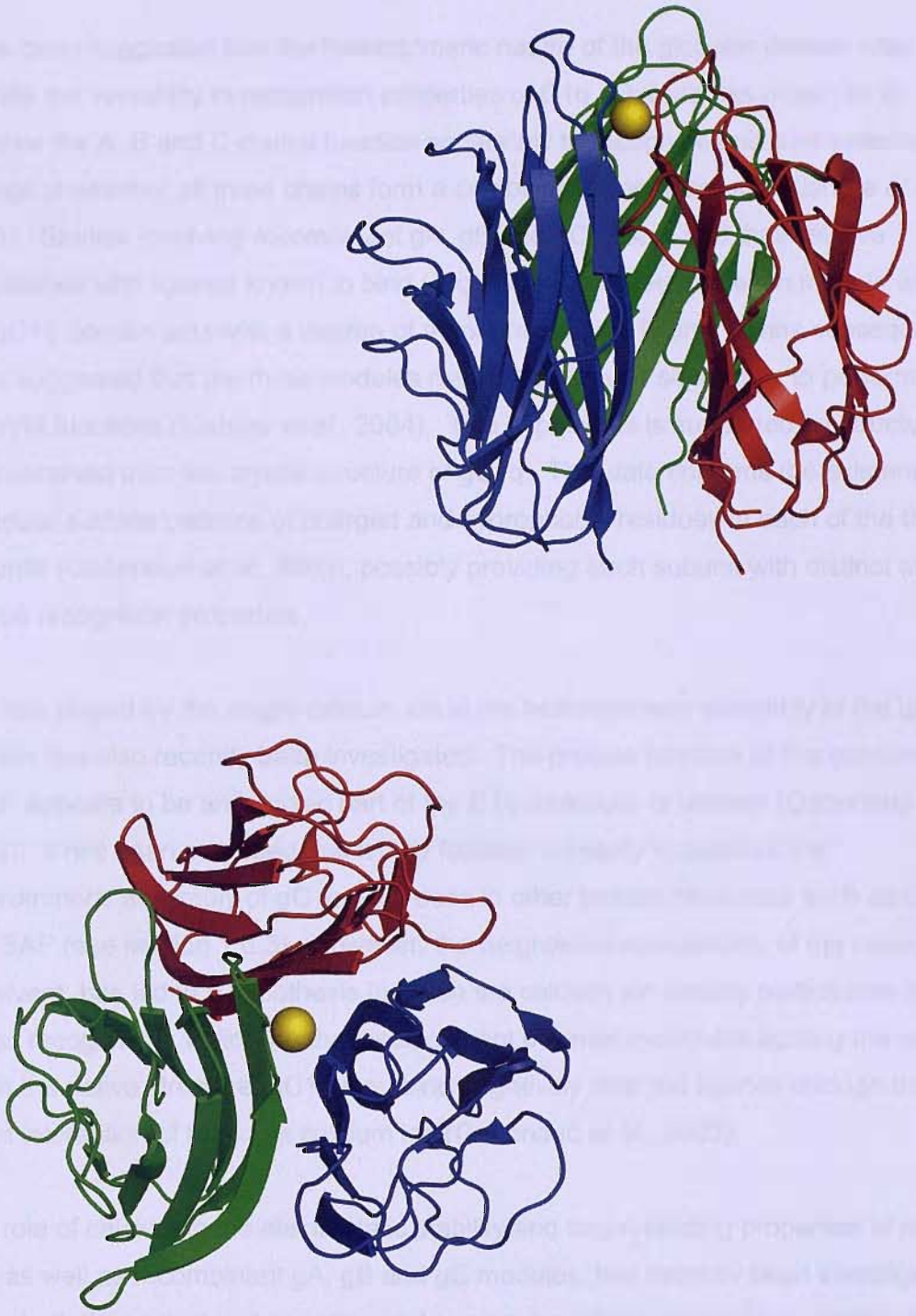


Figure 1.6 The X-ray crystal structure of the globular domain of C1q comprising the C-termini of chains A (red), B (blue), C (green) and a single calcium ion (yellow sphere). **Top right -** side view of the globular domain. **Bottom left** - view down the centre of the globular domain (Gaboriaud *et al.*, 2003).

It has been suggested that the heterotrimeric nature of the globular domain may provide the versatility in recognition properties of C1q. Debate has grown as to whether the A, B and C chains function separately to recognise and bind potential ligands or whether all three chains form a combined recognition unit (Kishore *et al.*, 2003). Studies involving recombinant gA, gB and gC chains and their relative specificities with ligands known to bind C1q, have illustrated that each module within the gC1q domain acts with a degree of autonomy to bind ligand. It has subsequently been suggested that the three modules may have evolved separately to perform different functions (Kishore *et al.*, 2004). This hypothesis is supported by structural data obtained from the crystal structure of gC1q. This data confirms the existence of particular surface patterns of charged and hydrophobic residues in each of the three subunits (Gaboriaud *et al.*, 2003), possibly providing each subunit with distinct and unique recognition properties.

The role played by the single calcium ion in the heterotrimeric assembly of the globular domain has also recently been investigated. The precise function of this calcium, which appears to be an intrinsic part of the C1q molecule, is unclear (Gaboriaud *et al.*, 2003). It has been proposed that it may function primarily to stabilise the heterotrimeric assembly of gC1q, as it does in other protein molecules such as CRP and SAP (see section 1.6.3). However, the heightened accessibility of the calcium ion to solvent, has led to a hypothesis in which the calcium ion directly participates in ligand recognition. Following the displacement of water molecules ligating the calcium ion in the native structure, gC1q may bind negatively charged ligands through the direct interaction of its single calcium ion (Gaboriaud *et al.*, 2003).

The role of calcium in the electrostatic stability and target-binding properties of native C1q as well as recombinant gA, gB and gC modules, has recently been investigated using both theoretical and experimental approaches (Roumenina *et al.*, 2005). The study reports that calcium primarily influences the target recognition properties of C1q toward IgG, IgM, CRP and PTX3. The study also proposes a model through which calcium may achieve this. In normal serum, calcium is an intrinsic part of the C1q molecule, its presence directing the electrical moment toward the top of the gC1q heterotrimer. Calcium binding to C1q facilitates recognition of negatively charged molecules in the initial phase of C1q-ligand interaction. However, once bound, the negative field of the ligand accelerates the removal of calcium from C1q. This generates mechanical stress in the C1q molecule and a structural change in the CLR region, leading to C1r activation (Roumenina *et al.*, 2005). The model concurs not only

with previous hypotheses but also with structural data currently available on the binding sites of CRP (see section 1.6.4.2).

1.2.7 CRP-Mediated Activation of Complement

CRP-mediated activation of complement was first demonstrated by studies involving the addition of *pneumoccocal* C-polysaccharide (PnC) to acute phase sera containing CRP (Kaplan & Volanakis, 1974). The addition resulted in the consumption of complement components C1, C4 and C2 as well as C3-9, suggesting that CRP-mediated activation of complement proceeds via the classical pathway. Also demonstrated was the formation of an effective C3-convertase and the binding of C3 and C1q to CRP-PnC complexes reacted with normal serum. These findings further highlighting the ability of CRP to activate the classical pathway of complement and suggested possible components to which CRP may bind. As previous studies had led to the identification of PC residues in PnC as the major reacting site for CRP (Brundish & Baddiley, 1968; Volanakis & Kaplan, 1971), the ability of CRP-phosphatidylcholine and CRP-sphingomyelin complexes were investigated for their ability to initiate complement activation. These complexes efficiently activated the classical pathway of complement (Kaplan & Volanakis, 1974), suggesting that in addition to mediating the elimination of microbial cells, CRP may also have a role in mediating the clearance of host cells.

The same group were responsible for identifying C1q as the complement component to which CRP binds and subsequently activates complement (Volanakis & Kaplan, 1974). When CRP, complexed with either PnC or PC-containing phospholipids, was added to guinea-pig serum, activation of complement failed. This effect could however, be overcome on addition of human C1q, suggesting not only that C1q is the specific ligand of the classical pathway for CRP, but also that there is an incompatibility between guinea-pig C1q and human CRP.

These studies were the first to identify the nature of complement activation by CRP and suggested that through the binding of previously identified PC-containing microbes including *pneumoccoci* (Brundish & Baddiley, 1968) and PC residues of specific eukaryotic cell membrane phospholipids (lecithin and sphingomyelin), may activate the classical pathway of complement via C1q.

Recent studies on CRP-mediated activation of complement have focused primarily on the structure of the complex and in particular, the location of the binding site on C1q. Despite the identification of the GLR to be the site at which C1q binds IgG (Marques, 1993), initial studies investigating the CRP binding site proposed that CRP bound to C1q at the CLR (Jiang *et al.*, 1991; Jiang *et al.*, 1992). Cross-inhibition studies involving the binding of CRP and IgG to C1q, suggested that these two proteins bind at different locations on the C1q molecule (Jiang *et al.*, 1991). More specifically, an epitope consisting of residues 81-97 positioned at the connecting strands of the collagenous portion just below the GLR, was identified in this study. Further to this work, the same research group published a second paper identifying two regions located exclusively on the A chain of C1q between residues 14-26 and 76-92 (Jiang *et al.*, 1992).

These studies were performed before the elucidation of the crystal structure of CRP (Shrive *et al.*, 1996) and did not benefit from the subsequent structural data presented. Analysis of this data led to a dramatic shift in opinion, suggesting that in complete contrast to previous hypotheses, the CRP-binding site is located on the globular heads of C1q (Agrawal *et al.*, 2001). As the PC binding sites are located on one face of the pentamer (B face) rather than on the perimeter of the ring, it is proposed that the only accessible region for C1q on a CRP pentamer would be on the opposing face (A face). This would be physically inaccessible to the neck or tail region of a C1q molecule (Agrawal *et al.*, 2001).

Recent support to this hypothesis has been provided by two different studies. Pentraxin 3 (PTX3), containing a CRP homology domain, has recently been shown to bind gC1q as well as recombinant ghA, ghB and ghC modules (Nauta *et al.*, 2003). The crystal structure of the GLR of C1q has also provided supporting evidence for this hypothesis (Gaboriaud *et al.*, 2003). Modelling studies on the two proteins have uncovered striking shape and electrostatic complementarities between the top of the GLR of C1q and a prominent cleft situated on the exposed face of a CRP pentamer (see section 1.6.4.2).

1.2.8 CRP and Removal of Apoptotic Cells

Recent studies have shown that as well as being involved in the removal of damaged cells, CRP may also have a pivotal role in ensuring the safe removal of cells undergoing apoptosis (Gershov *et al.*, 2000). Apoptosis, or programmed cell death, is

an essential process in the maintenance of normal tissue homeostasis and is thought to be a tightly regulated mode of cell death that occurs without inflammation. The process is characterised by the fragmentation of the nucleus and cellular constituents into membrane-enclosed apoptotic bodies or blebs.

Most of the experimental evidence seems to support a mechanism in which apoptotic cells are phagocytosed by macrophages and provoke the release of anti-inflammatory cytokines such as TGF-β (Voll *et al.*, 1997). This mechanism serves to remove apoptotic cells from the surrounding tissue and in doing so protects against local damage resulting from uncontrolled leakage of noxious contents. In contrast to this, necrotic cells appear to induce the release of proinflammatory cytokines and may promote the maturation of dendritic cells, leading to antigen presentation and an immune response (Voll *et al.*, 1997). If the process of apoptosis is disturbed, apoptotic cells can also proceed to a state of secondary necrosis, accompanied by swelling and eventual bursting of the cell. Since apoptotic cell blebs have been identified as a potential source of autoantigens such as DNA, RNA and chromatin (Casciola-Rosen *et al.*, 1994; Cline & Radic, 2004) it is essential that the process of apoptosis proceeds undisturbed. CRP has been implicated as an important mediating factor in this process.

The mechanism mediating the binding of CRP to damaged cells - flip-flop exchange of phospholipids from the inner to the outer membrane, has been shown to occur in cells undergoing apoptosis (Martin *et al.*, 1995). In light of this, and the observation that CRP can bind to small nuclear ribonucleoprotiens (Du Clos, 1989) and small ribonucleoprotein particles (Pepys *et al.*, 1994), a study was carried out to investigate a potential role for CRP in the clearance of apoptotic cells and apoptotic cell debris. The study demonstrated the binding of CRP to the surface of apoptotic cells in a calcium-dependant manner, and the subsequent activation of the classical pathway of complement (Gershov *et al.*, 2000). In accordance with previous observations (Berman *et al.*, 1986), the study revealed that CRP-mediated activation of complement was restricted to the formation of C3-convertase and assembly of the terminal complement components was reduced. This effect, in addition with an observed increase in phagocytosis of apoptotic cells by macrophages and a sustained production of TGF-β, suggests that CRP and the classical pathway of complement work in accord to promote the clearance of apoptotic cells in an anti-inflammatory manner.

Perhaps the most important factor mediating CRP-induced clearance of apoptotic cells is the interaction of the protein with phagocytes. Evidence supporting a direct interaction of CRP with phagocytic cells has been available for many years (Mortensen *et al.*, 1976). Despite this, only recently have the receptors for CRP been identified (Marnell *et al.*, 1995): the study reported a high affinity interaction of CRP with the IgG receptor, FcγRI (K_D = 0.81 x 10^{-9} M) (Bodman-Smith *et al.*, 2002). As cells lacking this receptor were still able to bind CRP, and up-regulation of the receptor did not significantly enhance CRP binding however, it was apparent that FcγRI does not exclusively mediate CRP binding to phagocytic cells. In a subsequent study to try and define the binding of CRP to phagocytic cells further (Bhardwaj *et al.*, 1999), the low affinity IgG receptor FcγRIIa was identified as the second receptor for CRP on phagocytic cells. The proposed hypothesis that CRP binds to phagocytic cells through FcγRs was further supported by studies in which mouse models lacking all three FcγRs (FcγRI, FcγRII and FcγRIII) failed to bind CRP (Stein *et al.*, 2000). Following on from these studies, the Du Clos group were able to show a direct relationship between these observations and a possible functional role for CRP (Mold *et al.*, 2002): CRP was shown to opsonise apoptotic cells for phagocytosis via FcγRs.

1.3 Possible Physiological Role of SAP

Despite its structure and association with the pathology of amyloidosis being relatively well defined, the physiological role of SAP remains unclear. Many ligands have been identified for SAP, including DNA, chromatin, C1q, and PE, but the physiological relevance of these studies has not been determined. The plasma concentration of human SAP is approximately 30 mg/L and remains constant during all forms of infection and tissue injury (Pepys *et al.*, 1997). Therefore, unlike CRP, SAP is not an acute phase protein in humans. As with CRP however, the absence of any known polymorphism or deficiency of SAP in humans is suggestive of an important physiological role.

1.3.1 SAP binding to DNA and Chromatin

SAP has been shown to bind, in a calcium-dependent manner and at physiological pH and ionic strength, to both DNA and native long chromatin (Pepys & Butler, 1987). This interaction appears to be specific to SAP as no other serum protein, including

CRP, bound or was eluted in significant quantities from immobilised DNA beads (Pepys & Butler, 1987). Subsequently, SAP has been classified as the single protein in whole serum to bind DNA.

Further to this work it was found that the interaction of SAP with long chromatin is sufficient to mediate the displacement of H1 histones (Butler *et al.*, 1990). Long chromatin is normally highly insoluble in buffers of physiological strength. Histones, responsible for compression of DNA and enabling its packaging into the cell nucleus, are however displaced by SAP, thereby promoting the solubilisation of chromatin (Butler *et al.*, 1990). Investigating the interaction of SAP with chromatin further, Hicks *et al.* (1992) reported the direct binding of SAP to histones and the ability of these complexes to activate the classical pathway of complement.

SAP also binds to extracellular chromatin *in vivo*, and in particular, studies have reported the association of SAP with unusual globular dermal deposits of nuclear material in skin biopsies from patients with the autoimmune disease systemic lupus erythematosus (SLE) (Breathnach *et al.*, 1989). SLE, a systemic autoimmune disease, is characterised by the presence of auto-antibodies that recognise nuclear antigens such as DNA, RNA and chromatin. It is a chronic inflammatory autoimmune disorder causing skin rashes, arthritis, anaemia, seizures or psychiatric illness, and affects a variety of different organs including the heart, kidney and lungs.

It appears that the characteristic binding of chromatin and DNA to SAP may be important in defining its as yet unknown physiological role (Birkerstaff *et al.*, 1999; Sorensen *et al.*, 2000). A plethora of results suggest that the role of SAP may be in the handling of nuclear material such as chromatin and DNA from dead or dying cells, thereby preventing formation of anti-DNA antibodies such as those observed in SLE. This hypothesis has been supported further by experiments involving SAP knockout mice. Both the development and breeding capabilities of these were not affected. They did, however, accumulate pathogenic antinuclear antibodies to DNA, chromatin and histones and also suffered from a high incidence of severe proliferative glomerulonephritis, a phenotype reminiscent of human SLE (Bickerstaff *et al.*, 1999). Despite the apparent ability of SAP to bind to C1q and activate complement *in vitro* (Ying *et al.*, 1992), the clearance of nuclear material by SAP appears to occur through a process independent of the classical pathway of complement as the rate of SAP-mediated chromatin degradation was unaffected in C1q knockout mice. In contrast to acting as an opsonin for chromatin uptake, it appears that SAP acts powerfully to

inhibit and control inappropriate, immunogenic processing of this ubiquitous autoantigen (Bickerstaff *et al.*, 1999).

Sorensen *et al.* (2000) provided the first reported existence of SAP-DNA complexes in human sera. In agreement with Pepys and Butler (1987), the study showed the interaction of SAP with DNA to be calcium-dependent, as complexes were not detected in EDTA plasma. Also investigated by this study were the relative concentrations of SAP-DNA complexes in the serum of healthy and SLE patients. It was reported that the concentration of these complexes in SLE sera was significantly lower than those of normal sera. These findings provide further support for the hypothesis that the physiological role of SAP is to prevent the development of antinuclear autoimmunity.

1.3.2 SAP and Complement

Bristow and Boakle (1986) were the first to provide evidence for the binding of SAP to the complement component C1q. This was followed by studies showing the activation of complement by supraphysiological amounts of aggregated SAP (200 µg/ml) (Butler *et al.*, 1990) and a full description of the observed binding of SAP to the collagen region of C1q (Ying *et al.*, 1993). The ability of aggregated SAP to activate complement *in vitro* is a well established phenomenon, but whether SAP-mediated activation of complement occurs *in vivo* remains uncertain (Pepys *et al.*, 1997). Since recent studies have now almost certainly established that CRP binds to the GLR of C1q, it seems unlikely that the two pentraxins would bind at alternative locations on the C1q molecule and therefore puts the findings of Ying *et al.* (1993) in considerable doubt.

The activation of complement by aggregated SAP is inhibited by the addition of DNA and chromatin (Butler *et al.*, 1990) and it has been suggested that this may be the result of the dissipation of these aggregates into single, ligand-bound SAP pentamers devoid of the ability to bind and activate complement. It has been hypothesised that this characteristic may be important to the role of SAP in the pathology of amyloidosis (Pepys *et al.*, 1997). SAP aggregates are found, according to electron microscopy studies, deposited along the axis of an amyloid fibril (Skinner *et al.*, 1982). SAP aggregated in this manner would then presumably be of the type competent in activating complement, as supported by various studies reporting the presence of complement components in amyloid deposits (Akiyama *et al.*, 1991; Tacnet-Delorme *et*

al., 2001). Complement activation in the vicinity of the amyloid deposit could therefore have a harmful effect on the surrounding cells, thereby contributing to the clinical effects of amyloidosis.

1.4 Associated Pathologies of CRP

The traditional use of CRP has been as a clinical acute phase marker of tissue injury, infection and inflammation. A high CRP value is a strong indicator of tissue damaging disease. In spite of this, due to the non-specific nature of the acute phase response, CRP levels can only be used for diagnostic purposes in conjunction with other laboratory and clinical information.

The recent development of high sensitivity assays has however, shown a predictive relationship between slightly elevated levels of CRP and the risk of developing future cardiovascular disease. Studies have been carried out to try and assess the association between baseline values of circulating markers of inflammation and future risk of coronary heart disease. They have shown that even in asymptomatic people, individuals with baseline CRP values in the top third of the distribution (mean 2.4 mg/L), have twice the future risk of a coronary event than those with values in the bottom third (mean 1.0 mg/L) (Danesh *et al.*, 2004). Similar relationships have also been shown to exist with other atherothrombotic events including stroke and peripheral vascular disease (Pepys & Berger, 2001).

Further to this work and further supporting a pathogenetic role for CRP in cardiovascular disease, are studies that have shown a possible specific association of the protein with cardiovascular biology and pathology. First, CRP has been shown to bind selectively to LDL, especially oxidised LDL (Chang *et al.*, 2002), and is actually found deposited in most atherosclerotic plaques (Zhang, 1999). Second, CRP is co-deposited with activated complement in all acute myocardial infarction lesions (Lagrand *et al.*, 1997). And human CRP and complement increase final myocardial infarction size in experimental models (Griselli *et al.*, 1999). It has therefore been suggested that CRP may have an important role in the pathogenesis of atherosclerosis and myocardial infarction.

1.4.1 Atherosclerosis

Atherosclerosis is a disease of the large arteries and has been characterised as the primary cause of heart disease and stroke. In Western societies, it is the underlying cause of approximately 50% of all deaths. Epidemiological studies have identified a number of genetic and environmental risk factors predisposing or contributing to the development of atherosclerosis in individuals (Lusis, 2000). Defining the precise nature of contributory molecular and cellular interactions to the aetiology of the disease has, however, proved more difficult due to the complex nature of atherosclerotic aetiology.

It was originally hypothesised that atherosclerosis was simply a process whereby fatty residues get deposited and build up on passive artery walls. Continued growth of the deposit or plaque would lead to invasion into the surrounding artery and obstruction of blood flow to the intended tissue. Since many heart attacks occur without warning and relatively few of the deposits expand so as to actually shrink the bloodstream to a pinpoint, it is now widely accepted that the mechanisms underlying the pathology of atherosclerosis are far more complex. Recent research has suggested a key role for inflammation, a process that is common to all phases of atherosclerosis, from plaque formation and growth to rupture and thrombosis (Epstein *et al.*, 1999). Inflammation- a process, which under normal circumstances of microbial infection and tissue injury is beneficial, appears to have harmful consequences in the case of atherosclerosis (Libby *et al.*, 2002).

1.4.2 Atherosclerosis and Inflammation

Recent theory suggests that atherosclerosis is an inflammatory disease: the characteristic lesions which define the pathology of atherosclerosis represent a series of highly specific cellular and molecular responses that together, can best be described as an inflammatory disease. (Ross, 1993).

Perhaps one of the most important molecules in terms of understanding the initiation of arterial inflammation and the pathology of atherosclerosis is low-density lipoprotein (LDL). Synthesised by the liver and intestines, LDL is responsible for the transport of cholesterol to target tissues, where it is used to repair membranes and produce steroids. As shown in Figure 1.7, a molecule of LDL consists of a core of

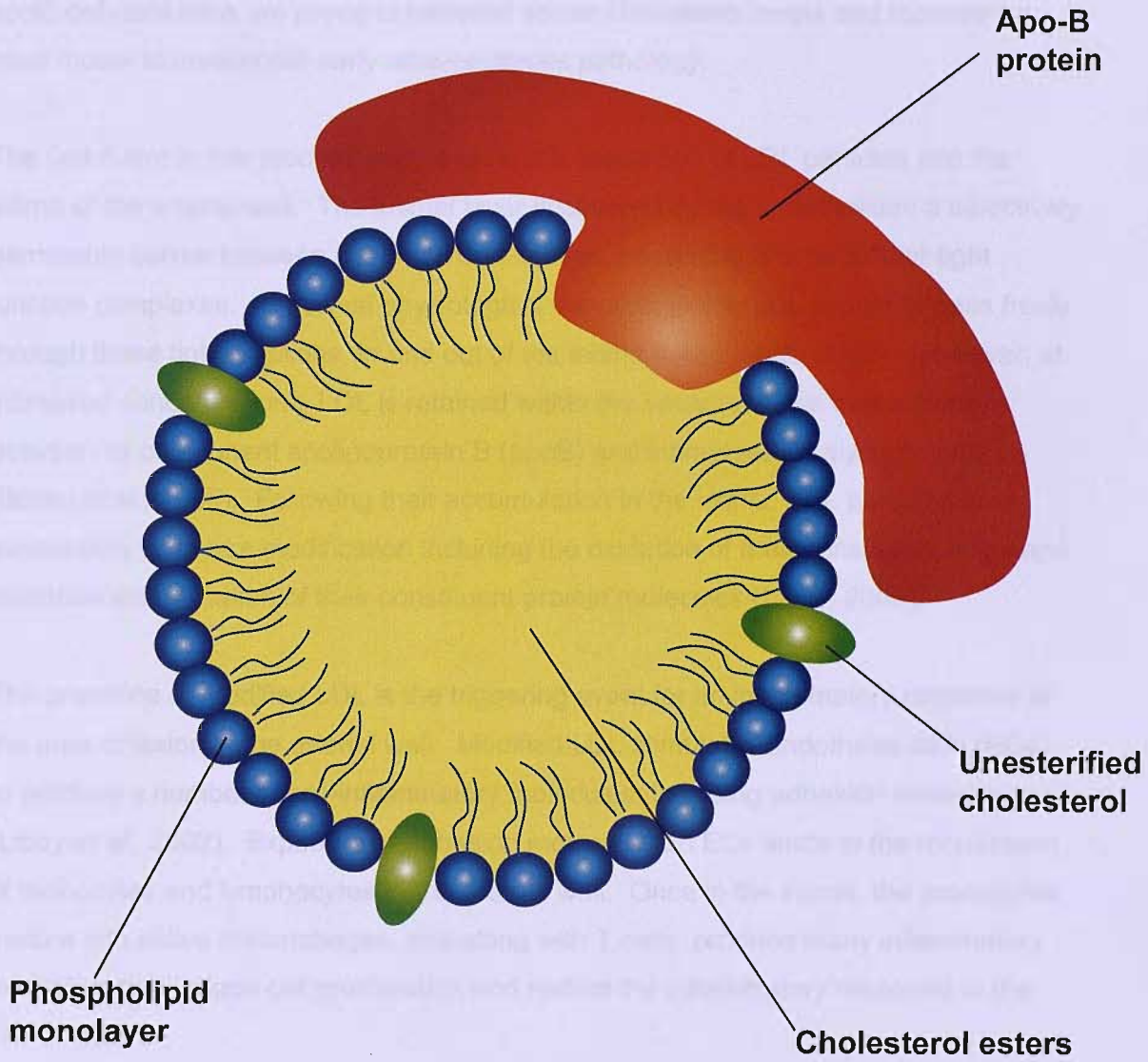


Figure 1.7 LDL molecule.

triacylglycerol and cholesterol esters surrounded by a monolayer of phospholipids interspersed with cholesterol and apolipoprotein B.

Studies investigating early lipid accumulation in apoE-deficient mice using electron microscopy have enabled the postulation of a model for the processes underlying atherosclerotic pathology (Tamminen *et al.*, 1999). ApoE is a lipoprotein present in very low-density lipoproteins (VLDL). VLDL, in contrast to LDL, is responsible for the removal of excess cholesterol from the blood and its transport to the liver. Hence, apoE-deficient mice are prone to elevated serum cholesterol levels and represent an ideal model to investigate early atherosclerotic pathology.

The first event in this process seems to be the deposition of LDL particles into the intima of the arterial wall. The intimal layer is covered by the endothelium: a selectively permeable barrier between the blood and tissues, consisting of intercellular tight junction complexes. At normal physiological concentrations, LDL is able to pass freely through these tight junctions, in and out of the intima (Libby *et al.*, 2002). However, at increased concentrations, LDL is retained within the vessel wall via interactions between its component apolipoprotein B (apoB) and intimal matrix glycoproteins (Boren *et al.*, 1998). Following their accumulation in the intima, LDL particles are susceptible to severe modification including the oxidation of their constituent lipids and oxidation and glycation of their constituent protein molecules (Lusis, 2000).

The presence of modified LDL is the triggering event for an inflammatory response at the area of lesion in the arterial wall. Modified LDL stimulates endothelial cells (ECs) to produce a number of pro-inflammatory molecules, including adhesion molecules (Libby *et al.*, 2002). Exposure of adhesion molecules on ECs leads to the recruitment of monocytes and lymphocytes to the arterial wall. Once in the intima, the monocytes mature into active macrophages, and along with T cells, produce many inflammatory mediators that induce cell proliferation and restrict the inflammatory response to the site of lesion.

In addition, macrophages also express on their surfaces scavenger receptors through which they are able to ingest modified LDLs. This is perhaps the most important process in the formation of the earliest form of the atherosclerotic plaque. Extensive modification of LDL, presumably by reactive oxygen species produced by endothelial cells and macrophages of the atherosclerotic lesion, prime the LDL for uptake by

macrophages. Macrophages engorged with LDL are frothy-looking and fat-laden and are thus called 'foam cells' (Lusis, 2000). These foam cells, along with constitutive T cells, form the 'fatty streak': the earliest form of the atherosclerotic plaque (Figure 1.8).

As the plaque progresses, inflammatory molecules such as cytokines and growth factors, promote further growth of the plaque and the formation of a fibrous cap over the lipid core. The core is formed by the migration of smooth muscle cells from the media to the top of the intima, where they multiply and produce a tough, fibrous, protective matrix. At this stage the plaque is relatively stable, surrounded by the fibrous cap which walls off the lipid core and its potentially harmful contents from the blood.

Plaques that are most vulnerable are those that possess a thinned cap, large lipid pool and many macrophages (Lusis, 2000). The process driving their vulnerability is inflammation. The strength of the cap is largely dependent on collagen fibres produced by smooth muscle cells. When an event triggers inflammation, for example infection and the subsequent induction of acute phase proteins, two processes occur that can compromise the strength of the cap and plaque structure. First, inflammatory mediators can stimulate macrophages to secrete collagenases and other proteases that degrade the fibrous cap. Second, they can also inhibit smooth muscle cells from producing collagen for the repair and maintenance of the cap (Lusis, 2000).

With the protective cap now weakened, it can be penetrated by blood that seeps through to the lipid core where it is able to react with coagulants such as macrophage-produced tissue factor. This can result in thrombosis and ultimately myocardial infarction or stroke (Epstein *et al.*, 1999).

1.4.3 CRP and Atherosclerosis

Studies showing the selective binding of CRP to low density lipoproteins (LDL), and in particular oxidised and partly degraded LDL as found in atheromatous plaques, were first to suggest a possible relationship between CRP and atherosclerosis (De Beer *et al.*, 1982). This finding has been supported by more recent observations involving CRP-specific immunofluorescence techniques in patient autopsies (Zhang *et al.*, 1999). In

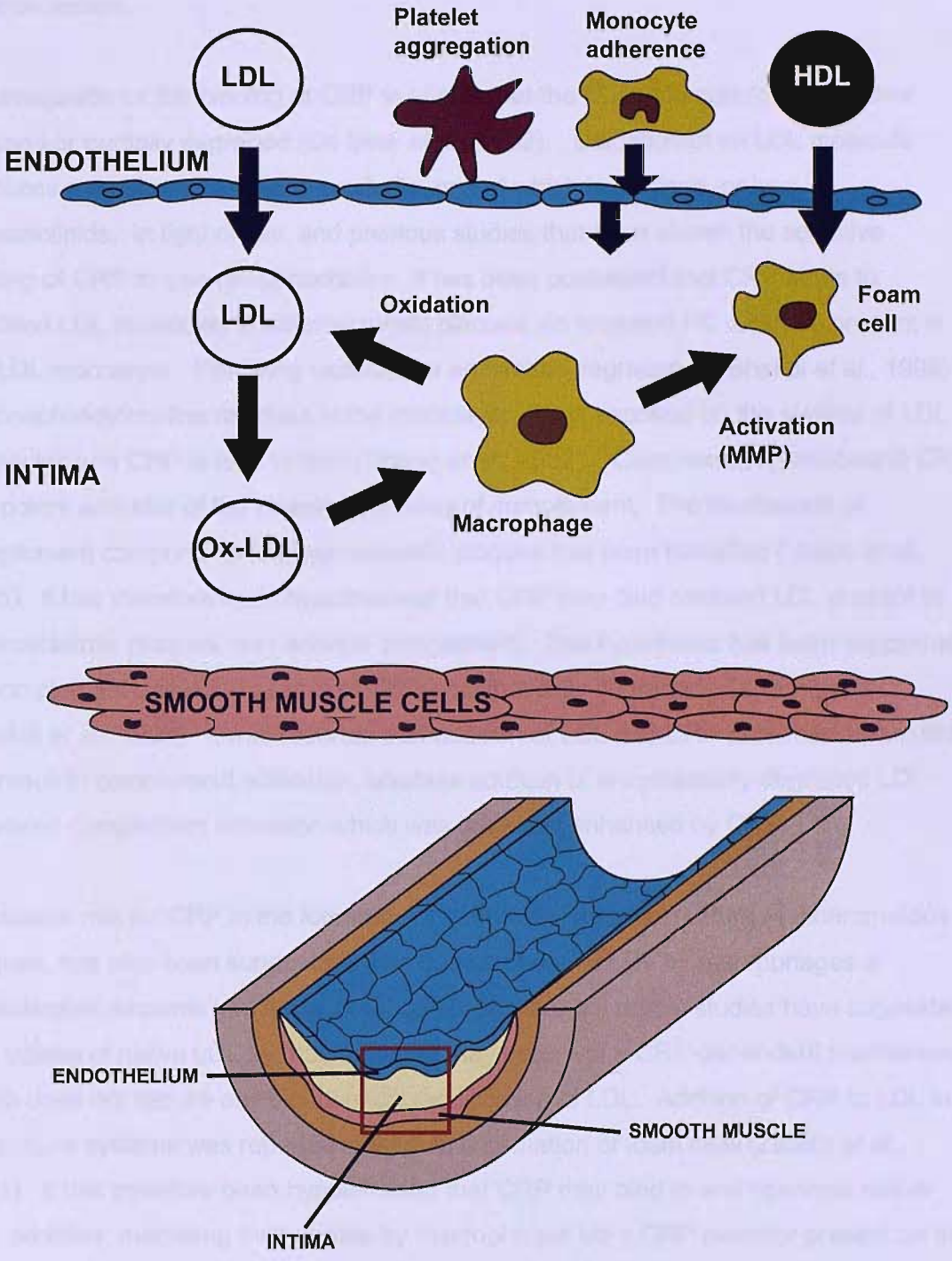


Figure 1.8 Proposed model for the formation of the 'fatty streak' in atherosclerosis (Lusis, 2000).

these studies, CRP was found deposited in atherosclerotic plaques and its distribution strongly correlated with the solubility of the intimal lipids at the various different stages of atherosclerosis.

A prerequisite for the binding of CRP to LDL is that the LDL molecule must be either oxidised or partially degraded (De Beer *et al.*, 1982). Oxidation of an LDL molecule produces a number of oxidative products, one of which is oxidised, or lyso-phospholipids. In light of this, and previous studies that have shown the selective binding of CRP to lyso-phosphocholine, it has been postulated that CRP binds to modified LDL molecules in atherosclerotic plaques via exposed PC residues present in the LDL monolayer. Following oxidation or enzymatic degradation (Bhakdi *et al.*, 1999) of phosphatidylcholine residues in the monolayer, PC is exposed on the surface of LDL molecules and CRP is able to bind (Chang *et al.*, 2002). Complexed, ligand-bound CRP is a potent activator of the classical pathway of complement. The localisation of complement components to atherosclerotic plaques has been identified (Vlaicu *et al.*, 1985). It has therefore been hypothesised that CRP may bind oxidised LDL present in atherosclerotic plaques, and activate complement. This hypothesis has been supported by complement activation assays of CRP-enzymatically degraded LDL complexes (Bhakdi *et al.*, 1999). It was reported that addition of LDL and CRP to human serum did not result in complement activation, whereas addition of enzymatically-degraded LDL provoked complement activation which was markedly enhanced by CRP.

A possible role for CRP in the formation of foam cells, a typical feature of atheromatous plaques, has also been suggested. The uptake of native LDL by macrophages in considerable amounts has never been demonstrated, but recent studies have suggested that uptake of native LDL by macrophages may occur via a CRP-dependent mechanism which does not require any biochemical modification of LDL. Addition of CRP to LDL in cell culture systems was reported to stimulate formation of foam cells (Zwaka *et al.*, 2001). It has therefore been hypothesised that CRP may bind to and opsonise native LDL particles, mediating their uptake by macrophages via a CRP receptor present on the surface of macrophages. This receptor has been identified as the low-affinity immunoglobulin receptor CD32 (Zwaka *et al.*, 2001).

In opposition to this hypothesis, however, another study has reported that the uptake of CRP/LDL complexes is not mediated by the binding of CRP to CD32 (Fu & Borensztajn, 2002). In an experiment designed to mimic the *in vivo* conditions of an arterial wall, immobilised CRP was able to bind and cluster LDL molecules in a calcium-dependent

fashion. The CRP-bound aggregated LDL was subsequently taken up by macrophages via a mechanism that seemed to proceed without an increase in CD32 expression. This study therefore suggested an alternative hypothesis by which CRP contributes to the formation of foam cells by its ability to aggregate and not opsonise LDL molecules. CRP has also been reported to stimulate tissue factor production by peripheral blood monocytes *in vitro* (Cermak *et al.*, 1993) and could therefore have a procoagulant role in atherogenesis. All of these studies, however, were performed *in vitro* and their relevance to atherogenesis *in vivo* remains to be determined.

1.4.4 CRP and Myocardial Infarction

The primary process underlying an acute myocardial infarction (MI) is ischemia. Ischemia is described as a lack of blood to a particular area of the body. In the case of myocardial infarction, ischemia results from an occlusion in a coronary artery causing a massive reduction in blood flow to a particular region of cardiac muscle.

Although ischemia is the major process underlying MI, inflammation is also a major contributing factor to the overall pathology of MI. Inflammation in and around the area of hypoxic necrosis makes a major contribution to the final size of the lesion and to the outcome of the patient (Entman *et al.*, 1991). The inflammatory response associated with MI is attributed to the fact that tissue necrosis is a potent activator of the acute phase response. The magnitude of this response appears to reflect the extent of myocardial necrosis (Pepys & Hirschfield & Pepys, 2003).

Two processes which are universal features and which are suggestive of initiation of an acute phase response following a MI, are the local activation of complement within two hours of ischemic injury, and the infiltration of neutrophils. The importance of complement activation in determining the severity of the infarct has been shown in experimental animal models. Inhibition of complement activation in these experimental models at the time of coronary artery occlusion, not only prevented neutrophil infiltration to the site of myocardial necrosis, but also markedly reduced the size of the infarct (Gill *et al.*, 2004).

CRP is always increased after acute MI with levels beginning to rise 4-6 hours after the onset of symptoms and reaching peak values after approximately 50 hours. (Kushner *et al.*, 1978; De Beer & Pepys, 1982). The peak value of CRP is a powerful predictor of post-infarct mortality and morbility. A peak value of serum CRP greater than 200 mg/L is

strongly associated with ventricular rupture; continually high CRP levels predict mortality over the next six months from all causes related to MI (Griselli *et al.*, 1999).

As discussed, CRP shows avid specific binding reactivity for exposed PC residues on the surface of cells following tissue necrosis, and is also able to activate complement through the binding of C1q. A hypothesis has therefore been proposed in which CRP-mediated complement activation contributes to the elevated infarct size and outcome in human clinical acute MI. CRP may bind to non-irremediably damaged cells and through its activation with complement, cause the opsonisation and direct cytotoxicity of these cells, increasing cell death and adding to the pro-inflammatory activity in the zone of direct ischemic necrosis (Hirschfield & Pepys., 2003).

This hypothesis has been supported by experiments involving rat models (Griselli *et al.*, 1999; Pepys *et al.*, 2006). Due to fundamental differences in human and rat forms of CRP, the rat represents an ideal model to investigate the *in vivo* effects attributable to the protein. Both rat and human CRP share extensive amino acid sequence homology, the same flattened 13-jelly roll topology and cyclic pentameric arrangement of protomers. Differences include: glycosylation; covalent linkage of two of the constituent subunits; circulation concentrations of approximately 300 mg/L rising to 900 mg/L during an acute phase response. Also, unlike human CRP which is able to activate complement in both humans and rats, rat CRP does not activate complement in either species (De Beer & Pepys, 1982).

Acute MI was induced in rat models by ligation of the coronary artery. Following induction, human CRP was injected intraperitoneally at 24 hour intervals for a five day period after which the rats were killed. On the fifth day following autopsy, it was reproducibly shown that injection of human CRP had increased infarct size by up to 40% (Griselli *et al.*, 1999). This was in comparison with control rat models in which either buffer alone, or the closely related pentraxin SAP, had been intraperitoneally injected.

This process appears to proceed in a complement-dependent manner as shown by the treatment of rat models with cobra venom factor at the time of coronary ligation and prior to injection with CRP. Cobra venom factor administered *in vivo* rapidly produces profound and sustained depletion of C3, with the absence of active C3 being reported as early as six hours after administration. Treatment of rat models with cobra venom factor completely diminishes the effect of CRP on infarct size even when administered two hours after coronary ligation (Griselli *et al.*, 1999). Immunohistochemical staining of rat

MIs on day five, revealed the deposition of human CRP and rat C3 and rat CRP in the infarcted myocardium.

CRP has also been shown to have a similar affect on cerebral infarct size following ischemic stroke (Gill *et al.*, 2004). CRP concentration is an independent predictor of survival after ischemic stroke in humans and it has therefore been proposed that CRP could enter the ischemic brain and contribute to cerebral tissue damage. Studies investigating the effect of CRP on infarct size following middle cerebral artery (MCA) occlusion used a similar experimental system to the one using rat models described above.

The MCA was permanently occluded in these rats. Without administration of CRP, the cerebral infarct volume in a typical rat model having undergone MCA occlusion, reaches maximum values after approximately 24 hours (Gill *et al.*, 1995). This was shown to proceed independently of complement as when cobra-venom factor was administered to rat models both at the time of occlusion and throughout the experiment, the infarct size was not significantly different to control rat models (Gill *et al.*, 2004).

To investigate the affect of human CRP on cerebral infarct size, human CRP was intraperitoneally injected into rat models at 5 minutes, 24 hours and 48 hours after MCA occlusion. As with the study on myocardial infaction described earlier, human CRP was shown to significantly increase cerebral infarct size compared to control models. Maximal infarct volumes were observed 72 hours after MAC occlusion (Gill *et al.*, 2004). It appears that ischemic damage to the blood brain barrier following a stroke may enable circulating plasma proteins including the administered human CRP and autologous rat complement, to gain access to cerebral tissue and exacerbate infarct size. This seems to provide further evidence to the hypothesis that CRP contributes to ischemic tissue damage and provides novel evidence to suggest that this not only occurs in the heart but may also occur in the brain. However further studies are needed to confirm the role of complement in the latter.

1.5 Amyloidosis & Role of SAP

1.5.1 Amyloidosis – An Introduction

Amyloidosis is a disorder of protein folding in which normally innocuous, soluble proteins polymerise to form abnormal, insoluble protein fibrils (Pepys & Hirschfield, 2003). Following extracellular deposition of these fibrils, they associate with plasma, extracellular matrix proteins and proteoglycans to form amyloid deposits. The growing mass of amyloid deposit invades the extracellular spaces of organs, destroying normal tissue architecture and function (Sipe, 1992).

The amyloidoses were first observed in the mid 19th century by pathologists who noticed the extracellular deposition of 'lardaceous deposits' or 'waxy degenerations' in the vital organs of diseased patients. These deposits were subsequently termed amyloid, meaning starch-like, due to their positive carbohydrate tests with iodine (Virchow, 1851). Despite the characterisation of protein and not carbohydrate as the major component of deposits, the name 'amyloid' has endured.

The underlying molecular abnormalities of amyloidosis can be acquired or hereditary and to date approximately 20 different proteins have been found to potentially form clinically or pathologically significant amyloid fibrils *in vivo* (Pepys & Hirschfield, 2003), as shown in Table 1. Small, focal depositions of amyloid in the brain, heart, seminal vesicles and joints are a universal accompaniment of ageing and are often clinically silent. Accumulative, systemic, large scale deposition of amyloid deposits, however, progressively disrupts the normal structure and function of the surrounding tissues leading to organ failure and often fatality (Pepys & Hirschfield, 2003).

Amyloidosis was once considered a rare disease but recent studies have highlighted the importance of processes involved in amyloid formation in a number of different conditions, some of which are very common. Statistics show that 1 in 1500 individuals die from light chain amyloidosis (systemic AL); whilst over 1 million people who are currently receiving haemodialysis are at risk of developing symptomatic β2 microglobulin amyloidosis. The role of local amyloid deposits in the cerebral blood vessels of Alzheimer's patients and in the islets of Langerhan in the pancreas of type II, mature onset diabetes mellitus is unclear, however the pathology of these conditions is always accompanied by amyloid deposition (Hirschfield, 2004).

Certain correlations exist between the type of fibril protein and clinical manifestations, however in many forms of amyloidosis there is no relationship between the fibril protein and the clinical outcome. Therefore, there are obviously additional genetic and/or environmental factors, distinct from the fibril protein determining when, where and whether the amyloid deposits form. The nature and identity of these factors remains unclear, as do the mechanisms by which amyloid deposits cause disease. A heavy load of amyloid in tissues is inevitably a burden but despite this, a poor correlation is observed between the amount of amyloid deposition and the level of organ dysfunction. Studies have therefore suggested that the pathology of certain forms of amyloidosis may evolve from a cytotoxic effect of fibres inducing cell apoptosis or necrosis in the surrounding tissues (Pepys & Hirschfield, 2003). Conclusive evidence of this is however, still to be obtained.

1.5.2 Fibril Structure

Despite the variation in composition with respect to the protein subunit, isolation of amyloid fibrils from clinically and biochemically diverse amyloid deposits has shown that amyloid fibrils share a common ultrastructure (Sunde *et al.*, 1997) and are characterised by three major criteria. First, all amyloid fibres stain with Congo red dye, exhibiting an apple-green birefringence when examined under cross-polarised light (Sipe & Cohen, 2000). This characteristic is exploited in the use of Congo red as a histochemical dye for amyloid deposits (Pepys & Hirschfield, 2003). Second, X-ray fibre diffraction studies have revealed that amyloid fibres exhibit a cross β-diffraction pattern arising from a repeating core structure composed predominantly of β-sheet (Sunde *et al.*, 1997), as shown in Figure 1.9. Two distinct reflections are common to this diffraction pattern: a strong, sharp meridional reflection corresponding to the fibre axis at 4.7 Å, thought to arise from the hydrogen-bonding distance between β-strands in the component β-sheet; and a more diffuse equatorial reflection around 10-12 Å, the variation of which is likely to depend on the side-chain content of the composite peptide (Sunde *et al.*, 1997). Third, electron microscopy studies of amyloid fibres reveal they are of variable length and approximately 60-100 Å in diameter (Stromer & Serpell, 2005).

Structural analysis of transthyretin amyloid fibrils via fibre diffraction methods and electron microscopy has led to a proposed model for the core molecular structure of all amyloid fibrils (Blake and Serpell, 1996; Serpell *et al.*, 2000a). The model suggests that each fibril is composed of two or more protofilaments (3 nm in diameter) running parallel and twisting about each other. This is in agreement with earlier studies

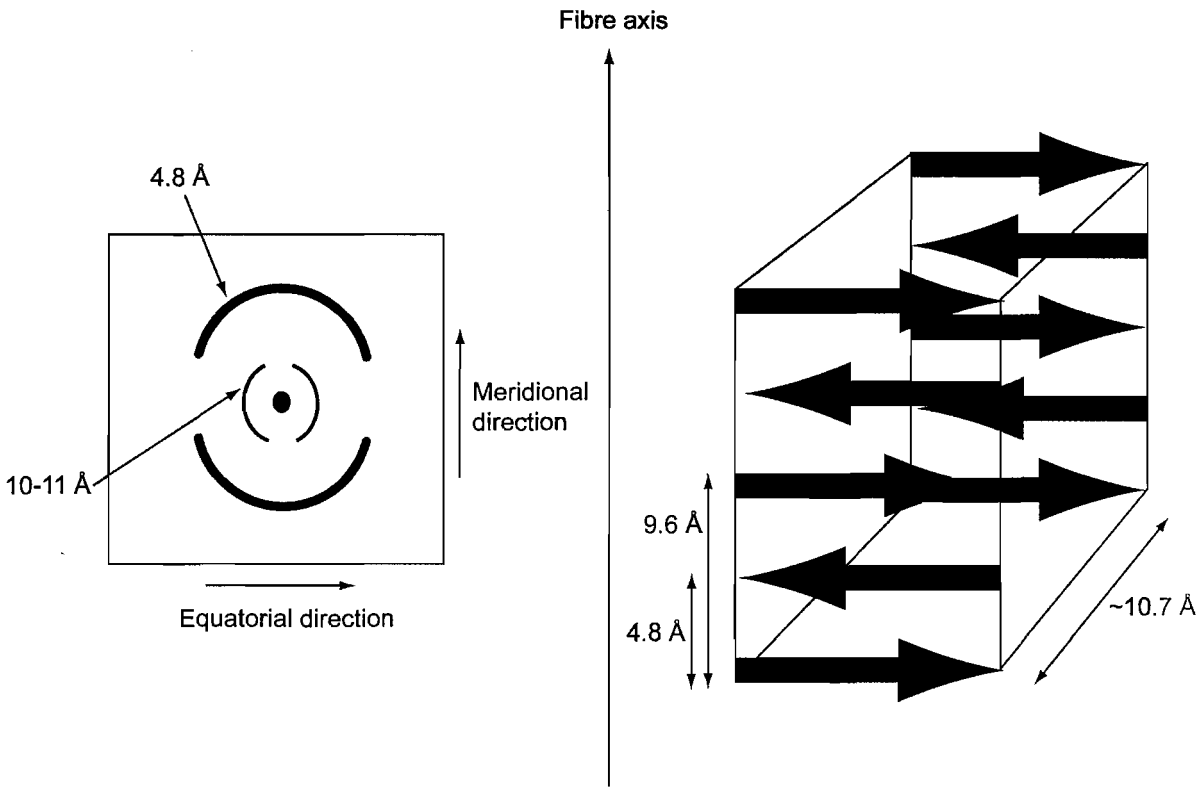


Figure 1.9 The characteristic cross-β relations. On the left is a schematic diagram of a typical fibre diffraction X-ray image. On the right is a proposal for the arrangement of two parallel β-sheets (Serpell *et al.*, 2000).

showing that thin sections of amyloid β (Aβ) protein, the amyloid fibril protein of Alzheimer's disease, show cross-sections containing five or six protofilaments (Fraser *et al.*, 1991a). Protofilaments are arranged around a hollow channel that runs through the centre of the amyloid fibre (Serpell *et al.*, 2000b). The precise number of protofilaments required to form a fibril is still in much debate as different studies have reported the presence of between two and six protofilaments in an amyloid fibre (Serpell *et al.*, 2000b). Fraser *et al.* (1991b) have however, reported that subfibrillar organisation is dependent on both the pH and ion composition of the surrounding environment, factors which may vary with staining technique. It has also been suggested that the composition of the amyloid fibre, in terms of the number and arrangement of protofilaments, may vary with the type of amyloid precursor protein (Stromer & Serpell, 2005). Protofilaments are composed of a number of β-sheets (4 in TTR fibrils) running parallel to the fibril axis and with their constituent β-strands positioned perpendicular to the fibril axis (Figure 1.10). This arrangement is characteristic of the cross β-structure observed in diffraction patterns of amyloid fibrils. A 15 Å twist exists between adjacent β-strands of the β-sheet creating a novel 115.5 Å β-helix consisting of a 24 β-strand helical repeat. The novel helical structure of the fibril enables hydrogen bonding between β-strands to extend the length of the fibril (Blake & Serpell, 1996). This characteristic may contribute to the observed rigidity and stability of the amyloid fibrils.

1.5.3 Fibril Formation

The precise nature and mechanism of fibrillogenesis (amyloid fibril formation) continues to elude researchers. It still remains unclear how the 20 or so known amyloidogenic protein precursors, some of which are vastly diverse in their native fold, all converge into the same cross β-structure known to be common to all amyloid fibres. The native structures of some of these proteins (immunoglobulin light chain; transthyretin; β_2-microglobulin) are known to be composed primarily of β-sheet, but analysis of their three dimensional structures reveals they still require major structural rearrangement in the formation of a cross β-structure. The conformational change in predominantly α-helical proteins such as insulin and some of the amyloidogenic lysozyme variants must be even greater.

Various studies have provided good evidence that destabilisation of the native form of the precursor protein is central in fibril formation. A model describing the structural

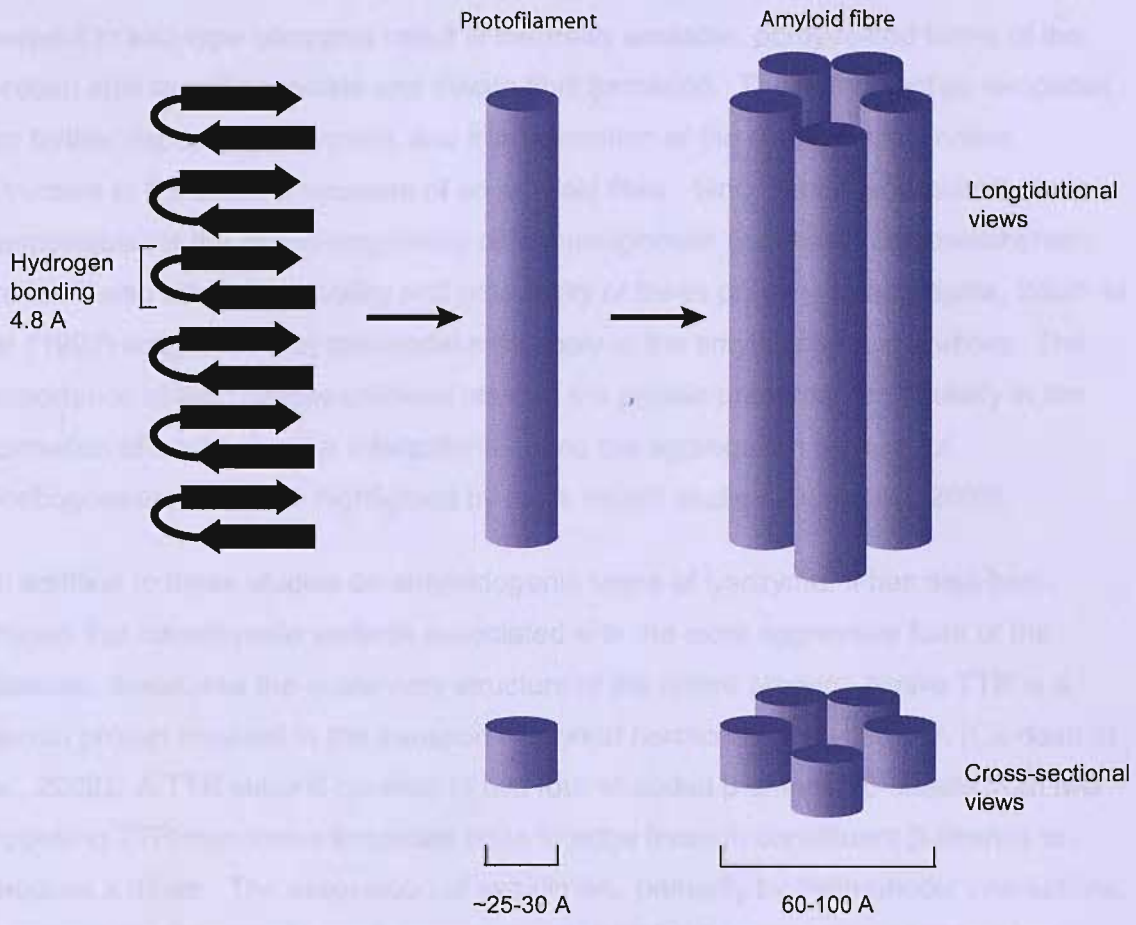


Figure 1.10 The different levels of amyloid fibril structure. The continuous hydrogen-bonded β-sheet within a protofilament is shown on the left; the organisation of protofilaments within an amyloid fibril is shown on the right (Serpell *et al.*, 2000a).

conversion of the two human lysozyme variants, both of which are known to be amyloidogenic, was first reported by Booth *et al.* (1997) and subsequently reaffirmed by Dumoulin *et al.* (2005), (Figure 1.11). In this model, amino acid substitutions with respect to wild-type lysozyme result in thermally unstable, partly-folded forms of the protein able to self-associate and initiate fibril formation. These then act as templates for further deposition of protein, and transformation of the mainly helical native structure to the cross-β structure of an amyloid fibre. Since amino acid substitutions responsible for the amyloidogenecity of immunoglobulin light chain and transthyretin variants also affect the stability and propensity of these proteins to aggregate, Booth *et al.* (1997) suggested that this model may apply to the amyloidoses as a whole. The importance of this partially unfolded state of the protein precursor, particularly in the formation of intermolecular interactions during the aggregation process of fibrillogenesis, has been highlighted by more recent studies (Dumoulin, 2005).

In addition to these studies on amyloidogenic forms of lysozyme, it has also been shown that transthyretin variants associated with the most aggressive form of the disease, destabilise the quaternary structure of the native protein. Native TTR is a serum protein involved in the transport of thyroid hormones and vitamin A (Cardoso *et al.*, 2002). A TTR subunit consists of two four-stranded β-sheets. β-sheets from two opposing TTR monomers associate edge to edge through constituent β-strands to produce a dimer. The association of two dimers, primarily by hydrophobic interactions, produces the functional homotetramer of native TTR (Correira *et al.*, 2006). It has subsequently been reported that the amyloidogenic variants of TTR exhibit diminished tetramer stability, as measured by a lower activation barrier for tetramer dissociation (Lashuel *et al.*, 1998). In addition to these findings, crystallisation studies have shown that in contrast to wild-type TTR (Blake & Swan, 1971), in which the unit cell contains a single dimer, the unit cell of the TTR variant contains eight monomers (Sebastiao *et al.*, 1998). It has therefore been suggested that the dissociation of TTR tetramers into component monomers might mediate TTR fibril formation (Lashuel *et al.*, 1998).

Recent studies have shown that non-amyloidogenic proteins are able to form amyloid fibres under conditions of low pH and high temperature (Rochet *et al.*, 2000). On the basis of these findings, Dobson (2002) has suggested that fibrillisation is a generic component of all protein molecules. The structures of non-aggregating functional globular proteins are dictated by side-chain interactions which mostly encase the main-chain of the protein in the centre of the molecule where it is unable to interact with other molecules. In contrast to globular proteins, the fibrillar form of these proteins is

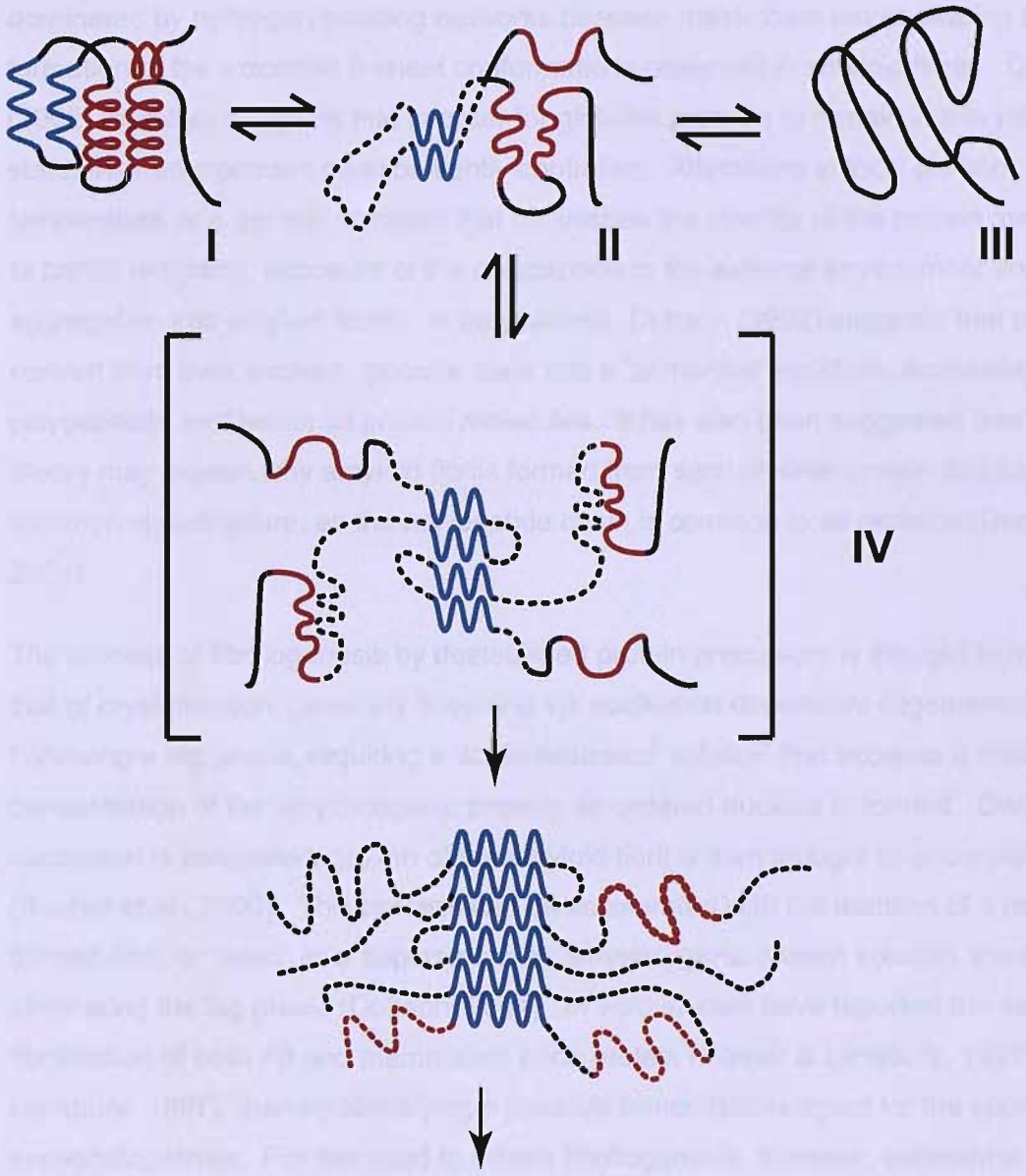


Figure 1.11 A possible mechanism by which fibrils may form in vivo. A partially unfolded intermediate (II) between folded (I) and unfolded (III) states self associates via the β-structural element (blue) to transiently form (IV). This allows further propagation as more intermediate associates to form a stable β-core structure (V). β-strands are shown in blue, helices in red, loops or coils in black, and undefined structure as dashed lines (Booth *et al.*, 1997).

dominated by hydrogen-bonding networks between main-chain atoms leading to the formation of the extended β-sheet conformations observed in amyloid fibrils. Dobson (2002) therefore suggests that in order for globular proteins to remain in this folded state, their environment must be tightly controlled. Alterations in local pH and temperature or a genetic mutation that diminishes the stability of the protein may lead to partial unfolding, exposure of the polypeptide to the external environment and aggregation into amyloid fibrils. In amyloidosis, Dobson (2002) suggests that proteins convert from their evolved, globular state into a 'primordial' structure, accessible to all polypeptides and hence all protein molecules. It has also been suggested that this theory may explain why amyloid fibrils formed from such diverse protein fold all share a common ultrastructure, as the polypeptide chain is common to all proteins (Dobson, 2004).

The process of fibrillogenesis by destabilised protein precursors is thought to mimic that of crystallisation: generally occurring via nucleation-dependent oligomerisation. Following a lag phase, requiring a 'supersaturated' solution that exceeds a critical concentration of the amyloidogenic protein, an ordered nucleus is formed. Once nucleation is completed, growth of the amyloid fibril is then thought to occur rapidly (Rochet *et al.*, 2000). This process can be accelerated with the addition of a pre-formed fibril, or 'seed', to a supersaturated amyloidogenic protein solution, thereby eliminating the lag phase (Dobson, 2004). *In vitro* studies have reported the seeded fibrillisation of both Aβ and mammalian prion protein (Harper & Lansbury, 1997; Lansbury, 1997), thereby identifying a possible transmissible agent for the spongiform encephalopathies. For this seed to initiate fibrillogenesis, however, substantial complementation with the host amyloidogenic protein would be required: a requirement which may explain the observed species barrier that exists in transmissible spongiform encephalopathies.

1.5.4 SAP and Amyloidosis

In vivo amyloid deposits are composed of a disease-specific component and a subset of components common to all amyloid deposits. The so-called disease-specific component comprises the protein precursor characterising a particular amyloid type. The second subset of components include SAP, proteoglycans, α1-antichymotrypsin, some complement components, apolipoprotein E, and various extracellular matrix or basement membrane proteins (Pepys& Hirschfield, 2003). Whilst the role of the majority of these components in the pathogenesis of amyloidosis remains unclear, the

importance of SAP in terms of its universality and quantity in amyloid deposits is unquestionable.

SAP has been shown to bind universally to all types of amyloid fibril *in vivo* and binding to AA and AL type fibrils has been demonstrated *in vitro* (Pepys *et al.*, 1979). Images of amyloid fibrils obtained via electron microscopy have confirmed the presence of SAP molecules stacked along the fibril axis (Figure 1.12). Apart from the fibril proteins themselves, SAP is the most abundant protein in amyloid deposits and can account for up to 15% of the total mass.

The nature and identity of the component to which SAP binds on amyloid fibrils is currently unknown, although two possible candidates have been suggested. An initial insight into a possible mechanism of SAP binding to amyloid fibrils was provided by experiments showing that millimolar concentrations of MOβDG inhibit binding of SAP to amyloid fibrils and dissociates SAP already bound to fibrils (Hind *et al.*, 1984). Since one of the primary components of amyloid deposits are glycosaminoglycans, it was subsequently hypothesised that SAP may recognise and bind to carbohydrate residues situated on the surface of amyloid deposits. Although MOβDG is not present in amyloid deposits, both heparin sulphate and dermatan sulphate have been shown to constitute a small but significant part of amyloid tissues. In a study carried out to investigate SAP binding to these glycosaminoglycans it was shown that SAP demonstrated specific binding to both heparin and dermatan sulphate in a calcium dependent manner and under physiological conditions (Hamazaki, 1987).

Although this may seem a highly plausible theory as to how SAP both recognises and binds to amyloid deposits, additional studies have shown that SAP binds to fibrils formed from pure proteins or peptides *in vitro* with the same affinity ($K_d \sim 10^{-8}$M) as intact fibrils *ex vivo*, suggesting the interaction of SAP is actually with a structural motif present in amyloid fibrils. This motif must be exclusive to the fibrillar form of the amyloidogenic protein as SAP shows reactivity with insulin, transthyretin and Aβ only when in their aggregated form and not in their native soluble state (Tennent *et al.*, 1995).

Various experimental findings and analysis of a theoretical model of an amyloid fibril suggest that this structural motif may take the form of turn and/or loop structures predicted to lie on the surface of the fibrils. Recognition of this motif by SAP is thought

Figure 1.12 Electron microscopy picture of SAP pentamers stacked along the axis of an amyloid fibril (shown with the arrow) (Skinner, 1982).

to occur via the double calcium binding site, since fibril binding is inhibited by the addition of MOβDG. Modelling studies have shown that the double calcium binding site can accommodate a β-turn and that this accommodation results in the positioning of an aspartic acid residue suitable for calcium binding. It has been suggested that the spacing between such turns may equal the distance between individual binding sites on the B face of the SAP pentamer, leading to a highly cooperative association. This may explain the significance of the inherent 5-fold symmetry of the SAP molecule on fibril recognition (Pepys *et al.*, 1997).

A link between SAP and amyloidogenesis was first identified by a group comparing mouse strains of varying resistance to experimental, casein-induced, AA amyloidosis. The study observed that following daily injections of casein to induce amyloidosis, elevated levels of SAP were sustained and persisted through pre-amyloid and amyloid phases. An insight into the role of SAP was also provided as the same effect was observed in strains that were relatively resistant to amyloidosis induction, suggesting that SAP may not be involved in the formation of amyloid but instead in its persistence post formation (Benson *et al.*, 1977).

In order to investigate the role of SAP in amyloidogenesis further, SAP-deficient mice were generated by targeted depletion of the SAP gene. These mice were then compared to SAP-sufficient control mice and evaluated for their ability to develop systemic amyloid (AA) amyloidosis. The mice contained no circulating SAP, but developed normally and were fertile. Following repeated daily injections of casein, the development of AA amyloid deposits was delayed and their quantity reduced in SAP deficient mice, compared to mice sufficient in SAP (Botto *et al.*, 1997). This therefore provides further evidence that SAP contributes to the pathogenesis of amyloidosis.

The lack of clinical or experimental evidence suggesting SAP is required for the conversion of soluble proteins into insoluble fibrillar form, together with the observed development of amyloidosis in SAP deficient mice, has suggested that the role of SAP in the pathology of amyloidosis is not in the formation of fibrils but in their stability once formed. This view has been strengthened by metabolic studies in patients suffering from systemic amyloidosis. These studies investigated the metabolism of SAP in these patients and found that its half-life was dramatically increased from 24 hours in the normal circulation, to 30 days in amyloid deposits (Hawkins *et al.*, 1990b). The only site of SAP catabolism *in vivo* is in hepatocytes, suggesting that phagocytic leukocytes lack receptors for SAP and are unable to recognise and break it down.

SAP is also highly resistant to proteolysis and this is thought to be due to the flattened β jelly-roll topology of the protomer in which the component β-strands are connected by compact loops tightly bound to the oligomeric body of the protein (see section 1.6.2).

Since amyloid deposits are composed only of autologous proteins and glycosaminoglycans, they should be easily degraded *in vivo*. Indeed, human neutrophil cathepsin G as well as whole phagocytic cells have been shown to digest amyloid fibrils *in vitro*. On addition of SAP, however, this effect was blocked (Tennent *et al.*, 1995). In conclusion of this plethora of evidence, it therefore appears that SAP may coat amyloid fibrils and mask the abnormal fibril structure from molecular and cellular mechanisms that should contribute to their clearance and removal.

1.5.5 SAP as a Diagnostic Tool for Amyloidosis

The importance of SAP in the pathogenesis of amyloidosis has resulted in its use as a diagnostic tool. The distribution of SAP can be visualised *in vivo* using scintigraphy with iodine-123 labelled SAP (^{123}I-SAP) (Hawkins *et al.*, 1988; Hawkins *et al.*, 1990a). This technique involves intravenously injecting the radiolabelled SAP into the patient and monitoring the distribution using a gamma camera (Figure 1.13). In a patient suffering from amyloidosis the radiolabelled SAP accumulates at sites of amyloid deposition, thereby preventing its clearance from the blood plasma pool by the liver and resulting in the dramatic increase in its half-life from 24 hours to 30 days (Hawkins *et al.*, 1990b). Scintigraphy, along with Congo red and immunohistochemical staining, remain the preferred methods for the diagnosis of amyloidosis and they are routinely used in conjunction to diagnose and to study the progression of the majority of amyloidoses.

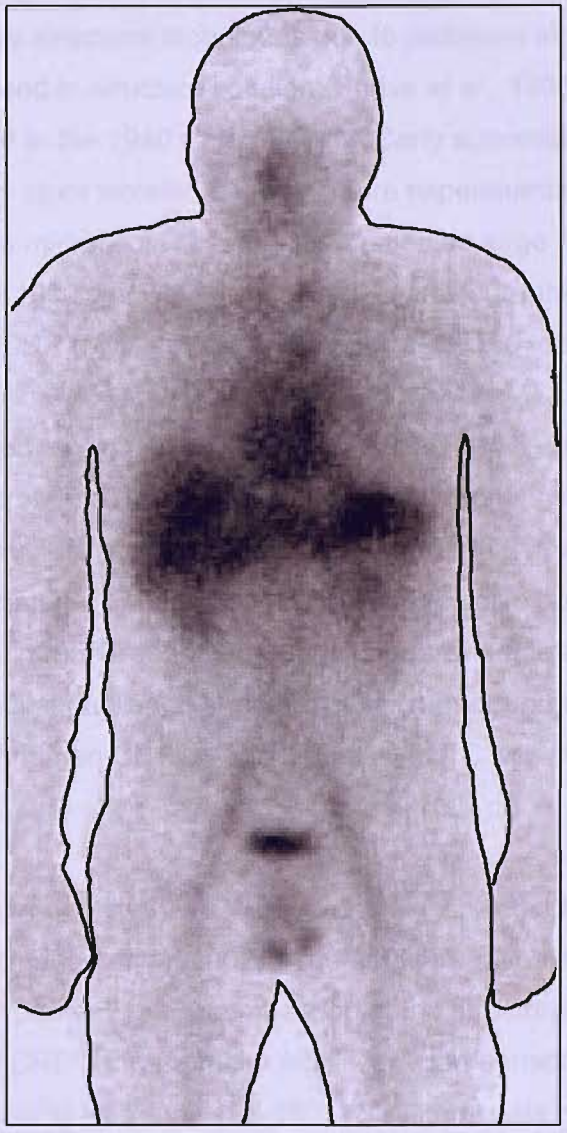


Figure 1.13 ^{123}I-SAP Scintigraphy body scan of a patient with systemic amyloidosis. Regions affected show up as dark grey or black (Hawkins *et al.,* 1988).

1.6 Structure of Pentraxins

SAP and CRP have proved to be extremely difficult subjects for structural determination via X-ray structural techniques due to problems in obtaining and reproducing crystals, and in structure solution (Shrive *et al.*, 1996). Crystallisation of CRP was first reported in the 1940's. In 1947, McCarty successfully crystallised CRP in an attempt to simply store isolated CRP for future experiments. Analysis of precipitated protein via microscopic examination revealed large rhomboid plates. It was not until the early 1990's however, that X-ray crystallographic data was collected on crystals of human CRP (Myles *et al.*, 1990) and rat CRP (Hopkins *et al.*, 1994).

The initial data obtained for human CRP by Myles *et al.* (1990) was of a calcium-depleted form of the protein. These crystals were found to be twinned and although unable to solve the structure of calcium-depleted CRP at the time, they were able to produce a pentameric model of CRP which was used to solve both the calcium-bound (Shrive *et al.*, 1996) and rat (Greenhough *et al.,* unpublished data) forms of the protein. Following on from the first published structure of calcium-bound CRP, a second higher resolution structure of human CRP co-crystallised with PC was published (Thompson *et al.*, 1999).

The structure of calcium-depleted CRP was eventually solved 12 years after the data was first collected (Ramadan *et al.*, 2002). The structure revealed a CRP decamer formed from two CRP pentamers. A comparison of the structures of calcium-bound and calcium-depleted CRP shows a major significant conformational change in a loop involved in calcium binding (residues 138-150). Structural data now available on human CRP provide a detailed description of the general structure of the protein as well as the basis of calcium and PC binding (Shrive *et al.*, 1996; Thompson *et al.*, 1999).

The growth of SAP crystals suitable for X-ray diffraction was first reported in 1986. Various reports of the successful growth of SAP crystals followed (O'Hara, 1988) and X-ray diffraction data was subsequently collected and reported (Wood *et al.*, 1988). Contrary to previous reports (Painter *et al.*, 1982; Perkins & Pepys, 1986) that had suggested that under physiological conditions SAP was in decameric form, these initial crystallisation experiments reported the presence of SAP in a pentameric form. Following the collection of X-ray data sets for the native form of SAP, attempts were made to produce suitable heavy atom derivatives.

The three-dimensional structure of SAP was finally solved by X-ray crystallographic methods and reported in 1994 (Emsley *et al.*, 1994). Subsequent co-crystallisations of SAP and various ligands have been reported: MOβDG (Emsley *et al.*, 1994; Thompson *et al.,* 2002), dAMP (Hohenester *et al.*, 1997), PE (Emsley *et al.*, 1994; Pye, 1997) and CPHPC (Pepys *et al.*, 2002; Purvis, 2002).

1.6.1 Pentamer Structure

A single molecule of CRP and SAP consists of five identical subunits or 'protomers' arranged into a pentameric structure, as shown in Figure 1.14. The five identical protomers are positioned symmetrically around a pore of approximately 30 Å running through the centre of the pentamer.In addition to all five protomers being structurally identical to each other, they also exhibit the same orientation within the pentamer, creating final CRP and SAP pentamers which have two distinct faces. On first inspection, the pentameric structures of CRP and SAP are identical. Structural data (Thompson *et al.*, 1999) has however, revealed that with respect to the five-fold axis running through the centre of the pentamer, the protomers forming the pentamer of CRP are all rotated towards this axis when compared to those of SAP. This characteristic is very apparent when attempting to superimpose the two pentamers and is thought to be responsible for some of the distinct properties exhibited by both proteins.

1.6.2 Protomer Structure

As predicted from their high sequence identity (51% overall) (Woo *et al.*, 1985), the protomeric structures of SAP and CRP are very similar. Despite this striking similarity however, there are specific features unique to and which characterise each protein.

The CRP and SAP protomers have a diameter of approximately 36 Å (Volanakis, 2001) and are composed of 206 and 204 amino acid residues respectively. In both proteins the majority of these residues form antiparallel β-strands arranged into two β-sheets. The overall morphology of a protomer can be described as a two-layered anti-parallel sheet displaying flattened jellyroll topology (Emsley *et al.*, 1994;Shrive *et al.*, 1996). Common to both proteins is the presence of a single long α-helix positioned against one of these β-sheets. Removal of this helix exposes the solitary disulphide bond formed between residues Cys36-Cys97 in CRP and Cys36-Cys95 in SAP.

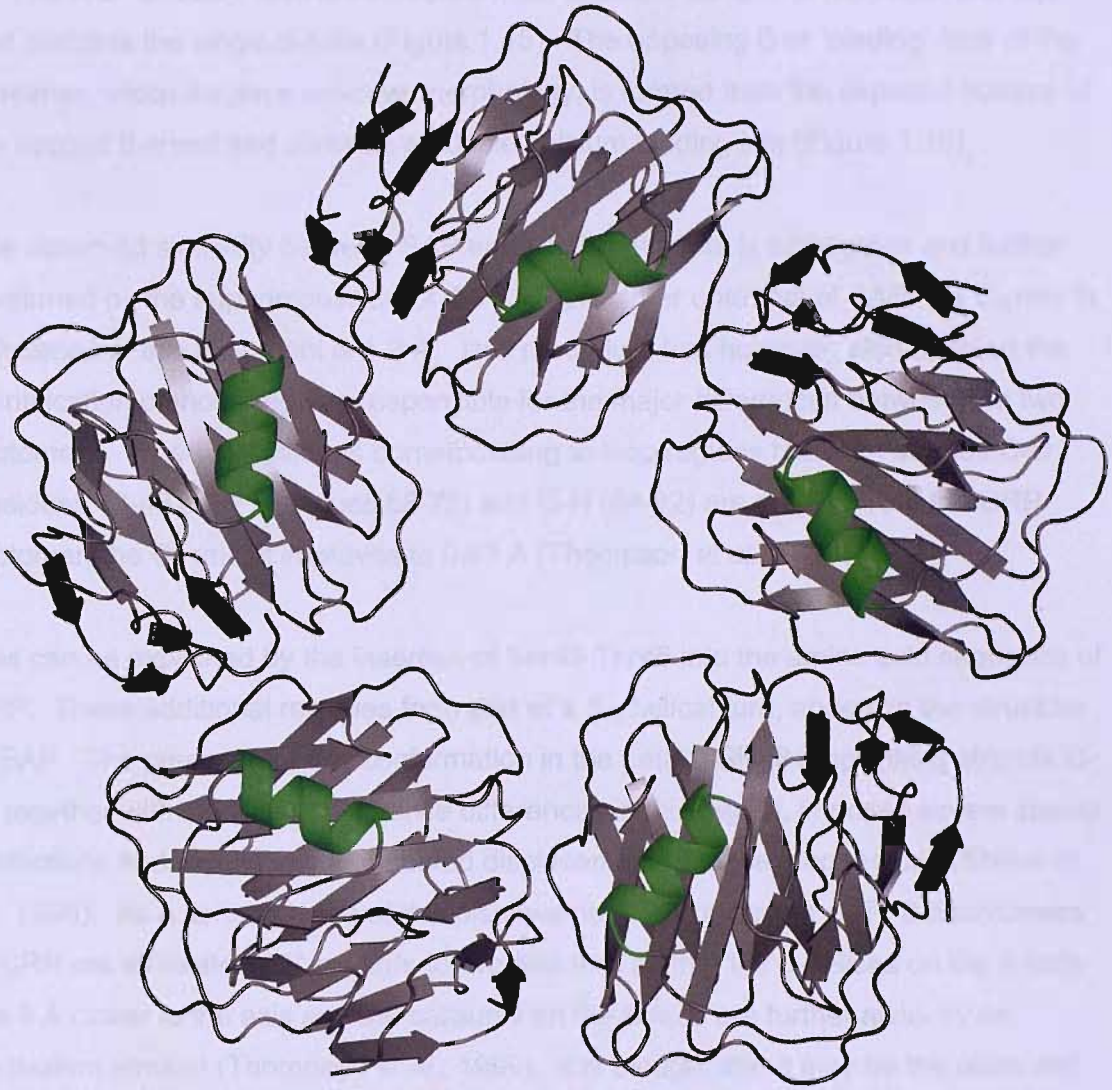


Figure 1.14 Pentamer structure.

A protomer of CRP and SAP exhibits two distinct faces given the nomenclature A and B. The A or 'affector' face is formed from the exposed surface of the flatter β-sheet and contains the single α-helix (Figure 1.15). The opposing B or 'binding' face of the protomer, which displays concave morphology, is formed from the exposed surface of the second β-sheet and contains a double calcium binding site (Figure 1.16).

The observed similarity between SAP and CRP protomers is highlighted and further confirmed by the superimposition of the CRP protomer onto that of SAP: the C_α rms fit calculated of this alignment is 1.3 Å. This procedure has however, also enabled the identification of those regions responsible for the major differences between the two protomers. When 19 residues corresponding to loop regions between strands C-D (residues 43-48), E-F (residues 68-72) and G-H (84-92) are omitted from the CRP protomer, the C_α rms fit improves to 0.83 Å (Thompson *et al.*, 1999).

This can be explained by the insertion of Ser45-Thr46 into the amino acid sequence of CRP. These additional residues form part of a 3_{10} helical turn, absent in the structure of SAP. The presence of this conformation in the Leu43-Gly48 loop linking strands C-D, together with significant sequence differences in this region, imposes severe spatial restrictions and results in the observed displacement of these loop regions (Shrive *et al.*, 1996). As a consequence of this displacement, in contrast to SAP, the protomers of CRP are all rotated 22° towards to five-fold axis so that the α-helices on the A face are 5 Å closer to the axis and the calciums on the B face are further away by an equivalent amount (Thompson *et al.*, 1999). It is thought that it may be the observed rotation of the CRP protomers with respect to the five-fold axis that is responsible for the difference in aggregation properties of the two proteins (see section 1.6.3.1).

It has been suggested that the topologies of the CRP protomer in the crystallographic structures to date, have been stabilised by crystal contacts and that when in solution, CRP protomers are able to rotate about their axis to enable multivalent binding of pentamers to surfaces of varying PC distribution. This may also be a dynamic property shared by SAP. When the rotation of 22° is imposed on the coordinates of the SAP protomer, minimal residue clashes result, suggesting that both proteins exhibit a dynamic capacity to shift the positions of their constituent protomers (Thompson *et al.*, 1999). This movement would redistribute the position of ligand binding sites on the B face, potentially having a significant role in the binding of SAP and CRP to cellular surfaces.

Figure 1.15 The A face. The A face of SAP (top) and CRP (bottom) showing the α-helix (shown in green) covering the solitary disulphide bond (shown in yellow).

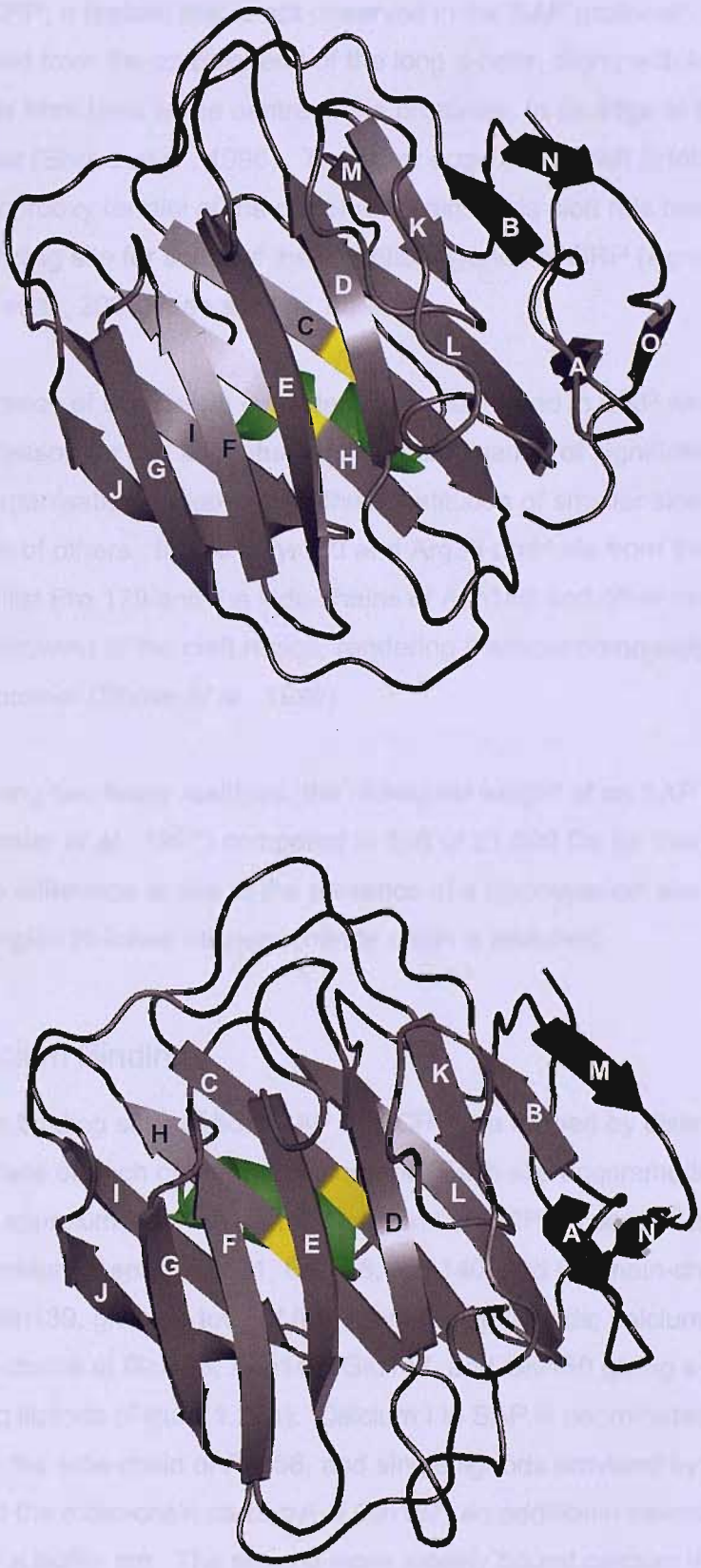


Figure 1.16 The B face. The B face of SAP (top) and CRP (bottom) with the β-strands labelled A to O from the N- to the C-terminus.

Another striking difference between the protomers is the appearance of a deep cleft or crevice in CRP; a feature that is not observed in the SAP protomer. One end of the cleft is formed from the carboxy end of the long α-helix, along with loop 177-182. The cleft extends from here at the centre of the protomer, to its edge at the central pore of the pentamer (Shrive *et al.*, 1996). The outer edge of the cleft is formed by parts of the amino and carboxy termini of the protomer chain. This cleft has been identified as a potential binding site for some of the identified ligands of CRP (Agrawal *et al.*, 2001;Bang *et al.*, 2005) (see section 1.6.4.2).

The appearance of this cleft is considerably accentuated in CRP when compared to SAP. The reason for this accentuation is a combination of significant differences in pentamer organisation, together with the substitution of smaller side-chains and the reorientation of others. In SAP, Tyr160 and Arg38 protrude from the would be floor of the cleft, whilst Pro 179 and the side-chains of Asn186 and other residues contribute to a severe narrowing of the cleft region, rendering it almost completely unrecognisable in the SAP protomer (Shrive *et al.*, 1996).

Despite having two fewer residues, the molecular weight of an SAP protomer is 25,462 Da (Hohenester *et al.*, 1997) compared to that of 21,500 Da for that of CRP (Liu *et al.*, 1982). This difference is due to the presence of a glycosylation site at SAP-Asn32 to which a complex N-linked oligosaccharide chain is attached.

1.6.3 Calcium Binding

The calcium binding sites of both SAP and CRP are formed by distinct loops protruding from the B face of each of the five protomers. Each site accommodates two calcium ions bound approximately 4 Å apart. Calcium I of CRP is coordinated by the side-chains of residues Asp60, Asn61, Glu138, Asp140, and the main-chain carbonyl oxygen of Gln139, giving a total of five coordinating ligands; calcium II is coordinated by the side-chains of Glu138, Asp140, Glu147, and Glu150 giving a total of four coordinating ligands (Figure 1.17a). Calcium I in SAP is coordinated by two ligands provided by the side-chain of Asp58, and single ligands provided by Asn59, Glu136, Asp138 and the main-chain carbonyl of Gln137; an additional seventh ligand is provided by a buffer ion. The second more loosely bound calcium II is coordinated by the side-chains of Glu136, Asp138, Gln148 and two water molecules (Figure 1.17b).

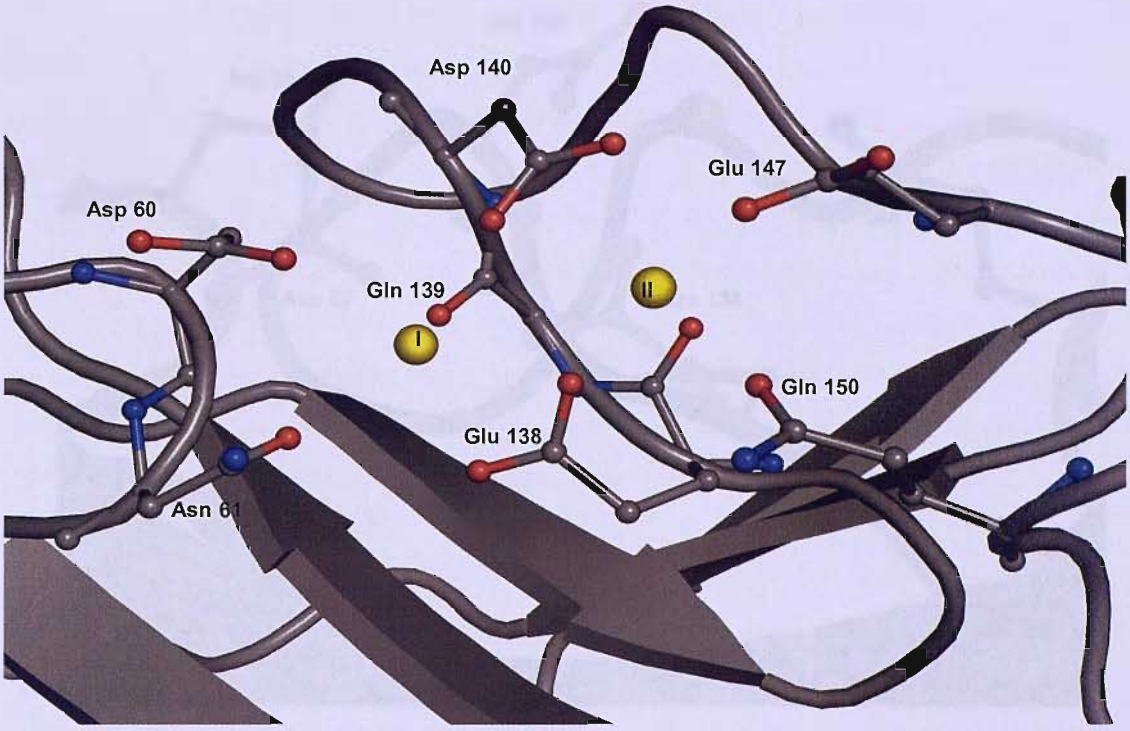


Figure 1.17a The calcium binding site of CRP. Calcium ions are shown in yellow.

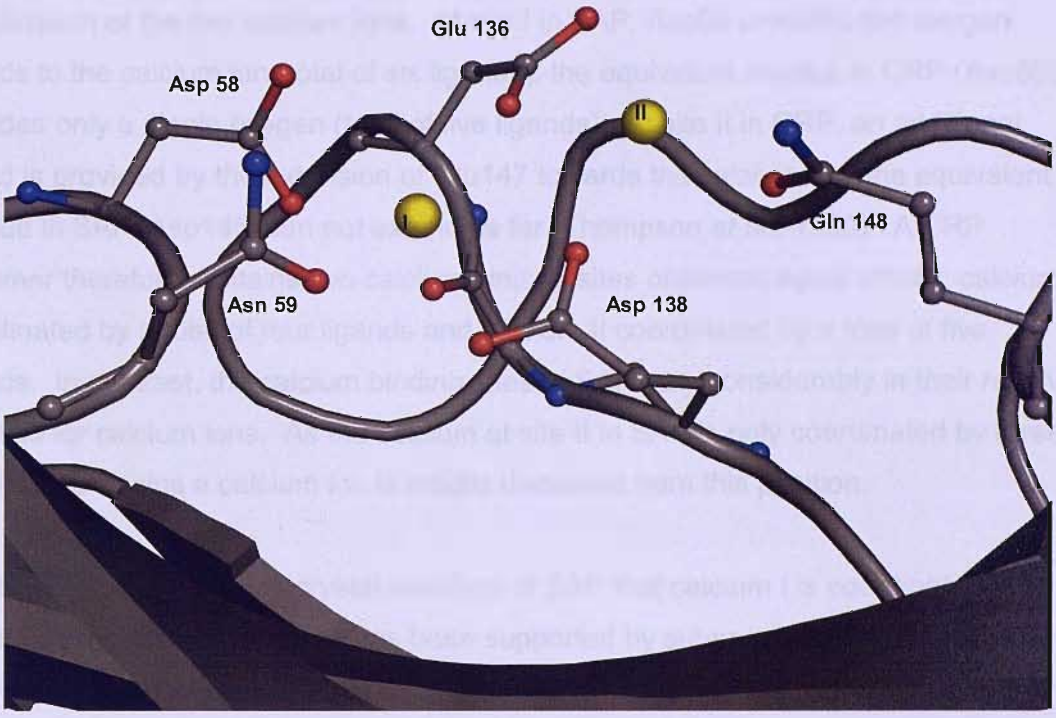


Figure 1.17b The calcium binding site of SAP. Calcium ions are shown in yellow.

Comparison of the calcium binding sites of CRP and SAP reveals aberrations in the coordination of the two calcium ions. At site I in SAP, Asp58 provides two oxygen ligands to the calcium ion (total of six ligands); the equivalent residue in CRP (Asp60) provides only a single oxygen (total of five ligands). At site II in CRP, an additional ligand is provided by the extension of Glu147 towards the calcium ion; the equivalent residue in SAP, Asp145, can not extend as far (Thompson *et al.*, 1999). A CRP protomer therefore contains two calcium binding sites of almost equal affinity; calcium I coordinated by a total of four ligands and calcium II coordinated by a total of five ligands. In contrast, the calcium binding sites of SAP vary considerably in their relative affinities for calcium ions. As the calcium at site II in SAP is only coordinated by three protein side-chains a calcium ion is readily displaced from this position.

The observation in the first crystal structure of SAP that calcium I is coordinated with greater strength than calcium II has been supported by subsequent crystal structures of SAP in which the second calcium binding site is occupied by a water molecule forming hydrogen bonds with Glu136, Asp138 and Gln148 (Thompson, 1997). It has been suggested that the lower binding affinity of the second binding site for calcium may be because the site is occupied by an alternative divalent ion; this has been supported by experiments showing that SAP also has affinity for barium, cobalt, copper, nickel, zinc and cadmium (Haupt & Heimburger, 1972; Potempa *et al.*,1985).

The ability of SAP to bind these alternative divalent ions appears to be strongly dependent on the pH of the surrounding environment; this appears to be of particular importance when considering copper binding to SAP. Binding of SAP to the yeast cell wall polysaccharide zymosan, was measured as a function of pH and divalent cation. The results showed that at pH 5.0, the percentage of SAP bound to zymosan in the presence of copper was increased six-fold, compared to the percentage bound in the presence of calcium. This may be of importance in the role of SAP in inflammatory processes, as a plethora of evidence has shown that serum concentrations of copper increase dramatically during inflammation and may contribute to pathways operative in the immune response (Milianino *et al.*, 1979).

Attempts at growing crystals of SAP in complex with copper of suitable quality for X-ray diffraction experiments have so far proved unsuccessful. These attempts have however, suggested that SAP may preferentially bind copper over calcium at one or both sites. When SAP-calcium crystals are soaked in a solution of copper chloride, the crystal lattice is apparently disturbed leading to the disintegration of crystals

(Thompson, unpublished data). It has therefore been suggested that divalent cations, alternative to calcium, in association with changes in micro-environmental pH, may influence the binding reactivities and biological effects initiated by SAP (Potempa *et al.*, 1985).

The presence of two calcium ions seems to be important in stabilising the structures of both SAP and CRP. Comparison of the structures of calcium-bound and calcium-depleted human CRP has highlighted the importance of calcium binding in both protein stability and ligand binding. Conformational changes in protomer structure following the binding of calcium, lead not only to the formation of a binding site for identified ligands such as PC; but also to an increase in protein stability. Analysis of calcium-depleted CRP indicates that the only site at which addition or loss of calcium causes significant allosteric or conformational changes, is in the mobile calcium-binding loop of residues 138-150 (Ramadan *et al.*, 2002). When calcium is absent from CRP, the loop protrudes from the surface of the protein and is vulnerable to proteolytic attack from the proteases, Nagarase and Pronase (Kinoshita *et al.*, 1989). Proteolytic cleavage of CRP by these proteases occurs at the same site, in a protease-sensitive region formed by residues Asn145, Phe146 and Glu147.

Binding of calcium at a concentration between 0.1 M and 1.0 M protects CRP from proteolysis by either Pronase or Nagarase. On binding calcium, the mobile loop region folds in to protect the protease sensitive site from proteolytic attack (Thompson *et al.*, 1999). Glu147 is now positioned so as to coordinate a calcium ion at site II and the binding avidity of the protein for PC is restored. The physiological relevance of these observations remains unclear however, one possible theory is that calcium concentration may serve not only to regulate the relative affinity of CRP for physiological ligands containing PC, but also through the 138-150 loop, regulate the susceptibility of the protein to proteolysis and denaturation ((Ramadan *et al.*, 2002).

The presence of the two calcium ions also seems to be important in stabilising the structure of the SAP protomer, as removal of the two calcium ions renders the protomer susceptible to proteolysis by α-chymotrypsin, trypsin, pronase and nagarase (Kinoshita *et al.*, 1992). The apparent resistance of calcium-bound SAP to proteolysis can be attributed to the compact nature of this form of the protein in which all but one of the loops connecting adjacent strands are in contact with the central β-sandwich; the only loop exempt from this contact is the long loop associated with calcium binding. When calcium is removed from SAP the conformation of this loop is altered and the

region is destabilised, enabling proteolytic cleavage at residues 144-145 and generating two fragments of 18 kDa and 7.5 kDa (Kinoshita *et al.*, 1992). To date, crystallisation trials of calcium free SAP have failed to produce robust, diffracting crystals.

1.6.3.1 Auto-aggregation

There has been much debate over the precise physiological oligomeric assembly of SAP and analysis of the assembly has proved difficult due to the characteristic auto-aggregation of isolated SAP when exposed to calcium. This aggregation is inhibited by all known calcium-dependent ligands, physiological conditions of albumin (Pepys *et al.*, 1997), physiological ionic strengths and supra-physiological concentrations of salt (600 mM). Analysis of pentamer stacking in some of the crystal forms of SAP currently available suggest that auto-aggregation may proceed via A face to B face stacking $(AB)_n$. The carboxyl group of residue Glu167 located on the N-terminus of the α-helix on face A protrudes away from the body of the protein and extends towards the calcium binding site located on the B face of an adjacent pentamer (Baltz *et al.*, 1982). Site-directed mutagenesis studies of Glu167 to either a glutamine or a serine completely abolishes calcium-dependent aggregation (Booth, 1998).

The observation that all known calcium-dependent ligands of SAP block auto-aggregation, can also be explained by this interaction. Ligands bound to SAP via the double calcium binding site presumably prevent the Glu167 calcium interaction and therefore abolish the ability of pentamers to stack in this manner. The analysis of a recent high salt, high calcium SAP structure revealed an obscure piece of electron density located at the binding site and it has been hypothesised that this may explain the apparent ability of high salt concentrations to prevent SAP auto-aggregation. Subsequent model building of various ligands into this electron density has revealed a chloride ion to be the most likely inhabitant. It has therefore been postulated that in high concentrations of salt, a chloride ion provides an additional ligand to the calcium ions, preventing interaction with the carboxyl group of Glu167 (Kolstoe, 2005).

In the absence of calcium, SAP forms stable decamers thought to derive from A to A face stacking of pentamers. This is consistent with the findings of both X-ray scattering studies (Ashton *et al.,* 1997) and protease sensitivity analysis of SAP in the absence of calcium (Kinoshita *et al.*, 1992). A to A face stacking of pentamers would expose the calcium binding loops on the B faces to proteases and would explain the observed susceptibility of SAP to proteolysis in the absence of calcium.

SAP in complex with 2'-deoxyadenosine-5'-monophosphate (dAMP) forms stable B face to B face decamers. Analysis of the pentamer to pentamer interface reveals that most of the decameric interactions are mediated by the ligand and more specifically by the stacking of adenine bases between the two pentamers (Hohenester *et al.*, 1997). Despite these occurrences of SAP decamer formation and previous uncertainty over the oligomeric assembly of SAP, it has been successfully demonstrated that SAP exists as a single, uncomplexed pentamer in human serum (Sorensen *et al.*, 1995). It therefore appears that decamer formation is purely an *in vitro* artefact resulting from ligand mediated contact or from electrostatic interactions between the two faces of SAP pentamers.

Interestingly, CRP does not share the characteristic auto-aggregation property of SAP in the presence of calcium. In CRP, the carboxyl group believed to mediate the aggregation (Glu167) is maintained in the form of Asp169. The decrease in side chain length together with the 22° subunit rotation with respect to SAP, may explain the difference in autoaggregation properties of the two pentraxins.

1.6.4 Ligand Binding to CRP

1.6.4.1 Phosphocholine Binding Site

Despite the identification of PC as a principle ligand of CRP as early as the 1930's, the lack of X-ray crystallographic data meant that the PC binding site was not characterised until the 1980's. The first attempt at localising the PC binding site on CRP was performed using immunoelectron microscopy (Roux *et al.*, 1983). An anti-CRP monoclonal antibody which binds at or near the PC binding site in the presence of calcium, was used to locate the position of the PC binding site on CRP by analysis of negatively stained CRP-Ab complexes by transmission electron microscopy. The study demonstrated that the PC-binding sites on all CRP protomers were located on one face of the molecule.

The elucidation of the crystal structure of CRP (Shrive *et al.*, 1996) enabled a more detailed proposition of the location of the PC binding site. Up to this point, speculation had centred on whether PC binding was mediated directly through a phosphate-calcium electrostatic interaction, or whether calcium had more of an indirect role, as an allosteric effector. The structural data obtained by Shrive *et al.* (1996) supported the former of these claims: direct interaction of phosphate at the calcium binding site, coupled with a hydrophobic interaction between the three choline methyl groups and a

hydrophobic pocket formed by Phe66, Leu64 and Thr76. This concurred with site-directed mutagenesis studies that had identified Thr76 as an important residue in PC binding (Agrawal *et al.*, 1992). The side-chains of residues Ser68, Ser74 and Glu81 are all positioned at the opposing end of the hydrophobic pocket to the calcium ions enabling, after reorientation of its side-chains, interaction of Glu81 with the choline nitrogen (Shrive *et al.*, 1996).

Although the CRP structure described by Shrive *et al.* (1996) did not contain any PC, a crystallographic artefact resulting from crystal packing produced relevant data for the CRP-PC complex. Due to crystal packing, two independent pentamers were arranged in a staggered decameric fashion. The resulting crystal contacts caused both a loss of calcium and the structural rearrangement of two of the protomers, one from each pentamer. Native, calcium-bound CRP protomers were termed type I; calcium-depleted CRP protomers were termed type II. Interaction of a displaced loop from the type II protomer with the proposed binding site of type I protomer mimicked PC binding. Glu147 (II) chelates the two calcium ions and Phe146: mimicking the methyl groups of choline, is directed towards the hydrophobic pocket (Shrive *et al.*, 1996).

The elucidation of the structure of CRP in complex with PC to 2.5 Å (Thompson *et al.*, 1999) confirmed both the location of the binding site and the mode of PC binding. The two terminal oxygens of the PC phosphate group interact directly with the two calcium ions bound to the B face of the protomer. The third oxygen of the phosphate group is left pointing away into the solvent, therefore allowing possible ester linkage of PC with other molecules. The remaining choline moiety runs along the surface of CRP, packed closely to Phe66, towards the side chain of Glu81. In this orientation, Phe66 is ideally positioned to provide stabilising hydrophobic contacts for the methyl groups of the choline moiety. In addition, with the PC positioned in this manner, the negatively charged side chain of Glu81 is positioned approximately 3.8 Å from the positively charged quaternary nitrogen of PC, providing further protein-ligand stabilising contacts (Thompson *et al.*, 1999).

The importance of these two residues at the PC binding site of CRP was confirmed using site-directed mutagenesis to produce a CRP mutant incapable of binding PC-containing ligands (Agrawal *et al.*, 2002). Production of a double CRP mutant in which both of these residues were mutated to alanine, rendered the CRP mutant unable to bind PC, as did the single mutation of Phe66 to an alanine. This observation, together with the finding that mutation of phenylalanine to another aromatic residue (Tyrosine)

produced a CRP molecule capable of binding PC in an identical manner to the wild-type, indicates that Phe66 is the major determinant in PC binding. Additional residues identified as being important to the geometry of the PC binding site are residues Gly79 and Asn61, which along with Glu81 form the lining of the hydrophobic cavity (Thompson *et al.*, 1999).

Comparison of the binding sites of CRP (Shrive *et al.*, 1996; Thompson *et al.*, 1999) and SAP (Emsley *et al.*, 1994) reveal minor variations in residue type at certain 'key positions' which may explain the differences in both ligand binding specificity and affinity of the two proteins. In particular, these aberrations can be used to explain the observed differences in the binding of PE and PC between the two proteins. Unlike CRP which is able to bind both PC and with a lower affinity PE, SAP binds to PE but fails to bind to PC residues. PE is identical in structure to PC apart from the absence of the three methyl groups attached to the quaternary nitrogen.

Possibly the most significant difference between the two proteins, which may explain the variability in ligand specificity, is the size and geometry of the hydrophobic pocket. Substitution of CRP residues Thr76 and Phe66 for two Tyr residues in SAP, in combination with various main-chain alterations, creates a significantly larger hydrophobic pocket in CRP (8.7 x 7 x 3.5 Å) able to accommodate the choline methyl groups of PC (Shrive, 1996). Another major alteration is the replacement of Glu81 in CRP with the equivalent residue Lys79 in SAP. In CRP, the negatively charged side-chain of Glu81 coordinates the positively charged quaternary nitrogen of PC; the equivalent residue in SAP, Lys79, has a positively charged side-chain and therefore does not favour this interaction (Thompson *et al.*, 1999). These observations may explain the decreased binding affinity of SAP for PC.

1.6.4.2 C1q Binding Site

CRP has been shown to activate the classical pathway of complement through the binding of the first complement component, Clq. Mutagenesis studies on CRP, the first to try and elucidate the precise location of the binding site, indicated that Asp112 was an important residue for recognition of Clq by CRP (Agrawal *et al.*, 1994). Substitution of this negatively charged residue with asparagine (polar), alanine (non-polar) or lysine (positively charged), resulted in a decreased avidity of ligand-bound CRP for Clq and also decreased the complement-activating efficiency of CRP (as measured by C3-fragment deposition).

Also implicated in the interaction with C1q are residues CRP-Lys114 (Agrawal *et al.*, 1994; Agrawal *et al.*, 2001; Bang *et al.*, 2005) and CRP-Glu88 (Agrawal *et al.*, 2001). Based on the previously identified C1q binding sites of IgG and IgM in which negatively charged residues play a critical role in C1q binding and complement activation, it has been suggested that the CRP binding site for C1q may also involve a ring of negatively charged residues, including CRP-Asp112 and CRP-Glu88, positioned toward the central pore of the pentamer. The apparent hindrance placed on C1q binding by the presence of CRP-Lys114 further supports this hypothesis.

Mutation of the positive charge of this lysine residue to threonine (polar), alanine (non-polar) or glutamine (negatively charged) increases the binding avidity of CRP for C1q by 2-3 fold (Agrawal *et al.*, 1994). It appears that the extended positive side-chain of CRP-Lys114 may interfere with the interaction between CRP-Asp112 and C1q. Further to this and following the suggested structure of the binding site, it has been proposed that this effect may result from the side-chain of CRP-Lys114 from a neighbouring CRP protomer (Volanakis, 2001). If this was indeed the case, it would also support a hypothesis in which more than one CRP protomer is required for the successful binding of C1q and that this binding may only take place following a significant conformational change in CRP thereby altering the position of the hindering CRP-Lys114.

In contrast to the contributions of CRP-Asp112 and CRP-Lys114, CRP-Glu88 has been proposed to have a direct role in the promotion of complement activation (Agrawal *et al.*, 2001). In order to initiate the complement cascade, Clq must undergo a conformational change resulting in the reorientation and mutual activation of the two C1r zymogens of the Clr_2Cls_2 pro-enzyme. Whilst mutation of CRP-Glu88 only modestly decreased the ability of CRP to bind C1q, the ability to form the C3 convertase was significantly reduced. Therefore it has been proposed that CRP-Glu88 strongly influences a conformational change in C1q required for the activation of complement (Agrawal *et al.*, 2001).

Elucidation of the crystal structure of CRP revealed that CRP-Asp112, CRP-Lys114 and CRP-Glu88 are all located in a predominant cleft situated on the A face of the CRP protomer (Shrive *et al.*, 1996). It seems that in comparison with the B face, the A face of the protomer is a more plausible site for Clq binding as in order to bind and activate complement, CRP must also bind an appropriate multivalent ligand through the PC binding site located on the B face (Kaplan & Volanakis, 1974; Volanakis, 1982). This

led to a hypothesis proposing that the cleft is a very important structural element in the binding of Clq to CRP. The cleft is approximately 24 Å long, 7.5 Å deep and 12.4 Å wide and is deep and narrow at its origin in the centre of the protomer becoming wider and shallower towards its edge at the central pore of the pentamer (Thompson *et al.*, 1999). The walls of the cleft are constructed from Ser5, Arg6, Gln203, Pro206, Trp187, Arg188, Asn160, G1y177, Leu176, Tyr175, His95 and Aspl12 and the bottom is lined by Asn158, His38, Leu37, Va194 and Asp112 (see Figure 1.18).

In order to investigate this hypothesis further, additional site-directed mutagenesis studies were performed on CRP residues forming this cleft. These studies have subsequently proposed the shallow end of the cleft bounded by the 112-114 loop, the C-terminus of the protomer and the C-terminus of the single α-helix of the protomer (residues 169-176), as the precise location of the C1q binding site (Agrawal *et al.*, 2001). More specifically, these studies have highlighted CRP-Asp169, CRP-Thr173, CRP-Leu176 and in particular CRP-Tyr175, as important contact residues for C1q, and CRP-Asn158 and CRP-His38 for a possible role in maintaining the correct geometry of the binding site (Bang *et al.*, 2005).

The exact stoichiometry of CRP-C1q binding remains unclear due to the current absence of a three-dimensional structure describing the interaction. Assuming that it is the globular head domains of C1q that bind CRP (see section 1.2.7), the dimensions of both molecules suggest that only a single globular C1q head can bind to a CRP pentamer bound flat to the surface of its ligand via the B face. The diameter of a C1q globular head has been estimated, through electron microscopy (Knobel *et al.*, 1975), to be 50 Å; while the overall dimensions of a CRP pentamer are approximately 102 Å outside diameter, 30 Å of this comprising the central pore, and 36 Å per protomer diameter. Assuming this hypothesis to be correct, multiple CRP molecules would therefore be required to bind all six C1q globular heads (Kishore & Reid, 2000).

The number of complexed CRP molecules required to form a complete C1q binding site ultimately depends on the number of C1q binding sites per CRP pentamer. The C_α of Tyr175 in all five protomers are arranged in a circle of 38 Å in diameter, whilst the C_α of Asp112 are also all arranged in a similar fashion (Agrawal *et al.*, 2001). These dimensions suggest that, following a significant conformational change on ligand binding to the B face of CRP, residues Asp112 and Tyr175 are brought closer together or exposed more fully to solvent, thereby creating the necessary geometry for the C1q

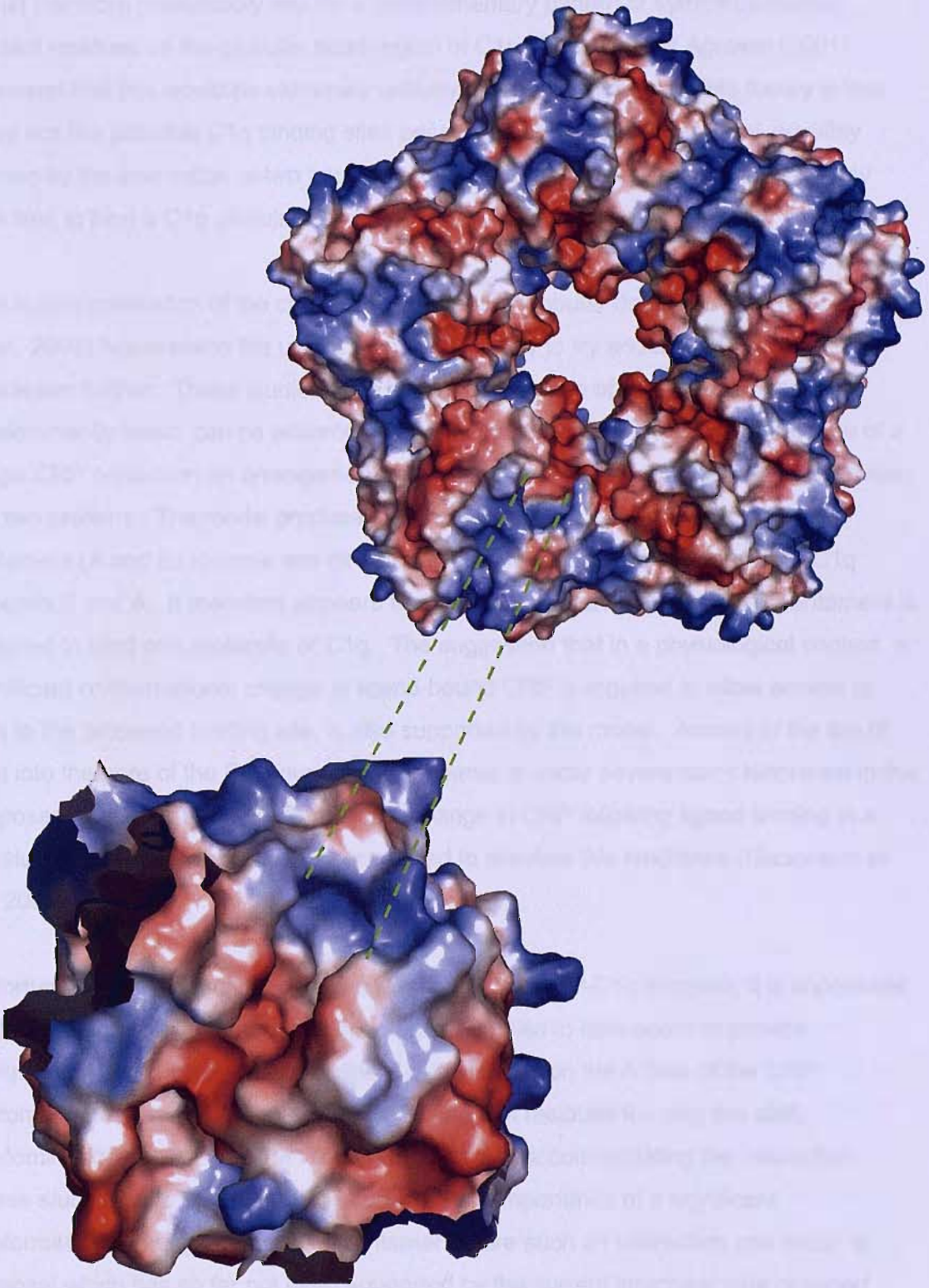


Figure 1.18 The proposed C1q binding site on the CRP protomer (Thompson *et al.*, 1999).

binding site. Such a binding site would however, exhibit pentameric symmetry and would therefore presumably require a complementary pattern of symmetry-related contact residues on the globular head region of C1q. The group of Agrawal (2001) proposed that this would be extremely unlikely and that a more plausible theory is that there are five possible C1q binding sites per pentamer. One of these sites, possibly formed by the interaction of two opposing CRP protomers, could then function at any one time to bind a C1q globular head.

The recent publication of the crystal structure of the globular domain of C1q (Gaboriaud *et al.*, 2003) has enabled the use of modelling studies to try and define the CRP-C1q interaction further. These studies have shown that the top of a single C1q head, predominantly basic, can be accommodated by the negatively charged central pore of a single CRP pentamer; an arrangement that exploits the striking complementary between the two proteins. The model proposed also enables Asp112 and Thr175, from two protomers (A and E) to come into direct contact with appropriate residues from C1q subunits B and A. It therefore appears that multi-point attachment of CRP pentamers is required to bind one molecule of C1q. The suggestion that in a physiological context, a significant conformational change in ligand-bound CRP is required to allow access of C1q to the proposed binding site, is also supported by the model. Access of the top of C1q into the pore of the PC-bound CRP pentamer is under severe steric hindrance in the proposed model. A slight conformational change in CRP following ligand binding in a physiological environment would be expected to alleviate this hindrance (Gaboriaud *et al.*, 2003).

Unfortunately, until structural data is available for the CRP-C1q complex, it is impossible to define the precise nature of this interaction. Studies to date seem to provide unequivocal evidence that the prominent cleft situated on the A face of the CRP protomer is the site of C1q binding to CRP, and that residues forming this cleft, predominantly Asp112 and Tyr175, are important in accommodating the interaction. These studies have however, also suggested the importance of a significant conformational change in the CRP pentamer before such an interaction can occur, a proposal which has so far not been supported by the current structural data obtained from CRP-ligand complexes. It is therefore difficult to predict whether suitable physiological conditions which are obviously required for the interaction to occur, can be sufficiently replicated for use in crystallisation studies.

1.6.4.3 Fcγ Receptor Binding Site

Structural studies involving site directed mutagenesis of CRP residues have recently been undertaken to try and characterise the precise location of the binding site for FcγR on CRP (Bang *et al.*, 2005). CRP-Leu176 was previously identified as an important contact residue as when it was substituted for a glutamate it caused abrogation of the binding of CRP to FcγRI but not FcγRIIa (Marnell *et al.*, 2005). Additional residues identified as being important in FcγR binding to CRP were CRP-Thr173, CRP-Asn186 and CRP-Lys114. It still remains unclear as to whether residues involved in binding FcγRI are among a subset of a group of residues also involved in binding to FcγRIIa, as although mutation of Thr173 and Asn186 affected binding to both receptors, mutation of Lys114 and Leu176 only affected the binding of Fcγl (Bang *et al.*, 2005).

The residues implicated in FcγR binding are located on the A face of the pentamer and in particular Thr173 and Leu176 are located at the C-terminal of the CRP α-helix forming one end of the cleft that runs through the A face of the protomer: the C1q binding site. The identification of these residues along with Lys114, also implicated in C1q binding, suggest that a portion of the C1q binding site may have been co-opted for binding to FcγR sometime during the evolution of the FcγR-mediated adaptive immune response (Bang *et al.*, 2005).

1.6.5 Ligand Binding to SAP

The characteristic affinity and calcium-dependent binding of SAP to agarose derived from marine algae, was originally used to isolate the protein from serum and amyloid extracts (Pepys *et al.*, 1977). It has subsequently been shown that SAP also binds to glycosaminoglycans, especially heparan sulphate and dermatan sulphate (Hamazaki, 1987), to mannose terminated glycans (Kubak *et al.*, 1988) and to glycans with pre-terminal galactose residues (Hamazaki, 1986). Despite showing binding affinity for these ligands, the most well defined carbohydrate ligand of SAP is the agarose-derived pyruvate acetal of galactose: methyl 4,6-O-(1-carboxyethylidene)-β-D-galactopyranoside (MOβDG) (Hind *et al.*, 1984).

SAP binds with high affinity to PE (Schwalbe *et al.*, 1992) but unlike CRP does not bind PC. This property is utilised in the one step preparation of SAP and CRP from blood plasma, via calcium-dependent affinity chromatography. Once bound to PE on the

column, the two pentraxins are then separated from each other via their affinity for PC, which when added even at high concentration, is unable to separate SAP from the affinity column (Pontet *et al.*, 1978; Hawkins *et al.*, 1991).

1.6.5.1 Binding of MOβDG and PE

SAP was first shown to bind the marine algae-derived polysaccharide, agarose, in a calcium-dependent manner (Pepys *et al.*, 1977). The precise nature of the agarose component to which SAP binds was however, not identified until 1984 (Hind *et al.*, 1984). Agarose is composed of repeating units of agarobiose (1,3-linked β-D-galacatopyranose and 1,4-linked 3,6 anhydro-α-*L*-galactopyranose) together with trace amounts of sulphate and pyruvate, the latter as the 4, 6-pyruvate acetal of β-D-galactopyranose (Hind *et al.*, 1984). In the study by Hind *et al.* (1984), the ability of SAP to bind to agarose of varying pyruvate content was investigated. The results showed that the amount of SAP bound to an agarose gel was directly related to their pyruvate content and not at all with their sulphate content. This binding was inhibited by methylation of the carboxy moiety of pyruvate and suggested for the first time, a possible mechanism through which SAP may bind to its ligands.

In an attempt to define a lower molecular weight ligand for SAP, the methyl 4.6-O-(1-carboxyethylidene)-β-D-galactopyranoside (the cyclic acetal MOβDG) was synthesised. This compound was shown to inhibit the binding of human, mouse and plaice SAP to pyruvate-rich agarose and amyloid fibrils (Hind *et al.*, 1984).

Co-crystallisation of SAP with MOβDG (Emsley *et al.*, 1994; Thompson *et al.*, 2002) and PE (Emsley *et al.*, 1994; Pye, 1997) enabled the initial analysis of SAP-ligand interactions. The acidic pyruvate carboxyl group of MOβDG binds directly to the two calcium ions of SAP (Emsley *et al.*, 1994): calcium I binds at a distance of 2.39 Å to the carboxyl oxygen of MOβDG, whilst calcium II binds at a distance of 2.45 Å to the second carboxyl oxygen (Thompson, 1997). Hydrogen bonds play a central role in coordinating ligands to SAP, and in particular, Gln148 and Asn59 have been identified as particularly important residues in the formation of these contacts: the amide nitrogens of Gln148 and Asn59 form hydrogen bonds with both O-4 and O-6 of the sugar ring of MOβDG (Emsley *et al.*, 1994). Gln148 is also involved in hydrogen bond formation with O3 of the sugar ring whilst Lys79 is involved in hydrogen bond formation with O1 (Thompson, 1997). Since Gln148 and Asn59 are also involved in coordinating the two calcium ions, it has therefore been hypothesised that in addition to the direct

interaction with the ligand of interest, the calcium ions may also orientate SAP residues Gln148 and Asn59 into the correct position for ligand binding.

The interaction of SAP with MOβDG is strengthened further by the positioning of the methyl group from the pyruvate moiety into a small hydrophobic pocket formed by residues Leu62, Tyr64 and Tyr74 (Thompson, 1997). In contrast to SAP, CRP binds with only weak affinity to MOβDG. Since the only major differences between the two proteins in the vicinity of MOβDG binding are the substitution of SAP-Lys79 and SAP-Tyr74 for Glu81 and Thr76 in CRP, it has been suggested that the hydrogen bond formed between SAP-Lys79 and O-1 of the sugar ring and the Van der Waals contacts formed with the hydrophobic pocket residue SAP-Tyr74, would not be available in a CRP-MOβDG complex (Thompson, 1997).

Analysis of the crystal structure of SAP bound to PE has shown the interactions between protein and ligand to be very similar to those involved in SAP binding to MOβDG. As in MOβDG it is an acidic functional group, the phosphate moiety, which is responsible for bridging the two calcium ions. Tyr64 and Tyr74 are positioned flanking the sides of the PE binding region with the aromatic side chain of Tyr64 situated below the amine group of the PE molecule. This amine group is also held in place by a hydrogen bond via a water molecule to one of the oxygen atoms of SAP-Glu66 (Pye, 1997). A previously solved low resolution structure of the SAP-PE complex (Emsley *et al.*, 1994; coordinates unavailable) had suggested that the amine group of PE was also in an appropriate position to hydrogen bond to the phenolic oxygen of SAP-Tyr74. More recent high resolution data have however, shown the average distance between the amine group and oxygen of Tyr74 to be 3.5 Å, making formation of this hydrogen bond impossible (Pye, 1997).

1.6.5.2 Binding of dAMP

The determination of the crystal structure of SAP (Emsley *et al.*, 1994) had suggested potential locations for the binding site of DNA. Co-crystallisation studies of SAP with the nucleotide ligand, 2'-deoxyadenosine-5'-monophosphate (dAMP) (Hohenester *et al.*, 1997), have however, ruled out the possibility of DNA following in the tradition of other SAP ligands, all of which interact directly with the calcium ions. In an attempt to identify novel ligands for SAP, various phosphorylated and sulphated compounds were screened and chosen based on their structural resemblance to characterised biological ligands of SAP, or because of their ability to inhibit binding of established ligands. The

mononucleotide dAMP was chosen because of its interest with respect to DNA recognition and because it was found to be as effective at inhibiting calcium-induced SAP precipitation as MOβDG and PE.

Co-crystallisation of SAP with dAMP led to the association of two pentamers to form a B-face to B-face SAP decamer. A single dAMP molecule was found positioned at each of the ten calcium binding sites in the SAP decamer. The decamer was found to be stable in solution as tested by subsequent gel filtration chromatography and electron microscopy studies (Hohenester *et al.*, 1997). Analysis of the decamer interface indicates that the only stabilising interactions between the two pentamers are mediated by the ten dAMP molecules. In particular these interactions arise from the stacking of adenine bases of adjacent dAMP molecules in between the two pentamers.

The binding of dAMP to SAP mimics that of previously identified ligands, MOβDG and PE. As in MOβDG and PE, it is an acidic functional group (phosphate ester moiety of dAMP) that is responsible for bridging the two calcium ions with calcium-oxygen distances of 2.4 Å and 2.5 Å (Hohenester *et al.*, 1997). Residues involved in hydrogen bonding to dAMP are identical to those involved in ligating MOβDG and PE. Asn59 forms a hydrogen bond with one of the phosphate oxygens of dAMP; and Tyr64 and Gln148 mediate contacts with the nucleotide sugar moiety, forming hydrogen bonds with the 2' oxygen atom of the deoxyribose ring. In the SAP-dAMP complex, residues removed from the immediate binding site, Asp145 and Ser147, provide additional hydrogen bonds to the ligand.

It was hoped that the crystal structure of SAP in complex with dAMP might provide an insight into the avid binding of DNA by SAP (Butler *et al.*, 1990). It was concluded from this study however, that the structure was unable to provide an explanation as to how this interaction might occur. A feature of DNA binding to SAP that can be concluded from the structure is that it appears to proceed at a site or sites other than the double calcium binding site: a contrast to all other known ligands of SAP (Hohenester *et al.*, 1997). A plausible location is the basic groove situated between adjacent protomers, as it is the correct size to accommodate a DNA double helix (Hohenester *et al.*, 1997).

Although unable to provide an insight into DNA binding to SAP, the dAMP mediated decamerisation of SAP was subsequently utilised as a model in the design and identification of potential inhibitors of SAP. The design of bivalent compounds able to

cross-link SAP pentamers B-face to B-face to form a decamer would abolish the availability of the double calcium binding site for ligand interaction.

1.6.5.3 SAP binding to R-1-[6-[R-2-carboxy-pyrrolidin-1-yl]-6-oxohexanoyl] pyrrolidine-2-carboxylic acid (CPHPC)

It has been hypothesised that SAP binds to amyloid fibres, thereby forming a protective coat and preventing proteolytic cleavage and possibly phagocytic-mediated clearance. A potential treatment for amyloidosis could therefore be the targeted depletion of SAP by the design of compounds that inhibit its ligand binding activity. One such compound, CPHPC (R-1-[6-[R-2-carboxy-pyrrolidin-1-yl]-6-oxohexanoyl] pyrrolidine-2-carboxylic acid (Ro 63-8695)), was identified following high throughput screening of the Roche compound library for inhibitors of SAP binding to Alzheimer's disease amyloid-β (Aβ) amyloid fibrils (Pepys *et al.*, 2002).

CPHPC consists of two D-proline residues linked by a four carbon aliphatic linker. Its palindromic nature enables it not only to block the ligand binding sites on SAP, but also cross-link SAP pentamers to form B face to B face decamers. A low resolution X-ray crystal structure of the SAP-CPHPC was originally solved to 3.3 Å and confirmed the presence of an SAP decamer reversibly cross-linked by five CPHPC molecules (Purvis, 2002; Pepys *et al.*, 2002). The carboxylate groups positioned at the opposing ends of the CPHPC molecule interact directly with the calcium binding sites, and as with the SAP-dAMP complex, there are no close contacts between the two pentameric SAP molecules. The complex is further stabilised by the highly cooperative binding of the pyrrolidine ring of CPHPC into the well-defined hydrophobic pocket of SAP, located adjacent to the two calcium ions. Since the alkyl chain that links the two proline head groups is too long to fit the electron density in its extended conformation, the linker adopts a kinked, eclipsed rotamer around the C2-C3 bond.

As the original crystal structure of the SAP-CPHPC complex was only solved at low resolution to 3.3 Å, an important feature, the identification of the proline peptide bond as either the *cis* or *trans* conformation, remained undetermined. This feature of CPHPC has important consequences for the potency of the drug *in vivo*. If the correct conformation of the peptide bond can be determined then it will be possible to 'lock' the molecule in this conformation. This is particularly important if the peptide bond is in the *cis* conformation, as from NMR studies, only 25% of the CPHPC head groups exist in this form in solution (Purvis, 2002), and therefore 75% of the drug administered will be of no use in inhibiting SAP ligand binding.

Chapter 2
X-ray Crystallographic Methods

2.1 X-ray Crystallography

The method used to determine the three-dimensional structures of ligand-complexed SAP and CRP as described in this report, was X-ray crystallography. This method of solving macromolecular structures has dramatically grown in speed and sophistication in the last 30 years and is now widely established as the most commonly used technique. In 1975 the number of protein crystal structures published totalled 78 and the number of these coordinates deposited in the protein databank (PDB) was just 18. In the year 2000, however, many more structures were published every week then were published in a whole year in 1975 (15) and the number of coordinates deposited in the PDB totalled an astonishing 13,500 (Hendrickson, 2000). This dramatic enhancement in both the use and productivity of X-ray crystallography as a method for macromolecular structure determination is the consequence of a number of factors, perhaps the most important of these being: the invention of molecular biology techniques enabling the production of recombinant protein; the advance in computer technology; and lastly the development of more intense X-ray sources specific for macromolecular crystallography.

The method involves interpreting the diffraction patterns of X-rays from a number of identical molecules in an ordered array such as a crystalline lattice. The molecules must be in a crystalline lattice as the diffraction from a single molecule would not be strong enough to measure. By placing the molecules in a lattice in which they are all in the same orientation, the diffracted waves from each molecule interfere constructively to produce a measurable resultant wave. The direction and intensity of these waves are recorded in the form of a diffraction pattern on a detector and this information is then fed to a computer. Simulating the action of a lens, the computer uses this information in combination with various mathematical functions to produce a graphical three-dimensional image of the molecule.

The majority of three-dimensional protein structures currently available have been solved using X-ray crystallography. Due to the nature of protein molecules compared to their inorganic counterparts, however, three significant problems must be overcome in order to determine crystallographically, the three-dimensional structure of a protein molecule. The first of these is the successful expression and purification of the protein in sufficient quantities to facilitate use in initial crystal trials. Second, on achieving this, crystals of the protein must be grown and must diffract to a resolution high enough to reveal atomic detail. For X-ray crystallographers this is often the rate-determining step

in protein structure solution. Finally, having completed these stages, the phases of the resultant waves need to be determined by one of three methods: molecular replacement, multiple isomorphous replacement or multi-wavelength anomalous dispersion phasing.

2.2 Protein Crystallisation

The first stage in elucidating the three dimensional structure of a protein is the growth of crystals deemed suitable for X-ray diffraction. Crystal growth results from the controlled precipitation of protein involving the slow approach to a degree of supersaturation. If the precipitation of the protein proceeds in an uncontrolled manner, for example if the crystallisation conditions are incorrect, then instead of leaving solution in the form of a crystal, the protein forms an amorphous precipitate which is of no use in protein crystallography. Supersaturation is a function of protein concentration and parameters affecting the solubility of the protein including pH, temperature and precipitants. Typically, crystals are obtained from solutions containing a precipitant such as polyethylene glycol (PEG), ammonium sulphate or 2-methyl-2,4- pentanediol (MPD) at a concentration just below that necessary to precipitate the protein.

Nucleation is the first phase in the crystallisation process and is characterised by the formation of well ordered protein aggregates. These aggregates assemble until a critical nuclear size is reached and the nuclei become stabilised. Once stable, the nuclei can then enter the growth phase proceeding via the addition of protein molecules to the crystalline lattice resulting in the formation of a large well ordered crystal. Crystal size is limited by growth defects, the presence of contaminates and the depletion of protein molecules (Ducruix, 1992).

The processes driving crystal growth are best illustrated in the form of a phase diagram separated into saturated and unsaturated regions (Figure 2.1). The solubility curve separates regions that will support crystallisation (supersaturated solutions) from regions that will not (unsaturated solutions). The region of supersaturation can be further divided into three distinct zones: precipitation, nucleation and the metastable zone. In those conditions that support precipitation, the protein partitions from solution into an amorphous precipitate that is of no use in crystallography. Conditions in the

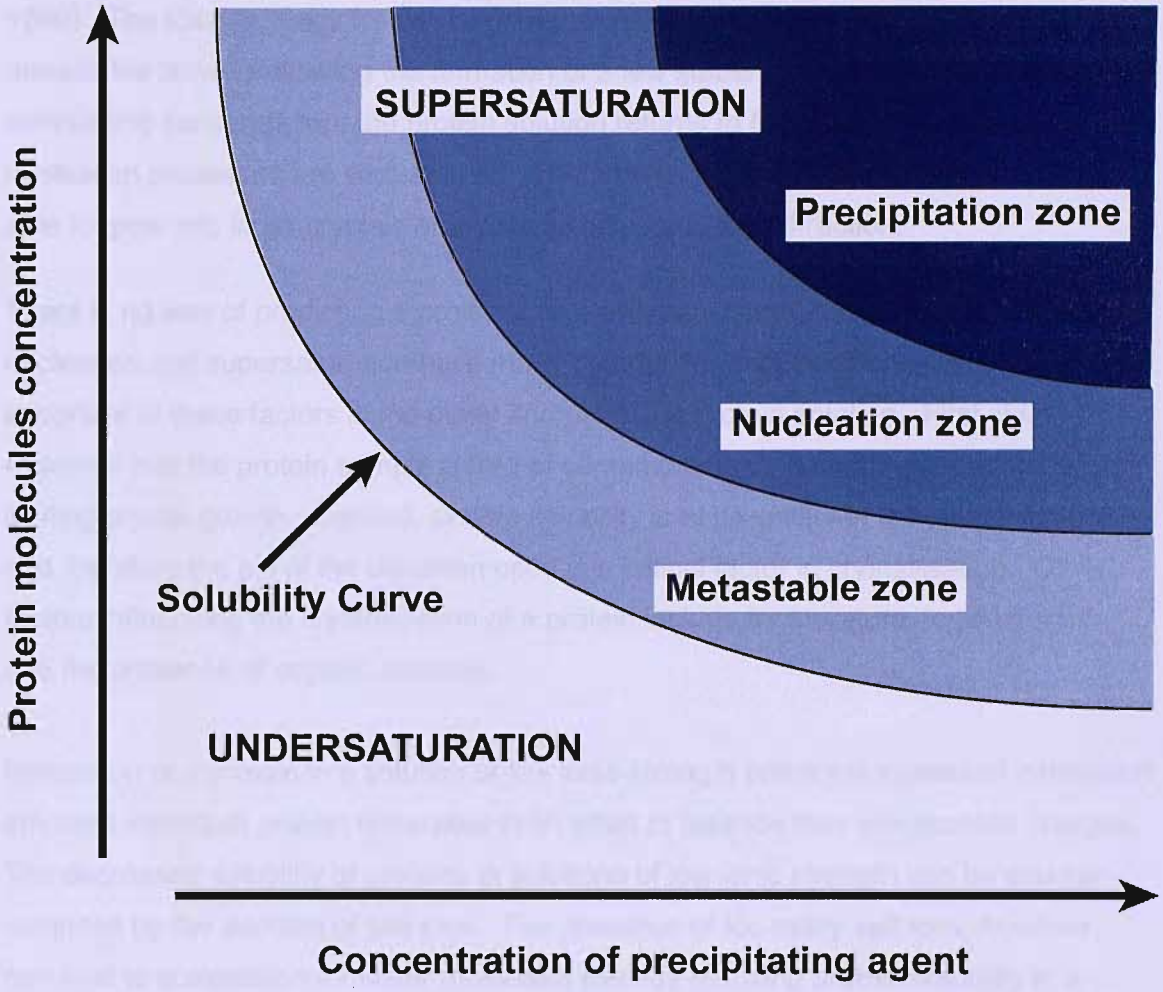


Figure 2.1 Phase diagram of the crystallisation process showing the affect of protein and salt concentration on crystal growth.

nucleation zone promote the spontaneous formation and growth of stable nuclei thereby resulting in the partition of protein from solution into the crystalline state. Finally, the metastable zone includes those conditions able to support the growth of stable nuclei but are unable to promote the initiation of spontaneous growth processes. Subsequently, protein in the metastable zone is only able to proceed to nucleation on exposure to mechanical shock or after the addition of a crystal seed (McPherson, 1990). The ideal strategy for crystal growth is for nucleation to occur just beyond the metastable zone. Following the formation of a few stable nuclei and as a result of diminishing concentration, the protein solution returns to the metastable zone and nucleation processes are sequestered. The small numbers of stable nuclei are thus able to grow into large crystals of a suitable size for X-ray diffraction.

There is no way of predicting a protein's crystallisation conditions as the processes of nucleation and supersaturation have many contributing factors. Perhaps the most important of these factors is the purity and pH of the protein solution. First, it is essential that the protein sample is free of contaminants so as not to introduce defects limiting crystal growth. Second, protein solubility is at its lowest at the isoelectric point and therefore the pH of the condition used is a critical factor in crystallisation. Other factors influencing the crystallisation of a protein include temperature, ionic strength and the presence of organic solvents.

Immersion of a protein in a solution of low ionic strength promotes increased interaction amongst individual protein molecules in an effort to balance their electrostatic charges. The decreased solubility of proteins in solutions of low ionic strength can be counter-balanced by the addition of salt ions. The presence of too many salt ions, however, can lead to competition for water molecules thereby reducing protein solubility in a process referred to as 'salting out'. Determining the correct ionic strength for the protein of interest is therefore an essential process in the quest for suitable crystallisation conditions.

The addition of organic solvents including ethanol, 2-methyl-2,4-pentane-di-ol (MPD) and polymers such as polyethylene glycol (PEG) also has a dehydrating effect on the protein solution similar to that of salt ions. A further benefit of the presence of organic solvents is their ability to reduce the dielectric constant of solutions, thereby increasing the strength of electrostatic forces between protein molecules and promoting their self-association. Substrates, coenzymes and inhibitors of the protein can also aid in the crystallisation process as they often enable the protein to adopt a more stable

conformation and can provide additional sites of contact for molecules in the crystalline lattice.

These parameters are often varied to optimise crystal growth and formation with the aid of commercial screens. Commercially available screens are available for initial crystallisation trials from a number of companies including Molecular Dimensions, Hampton and JenaBioscience. Although the strategy employed by these various screens is the same – a grid system in which pH, precipitant and salt concentration are varied, the agents utilised differ between company. The crystallisation screens of choice in this report were Molecular Dimensions and JenaBioscience.

Although the use of crystallisation screens initially provides so-called 'hits' for conditions that promote crystal formation by the target protein, the crystals produced are often too small for X-ray diffraction and therefore have to be optimised by the crystallographer. This involves screening around the 'hit' conditions and may also involve alterations to protein concentration, incubation temperature and if all else fails 'seeding'. Seeding is a technique whereby previously nucleated crystals are introduced into a solution equilibrated at a lower level of supersaturation. Since the conditions required for nucleation are vastly different to those of growth, i.e. high levels of supersaturation for nucleation and lower levels for growth, seeding represents an ideal method whereby nucleation can be uncoupled from growth and the requirements of each stage can fully satisfied (Bergfors, 2003). This technique, however, is often unreliable and is therefore only used as a last resort in the optimisation process, particularly if nucleated crystals are scarce.

The most commonly used technique to grow protein crystals is the method of vapour diffusion using a hanging drop. Purified protein, typically 2-5 µl, is mixed with an equal volume of the crystallisation solution on a pre-siliconised cover slip, giving a precipitant concentration approximately 50% of that required for crystallisation. The cover slip is placed over the well containing a volume, typically 1 ml, of the crystal screen solution and the system is sealed with grease (Figure 2.2). Since the concentration of the precipitant in the drop is less than that of the well solution, vapour diffusion occurs in the closed system and water is transferred from the drop to the solution in the well. This continues until a state of equilibrium is reached, maintaining the protein at the optimal precipitant concentration. After this the system proceeds to supersaturation: the required state for nucleation and subsequent growth of the crystal (Ducruix & Giege, 1992).

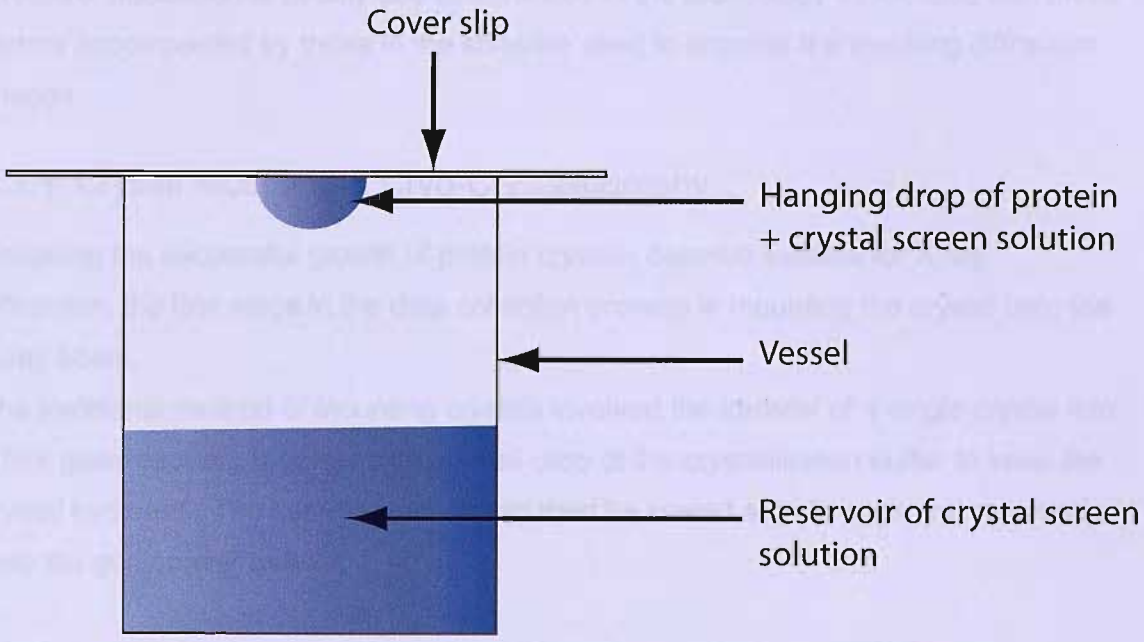


Figure 2.2 The hanging drop method of protein crystallisation.

2.3 Data Collection & Processing

The quality of the diffraction data and therefore the success of the collection process is generally assessed by the accuracy of the measured intensities and the completeness of the data set. These are reliant on a number of factors in the collection process including the amount of time available for the experiment (usually limited), the practically available resolution of the data and the resistance of the crystal to radiation damage. Recent progress in macromolecular crystallography as a technique for structure elucidation is largely due to advances in the technology associated with these factors accompanied by those in the software used to process the resulting diffraction images.

2.3.1 Crystal Mounting & Cryo-Crystallography

Following the successful growth of protein crystals deemed suitable for X-ray diffraction, the first stage in the data collection process is mounting the crystal onto the X-ray beam.
The traditional method of mounting crystals involved the transfer of a single crystal into a fine glass capillary together with a small drop of the crystallisation buffer to keep the crystal hydrated. The capillary tube would then be sealed at both ends and mounted onto the goniometer head.

Despite being highly successful, the recent development of cryo-crystallography has meant that this technique has been made mostly redundant amongst crystallographers in favour of the loop-mounting technique. The advantages of collecting X-ray diffraction data at very low temperatures such as those provided by gaseous nitrogen (boiling point -196 °C) have been known by crystallographers for many years. Theory suggests that collecting diffraction data at cryogenic temperatures will have two distinct advantages.

First, it is thought to increase the molecular order of crystals, thereby improving diffraction. Second, compared to collection at room temperature, cryo-temperatures are thought to reduce radiation damage thereby increasing the durability of the crystal (Haas & Rossman, 1970; Petsko, 1975; Garman, 1999). At cryo-temperatures, free radicals produced from incident radiation are thought to have slower diffusion rates culminating in decreased motility through the crystal and thus limiting radiation damage to the component protein molecules (Garman, 2003).

Developing practical methods by which crystals can be successfully frozen has proved more difficult. Early attempts at freezing crystals were highly unsuccessful due to the formation of ice crystals that disrupted the crystalline lattice (Garman & Schneider, 1997). Recent developments have included flash freezing methods in the presence of an ice-preventing agent, or use of a cryo-protectant such as glycerol. In this method, the crystal is taken from the drop and is immersed in the desired cryo-protectant for 5-15 seconds in order to wash off the crystallisation buffer. Often the crystallisation buffer already contains a suitable cryo-protectant such as a low molecular weight PEG or MPD and preparation of the crystal simply involves determining the optimal concentration of these agents to prevent ice formation. The crystal is then removed from the cryo-protectant using a glass wool or nylon fibre loop of suitable size for the target crystal, where it remains immersed in a thin film of the cryo-protectant. The loop is then dipped into liquid nitrogen after which if the correct concentration of cryo-protectant has been used, the liquid film in the loop vitrifies into a glass and remains clear. For the duration of data collection, the loop is then transferred to a goniometer where it can be vitrified in a stream of nitrogen gas at a temperature of 100 K.

2.3.2 X-ray Sources & Detectors

X-rays are a form of electromagnetic radiation of wavelengths 0.1-100 Å. Since an average bond length is typically 1.5 Å, X-rays in the range of 0.6 - 3.0 Å are ideal for use in atomic scale structure determination as their relatively short wavelengths enable them to interact with, and be diffracted by the electron clouds surrounding atoms. In contrast to this, visible light has a relatively long wavelength of approximately 600 nm and therefore is not scattered by electrons in this manner. As a result of this, light microscopes are only able to resolve detail to approximately 200 nm.

There are three types of X-ray source: X-ray tubes, rotating anodes and particle storage rings producing radiation in the X-ray region. X-ray tubes and rotating anodes generate their X-rays by bombarding a metal target, most commonly copper, with a beam of electrons. These electrons are produced by heating a filament (cathode) and are accelerated by electrically charged plates so as to promote collision with the anode made of the target metal. Following collision with one of these high-energy electrons, an electron from a low-lying orbital in a target metal atom is displaced creating a vacancy in this orbital. This is then filled by an electron from a higher orbital dropping into the vacancy and in doing so emitting its excess energy as an X-ray photon.

Each metal target produces a characteristic emission spectrum the peaks of which indicate the intensity of the emission at each given wavelength. The emission spectrum of copper consists of two peaks with the nomenclature K_α and K_β corresponding to the orbital that the electrons fall from. Electrons dropping from the L shell to replace displaced K electrons (K_α transition) emit X-rays of wavelength 1.54 Å; the M to K transition produces the K_β peak at a wavelength of 1.39 Å. Generation of a monochromatic X-ray beam is an essential requirement for the collection of interpretable X-ray diffraction data and it is therefore necessary to employ filtering devices or 'monochromators' in order to remove the unwanted radiation (Whittaker, 1981).

In recent years, electron synchrotrons or particle storage rings have emerged as the most popular of X-ray sources for two reasons. First, the X-rays produced are both more powerful and more intense than those of a rotating anode source and can therefore be used to produce diffraction data to a better quality and higher resolution (Lindley, 1999). Second, the wavelength of the X-ray source is tunable, a characteristic which has become highly desirable with the increase in MAD/SAD experiments. A rise in the number of laboratories participating in X-ray diffraction experiments due to the now widely accepted importance of structural information to molecular biology has spawned the development of a multitude of new synchrotron beamlines. As a result there are now 50 active beam lines located worldwide which are dedicated to macromolecular crystallography.

A synchrotron is a machine that accelerates electrons or positrons to velocities of more than 99% of the speed of light. Once orbiting they are maintained in this state under high vacuum in 'storage rings'. The particles are driven by energy from radio-frequency transmitters and sustained in a circular motion by powerful magnets. As the electrons are maintained in this circular motion around the synchrotron ring, they emit an intense X-ray beam. Additional bending of this beam, culminating in an increase in radiation intensity, can be achieved via the use of accessory devices including wigglers. These act to deflect the particle beam from side to side via the poles from a number of magnets, emitting a wide horizontal fan of X-rays that can feed several work stations located around the ring at any one time. Prior to exposure of the X-rays to the protein crystal, the beam must be optimised by a number of optical devices installed between the source and sample; of these devices monochromators and mirrors are two of the most commonly used. A monochromator at a synchrotron consists of a single crystal of silicon, germanium or carbon (diamond), the X-ray diffraction from which

selects a single wavelength from the source spectrum. A mirror is then used to focus the beam onto the protein crystal.

As the strength of X-ray beams increased dramatically with the development of third generation synchrotrons, the pressure to improve detector characteristics also increased. With exposure times falling from hours to seconds, this also led to the requirement for enhanced detector readout speed. In third generation synchrotrons today, the most commonly used detectors to measure diffracted X-rays are charge coupled devices (CCDs). These represent an enhancement both in terms of count rate performance and read-out time. CCDs are fundamentally photon counters: coated in an energy converter they accumulate charge in direct proportion to the amount of light that strikes them (Westbrook & Naday, 1997). For the purpose of crystallography, the CCD is coated with phosphors emitting visible light when exposed to impinging diffracted X-rays. The visible light photons are transferred to the CCD via a tapered bundle of optical fibres where they are converted into charge carriers and stored in the constituent pixels of the CCD. When an entire diffraction pattern has been recorded the stored charge accumulated in the pixels is converted into a digital signal for subsequent processing on a computer.

There are two main types of X-ray camera used in the collection of diffraction data: the precession camera and the oscillation/rotation camera. Although the precession camera produces the simplest diffraction patterns it has recently been surpassed by the more efficient oscillation method. The major difference between these techniques is that in precession photography the crystal, screen, and film are moved in a precessing motion about the X-ray beam. In contrast, the oscillation method involves rotating the crystal about an axis perpendicular to the X-ray beam and oscillating it back and forth by a few degrees whilst the detector remains fixed. As a result of these differences, the precession camera reveals an undistorted image of the reciprocal lattice whilst the oscillation camera generates distorted images of the lattice. Each oscillation image is composed of a pattern of concentric, nearly circular regions called lunes. In comparison to precession photography in which the entire image is of a single plane in reciprocal space, in oscillation photography each lune arises from a single plane of reciprocal space. Traditionally, the processing of precession diffraction data was much more straightforward than those acquired via oscillation methods due to the lack of distortion in the images. The advance and virtual automation of data processing techniques has however, made interpretation of oscillation images much

easier and has led to this technique being the method of choice for the majority of diffraction experiments.

X-ray diffraction data for both of the structures reported here were collected at the third generation synchrotron at the ESRF in Grenoble, France on beamlines ID14-1 and ID23-2. ID 14-1 provides a monochromatic X-ray beam with a fixed wavelength of 0.934 Å. ID23 is composed of two independent end-stations: ID23-1 and ID23-2. ID23-1 is a fully tunable beamline that can be used for MAD experiments, whilst ID23-2 is a fixed energy beamline of wavelength 0.873 Å. The detectors present in the workstations of both ID23-2 and ID14-1 are CCD, however, ID23-2 utilises a MARCCD detector whilst ID14-1 utilises an ADSC CCD detector.

2.3.3 Data Processing

The object of data processing is to produce a data set or reflection file (.mtz) consisting of the Miller indices (hkl) of all reflections recorded by the diffraction experiment, together with an estimate of each of their intensities (I_{hkl}) and their standard uncertainties (σI_{hkl}). This requires a prediction of not only which diffraction image a reflection will appear on (as typically reflections are recorded over several images and therefore will only be partially recorded on any one image), but also the precise location of each reflection on a particular image. The process can be divided into three stages: autoindexing, post-refinement and integration.

Autoindexing is the first stage in data processing and involves determining the unit cell parameters, point group and orientation of the crystal with respect to the incident X-ray beam. A unit cell is described as the smallest and simplest volume element that is completely representative of the whole crystal. The unit cell dimensions can be defined by the lengths of three edges a, b and c; and three angles α, β and γ (Figure 2.3). The value of these dimensions describes the morphology of the unit cell adopting one of seven crystal systems: triclinic, monoclinic, orthorhombic, trigonal, tetragonal, hexagonal or cubic.

The unit cell can be subdivided into asymmetric units – the smallest unit of the crystal that can generate the complete crystal structure by means of its symmetry operations. The symmetry of a unit cell, i.e. the arrangement of asymmetric units, is described by its space group. Each space group is represented by a symbol such as $P2_12_12_1$, the letter at the beginning representing the lattice type and the numbers representing

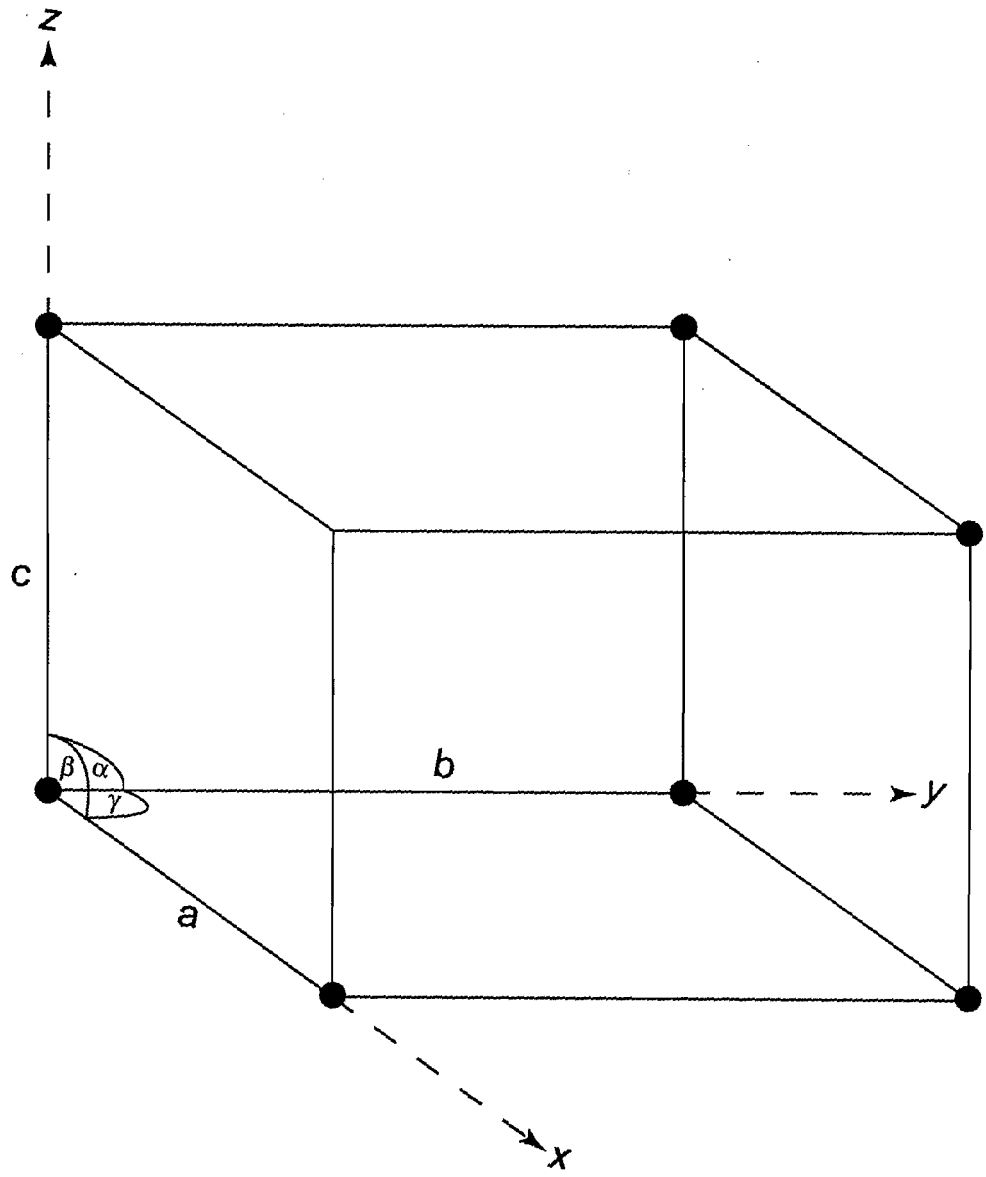


Figure 2.3 A unit cell.

operations defining the symmetry of the cell. Ideally the lattice type with the smallest possible volume, a primitive (P) lattice is chosen. This type of lattice contains one lattice point at each corner around which the asymmetric units are arranged. Other lattice types include: C, F, I or R. Lattices of these types are said to be face-centred or body-centred, containing in addition to the lattice points at the corners of the cell, extra points in the centre of one face (C), at the centre of all faces (F) or at the centre of the cell (I). Combination of these five types of centering (P, C, F, I and R) together with the seven crystal systems, gives rise to 14 Bravais lattices, shown in Figure 2.4.

Following the auto-indexing process, the initial estimates of the cell parameters need to be refined in a process termed post-refinement. The process requires the integration of one or two segments of data each containing two or three diffraction images. The third stage is the integration of all the collected diffraction images and simultaneous refinement of crystal and detector parameters to generate the full list of reflections (hkl) and associated intensities (I_{hkl}).

There are a number of programs currently being used to process X-ray diffraction data including: Mosflm (Leslie, 1992), XDS (Kabsch, 1988), HKL2000 and Denzo (Otwinowski, 1997). The basic principle of these programs is similar but the methods employed by each are subtly different. The program of choice in this report was Mosflm. To begin, Mosflm requires the construction of an input file containing information about the data collection experiment. In this file are details of the detector type, crystal-to-detector distance, wavelength of the X-ray source, beam coordinates and location of the diffraction images.

The first task of the auto-indexing function in Mosflm is to select a number of reflections on the diffraction image to be included in the auto-indexing process. The process works most reliably with the inclusion of 200+ reflections per image but typically >100 reflections are sufficient (Leslie, 1999). Reflections are chosen based on an I/σI cutoff value as chosen by the user (default = 20). The diffraction patterns produced by an oscillation experiment represent a distorted projection of the *reciprocal* lattice of the crystal. The reciprocal lattice of the diffraction pattern is related to the crystal, or *real* lattice, by an inverse relationship. Within this reciprocal lattice exists a reciprocal unit cell. When the lattice angles are all 90 the a* axis of the reciprocal cell lies along a; axis b* along b; and axis c* lying along c. The lengths of a*, b* and c* are the reciprocal of the real unit cell i.e. a* = 1/a and their units are 1/Å.

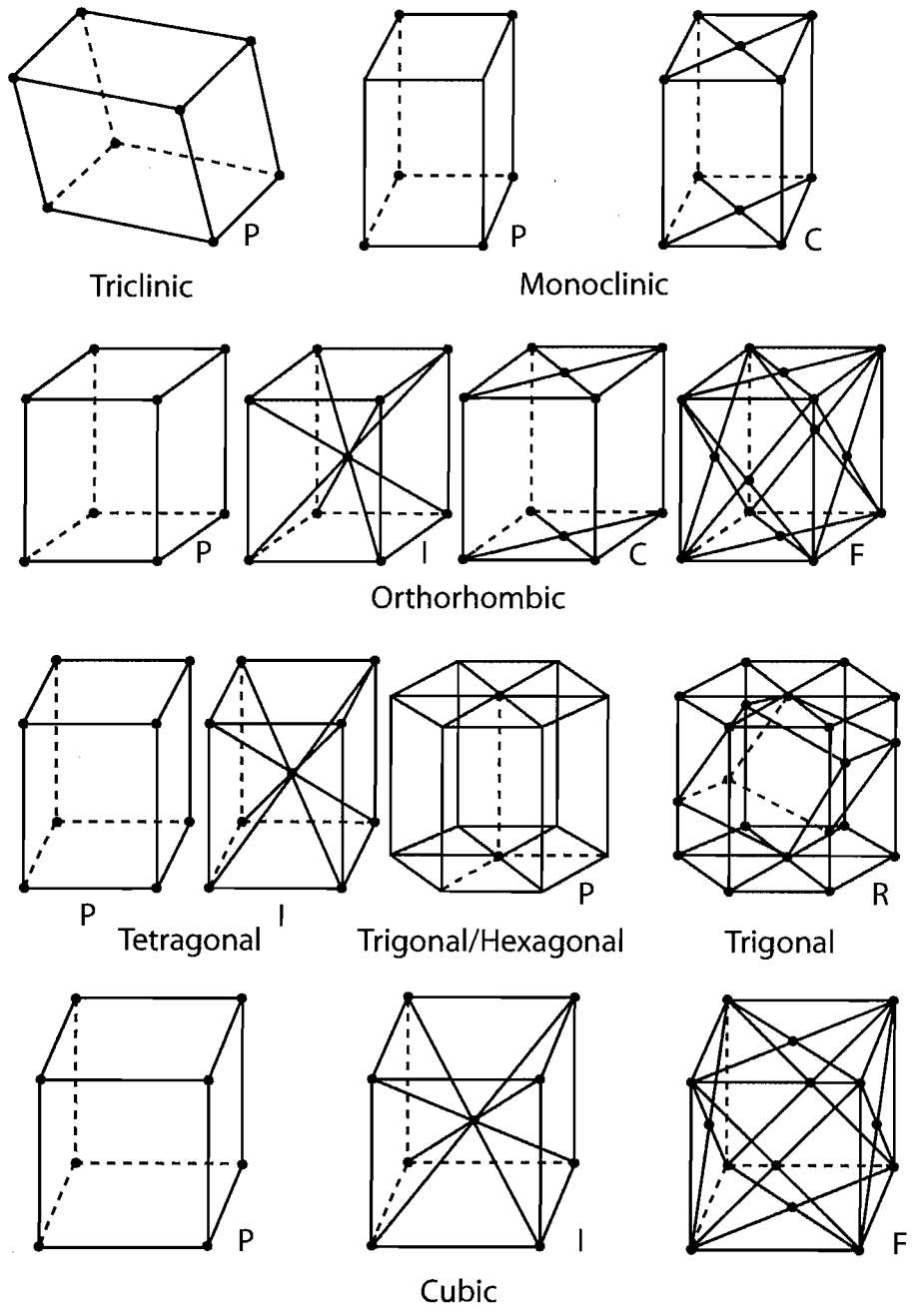


Figure 2.4 The 14 Bravais lattice types.

The autoindexing process in Mosflm works by mapping the chosen reflections on the diffraction image back to their equivalent reciprocal-lattice vectors in reciprocal space. From these vectors it is then possible to estimate the real unit cell spacing. The result of this process will be a basic triclinic cell, void of the true lattice symmetry. Subsequently, orientation matrices corresponding to all possible lattice types are successively positioned over the diffraction pattern to assess the best fit. The possible solutions are displayed in the form of a table with an associated penalty – a measure of the cell distortion required to overlay diffraction images. Typically, the solution with the highest symmetry and lowest penalty is chosen.

Following post-refinement of the unit cell dimensions and integration of all the diffraction images, the next stage is the scaling of intensity data. Various programs are available for the scaling of data but the program of choice in this report was Scala (CCP4, 1994). Depending on the symmetry of the crystal, diffraction images will contain so-called symmetry-related reflections. Ideally, as predicted by Friedel's Law, these reflections should be identical in intensity. Due to various physical factors in the experiment including variability in X-ray beam intensity, however, radiation damage to the crystal, crystal slippage, and error in space group assignment, the intensities of these reflections may not be consistent. During the scaling process symmetry-related reflections are compared and from these a scale factor is calculated and applied to the data to create uniformity in intensities across all images. Analysis of the agreement between equivalent reflections after scaling can be used to assess the overall quality of the data and identify parts of the data that are to be discarded.

There are a number of values from the scaling process that can be used to assess the quality of the diffraction data. Arguably, R_{merge} is the most important of these values and is a measure of the discrepancy in intensity value between symmetry-related reflections. If the intensity measurements between these pairs are accurately reproducible R_{merge} will be relatively small (typically <10%); higher values of R_{merge} (20%>) indicate poor data or incorrect space group assignment. Other statistics used to measure the quality of the data include the average I/σI, multiplicity and completeness. These merged intensities are then converted to structure factor amplitudes using the program Truncate (French & Wilson, 1978), the product of which is a reflection file (.mtz) that can be used for the phasing process.

Following processing through Truncate, the reflection file can be entered into the program HKLVIEW for the identification of higher symmetry elements. It is essential for

phase determination that the correct space group be chosen. The auto-indexing process may suggest P222 as a possible solution, however it is unable to determine whether one, two or all three of these axis are also subject to a translational symmetry operation or *screw* axis, and will therefore list P222, $P2_122$, $P2_12_12$ and $P2_12_12_1$ as possible solutions. In a space group such as $P2_1$, the presence of the screw means that not only does the axis rotate 180° to generate the 2-fold symmetry, but it also translates half a cell along in the direction of the axis to which the screw is parallel. Certain symmetry elements such as the screw axis, reveal themselves in the diffraction pattern by causing specific reflections to be missing. For example a two-fold screw axis along the b axis will result in all reflections *0k0* having odd values of *k* to be missing. These systematic absences can be viewed with the program HKLVIEW and can confirm that the correct space group has been chosen.

2.4 Phase Determination

2.4.1 The 'Phase Problem'

The next step in the process of structure determination is the extraction of structural information from the list of reflection intensities. The goal of this process is a function describing the electron density surrounding an average molecule in the unit cell, $\rho_{x,y,z}$.

$$\rho(x,y,z) = \frac{1}{V}\sum_h\sum_k\sum_l F_{hkl}e^{-2\pi i(hx+ky+lz)}$$

The crystallographer can obtain the electron density in the unit cell by constructing a Fourier series using structure factors F_{hkl}. The structure factor is itself a periodic function describing a complicated wave, consisting of a frequency, a phase, and an amplitude. These parameters are available from the diffraction data: the frequency is that of the X-ray source; the amplitude is equal to $(I_{hkl})^{1/2}$ the square root of the measured intensity for reflection hkl, and is directly measurable from the diffraction pattern. Unfortunately the phase is not directly measurable from the diffraction pattern but is essential in the calculation of electron density maps and determination of positions of atoms within the protein molecule. The absence of phase information in the diffraction data has meant that macromolecular crystallographers have had to devise methods by which the so-called 'phase problem' can be overcome. The three most commonly used phasing techniques are multiple isomorphous replacement (MIR),

single/multi-wavelength anomalous dispersion (MAD/SAD), and molecular replacement (MR). The structures presented in this report were solved using the phasing technique of MR.

2.4.2 Molecular Replacement

The method of MR involves using a previously solved protein structure (search model) to determine the orientation and position of the new protein molecule (target) in the unit cell. These calculated phases can then be applied to the target molecule in order to solve the structure of the new protein. The technique of molecular replacement requires that the search and target molecules share a reasonable amount of sequence identity (approximately > 25%). The result of molecular replacement is generally a preliminary solution which may require further optimisation using rigid body refinement. Since structures solved by molecular replacement are vulnerable to model bias, they can be entered into cycles of refinement, map calculation and model building, thereby improving model accuracy and reducing model bias.

The method of molecular replacement fundamentally involves determining a rotation matrix (*C*) and a translation vector (*d*) to apply to the coordinates of the search model *x*:

$$[C]x + d = x'$$

(*x*' = target structure)

This method generally requires a six dimensional search involving three rotation and three translation parameters. In order to save computing time and thus improve efficiency, the procedure is separated into two stages. First, the model is rotated without considering its crystallographic symmetry mates until the best agreement with the observed data (diffraction data) is obtained. Second, once the correct orientation has been found, it is translated within the parameters imposed by the crystallographic symmetry to obtain the best agreement with the diffraction data.

There are a number of molecular replacement programs available to the crystallographer including Molrep (Vagin & Teplyakov, 1997), X-PLOR/CNS (Brunger, 1992) and Amore (Navaza, 1994). These programs vary in both the type of algorithm used for the six dimensional search and in their relative CPU times. There are two commonly used approaches to the rotation search: the 'traditional rotation search' and

the 'direct rotation search'. Both of these methods fundamentally involve the comparison of a Patterson function calculated for both the search model and the observed diffraction data (Figure 2.5).

A Patterson function (Patterson, 1934) is the Fourier transform of the experimental intensities and requires no phase information. The Patterson function can be expressed as follows:

$$P_{(u,v,w)} = 1/V \sum_h \sum_k \sum_l \left|F_{hkl}\right|^2 e^{-2\pi i(hu+kv+lw)}$$

The Patterson map that is generated by the function consists of a number or peaks corresponding to vectors between atoms. A simple two-dimensional Patterson map can be constructed by drawing all possible vectors between atoms in one unit cell (self-vectors). Each atom is positioned at an origin, and vectors with their tails located at the origin are drawn between this atom and all other atoms in the unit cell. The head of each vector is the location of a peak in a Patterson map. In Patterson space the translation vector, d, is irrelevant since all self-vectors are shifted to the origin. Therefore, the resulting map is independent of the position of the molecule within the unit cell and enables the rotation function to be carried out separately from the translation function.

A basic conceptual distinction can be made to explain the fundamental differences between the traditional and direct methods of a rotation function. In the traditional rotation search, the Patterson maps produced for the search model and observed data are rotated with respect to each other and then superimposed. This is in comparison with the direct method in which the search model itself is rotated and structure factors are recomputed for each sampled orientation of the model with respect to the observed data (Grosse-Kunstleve & Adams, 2001). Another fundamental difference between the two methods is the type of unit cell in which the search model is placed. In a traditional rotation search, the search model is typically placed in an artificial orthogonal unit cell. Therefore, in order to minimise the number of inter-molecular vectors used in the rotation function a radius of integration needs to be specified by the crystallographer. However, this inevitably leads to the exclusion of a small number of intra-molecular vectors which can be detrimental to the search. The value chosen for the radius of integration is normally decided after a period of trial and error but suggested values have included 75-80% of the diameter of the protein

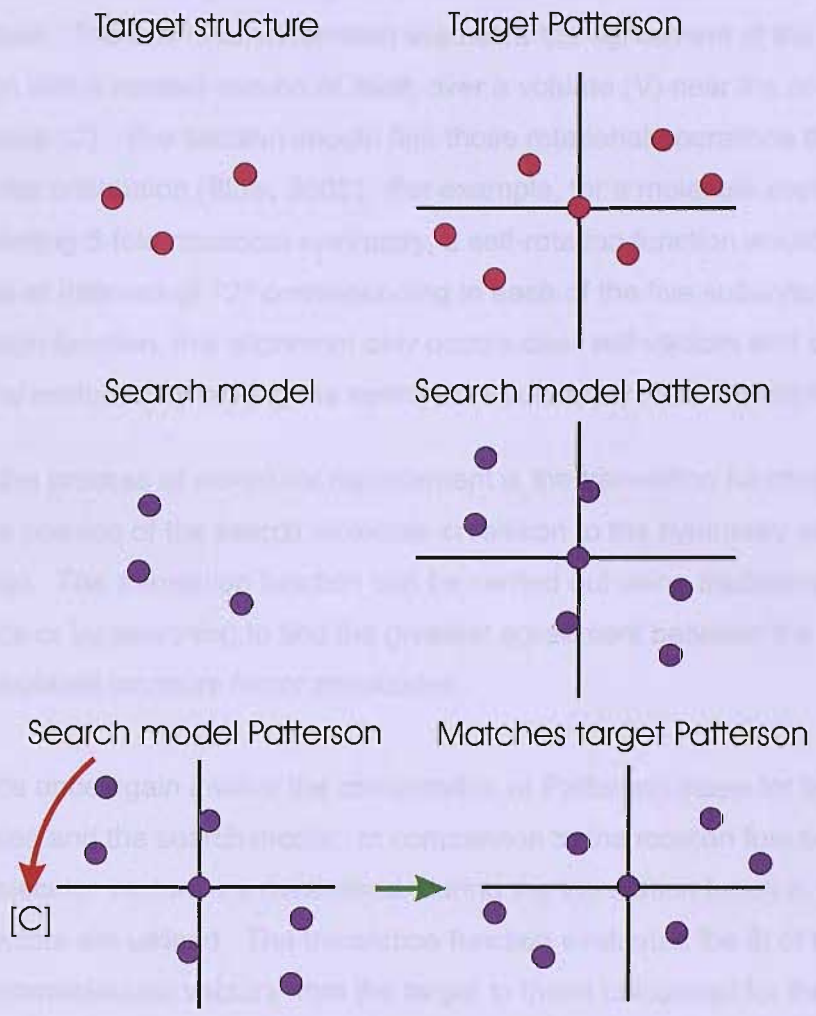


Figure 2.5 Simplified diagram showing the construction of a Patterson map. The Patterson map of the search model is related to that of the target by [C].

A self-rotation function involves generating a Patterson function for the molecule and rotating it upon itself. The self-rotation function evaluates the agreement of the Patterson function with a rotated version of itself, over a volume (V) near the origin, for all possible rotations (C). The function should find those rotational operations that align subunits to a similar orientation (Blow, 2002). For example, for a molecule such as CRP or SAP exhibiting 5-fold rotational symmetry, a self-rotation function would expect to generate peaks at intervals of 72° corresponding to each of the five subunits. As with a cross-rotation function, this alignment only occurs over self-vectors and cross-vectors need to be excluded by limiting the search to a volume around the origin.

The next step in the process of molecular replacement is the translation function which aims to define the position of the search molecule in relation to the symmetry elements of the target crystal. The translation function can be carried out using traditional Patterson methods or by searching to find the greatest agreement between the observed and calculated structure factor amplitudes.

Patterson methods once again involve the computation of Patterson maps for both the observed intensities and the search model. In comparison to the rotation function in which the intramolecular vectors are maximised, during the translation function the intermolecular vectors are utilised. The translation function evaluates the fit of the configuration of intermolecular vectors from the target to those calculated for the oriented search molecule for all possible displacement vectors. The Patterson maps are placed over each other until the configuration of intermolecular vectors from each map fit match each other. The vector through which the configuration has been displaced, from its starting position to the position of best fit, determines the location of the model's origin with respect to the rotational axis in the crystal.

An R factor search can also be used to evaluate the translation function. The R factor is calculated by monitoring the correspondence between the expected structure factor amplitudes from the model in a given trial location and the actual amplitudes derived from the observed data. The R factor compares the agreement between the two sets of amplitudes as follows:

$$R = \frac{\sum \left| |F_{obs}| - |F_{calc}| \right|}{\sum |F_{obs}|}$$

For each reflection the difference between the observed structure factors F_{obs}, and calculated amplitudes from the model in the trial location F_{calc} can be calculated. The magnitudes of the difference for each reflection are then added together and divided by the sum of F_{obs}. If this agreement is close then the sum of the differences will be smaller compared to the sum of the intensities themselves, and R is therefore small. A perfect agreement between the observed and calculated structure factors will result in a value of zero for the R factor. A value of between 0.3 and 0.4 is respectable for a molecular replacement solution.

The correlation coefficient is an alternative parameter used to assess the success of the translation function and like the R factor search measures the agreement between observed and calculated structure factors:

$$CC = \frac{\left\langle |F_o||F_c| - \langle |F_o| \rangle \langle |F_c| \rangle \right\rangle}{\left(\left\langle |F_o|^2 - \langle |F_o| \rangle^2 \right\rangle \left\langle |F_c|^2 - \langle |F_c| \rangle^2 \right\rangle \right)^{1/2}}$$

In contrast to the R factor search, the higher the correlation coefficient value the better the solution found.

The programs of choice in this report were Molrep and CNS. Molrep can be used as an automated program in which it automatically chooses the required parameters, selects the solution and presents the model correctly oriented and positioned in the unit cell. Alternatively, if the user requires more input, the rotation and translation functions can be performed separately using parameters specified by the crystallographer. Molrep essentially employs a traditional rotation function followed by a translation search evaluated by monitoring the R factor and correlation coefficient. Since its development in 1997 (Vagin & Teplyakov, 1997) Molrep has grown in popularity, primarily as a consequence of vastly reduced CPU times when compared to CNS and minimised demands on computational memory as compared to Amore.

An additional feature of the program is the implementation of a multi-copy search to the translation function. This may be of particular use for oligomeric structures such as SAP and CRP and as a result was utilised in this report. The method involves using a monomer such as a CRP or SAP protomer as the initial search model. Since SAP and CRP are pentamers, each one of the protomers in the asymmetric unit must be

positioned with respect to the same origin. Therefore, following the first translation search the position of the first molecule can be fixed at the origin allowing the position of the second protomer to be positioned with respect to the same origin. This so-called dyad can then be used as the search model for the subsequent search of the third monomer (Vagin & Taplyakov, 2000). The process is repeated until the positions of all monomers are located.

In order for a successful MR solution to be obtained, it is essential that the orientation found is as close to the correct one as possible before it is entered into the TF. Although programs such as Molrep and Amore are computationally more efficient, it has been suggested that this may be at the cost of the accuracy of the rotation function (DeLano & Brunger, 1995). As a result, for more difficult problems the MR package provided by CNS is perhaps considered a more effective approach.

The MR program implemented in CNS offers the user the choice of a traditional rotation search or a direct rotation search. Traditional rotation searches involve the direct comparison of Patterson maps for the observed data and search model, a function that is evaluated in CNS by the term:

$$Rot(\Omega) = \int_{u} P_{obs}(u) P_{model}(\Omega_u) du$$

where, $Rot(\Omega)$ = *Patterson correlation*
Ω = *orientation*
P_{obs} = *observed Patterson*
P_{model} = *model Patterson*
u = *location vector in Patterson space*

The direct rotation search in CNS is evaluated by the calculation of a Patterson correlation coefficient. Patterson correlation (PC) is the linear correlation coefficient between observed and calculated squared normalised structure factors ($|E|^2$) and is a measure of the phase accuracy of a partial atomic model (Hauptman, 1982). PC correlates observed and calculated structure factors where the latter are calculated for each sampled orientation of the search model in a P1 unit cell with dimensions identical to the target cell. The term used to evaluate the correlation coefficient is:

$$CC(\Omega) = \frac{\sum_{H} \left(X_{H,obs} - \langle X_{obs} \rangle\right)\left(X_{H,\Omega} - \langle X_{\Omega} \rangle\right)}{\left[\sum_{H} \left(X_{H,obs} - \langle X_{obs} \rangle\right)^2\right]^{1/2} \left[\sum_{H} \left(X_{H,\Omega} - \langle X_{\Omega} \rangle\right)^2\right]^{1/2}}$$

In CNS a powerful technique of PC refinement is inserted in between the rotation and translation searches therefore enabling the improvement of the orientation of the search model. During the procedure, a large number of rotation peaks are chosen (it is assumed that correct orientation will be present in the selected peaks). These solutions are then refined with respect to a number of parameters that are expected to be dissimilar between the observed data and the search model i.e. large rigid groups, domains, subdomains or secondary structural elements. PC has been shown to improve the discrimation between correct and incorrect orientations, therefore enabling location of a correct peak in a noisy rotation function. Compared to other MR programs such as Amore, CNS has a higher signal-to-noise ratio and it has been shown that a rotation-function peak that is within 15 ° of the correct orientation will converge to it following PC refinement (DeLano & Brunger, 1995). This is therefore of particular use in solutions where the search model is only a partial model for the target crystal structure.

2.5 Refinement & Model Building

The process of model building and structure refinement is iterative, continually switching between the real space of the electron density maps and the reciprocal space of the structure factors. In real space, the improvement and removal of errors from the electron density is performed in the fitting of the model to the map or in the adherence of the model to expected bond lengths and angles. In reciprocal space, the improvement and removal of errors involves assessing the reliability of the phases and agreement of the calculated structure factors with measured intensities. The Fourier transform permits the continual shift between real and reciprocal space in an attempt to nurture the model into agreement with the structure.

2.5.1 Structure Refinement

The purpose of structure refinement is to produce the best fit of the structure to the experimental observations. The progression of the refinement procedure is followed by measuring the discrepancies between the calculated structure factors (from the model

structure) and the observed structure factors (from the diffraction data). This is presented in the form of the refinement parameter or R factor. The purpose of refinement processes are to alter atomic positions (x,y,z) in the model to improve the fit of the calculated structure factors to the observed structure factors, thereby reducing the R factor:

$$R = \frac{\sum \left[|F_{obs}| \right] - \left[|F_{calc}| \right]}{\sum \left[|F_{obs}| \right]}$$

The value of the R factor will fall towards zero as the agreement between the observed and calculated structure factor amplitudes improves. The final value however, depends on a combination of the degree of order in the crystals, and the resolution and quality of the diffraction data.

The parameters of each atom included in the process of refinement include: occupancy, temperature factor (B-factor) and atomic position. The refinement of protein molecules can be particularly problematic in comparison with their inorganic counterparts. This is due to the large number of atoms present in a protein structure resulting in a poor observations to parameters ratio. In order to combat this problem, a number of constraints and restraints can be employed in the refinement process (Konnert, 1976). A constraint is a fixed value for a certain parameter, for example if the occupancy of a particular atom is set or constrained to 1.0 during refinement. In comparison to this, with restrained refinement, the parameter imposed on a particular atom is flexible e.g. all bond lengths and bond angles have to be within a specified range of values. Ideal values for these parameters are obtained from the average values of bond lengths and angles of previously solved small organic molecules. The inclusion of these restraints and constraints to the stereochemical information of the protein structure is equivalent to increasing the number of observations and also helps to maintain the correct stereochemistry of the model.

The level of constraint or restraint applied to a protein structure during refinement is dependent upon the resolution and quality of data. As the resolution of the data decreases the number of recorded reflections also decreases and therefore there are not enough observations to permit the refinement of all atomic parameters individually. At a relatively low resolution the number of observations would permit refinement of the main-chain conformational angles (torsion angles) only, therefore generating two

variables per amino acid: φ and ψ. At a slightly better resolution, side-chain conformational angles χ, can also be refined but with the component bond angles, bond lengths, and planarities tightly restrained and a single B factor calculated for the whole structure. This will generate, on average, four variables per amino acid. Further improvement of the resolution enables the relaxation of α-carbon atoms around which there is more flexibility, and calculation of two separate B factors for each amino acid (one for the main-chain atoms and the second for the side-chain atoms) generating three more variables per amino acid. At a resolution of approximately 2.5 Å separate B factors can be calculated for every atom; at approximately 2.0 Å cartesian angles can be refined for each atom; and finally at about 1.5 Å resolution anisotropic B factors showing the density distribution of individual atoms and the particular direction in which they are free to move, are added as parameters.

There are many programs available to carry out protein structure refinement, three of the most commonly used are: SHELXL (Sheldrick & Schneider, 1997), CNS (Brunger *et al.*, 1998) and Refmac (Murshodov *et al.*, 1997). The program SHELXL can be used on structures with a resolution of 2.5 Å or better and uses traditional least-squares methods to carry out the refinement process. In crystallography, the parameters of interest are the positions of all atoms (x,y,z) in the model that best fit the observed structure factor amplitudes. These positions are used to calculate structure factors and therefore can be used to compute expected structure factor amplitudes, F_{calc}, for the current model. The goal of refinement is to find a set of atom positions that give F_{calc} as close as possible to the F_{obs}. The method of least squares finds the model parameters, including atomic positions, B factors and occupancies, that minimise the sum of the squares of the differences between corresponding F_{calc} and F_{obs} (Rhodes, 1993). The difference between the observed amplitude F_{obs} and the measured amplitude F_{calc} for reflection hkl can be defined as $(F_{obs} - F_{calc})_{hkl}$. Least squares methods seek to minimise the function Φ, where

$$\phi = \sum_{hkl} w_{hkl} \left(\left| F_o - F_c \right| \right)^2_{hkl}$$

The function Φ is the sum of the squares of differences between observed and calculated amplitudes taken over all reflections hkl. Each difference is weighted by the term w_{hkl}, the value of which depends on reliability of the corresponding observed intensity.

One major problem associated with least squares methods is that the complicated function Φ exhibits many local minima corresponding to variations in model conformation all of which may minimise Φ to varying degrees. With least squares refinement the local minimum closest to the starting point is located and therefore it is essential that the starting model parameters are as close to the global minimum (the conformation that gives the highest agreement between F_{obs} and F_{calc}) as possible. If this is not the case then the refinement will converge to an incorrect local minimum from where it will be unable to extract itself. The maximum distance from the global minimum from which the refinement can converge is calculated by $d_{min}/4$ and is called the radius of convergence. d_{min} corresponds to the lattice-plane spacing of the reflection with the highest resolution, and therefore although adding higher resolution data will provide more detail on the overall structure of the protein it will also decrease the radius of convergence. The consequence of this is that on adding the higher resolution terms, the model must be ever closer to the global minimum.

There are various ways of increasing the radius of convergence, thereby increasing the probability of finding the global minimum. The first of these is to use the various constraints and restraints described earlier. The function Φ can be described with addition of bond angle and bond length restraints to the refinement process. The effect of these restraints is to penalise adjustments to parameters that make the model less realistic. The use of least squares methods is limited as they only permit the process of refinement to proceed in an energetically downward fashion. This is acceptable for structures in a conformation that is close to the global minimum at the beginning of refinement, as normally the process only needs to progress in a downward direction to achieve the desired global minimum. A problem occurs if the starting point of refinement is not close enough to the global minimum and becomes stuck in a local minimum. Here, it is unable to pass the requested energy barriers without manual intervention in the form of model building; or the addition of energy via the method of simulating annealing as implemented in the X-PLOR/CNS package (Brunger *et al.*, 1997). Both of these methods allow the refinement to converge to the global minimum.

The progress of the refinement process has traditionally been monitored by following the fall of the R factor. This, however, was criticised by Brunger (1992) who suggested that bias may be introduced into the model when the same data is used for both monitoring and refinement. In a situation of low data to parameter ratio, typical of protein crystallography, the R factor can be lowered by refining an increased number of parameters without improving the reality of the model. To combat this problem, a small

fraction of reflections (typically 5%) are selected and deleted from the data used for refinement. These reflections, the free R or test set, are used to calculate an R_{free} factor:

$$R_{free} = \frac{\sum_{test} \left| |F_{obs}| - |F_{calc}| \right|}{\sum_{test} |F_{obs}|}$$

Although this represents a loss of experimental data, the great advantage of calculating an R_{free} is that these observations are not available to the refinement procedure and improvements to the R_{free} value therefore reflect genuine improvements to the model, rather than overfitting to noise which does not adequately define the model.

2.5.2 Model Building

Manual model building is commonly used to cross the energy barriers of refinement, particularly if the model remains trapped in a local minimum. The requirement for this method is the calculation of electron density, $\rho_{(x,y,z)}$, so that the model can be rebuilt into the density wherever necessary. Typically, two types of electron density map are computed: a $2F_o$-F_c map and a F_o-F_c map. The F_o-F_c map provides information about the new model and depending on which of F_o or F_c is larger, the resulting Fourier term can be either positive or negative. Positive density in a region of the map indicates that the contribution of the observed intensities (F_o) to ρ is greater than that of the calculated structure factors (F_c). This suggests that the unit cell contains more electron density in this region than implied by the model (F_c) and requires the movement of atoms towards this area in order to increase the electron density. In contrast, negative density indicates that the model implies more electron density in the region than the unit cell actually contains and atoms must be moved away from this area. As the F_o-F_c map on its own can be difficult to interpret, a more interpretable map, the $2F_o$-F_c, is also computed. In calculating this map, the influence of F_c are still reduced but to a lesser extent than the F_o-F_c map. If errors in the model are minimal, this map is completely positive and depicts a molecular surface covering all the atoms of the molecule. Initial electron density maps frequently contain many errors due to the inaccurate estimation of phases used to calculate them. There are various methods used to reduce the noise from maps thereby making them more interpretable: NCS averaging (Bricogne, 1974), solvent flattening (Wang, 1985) and histogram matching (Zhang & Main, 1990). From the modified function $\rho_{(x,y,z)}$, predicted values for phases can be

calculated and can be combined with F_{obs} to produce a new electron density distribution, which in turn can be modified.

Chapter 3
Structural Studies of Serum Amyloid P Component in Complex with CPHPC

3.1 Introduction to SAP - CPHPC Complex

SAP is found associated with all types of amyloid and comprises approximately 14% of the total mass of the amyloid deposit. The role of SAP in the pathogenesis of amyloidosis is thought not to be in the formation of these deposits but in the maintenance of their stability. It has been postulated that SAP binds to amyloid fibres thereby forming a protective coat and preventing proteolytic cleavage and possibly phagocyte-mediated clearance. Therefore, it has been suggested that a treatment for amyloidosis could be the targeted depletion of SAP by the design of compounds that inhibit SAP-ligand binding.

The X-ray crystal structure of SAP was first solved by Emsley *et al.* (1994) and revealed a pentameric molecule composed of five identical subunits or protomers arranged around a central hollow pore. A host of co-crystallisation studies of SAP with various ligands followed, including natural ligands such as PE and MOβDG and synthetic ligands including dAMP, *D*-proline, *L*-proline and methyl-malonic acid. It is these co-crystallisation studies and in particular the SAP-dAMP structure that led to the design and synthesis of a potential inhibitory compound for SAP-ligand binding.

Co-crystallisation of SAP with dAMP revealed a B-face to B-face decamer formed via stacking of the adenine bases in between the two pentamers. The arrangement of SAP pentamers in this manner with their binding sites inaccessible to incoming ligand, provided an ideal model for the design of cross-linking compounds that would result in decamerisation of SAP molecules. In addition to inhibiting ligand binding to the B face of SAP molecules, it was also hoped that the formation of an unnatural decameric complex would result in the clearance of SAP by the immune system at sites of amyloidosis.

To begin the search for a potential drug molecule for SAP a high throughput assay for inhibitors of SAP binding to immobilised Alzheimer's disease amyloid-β (Aβ) amyloid fibrils was used to screen the Roche compound library. The most potent of these identified compounds was CPHPC (R-1-[6-[R-2-carboxy-pyrrolidin-1-yl]-6-oxohexanoyl] pyrrolidine-2-carboxylic acid (Ro 63-8695)) consisting of two *D*-proline residues linked by an aliphatic linker (Figure 3.1). The palindromic nature of CPHPC enables it not only to block the ligand binding sites on SAP but also to cross-link SAP pentamers to form B face to B face decamers. A low resolution X-ray crystal structure of the SAP-

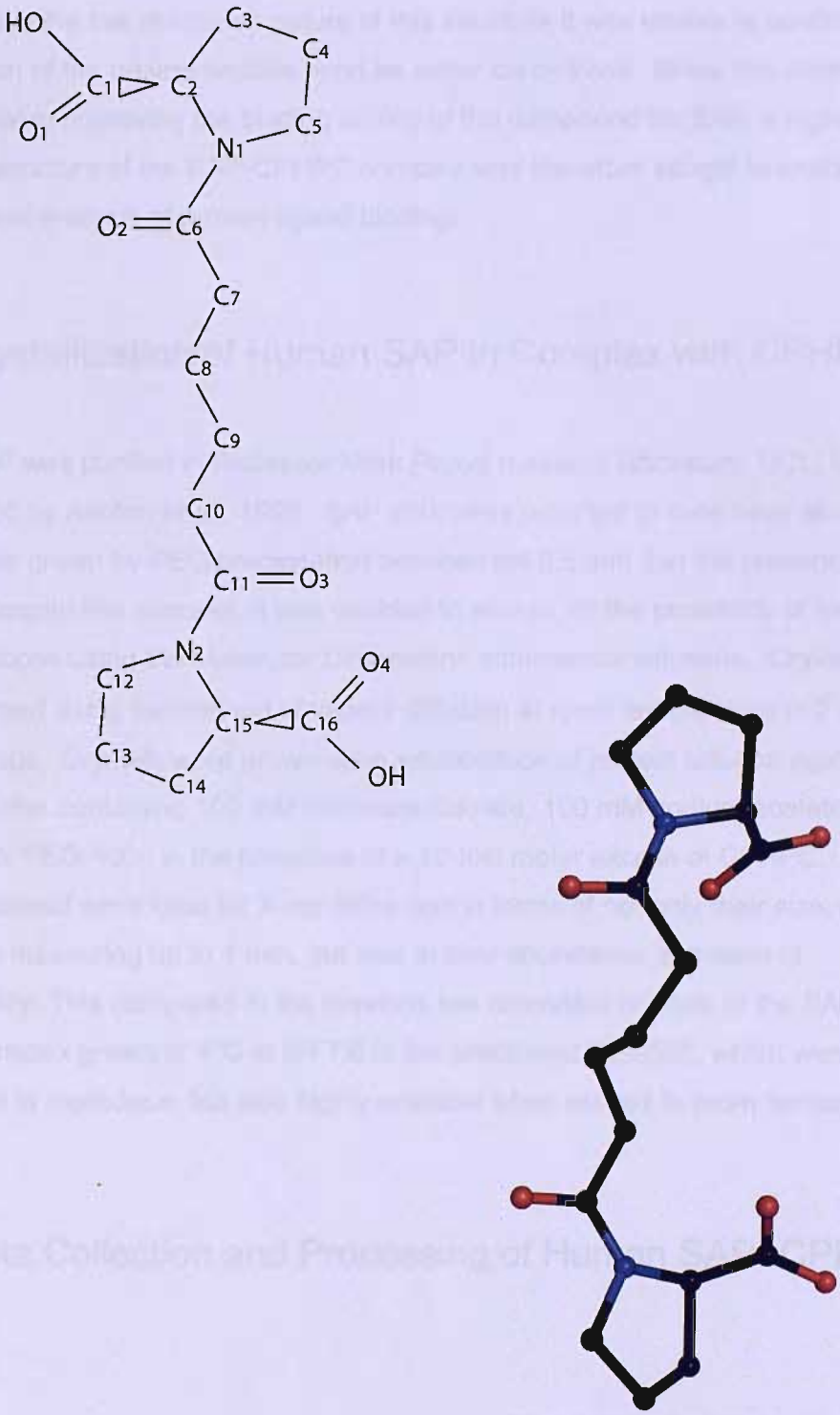


Figure 3.1 The structure of CPHPC. **Top left** Chemdraw representation of CPHPC. **Bottom right** Pymol representation of CPHPC.

CPHPC complex was originally solved to 3.3 Å and confirmed the presence of an SAP decamer reversibly cross-linked by five CPHPC molecules (Pepys *et al.*, 2002). However, due the low resolution nature of this structure it was unable to confirm the conformation of the proline peptide bond as either *cis* or *trans*. Since this information could be vital in improving the binding affinity of the compound for SAP, a higher resolution structure of the SAP-CPHPC complex was therefore sought to enable a more detailed analysis of protein-ligand binding.

3.2 Crystallisation of Human SAP in Complex with CPHPC

Human SAP was purified in Professor Mark Pepys research laboratory, UCL, London, as described by Ashton *et al.*, 1997. SAP structures reported to date have all resulted from crystals grown by PEG precipitation between pH 5.5 and 8 in the presence of calcium. Despite this success, it was decided to screen for the possibility of identifying novel conditions using the Molecular Dimensions commercial solutions. Crystallisation was performed using the method of vapour diffusion at room temperature in 2 μl hanging drops. Crystals were grown upon equilibration of protein solution against reservoir buffer containing 100 mM cadmium chloride, 100 mM sodium acetate pH 4.6 and 30% v/v PEG 4000 in the presence of a 10-fold molar excess of CPHPC. The crystals obtained were ideal for X-ray diffraction in terms of not only their size, with dimensions measuring up to 1 mm, but also in their abundance and ease of reproducibility. This compared to the previous low resolution crystals of the SAP-CPHPC complex grown at 4°C at pH 7.6 in the precipitant PEG550, which were not only difficult to reproduce, but also highly unstable when moved to room temperature.

3.3 Data Collection and Processing of Human SAP- CPHPC to 1.6 Å

Crystals were flash-cooled in liquid nitrogen using 30% (v/v) glycerol as the cryo-protectant. Diffraction to ~2.0 Å resolution was observed on the in-house rotating anode X-ray source. Diffraction data were collected to 1.6 Å on BM 14.1 (λ = 0.934 Å) using an ADSC detector at ESRF, Grenoble, France (Figure 3.2). Two separate data sets were collected: a low resolution set (81-1.9 Å) and a high resolution set (81-1.6 Å). A total of 90 1° oscillation frames were measured, with an exposure time of 10 seconds

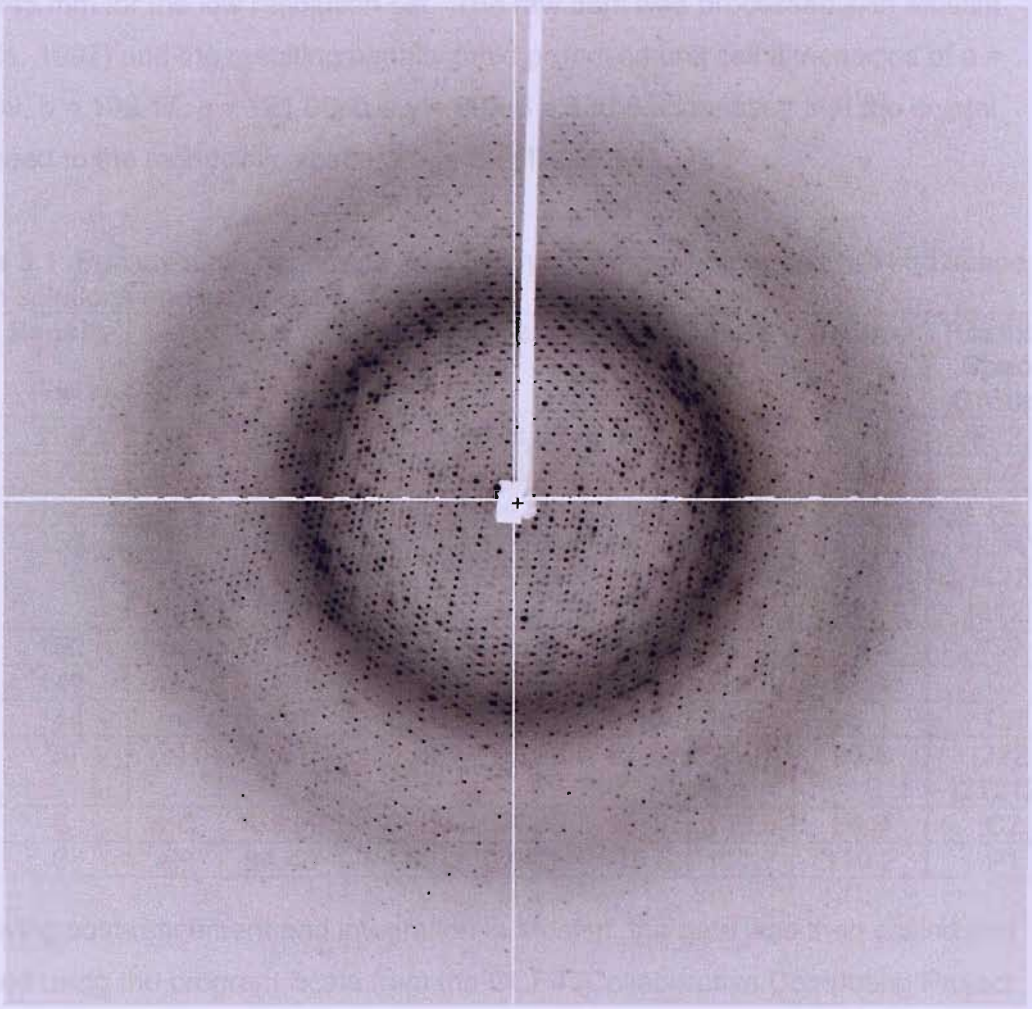


Figure 3.2 A diffraction image from the SAP-CPHPC crystal to 1.6 Å.

per frame and a crystal to detector distance of 135 mm for the high resolution data set and 245 mm for the low resolution set. The raw data was processed with Mosflm (Leslie, 1997) and the resulting penalty table proposed unit cell dimensions of a = 155.09, b = 108.17, c = 121.05, $\alpha = \gamma = 90°$, $\beta = 138.8$ suggesting that the crystal belonged to the monoclinic space group C2 (Table 3.1).

Table 3.1 Penalty table generated by Mosflm showing possible unit cell and space group solutions and associated penalties.

No.	Penalty	Latt.	a	b	c	alpha	beta	gamma	Possible Space Groups
10	209	oF	121.05	149.10	148.58	93.2	88.3	88.4	F222
9	172	mC	155.09	108.17	102.23	89.8	128.8	90.2	C2
8	162	aP	102.23	94.40	94.69	69.8	59.0	59.2	P1
7	157	tI	102.23	108.17	121.05	89.9	87.6	90.2	I4, I41, I422, I4122
6	152	mC	121.05	149.10	97.23	88.5	130.2	88.4	C2
5	149	mC	121.05	149.10	94.40	93.6	128.1	88.4	C2
4	144	mI	94.40	149.10	97.23	91.5	101.7	86.4	C2
3	59	oI	108.17	121.05	102.23	87.6	89.8	90.1	I222, I212121
2	**8**	**mC**	**155.09**	**108.17**	**121.05**	**90.1**	**138.8**	**89.8**	**C2**
1	0	aP	94.40	94.69	97.23	115.6	101.7	110.2	P1

Following post-refinement and integration in Mosflm, the data was then scaled and merged using the program Scala from the CCP4 (Collaborative Computing Project Number 4, 1994) suite. The scaled and merged intensities generated from Scala were converted to structure factor amplitudes using Truncate (CCP4, 1994). The data statistics are shown in Table 3.2. The choice of space group was confirmed after analysis of the data first in P1 to confirm the presence of a 2-fold axis and then in P2 to confirm the presence of systematic absences for reflections h+k = odd (see Figure 3.3).

3.4 Molecular Replacement Studies of Human SAP-CPHPC to 1.6 Å

3.4.1 Rotation Search

Prior to the rotation search, solvent content calculations were performed (Matthews, 1968) to obtain the number of molecules per asymmetric unit of the SAP-CPHPC crystal. The calculation clearly identified the presence of a single pentamer in the

asymmetric unit with a solvent content of 63 %. The equivalent calculation for two pentamers per asymmetric unit was 7.99 %.

Table 3.2 Data Processing Statistics for the SAP-CPHPC crystal.

Parameter	**Value**
Space group	C2
Unit cell	a = 154.31Å, b = 108.12Å, c = 120.28Å α = 90°, β = 138.40°, γ = 90°
Resolution range (Å)	80-1.6
Measured reflections	2019227
Unique reflections	171493
Multiplicity	5.0
Completeness (%)	99.3
Rmerge (outer resolution shell) (%)	9.8 (22.5)
Mean (I)/sd(I) (outer resolution shell)	3.9 (3.2)
Solvent Content (%)	63

Initial phases were obtained by molecular replacement calculations using Molrep. The search model used in the rotation search of the 1.6 Å data was an SAP pentamer from the refined structure of human SAP to 2.0 Å with the calcium atoms omitted (accession code 1SAC; Emsley *et al.*, 1994). The cross rotation search was performed using reflections within the resolution range 80.56 Å to 3.0 Å and yielded five significant solutions (Table 3.3):

Table 3.3 The ten highest peaks following cross rotation calculations performed using data between 80 Å and 3 Å (α, β, γ correspond to CCP4 Eulerian angles).

Peak Number	α	β	γ	Rf/σ
1	324.90	0.00	34.90	20.87
2	285.32	65.98	105.34	19.49
3	74.57	66.01	254.46	19.49
4	220.38	56.50	319.59	19.20
5	139.68	56.48	40.28	19.12
6	170.29	44.98	10.33	3.55
7	276.82	40.57	99.18	3.37
8	114.71	90.00	65.56	3.35
9	240.26	77.28	300.70	3.34
10	86.17	25.89	263.75	3.34

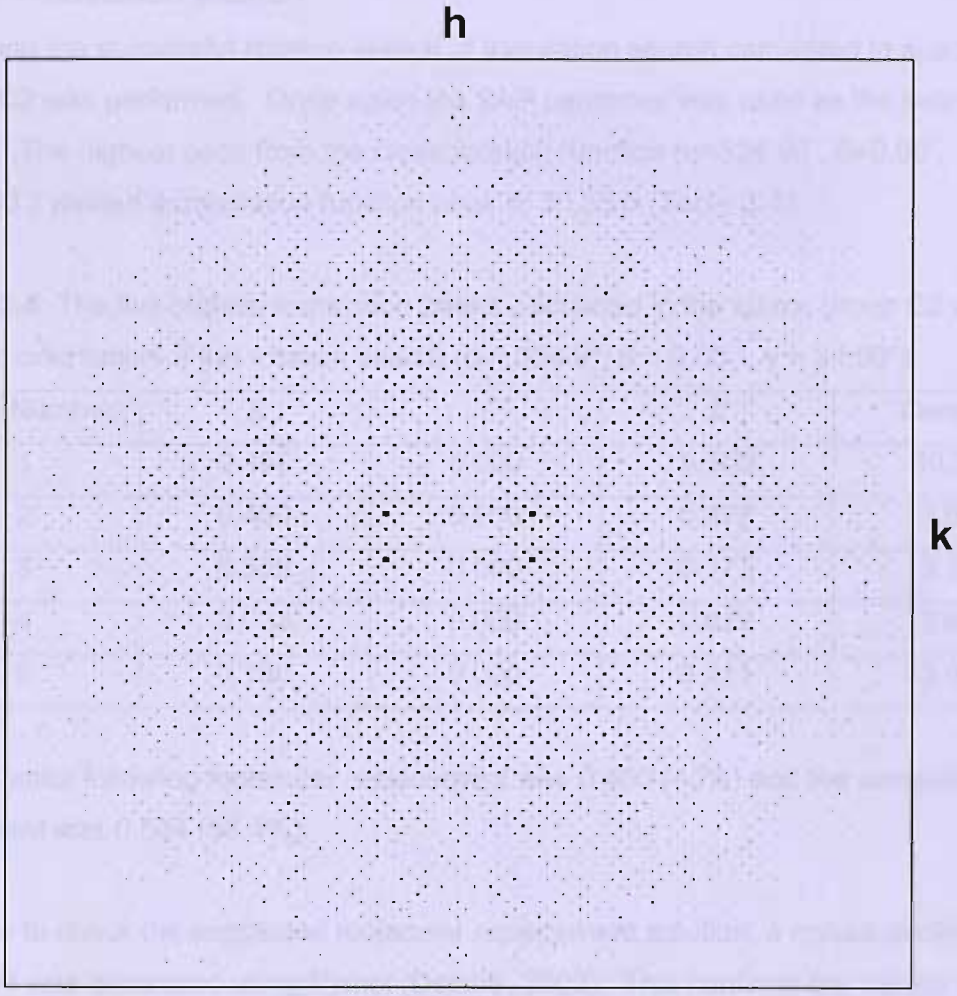


Figure 3.3 Precession camera image of the SAP-CPHPC diffraction data generated using HKLVIEW.

3.4.2 Translation Search

Following the successful rotation search, a translation search calculated in space group C2 was performed. Once again the SAP pentamer was used as the search model. The highest peak from the cross rotation function (α=324.90°, β=0.00°, γ=34.90°) yielded a translation function peak of 30.35 σ (Table 3.4).

Table 3.4 The five highest translation peaks calculated in the space group C2 using the first orientation of the rotation search (α = 324.9°, β = 0.00 °, γ = 34.90°).

Peak Number	X	Y	Z	Dens/σ
1	0.498	0.000	0.500	30.35
2	0.490	0.000	0.472	3.92
3	0.309	0.000	0.173	3.15
4	0.033	0.000	0.022	3.09
5	0.180	0.000	0.411	3.02

The R factor following molecular replacement was 0.400 (40%) and the correlation coefficient was 0.584 (58.4%).

In order to check the suggested molecular replacement solution, a crystal packing diagram was generated using Pymol (Delano, 2002). This confirms the validity of the presented solution, indicating that the asymmetric unit contains only a single pentamer and subsequent SAP decamers are formed utilising crystallographic symmetry (Figure 3.4).

3.5 Refinement & Model Building of SAP-CPHPC at 1.6 Å

The model was refined with the program CNS using 172501 reflections in the resolution range 81.6-1.6 Å. The test set of reflections used for the R_{free} calculation (5% of reflections) was chosen using the CCP4 program freerflag. The first round of refinement consisted of rigid body, minimisation and individual B factor refinement. The R factor after the initial round of rigid body refinement was 29.5% (R_{free} = 32.5%) and decreased to 28.7% (R_{free} = 31.4%) after minimisation and 28.5% (R_{free} = 30.4%) after group B factor refinement.

Following the first round of refinement, the resulting model was used to calculate sigmaA weighted 2*Fo*-*Fc* and *Fo*-*Fc* electron density maps. The maps were calculated

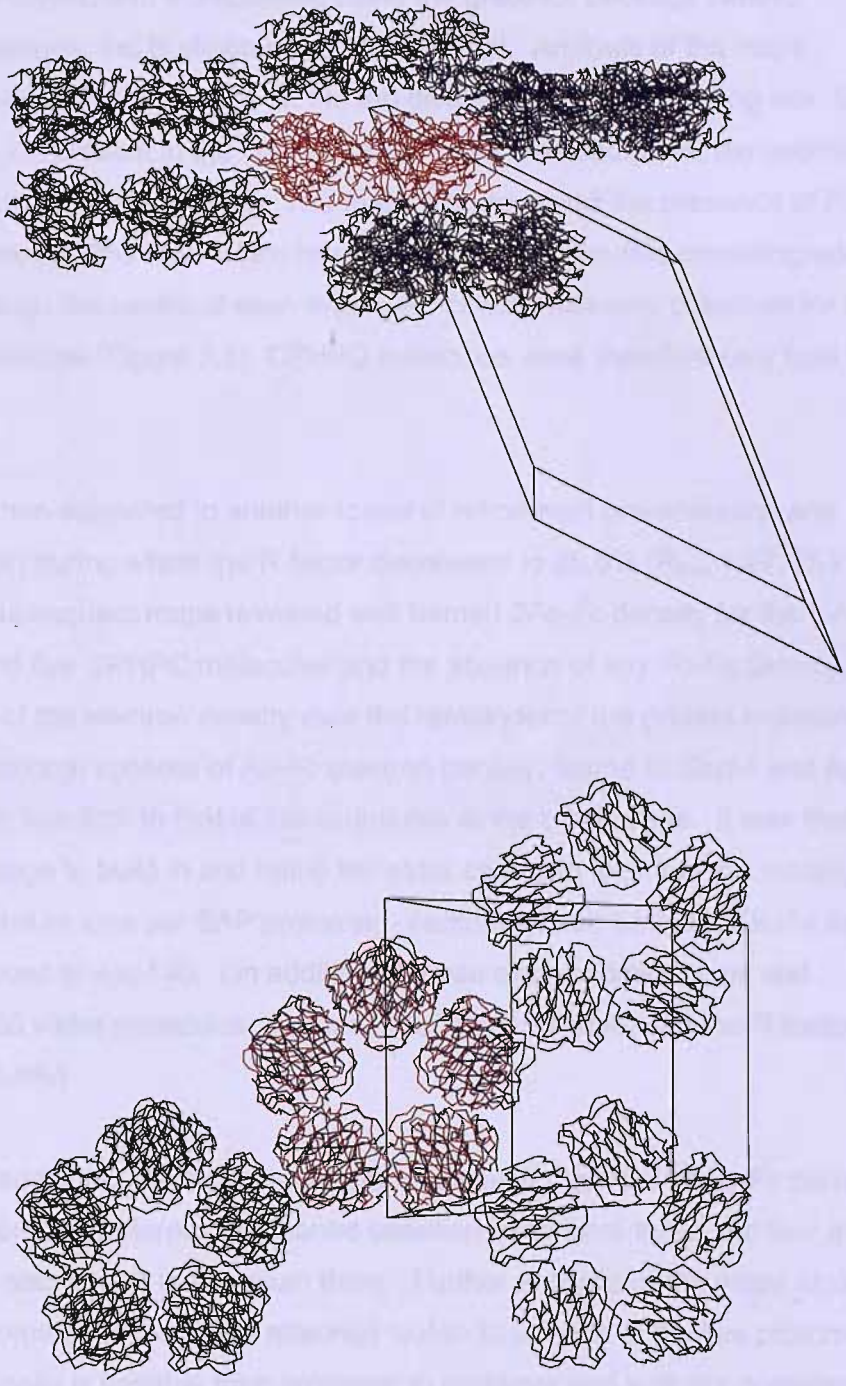


Figure 3.4 Packing diagram for the SAP-CPHPC crystal. **Top** Side view of the unit cell showing a single pentamer highlighted in red with its symmetry-related pentamer positioned above the 2-fold axis. **Bottom** View down the 5-fold axis of the decamer.

using the CNS program model_map.inp and were converted into X-PLOR format using the CCP4 program Mapman (Kleywegt & Jones, 1994). The maps generated by Mapman were analysed and manipulated using the graphics package Quanta (Molecular Simulations, Inc Burlington, Massachusetts). Analysis of the maps indicated well-defined *Fo-Fc* density for the ten divalent ions at the binding site. Since the only divalent ion present in the crystallisation buffer was cadmium, ten cadmium ions were built into the *Fo-Fc* density. The maps also revealed the presence of *Fo-Fc* density for the five CPHPC molecules; however, due to the two-fold crystallographic axis running through the centre of each molecule, density was only observed for half of each CPHPC molecule (Figure 3.5). CPHPC molecules were therefore only built in up to C8.

The model was then subjected to another round of refinement (minimisation and individual B factor) during which the R factor decreased to 25.6% (R_{free} = 27.4%). Analysis of the subsequent maps revealed well formed *2Fo-Fc* density for the ten cadmium ions and five CPHPC molecules and the absence of any *Fo-Fc* density. Further analysis of the electron density over the remainder of the protein molecule also revealed two additional spheres of *Fo-Fc* electron density, bound to Glu14 and Asp145 of each protomer, identical to that of the cadmiums at the binding site. It was therefore decided at this stage to build in and refine ten extra cadmium ions into the model, totalling four cadmium ions per SAP protomer - cadmium three bound to Glu14 and cadmium four bound to Asp145. On addition of these extra cadmium ions and approximately 700 water molecules, a round of refinement decreased the R factor to 24.3% (R_{free} = 25.9%).

Following these additions, the subsequent maps revealed a sphere of *Fo-Fc* density for a fifth cadmium ion per protomer, positioned between cadmiums three and four and water molecules also bound to cadmium three. Further analysis of the maps also revealed *Fo-Fc* density for the sugar attached to Asn32 in each of the five protomers. However, this density is variable from protomer to protomer and was not considered good enough in any of the protomers to build in to. A fourth round of minimisation and individual B factor refinement produced well presented *2Fo-Fc* density for the fifth cadmium ion. A subsequent water search resulted in the addition of approximately 300 extra water molecules, totalling approximately 1000 water molecules for the structure. The resulting R factor after a fifth round of refinement was 21.32% (R_{free} = 22.66%). The refinement statistics for the model are shown in Table 3.5.

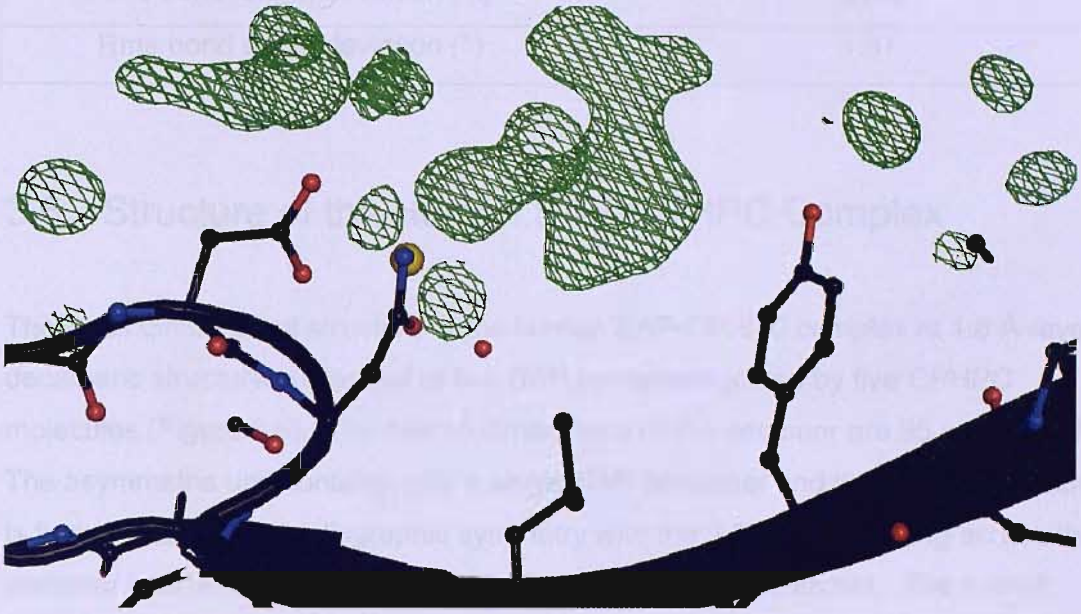


Figure 3.5 Electron density (Fo-Fc) contoured at 3σ of the calcium binding site before the addition of CPHPC. The density clearly shows the presence of the proline ring and carboxylic acid group of CPHPC.

Table 3.5 Refinement statistics for the SAP-CPHPC structure.

Resolution range (Å)	80 – 1.6
R-factor (%)	21.32
R-free (%)	22.66
Number of reflections in the working set	162477
Number of reflections in the test set	8737
Rms bond length deviation (Å)	0.005
Rms bond angle deviation (°)	1.37

3.6 Structure of the Human SAP-CPHPC Complex

The three dimensional structure of the human SAP-CPHPC complex at 1.6 Å reveals a decameric structure composed of two SAP pentamers joined by five CPHPC molecules (Figure 3.6). The overall dimensions of the decamer are 95 x 95 x 73 Å. The asymmetric unit contains only a single SAP pentamer and therefore the decamer is formed utilising crystallographic symmetry with the 2-fold axis running across the decamer interface perpendicular to the 5-fold axis of the pentamer. The overall structure of the SAP-CPHPC complex reported here is identical to the previously reported 3.2 Å description of the complex. However, due to the increase in resolution, the 1.6 Å structure is able to provide a more detailed description of the complex. Of particular importance has been the elucidation of the peptide bond conformation in CPHPC, information that may enable an increase in drug affinity.

The structure of the SAP-CPHPC complex at 1.6 Å shows good stereo-chemical properties with 86.8 % of residues in the most favoured regions of the Ramachandran plot, 13.2 % in the additional allowed regions and 0 % in disallowed regions, see Figure 3.7. A comparison of the temperature factors for the five SAP subunits shows a good B-factor distribution and an overall B-factor value of 18.12 $Å^2$. As shown in Figure 3.8 the biggest difference in B factor distribution across the five protomers is observed between residues 160-180. Subunits A and B have considerably lower B factor values across this residue range when compared to subunits C, D and E. In subunits A and B these residues are involved in extensive crystal contacts to neighbouring molecules. In contrast, in subunits C, D and E they are not involved in such interactions thereby explaining their greater mobility and increased B-factor values.

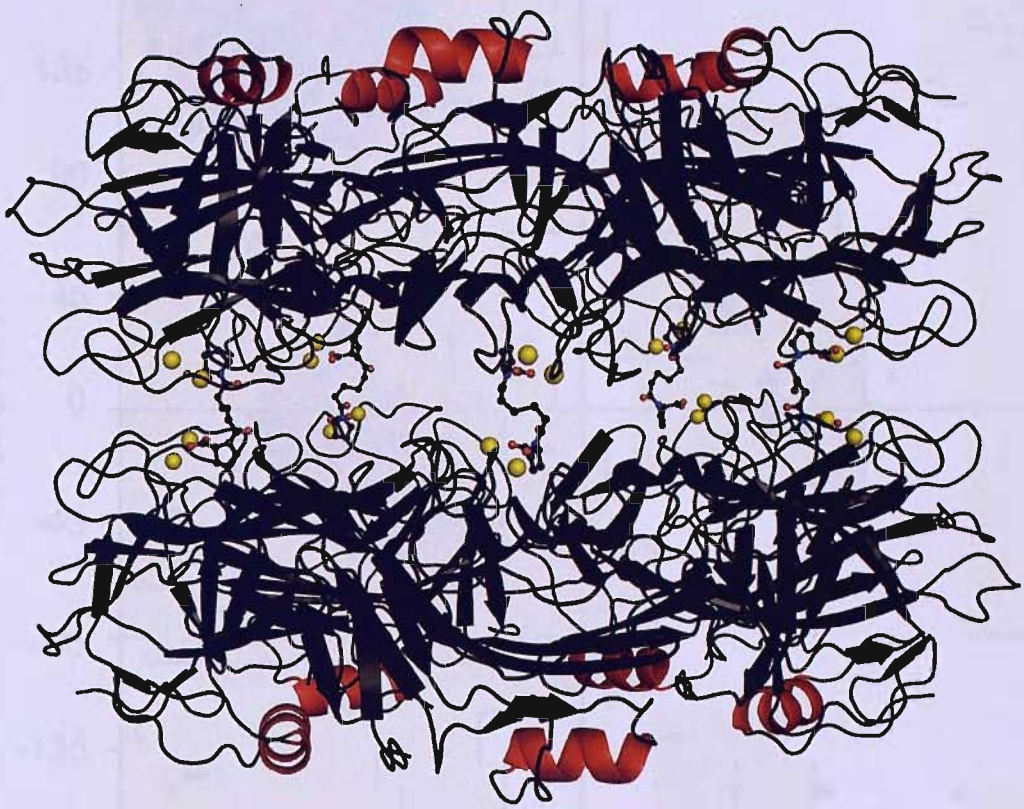


Figure 3.6 The SAP-CPHPC decamer showing two SAP pentamers joined B face to B face by five CPHPC molecules.

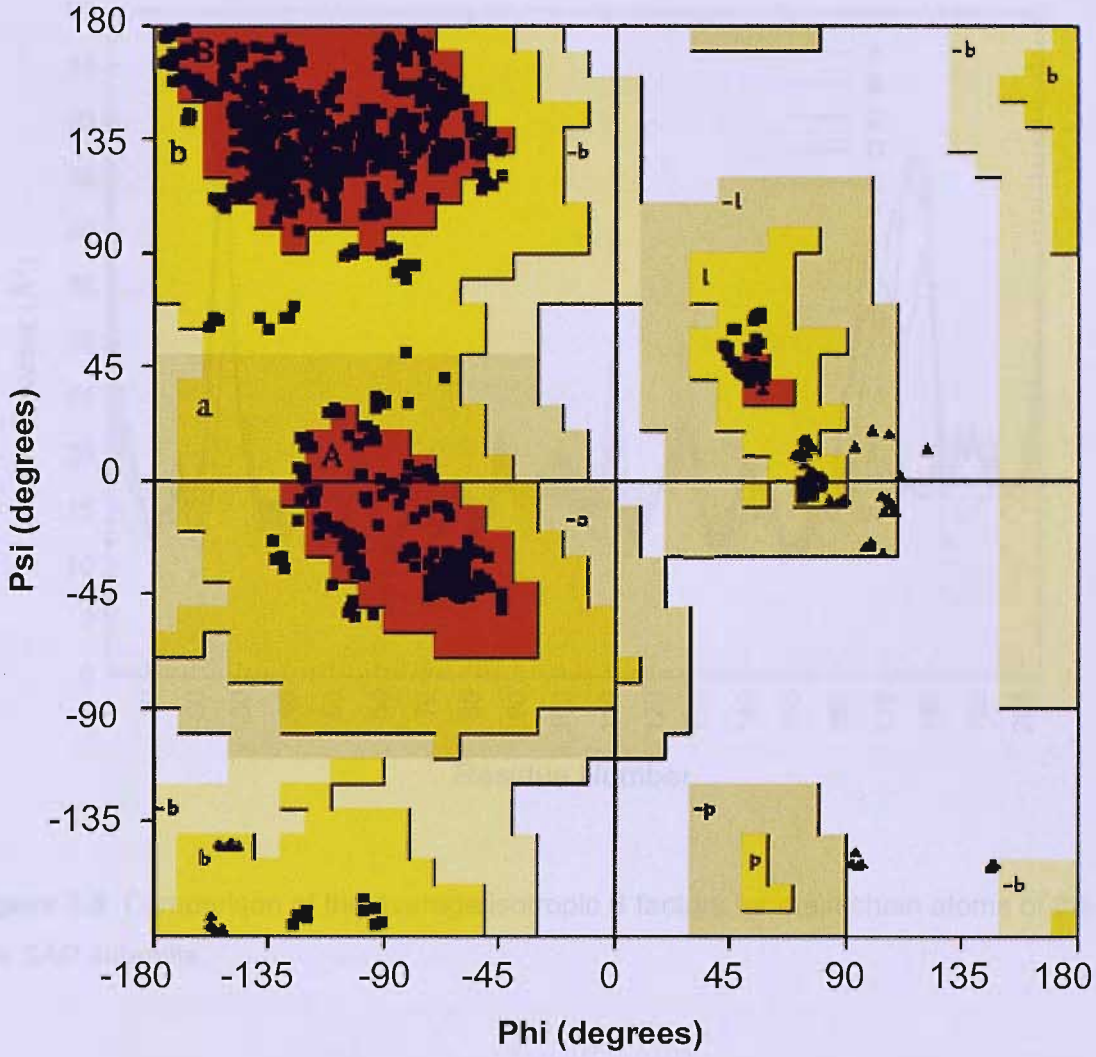


Figure 3.7 Ramachandran plot for SAP-CPHPC.

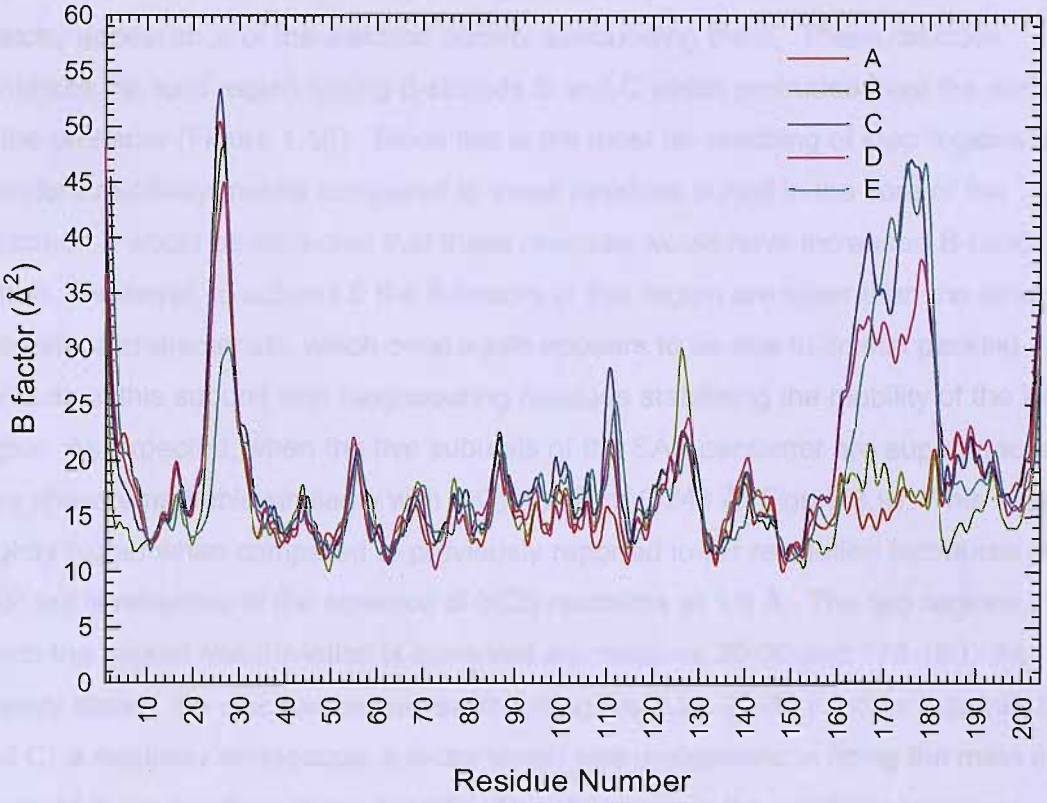


Figure 3.8 Comparison of the average isotropic B factors for main chain atoms of the five SAP subunits.

The graph also shows that region 20-30 in subunits A, B, C, and D is more disordered when compared to the remainder of the protein, an observation that correlates with the sketchy appearance of the electron density surrounding them. These residues constitute the loop region linking β-strands B and C which protrudes from the surface of the protomer (Figure 1.16). Since this is the most far-reaching of loop regions and is therefore relatively mobile compared to those residues buried in the core of the protomer, it would be expected that these residues would have increased B-factor values. However, in subunit E the B-factors of this region are lower than the other four subunits, a characteristic which once again appears to be due to crystal packing contacts of this subunit with neighbouring residues stabilising the mobility of the loop region. As expected, when the five subunits of the SAP pentamer are superimposed they show remarkable similarity with a C_α rms fit of 0.243 Å (Figure 3.9). This may be slightly higher when compared to previously reported lower resolution structures of SAP but is reflective of the absence of NCS restraints at 1.6 Å. The two regions in which the largest rms deviation is observed are residues 20-30 and 170-180. As already stated, the electron density surrounding residues 20-30 in some subunits (B and C) is relatively ambiguous, a factor which was problematic in fitting the main chain atoms of these residues to the density. This may explain the relatively large rms deviation observed in this region across the five subunits.

The overall fold of the component SAP pentamer is very similar to that already described (see section 1.6). Each pentamer consists of five protomers arranged symmetrically around a central hollow pore of approximately 25 Å. Characteristic of SAP, the protomers are composed primarily of β-sheet, the component β-strands of which exhibit jelly-roll topology. A single α-helix is located on the A face of the molecule, positioned over the solitary disulphide bond of the protomer, all of which are intact in the reported structure.

A comparison of the SAP pentamer from the SAP-CPHPC complex with that of the 2.0 Å SAP pentamer originally solved by Emsley *et al.*, 1994 (1SAC.pdb) confirms the observed similarity between the reported structure and previously presented structures of SAP. Figure 3.10 shows a comparison of the average isotropic B-factors for all five subunits between the structures of 1SAC and the SAP-CPHPC complex. Analysis of the graph reveals an almost identical distribution of B-factor values across the SAP protomers from the two structures. Regions of increased B-factor value are similar in both and as expected are localised to the highly flexible loop regions adjoining the component β-strands of the protomer.

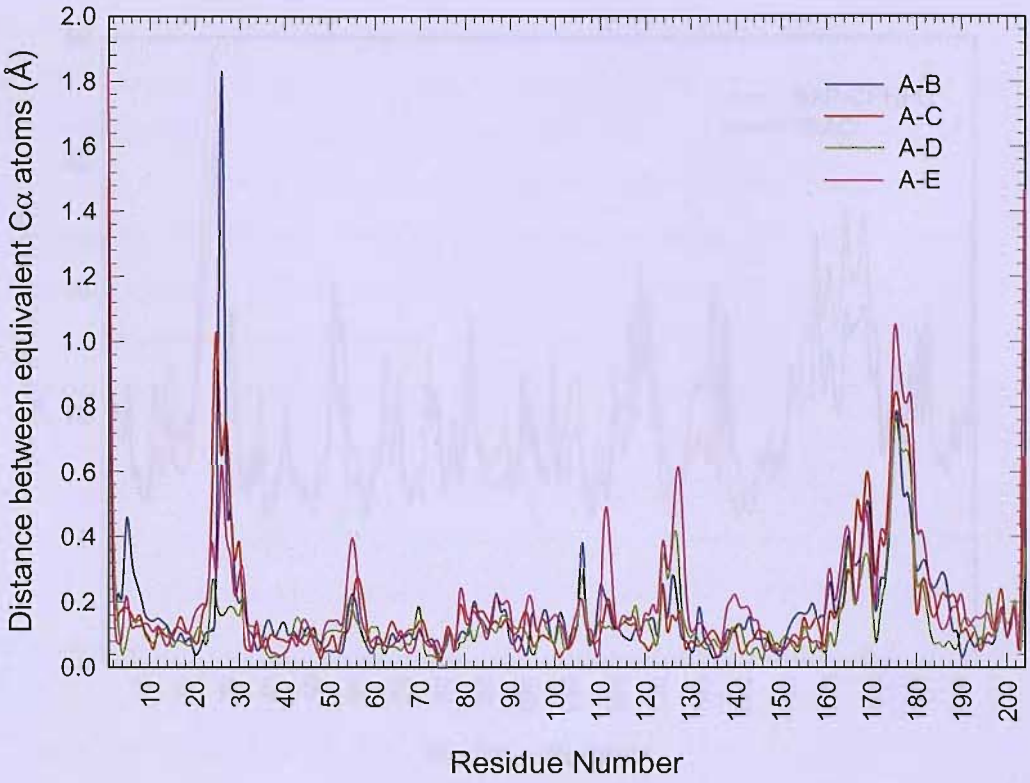


Figure 3.9 Comparison of C_α atom positions between all five SAP protomers in the SAP-CPHPC complex.

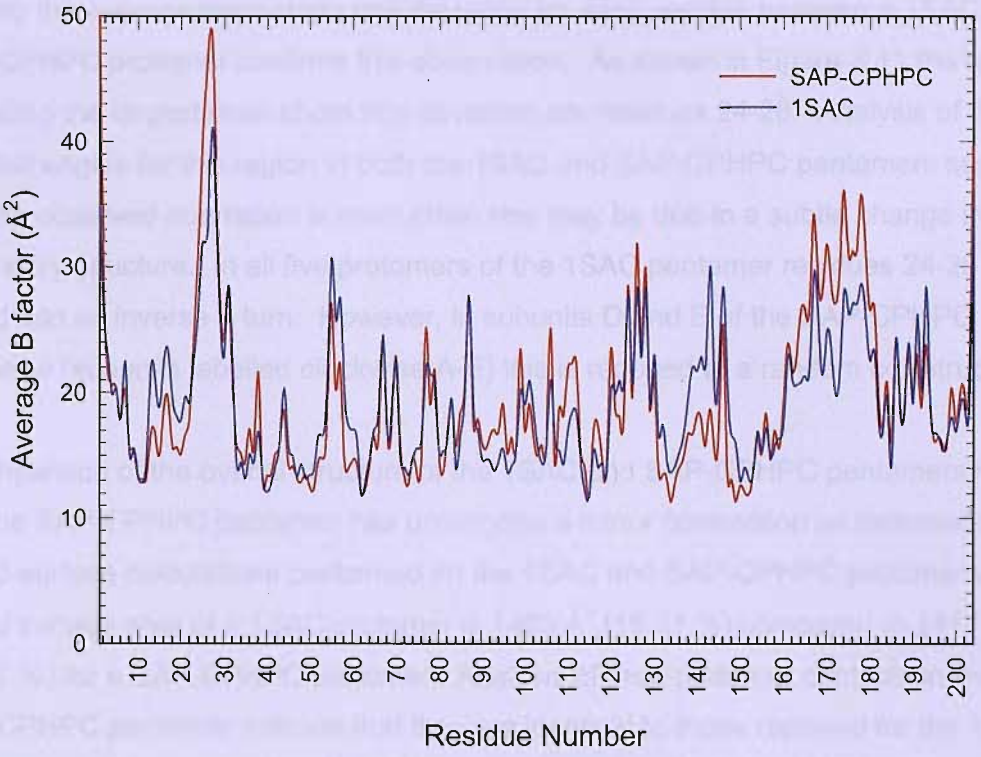


Figure 3.10 Comparison of average isotropic B factors for all residues between the SAP subunits of the SAP-CPHPC complex and 1SAC.

The average C_α rms fit of the five protomers from the 1SAC and SAP-CPHPC pentamers is 0.269 Å suggesting that as expected, the two structures are very similar. Plotting the average main chain rms deviation for each residue between a 1SAC and a SAP-CPHPC protomer confirms this observation. As shown in Figure 3.11 the region exhibiting the largest main chain rms deviation are residues 24-26. Analysis of the dihedral angles for this region in both the 1SAC and SAP-CPHPC pentamers suggests that the observed aberration in main chain rms may be due to a subtle change in secondary structure. In all five protomers of the 1SAC pentamer residues 24-26 are folded into an inverse γ-turn. However, in subunits D and E of the SAP-CPHPC pentamer (subunits labelled clockwise A-E) this is reduced to a random coil structure.

A comparison of the overall structure of the 1SAC and SAP-CPHPC pentamers reveals that the SAP-CPHPC pentamer has undergone a minor contraction as indicated in buried-surface calculations performed on the 1SAC and SAP-CPHPC protomers. The buried surface area of a 1SAC protomer is 1463 $Å^2$ (16.31 %) compared to 1417 $Å^2$ (15.77 %) for a SAP-CPHPC protomer. Analysis of inter-protomer contacts in the SAP-CPHPC pentamer indicate that they are identical to those reported for the 1SAC pentamer with the exception of the salt bridge formed between Glu 153 and Lys 116 which is absent from all five protomers in the SAP-CPHPC pentamer. The reason for the observed contraction and loss of an inter-protomer salt bridge is unlikely to be an artefact of CPHPC binding since it occurs perpendicular to the plane of the pentamer. It could be explained by the difference in the pH of the two crystallisation conditions for each SAP structure. 1SAC was crystallised at a pH of approximately 7, conditions that would permit the formation of the salt bridge. However, the 1.6 Å structure of the SAP-CPHPC complex was grown at pH 4.6 which is very close to the p*k* of glutamate.

The double cadmium binding site exhibits six ligands for both cadmium ions. Cadmium I is coordinated by single ligands provided by the side-chains of Glu136, Asn59, Asp138 and the main-chain carbonyl of Gln137. Contrary to the descriptions of ion coordination in 1SAC and other structures of SAP, Asp58 provides only a single ligand to Cadmium I in the reported structure. The distance between the carboxylate oxygen of Asp58 and Cadmium I is on average 3 Å and is unlikely to provide an additional ligand. The sixth ligand is therefore provided by a terminal carboxyl oxygen of CPHPC, a position that is occupied by the acetate buffer ion in 1SAC. Cadmium II is coordinated by six ligands provided by the side chains of Gln148, Glu136 and Asp138, two water molecules and the second oxygen of the terminal carboxylate of CPHPC. The coordination spheres of the two cadmium ions are therefore arranged as a

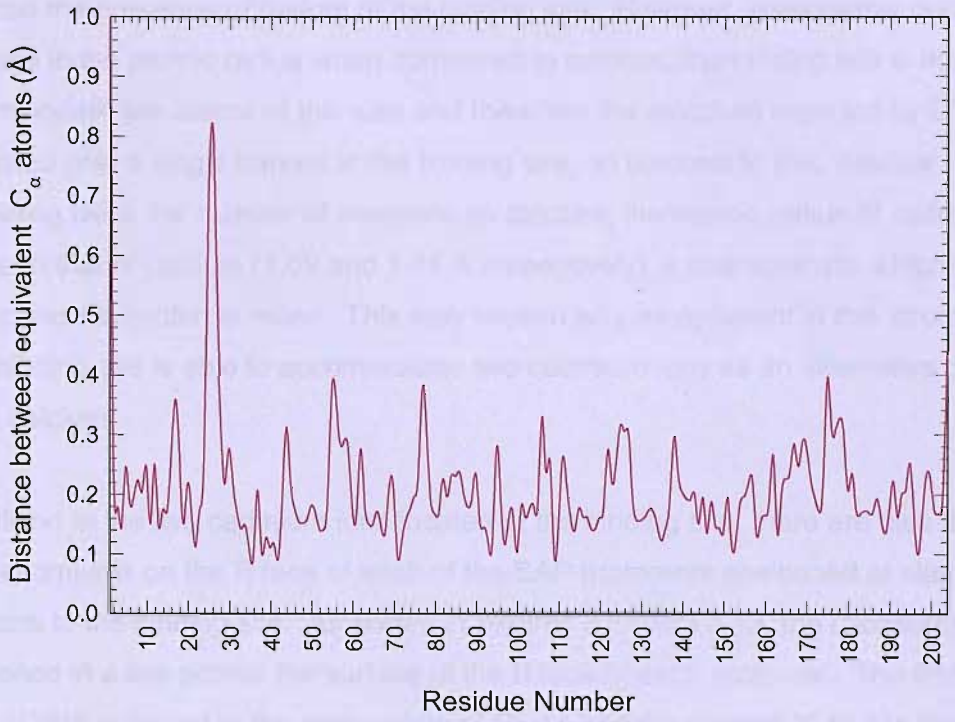


Figure 3.11 Average C_α distances between superimposed subunits of the SAP-CPHPC and 1SAC pentamers.

distorted octahedron. The two cadmium ions are positioned approximately 3.7 Å apart and as in 1SAC are bridged by the side-chains of Glu136 and Asp138.

The 1.6 Å SAP-CPHPC complex is the first SAP structure presented with two alternative divalent ions coordinated at the binding site. O'Hara (1988) had previously reported the presence of barium at the binding site. However, presumably due to the increase in the atomic radius when compared to calcium, the binding site is not able to accommodate two atoms of this size and therefore the structure reported by O'Hara contained only a single barium at the binding site. In contrast to this, despite containing twice the number of electrons as calcium, the atomic radius of cadmium is similar to that of calcium (1.09 and 1.14 Å respectively), a characteristic which reflects the increase in proton number. This may explain why as apparent in this structure, the SAP binding site is able to accommodate two cadmium ions as an alternative divalent ion to calcium.

In addition to the two cadmium ions located at the binding site, there are also three extra cadmiums on the B face of each of the SAP protomers positioned at alternative locations to the binding site. As shown in Figures 3.12 and 3.13, the cadmiums are positioned in a line across the surface of the B face of each protomer. The first of these (CdIII) is bound to the carboxylate of Glu14 and the second (CdIV) to the carboxylate of Asp145. The third (CdV) is positioned in between these two cadmiums devoid of a direct protein ligand but bound by three water molecules that provide an anchor to the main body of the protein. There is also a fourth additional cadmium (CdVI) bound to the carboxyl of Asp58 that is observed in subunit E, and possibly in subunit D with a lower occupancy. This cadmium appears to be directly involved in linking the pentamers as it is also bound to Asp138 in the adjacent pentamer. Due to the low occupancy of this ion across the five subunits, it is however, difficult to say whether the ion significantly contributes to decamer stability. Presumably due to the location of these cadmiums on the surface of the protein, in comparison to cadmium I and cadmium II (average B factors CdI = 13.88 $Å^2$ CdII = 14.03 $Å^2$), the thermal motion of cadmiums III, IV and V is increased with the average B factor of CdIII 24.78 $Å^2$, CdIV 23.20 $Å^2$ and CdV 36.94 $Å^2$.

The double cadmium binding sites of adjacent SAP pentamers are positioned approximately 9.3 Å apart and are bridged by a single CPHPC molecule. The terminal carboxyl groups of the CPHPC molecule bridge the two cadmium ions at the binding site (CdI and CdII) with an average O-Cd distance of 2.3 Å and are involved in

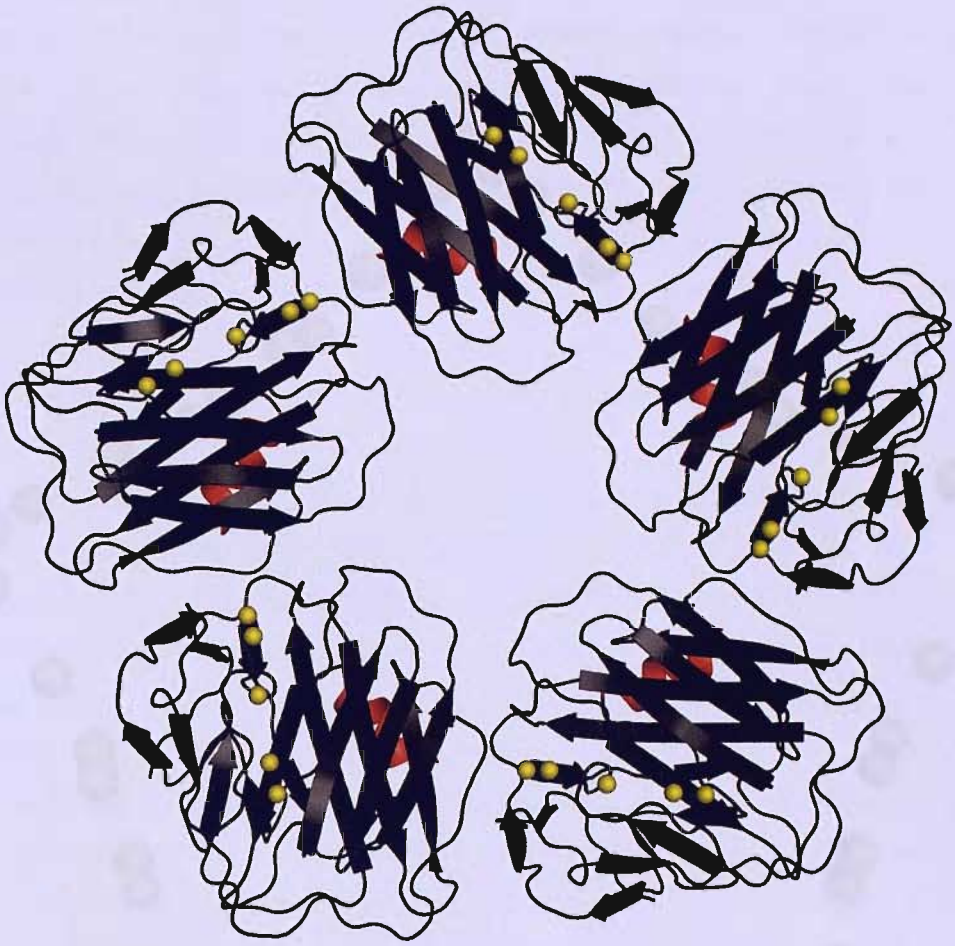


Figure 3.12 The distribution of all cadmium ions (yellow spheres) across the surface of the SAP pentamer.

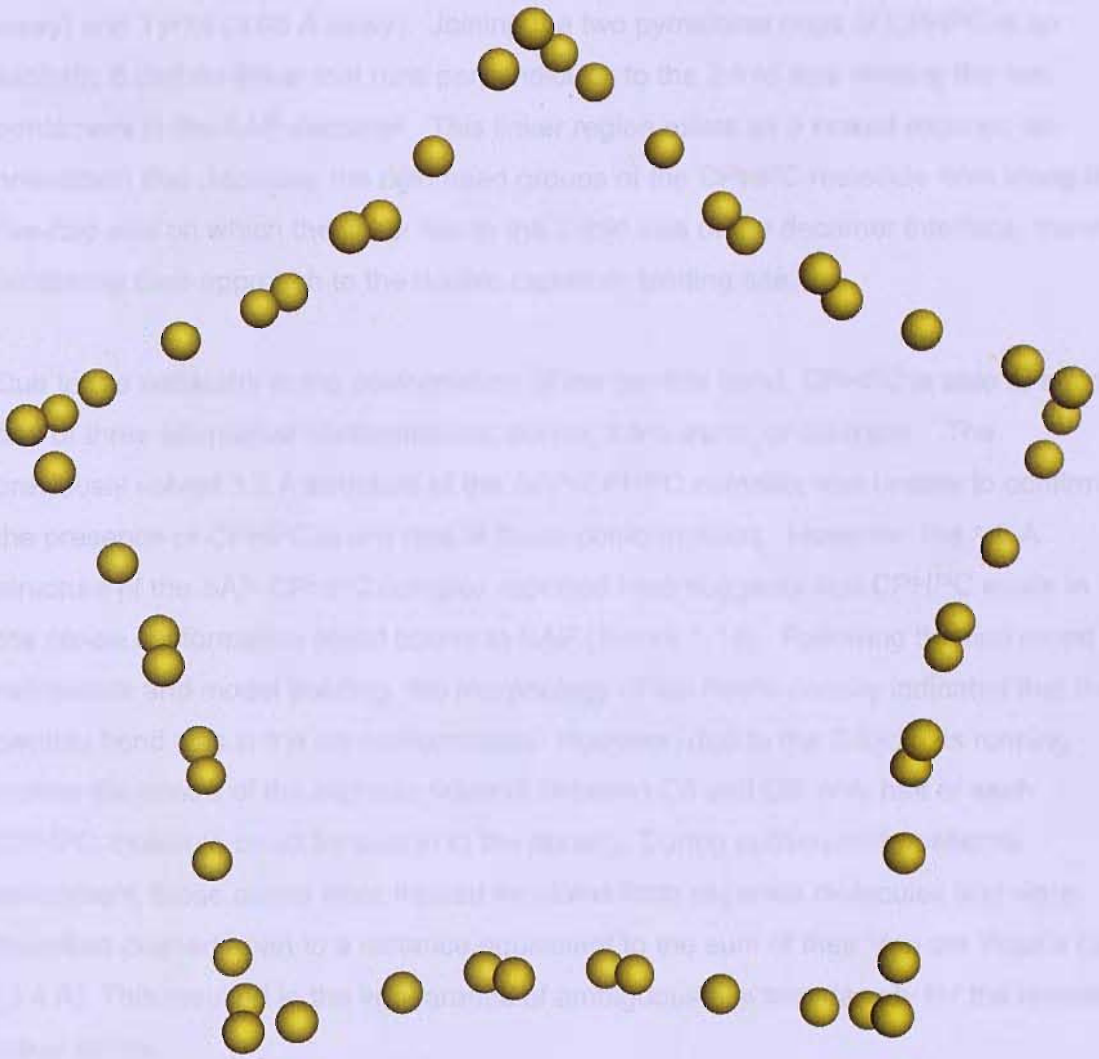


Figure 3.13 Distribution of the cadmium ions (yellow spheres) across the SAP decamer. The 2-fold axis creates a continuous path of cadmiums arranged in a pentagonal path around the SAP-CPHPC complex.

hydrogen bond formation with Gln 148 and Asn 59. With the terminal carboxyl group orientated in this manner, the pyrrolidine ring of CPHPC is ideally positioned into the adjacent hydrophobic pocket formed by residues Leu62 (3.64 Å away), Tyr64 (3.42 Å away) and Tyr74 (3.68 Å away). Joining the two pyrrolidine rings of CPHPC is an aliphatic 6 carbon linker that runs perpendicular to the 2-fold axis relating the two pentamers in the SAP decamer. This linker region exists as a kinked rotamer, an orientation that displaces the rigid head groups of the CPHPC molecule from along the five-fold axis on which the linker lies to the 2-fold axis of the decamer interface, thereby facilitating their approach to the double cadmium binding site.

Due to the variability in the conformation of the peptide bond, CPHPC is able to adopt one of three alternative conformations: *cis-cis*; *trans-trans*; or *cis-trans*. The previously solved 3.2 Å structure of the SAP-CPHPC complex was unable to confirm the presence of CPHPC in any one of these conformations. However, the 1.6 Å structure of the SAP-CPHPC complex reported here suggests that CPHPC exists in the *cis-cis* conformation whilst bound to SAP (Figure 3.14). Following the first round of refinement and model building, the morphology of the *F*o-*F*c density indicated that the peptide bond was in the *cis* conformation. However, due to the 2-fold axis running across the centre of the aliphatic linker in between C8 and C9, only half of each CPHPC molecule could be built in to the density. During subsequent positional refinement, these atoms were treated as atoms from separate molecules and were therefore pushed apart to a distance equivalent to the sum of their Van der Waal's radii (3.4 Å). This resulted in the appearance of ambiguous electron density for the terminal linker atoms.

In order to confidently define the positions of these linker atoms, it was decided to restrict Van der Waal's repulsion by fixing the position of atom C8 during refinement. The resulting model produced well-defined 2*F*o-*F*c density for CPHPC in the *cis* conformation in three of the five subunits (B, D and E). Positioning of the atoms in the *trans* conformation in subunits B, D and E forces linker atoms C7 and C8 out of the electron density where they are surrounded by strong negative density peaks (Figure 3.15). In addition to this negative density surrounding atoms in the *trans* conformation, positive density is observed at those positions where linker atoms in the *cis* conformation would protrude from the pyrrolidine ring. The electron density for the linker region in subunits A and C remains inconclusive. This is most likely a consequence of fixing atom C8 during refinement, as this atom is potentially forced to adopt an incorrect conformation.

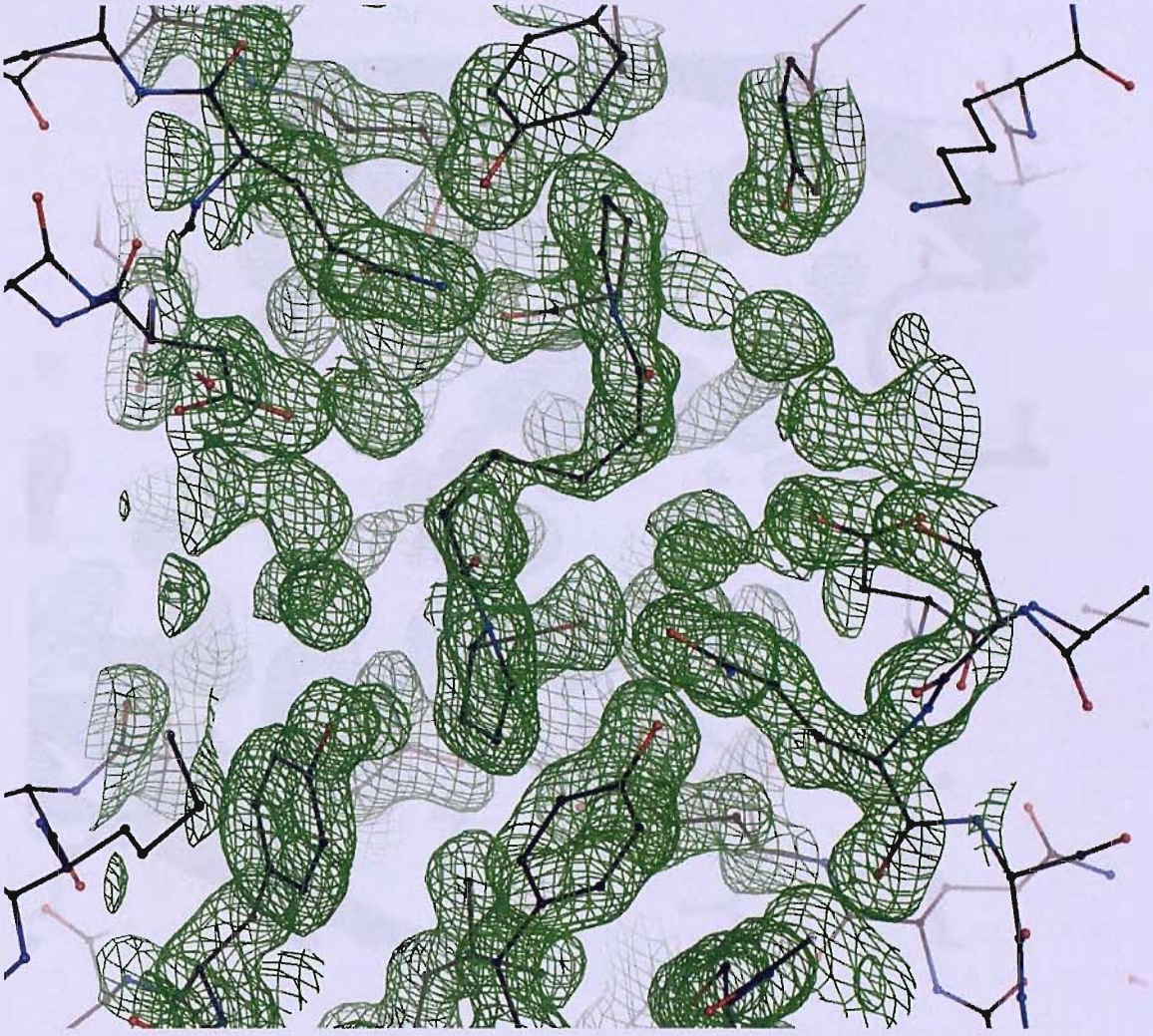


Figure 3.14 Electron density (2Fo-Fc) contoured at 1σ, and the fitted CPHPC molecule in the *cis* conformation.

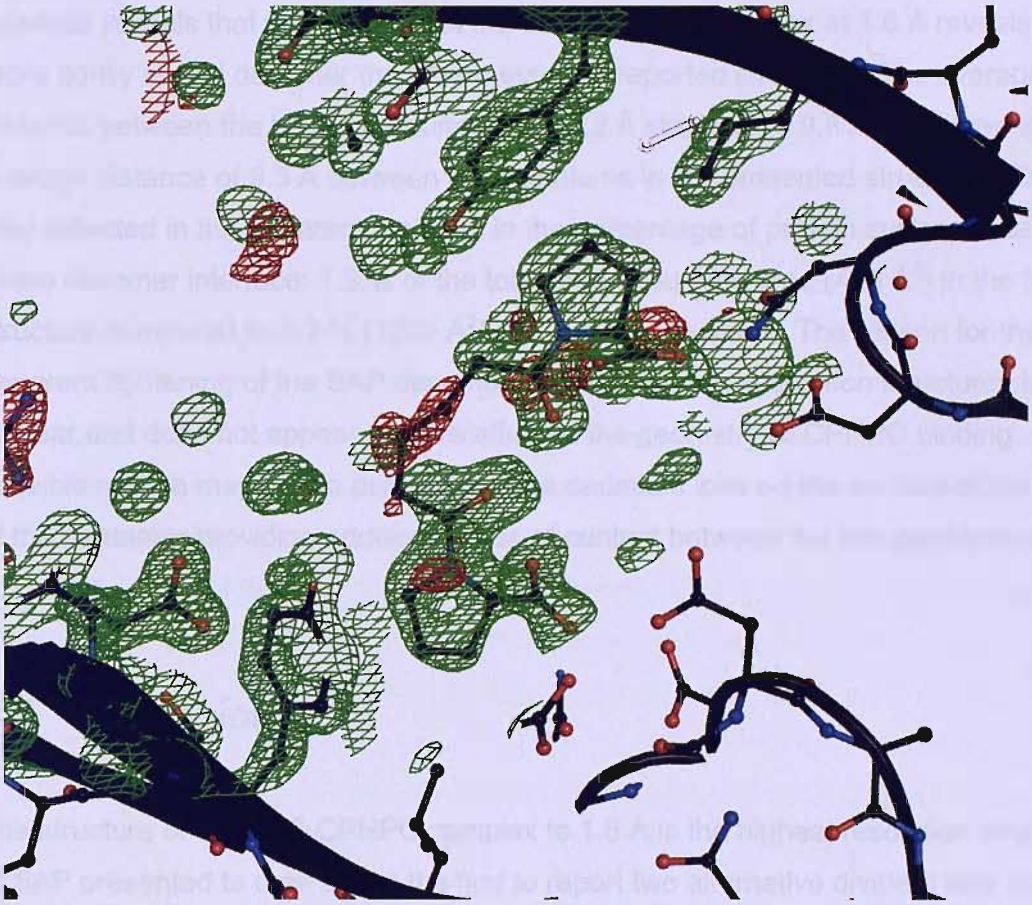


Figure 3.15 2Fo-Fc density (green; 1.5 σ) and negative Fo-Fc density (red; -2 σ) of the fitted CPHPC molecule in the *trans* conformation.

In comparison to the 3.2 Å structure of the SAP-CPHPC complex which reported an absence of any protein-protein interactions across the decamer interface, the structure at 1.6 Å reveals the formation of two hydrogen bonds formed between Arg77 and Phe144 of each symmetry-related protomer pair and hydrophobic interactions between Arg77 and Lys143. Analysis of the cadmium to cadmium distances across the interface reveals that the structure of the SAP-CPHPC complex at 1.6 Å reveals a more tightly bound decamer than the previously reported structure. The average distance between the bound calciums in the 3.2 Å structure is 9.8 Å compared to an average distance of 9.3 Å between the cadmiums in the presented structure. This is also reflected in the apparent increase in the percentage of protein surface area buried at the decamer interface: 1.3 % of the total protein surface area (448 $Å^2$) in the 3.2 Å structure compared to 3.2 % (1204 $Å^2$) in the 1.6 Å structure. The reason for the apparent tightening of the SAP decamer between the two resolution structures is unclear and does not appear to have affected the geometry of CPHPC binding. One possible reason may be the presence of the cadmium ions on the surface of the B face of the pentamer providing additional sites of contact between the two pentamers.

3.7 Discussion

The structure of the SAP-CPHPC complex to 1.6 Å is the highest resolution structure of SAP presented to date and is the first to report two alternative divalent ions at the binding site. The substitution of native metal ions for cadmium has been reported in a number of other X-ray crystal structures (Blackwell *et al.*, 1994; Huang *et al.*, 2002). Cadmium has also been used as an additive to the crystallisation conditions and has been reported as being a requirement for the successful growth of certain protein crystals (Zanotti *et al.*, 1998; Enguita *et al.*, 2003). In response to these various structures describing cadmium as an essential ingredient to the crystallisation cocktail, a study was carried out to specifically investigate the influence of cadmium ions on the process of crystallisation (Trakhanov *et al.*, 1998). As in the SAP-CPHPC complex, the study reports the deposition of cadmium ions at the interface between neighbouring protein molecules bound by the carboxyl groups of aspartic or glutamic acid residues and supplementary water molecules. However, in contrast to the SAP-CPHPC complex these cadmiums are involved in forming direct contacts between the two protein molecules across the interface. It appears that in the SAP-CPHPC complex the cadmiums are not involved in direct contacts but may instead provide additional

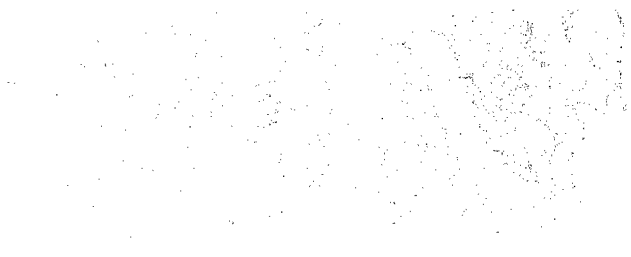
sites for well-ordered connecting water networks between the protein molecules. The formation of well-ordered water networks sandwiched between the two pentamers of the SAP decamer would explain the observed increase in crystal quality compared to the 3.2 Å structure. However, the reason for the apparent specificity for cadmium in improving protein crystallisation over other divalent ions such as Zn^{2+}, Ca^{2+} or Mg^{2+} remains unclear.

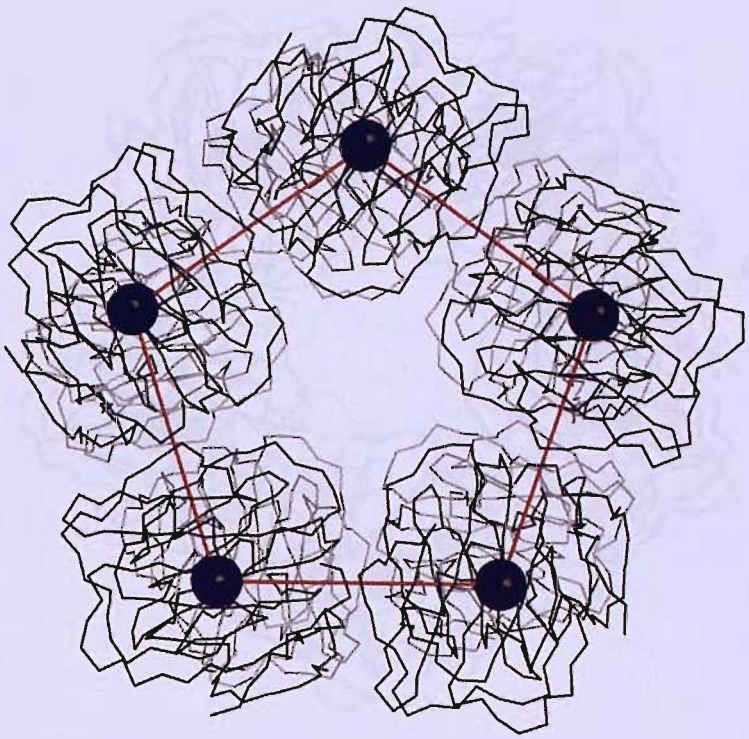
The previously reported structure of the SAP-CPHPC complex at 3.2 Å was unable to confirm the confirmation of the peptide bond in CPHPC as either *cis* or *trans*. The 1.6 Å structure of the complex reported here strongly suggests that the preferred conformation of the peptide bond is in the *cis* form. This finding has important implications for the success of the compound as a potential drug in the treatment of amyloidosis. Since only 25% of the population of *X*-Pro (*X* = any amino acid) peptide bonds are present in the *cis* conformation in free solution (Purvis, 2002), the rate of SAP decamerisation will be dependent on the rate of isomerisation. The mechanism and rate of *cis-trans* isomerisation is intrinsically a slow process and has been identified as one of two rate-limiting steps in protein folding. Identification of the preferred peptide conformation in CPHPC as *cis-cis* will enable the design of methods by which it can be 'locked' in this form thus improving drug efficacy. A possible method of locking CPHPC in the *cis-cis* form could be the transformation of the peptide bond to a double bond, thus preventing the rotation around this bond that is required for isomerisation.

Since the SAP-CPHPC structure has been solved to 1.6 Å, another research group has published two structures of SAP decamerised by two different bivalent ligands (Ho *et al.*, 2005). The two ligands presented are both simple but potent bivalent inhibitors of SAP that incorporate linkers of different flexibility, a characteristic that seems to have a dramatic effect on the potency of the inhibitory compound. Compounds 1 and 2, as presented by the group, are based on the SAP ligand MOβDG. Both compounds contain identical end groups of the cyclic pyruvate ketal of galactose and are joined by linkers of identical chain length. However, in compound 2 an additional piperazine ring positioned in the centre of the linker prevents rotation around the three central C-C bonds therefore restricting the flexibility of the molecule. This decrease in the degree of freedom around the linker structure seems to result in a 30-fold increase in the binding affinity of compound 2 (0.12 μM) when compared to compound 1 (3.8 μM). This may be an interesting feature in the continual optimisation of CPHPC as a potential therapeutic treatment for amyloidosis. The introduction of a ring structure into

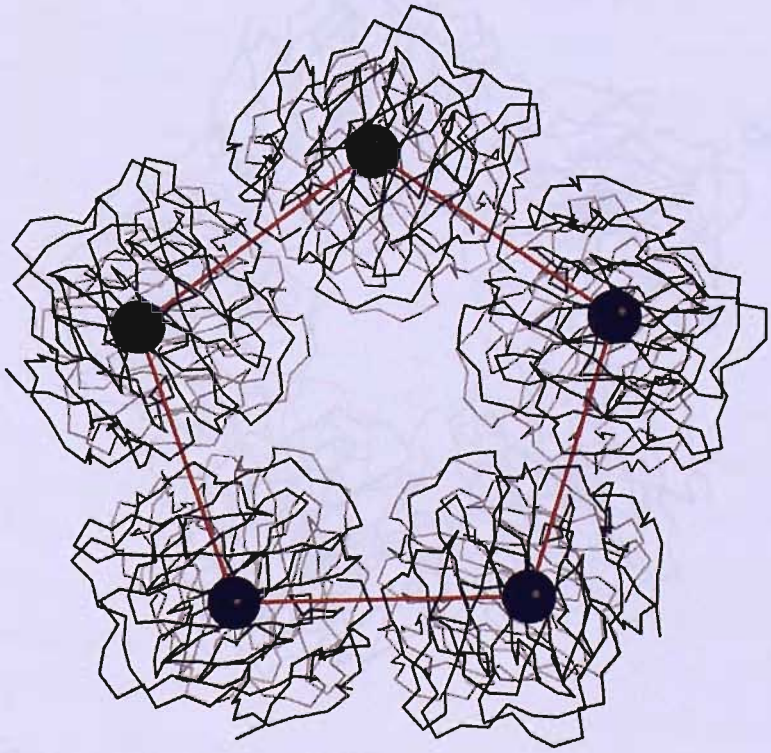
the linker region may act to stabilise the compound in a similar way. CPHPC in its current form is like compound 1 in that it contains three central single bonds around which there are three additional degrees of torsional rotational freedom. In theory, these single bonds in CPHPC could therefore adopt any one of three staggered rotamers, a feature that according to the thermodynamic model proposed by Hoo *et al.* (2005) is expected to reduce the stability of the decameric complex ($K_{decamer}$) by a factor of approximately 3^3. However, in CPHPC it may be that the conformation of the linker is already restricted by the presence of a kinked rotamer between C8 and C9. The presence of this kink seems to be essential in allowing the head groups to be correctly orientated in order to interact with the double metal binding sites of the opposing SAP pentamers. In order to produce the necessary kink, the bond between C8 and C9 can only exist in an eclipsed conformation thereby restricting the degrees of freedom around this bond.

It also appears from the analysis of all these structures that the component SAP pentamers are free to adopt any one of a number of rotations with respect to each other within the decamer. Alignment of the SAP decamers formed by CPHPC, compound 1 and compound 2 with that of the dAMP structure reveals a range of relative rotational arrangements of the pentamers within these structures (Figure 3.16). The calcium binding sites of the opposing pentamers in the dAMP complex are closely aligned. In comparison, the metal ions in the SAP-CPHPC, SAP-compound1 and SAP-compound2 complexes are offset from each other and subsequently induce a rotational shift in one pentamer with respect to the other. When compared to the dAMP complex, the SAP-CPHPC pentamers are rotated 8° with respect to each other and the pentamers of the compound 1 and 2 complexes are rotated 26° and 22° respectively.





dAMP



CPHPC

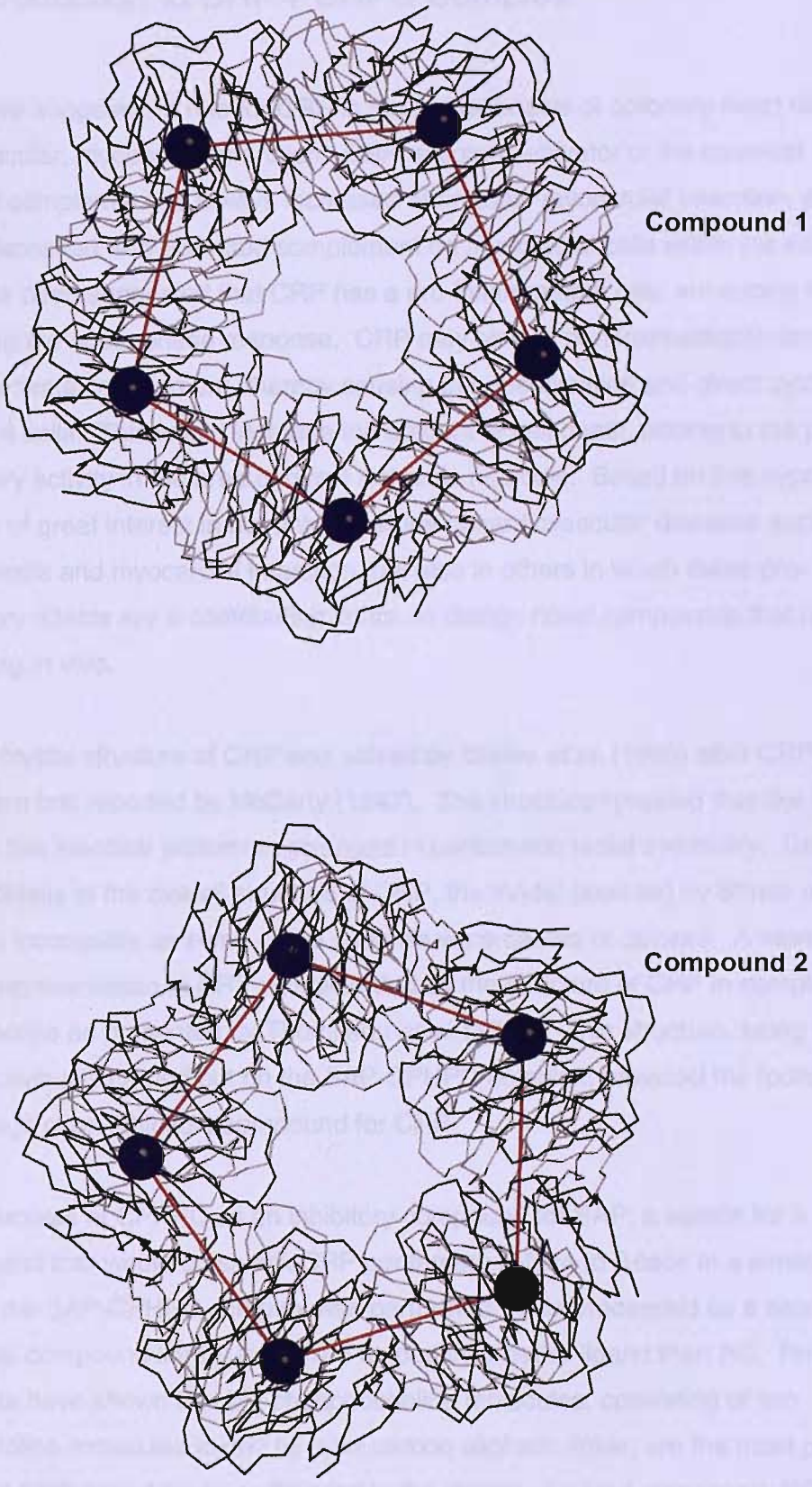


Figure 3.16 Decameric structures of SAP bound to dAMP, CPHPC, compound 1 and compound 2. The orientation of the lower pentamer (coloured grey) is the same for all molecules. The centre of mass for each protomer is shown as a large black sphere.

4.1 Introduction to CRP-PCHPC Complex

Studies have suggested a role for CRP in the pathogenesis of coronary heart disease and in particular, myocardial infarction. CRP, a potent activator of the classical pathway of complement, is always increased after acute myocardial infarction, and is found co-deposited with activated complement on myocardial cells within the infarcted area. It has been suggested that CRP has a pro-inflammatory role; enhancing tissue injury during the acute phase response. CRP may bind to non-irremediably damaged cells and activate complement, thereby causing the opsonisation and direct cytotoxicity of the target cells. This would increase the amount of cell death, adding to the pro-inflammatory activity in the area of direct ischemic necrosis. Based on this hypothesis, it would be of great interest in not only the study of cardiovascular diseases such as atherosclerosis and myocardial infarction, but also in others in which these pro-inflammatory effects are a contributing factor, to design novel compounds that block CRP binding *in vivo*.

The X-ray crystal structure of CRP was solved by Shrive *et al.* (1996) after CRP crystals were first reported by McCarty (1947). The structure revealed that like SAP, it consists of five identical protomers arranged in pentameric radial symmetry. Despite providing details of the overall structure of CRP, the model provided by Shrive *et al.* (1996) was incomplete as some of the subunits were devoid of calcium. A more complete representation of CRP was provided by the structure of CRP in complex with phosphocholine as presented by Thompson *et al.* (1999). This structure, along with the work previously carried out on the SAP-CPHPC complex, provided the foundation for the design of an inhibitory compound for CRP.

After the success of CPHPC as an inhibitory compound for SAP, a search for a bivalent ligand that would cross-link CRP pentamers B-face to B-face in a similar manner to the SAP-CPHPC complex was performed. This proceeded as a search of an industrial compound library that failed to identify a better ligand than PC. Previous experiments have shown that bis-phosphocholine molecules, consisting of two phosphocholine molecules joined by a six carbon aliphatic linker, are the most potent inhibitors of CRP-ligand binding. This led to the design of a lead compound, PCHPC (phosphocholine-hexane-phosphocholine [also known as bis-phosphocholine hexane in the paper]) as shown in Figure 4.1. Subsequent studies have shown the binding

Chapter 4
Structural Studies of C-Reactive Protein in Complex with PCHPC

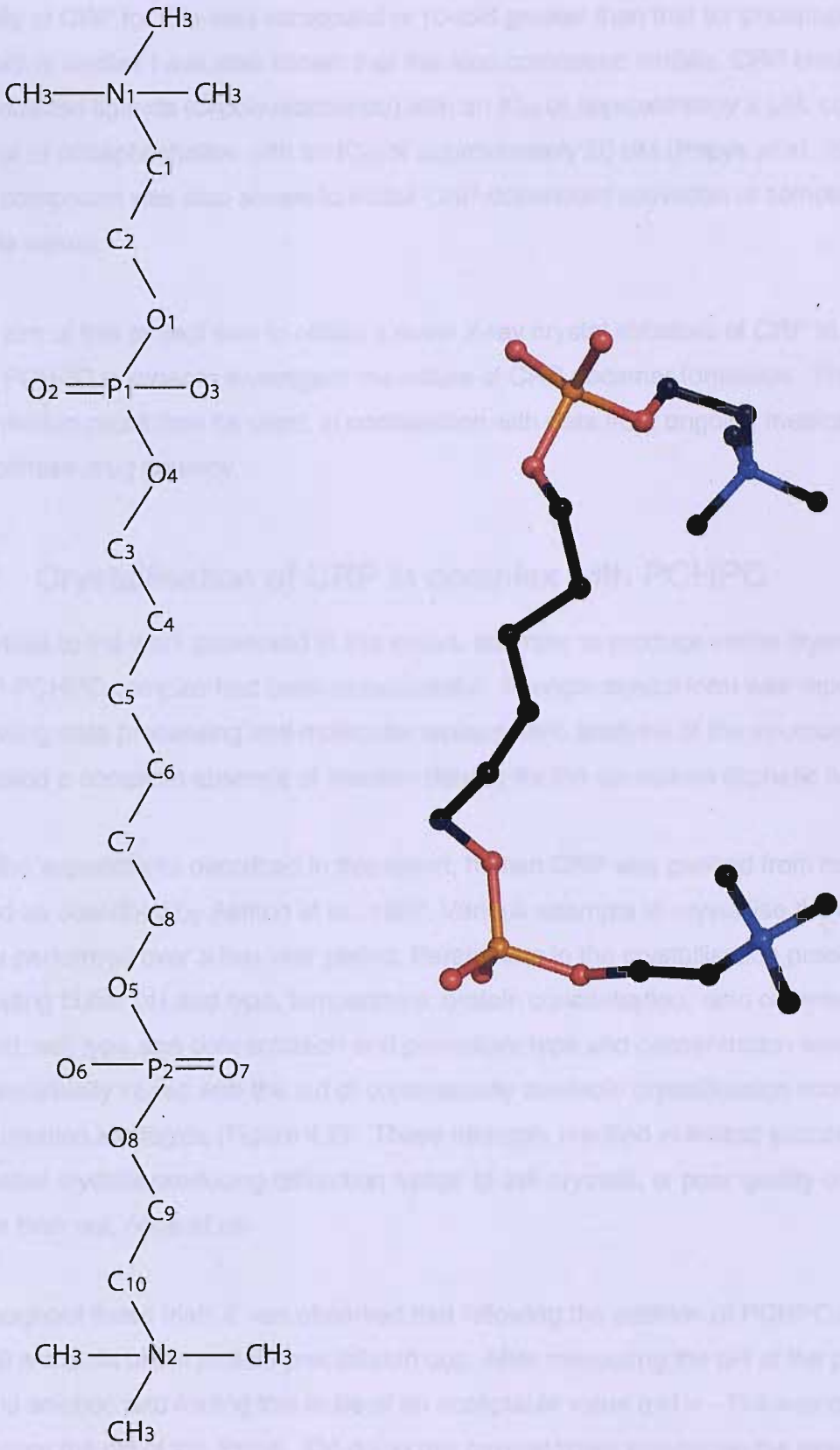


Figure 4.1 Structure of PCHPC. **Left** Chemdraw representation of PCHPC. **Right** Pymol representation of PCHPC.

affinity of CRP for this lead compound is 10-fold greater than that for phosphocholine. Inhibitory studies have also shown that this lead compound inhibits CRP binding to immobilised ligands (C-polysaccharide) with an IC_{50} of approximately 2 μM, compared to that of phosphocholine with an IC_{50} of approximately 20 μM (Pepys *et al.*, 2006). The compound was also shown to inhibit CRP-dependent activation of complement in whole serum.

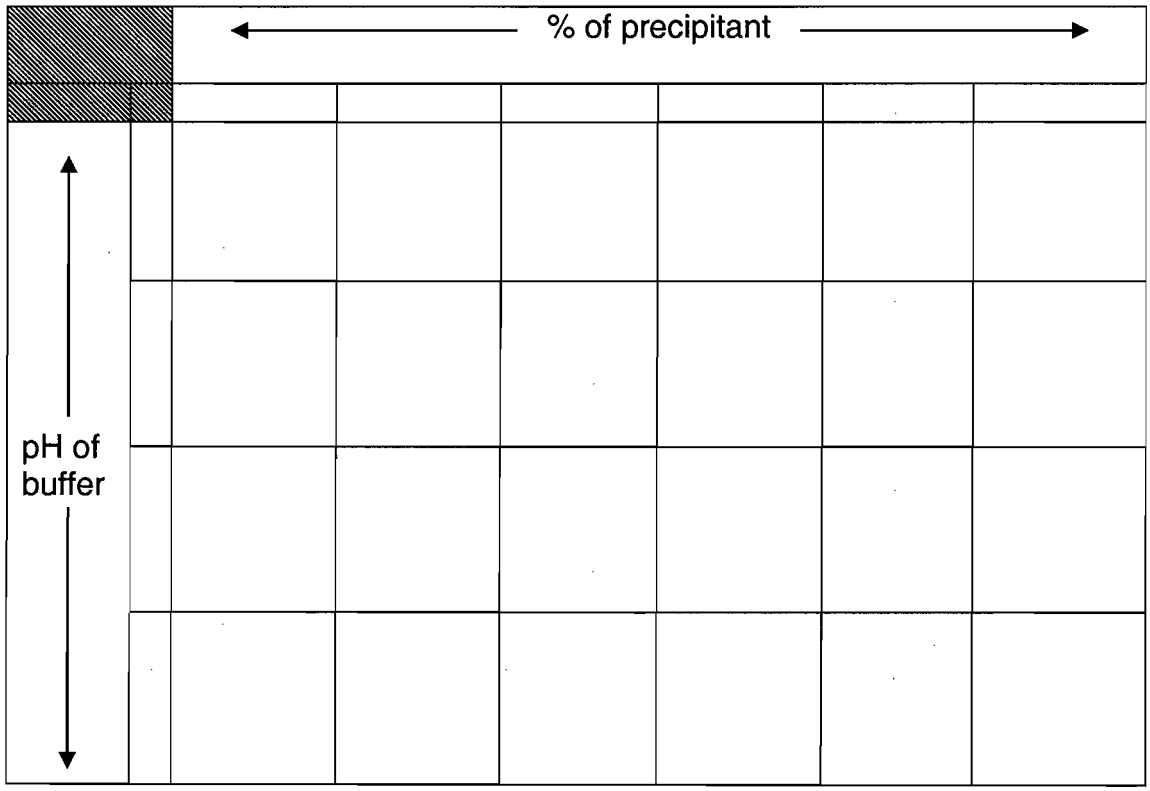
The aim of this project was to obtain a novel X-ray crystal structure of CRP in complex with PCHPC in order to investigate the nature of CRP decamer formation. This information could then be used, in combination with data from ongoing medical trials, to optimise drug potency.

4.2 Crystallisation of CRP in complex with PCHPC

Previous to the work presented in this report, attempts to produce viable crystals of the CRP-PCHPC complex had been unsuccessful. A single crystal form was reported but following data processing and molecular replacement, analysis of the structure revealed a complete absence of electron density for the six carbon aliphatic linker.

For the experiments described in this report, human CRP was purified from human blood as described by Ashton et al., 1997. Various attempts to crystallise the complex were performed over a two year period. Parameters in the crystallisation process including buffer pH and type, temperature, protein concentration, ratio of protein to ligand, salt type and concentration and precipitant type and concentration were systematically varied with the aid of commercially available crystallisation screens and optimisation strategies (Figure 4.2). These attempts resulted in limited success: potential crystals producing diffraction typical of salt crystals, of poor quality or more often than not, none at all.

Throughout these trials it was observed that following the addition of PCHPC to CRP, small amounts of the protein precipitated out. After measuring the pH of the protein-ligand solution and finding this to be of an acceptable value (pH = ~7) it was decided to measure the pH of the ligand. On doing this several times and taking the average value, the pH of the ligand was found to be pH 2.0. As a result, for subsequent crystallisation trials, the pH of the PCHPC was adjusted to pH 7.5 by addition of 100 mM Tris/HCl, 10 mM calcium acetate and 140 mM sodium chloride prior to mixing with CRP.



Buffers

- Bis-Tris (pH 6.6 – 7.2)
- Ches (pH 8.6 – 9.2)
- Citric Acid (pH 4.0)
- Mes (pH 6.2 – 6.8)
- Mops (pH 7.2 – 7.8)
- Sodium Acetate (pH 4.4 – 5.4)
- Sodium Hepes (pH 4.0 – 9.0)
- Pipes (pH 6.8 – 7.4)
- Tris (pH 8.2 – 8.8)

Salts

- Sodium phosphate (0.1 M)
- Potassium phosphate (0.1M)
- Ammonium phosphate (0.1-0.2 M)
- Calcium acetate (1 M)
- Sodium chloride (1 M)
- Sodium acetate (0.1 M)

Precipitants

- MPD (6-60%)
- Jeffamine (5-30%)
- Sodium chloride (14-26%)
- Propanol (10-60%)

Temperatures

- 4 °C
- 12 °C
- 25 °C

Protein Concentration

- 4 mg/ml
- 40 mg/ml

Ligand:protein ratio

- 10-fold molar excess
- 50-fold molar excess
- 100-fold molar excess

Figure 4.2 Crystallisation conditions screened for the CRP-CPHPC complex.

Following the production of the highly robust crystals of the SAP-CPHPC complex, it was decided to design crystallisation screens for the CRP-PCHPC complex based on the previously reported conditions of the SAP-CPHPC crystals. Analysis of the SAP-CHPC complex reveals the apparent importance of the high concentration of cadmium (100 mM cadmium chloride) in the crystallisation buffer. In addition to the cadmium ions at the binding site mediating ligand binding, cadmium ions are also observed at the interface between SAP pentamers appearing to sandwich the molecules together. These interactions are proposed to stabilise the structure of the decamer therefore producing well-diffracting, more robust crystals. It was therefore hypothesised that high concentrations of a divalent ion may have a similar effect on the CRP-PCHPC complex. Since cadmium had been so successful in the crystallisation of SAP-CPHPC, it was used as the divalent ion in initial optimisation screens. However, due to a lack of crystallisation 'hits' it was decided to revert to calcium as the divalent ion of choice.

The importance of the low pH and presence of acetate in the crystallisation buffer were also considered. The working hypothesis in the design of a suitable crystallisation solution for CRP-PCHPC was that at a low pH, the carboxyl groups on those amino acid residues involved in ligand contact, are on the edge of dissociation (pKa 4.5). Thereby hydrogen ions might compete for oxygen binding with the ligand consequently slowing down the process of crystallisation. The same theory was applied to the presence of acetate within the crystallisation buffer: acetate, a potential ligand for SAP and CRP, may compete with PCHPC (CPHPC for SAP) thereby slowing down the process of crystallisation. MPD was chosen over PEG as the precipitant of choice for CRP, as previous successful crystallisations of CRP have contained MPD as the precipitant. These crystallisation screens are shown in Figure 4.3.

During setting up of these crystallisation conditions for the first time at room temperature, tiny, barely visible (~5 µM) crystal fragments were observed in a number of wells containing 50-56% MPD, 100 mM sodium acetate pH 4.6 and 0.04 M calcium chloride at a CRP concentration of 4 mg/ml. In an attempt to increase the size of the crystals, these conditions were subsequently screened and optimised, with varying pH and amounts of sodium acetate, calcium chloride and MPD. CRP-PCHPC crystals were eventually grown using the hanging drop method in 100 mM sodium acetate pH 4.6, 50 mM calcium chloride, 52% MPD at room temperature in the presence of 10-fold molar excess of pH-adjusted PCHPC (Figure 4.4).

		% MPD					
		48%	50%	52%	54%	56%	58%
pH of 0.1 M sodium acetate	4.6						
	4.8						
	5.0						
	5.2						

- Six optimization grids with the addition of 0 – 0.1 M calcium chloride.
- For CRP concentrations of 4, 10, 20 and 40 mg/ml.

Figure 4.3 Optimisation of CRP-PCHPC crystallisation conditions based on those of the SAP-CPHPC structure.

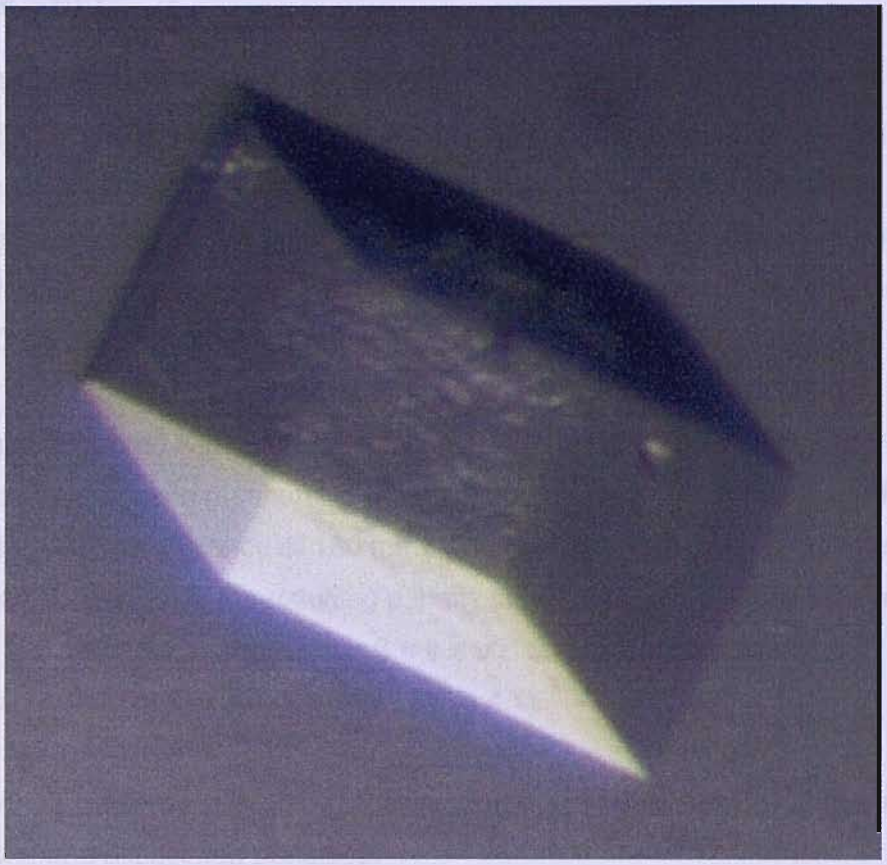


Figure 4.4 Crystal of CRP in complex with PCHPC, grown in 100 mM sodium acetate pH 4.6, 50 mM calcium chloride, 52% MPD in the presence of 10-fold molar excess of pH-adjusted PCHPC.

4.3 Data set I

4.3.1 Data Collection and Processing

Initial diffraction experiments were performed on the in-house rotating anode X-ray source so as to test the ability of the crystals to diffract X-rays. Diffraction was obtained to approximately 2.8 Å and therefore the crystals were flash-cooled in liquid nitrogen for further diffraction experiments. Synchrotron diffraction data from the CRP-PCHPC crystal was collected with a MAR CCD detector at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France on station ID23 (λ = 0.934 Å). The crystal diffracted X-rays to a resolution of 2.3 Å (Figure 4.5). A total of 190 1.0° oscillation frames were measured, with an exposure time of 3 seconds per frame and a crystal to detector distance of 180 mm. The raw data was processed with MOSFLM (Leslie, 1992) and the resulting penalty table proposed unit cell dimensions of a = 96.00, b = 158.23, c = 165.33, $\alpha = \beta = \gamma = 90°$ suggesting that the crystal belonged to an orthorhombic space group (Table 4.1).

Table 4.1 Penalty table generated by Mosflm showing possible unit cell and space group solutions and associated penalties.

No.	Penalty	Latt.	a	b	c	alpha	beta	gamma	Possible Space Groups
10	39	tP	160.22	166.17	96.57	90.3	89.8	90.4	P4, P41, P42, P43
9	36	mC	229.99	231.68	96.57	90.4	90.1	87.9	C2
8	36	oC	229.99	231.68	96.57	90.4	90.1	87.9	C222, C2221
7	33	mC	231.68	229.99	96.57	89.9	90.4	92.1	C2
6	**7**	**oP**	**96.57**	**160.22**	**166.17**	**90.4**	**90.3**	**89.8**	**P222, P2221, P21212, P212121**
5	6	mP	96.57	166.17	160.22	90.4	89.8	90.3	P2, P21
4	5	mP	96.57	160.22	166.17	90.4	90.3	89.8	P2, P21
3	4	mP	160.22	96.57	166.17	90.3	90.4	89.8	P2, P21
2	1	ap	96.57	160.22	166.17	90.4	90.3	89.8	P1
1	0	aP	96.57	160.22	166.17	89.6	89.7	89.8	P1

Following processing with Mosflm the data was scaled and merged using the program Scala from the CCP4 (Collaborative Computing Project Number 4, 1994) suite. The scaled and merged intensities generated from Scala were converted to structure factor

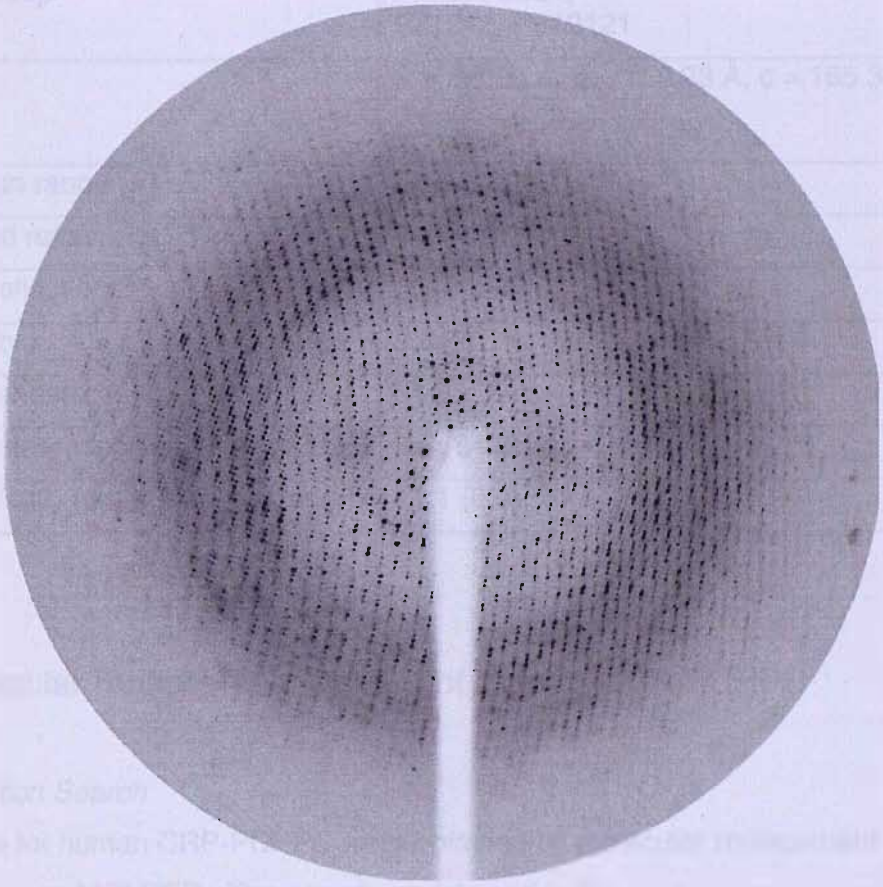


Figure 4.5 Diffraction image collected from the CRP-PCHPC crystal with 1.0° oscillations; collected at the ESRF, Grenoble, France.

amplitudes using Truncate (CCP4, 1994). The statistics of the data set are shown in Table 4.2:

Table 4.2 Data Processing Statistics for the CRP-PCHPC crystal, collected at 1.0° oscillations.

Parameter	Value
Space group	P222, P2221, P21212, P212121
Unit cell	a = 96.00 Å, b = 158.23 Å, c = 165.33 Å α = 90°, β = 90°, γ = 90°
Resolution range (Å)	55 – 2.3
Measured reflections	640652
Unique reflections	101840
Multiplicity	3.0
Completeness (%)	81.0
R_{merge} (outer resolution shell) (%)	8.7 (84.5)
Mean (I)/sd(I) (outer resolution shell)	11.1 (0.9)

4.3.2 Molecular Replacement Studies of Human CRP-PCHPC

4.3.2.1 Rotation Search

Initial phases for human CRP-PCHPC were obtained by molecular replacement calculations using MOLREP. The search model used in the rotation search was a single pentamer from the refined structure of human CRP to 2.0 Å with the calcium atoms omitted (accession code 1B09; Thompson *et al.*, 1999). The solvent content of the unit cell was estimated to be 75% for a single CRP pentamer in the asymmetric unit and 50% for two pentamers in the asymmetric unit (Matthews, 1968). From this calculation it was therefore concluded that there were two CRP pentamers in the asymmetric unit. The rotation search was performed using reflections within the resolution range 46 Å to 3 Å and yielded only one significant solution (Table 4.3).

Table 4.3 The ten highest peaks following rotation calculations performed using data between 46 Å and 3 Å (α, β, γ correspond to CCP4 Eulerian angles).

Peak Number	α	β	γ	Rf/σ
1	112.88	90.00	79.89	10.20
2	99.76	87.42	64.53	8.79
3	132.32	90.00	112.06	8.71
4	102.80	85.62	59.04	8.67
5	73.59	84.65	28.08	8.52
6	97.44	90.00	67.29	8.45
7	133.61	90.00	219.69	8.37
8	143.67	90.00	121.43	7.96
9	138.32	90.00	115.06	7.92
10	154.13	90.00	65.86	7.81

Despite various attempts, successful translation solutions could not be found using this data set. It was thought that this may be due to the low completeness value of the data or that the translation function was being carried out in an incorrect space group. Due to missing reflections, analysis of the data for screw axes using HKLVIEW on plane *h0l* was unable to provide evidence as to the correct space group ($P2_1$, $P2_122$, $P2_12_12$ or $P2_12_12_1$). Despite attempting translation searches in all possible space groups, adjusting all possible parameters of the search including radius of integration, resolution range, search model and even molecular replacement program (CNS, Refmac) a suitable solution could not be attained using this data set. It was therefore decided to collect diffraction data using a smaller oscillation angle (0.5°) in an attempt to reduce spot overlap, thereby increasing data completeness.

4.3.2.2 Self Rotation

Due to problems in finding a molecular replacement solution it was decided to perform a self-rotation function in order to determine the nature of any non-crystallographic symmetry elements. This was performed using the CCP4 program POLARRFN. Reflections within the range 15-2.3 Å were included in the calculation with a radius of integration of 20 Å. Stereographic images calculated with spherical polar angles revealed the presence of peaks in the $\kappa = 72$ section, as expected for the presence of a single non-crystallographic 5-fold axis, Figure 4.6. The images also revealed that the 5-fold axis of the pentamer is close to the crystallographic 2-fold of the b axis.

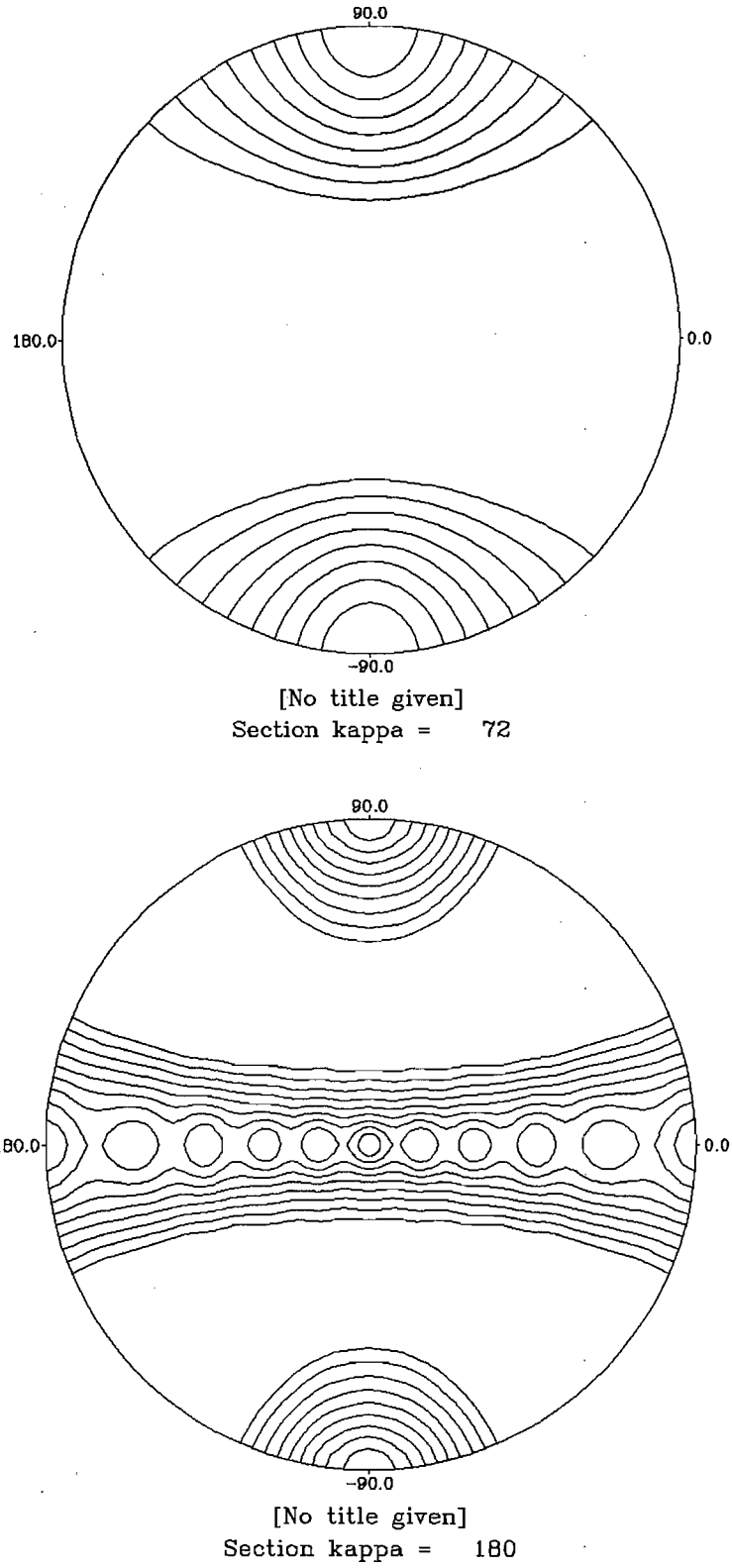


Figure 4.6 Self-rotation function in spherical polar angles calculated using Polarrfn.

4.4 Data Set II

4.4.1 Data Collection and Processing

A second data set was collected using the same crystal at station ID23 at the ESRF in Grenoble, France. A total of 185 0.5° oscillation frames were measured, with an exposure time of 3 seconds per frame and a crystal to detector distance of 200 mm, Figure 4.7. The raw data was once again processed with Mosflm (Leslie, 1992) and scaled and merged using the program Scala from the CCP4 (Collaborative Computing Project Number 4, 1994) suite. The scaled and merged intensities generated from SCALA were converted to structure factor amplitudes using Truncate (CCP4, 1994). The statistics of the data set are shown in Table 4.4:

Table 4.4 Data Processing Statistics for the CRP-PCHPC crystal, collected at 0.5° oscillations.

Parameter	Value
Space group	P222, P2221, P21212, P212121
Unit cell	a = 96.00Å, b = 158.22 Å, c = 165.33Å α = 90°, β = 90°, γ = 90°
Resolution range (Å)	60 – 2.3
Measured reflections	842376
Unique reflections	112504
Multiplicity	7.5
Completeness (%)	99.7
R_{merge} (outer resolution shell) (%)	11.7 (73.8)
Mean (I)/sd(I) (outer resolution shell)	17.0 (2.5)

The data was processed in P222 and analysed using the program HKLVIEW, revealing the presence of three screw axes along *h*, *k* and *l* (Figure 4.8); it was therefore decided to carry out molecular replacement in space group $P2_12_12_1$.

4.4.2 Molecular Replacement Studies

4.4.2.1 Rotation Search

Once again the rotation and translation searches were performed using the program MOLREP. Rotation searches were carried out on the data using both a CRP pentamer and a single CRP protomer (1B09; Thompson *et al.*, 1999), with calciums omitted, as

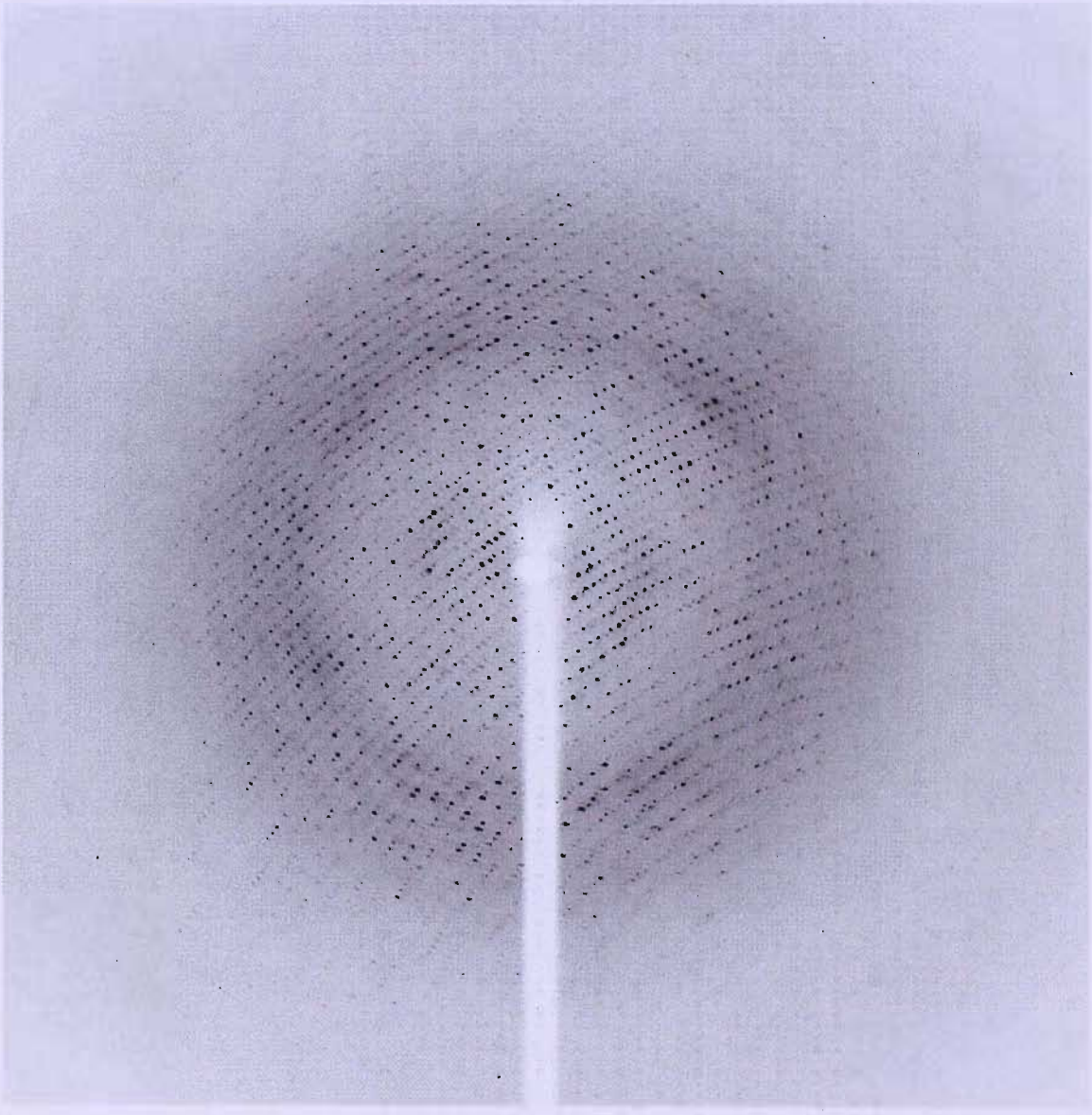


Figure 4.7 Diffraction image collected from the CRP-PCHPC crystal with 0.5° oscillations; collected at the ESRF, Grenoble, France.

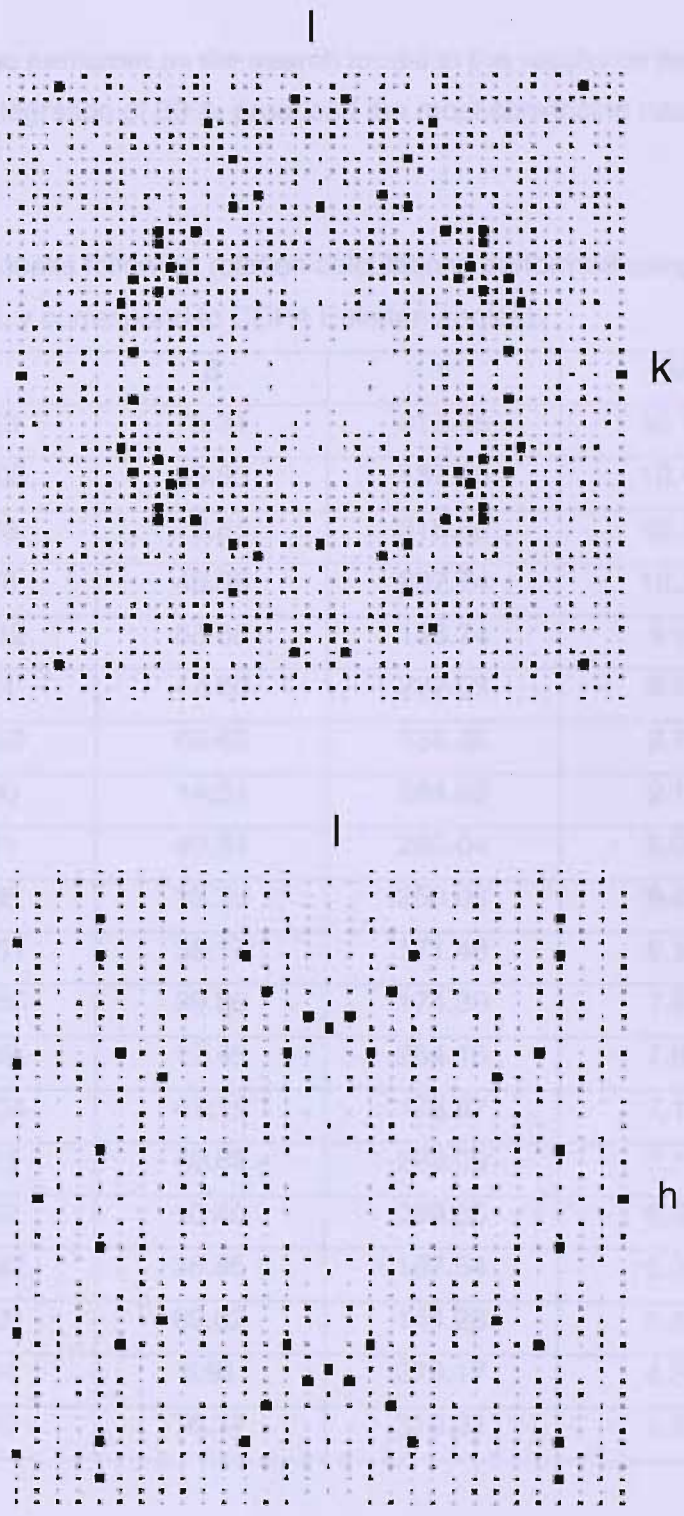


Figure 4.8 Precession camera images of the CRP-PCHPC diffraction data generated using HKLVIEW in P222. **Top** The 0kl plane . **Bottom** h0l plane. Systematic absences are observed along along all three axis indicating that the correct space group is $P2_12_12_1$.

the search model. Using the pentamer as the search model in the resolution range 3-58 Å and with a radius of integration of 30 Å, produced the most convincing rotation solution (Table 4.5).

Table 4.5 The ten highest peaks following rotation calculations performed using data between 3 Å and 58 Å (α, β, γ correspond to CCP4 Eulerian angles).

Peak Number	α	β	γ	Rf/σ
1	25.23	76.31	317.45	10.77
2	155.05	65.83	137.25	10.41
3	11.43	76.51	316.68	10.37
4	25.75	48.06	287.74	10.34
5	169.12	65.65	136.74	9.95
6	39.21	47.89	287.93	9.83
7	169.63	65.63	136.82	9.78
8	61.30	14.54	284.59	9.16
9	28.11	49.34	285.04	9.06
10	73.66	13.79	250.08	8.49
11	133.51	38.17	171.48	8.15
12	130.55	39.86	174.39	7.94
13	69.69	12.45	254.15	7.86
14	129.04	41.15	176.87	7.10
15	41.73	49.62	285.09	7.10
16	63.62	10.80	259.05	6.98
17	136.42	36.85	167.54	6.31
18	151.01	69.02	144.23	5.63
19	42.64	6.91	279.17	4.34
20	39.30	76.37	318.21	4.27

4.4.2.2 Translation Search

Despite finding this rotation solution, once again the translation search proved less successful. The first of two the pentamers was repeatedly found on successive attempts, but when searching for the second pentamer, Molrep was unable to improve the model (R factor and correlation coefficient) with any of the possible solutions calculated. A translation solution was found using the first and third peaks from the rotation function, at a resolution range of 3-58 Å and a radius of integration of 30 Å (Tables 4.6 and 4.7).

Table 4.6 The five highest translation peaks calculated in the space group $P2_12_12_1$ using the first orientation of the rotation search ($\alpha = 25.23°$, $\beta = 76.31°$, $\gamma = 317.45°$).

Peak Number	X	Y	Z	Dens/σ
1	0.064	0.370	0.274	24.25
2	0.035	0.370	0.274	11.50
3	0.064	0.022	0.274	11.06
4	0.063	0.496	0.274	10.70
5	0.063	0.870	0.161	10.65

Table 4.7 The five highest translation peaks calculated in the space group $P2_12_12_1$ using the third orientation of the rotation search ($\alpha = 11.43°$, $\beta = 76.51°$, $\gamma = -43.32°$).

Peak Number	X	Y	Z	Dens/σ
1	0.515	0.904	0.711	96.74
2	0.515	0.406	0.711	38.12
3	0.015	0.904	0.711	29.92
4	0.015	0.406	0.711	29.89
5	0.516	0.907	0.678	25.13

Despite having a relatively low correlation coefficient and R factor of 35.7 and 51.8 respectively, the suggested model was checked using a packing diagram generated by Pymol (Delano, 2002) shown in Figure 4.9. The diagram confirmed the validity of the solution and confirmed the presence of two CRP pentamers in the asymmetric unit. It was therefore decided to proceed with the refinement process using this solution as the model.

4.4.3 Refinement of CRP-PCHPC Model

The model was initially refined with the program CNS using 111987 reflections in the resolution range 15-2.3 Å. The test set of reflections used for the R_{free} calculation was chosen using the CNS program make_cv.inp. The first round of refinement consisted of rigid body refinement, simulated annealing (slow cool from 5000 to 500 K in 25 K decrements) and grouped B factor refinement. Full strength NCS restraints were imposed on the five CRP protomers for the first round of refinement. Rigid body refinement surprisingly reduced the R factor by 10% from 46.38 to 37.11% (R_{free} 46.85-

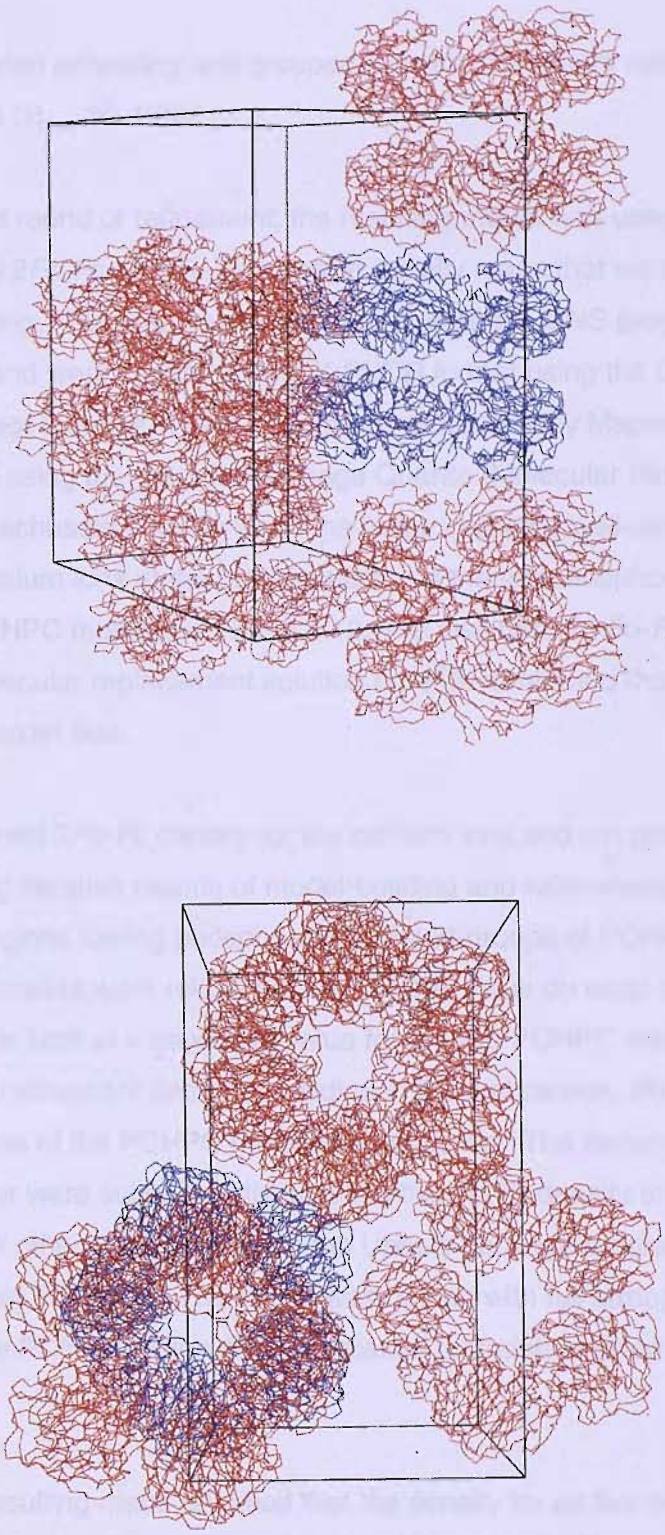


Figure 4.9 Packing diagram for the CRP-PCHPC crystal. **Top** Side view of the unit cell showing a CRP decamer highlighted in blue. **Bottom** View down the 5-fold axis of the decamer.

36.76%). Simulated annealing and grouped B factor refinement reduced the R factor further to 28.66% (R_{free} 30.64%).

Following the first round of refinement, the resulting model was used to calculate sigmaA weighted 2*Fo*-*Fc* and *Fo*-*Fc* electron density maps that were subjected to five-fold NCS averaging. The maps were calculated using the CNS program model_map.inp and were converted into X-PLOR format using the CCP4 program Mapman (Kleywegt & Jones, 1994). The maps generated by Mapman were analysed and manipulated using the graphics package Quanta (Molecular Simulations, Inc Burlington, Massachusetts). Analysis of the maps indicated well-defined *Fo*-*Fc* density for the twenty calcium ions at the binding sites and the ten phosphocholine head groups of the PCHPC molecules Figure 4.10. The well defined *Fo*-*Fc* density not only validated the molecular replacement solution but also confirmed that the maps were not affected by model bias.

Despite well-defined 2*Fo*-*Fc* density for the calcium ions and ten phosphocholine moieties following iterative rounds of model-building and refinement, density for the aliphatic linker regions joining phosphocholine head-groups of PCHPC remained absent. NCS restraints were relaxed to the minimal value on each of the protein subunits, but were kept at a maximum value for the five PCHPC molecules. Following the fifth round of refinement and after building in the first carbon, *Fo*-*Fc* density for the linker region in one of the PCHPC molecules appeared. The remainder of the carbon atoms in the linker were subsequently built into the *Fo*-*Fc* density and were copied across to the four other molecules using the Upsala package program LSQKAB and the CCP4 program PDBSET. The model was refined with full strength NCS restraints applied to the five PCHPC molecules and relaxed restraints applied to the ten protein subunits.

Analysis of the resulting maps revealed that the density for all five of the PCHPC molecules had completely disappeared and that the carbon atoms comprising the aliphatic linker had moved to positions of disallowed bond length and angle. It was therefore apparent from this observation that the five PCHPC molecules in the model were quite different between subunits and that subsequent refinement rounds should proceed with minimal NCS restraints imposed on them. Following this protocol in the subsequent refinement round resulted in the appearance of well-defined 2*Fo*-*Fc* electron density for all six carbon atoms in the aliphatic linker of all five PCHPC molecules (Figure 4.11).

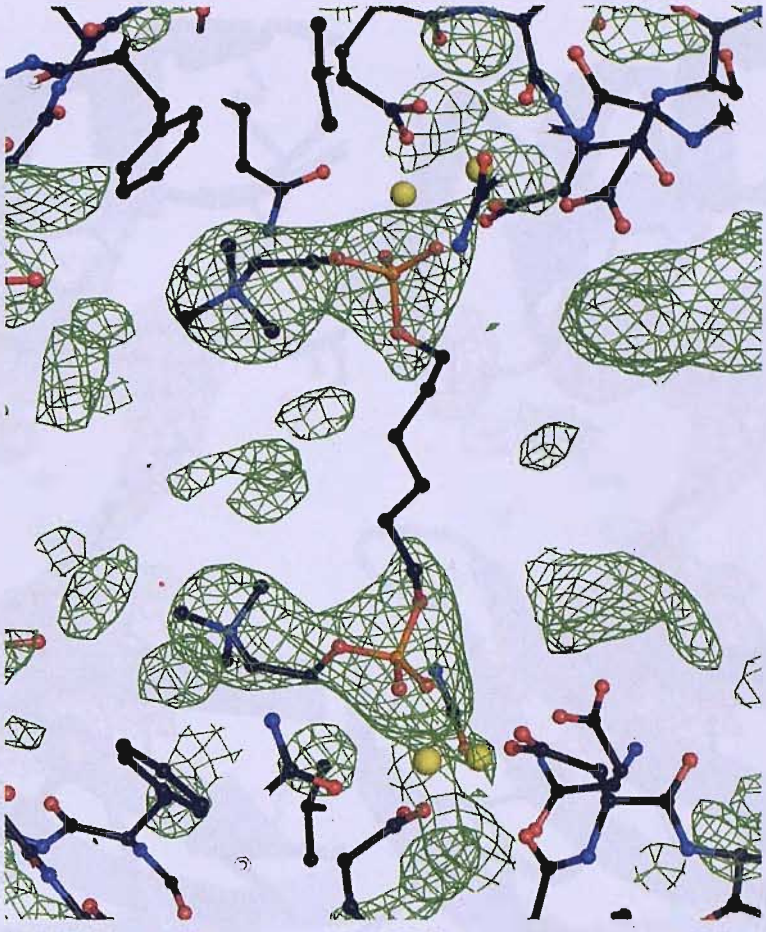


Figure 4.10 Electron density (Fo-Fc) contoured at 3 σ of the calcium binding site before the addition of PCHPC. Overlay of the PCHPC molecule clearly shows the density covering the choline moiety of the phosphocholine residues.

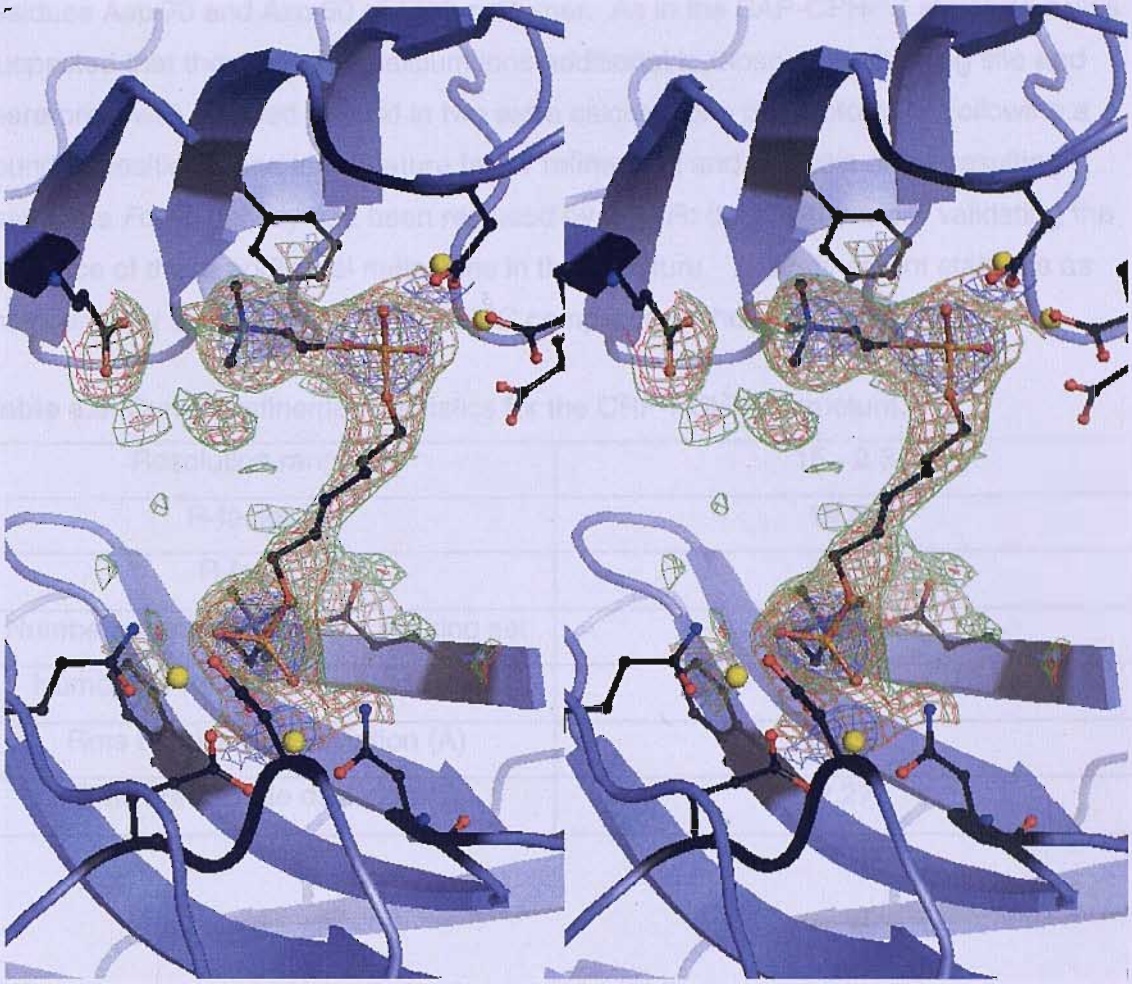


Figure 4.11 Stereo view of electron density (2Fo-Fc) contoured at 3σ (blue), 1σ (red) and 0.75σ (green), and the fitted PCHPC molecule.

The maps also revealed the presence of two large spheres of *Fo-Fc* density bound to residues Asp 70 and Asp 60 of each protomer. As in the SAP-CPHPC structure it was suspected that these may be calcium ions additional to those at the binding site and therefore it was decided to build in two extra calcium ions per protomer. Following a round of positional and temperature factor refinement and analysis of the resulting maps, the *Fo-Fc* density had been replaced by *2Fo-Fc* density therefore validating the presence of these additional metal ions in the structure. The refinement statistics as they currently stand for the CRP-PCHPC complex are shown in Table 4.8.

Table 4.8 Current refinement statistics for the CRP-PCHPC structure.

Resolution range (Å)	15 - 2.3
R-factor (%)	19.02
R-free (%)	19.74
Number of reflections in the working set	106338
Number of reflections in the test set	5649
Rms bond length deviation (Å)	0.037
Rms bond angle deviation (°)	2.27

4.5 Structure of Human CRP-PCHPC Complex

The X-ray crystal structure of the CRP-PCHPC complex reveals a decamer composed of two CRP pentamers cross-linked via their five phosphocholine binding sites by five PCHPC molecules (Figure 4.12). The overall dimensions of the complex are 93 x 93 x 75 Å, and the model shows good stereo-chemical properties with 84.8 % of residues in most favoured regions, 14.6 % in additional allowed regions, 0.5 % in generously allowed regions and 0.1 % in disallowed regions (Figure 4.13).

A comparison of temperature factors for residues across all ten CRP subunits in the complex (Figure 4.14) reveals a relatively high average value of 47.5 $Å^2$. This observation suggests that the CRP decamer formed on binding PCHPC, is a highly flexible complex. As shown on the graph, regions of the protomer that seem to have particularly high temperature factor values are residues 24-26, 90-94, and 172-178. As might be expected, these residues belong to flexible loop regions at the surface of the protomer, adjoining adjacent strands. Residues 24-26 constitute a loop linking strands B and C, residues 90-94 link strands G and H, and residues 172-178 are found at the

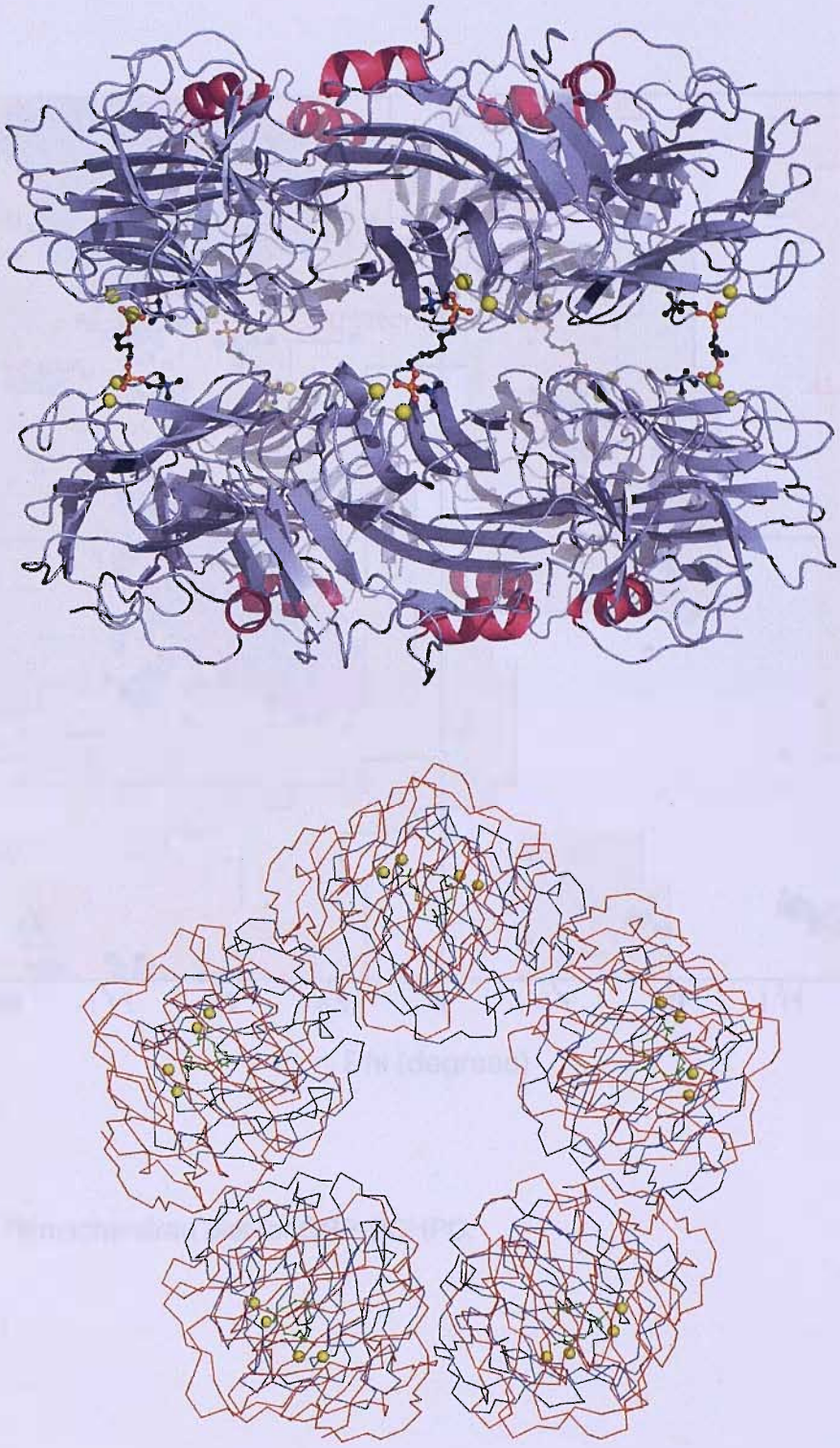


Figure 4.12 Structure of the CRP-PCHPC complex. **Top** View of the complex perpendicular to the five-fold axis showing two CRP pentamers joined B face to B face via five PCHPC molecules. A-face helices are shown in pink. **Bottom** View down the five-fold axis with one pentamer shown in blue and one in red. Calciums are shown in yellow and PCHPC molecules in green.

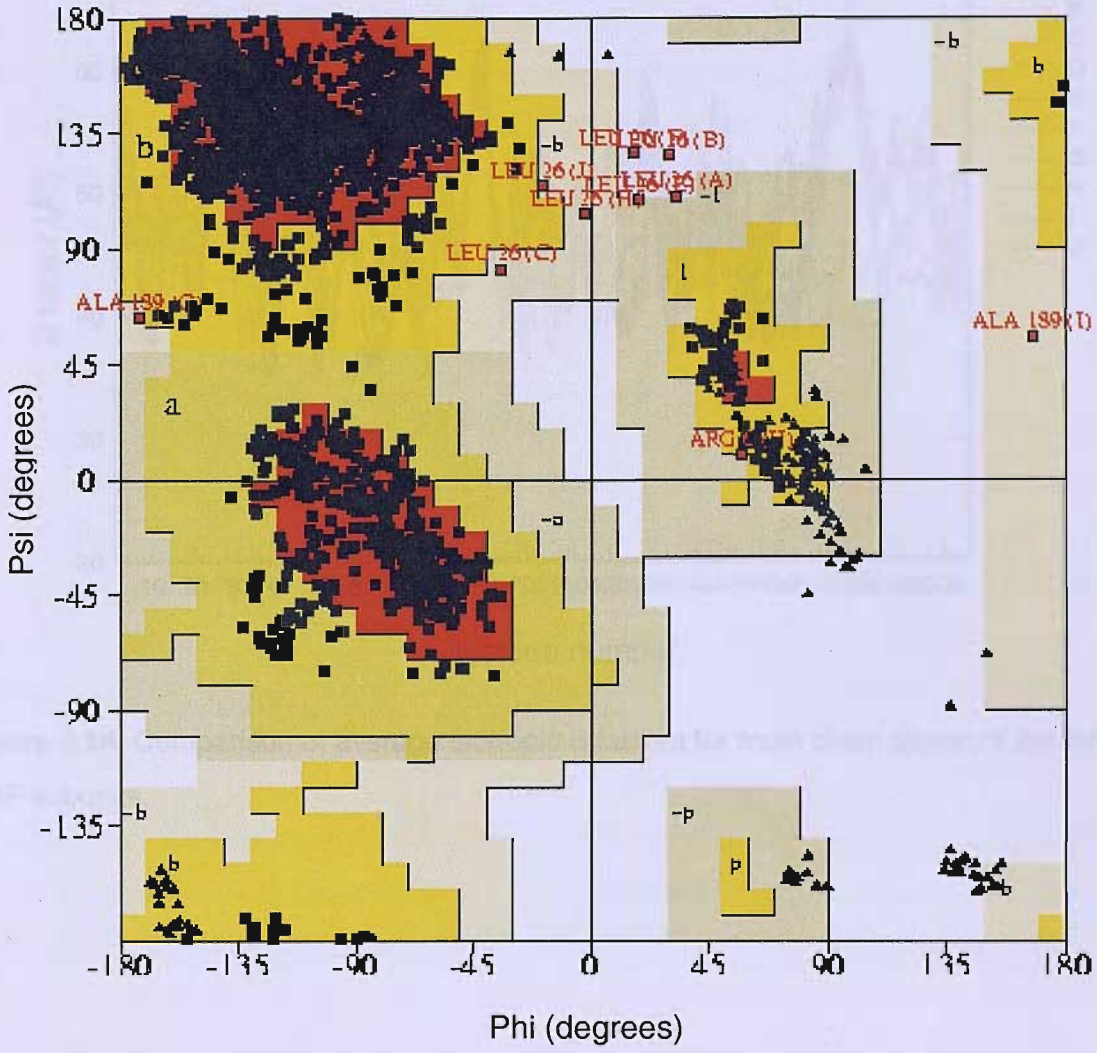


Figure 4.13 Ramachandran plot for CRP-PCHPC.

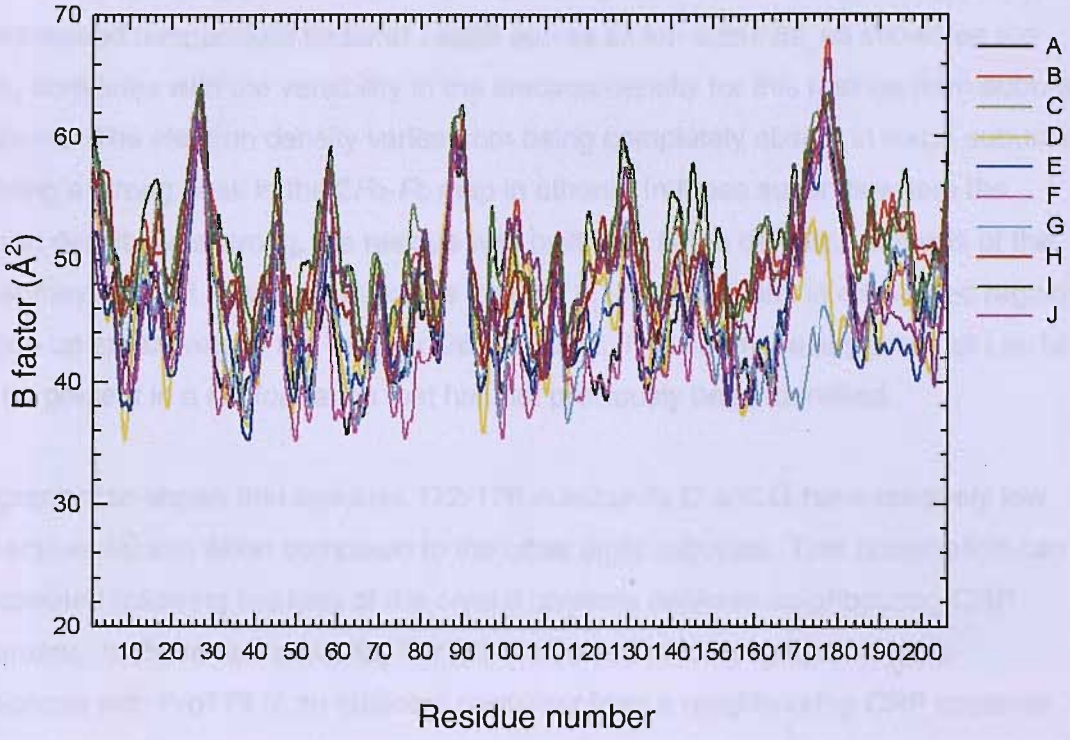


Figure 4.14 Comparison of average isotropic B factors for main chain atoms of the ten CRP subunits.

C-terminus of the single α-helix and the loop linking this to the 3-10 helix (Figure 1.16). The increased temperature factor of Leu26 across all ten subunits, as shown on the graph, correlates with the variability in the electron density for this residue from subunit to subunit. The electron density varies from being completely absent in some subunits to having a strong peak in the 2*Fo*-*Fc* map in others. In those subunits where the electron density was strong, the residue was built in to fit the density. Analysis of the Ramachandran plot, however, indicates that the 0.1% of residues in disallowed regions is made up exclusively of the ten Leu126 residues. This seems to suggest that Leu126 may be present in a conformation that has not previously been identified.

The graph also shows that residues 172-178 in subunits D and G have relatively low temperature factors when compared to the other eight subunits. This observation can be explained following analysis of the crystal contacts between neighbouring CRP pentamers. In these two subunits, Thr173 is involved in forming hydrophobic interactions with Pro179 of an adjacent pentamer from a neighbouring CRP decamer. The presence of these crystal contacts stabilises the loop region, therefore explaining the lower temperature factors of these residues.

Superimposition of the ten CRP subunits from the CRP-PCHPC complex reveals a Cα rms fit of 0.325 Å and as expected shows a strong similarity between subunits (Figure 4.15). Regions that exhibit the largest rms deviation (residues 24-26, 90-94 and 172-178) correspond to those with high temperature factors, indicating that these regions are highly flexible and may be present in alternative conformations in the ten subunits.

The component CRP pentamers in the reported structure are, as expected, very similar to those of Shrive *et al.* (1996) and the free phosphocholine complex (Thompson *et al.*, 1999). The pentamers are disc-like with their component subunits displaying pentameric radial symmetry. The B-face of a pentamer is made up exclusively of β-sheet and contains the five calcium binding sites. The opposite face of the pentamer (A face) contains five α-helices, one per subunit. As described in the two previously solved CRP structures, displayed on the A face of the two CRP pentamers is a characteristic deep cleft that runs from the centre of each protomer to the central pore of the molecule.

Protomer overlay between the CRP-PCHPC and CRP-PC structures produces an rms fit of approximately 0.48 Å, but a considerably worse fit is observed when overlaying

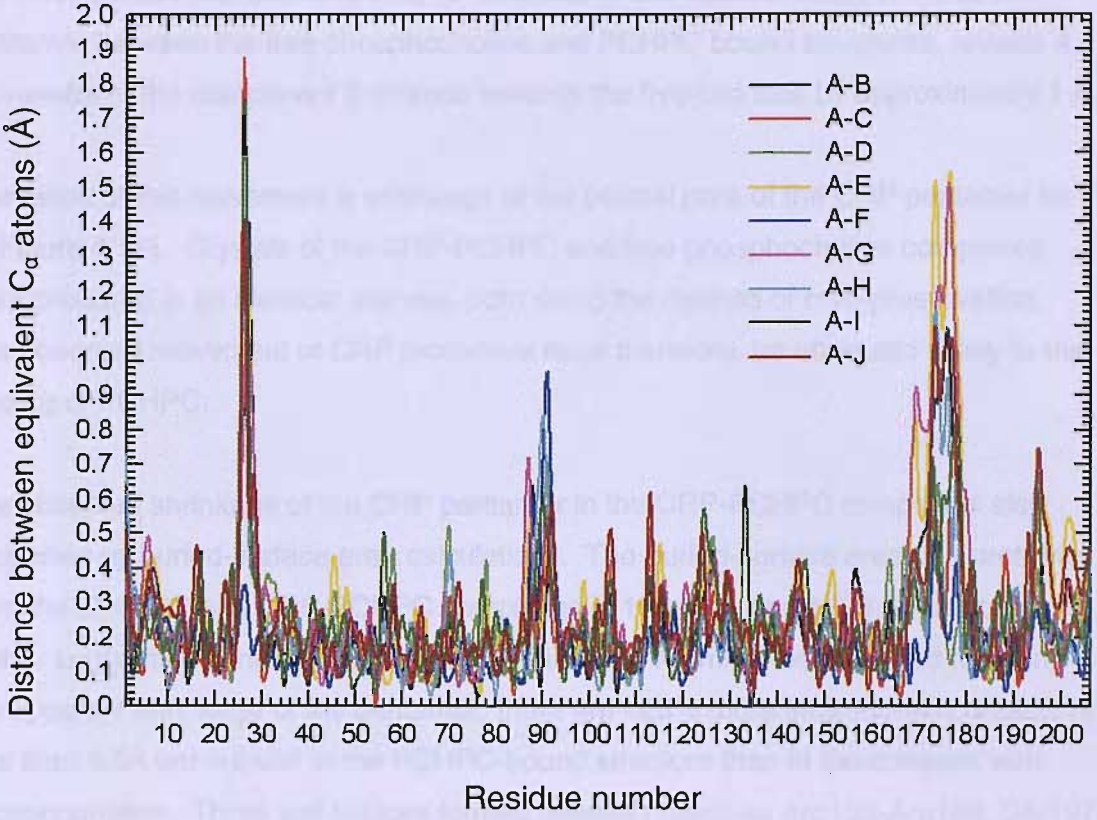


Figure 4.15 Comparison of C_α atom positions between CRP protomers in the CRP-PCHPC complex.

complete pentamers, rms fit of 0.85 Å. Analysis of this apparent shift in the CRP pentamer between the free phosphocholine and PCHPC bound structures, reveals a movement of the component β-strands towards the five-fold axis by approximately 1 Å.

The result of this movement is shrinkage of the central pore of the CRP pentamer by 2 Å (Figure 4.16). Crystals of the CRP-PCHPC and free phosphocholine complexes were prepared in an identical manner, both using the method of cryo-preservation. The observed movement of CRP protomers must therefore, be attributed solely to the binding of PCHPC.

The observed shrinkage of the CRP pentamer in the CRP-PCHPC complex is also supported by buried-surface area calculations. The buried-surface area of a protomer from the CRP-PC and CRP-PCHPC complexes is 1584.4 Å and 1733.6 Å respectively, further suggesting shrinkage of the CRP pentamer. As might be expected following the apparent shrinkage of the pentamer, there are ~25% more inter-atomic contacts of less than 3.5Å per subunit in the PCHPC-bound structure than in the complex with phosphocholine. Three salt bridges formed between residues Arg155-Arg188, Glu197-Lys123 and Lys201-Glu101 in the phosphocholine-free complex are conserved in the CRP-PCHPC structure and also present are two additional salt bridges involving Glu42-Lys119 and Glu42-Arg116.

Plotting the main-chain rms deviation between protomers taken from the CRP-PCHPC complex and CRP-PC complexes highlights the overall similarity of the two structures. Regions exhibiting the largest main chain rms deviation are residues 26-30, 86-90 and 172-180. Analysis of the phi and psi angles for these residues indicates that there are no changes to the secondary structure of CRP in the CRP-PCHPC complex when compared to that of the CRP-PC complex. The observed differences in these regions must therefore, reflect the high flexibility of these regions and suggests that the component residues are able to adopt a number of different conformations. This is consistent with the observed B factors of residues in these regions.

The ligand binding site contains two calcium ions that are positioned approximately 4 Å apart, each bound by a total of six ligands. Calcium I is bound by five protein ligands provided by the side-chains of Asp60, Asn61, Glu138, Asp140, and the main-chain of Gln139; the sixth ligand is provided by a terminal oxygen of the PCHPC molecule. Calcium II is bound by four protein ligands provided by the side-chains of Gln138, Asp140, Glu147 and Gln150. The fifth and sixth ligands are provided by a water

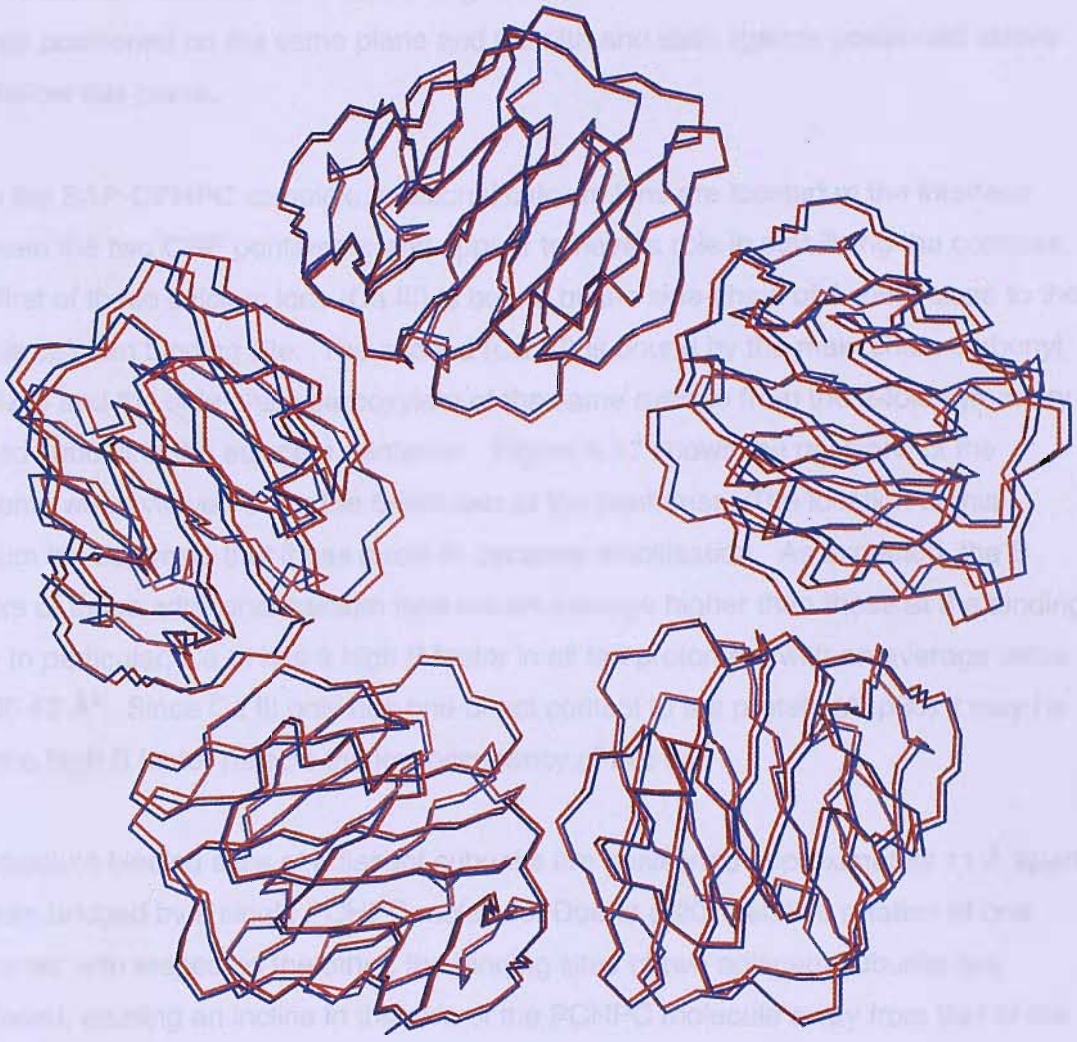


Figure 4.16 Shrinkage of the CRP pentamer in the CRP-PCHPC complex (shown in red) compared to that of the CRP pentamer from the CRP-PC complex (shown in blue).

molecule and a terminal oxygen of the PCHPC molecule. The coordination spheres of the two calcium ions are therefore arranged as an octahedron with four of the six ligands positioned on the same plane and the fifth and sixth ligands positioned above and below this plane.

As in the SAP-CPHPC complex, additional calcium ions are located at the interface between the two CRP pentamers and appear to have a role in stabilising the complex. The first of these calcium ions (Ca III) is bound by the side-chain of Asp60, close to the double calcium binding site. The second (Ca IV) is bound by the main chain carbonyl of Glu70 and the side-chain carboxylate of the same residue from the 2-fold symmetry related subunit in the adjacent pentamer. Figure 4.17 shows the positions of the calciums when viewed down the 5-fold axis of the pentamer. The location of this calcium ion suggests that it has a role in decamer stabilisation. As expected, the B factors of these additional calcium ions are on average higher than those at the binding site. In particular, Ca III has a high B factor in all ten protomers with an average value of 105.62 $Å^2$. Since Ca III only has one direct contact to the protein (Asp60) it may be that the high B factor reflects the low occupancy of this ion

The calcium binding sites of adjacent subunits are positioned approximately 11 Å apart and are bridged by a single PCHPC molecule. Due to a 20 ° relative rotation of one pentamer with respect to the other, the binding sites of two adjacent subunits are displaced, causing an incline in the axis of the PCHPC molecule away from that of the five-fold axis of the decamer. It is unclear as to whether this observation is an artefact of PCHPC binding or if decamer formation requires the pentamers to adopt this arrangement therefore inducing the incline in the axis of the PCHPC molecule.

The calcium ions at the binding site are bridged by the two free oxygen atoms of the terminal phosphate group in PCHPC. The choline moiety of the ligand is attached to the third oxygen atom and as in the previously reported structure of CRP-PC, runs along the surface of the protomer forming hydrophobic contacts with Phe66 and electrostatic interactions with the side-chain of Glu81. Attached to the fourth oxygen is the aliphatic linker. As opposed to the phosphocholine component and first carbon of the linker which have very good electron density, the density for the four central carbon atoms is relatively weak. Presumably, this is due to high flexibility in the cross-linker. As in the SAP-CPHPC complex, in order to fit the linker atoms into the available space and to facilitate the required approach to the phosphate groups, the C3-C4 bond had to be built in an unfavourable eclipsed rotamer.

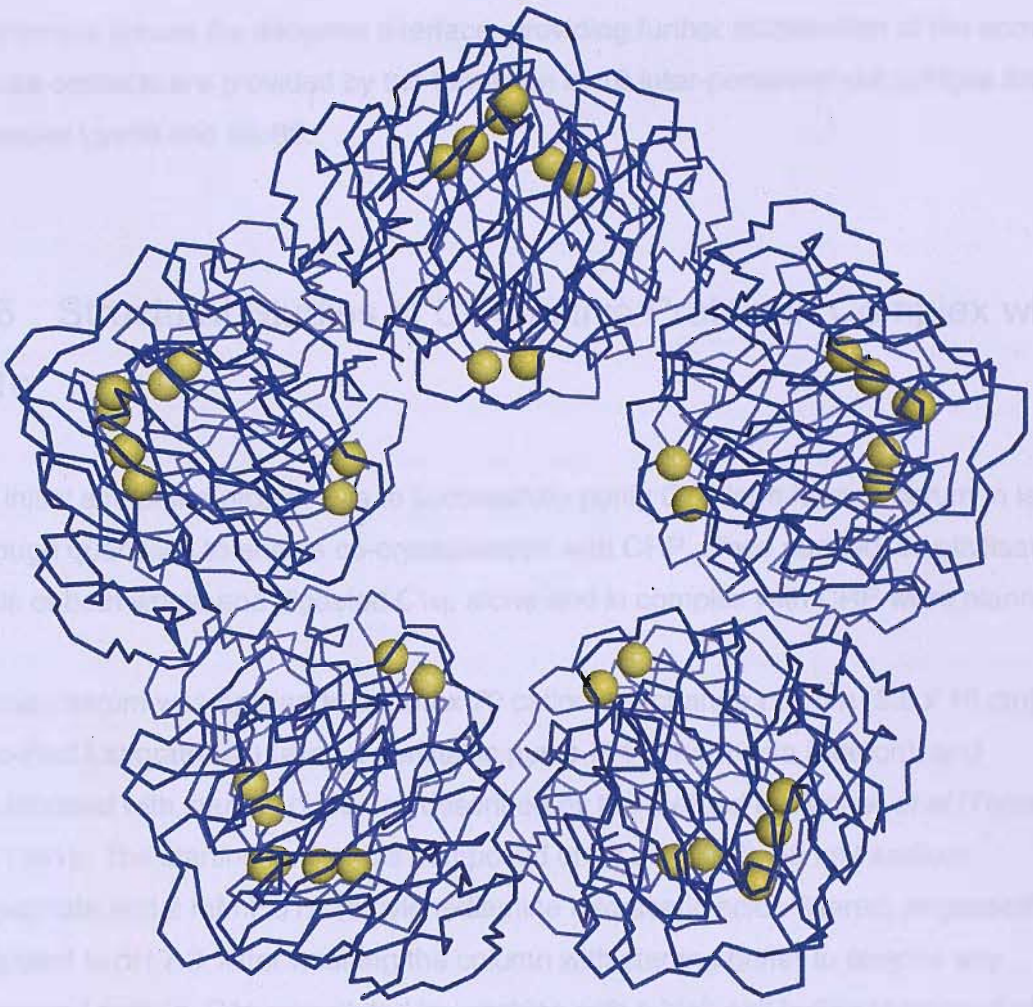


Figure 4.17 Distribution of all calcium ions across the CRP decamer.

In addition to decamer stabilisation resulting from the cross-linking effect of PCHPC and the extra calcium ions, protein-protein contacts are also found between the CRP pentamers across the decamer interface, providing further stabilisation of the complex. These contacts are provided by the formation of 10 inter-pentamer salt bridges formed between Lys69 and Glu85.

4.6 Structural Studies of C-Reactive Protein in Complex with C1q

An initial aim of the project was to successfully purify C1q from human serum in large enough quantities to enable co-crystallisation with CRP. Once purified, crystallisation trials of both whole and digested C1q, alone and in complex with CRP were planned.

Human serum was applied to a Biorex 70 cationic exchange column (5.6 x 18 cm) (Bio-Rad Laboratories) using a peristaltic pump P-1 (Pharmacia Biotech), and equilibrated with starting buffer, as described by the method of Tenner *et al* (Tenner *et al*, 1981). The starting buffer was composed of 82 mM NaCl, 50 mM sodium phosphate and 2 mM EDTA (ethylenediamine tetraacetic acid), filtered, degassed and adjusted to pH 7.3. After washing the column with starting buffer to remove any unwanted protein, C1q was eluted by washing with a high salt buffer composed of 50 mM sodium phosphate, 500 mM NaCl and 2 mM EDTA.

Fractions obtained from ion-exchange chromatography were pooled and concentrated using a stirred cell (Amicon). Nitrogen gas at a pressure of 50 bars was applied to the cell, driving molecules with a molecular weight <10kDa through the membrane. The samples collected were then further concentrated by centrifugation. The samples were transferred to centricoms in 2ml volumes, placed in a Sigma 4K10 centrifuge (B.Braun Biotech) at 4°C and centrifuged at 5000 G force for approximately 30 mins.

Electrophoresis of the samples was carried out under denaturing conditions, thus separating any C1q into its three component polypeptide chains A, B and C. The denaturing agent used was sodium dodecyl sulphate (SDS). Following analysis of the gel and a positive identification of C1q, the final step of the purification process was gel filtration. This was used to assess and further refine the purity of the C1q sample. A

Superose 6 column was used in the purification of C1q. The column was first equilibrated with buffer composed of 2 mM EDTA, 50 mM sodium phosphate and 500 mM NaCl. 200 µl samples of the protein were then injected into the column one at a time and the elutions recorded. After analysis of the resulting peaks by SDS–PAGE, the fractions containing C1q were pooled.

The partial digestion of C1q protein was carried out using a modified version of a method by Reid, 1976. 1/3 mg of pepsin was added to 300 µl of 100 mM sodium acetate pH 4.5. This was then added to 200 µl of purified C1q protein, and placed in a water bath at 37°C. In order to follow the digestion process, 10 µl samples were extracted from the reaction mixture at hourly intervals for a total of 20 hours. SDS–PAGE electrophoresis was then carried out on the samples and the gels analysed for the digestion products.

Despite various attempts at the purification of C1q from human serum, none of them yielded quantities large enough to pursue crystallisation trials. The final samples also seemed to be contaminated with an unknown protein, a factor that would have hindered the crystallisation process. In order to remove this contaminating protein from the sample, It was planned to introduce an alternative step into the purification process. It was decided to attempt ammonium sulphate precipitation to concentrate the sample after ion-exchange chromatography instead of centrifugation. However, as the project moved away from C1q, towards the crystallisation of SAP-CPHPC and CRP-PCHPC, this was never fully attempted.

4.7 Discussion

The crystal structure of the CRP-PCHPC complex validates the original hypothesis behind the design of PCHPC: a decamer of two CRP pentamers crosslinked B face to B face by five PCHPC molecules. As with the design of its sister compound CPHPC, it is proposed that decamer formation not only renders CRP unable to bind to its natural ligand, but also leads to the clearance of CRP molecules in this 'unnatural' decameric form. This work therefore represents the first rational, structure-led design of a drug molecule targeted at the inhibition of human CRP and identifies a potentially valid therapeutic approach to cardioprotection in acute MI, neuroprotection in stroke, and

other inflammatory, infective and tissue-damaging conditions associated with elevated levels of CRP. Statin therapy currently administered for the prevention of cardiovascular disease via a LDL-lowering action, has been shown to decrease CRP levels. However, this is acknowledged as a pleiotropic effect not part of the original hypothesis describing statin therapy, and of which the precise mechanism of action remains unclear.

Although PCHPC is still to enter clinical testing, preclinical experiments in MI rat models have provided encouraging results. PHCPC infusion initiated prior to coronary artery ligation followed by subcutaneous injection of human CRP, resulted in no deaths compared to models receiving only human CRP in which there was an increased mortality compared to control. In addition, infarct size was significantly larger in rats receiving CRP only, when compared to rats receiving PCHPC and CRP in which infarct size was the same as control rats.

The structural information provided in this report detailing precise atomic interactions and the general morphology of the complex will enable the optimisation of PCHPC. A disappointing aspect of the research has been the magnitude of the increase in the binding affinity of PCHPC when compared to the affinity to which monovalent PC binds CRP. The work on the SAP-CPHPC complex showed a 1000-fold increase in the binding affinity of the bivalent form of N-acetyl D-proline (15 μM) compared to the monovalent form (10 nM). The work on the CRP-PCHPC complex has however, only shown a 10-fold increase in the binding of PCHPC (300 nM) compared to PC (10-20 μM).

Since the addition of the aliphatic linker is the only difference between PC and PCHPC, it seems logical to assume that this may be the source of the apparent loss in affinity. A 6 carbon aliphatic linker (2 extra carbons) was used in the design of PCHPC in order to accommodate the 22° relative rotation of CRP and SAP protomers. This rotation culminates in the calcium binding sites of CRP being further away from each other and is therefore important in the design of potential ligands. As the structure reveals an unfavourable eclipsed rotamer about the C3-C4 bond, it was hypothesised that a shorter five carbon linker may provide an increase in binding affinity. Subsequent isothermal titration calorimetry studies to assess the binding of 1,5-bis(phosphocholine)-pentane to CRP have however, revealed that it binds with a substantially lower affinity (Kd ~ 3.7 μM), therefore disparaging this as a potential theory.

The work of Ho *et al.*, 2005 as described in section 3.7, suggests that the introduction of a more rigid group such as a piperazine ring into the linker may increase the binding affinity of a bivalent ligand. This was also disparaged as a method of increasing the affinity of PCHPC as bis(phosphocholine)-dimethylcyclohexane failed to bind CRP at all. An alternative method that has not been attempted to date is the utilisation of the small cavity formed opposite to PC by Glu81, Gly79, Asn61 and Thr76, originally identified by Thompson *et al.* (1999). The addition of a bulky group such as methyl to C2 in PCHPC would be expected to increase the affinity of the compound for CRP.

The reason for the shrinkage of the CRP pentamer relative to the CRP-PC structure is also unclear. Identical cryopreservation techniques were used for both structures and therefore the shrinkage must be a direct consequence of PCHPC binding. This, together with the increased B factor values across all residues in the structure, reveal that CRP-PCHPC is a highly flexible complex. This observation provides support to the theory that CRP and SAP protomers are able to alter their positions under the constraints of the pentamer to enable binding to multivalent ligands.

Chapter 5
General Discussion

The work presented in this report demonstrates the importance of structural biology to modern day drug design. The detailed molecular information provided by a three-dimensional structure enables the design and optimisation of compounds targeted at specific proteins of interest. There are advantages to both the structure-led or 'rational' design, and automated screening methods of drug identification. Screening of a compound library ensures that there is no bias in the selection of compound and may lead to the identification of novel ligands. This method proved beneficial in the case of SAP as it not only identified a novel ligand, but also suggested a recognition pattern through which SAP might bind to amyloid fibrils. In contrast to this was its unsuccessful use in the identification of a suitable compound for CRP.

The origin of this work was the elucidation of the X-ray crystal structure of SAP in 1994 (Emsley *et al.*, 1994). This enabled the characterisation and identification of known and novel ligands of SAP including PE, MOβDG, dAMP, and N-acetyl D-proline. Following screening of the Roche compound library for inhibitors of SAP binding to Aβ amyloid fibrils, the bivalent form of the first lead compound was more potent than the monovalent version. The molecular mechanism responsible for this greater potency was suggested by the previously solved structure of the SAP-dAMP complex, containing pairs of pentameric SAP molecules interacting face to face as a result of base-stacking interactions between the bound dAMP molecules. The crystal structure of the SAP-CPHPC complex (Purvis, 2002) subsequently confirmed this mechanism, and was to play an important role in the design of inhibitory compounds for CRP.

The search for a candidate compound for CRP began with an automated search of the Pfizer compound library that was unable to identify a target that bound with significant affinity (Pepys *et al.*, 2006). As a consequence, structural methods based on the previously solved CRP-PC and SAP-CPHPC structures were employed and suggested a bivalent compound of two PC residues joined by a six carbon aliphatic linker (PCHPC). In summation, X-ray crystallographic methods of structure determination have dramatically influenced the design of inhibitory compounds for both SAP and CRP, and are currently enabling the structure-led optimisation of both CPHPC and PCHPC.

The design of multivalent ligands has become increasingly popular in recent years due primarily to the enhanced binding affinity relative to corresponding monovalent interactions. A number of approaches have been employed in the design of these

compounds and this section will discuss how these alternative methods may be applied to the design of compounds for SAP and CRP.

Perhaps the most relevant and most interesting of these studies with respect to SAP and CRP, is research into the shiga-like toxins (SLT) from *E.coli* and cholera toxins (CT). Both of these belong to a subset of bacterial AB_5 toxins consisting of an enzymatically active A subunit that gains entry to cells after oligosaccharide recognition by the B_5 homopentamer. This recognition process and more specifically the B5 homopentamer, has been at the centre of the majority of structure-led drug design and therefore can be directly applied to the design of ligands for SAP and CRP.

The work on the design of an inhibitory molecule for shiga-like toxins began with the publication of the crystal structure of the SLT B-subunit pentamer in complex with a Gb3 trisaccharide analogue (glycolipid Gb3 is the toxin receptor on mammalian cells) (Ling *et al.*, 1998). As in SAP and CRP, the pentamer consists of five non-covalently associated subunits or protomers forming a doughnut-shaped B-subunit, and each protomer contains three saccharide-binding sites so that all 15 binding sites are aligned on the same face of the toxin. The opposing face of the pentamer is attached to the A subunit, therefore allowing all 15 binding sites to engage with ligand.

Binding studies, following the synthesis of a monomeric inhibitor, produced only a millimolar dissociation constant and subsequent functional group modification failed to increase the affinity constant to nanomolar activity. The group therefore switched the focus of their research to the design of a multivalent inhibitor tailored to the structure of the B-subunit pentamer.

The resultant compound, named STARFISH, consists of a pentameric display of bridged trisaccharide dimers at the tips of five appropriately oriented arms that radiate from a central glucose core (Kitov *et al.*, 2000) (Figure 5.1). Binding studies showed a significant increase in the affinity constant of STARFISH (0.4 nM for SLT-1; 6 nM for SLT-II) compared to both the univalent trisaccharide (mM) and the bridged trisaccharide dimer (10^{-5} M). The original hypothesis behind the design of

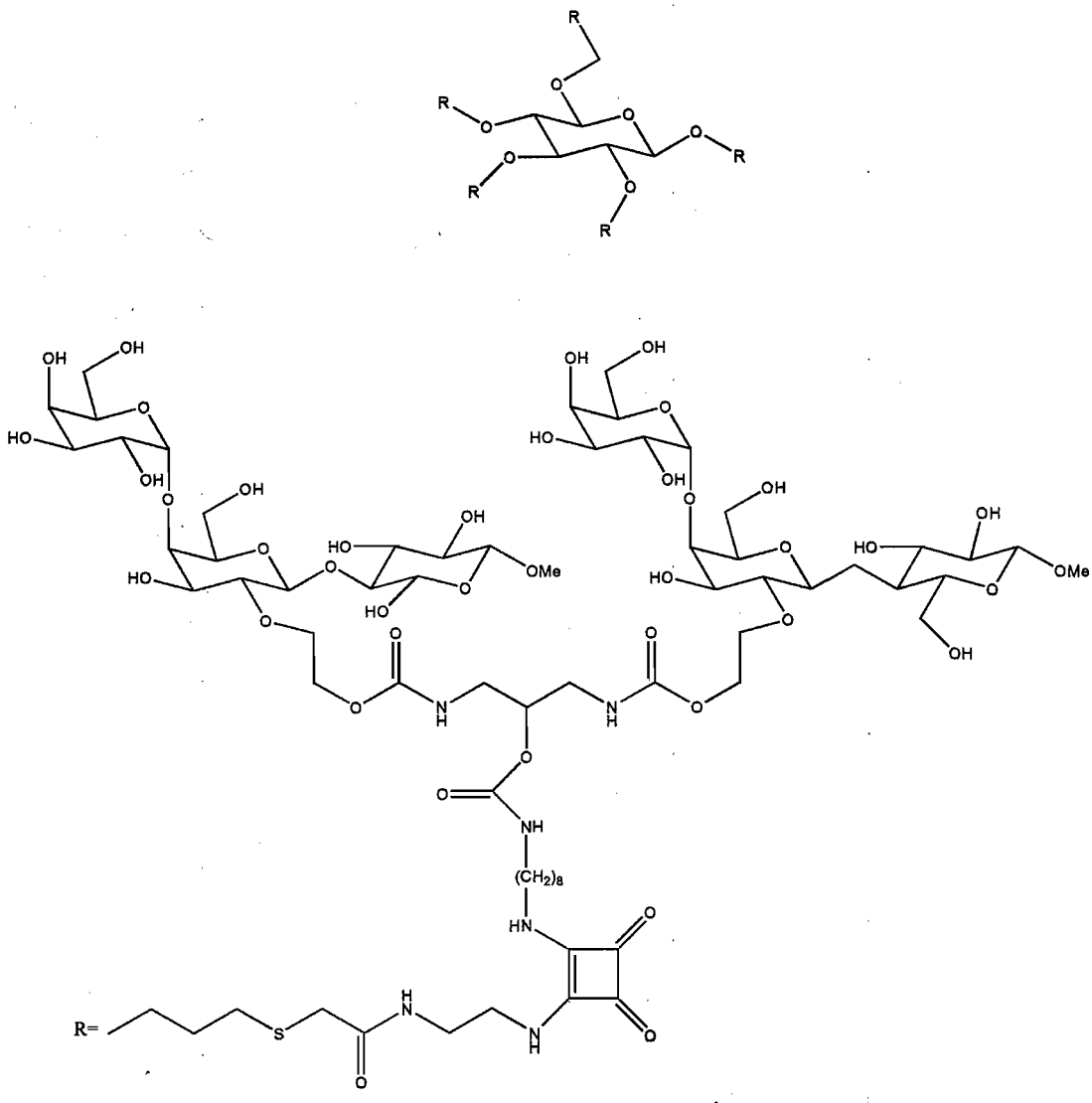


Figure 5.1 The STARFISH molecule..

this compound was that the divalent trisaccharide would simultaneously occupy sites 1 and 2 of a single B subunit; however, the crystal structure of the B-subunit pentamer/STARFISH complex revealed a different mode of binding. It revealed that one STARFISH molecule binds to not one but two B-subunit pentamers from two separate toxin molecules. Instead of bridging sites 1 and 2 on the same subunit the trisaccharides cross-link two subunits from separate toxin molecules in a mode that mirrors that of CPHPC and PCHPC.

Owing to the strong similarities between this and the work presented on CPHPC and PCHPC, the design of a multivalent ligand targeted at the pentameric structures of SAP and CRP incorporating CPHPC and PCHPC would be an attractive possibility. The preservation of the pentameric structures of SAP and CRP throughout evolution suggests that it is important to the physiological role of these proteins and in particular in its specificity for biological ligands. SAP binding to amyloid fibrils and CRP binding to membranes are both thought to be pattern-specific, regulated by the surface distribution of ligand. Therefore it is in my opinion that the design of inhibitory molecules must focus on the incorporation of the apparent physiological requirement for a symmetrical, pentameric distribution of ligand.

Although both CPHPC and PCHPC represent successful inhibitory compounds, further optimisation is still both possible and required. The ability of CPHPC to clear SAP from the circulation and therefore disrupt the equilibrium between plasma and amyloid pools has been proven (Pepys *et al.*, 2002). However, the ability of CPHPC to directly compete with SAP binding to amyloid fibrils is still in question. The design of an inhibitory compound that mimics the physiological ligand and therefore enables direct competition with amyloid fibrils would potentially reduce the stability of amyloid deposits and promote their regression.

Methods to improve the affinity constant of PCHPC for CRP would also be of interest and therefore the observed increase in the affinity constant of STARFISH compared to its monovalent equivalent is a factor that must be considered in the optimisation of PCHPC. PCHPC molecules displayed in a similar manner to STARFISH may more closely resemble the *in vivo* arrangement of its natural ligand on the surface of membranes, a factor that may also produce an increase in binding affinity.

The methods employed in the design of the STARFISH compound can also be directly applied to the design of similar compounds for SAP and CRP. As expected, analysis of the pentameric structures of all three proteins reveals that the respective distances from their ligand binding sites to the centre of the pentameric complex are very similar (SAP and CRP = ~35 Å; B5 pentamer = ~30 Å). Therefore the same glucose core could be used with the addition of five tailored arms, appropriately oriented to allow approach to the ligand binding site.

In the STARFISH molecule the arms are attached to the trisaccharide dimer through oxygen atoms positioned at the centre of the tether regions. This may pose a problem in the design of similar compounds for SAP and CRP as both CPHPC and PCHPC have aliphatic linkers with even numbers of carbon atoms. The attachment of an arm to a central atom in these molecules would not be possible without alteration of the linker region. Alteration of the tether length has already proved detrimental to the optimisation of PCHPC and it is likely that it would have a similar effect on CPHPC. Further investigation would be required to find a suitable solution to this problem.

An alternative method to the one used for the shiga-like toxins is that used for the design of compounds for the cholera toxins. The first step in the pathway of this toxin is the recognition and binding of the gangloside GM1 receptor in the small intestine. Inhibitors of this interaction are therefore an attractive route for the treatment of cholera (Pixens *et al.*, 2004).

In contrast to shiga-like toxins in which ligand design has centred around a compound with long arms radiating from a small core such as glucose, recent work on the cholera toxins has focused on the design of a compound with a large core such as that provided by a large ring of head-tail peptides. It has been proposed that the use of cyclic peptides in the design of multivalent ligands may provide two advantages over their small core counterparts. First, is the ability to incorporate a wide range of amino acids with differing side chains for the attachment of linkers and monovalent ligands. Second, the length of the cyclic peptide can be easily manipulated allowing for the vast diversity in the geometric requirements of different protein molecules (Zhang *et al.*, 2004).

The limiting factor in the use of these compounds for the design of inhibitory molecules for SAP and CRP is the maximum numbers of atoms that can be

incorporated into the cyclic peptide for it to remain in an expanded conformation. As shown by the work of Zhang *et al.* (2004) this seems to be dependant on the type of amino acid used in the construction of the ring. This work demonstrated the successful synthesis of 30-, 40-, and 50-atom rings constructed from lysine residues alternated with flexible amino acids such as glycine, γ-aminobutyric acid, or ε-aminohexanoic acid. These amino acids were chosen so as to enable the variation in ring size and also for the increased likelihood that they would adopt expanded conformations in solution.

Following synthesis, these compounds were assayed for their ability to inhibit the binding of the cholera toxin B pentamer to gangloside coated plates. The results suggest that the cyclic peptides retain an expanded conformation in solution as when compared to a smaller pentacyclen core previously assayed, progressively shorter linkers were required for each of the three ring sizes to achieve optimal inhibition.

Both of the techniques described could potentially be used in the optimisation of CPHPC and PCHPC as drug molecules for SAP and CRP. The size of these resultant compounds would not pose a problem in their use as potential drug molecules as they would not have to cross any biological membrane to exert their effect.

Chapter 6
References

Agrawal A, Xu Y, Ansardi D, Macon KJ, and Volanakis JE (1992). Probing the phosphocholine-binding site of human C-reactive protein by site-directed mutagenesis. *J Biol Chem,* **267,** 25353-25358.

Agrawal A and Volanakis JE (1994). Probing the C1q-binding site on human C-reactive protein by site-directed mutagenesis. *J Immunol,* **152,** 5404-5410.

Agrawal A, Shrive AK, Greenhough TJ, and Volanakis JE (2001). Topology and structure of the C1q-binding site on C-reactive protein. *J Immunol,* **166,** 3998-4004.

Agrawal A, Simpson MJ, Black S, Carey MP, and Samols D (2002). A C-reactive protein mutant that does not bind to phosphocholine and pneumococcal C-polysaccharide. *J Immunol,* **169,** 3217-3222.

Akiyama H, Yamada T, Kawamata T, and McGeer PL (1991). Association of amyloid P component with complement proteins in neurologically diseased brain tissue. *Brain Res,* **548,** 349-352.

Anderson JK, Stroud RM, and Volanakis JE (1978). Studies on the binding specificity of human C-reactive protein for phosphorylcholine. *Fed Proc,* **37,** 1495.

Arlaud GJ, Gaboriaud C, Thielens NM, Rossi V, Bersch B, Hernandez JF, and Fontecilla-Camps JC (2001). Structural biology of C1: dissection of a complex molecular machinery. *Immunol Rev,* **180,** 136-145.

Ashton AW, Boehm MK, Gallimore JR, Pepys MB, and Perkins SJ (1997). Pentameric and decameric structures in solution of serum amyloid P component by X-ray and neutron scattering and molecular modelling analyses. *J Mol Biol,* **272,** 408-422.

Baltz ML, De Beer FC, Feinstein A, and Pepys MB (1982). Calcium-dependent aggregation of human serum amyloid P component. *Biochim Biophys Acta,* **701,** 229-236.

Bang R, Marnell L, Mold C, Stein MP, Clos KT, Chivington-Buck C, and Clos TW (2005). Analysis of binding sites in human C-reactive protein for Fc{gamma}RI, Fc{gamma}RIIA, and C1q by site-directed mutagenesis. *J Biol Chem,* **280,** 25095-25102.

Benson MD, Scheinberg MA, Shirahama T, Cathcart ES, and Skinner M (1977). Kinetics of serum amyloid protein A in casein-induced murine amyloidosis. *J Clin Invest,* **59,** 412-417.

Bergfors T (2003). Seeds to crystals. *J Struct Biol,* **142,** 66-76.

Berman S, Gewurz H, and Mold C (1986). Binding of C-reactive protein to nucleated cells leads to complement activation without cytolysis. *J Immunol,* **136,** 1354-1359.

Bhakdi S, Torzewski M, Klouche M, and Hemmes M (1999). Complement and atherogenesis: binding of CRP to degraded, nonoxidized LDL enhances complement activation. *Arterioscler Thromb Vasc Biol,* **19,** 2348-2354.

Bharadwaj D, Stein MP, Volzer M, Mold C, and Du Clos TW (1999). The major receptor for C-reactive protein on leukocytes is fcgamma receptor II. *J Exp Med,* **190,** 585-590.

Bickerstaff MC, Botto M, Hutchinson WL, Herbert J, Tennent GA, Bybee A, Mitchell DA, Cook HT, Butler PJ, Walport MJ, and Pepys MB (1999). Serum amyloid P component controls chromatin degradation and prevents antinuclear autoimmunity. *Nat Med,* **5,** 694-697.

Biesecker G, Lachmann P, and Henderson R (1993). Structure of complement poly-C9 determined in projection by cryo-electron microscopy and single particle analysis. *Mol Immunol,* **30,** 1369-1382.

Black S, Kushner I, and Samols D (2004). C-reactive Protein. *J Biol Chem,* **279,** 48487-48490.

Blackwell KA, Andersen BF, and Baker EN (1994). Metal substitution in a blue-copper protein: the crystal structure of cadmium-azurin at 1.8 A resolution. *Acta Crystallogr D,* **50,** 263-270.

Blake C and Serpell L (1996). Synchrotron X-ray studies suggest that the core of the transthyretin amyloid fibril is a continuous beta-sheet helix. *Structure,* **4,** 989-998.

Blake CC and Swan ID (1971). X-ray analysis of structure of human lysozyme at 6 A resolution. *Nat New Biol,* **232,** 12-15.

Blow DM (1985). Introduction to rotation and translation functions. Proceedings of the Daresbury Study Weekend. Molecular replacement.

Blow DM (2002). *Outline of crystallography for biologists.* Oxford University Press.

Bodman-Smith KB, Melendez AJ, Campbell I, Harrison PT, Allen JM, and Raynes JG (2002). C-reactive protein-mediated phagocytosis and phospholipase D signalling through the high-affinity receptor for immunoglobulin G (FcgammaRI). *Immunology,* **107,** 252-260.

Booth DR, Sunde M, Bellotti V, Robinson CV, Hutchinson WL, Fraser PE, Hawkins PN, Dobson CM, Radford SE, Blake CC, and Pepys MB (1997). Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature,* **385,** 787-793.

Booth DR (1998). Analysis of autoaggregation and ligand binding sites of SAP component by in vitro mutagenesis. *Amyloid & Amyloidosis,* 23-25.

Boren J, Olin K, Lee I, Chait A, Wight TN, and Innerarity TL (1998). Identification of the principal proteoglycan-binding site in LDL. A single-point mutation in apo-B100 severely affects proteoglycan interaction without affecting LDL receptor binding. *J Clin Invest,* **101,** 2658-2664.

Botto M, Hawkins PN, Bickerstaff MC, Herbert J, Bygrave AE, McBride A, Hutchinson WL, Tennent GA, Walport MJ, and Pepys MB (1997). Amyloid deposition is delayed in mice with targeted deletion of the serum amyloid P component gene. *Nat Med,* **3,** 855-859.

Breathnach SM, Kofler H, Sepp N, Ashworth J, Woodrow D, Pepys MB, and Hintner H (1989). Serum amyloid P component binds to cell nuclei in vitro and to in vivo deposits of extracellular chromatin in systemic lupus erythematosus. *J Exp Med,* **170,** 1433-1438.

Bricogne G (2006). Geometric sources of redundancy in intensity data and their use for phase determination. *Acta Crystallogr A,* **30,** 395-405.

Bristow CL and Boackle RJ (1986). Evidence for the binding of human serum amyloid P component to Clq and Fab gamma. *Mol Immunol,* **23,** 1045-1052.

Brundish DE and Baddiley J (1968). Pneumococcal C-substance, a ribitol teichoic acid containing choline phosphate. *Biochem J,* **110,** 573-582.

Brunger AT (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature,* **355,** 472-475.

Brunger AT, Adams PD, and Rice LM (1997). New applications of simulated annealing in X-ray crystallography and solution NMR. *Structure,* **5,** 325-336.

Brunger AT, Adams PD, and Rice LM (1998). Recent developments for the efficient crystallographic refinement of macromolecular structures. *Curr Opin Struct Biol,* **8,** 606-611.

Butler PJ, Tennent GA, and Pepys MB (1990). Pentraxin-chromatin interactions: serum amyloid P component specifically displaces H1-type histones and solubilizes native long chromatin. *J Exp Med,* **172,** 13-18.

Cardoso I, Goldsbury CS, Muller SA, Olivieri V, Wirtz S, Damas AM, Aebi U, and Saraiva MJ (2002). Transthyretin fibrillogenesis entails the assembly of monomers: a molecular model for in vitro assembled transthyretin amyloid-like fibrils. *J Mol Biol,* **317,** 683-695.

Carroll MC (2004). The complement system in regulation of adaptive immunity. *Nat Immunol,* **5,** 981-986.

Casciola-Rosen LA, Anhalt G, and Rosen A (1994). Autoantigens targeted in systemic lupus erythematosus are clustered in two populations of surface structures on apoptotic keratinocytes. *J Exp Med,* **179,** 1317-1330.

Cermak J, Key NS, Bach RR, Balla J, Jacob HS, and Vercellotti GM (1993). C-reactive protein induces human peripheral blood monocytes to synthesize tissue factor. *Blood,* **82,** 513-520.

Chang MK, Binder CJ, Torzewski M, and Witztum JL (2002). C-reactive protein binds to both oxidized LDL and apoptotic cells through recognition of a common ligand: Phosphorylcholine of oxidized phospholipids. *Proc Natl Acad Sci U S A,* **99,** 13043-13048.

Cline AM and Radic MZ (2004). Apoptosis, subcellular particles, and autoimmunity. *Clin Immunol,* **112,** 175-182.

Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D,* **50,** 760-763.

Cooper NR (1985). The classical complement pathway: activation and regulation of the first complement component. *Adv Immunol,* **37,** 151-216.

Correia BE, Loureiro-Ferreira N, Rodrigues JR, and Brito RM (2006). A structural model of an amyloid protofilament of transthyretin. *Protein Sci,* **15,** 28-32.

Danesh J, Wheeler JG, Hirschfield GM, Eda S, Eiriksdottir G, Rumley A, Lowe GD, Pepys MB, and Gudnason V (2004). C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. *N Engl J Med,* **350,** 1387-1397.

De Beer FC and Pepys MB (1982). Isolation of human C-reactive protein and serum amyloid P component. *J Immunol Methods,* **50,** 17-31.

De Beer FC, Soutar AK, Baltz ML, Trayner IM, Feinstein A, and Pepys MB (1982). Low density lipoprotein and very low density lipoprotein are selectively bound by aggregated C-reactive protein. *J Exp Med,* **156,** 230-242.

DeLano WL and Brunger AT (1995). The direct rotation function: Patterson correlation search applied to molecular replacement. *Acta Crystallogr D,* **51,** 740-748.

DeLano WL (2002). The Pymol Molecular Graphics System. Delano Scientific, San Carlos, CA, USA.

Dobson CM (2002). Getting out of shape. *Nature,* **418,** 729-730.

Dobson CM (2004). Chemical space and biology. *Nature,* **432,** 824-828.

Du Clos TW (1989). C-reactive protein reacts with the U1 small nuclear ribonucleoprotein. *J Immunol,* **143,** 2553-2559.

Ducruix A and Giege R (1992). *Crystallization of nucleic acids and proteins. A practical approach.* Oxford University Press.

Dumoulin M, Canet D, Last AM, Pardon E, Archer DB, Muyldermans S, Wyns L, Matagne A, Robinson CV, Redfield C, and Dobson CM (2005). Reduced global cooperativity is a common feature underlying the amyloidogenicity of pathogenic lysozyme mutations. *J Mol Biol,* **346,** 773-788.

Emsley J, White HE, O'Hara BP, Oliva G, Srinivasan N, Tickle IJ, Blundell TL, Pepys MB, and Wood SP (1994). Structure of pentameric human serum amyloid P component. *Nature,* **367,** 338-345.

Enguita FJ, Rodrigues L, Archer M, Sieker L, Rodrigues A, Pohl E, Turner DL, Santos H, and Carrondo MA (2003). Crystallization and preliminary X-ray characterization of cytochrome c" from the obligate methylotroph Methylophilus methylotrophus. *Acta Crystallogr D,* **59,** 580-583.

Entman ML, Michael L, Rossen RD, Dreyer WJ, Anderson DC, Taylor AA, and Smith CW (1991). Inflammation in the course of early myocardial ischemia. *FASEB J,* **5,** 2529-2537.

Epstein SE, Zhou YF, and Zhu J (1999). Infection and atherosclerosis: emerging mechanistic paradigms. *Circulation,* **100,** e20-e28.

Esser AF (1994) The membrane attack complex of complement. Assembly, structure and cytotoxic activity. *Toxicology,* **87,** 229-247.

Evans PR (1997). Joint CCP4 and ESF-EACBM Newsl. *Protein Crystallogr,* **33,** 22-24.

Fraser PE, Nguyen JT, Surewicz WK, and Kirschner DA (1991). pH-dependent structural transitions of Alzheimer amyloid peptides. *Biophys J,* **60,** 1190-1201.

Fraser PE, Duffy LK, O'Malley MB, Nguyen J, Inouye H, and Kirschner DA (1991). Morphology and antibody recognition of synthetic beta-amyloid peptides. *J Neurosci Res,* **28,** 474-485.

French GS and Wilson KS (1978). On the treatment of negative intensity observations. *Acta Crystallogr A,* **34,** 517-525.

Fu T and Borensztajn J (2002). Macrophage uptake of low-density lipoprotein bound to aggregated C-reactive protein: possible mechanism of foam-cell formation in atherosclerotic lesions. *Biochem J,* **366,** 195-201.

Fujita T, Endo Y, and Nonaka M (2004). Primitive complement system--recognition and activation. *Mol Immunol,* **41,** 103-111.

Gaboriaud C, Juanhuix J, Gruez A, Lacroix M, Darnault C, Pignol D, Verger D, Fontecilla-Camps JC, and Arlaud GJ (2003). The crystal structure of the globular head of complement protein C1q provides a basis for its versatile recognition properties. *J Biol Chem,* **278,** 46974-46982.

Garman E (1999). Cool data: quantity AND quality. *Acta Crystallogr D,* **55,** 1641-1653.

Garman E (2003). 'Cool' crystals: macromolecular cryocrystallography and radiation damage. *Curr Opin Struct Biol,* **13,** 545-551.

Garman EF and Schneider TR (1997). Macromolecular Cryocrystallography. *J Appl Crystallogr,* **30,** 211-217.

Gershov D, Kim S, Brot N, and Elkon KB (2000). C-Reactive protein binds to apoptotic cells, protects the cells from assembly of the terminal complement components, and sustains an antiinflammatory innate immune response: implications for systemic autoimmunity. *J Exp Med,* **192,** 1353-1364.

Gill R, Sibson NR, Hatfield RH, Burdett NG, Carpenter TA, Hall LD, and Pickard JD (1995). A comparison of the early development of ischaemic damage following permanent middle cerebral artery occlusion in rats as assessed using magnetic resonance imaging and histology. *J Cereb Blood Flow Metab,* **15,** 1-11.

Gill R, Kemp JA, Sabin C, and Pepys MB (2004). Human C-reactive protein increases cerebral infarct size after middle cerebral artery occlusion in adult rats. *J Cereb Blood Flow Metab,* **24,** 1214-1218.

Griselli M, Herbert J, Hutchinson WL, Taylor KM, Sohail M, Krausz T, and Pepys MB (1999). C-reactive protein and complement are important mediators of tissue damage in acute myocardial infarction. *J Exp Med,* **190,** 1733-1740.

Grosse-Kunstleve RW and Adams PD (2001). Patterson correlation methods: a review of molecular replacement with CNS. *Acta Crystallogr D,* **57,** 1390-1396.

Haas DJ and Rossman MG (1970). Crystallographic studies on lactate dehydrogenase at –75ºC. *Acta Crystallogr B,* **26,** 998-1004.

Hack CE, Wolbink GJ, Schalkwijk C, Speijer H, Hermens WT, and van den BH (1997). A role for secretory phospholipase A2 and C-reactive protein in the removal of injured cells. *Immunol Today,* **18,** 111-115.

Hamazaki H (1987). Ca2+-mediated association of human serum amyloid P component with heparan sulfate and dermatan sulfate. *J Biol Chem,* **262,** 1456-1460.

Harper JD and Lansbury PT, Jr. (1997). Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. *Annu Rev Biochem,* **66,** 385-407.

Haupt H and Heimburger N (1972). [Human serum proteins with high affinity for carboxymethylcellulose. I. Isolation of lysozyme, C1q and 2 hitherto unknown -globulins]. *Hoppe Seylers Z Physiol Chem,* **353,** 1125-1132.

Hauptman H (1982). On integrating the techniques of direct methods and isomorphous replacement. I. The theoretical basis. *Acta Crystallogr A,* **38,** 289-294.

Hawkins PN, Myers MJ, Epenetos AA, Caspi D, and Pepys MB (1988). Specific localization and imaging of amyloid deposits in vivo using 123I-labeled serum amyloid P component. *J Exp Med,* **167,** 903-913.

Hawkins PN, Wootton R, and Pepys MB (1990). Metabolic studies of radioiodinated serum amyloid P component in normal subjects and patients with systemic amyloidosis. *J Clin Invest,* **86,** 1862-1869.

Hawkins PN, Lavender JP, and Pepys MB (1990). Evaluation of systemic amyloidosis by scintigraphy with 123I-labeled serum amyloid P component. *N Engl J Med,* **323,** 508-513.

Hawkins PN, Tennent GA, Woo P, and Pepys MB (1991). Studies in vivo and in vitro of serum amyloid P component in normals and in a patient with AA amyloidosis. *Clin Exp Immunol,* **84,** 308-316.

Hendrickson WA (2000). Synchrotron crystallography. *Trends Biochem Sci,* **25,** 637-643.

Hicks PS, Saunero-Nava L, Du Clos TW, and Mold C (1992). Serum amyloid P component binds to histones and activates the classical complement pathway. *J Immunol,* **149,** 3689-3694.

Hind CR, Collins PM, Renn D, Cook RB, Caspi D, Baltz ML, and Pepys MB (1984) Binding specificity of serum amyloid P component for the pyruvate acetal of galactose. *J Exp Med,* **159,** 1058-1069.

Hirschfield GM and Pepys MB (2003). C-reactive protein and cardiovascular disease: new insights from an old molecule. *QJM,* **96,** 793-807.

Hirschfield GM (2004). Amyloidosis: a clinico-pathophysiological synopsis. *Semin Cell Dev Biol,* **15,** 39-44.

Ho JG, Kitov PI, Paszkiewicz E, Sadowska J, Bundle DR, and Ng KK (2005). Ligand-assisted aggregation of proteins. Dimerization of serum amyloid P component by bivalent ligands. *J Biol Chem,* **280,** 31999-32008.

Hohenester E, Hutchinson WL, Pepys MB, and Wood SP (1997). Crystal structure of a decameric complex of human serum amyloid P component with bound dAMP. *J Mol Biol,* **269,** 570-578.

Hopkins M, Flanagan PA, Bailey S, Glover ID, Myles DA, and Greenhough TJ (1994). Crystallization and preliminary X-ray analysis of C-reactive protein from rat. *J Mol Biol,* **235,** 767-771.

Huang KF, Chiou SH, Ko TP, Yuann JM, and Wang AH (2002). The 1.35 A structure of cadmium-substituted TM-3, a snake-venom metalloproteinase from Taiwan habu: elucidation of a TNFalpha-converting enzyme-like active-site structure with a distorted octahedral geometry of cadmium. *Acta Crystallogr D,* **58,** 1118-1128.

Jiang H, Lint TF, and Gewurz H (1991). Defined chemically cross-linked oligomers of human C-reactive protein: characterization and reactivity with the complement system. *Immunology,* **74,** 725-731.

Jiang H, Robey FA, and Gewurz H (1992). Localization of sites through which C-reactive protein binds and activates complement to residues 14-26 and 76-92 of the human C1q A chain. *J Exp Med,* **175,** 1373-1379.

Kabsch W(1988). XDS program. *J Appl Crystallogr,* **21,** 916-924.

Kaplan MH and Volanakis JE (1974). Interaction of C-reactive protein complexes with the complement system. I. Consumption of human complement associated with the reaction of C-reactive protein with pneumococcal C-polysaccharide and with the choline phosphatides, lecithin and sphingomyelin. *J Immunol,* **112,** 2135-2147.

Kinoshita CM, Ying SC, Hugli TE, Siegel JN, Potempa LA, Jiang H, Houghten RA, and Gewurz H (1989). Elucidation of a protease-sensitive site involved in the binding of calcium to C-reactive protein. *Biochemistry,* **28,** 9840-9848.

Kinoshita CM, Gewurz AT, Siegel JN, Ying SC, Hugli TE, Coe JE, Gupta RK, Huckman R, and Gewurz H (1992). A protease-sensitive site in the proposed Ca(2+)-binding region of human serum amyloid P component and other pentraxins. *Protein Sci,* **1,** 700-709.

Kishore U and Reid KB (2000). C1q: structure, function, and receptors. *Immunopharmacology,* **49,** 159-170.

Kishore U, Gupta SK, Perdikoulis MV, Kojouharova MS, Urban BC, and Reid KB (2003). Modular organization of the carboxyl-terminal, globular head region of human C1q A, B, and C chains. *J Immunol,* **171,** 812-820.

Kishore U, Ghai R, Greenhough TJ, Shrive AK, Bonifati DM, Gadjeva MG, Waters P, Kojouharova MS, Chakraborty T, and Agrawal A (2004). Structural and functional anatomy of the globular domain of complement protein C1q. *Immunol Lett,* **95,** 113-128.

Kitov PI, Sadowska JM, Mulvey G, Armstrong GD, Ling H, Pannu NS, Read RJ, and Bundle DR (2000). Shiga-like toxins are neutralized by tailored multivalent carbohydrate ligands. *Nature,* **403,** 669-672.

Kleywegt GJ and Jones TA (1994). From first map to final model. SERC Daresbury Laboratory, Daresbury, UK.

Knobel HR, Villiger W, and Isliker H (1975). Chemical analysis and electron microscopy studies of human C1q prepared by different methods. *Eur J Immunol,* **5,** 78-82.

Konnert JH (2006). A restrained-parameter structure-factor least-squares refinement procedure for large asymmetric units. *Acta Crystallogr A,* **32,** 614-617.

Kubak BM, Potempa LA, Anderson B, Mahklouf S, Venegas M, Gewurz H, and Gewurz AT (1988). Evidence that serum amyloid P component binds to mannose-terminated sequences of polysaccharides and glycoproteins. *Mol Immunol,* **25,** 851-858.

Kushner I, Broder ML, and Karp D (1978). Control of the acute phase response. Serum C-reactive protein kinetics after acute myocardial infarction. *J Clin Invest,* **61,** 235-242.

Kushner I (1982). The phenomenon of the acute phase response. *Ann N Y Acad Sci,* **389,** 39-48.

Lagrand WK, Niessen HW, Wolbink GJ, Jaspars LH, Visser CA, Verheugt FW, Meijer CJ, and Hack CE (1997). C-reactive protein colocalizes with complement in human hearts during acute myocardial infarction. *Circulation,* **95,** 97-103.

Lansbury PT, Jr. (1997). Inhibition of amyloid formation: a strategy to delay the onset of Alzheimer's disease. *Curr Opin Chem Biol,* **1,** 260-267.

Lashuel HA, Lai Z, and Kelly JW (1998). Characterization of the transthyretin acid denaturation pathways by analytical ultracentrifugation: implications for wild-type, V30M, and L55P amyloid fibril formation. *Biochemistry,* **37,** 17851-17864.

Lee GW, Lee TH, and Vilcek J (1993). TSG-14, a tumor necrosis factor- and IL-1-inducible protein, is a novel member of the pentaxin family of acute phase proteins. *J Immunol,* **150,** 1804-1812.

Leslie AGW (1992). Jnt CCP4/ESF-EACMB Newsletter. *Protein Crystallogr,* **26.**

Leslie AGW (1997). Mosflm User's Guide. MRC Laboratory of Molecular Biology, Cambridge.

Leslie AGW (1999). Integration of macromolecular diffraction data. *Acta Crystallogr D,* **55,** 1696-1702.

Libby P, Ridker PM, and Maseri A (2002). Inflammation and atherosclerosis. *Circulation,* **105,** 1135-1143.

Lindley PF (1999). Macromolecular crystallography with a third-generation synchrotron source. *Acta Crystallogr D,* **55,** 1654-1662.

Ling H, Boodhoo A, Hazes B, Cummings MD, Armstrong GD, Brunton JL, and Read RJ (1998). Structure of the shiga-like toxin I B-pentamer complexed with an analogue of its receptor Gb3. *Biochemistry,* **37,** 1777-1788.

Lusis AJ (2000). Atherosclerosis. *Nature,* **407,** 233-241.

Marnell L, Mold C, and Du Clos TW (2005). C-reactive protein: ligands, receptors and role in inflammation. *Clin Immunol,* **117,** 104-111.

Marnell LL, Mold C, Volzer MA, Burlingame RW, and Du Clos TW (1995). C-reactive protein binds to Fc gamma RI in transfected COS cells. *J Immunol,* **155,** 2185-2193.

Martin SJ, Reutelingsperger CP, McGahon AJ, Rader JA, van Schie RC, LaFace DM, and Green DR (1995). Early redistribution of plasma membrane phosphatidylserine is a general feature of apoptosis regardless of the initiating stimulus: inhibition by overexpression of Bcl-2 and Abl. *J Exp Med,* **182,** 1545-1556.

Matthews BW (1968). Solvent content of protein crystals. *J Mol Biol,* **33,** 491-497.

McCarty M (1947). The occurrence during acute infection of a protein not normally present in blood. *J Exp Med,* **85,** 491-498.

McPherson A (1990). Current approaches to macromolecular crystallization. *Eur J Biochem,* **189,** 1-23.

Milanino R, Conforti A, Fracasso ME, Franco L, Leone R, Passarella E, Tarter G, and Velo GP (1979). Concerning the role of endogenous copper in the acute inflammatory process. *Agents Actions,* **9,** 581-588.

Mold C, Baca R, and Du Clos TW (2002). Serum amyloid P component and C-reactive protein opsonize apoptotic cells for phagocytosis through Fcgamma receptors. *J Autoimmun,* **19,** 147-154.

Morgan BP, Marchbank KJ, Longhi MP, Harris CL, and Gallimore AM (2005). Complement: central to innate immunity and bridging to adaptive responses. *Immunol Lett,* **97,** 171-179.

Mortensen RF, Osmand AP, Lint TF, and Gewurz H (1976). Interaction of C-reactive protein with lymphocytes and monocytes: complement-dependent adherence and phagocytosis. *J Immunol,* **117,** 774-781.

Muller-Eberhard HJ (1986). The membrane attack complex of complement. *Annu Rev Immunol,* **4,** 503-528.

Murshudov GN, Vagin AA, and Dobson EJ (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr D,* **53,** 240-255.

Myles DA, Rule SA, DeLucas LJ, Babu YS, Xu Y, Volanakis JE, Bugg CE, Bailey S, and Greenhough TJ (1990). Rotation function studies of human C-reactive protein. *J Mol Biol,* **216,** 491-496.

Narkates AJ and Volanakis JE (1982). C-reactive protein binding specificities: artificial and natural phospholipid bilayers. *Ann N Y Acad Sci,* **389,** 172-182.

Nauta AJ, Bottazzi B, Mantovani A, Salvatori G, Kishore U, Schwaeble WJ, Gingras AR, Tzima S, Vivanco F, Egido J, Tijsma O, Hack EC, Daha MR, and Roos A (2003). Biochemical and functional characterization of the interaction between pentraxin 3 and C1q. *Eur J Immunol,* **33,** 465-473.

Navaza J (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallogr A,* **50,** 157-163.

O'Hara BP (1988). University of London.

Otwinowski Z and Minor W (1997). Processing of X-ray diffraction data collected in oscillation mode, in *Methods Enzymol*, **276,** Macromolecular Crystallogr Part A, eds. Carter CD and Sweet RM, Academic Press: New York pp307-326.

Painter RH, De E, I, Massey A, Pinteric L, and Stern SB (1982). The structure and binding characteristics of serum amyloid protein (9.5S alpha 1-glycoprotein). *Ann N Y Acad Sci,* **389,** 199-215.

Patterson AL (1934). A Fourier series method for the determination of the components of interatomic distances in crystals. *Phys Rev*, **46,** 372-376.

Pepys MB, Dash AC, and Ashley MJ (1977). Isolation of C-reactive protein by affinity chromatography. *Clin Exp Immunol,* **30,** 32-37.

Pepys MB, Dyck RF, De Beer FC, Skinner M, and Cohen AS (1979). Binding of serum amyloid P-component (SAP) by amyloid fibrils. *Clin Exp Immunol,* **38,** 284-293.

Pepys MB and Baltz ML (1983). Acute phase proteins with special reference to C-reactive protein and related proteins (pentaxins) and serum amyloid A protein. *Adv Immunol,* **34,** 141-212.

Pepys MB and Butler PJ (1987). Serum amyloid P component is the major calcium-dependent specific DNA binding protein of the serum. *Biochem Biophys Res Commun,* **148,** 308-313.

Pepys MB, Booth SE, Tennent GA, Butler PJ, and Williams DG (1994). Binding of pentraxins to different nuclear structures: C-reactive protein binds to small nuclear ribonucleoprotein particles, serum amyloid P component binds to chromatin and nucleoli. *Clin Exp Immunol,* **97,** 152-157.

Pepys MB, Booth DR, Hutchinson WL, Gallimore JR, Collins PM, and Hohenester E (1997). Amyloid P component. A critical review. *Amyloid,* **4,** 274-295.

Pepys MB and Berger A (2001). The renaissance of C reactive protein. *BMJ,* **322,** 4-5.

Pepys MB, Herbert J, Hutchinson WL, Tennent GA, Lachmann HJ, Gallimore JR, Lovat LB, Bartfai T, Alanine A, Hertel C, Hoffmann T, Jakob-Roetne R, Norcross RD, Kemp JA, Yamamura K, Suzuki M, Taylor GW, Murray S, Thompson D, Purvis A, Kolstoe S, Wood SP, and Hawkins PN (2002). Targeted pharmacological depletion of serum amyloid P component for treatment of human amyloidosis. *Nature,* **417,** 254-259.

Pepys MB and Hirschfield GM (2003). C-reactive protein: a critical update. *J Clin Invest,* **111,** 1805-1812.

Pepys MB, Hirschfield GM, Tennent GA, Gallimore JR, Kahan MC, Bellotti V, Hawkins PN, Myers RM, Smith MD, Polara A, Cobb AJ, Ley SV, Aquilina JA, Robinson CV, Sharif I, Gray GA, Sabin CA, Jenvey MC, Kolstoe SE, Thompson D, and Wood SP (2006). Targeting C-reactive protein for the treatment of cardiovascular disease. *Nature,* **440,** 1217-1221.

Perkins SJ and Pepys MB (1986). X-ray and neutron solution scattering studies on CRP and SAP. *Protide Biol Fluids,* **34,** 323-326.

Petsko GA (1975). Protein crystallography at sub-zero temperatures: cryo-protective mother liquors for protein crystals. *J Mol Biol*, **96,** 381-392.

Pickens JC, Mitchell DD, Liu J, Tan X, Zhang Z, Verlinde CL, Hol WG, and Fan E (2004). Nonspanning bivalent ligands as improved surface receptor binding inhibitors of the cholera toxin B pentamer. *Chem Biol,* **11,** 1205-1215.

Pontet M, Engler R, and Jayle MF (1978). One step preparation of both human C-reactive protein and Clt. *FEBS Lett,* **88,** 172-175.

Potempa LA, Kubak BM, and Gewurz H (1985). Effect of divalent metal ions and pH upon the binding reactivity of human serum amyloid P component, a C-reactive protein homologue, for zymosan. Preferential reactivity in the presence of copper and acidic pH. *J Biol Chem,* **260,** 12142-12147.

Ramadan MA, Shrive AK, Holden D, Myles DA, Volanakis JE, DeLucas LJ, and Greenhough TJ (2002). The three-dimensional structure of calcium-depleted human C-reactive protein from perfectly twinned crystals. *Acta Crystallogr D,* **58,** 992-1001.

Reid KB and Porter RR (1976). Subunit composition and structure of subcomponent C1q of the first component of human complement. *Biochem J,* **155,** 19-23.

Reid KB and Porter RR (1981). The proteolytic activation systems of complement. *Annu Rev Biochem,* **50,** 433-464.

Reid KB (1983). Proteins involved in the activation and control of the two pathways of human complement. *Biochem Soc Trans,* **11,** 1-12.

Reid KB and Thompson RA (1983). Characterization of a non-functional form of C1q found in patients with a genetically linked deficiency of C1q activity. *Mol Immunol,* **20,** 1117-1125.

Rhodes G (1993). *Crystallography made crystal clear: a guide for users of macromolecular models.* Academic Press Inc.

Ridker PM (2001). High-sensitivity C-reactive protein: potential adjunct for global risk assessment in the primary prevention of cardiovascular disease. *Circulation,* **103,** 1813-1818.

Rochet JC, Conway KA, and Lansbury PT, Jr. (2000). Inhibition of fibrillization and accumulation of prefibrillar oligomers in mixtures of human and mouse alpha-synuclein. *Biochemistry,* **39,** 10619-10626.

Ross R (1993). The pathogenesis of atherosclerosis: a perspective for the 1990s. *Nature,* **362,** 801-809.

Roumenina LT, Kantardjiev AA, Atanasov BP, Waters P, Gadjeva M, Reid KB, Mantovani A, Kishore U, and Kojouharova MS (2005). Role of Ca2+ in the electrostatic stability and the functional activity of the globular domain of human C1q. *Biochemistry,* **44,** 14097-14109.

Roux KH, Kilpatrick JM, Volanakis JE, and Kearney JF (1983). Localization of the phosphocholine-binding sites on C-reactive protein by immunoelectron microscopy. *J Immunol,* **131,** 2411-2415.

Schwalbe RA, Dahlback B, Coe JE, and Nelsestuen GL (1992). Pentraxin family of proteins interact specifically with phosphorylcholine and/or phosphorylethanolamine. *Biochemistry,* **31,** 4907-4915.

Sebastiao MP, Saraiva MJ, and Damas AM (1998). The crystal structure of amyloidogenic Leu55 --> Pro transthyretin variant reveals a possible pathway for transthyretin polymerization into amyloid fibrils. *J Biol Chem,* **273,** 24715-24722.

Sellar GC, Blake DJ, and Reid KB (1991). Characterization and organization of the genes encoding the A-, B- and C-chains of human complement subcomponent C1q. The complete derived amino acid sequence of human C1q. *Biochem J,* **274 (Pt 2),** 481-490.

Serpell LC, Sunde M, Benson MD, Tennent GA, Pepys MB, and Fraser PE (2000). The protofilament substructure of amyloid fibrils. *J Mol Biol,* **300,** 1033-1039.

Serpell LC, Blake CC, and Fraser PE (2000). Molecular structure of a fibrillar Alzheimer's A beta fragment. *Biochemistry,* **39,** 13269-13275.

Sheldrick GM and Schneider TR (2006). SHELXL: High resolution refinement. *Methods Enzymol,* **277,** 319-343.

Shrive AK, Holden D, Myles DA, and Greenhough TJ (1996). Structure solution of C-reactive proteins: molecular replacement with a twist. *Acta Crystallogr D,* **52,** 1049-1057.

Shrive AK, Metcalfe AM, Cartwright JR, and Greenhough TJ (1999). C-reactive protein and SAP-like pentraxin are both present in Limulus polyphemus haemolymph: crystal structure of Limulus SAP. *J Mol Biol,* **290,** 997-1008.

Sim RB and Tsiftsoglou SA (2004). Proteases of the complement system. *Biochem Soc Trans,* **32,** 21-27.

Sipe JD (1992). Amyloidosis. *Annu Rev Biochem,* **61,** 947-975.

Sipe JD and Cohen AS (2000). Review: history of the amyloid fibril. *J Struct Biol,* **130,** 88-98.

Skinner M, Sipe JD, Yood RA, Shirahama T, and Cohen AS (1982). Characterization of P-component (AP) isolated from amyloidotic tissue: half-life studies of human and murine AP. *Ann N Y Acad Sci,* **389,** 190-198.

Sorensen IJ, Andersen O, Nielsen EH, and Svehag SE (1995). Native human serum amyloid P component is a single pentamer. *Scand J Immunol,* **41,** 263-267.

Sorensen IJ, Holm NE, Schroder L, Voss A, Horvath L, and Svehag SE (2000). Complexes of serum amyloid P component and DNA in serum from healthy individuals and systemic lupus erythematosus patients. *J Clin Immunol,* **20,** 408-415.

Srinivasan N, White HE, Emsley J, Wood SP, Pepys MB, and Blundell TL (1994). Comparative analyses of pentraxins: implications for protomer assembly and ligand binding. *Structure,* **2,** 1017-1027.

Stein MP, Edberg JC, Kimberly RP, Mangan EK, Bharadwaj D, Mold C, and Du Clos TW (2000). C-reactive protein binding to FcgammaRIIa on human monocytes and neutrophils is allele-specific. *J Clin Invest,* **105,** 369-376.

Stromer T. and Serpell LC (2005). Structure and morphology of the Alzheimer's amyloid fibril. *Micros Res Tech,* **67,** 210-217.

Stryer L (1999). *Biochemistry.* WH Freeman & Co. New York.

Sunde M, Serpell LC, Bartlam M, Fraser PE, Pepys MB, and Blake CC (1997). Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J Mol Biol,* **273,** 729-739.

Tacnet-Delorme P, Chevallier S, and Arlaud GJ (2001). Beta-amyloid fibrils activate the C1 complex of complement under physiological conditions: evidence for a binding site for A beta on the C1q globular regions. *J Immunol,* **167,** 6374-6381.

Tamminen M, Mottino G, Qiao JH, Breslow JL, and Frank JS (1999). Ultrastructure of early lipid accumulation in ApoE-deficient mice. *Arterioscler Thromb Vasc Biol,* **19,** 847-853.

Tennent GA, Lovat LB, and Pepys MB (1995). Serum amyloid P component prevents proteolysis of the amyloid fibrils of Alzheimer disease and systemic amyloidosis. *Proc Natl Acad Sci U S A,* **92,** 4299-4303.

Thompson D, Pepys MB, and Wood SP (1999). The physiological structure of human C-reactive protein and its complex with phosphocholine. *Structure Fold Des,* **7,** 169-177.

Thompson D, Pepys MB, Tickle I, and Wood S (2002). The structures of crystalline complexes of human serum amyloid P component with its carbohydrate ligand, the cyclic pyruvate acetal of galactose. *J Mol Biol,* **320,** 1081-1086.

Thompson D (1997). University of London.

Tillet WS and Francis T (1930). Serological reactions in pneumonia with non-protein somatic fraction of pneumococcus. *J Exp Med,* **52,** 561-571.

Tillet WS, Geobel WF, and Avery OT (1930). Chemical and immunological properties of a species-specific carbohydrate of pneumococci. *J Exp Med,* **52,** 895-900.

Trakhanov S, Kreimer DI, Parkin S, Ames GF, and Rupp B (1998). Cadmium-induced crystallization of proteins: II. Crystallization of the Salmonella typhimurium histidine-binding protein in complex with L-histidine, L-arginine, or L-lysine. *Protein Sci,* **7,** 600-604.

Vagin A and Teplyakov A (1997). MOLREP:an Automated Program for Molecular Replacement. *J Appl Crystallogr,* **30,** 1022-1025.

Vagin A and Teplyakov A (2000). An approach to multi-copy search in molecular replacement. *Acta Crystallogr D,* **56,** 1622-1624.

Vlaicu R, Niculescu F, Rus HG, and Cristea A (1985). Immunohistochemical localization of the terminal C5b-9 complement complex in human aortic fibrous plaque. *Atherosclerosis,* **57,** 163-177.

Volanakis JE and Kaplan MH (1971). Specificity of C-reactive protein for choline phosphate residues of pneumococcal C-polysaccharide. *Proc Soc Exp Biol Med,* **136,** 612-614.

Volanakis JE and Kaplan MH (1974). Interaction of C-reactive protein complexes with the complement system. II. Consumption of guinea pig complement by CRP complexes: requirement for human C1q. *J Immunol,* **113,** 9-17.

Volanakis JE and Wirtz KW (1979). Interaction of C-reactive protein with artificial phosphatidylcholine bilayers. *Nature,* **281,** 155-157.

Volanakis JE (1982). Complement activation by C-reactive protein complexes. *Ann N Y Acad Sci,* **389,** 235-250.

Volanakis JE (2001). Human C-reactive protein: expression, structure, and function. *Mol Immunol,* **38,** 189-197.

Voll RE, Herrmann M, Roth EA, Stach C, Kalden JR, and Girkontaite I (1997). Immunosuppressive effects of apoptotic cells. *Nature,* **390,** 350-351.

Walport MJ (2001). Complement. Second of two parts. *N Engl J Med,* **344,** 1140-1144.

Wang BC (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol,* **115,** 90-112.

Westbrook EM and Naday I (1997). Charge-coupled device-based area detectors. *Methods Enzymol,* **276,** 244-268.

Whittaker EJW (2006). *Crystallography: An introduction for Earth Science students.* Pergamon Press, Oxford.

Woo P, Korenberg JR, and Whitehead AS (1985). Characterization of genomic and complementary DNA sequence of human C-reactive protein, and comparison with the complementary DNA sequence of serum amyloid P component. *J Biol Chem,* **260,** 13384-13388.

Wood SP, Oliva G, O'Hara BP, White HE, Blundell TL, Perkins SJ, Sardharwalla I, and Pepys MB (1988). A pentameric form of human serum amyloid P component. Crystallization, X-ray diffraction and neutron scattering studies. *J Mol Biol,* **202,** 169-173.

Ying SC, Marchalonis JJ, Gewurz AT, Siegel JN, Jiang H, Gewurz BE, and Gewurz H (1992). Reactivity of anti-human C-reactive protein (CRP) and serum amyloid P component (SAP) monoclonal antibodies with limulin and pentraxins of other species. *Immunology,* **76,** 324-330.

Ying SC, Gewurz AT, Jiang H, and Gewurz H (1993). Human serum amyloid P component oligomers bind and activate the classical complement pathway via residues 14-26 and 76-92 of the A chain collagen-like region of C1q. *J Immunol,* **150,** 169-176.

Zanotti G, Panzalorto M, Marcato A, Malpeli G, Folli C, and Berni R (1998). Structure of pig plasma retinol-binding protein at 1.65 A resolution. *Acta Crystallogr D,* **54,** 1049-1052.

Zhang KYJ and Main P (1990). The use of Sayre's equation with solvent flattening and histogram matching for phase extension and refinement of protein structures. *Acta Crystallogr,* **46,** 377-381.

Zhang YX, Cliff WJ, Schoefl GI, and Higgins G (1999). Coronary C-reactive protein distribution: its relation to development of atherosclerosis. *Atherosclerosis,* **145,** 375-379.

Zhang Z, Liu J, Verlinde CL, Hol WG, and Fan E (2004). Large cyclic peptides as cores of multivalent ligands: application to inhibitors of receptor binding by cholera toxin. *J Org Chem,* **69,** 7737-7740.

Zwaka TP, Hombach V, and Torzewski J (2001). C-reactive protein-mediated low density lipoprotein uptake by macrophages: implications for atherosclerosis. *Circulation,* **103,** 1194-1197.