UNIVERSITY OF SOUTHAMPTON

AMYLOID RECOGNITION BY SERUM AMYLOID P COMPONENT

by

Alan Purvis

A thesis submitted for the degree of
DOCTOR OF PHILOSOPHY

March 2002

Department of Biochemistry

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SCIENCE
DIVISION OF BIOCHEMISTRY

Doctor of Philosophy

AMYLOID RECOGNITION BY SERUM AMYLOID P COMPONENT

by Alan Purvis

The X-ray crystal structures of serum amyloid P component (SAP) with bound (R)-1-[6-[(R)-2-carboxy-pyrrolidin-1-yl]-6-oxo-hexanoyl]-pyrrolidine-2-carboxylic acid (Ro 63-8695) and related components have been solved to elucidate the molecular basis of the action of Ro 63-8695, a potential amyloid mobilizing drug for treatment of human amyloid disease. The structure of SAP in the presence of N-acetyl-D-proline has been determined to a resolution of 2.4Å using a previously solved SAP structure as the phasing model (unit cell dimensions a=96.1Å, b=70.8Å, c=103.6Å, and β=96.8°). The carboxyl group of N-acetyl-D-proline is bound in the double calcium-binding site of each subunit, orientating the pyrrolidine ring into the adjacent hydrophobic pocket formed between Leu62, Tyr64, and Tyr74. The structure of SAP co-crystallised with Ro 63-8695 has been determined to a resolution of 3.2Å by molecular replacement (unit cell dimensions a=b=230.9Å and c=281.4Å). This shows the formation of a ligand-induced decamer, where two SAP pentamers are reversibly cross-linked by five Ro 63-8695 molecules. Binding of the Ro 63-8695 molecule head group shows close superposition with the higher resolution N-acetyl-D-proline structure. The alkyl linker adopts a kinked rotamer about carbons 2 and 3 to facilitate binding of the head groups to the two-fold axis related subunits. The best fit of the electron density is found when both peptide bonds preceding the pyrrolidine ring adopt a cis conformation. Nuclear magnetic resonance spectroscopy has estimated this cis-cis isomer to contribute only ~6% of the Ro 63-8695 population in free solution. SAP has also been found to enhance the refolding yield of denatured lactate dehydrogenase and protects against enzyme inactivation during agitation through a calcium independent site.

PUBLICATIONS

Molecular chaperone properties of serum amyloid P component. Coker, A. R., Purvis, A., Baker, D., Pepys, M. B., Wood, S. P. (2000) *FEBS Letters* **473**, 199-202.

Targeted pharmacological depletion of serum amyloid P component (SAP) for treatment of human amyloidosis. Pepys, M. B., Herbert, J., Hutchinson, W. L., Tennent, G. A., Lachmann, H. J., Gallimore, J. R., Bartfai, T., Alanine, A., Hertel, C., Hoffmann, T., Jakob-Roetne, R., Norcross, R. D., Kemp, J. A., Yamamura, K., Suzuki, M., Taylor, G. W., Murray, S., Thompson, D., Purvis, A., Kolstoe, S., Wood, S. P. & Hawkins, P. N. (2002) *Nature* **417**, 254-259.

CONTENTS

CHAPTER 1: INTRODUCTION

CHAPTER 2: CRYSTALLOGRAPHIC METHODS

CHAPTER 3: STRUCTURE DETERMINATION OF N-ACETYL-D-PROLINE & Ro 63-8695 BOUND TO SAP

CHAPTER 5: MOLECULAR CHAPERONE PROPERTIES OF SERUM AMYLOID P COMPONENT

APPENDIX

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

CHAPTER 2: CRYSTALLOGRAPHIC METHODS

CHAPTER 3: STRUCTURE DETERMINATION OF N-ACETYL-D-PROLINE & Ro 63-8695 BOUND TO SAP

CHAPTER 4: STRUCTURE DETERMINATION OF THE SAP-Ro 63-8695 COMPLEX

CHAPTER 5: MOLECULAR CHAPERONE PROPERTIES OF SERUM AMYLOID P COMPONENT

LIST OF TABLES

CHAPTER 1: INTRODUCTION

CHAPTER 2: CRYSTALLOGRAPHIC METHODS

CHAPTER 3: STRUCTURE DETERMINATION OF N-ACETYL-D-PROLINE & Ro 63-8695 BOUND TO SAP

CHAPTER 4: STRUCTURE DETERMINATION OF THE SAP-Ro 63-8695 COMPLEX

CHAPTER 5: MOLECULAR CHAPERONE PROPERTIES OF SERUM AMYLOID P COMPONENT

ABBREVIATIONS

^{123}I-SAP	iodine-123 labelled SAP
Å	Ångström (10^{-10} metre)
°C	degree Celsius
AA	amyloid fibrils derived from serum amyloid A
a, b, c	unit cell dimensions
Aβ	Alzheimer's disease β-protein peptide
Agarobiose	1,3- linked β-D-galactopyranose with 1,4-linked 3,6-anhydro -α-L-galactopyranose
AL	amyloid fibrils derived from monoclonal immunoglobulin light chains
α, β, γ	depending upon the context, unit cell angles or Eulerian angles
AP	amyloid P component
BaFBr	barium halide phosphor
B-factor	temperature factor
C4BP	C4b-binding protein
Captopril	(S)-1-[(S)-3-mercapto-2-methylpropionyl]-pyrrolidine-2-carboxylic acid
CC	correlation coefficient
CCD	charge-coupled device
CCP4	Collaborative Computational Project Number 4
CJD	Creutzfeldt-Jakob disease
CLR	collagen-like region
CNS	Crystallography & NMR System
ConA	concanavalin A
CRP	C-reactive protein
D_2O	deuterated water
Da	dalton
dAMP	2'-deoxyadenosine-5'-monophosphate
δ	chemical shifts
DNA	deoxyribonucleic acid

EDTA	ethylenediaminetetraacetic acid
EMBL	European Molecular Biology Laboratory
ESRF	European Synchrotron Radiation Facility
Eu	europium
FFT	fast Fourier transform
GABA	γ–aminobutyric acid
GAL	galactose
GBM	glomerular basement membrane
HA	haemagglutinin
h, k, l	reciprocal space coordinates
(h, k, l)	Miller indices for lattice planes within real space
IC_{50}	half-maximal inhibitory concentration
IUPAC	International Union of Pure and Applied Chemistry
K	degree Kelvin
K_d	dissociation constant
λ	wavelength
LDH	Lactate dehydrogenase
MAD	multiple wavelength anomalous dispersion
MAN	mannose
MBP-A	mannose-binding protein A
MIR	multiple isomorphous replacement
MOβDG	methyl 4,6-O-(1-carboxyethylidene)-β-D-galactopyranoside
MOPS	3-(N-morpholino)propanesulponic acid
MPD	2-methyl-2,4-pentanediol
MR	molecular replacement
N-acetyl-D-proline	(R)-1-acetyl-pyrrolidine-2-carboxylic acid
NADH	reduced form of nicotinamide adenine dinucleotide
NAG	N-acetylglucosamine
NARP	neuronal activity-related pentraxins
NCS	non-crystallographic symmetry
NMR	nuclear magnetic resonance
PC	phosphocholine
PE	phosphoethanolamine

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
Ro 15-3479	(R)-1-[(S)-3-mercapto-2-methylpropionyl]-pyrrolidine-2-carboxylic acid
Ro 63-3300	(R)-1-[(S)-3-[(S)-3-[(R)-2-carboxy-pyrrolidin-1-yl]-2-methyl-3-oxo-propyldisulphanyl)-2-methyl-propionyl]-pyrrolidine-2-carboxylic acid
Ro 63-8695	(R)-1-[6-[(R)-2-carboxy-pyrrolidin-1-yl]-6-oxo-hexanoyl]-pyrrolidine-2-carboxylic acid
ROESY	rotating frame Overhauser effect spectroscopy
rpm	revolutions per minute
rps	revolutions per second
SAP	serum amyloid P component
SIA	sialic acid
SLE	systemic lupus erythematosus
snRNPs	small nuclear ribonucleoprotein particles
TAP	tissue amyloid P component
TOCSY	total correlation spectroscopy
Tris	Tris(hydroxymethyl)aminomethane
x, y, z	unit cell coordinates (real space)

ACKNOWLEDGEMENTS

I would like to thank my supervisor Professor Steve Wood for his help and guidance throughout the duration of my research. I also wish to extend my gratitude to my industrial supervisor Professor Mark Pepys who has both funded this research and supplied copious amounts of protein throughout it. I am especially grateful to both Doctor Darren Thompson and Doctor Fiyaz Mohammed for their help on the practical and theoretical aspects of X-ray crystallography, their continuing encouragement, and above all their friendship. I would also like to thank Doctor Jon Cooper and Doctor Peter Erskine for their guidance during various synchrotron trips, Doctor Alun Coker and Doctor Douglas Baker for their help on the refolding work, Doctor Matt Crump for his assistance with the NMR spectroscopy, and all of the members of the crystallography group, past and present, for their help and support at various times during the past three years.

I would also like to thank my family for their continuing support, but more importantly I would like to thank my wife for her love, support, and understanding.

This thesis is dedicated to the memory of Bettina Muriel Mason

18th January 1922 to 20th July 1997

CHAPTER 1

INTRODUCTION

1.1 Introduction

Human serum amyloid P component (SAP) is a normal plasma glycoprotein composed of five identical subunits. The plasma concentration of SAP is tightly regulated by the liver hepatocytes to 32mg/l in men, and slightly lower at 24mg/l in women (Nelson *et al* 1991b; Pepys *et al* 1997). SAP was initially isolated and characterised by several different groups and has been known as 9.5S α_1-glycoprotein (Haupt *et al* 1972) and C1t (Assimeh & Painter 1975a; Assimeh & Painter 1975b). SAP is a member of the pentraxin family of evolutionarily ancient and highly conserved proteins, which also include the acute phase reactant C-reactive protein (CRP). The tight regulation, structural conservation, and apparent lack of polymorphism or deficiency in humans suggest that SAP has important biological functions. Ligand recognition and binding by SAP has been reported to contribute to the pathogenesis of amyloidosis, the function of numerous basement membranes, host defence against infection, and in the handling of nuclear autoantigens.

This work is concerned with amyloidosis, a disorder of protein folding, where normally soluble proteins are deposited extracellularly in tissues as insoluble fibrils leading to tissue damage and disease. Specifically, it concerns the involvement of SAP in amyloidosis and a new potential therapeutic approach to the treatment of the disease, as well as the influence of SAP upon protein folding.

1.2 Pentraxins

Human SAP shares 51% sequence identity and 66% homology with human CRP (Emsley 1994; Srinivasan *et al* 1994). These are the primary members of the pentraxin

superfamily: A group of closely related plasma proteins with calcium-dependent ligand binding that exhibit five-fold rotational symmetry. Similar plasma proteins have been found in all types of vertebrates and some invertebrates (Pepys *et al* 1978), including the phylogenetically ancient *Limulus polyphemus* (the horseshoe crab) (Amatayakul-Chantler *et al* 1993; Quigley *et al* 1994; Shrive *et al* 1999; Tennent *et al* 1993b). Several other proteins that contain the pentraxin homology domain but show low sequence identities have also been identified (Gewurz *et al* 1995). These more distantly related or 'long' pentraxins include the human cytokine controlled protein PTX-3 (also known as the TNF-stimulated gene-14 protein) (Bottazzi *et al* 1997; Lee *et al* 1994), neuronal activity-related pentraxins (NARP) (Tsui *et al* 1996), the *Cavia porcellus* sperm acrosome component apexin (Noland *et al* 1994; Reid & Blobel 1994), and a protein from *Xenopus Laevis* (Seery *et al* 1993). The complete sequences of a large number of pentraxins have currently been determined, although the three-dimensional structures of only human SAP (Emsley *et al* 1994), human CRP (Shrive *et al* 1996; Thompson *et al* 1999), and *Limulus* SAP (Shrive *et al* 1999) have been solved.

There is a remarkable amount of similarity in appearance under the electron microscope, subunit size, and sequence identity between the vertebrate pentraxins (Osmand *et al* 1977; Pepys *et al* 1978) (see table 1.1 and figure 1.1). The pairwise alignment (see table 1.1) shows that SAP or CRP-type pentraxins from different species have approximately 71% sequence identity, significantly higher than the similarity between those from the same species (~49%). Multiple sequence alignments (see figure 1.1) and comparative modelling upon the known SAP fold (Srinivasan *et al* 1994) suggest that they have a similar arrangement of β-sheets with a high conservation of residues involved in calcium binding, even in *Limulus* CRP that shares less than 31% sequence identity with the vertebrate pentraxins. Ligand binding and protomer interacting regions are also highly

Table 1.1. Sequence identity table of several SAP and CRP-type pentraxins from different sources. The percentage identities were calculated using Malign 4.0 (Wheeler & Gladstein 1994).

Percentage identities:		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	Subunit size (amino acids)
1	SAP: HUMAN	100	68.1	69.1	71.1	73.0	51.5	46.1	49.0	46.6	48.0	29.4	204
2	SAP: GOLDEN HAMSTER	68.1	100	71.1	76.0	63.7	50.5	48.5	49.0	46.4	49.0	29.2	212
3	SAP: MOUSE	69.1	71.1	100	79.9	69.6	52.0	45.6	47.1	45.6	45.1	27.0	204
4	SAP: RAT	71.1	76.0	79.9	100	66.7	51.0	48.5	49.5	47.1	47.6	30.8	208
5	SAP: GUINEA PIG	73.0	63.7	69.6	66.7	100	53.9	49.0	50.5	47.1	50.5	30.4	204
6	CRP: HUMAN	51.5	50.5	52.0	51.0	53.9	100	70.9	70.9	66.5	73.3	28.2	206
7	CRP: GOLDEN HAMSTER	46.1	48.5	45.6	48.5	49.0	70.9	100	78.2	75.7	70.9	30.1	206
8	CRP: MOUSE	49.0	49.0	47.1	49.5	50.5	70.9	78.2	100	72.3	67.5	29.1	206
9	CRP: RAT	46.6	46.4	45.6	47.1	47.1	66.5	75.7	72.3	100	65.0	26.1	211
10	CRP: GUINEA PIG	48.0	49.0	45.1	47.6	50.5	73.3	70.9	67.5	65.0	100	26.7	206
11	CRP1: HORSESHOE CRAB	29.4	29.2	27.0	30.8	30.4	28.2	30.1	29.1	26.1	26.7	100	218

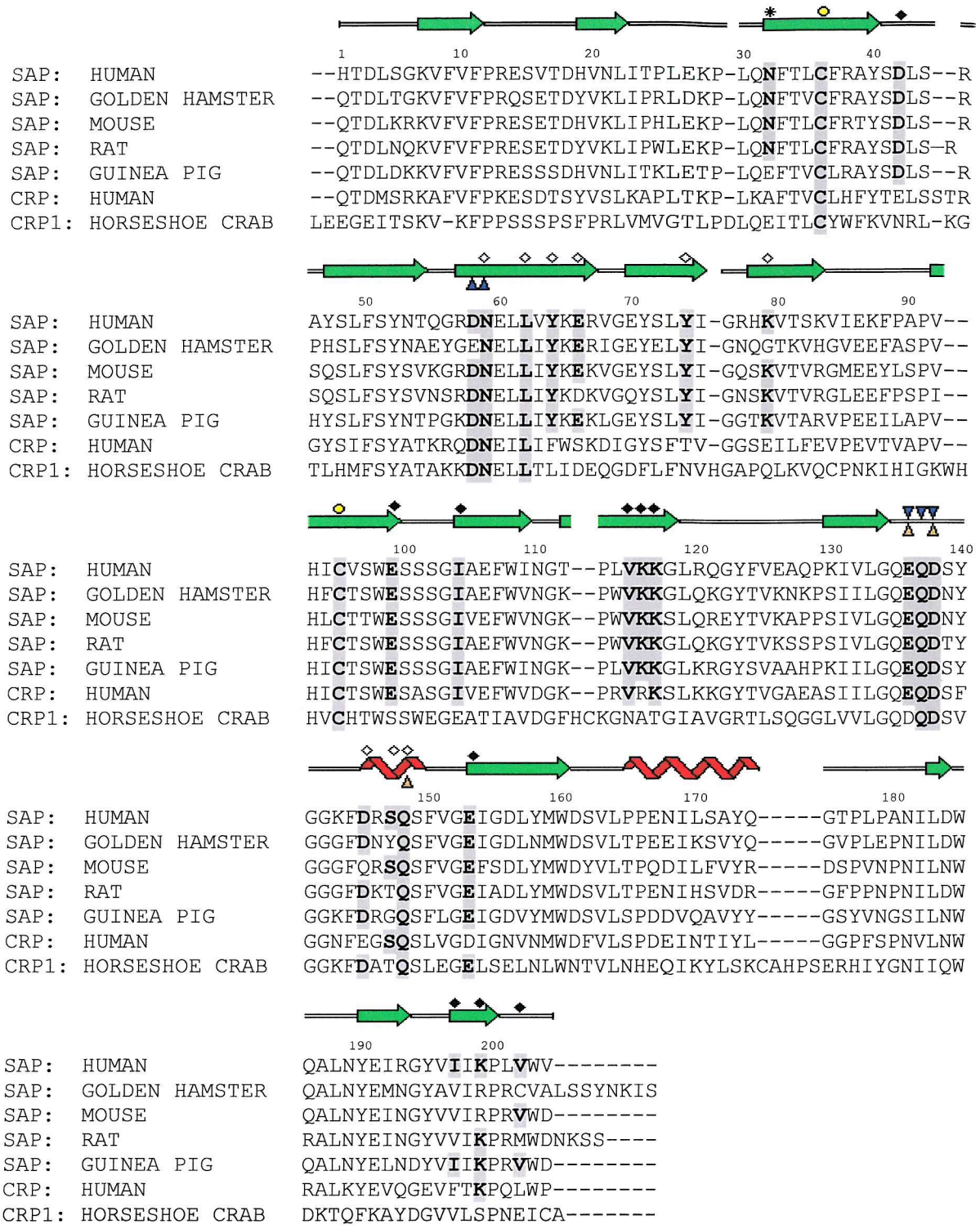


Figure 1.1. Primary sequence alignment of pentraxins from several different sources. The alignment was performed using Malign 4.0. The secondary structure and numbering of human SAP is shown with the important structural and binding residues highlighted; blue and orange triangles signify calcium I and II binding site residues respectively, open squares residues involved in ligand binding, closed squares residues involved in subunit contacts, star the glycosylation site, and the two yellow circles the disulphide bridge.

conserved within the SAP or CRP-type pentraxins, however, there are some significant differences between them responsible for both the similarities and differences in subunit arrangement and binding properties.

Human SAP and CRP are characterised by their different calcium-dependent binding specificities for phosphocholine (PC) and phosphoethanolamine (PE) (Schwalbe *et al* 1992), where SAP is only able to bind the latter while CRP binds both with high affinity (Emsley 1994). In most cases SAP-like pentraxins also bind the carbohydrate methyl 4,6-O-(1-carboxyethylidene)-β-D-galactopyranoside (MOβDG) (Hind *et al* 1984b), amyloid fibrils (Pepys *et al* 1979), and oligosaccharide moieties (Loveless *et al* 1992), while CRP-like pentraxins bind PC and pneumococcal C-polysaccharides (Gotschlich *et al* 1982; Volanakis & Kaplan 1971). However, SAP-type pentraxins from hamster display both amyloid and PC binding properties (Coe *et al* 1981; Coe & Ross 1985; Tennent *et al* 1993a).

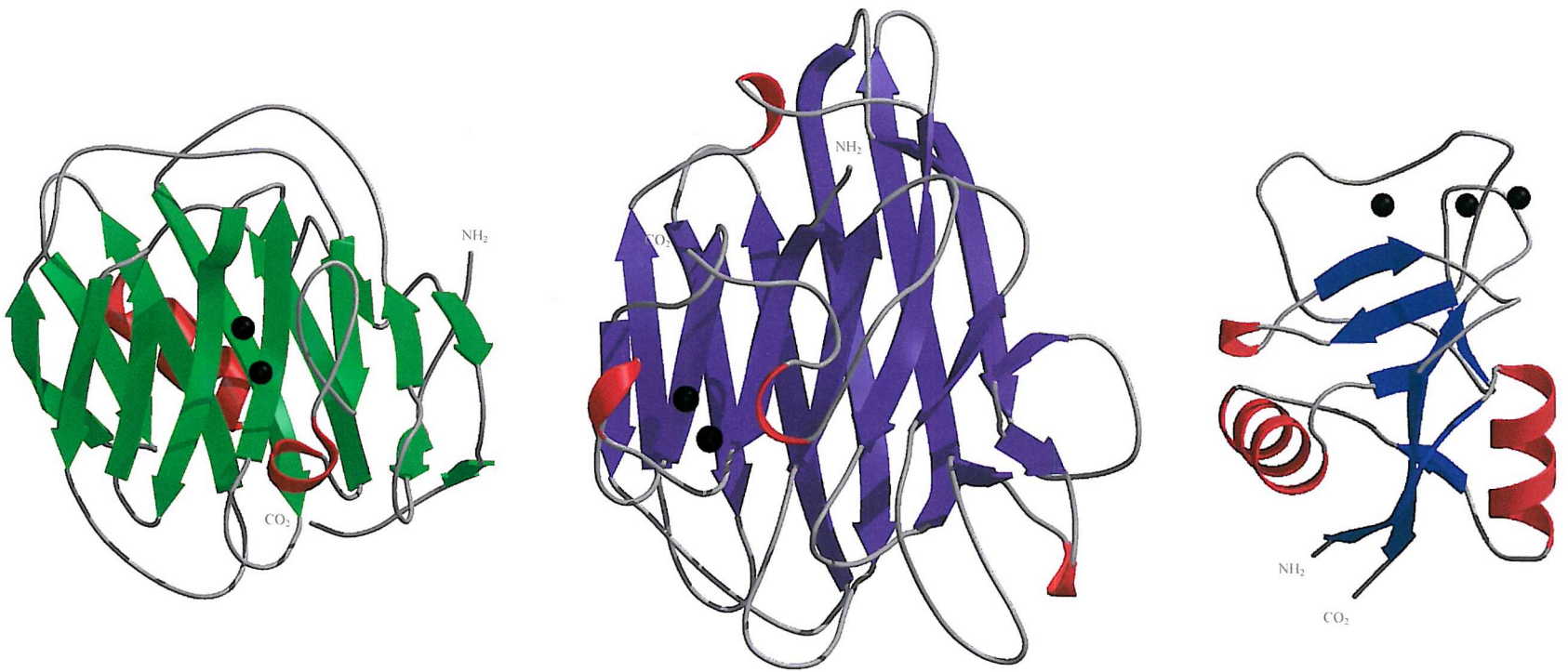
The pentraxins have been highly conserved throughout evolution. Stable conservation of both the SAP and CRP-type pentraxins within vertebrate evolution implies that they were duplicated from a common ancestral gene prior to the divergence of eutherian mammals and marsupials, as recent as 200 million years ago (Rubio *et al* 1993). While the identification of several sequence variants in *L. polyphemus* suggests the presence of a common ancestor in chordates and arthropods over 500 million years ago (Shrive *et al* 1999).

1.3 Lectins

SAP was first demonstrated to bind agarose derived from marine algae in 1977 (Pepys *et al* 1977). Agarose is composed of repeating units of agarobiose (1,3-linked β-D-

galactopyranose with 1,4-linked 3,6-anhydro-α-L-galactopyranose) and contains different trace amounts of pyruvate in a 4,6-cyclic pyruvate acetal of galactose depending upon the source. This galactoside is required for binding. It was synthesised and given the acronym MOβDG (Hind *et al* 1984b). SAP has now been shown to bind a wide range of pathogenic bacteria via oligosaccharide recognition *in vitro* and is a well-established member of the lectin family.

Lectins are sugar-binding proteins other than enzymes and antibodies that show some tertiary fold similarities but low sequence homology (Arason 1996; Rini 1995). There are six different lectin families classified by structure and functional similarities with legume and cereal lectins occurring in plants, while P-, S-, and C-type lectins, and pentraxins occur in animals. Oligosaccharide side-chain recognition forms the basis of many important interactions involved in fertilisation, development, cellular trafficking, leukocyte migration, and pathogen recognition. It is also the mechanism by which viral and bacterial cells attach to and infect host cells. Both the pentraxins and C-type lectins are able to discriminate between self and non-self and are implicated in host defence. The crystal structures of several metal-dependent lectins have been solved, and show different structural and functional similarities to SAP (see figure 1.2). Legume lectins show a similar tertiary fold to the flattened β-jellyroll of SAP, although the metal-sugar binding site differs in both location and function with metal ions required to maintain the site integrity but not directly involved in carbohydrate binding (Loris *et al* 1998). C-type lectins however, contain an unrelated topology to SAP but show functional similarities with carbohydrate binding via the formation of direct coordination bonds with calcium and hydrogen bonds with surrounding residues (Kolatkar & Weis 1996).



Serum amyloid P component
(Pentraxin)

Concanavalin A
(Legume lectin)

Mannose-binding protein A
(C-type lectin)

Figure 1.2. Structures of the human pentraxin SAP, the legume lectin concanavalin A (ConA) (Parkin *et al* 1996), and the C-type lectin mannose-binding protein A (MBP-A) (Weis *et al* 1992) illustrated as ribbon diagrams (prepared using Bobscript (Esnouf 1997) and Raster3D (Merritt & Murphy 1994)). The β-strands are represented as green, purple or blue arrows, α-helices as red ribbons, loops as grey cords, and metal ions as black spheres. SAP and ConA show similar topology within the two antiparallel β-sheets although there is significant divergence in terminal positions, metal binding location, and the length and conformation of loop regions. MBP-A does not show any structural similarities with either SAP or ConA.

SAP has now been shown to bind various bacteria *in vitro*, suggesting a possible role in host defence against infection (Hind *et al* 1985). Several different terminal oligosaccharide residues are recognized by SAP (Kubak *et al* 1988; Loveless *et al* 1992), including N-acetyl-D-galactosamine, D-mannose, 6-phosphate-mannose, glucuronic acid, and certain sulphated polysaccharides. These are thought to be involved in *Klebsiella rhinoscleromatis* and *Streptococcus pyogenes* binding, although the cell wall composition of the latter is currently unknown (Hind *et al* 1985). Recent *in vivo* studies using SAP knockout mice have shown that SAP actually enhances *S. pyogenes* infection, but significantly increases the resistance to *Escherichia coli* and *Salmonella typhimurium* which are not recognised by SAP (Pepys 1998). In fact, heavy binding of SAP to *S. pyogenes*, approximately 6,000 molecules per organism, is believed to mask the pathogen, protecting and stabilising it from the host's defence mechanism.

Carbohydrate recognition is not limited to bacterial pathogens, as influenza A virus haemagglutinin (HA) (Andersen *et al* 1997) and zymosan (Potempa *et al* 1985), a yeast cell wall polysaccharide derived from S*accharomyces cerevisiae*, are both bound by SAP. The binding of the influenza A virus is believed to occur through the terminal mannose of the HA1 protein, resulting in the inhibition of haemagglutination and infection *in vitro*. While, the binding of SAP to zymosan occurs with a binding constant (K_d 3.12 x 10^{-5}M (Potempa *et al* 1985)) comparable to a weak antibody reaction, this has recently been demonstrated to opsonize zymosan for phagocytosis *in vitro* (Bharadwaj *et al* 2001).

Several endogenous polysaccharides are also recognised by SAP; these include the glycosaminoglycans heparan sulphate and dermatan sulphate (Hamazaki 1987; Hamazaki 1989) which are frequently associated with amyloid fibrils, and the blood plasma anticoagulant heparin (Li *et al* 1994). Oligosaccharide recognition may also be involved in the binding of endogenous glycoproteins.

1.4 Amyloidosis

Amyloid was first described in the mid nineteenth century when pathologists at the time noticed that the vital organ parenchyma had been replaced by extracellular 'lardaceous deposits' or 'waxy degenerations' in diseased organs. These were subsequently termed amyloid or 'starch like' deposits due to positive carbohydrate tests with iodine (Virchow 1854). This was initially misleading as they are in fact extracellular deposits of insoluble protein associated with non-covalently bound glycosaminoglycans. The abnormal accumulation of amyloid is a major component of a large number of heterogeneous disorders known as the amyloidoses, but also accompanies a wide range of other human conditions such as cancer, diabetes, rheumatoid arthritis, and chronic renal dialysis (Gillmore *et al* 1997). These disorders can be either hereditary or acquired with deposits being focal, localised or systemic. Focal or small deposits are clinically silent and often incidental in the elderly (Rowe *et al* 1984), while systemic or localised deposits progressively disrupt normal tissue structure and function eventually leading to organ failure and death (Glenner 1980b; Glenner 1980a). The human cerebral forms of amyloidosis, Alzheimer's disease and Creutzfeldt-Jakob disease (CJD), have gained the most publicity over the past few decades due to the severity of the mental deterioration and the gradual increase in clinical cases.

1.4.1 Amyloid Deposits

Amyloid deposits are comprised of two distinct ultrastructures, rigid unbranched protein fibrils and a pentameric component known as amyloid P component (AP) (Bladen *et al* 1966; Cathcart *et al* 1967; Cohen & Calkins 1959; Glenner & Bladen 1966). The

insoluble protein fibrils are formed from the self-association of normally soluble protein precursors in an abnormal or non-native state (Kelly 1997). It is these precursors that vary depending upon the disorder and provides the basis for clinical classification of different forms of amyloidosis (see table 1.2). To date only 17 to 18 amyloidogenic proteins have been found to form fibrils *in vivo* (Tan & Pepys 1994). Although, *in vitro* studies on synthetic and non-amyloidogenic proteins suggest that all polypeptides can spontaneously self-associate into fibrils when subjected to acid denaturation and/or proteolytic cleavage (Jiménez *et al* 1999; Sunde *et al* 1997). These non-native fibrils show similar microscopic structural characteristics, X-ray diffraction patterns, and the same characteristic green-birefringence along the fibril axis when stained with Congo red and viewed with crossed polars (Puchtler *et al* 1962).

The pentameric component, AP, binds to all types of amyloid fibrils and is found universally associated with all types of amyloid deposits *in vivo* (Hawkins *et al* 1988; Hawkins *et al* 1990a; Pepys *et al* 1982). This protease resistant protein comprises up to 14% of the mass of amyloid deposits *in vivo* (Skinner *et al* 1980), and was found to be derived from, and identical to, the circulating form SAP (Painter *et al* 1982; Pepys *et al* 1982; Prelli *et al* 1985). Although not required for fibril formation or amyloid deposition *in vivo*, SAP has been shown to significantly contribute to the pathogenesis of amyloidosis. This was clearly demonstrated using the recently developed SAP knockout mice, where the absence of SAP significantly delayed and reduced the induction of AA amyloidosis via repeated casein injections (Botto *et al* 1997). SAP is believed to bind the abnormal fibril conformation initially reducing the rate of fibril formation (Janciauskiene *et al* 1995), but once formed contributes to fibril persistence by binding along the fibril axis and self-associating into perpendicular stacks (see figure 1.3). This effectively forms a protective

Table 1.2. Clinical classification of amyloidosis syndromes (Gillmore *et al* 1997).

Clinical syndrome		Fibril precursor
Alzheimer's disease		β-protein derived from β-amyloid precursor protein
Cardiomyopathy with persistent atrial standstill		Unknown
Cerebral amyloid angiopathy		β-protein derived from β-amyloid precursor protein
Cutaneous deposits		Unknown
Dialysis-related amyloid		β_2-microglobulin derived from high plasma levels
Down's syndrome		β-protein derived from β-amyloid precursor protein
Endocrine amyloidosis, associated with APUDomas		Peptide hormones or fragments thereof (e.g. precalcitonin in medullary carcinoma of the thyroid)
Familial British dementia		BRI gene genetic variant
Familial Mediterranean fever		AA derived from serum amyloid A
Focal senile amyloidosis:	atria of the heart	Atrial natriuretic peptide
	brain	β-protein
	joints	Unknown
	seminal vesicles	Seminal vesicle exocrine protein
	prostate	β_2-microglobulin
Hereditary cerebral haemorrhage with amyloidosis (cerebral amyloid angiopathy); autosomal dominant:		
	Icelandic type	Cystatin C, fragment of genetic variant
	Dutch type	β-protein derived from genetic variant in β-amyloid precursor protein
Local nodular amyloidosis (skin, respiratory tract, urogenital tract, etc) associated with focal immunocyte dyscrasia		AL fibrils derived from monoclonal immunoglobulin light chains
Muckle-Well's syndrome, nephropathy, deafness, urticaria, limb pain		AA derived from serum amyloid A
Non-neuropathic, prominent visceral involvement (Ostertag-type); autosomal dominant		ApoA1, lysozyme, or fibrinogen α-chain genetic variants
Ocular and orbital amyloid		Unknown
Predominant cranial nerve involvement with lattice corneal dystrophy; autosomal dominant		Gelsolin, fragment of genetic variants
Predominant cardiac involvement, no clinical neuropathy; autosomal dominant		Transthyretin genetic variants
Predominant peripheral nerve involvement, familial amyloid polyneuropathy; autosomal dominant		Transthyretin or Apolipoprotein A1 genetic variants
Primary localized cutaneous amyloid		? Keratin derived
Reactive systemic AA amyloidosis, associated with chronic active diseases		AA fibrils derived from serum amyloid A protein
Senile systemic amyloidosis		Transthyretin derived from plasma Transthyretin
Spongiform encephalopathies, prion diseases		Prion protein derived from prion protein precursor
Systemic AL amyloidosis, associated with immunocyte dyscrasia, myeloma, monoclonal gammopathy, occult dyscrasia		AL fibrils derived from monoclonal immunoglobulin light chains
Type II diabetes mellitus		Islet amyloid polypeptide, amylin, derived from its precursor protein

Figure 1.3. Electron micrograph of amyloid fibrils isolated from mouse tissue, lightly washed and negatively stained with phosphotungstate. Single rigid fibrils of 80 to 100Å in width are seen throughout the preparation. SAP is also present and can be seen in different views, from the side in stacks perpendicular to the fibril axis (open arrows), and from the top as pentagonal units (arrows) (Skinner *et al* 1982).

coat that stabilises and protects the abnormal structure from proteolytic cleavage, and possibly masks it from phagocytic clearance *in vivo* (Tennent *et al* 1995). Congo red and free glycosaminoglycans can inhibit the binding of SAP to amyloid fibrils *in vitro* (Gupta-Bansal & Brunden 1998). However, the protein coat can be stripped away by ethylenediaminetetraacetic acid (EDTA) or by calcium-dependent SAP ligands such as MOβDG (Hind *et al* 1984a).

The distribution of SAP can be visualised *in vivo* by scintigraphy with iodine-123 labelled SAP (^{123}I-SAP) (Hawkins *et al* 1988; Hawkins *et al* 1990a). The radiolabelled SAP is intravenously injected into the patient and the distribution monitored using a gamma camera. This non-invasive macroscopic imaging technique is sufficiently sensitive to quantify and locate the distribution of SAP within amyloid deposits from systemic and AL type amyloidosis. A normal healthy subject will show rapid distribution of ^{123}I-SAP within the blood plasma pool, where it has a half-life of 24 hours. Patients suffering from amyloidosis show rapid incorporation of radioactivity into amyloid deposits preventing the clearance of ^{123}I-SAP via the liver hepatocytes and dramatically increasing the half-life to 30 days (Hawkins *et al* 1990b). Figure 1.4 shows SAP scintigraphy scans from 2 young patients with different forms of systemic amyloidosis. SAP scintigraphy is the only method currently available for studying the progression and effect of potential treatments upon amyloid deposition *in vivo*, and is now routinely used in conjunction with histological examinations with Congo red and immunohistochemical staining in the diagnosis of amyloidosis. However, not all types of amyloidosis can be identified this way. For instance cardiac and cerebral amyloid is poorly visualised by scintigraphy, while biopsies are often too invasive to be taken.

Glycosaminoglycans and proteoglycans have been found in all types of amyloid deposits in which they have been sought (Glenner 1980b; Snow & White 1989). These normal extracellular matrix components are non-covalently bound to the fibrils and bind along the fibril axis. Proteoglycan binding is inhibited by Congo red, free glycosaminoglycans, and prevents SAP binding *in vitro* (Gupta-Bansal & Brunden 1998). Recent experiments have suggested that proteoglycans accumulate prior to fibril deposition and are actually up regulated during amyloidosis. Hence they may have a role in fibril formation and/or persistence *in vivo* (Stenstad *et al* 1994). Binding of the proteoglycans

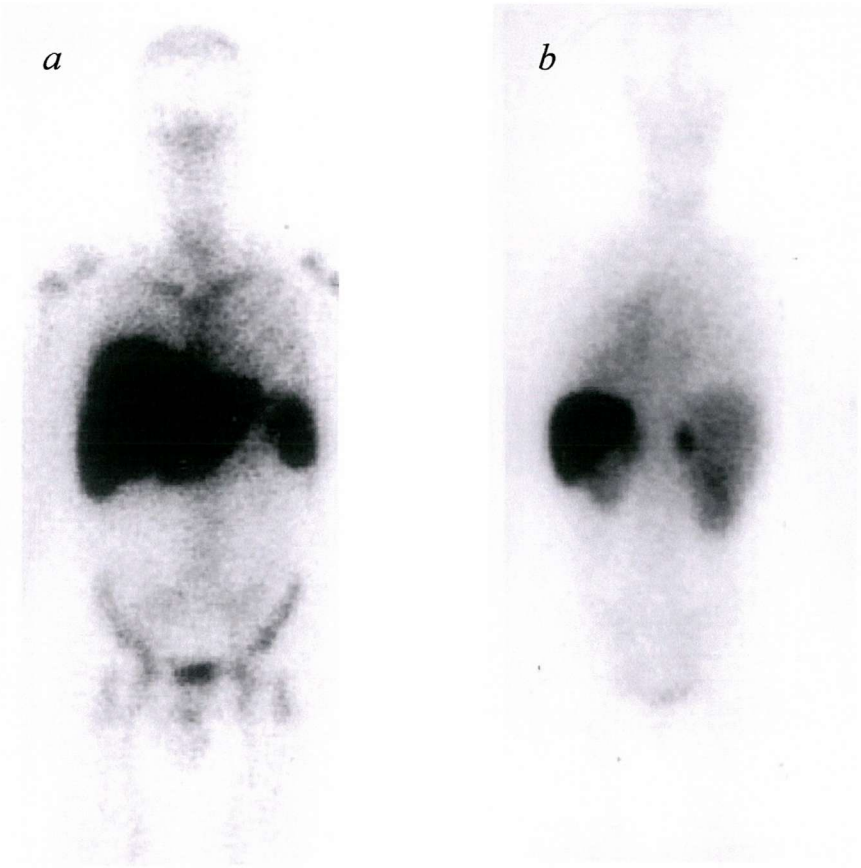


Figure 1.4. ^{123}I-SAP scintigraphy body scans from two patients with systemic amyloidosis. *a*, Anterior scan from a 26-year-old man showing substantial liver, spleen and bone marrow amyloid characteristic of systemic AL amyloidosis. *b*, Posterior scan from a 34-year-old woman with rheumatoid arthritis showing AA amyloid deposits in the spleen, adrenals, and kidneys (Gillmore *et al* 1997).

have also been demonstrated to protect the fibrils from both proteolytic (Gupta-Bansal *et al* 1995) and cellular (Shaffer *et al* 1995) degradation.

Other amyloid components show high fibril specificity. Apolipoprotein E, α_1-antichymotrypsin, ubiquitin, interleukin-1, and lysosomal hydrolases for example are all found associated with the cerebral deposits of Alzheimer's disease (Ashall & Goate 1994).

1.4.2 Fibril Structure

Amyloid fibrils appear as long unbranched rigid fibrils of indeterminate length with diameter 70 to 120Å when viewed in the electron microscope (Cohen *et al* 1982). Early X-ray fibre diffraction patters of isolated fibrils, orientated with the fibril axis, gave simple diffraction patterns with sharp 4.7 to 4.8Å meridional and diffuse 10Å equatorial reflections (Bonar *et al* 1967; Eanes & Glenner 1968). These diffraction patterns arise from the molecular spacing of well-ordered repeating elements, and is characteristic of the cross-β structure first described for silk from the insect *Chrysopa flava* (Geddes *et al* 1968). This cross-β structure is comprised of numerous polypeptide chains arranged in β-sheets running the length of the fibril with strands perpendicular to the fibril axis. Hence, the meridional reflection is derived from the separation distance between hydrogen bonding strands, while the equatorial reflection is derived from the separation distance between β-sheets.

The recent interpretation of electron microscopy imaging and high-resolution synchrotron diffraction patterns was used to derive a theoretical model for the substructure of a transthyretin amyloid fibril (Blake & Serpell 1996; Serpell *et al* 1995). These fibrils are of uniform width (130Å) and contain 4 protofilaments, measuring 40 to 50Å in diameter, arranged in a square array interwinding around a central hollow core. Each protofilament is believed to be comprised of 4 parallel β-sheets containing a common 15° twist between adjacent strands, assumed to be right handed, creating a novel 115.5Å or 24

strand repeating β-helix (see figure 1.5). This allows the sheets to hydrogen bond throughout the entire length of the fibril forming a rigid stable structure enclosing and stabilised by the hydrophobic core. The cross-β structure is assumed to be comprised of antiparallel sheets although this cannot be experimentally verified, and may therefore be parallel or a mixture of both.

The native structure of the transthyretin tetramer has previously been solved by X-ray crystallography (Blake *et al* 1978; Sebastião *et al* 1998). Each protomer is comprised of 8 β-strands arranged in 2 β-sheets with a 25° twist (see figure 1.5). Dimers are formed through the association of the end strands creating two elongated β-sheets. When compared to the protofilament model (see figure 1.5), it is clear that an altered monomer or dimer could form the basic building block of a protofilament, although the 25° β-sheet twist and 2 of the strands from each protomer would need to be reduced in order to create the 24 strand repeating unit (Blake & Serpell 1996).

The cross-β structure is able to accommodate various polypeptide chain lengths by altering the strand and repeating unit characteristics while maintaining the core structure. Similar synchrotron diffraction patterns have been obtained from several *ex vivo* and synthetic fibrils and are consistent with the proposed structure for a protofilament (Jiménez *et al* 1999; Sunde *et al* 1997).

1.4.3 Fibril Formation

The formation of amyloid fibrils *in vitro* resembles the mechanistic process of nucleation-dependent polymerisation seen in protein crystallography (Jarrett & Lansbury, Jr. 1993). Nucleation occurs through the association of native β-structural elements via the slow controlled aggregation of soluble partially folded intermediates, creating the characteristic cross-β sheet structure via the conversion of native α-helix to β-sheet

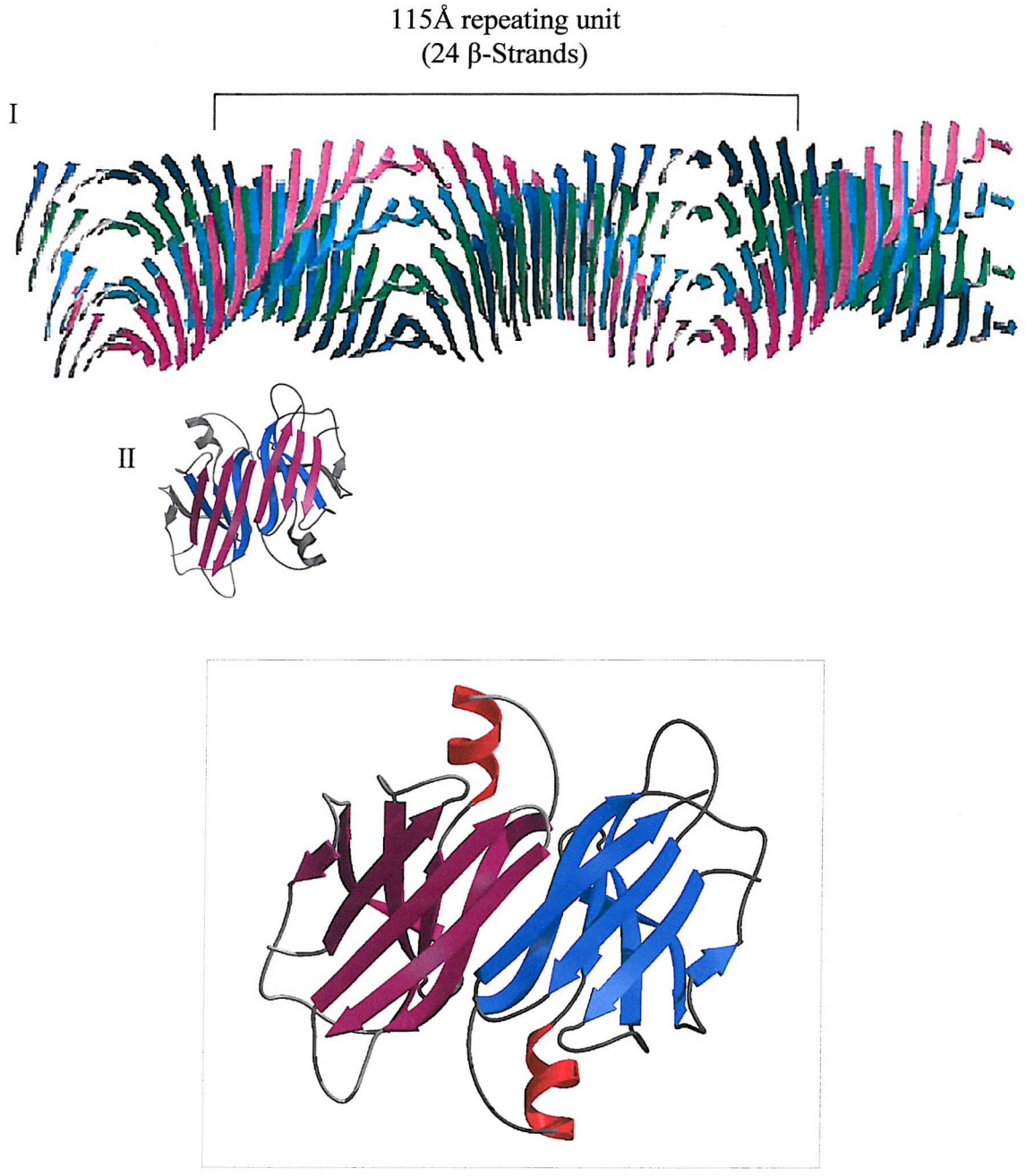


Figure 1.5. Comparison of the molecular model of a protofilament from a transthyretin amyloid fibril (Sunde *et al* 1997) with the three-dimensional structure of the native protein from which it is derived (prepared using Bobscript and Raster3D). The protofilament is comprised of four β-sheets running parallel to the protofilament axis with their component strands perpendicular, represented as different colours with the arrows illustrating the paths of the β-strands (I). The 15˚ inter-strand twist creates a 24 β-strand (or 115.5Å) repeating helix, which is derived from 2 elongated β-sheets that comprise the transthyretin dimer, coloured with respect to their sheet composition (II). This requires 2 strands at either end of the elongated sheets (coloured grey) and the 25˚ β-sheet twist to be reduced. The three-dimensional structure of the transthyretin dimer is also shown in the enclosed region where the β-strands are coloured with respect to their monomer composition, α-helices as red ribbons, and loops as grey cords (Blake *et al* 1978).

(see figure 1.6) (Booth *et al* 1997). This continues at a slow rate until the entropy cost is exceeded by the number of new interactions created during the association, thus making further additions thermodynamically favourable resulting in rapid polymerisation (Naiki & Gejyo 1999). The rate of nucleation and polymerisation is dependent upon the local precursor concentration and the kinetic stability of the native fold relative to the partially folded intermediate. This can be altered by point mutations or chemical modifications at structurally vital locations (Canet *et al* 1999; McCutchen *et al* 1993; McCutchen *et al* 1995; Orpiszewski & Benson 1999; Roher *et al* 1993). The protein stability can also be altered via interactions with other amyloid components (Soto *et al* 1995) and environmental factors (Bush *et al* 1994; Creighton 1996; Wood *et al* 1996). For example, the glycosaminoglycan heparan sulphate significantly alters the secondary structural elements of some precursor proteins to a more β-sheet conformation (Fraser *et al* 1992; McCubbin *et al* 1988), while SAP inhibits nucleation and possibly polymerisation by increasing the solubility of the precursors (Janciauskiene *et al* 1995), effectively reducing the concentration of available proteins to associate.

Fibril formation can also be induced both *in vivo* and *in vitro* by the formation or presence of a nidus (originally known as an amyloid enhancing factor) around which fibrillogenesis can occur (Kisilevsky & Fraser 1996). Conformational elements of the nidus are believed to provoke the amyloid protein into the fibril conformation, increasing the initial rate of nucleation when in the presence of sufficient concentrations of the precursor. The precise nature of the nidus is unclear although there are many different proteins that may have this function. Preformed fibrils can act as an endogenous nidus or seed (Johan *et al* 1998). This has been demonstrated by the induction of amyloid deposition in hamsters and mice upon injection of sonicated amyloid fibrils (Kisilevsky & Boudreau 1983; Niewold *et al* 1987). This presents a possible model for the amplification

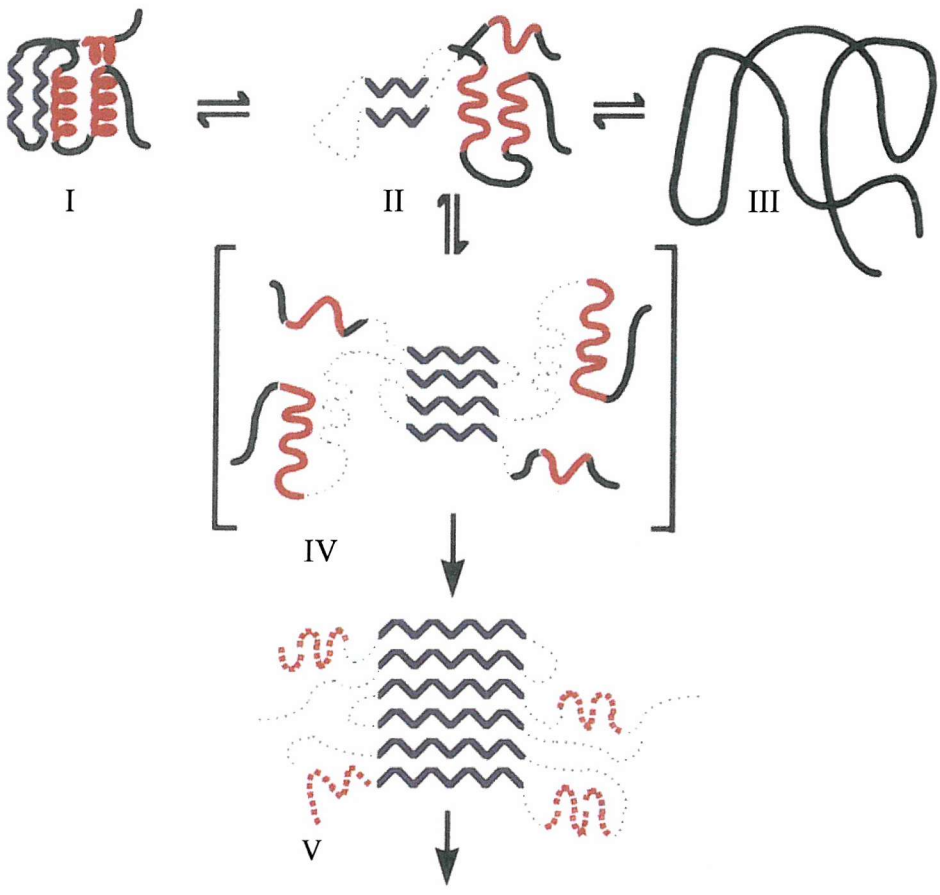


Figure 1.6. Proposed mechanism for fibril formation. The partially folded intermediate state (II) is structurally distinct from the native (I) and unfolded (III), allowing self-association via the β-structural elements. This produces a template for further additions (IV) eventually leading to the formation of a stable β-core structure (V). β-strands are represented in blue, helical structures in red, loops/coils in black, and the dotted lines are undefined (Booth *et al* 1997).

of amyloid *in vivo* via fibril fragmentation, but also identifies a possible transmissible agent. An exogenous seed would require substantial complementation between the seed and amyloidogenic protein for fibrillogenesis to occur, although the conformational differences would significantly reduce the initial rate of nucleation (Jarrett & Lansbury, Jr. 1993). This helps to explain the species barrier and pathogenic nature of the prion proteins in transmissible spongiform encephalopathies.

1.4.4 Potential Treatments for Amyloidosis

The production and formation of amyloid is believed to be fundamental to the pathology of amyloidosis. Although there is conflicting evidence about the precise mode of toxicity, there is a general consensus that this is either directly or indirectly due the presence of the amyloid fibrils. Huge advances in the diagnosis and characterisation of the amyloidoses have lead to the development of new approaches to promote regression and inhibit formation. However, there are currently still only a few type specific treatments available (Gillmore *et al* 1997).

It is now apparent from ^{123}I-SAP scintigraphy studies upon type specific treatments (Hawkins 1994; Persey *et al* 1996), and induced amyloidosis in mice (Kisilevsky & Boudreau 1983), that you can arrest or regress amyloid deposition by reducing the availability of the fibril precursor. Amyloid deposits are therefore dynamic and not inert as previously believed. This has sparked great interest, identifying the synthesis and clearance of these amyloidogenic proteins as a major therapeutic target. A novel approach at reducing the concentration of available precursors is currently being developed, whereby the production of antibodies against the precursor has been found to enhance clearance and/or prevent deposition *in vivo* (Schenk *et al* 1999). Initially achieved via immunization with the amyloidogenic precursor itself, this has lead to the development of

nonamyloidogenic derivatives that are currently being developed for therapeutic use (Sigurdsson *et al* 2001).

Broadband targets for potential treatments are centred upon the common ultrastructures present in all types of amyloid deposits. By directly targeting the common cross-β structure it may be possible to sufficiently reduce the rate of fibril nucleation and/or polymerisation. Several glycosaminoglycan mimics (Kisilevsky *et al* 1995) and other low molecular weight molecules (Howlett *et al* 1999a; Howlett *et al* 1999b; Kuner *et al* 2000) have already been shown to affect fibril nucleation and/or polymerisation *in vitro*. The amyloid fibrils can also be targeted indirectly by removing the protective protein coat and exposing the abnormal fibrillar conformation to the endogenous protein clearance mechanisms. Therefore, calcium-dependent ligands that inhibit or dissociate SAP binding to amyloid fibrils should prevent new amyloid deposition and accelerate the regression of existing amyloid (Pepys *et al* 1996; Tennent *et al* 1995).

1.5 Human Serum Amyloid P Component

The crystal structure of human SAP was solved to high resolution in 1994 using multiple isomorphous replacement (MIR) (Emsley *et al* 1994; O'Hara 1992). The molecular structure of SAP has provided great insight into the physical and biological properties of the protein. Subsequent co-crystallisations with low molecular weight ligands have also revealed important details about the interactions involved in ligand binding.

The atomic coordinates for the original crystal structure of SAP are currently available from the Protein Data Bank (Berman *et al* 2000), entry code 1SAC. Several other structures have subsequently been determined by molecular replacement (MR) (Hohenester *et al* 1997; Pye 2000; Thompson 1997; Thompson 2000). A summary of the processing

and refinement statistics for the different SAP structures can be found at the end of this section in table 1.3.

1.5.1 Protomer Structure

SAP is an oligomeric glycoprotein composed of identical 204 residue protomers (molecular weight 25,462Da) coded for by a single gene on chromosome 1 (Mantzouranis *et al* 1985). Each protomer contains 15 β-strands and 2 α-helices (see figure 1.7). The β-strands, labelled A to O from the N- to C-terminus, are arranged in two large antiparallel sheets of strands ACHIJLMO and BDEFGKN in the form of a flattened β-barrel with jellyroll topology (see figure 1.7). Strands ABCDKLMNO display the classic jellyroll topology while strands EFGHIJ form a β-meander at one end. The tertiary structure is stabilised by the hydrophobic core created between the opposing β-sheets, the N/O end of which is slightly open and accessible to solvent. The larger of the two α-helices (residues 165-174) is held against the exposed surface of the flatter sheet (strands ACHIJLMO) by several hydrophobic contacts and a hydrogen bond between Arg38 and Tyr173. Underneath the helix lies the only disulphide bridge between Cys36 and Cys95 of adjacent C and H strands respectively. Strand C also contains a glycosylation site at residue 32. The N-linked oligosaccharide bound is a complex bi-antennary chain and makes up 8% of the molecular weight of the protein (see figure 1.8) (Emsley 1994). The opposing buckled surface (strands BDEFGKN) contains a short five-residue helix (residues 145-149), a small hydrophobic pocket, and a double calcium-binding site.

Non-repetitive elements make up the remainder of the secondary structure of SAP. This involves 96 residues and comprises of a number of tight turns and hairpins that reverse the direction of the polypeptide chain enabling the tight antiparallel β-sheet structure to be formed. These include four tight β-hairpins between strands D-E, E-F, F-G, and H-I that

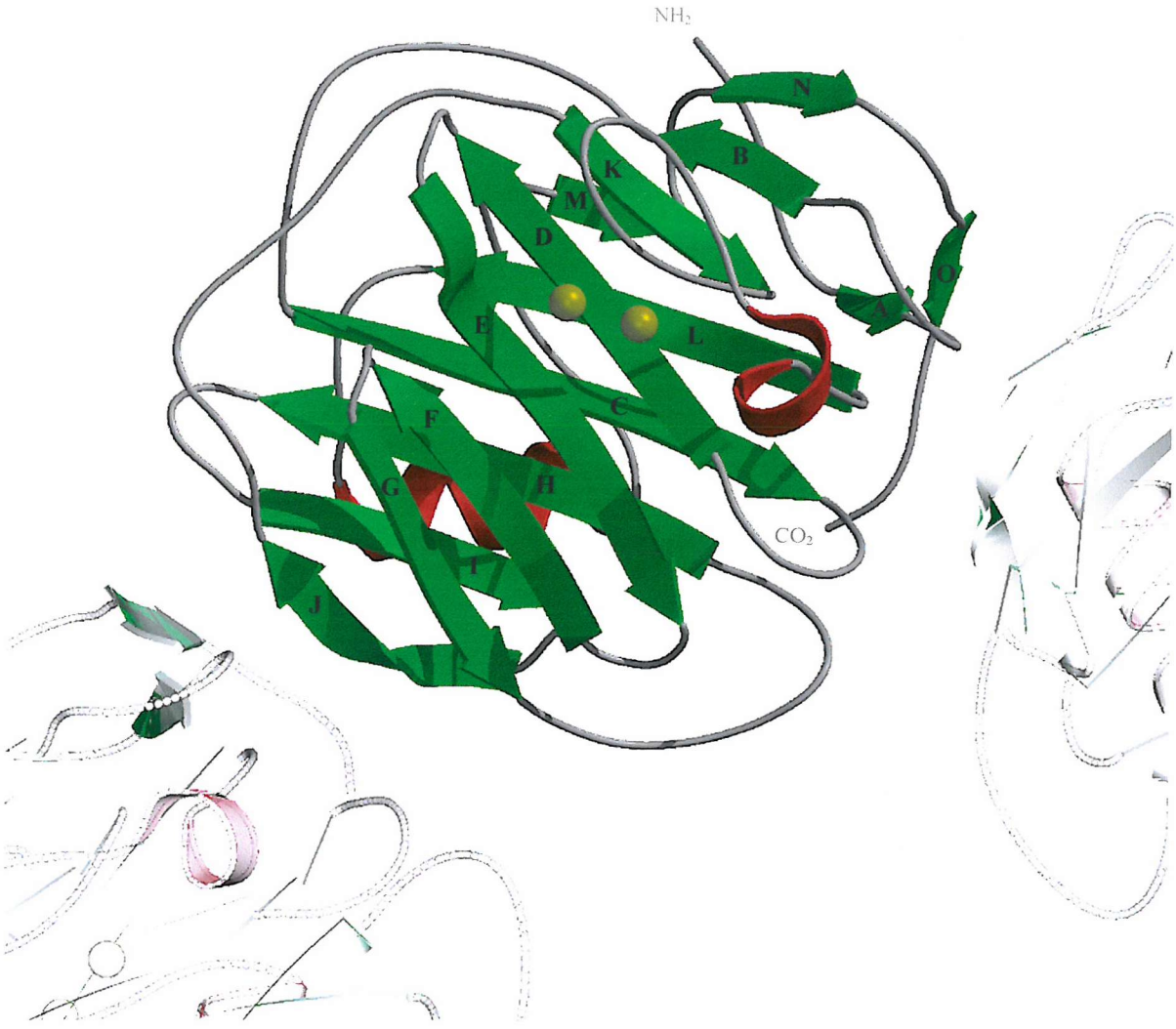


Figure 1.7. Three-dimensional structure of an SAP protomer viewed from the double calcium-binding site (prepared using Bobscript and Raster3D). The β-strands, labelled A to O from the N- to C-terminus, are arranged in two antiparallel sheets in the form of a flattened β-barrel with jellyroll topology, creating two distinct solvent exposed surfaces, the double calcium-binding and opposite helix bearing surface. The β-strands are represented as green arrows, α-helices as red ribbons, loops as gray cords, and calcium ions as yellow spheres.

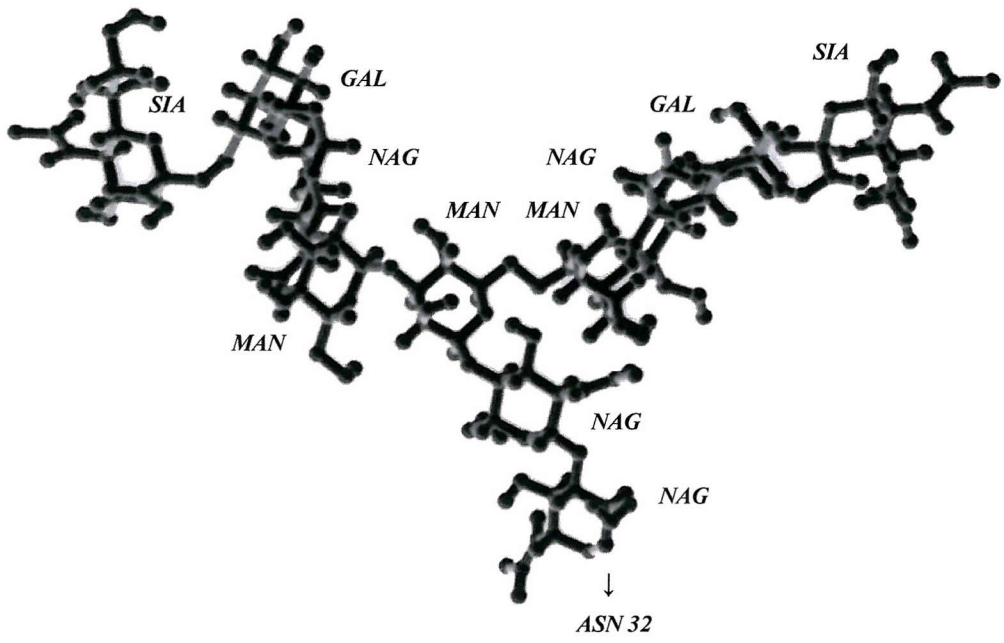


Figure 1.8. Graphical representation of the complex oligosaccharide N-linked to Asn32 (theoretical conformation) (prepared using SETOR (Evans 1993)). The pentasaccharide core consists of three mannose (MAN) and two N-acetylglucosamine (NAG) sugar residues with a N-acetylglucosamine, galactose (GAL), and sialic acid (SIA) cap built on.

contain a type II, II, I′, and I β-turn respectively. The two α-helices provide the linking regions either side of strand L, while the remaining strands are connected via loop structures.

1.5.2 Calcium Binding

The loop and helical region between strands K and L, and the N-terminal residues of strand E create the double calcium-binding site (see figure 1.9). Calcium I is coordinated by the side-chains of Asp58, Asn59, Glu136, Asp138, and by the main chain carbonyl of

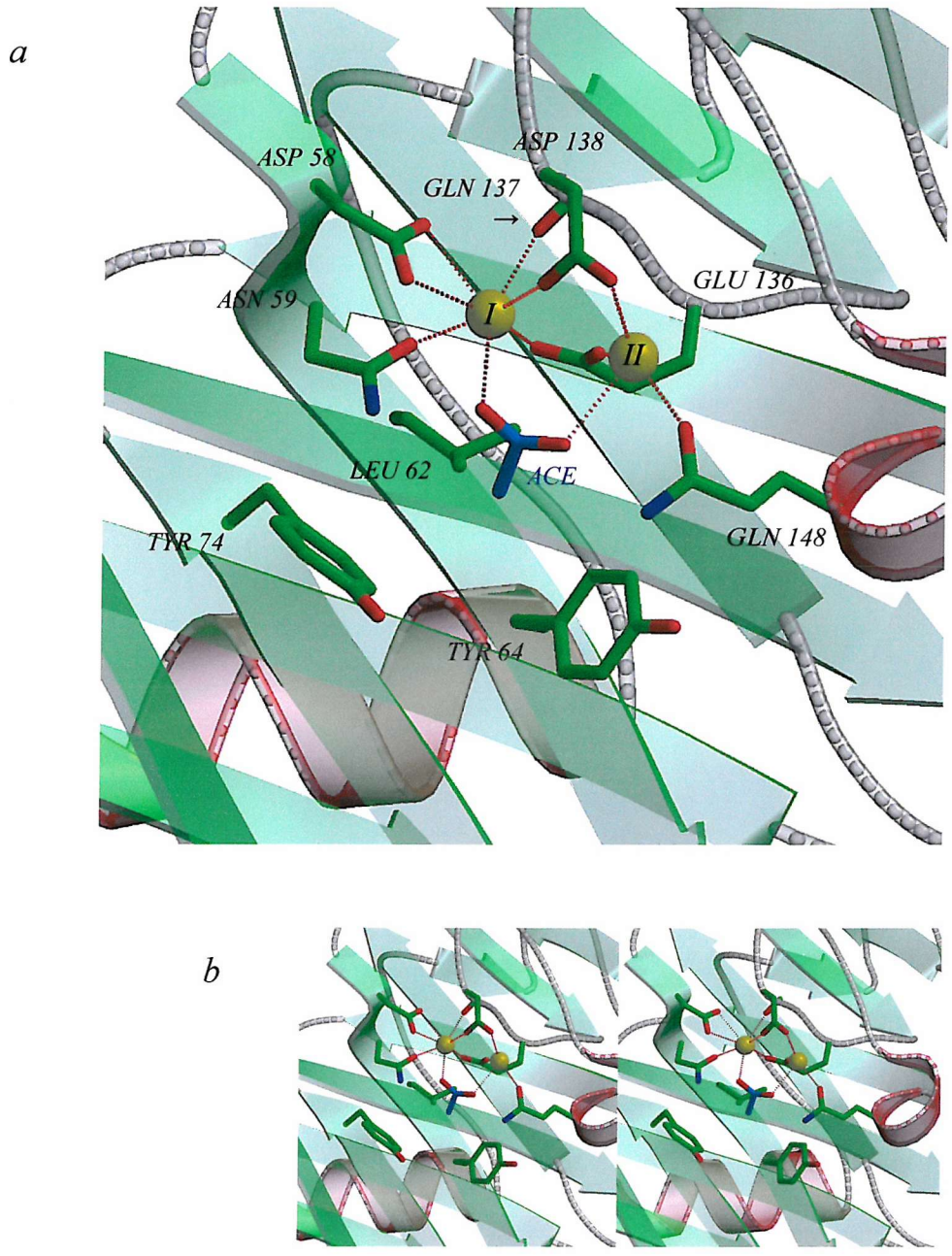


Figure 1.9. Three-dimensional representation of the double calcium-binding site of SAP viewed as an orthogonal projection (*a*) and stereoview (*b*) (prepared using Bobscript and Raster3D). The two calcium ions receive ligands from the surrounding protein side-chains, two water molecules (not shown), and an acetate ion (ACE) from the solvent (indicated by dotted red lines). The acetate ligand bridges the two calcium ions orientating it over the adjacent hydrophobic pocket (created by the side-chains of Leu62, Tyr64, and Tyr74). Secondary structural elements are represented as previously described with the important side-chains shown as stick models and labelled accordingly.

Gln137. Asp58 provides two ligands while the seventh coordination site is filled by a buffer ion. Calcium II is coordinated by the side-chains of Glu136, Asp138, Gln148, and by two water molecules. Residue Glu136 and Asp138 bridge the 2 calcium ions holding them ~4Å apart. The second calcium is more loosely bound and a water molecule, hydrogen-bonding to Glu136, Asp138, and Gln148, can occupy the site in some structures (Thompson 1997).

Metal ion binding at these two sites is not just restricted to calcium. SAP shows affinity for several other ions including barium, cobalt, copper, nickel, and zinc (Pepys *et al* 1997). In contrast to calcium, the larger barium ion displays a greater affinity for site II over site I (Thompson 1997).

When both calcium sites are unoccupied the subunits are susceptible to limited proteolysis with α-chymotrypsin and other proteases (trypsin, pronase, and nagarse protease) at residues 144-145, cleaving the subunits into 2 fragments of 18kDa and 7.5kDa (Kinoshita *et al* 1992). Although the fragments remain together under nondenaturing conditions, cleavage prevents calcium-dependent binding. Hamster SAP, rabbit CRP, and human CRP are also cleaved in the same homologous position as SAP (Kinoshita *et al* 1989; Kinoshita *et al* 1992). The removal of both calcium ions in human CRP has been shown to destabilise the small loop region involved in calcium binding, allowing it to protrude away from the body of the protein (Shrive *et al* 1996). This is believed to occur in SAP, which would expose the equivalent loop, residues 138-148, to proteolytic attack.

1.5.3 Oligomeric Structure

In the past there has been some confusion about the oligomeric state of SAP *in vivo*, however experiments conducted under physiological condition demonstrate that SAP forms stable pentamers in serum (Hutchinson *et al* 2000; Sørensen *et al* 1995). The pentameric

assembly is comprised of 5 subunits arranged in a ring, 95Å in diameter and 35Å deep, with radial symmetry creating a central pore 20Å wide (see figure 1.10). The β-sheets from each subunit are tightly held nearly perpendicular to the radial symmetry by several non-covalent interactions. This creates two separate and distinct faces; the 'alpha' or A-face contains the five larger α-helices and has a net negative charge, while the opposing 'binding' or B-face contains the double calcium-binding sites and has a net positive charge (Pepys *et al* 1997).

SAP forms stable decamers at physiological pH and ionic strength in the absence of calcium, or presence of the chelator EDTA (Painter *et al* 1982; Pepys *et al* 1978). The decameric assembly is comprised of two pentamers interacting face-to-face. It is not known whether the decamer is formed through the association of A or B faces, although X-ray and neutron solution scattering infers A-face stacking (Ashton *et al* 1997) which would be consistent with proteolytic cleavage in the absence of calcium. The crystal structure of this complex is yet to be solved (Pye 2000). Two ligand-induced decameric structures have been solved. These calcium dependent ligands are MOβDG which facilitates A-B face decamerization (Thompson 1997) and 2'-deoxyadenosine-5'-monophosphate (dAMP) which link the B-faces together (Hohenester *et al* 1997).

1.5.4 Aggregation

Isolated human SAP rapidly autoaggregates in solution at physiological pH in the presence of free calcium ions (Baltz *et al* 1982; Schwalbe *et al* 1990). This is concentration dependent and is believed to occur via A-B face stacking, where protruding glutamic acids (residue 167) on the A-face are thought to interact with the bound calcium ions on the B-face of an adjacent pentamer. Several Glu167 mutants have been produced and expressed in a baculovirus system, two of which (serine or glutamine) completely abolish calcium-

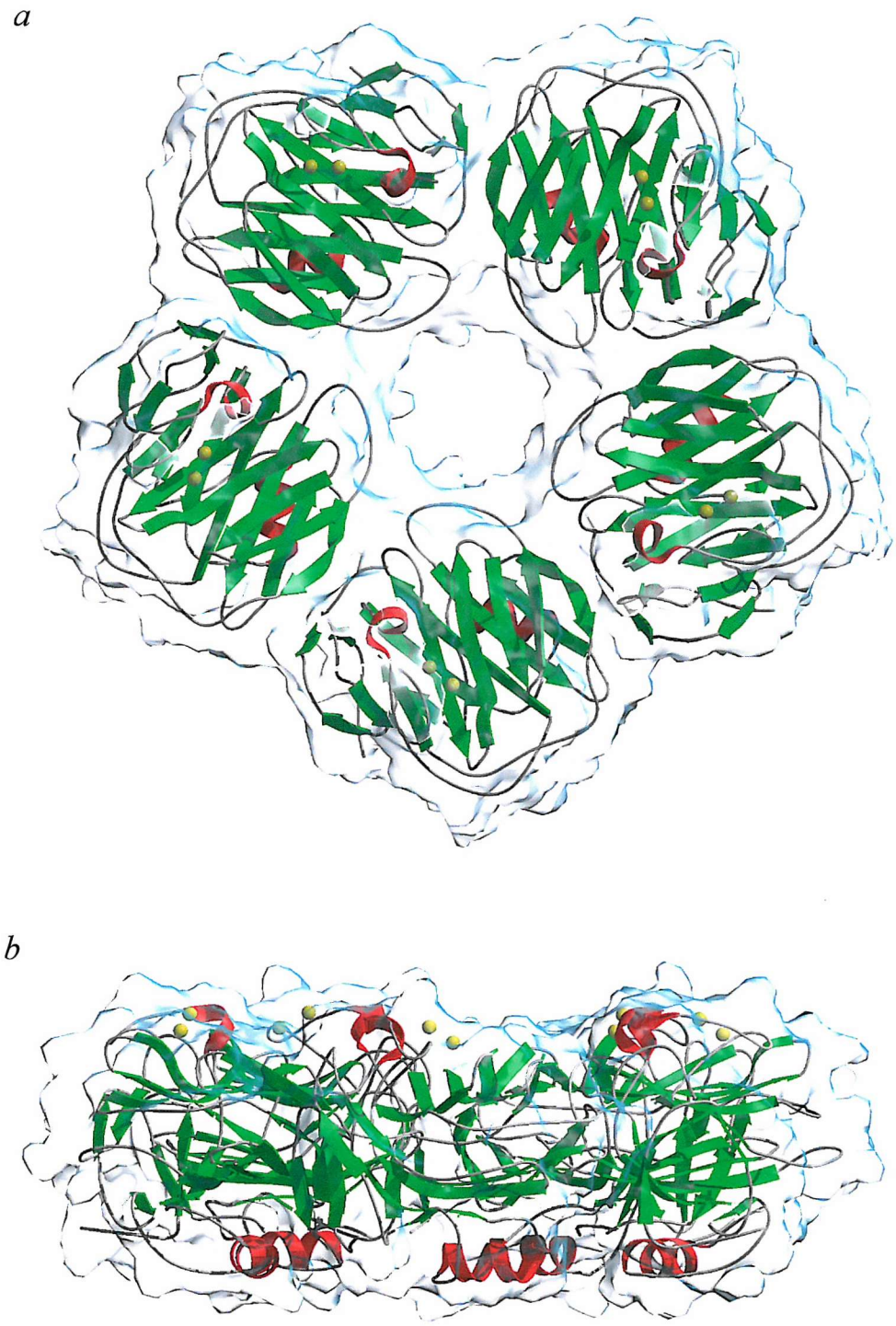


Figure 1.10. Three-dimensional structure of an SAP pentamer viewed horizontal (*a*) and perpendicular (*b*) to the radial symmetry (prepared using Bobscript, Grasp (Nicholls *et al* 1991), GL Render (Esser 2001), and Raster3D). The subunits are arranged in a ring with their β-sheets held nearly perpendicular to the radial symmetry, creating a calcium binding and opposite α-helical face. Secondary structural elements are represented as previously described within the transparent surface.

dependent aggregation (Pepys *et al* 1997). Interestingly mouse and guinea pig SAP contain a natural mutation at this point (see figure 1.1). Aggregation is a major problem when carrying out calcium-dependent binding studies. Unfortunately sources of non-aggregating protein are unable to produce sufficient yields, but aggregation can be inhibited by the presence of calcium-dependent ligands and this can be used to identify new ligands (Schwalbe *et al* 1992). Aggregation can be prevented by using buffer systems with a pH of 5.5 or below (Ashton *et al* 1997), or by the presence of bovine or human serum albumin at serum concentrations of 4% (w/v) or above (Hutchinson *et al* 2000; Pepys *et al* 1997). The effects of other serum metal ions upon aggregation, such as copper and zinc, have yet to be studied.

1.5.5 Ligand Binding

The affinity of SAP for agarose derived from marine algae was originally used in the calcium-dependent isolation of SAP from serum and amyloid extracts (Pepys *et al* 1977). The galactoside responsible for agarose binding was later synthesised as MOβDG. The R-isomer of MOβDG was found to bind with the highest affinity (K_d ~10^{-3}M (Emsley 1994)) and is able to completely inhibit calcium-dependent aggregation of SAP. MOβDG has subsequently been used to identify and study important biological interactions *in vitro*. SAP is able to bind PE with higher affinity than MOβDG, and can be used in the one step preparation of SAP and CRP via calcium-dependent affinity chromatography upon PE covalently immobilized through its amino group (de Beer & Pepys 1982; Hawkins *et al* 1991; Pontet *et al* 1978). This is currently the method of choice and allows easy separation of the two pentraxins via their affinity for PC, which even at high concentrations is unable to displace SAP from the affinity column.

In vivo and *in vitro* studies have shown that SAP associates with all types of amyloid fibrils. SAP can be fully dissociated from amyloid by MOβDG or by removing the free calcium ions with EDTA. SAP has been shown to bind the major glycosaminoglycans associated with amyloid fibrils when free in solution. When tightly bound to amyloid fibrils, glycosaminoglycans may to some degree be involved in SAP binding (Nelson *et al* 1991a; Stenstad *et al* 1993). However, SAP binds fibrils derived from pure proteins or polypeptides *in vitro* with the same affinity as *ex vivo* fibrils (K_d ~10^{-8}M (Pepys 1998)) (Tennent *et al* 1995). Amyloidogenic proteins can inhibit fibril binding when in their fibrillar conformation but not their native soluble state, suggesting a specific fibrillar motif is recognized by SAP (Lovat 1997). Although both fibrillar and nonfibrillar forms of β-protein (Aβ), the precursor protein associated with Alzheimer's disease, are able to bind SAP, only the fibrillar form is rendered resistant to proteolysis with pronase (Gupta-Bansal & Brunden 1998; Tennent *et al* 1995). Interestingly, Congo red and free glycosaminoglycans only prevent the association of SAP with the fibrillar form of Aβ suggesting a different motif or mechanism of binding to the nonfibrillar form, although both are dependent upon the presence of calcium ions. Based upon the experimental results and examination of the theoretical model of an amyloid fibril it was suggested that the turn and/or loop structures, expected to lie on the surface of the fibril, are recognised by SAP through the double calcium-binding sites (Pepys 1986). In fact sequences which form β-turns are capable of binding cations with high selectivity for calcium over sodium, potassium, and magnesium (Kula *et al* 1977; Urry 1974), and some β-turns can be docked into the calcium-binding site of SAP (Pepys *et al* 1997).

SAP is a normal matrix component of numerous basement membranes including the glomerular basement membrane (GBM), alveolar basement membranes, and the sweat gland basement membranes (al Mutlaq *et al* 1993; Dyck *et al* 1980b). The most well

characterised tissue amyloid P component (TAP) is that of the GBM, which is confined to the *lamina rara interna* but can only be released upon digestion of the membrane with collagenase (Dyck *et al* 1980b). TAP was found to constitute ~10% of the protein released, and was covalently associated with peptide fragments believed to be derived from collagen, although other matrix components should not be discounted. Several isolated extracellular matrix components have been found to bind SAP non-covalently in the presence of calcium, these include fibronectin (de Beer *et al* 1981), laminin (Zahedi 1997), type IV collagen (Zahedi 1996), and the glycosaminoglycans and proteoglycans (Hamazaki 1987). The significance of TAP is currently unknown although a range of glomerular diseases have disrupted or abnormal distribution patterns of TAP in the GBM (Dyck *et al* 1980a; Melvin *et al* 1986). Several possible functions of TAP have been suggested including the protection of sensitive regions or structures from proteolysis, aiding or facilitating the formation of protein complexes from other matrix components, or acting as a focal point for the recognition and binding of other ligands associated with the tissues.

SAP is the only serum protein found to bind deoxyribonucleic acid (DNA), native long chromatin, and H1-stripped chromatin under physiological conditions (Butler *et al* 1990; Pepys & Butler 1987). Binding to chromatin occurs with sufficient affinity to displace the H1 histones, which are required for condensation of higher order chromatin, thereby increasing the solubility of long chromatin that is otherwise profoundly insoluble in buffers of physiological ionic strength. The optimum solubility occurs at an SAP:chromatin (w/w) ratio of between 1:1 and 1:2 (Butler *et al* 1990). The presence of SAP has also been demonstrated to significantly reduce the rate of chromatin degradation both *in vivo* and *in vitro* (Bickerstaff *et al* 1999). SAP also binds apoptotic cells bearing chromatin fragments and the nuclear components released by necrosis *in vivo* (Breathnach *et al* 1989; Hintner *et al* 1988). Furthermore, the liver has been found to be the main site of

circulating nucleosome uptake (Burlingame *et al* 1996; Gauthier *et al* 1996), which is also the only site of SAP catabolism *in vivo*, indicating that SAP plays an important role in the handling of chromatin and DNA from dead or dying cells. The deletion of the SAP gene in mice has demonstrated the importance of SAP binding to DNA and chromatin. These SAP knockout mice were found to spontaneously develop antinuclear antibodies and antibodies to DNA, chromatin, and histones and had a high incidence of glomerulonephritis, a phenotype resembling the human autoimmune disease systemic lupus erythematosus (SLE) (Bickerstaff *et al* 1999). Interestingly the rate of chromatin degradation was found to be normal in C1q knockout mice, indicating that the complement pathway is not involved (Botto *et al* 1998) although they were found to develop glomerulonephritis but not antibodies against DNA or chromatin. Finally the existence of small amounts of SAP-DNA complexes have been confirmed in human sera (Sørensen *et al* 2000), where the levels show a negative correlation with anti-DNA antibodies in SLE patients.

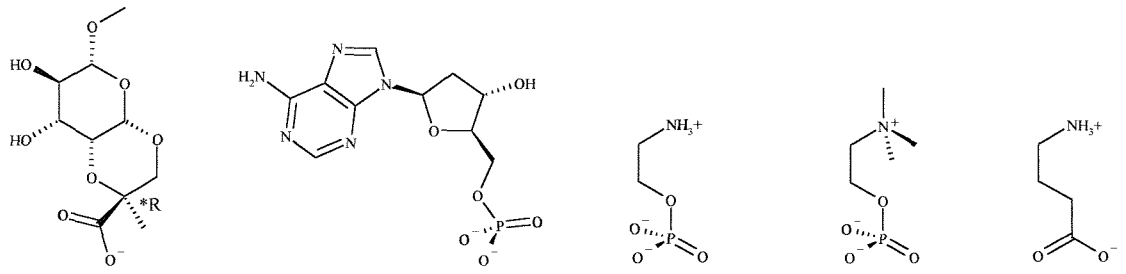
The recognition motif responsible for DNA binding is currently unknown although several interesting observations have been noted. Although binding is calcium-dependent, low MOβDG or PE concentrations were found to enhance DNA binding suggesting an allosteric effect (Pepys *et al* 1997). Model building studies found it impossible to construct a double stranded DNA-SAP complex interacting directly through the phosphate backbone and bound calcium ions of SAP, however the basic groove between protomers were found to be of sufficient size to accommodate the double stranded helix (Pepys *et al* 1997). Finally, residues 123-132 are highly conserved within the pentraxins that bind DNA, and show high sequence homology with H1/H5 and H4 histone sequences suggesting that these residues are involved in DNA binding via the minor grove (Turnell *et al* 1988).

Chemical cross-linking experiments have shown that SAP oligomers are able to bind the complement component C1q at the collagen-like region (CLR), residues 14-26

and/or 79-92 (Ying *et al* 1993). Binding is thought to be responsible for activation of the classical complement pathway by aggregated and oligomeric SAP *in vitro*. The only ligand complex currently found to activate this pathway is with the isolated H2A histone (Hicks *et al* 1992). SAP also binds the complement inhibitor C4b-binding protein (C4BP) (Evans, Jr. & Nelsestuen 1995; Schwalbe *et al* 1990) forming a complex that is reported to be a less effective inhibitor of complement (García de Frutos & Dahlbäck 1994). However, it is not clear if SAP is able to activate complement *in vivo*.

SAP is capable of binding a number of other biological ligands in the presence of calcium. The initial discovery of agarose binding led to the identification of several carbohydrates recognised by SAP, these are described in detail in section 1.3. Curiously, free SAP selectively binds immobilised CRP although the reverse does not occur (Swanson *et al* 1992). This is probably an artefact caused by structural changes during the immobilization process. Other calcium-dependent ligands include high density lipoproteins and very low density lipoproteins (Li *et al* 1998), isolated H1 and H2A histones (Hicks *et al* 1992), calumenin (Vorum *et al* 2000), dAMP, and the amino acid neurotransmitter γ-aminobutyric acid (GABA) (Pye 2000).

Structural information concerning the binding of the larger biological ligands is very limited as co-crystallisation attempts with SAP have currently been unsuccessful. However five low molecular weight ligands have been co-crystallised. These are MOβDG (Thompson 1997), dAMP (Hohenester *et al* 1997), PE (Pye 2000), PC (Thompson 2000), and GABA (Pye 2000) (see figure 1.11 and table 1.3).



MOβDG (R isomer) dAMP PE PC GABA

Figure 1.11. Ligands currently crystallised with SAP.

1.5.6 Ligand-Binding Site

All of the ligands currently identified bind with calcium-dependency, suggesting the bound calcium ions on the B-face are the primary site of interaction. The structural analyses of subsequent co-crystallisations have revealed some important details about the site involved. Carboxylate or phosphate groups were found to bridge the two calcium ions completing their coordination spheres. This orientates the rest of the ligand over the adjacent hydrophobic pocket (7.3Åx6.2Åx3.3Å) formed between the side-chains of Leu62, Tyr64, and Tyr74 (see figure 1.9), bringing the molecule to within hydrogen bonding distance of the surrounding side-chains (Asn59, Tyr64, Glu66, Tyr74, Lys79, Asp145, Ser147, and Gln148). It is this local area, through the ligation of the two bound calcium ions and the formation of surrounding hydrogen bonds, that is thought to be involved in the majority of ligand binding (see figure 1.12).

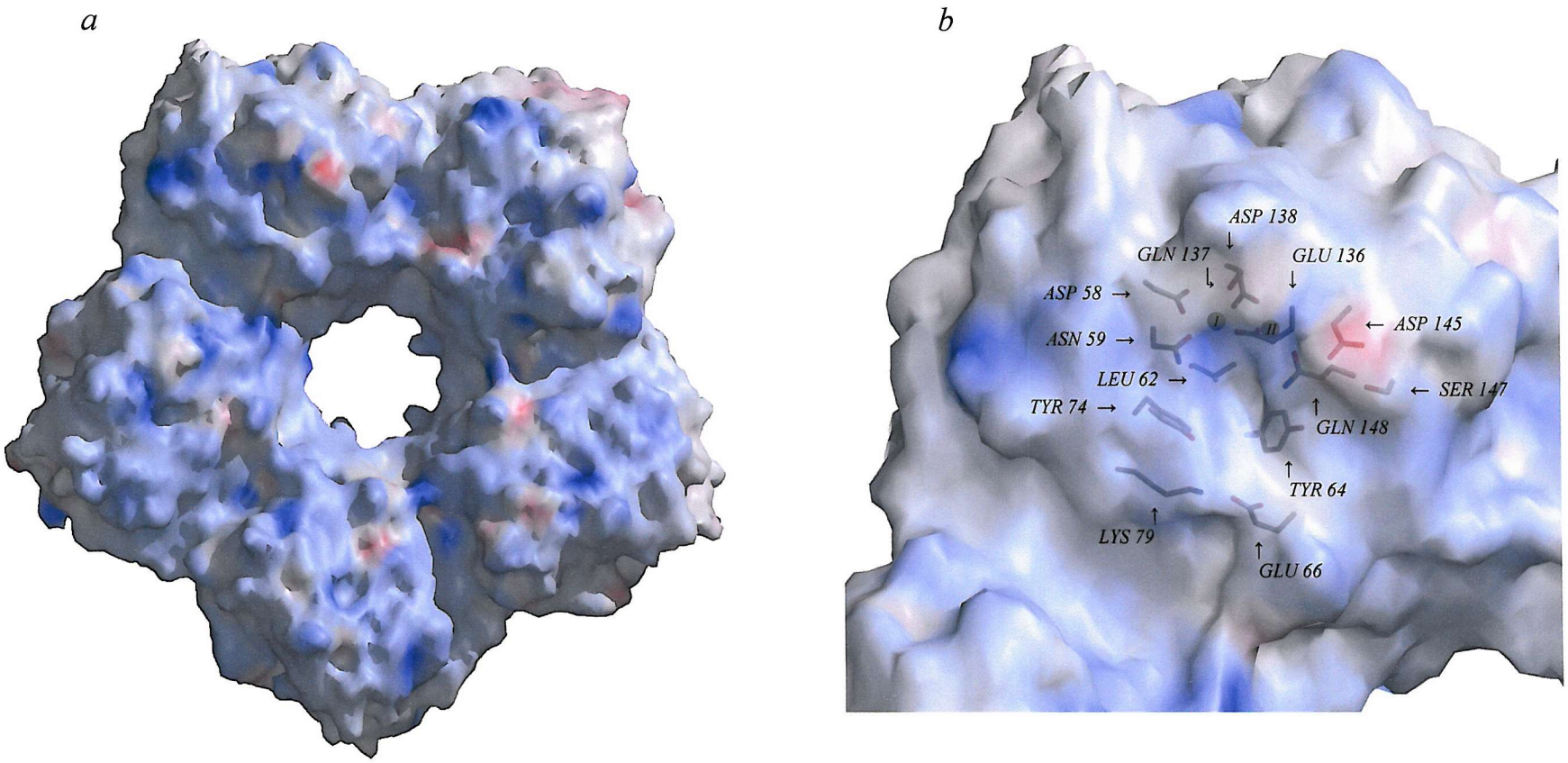


Figure 1.12. Surface representation of SAP illustrating the charge distribution over the B-face (prepared using Bobscript, Grasp, GL Render, and Raster3D). *a*, Whole pentameric face. *b*, Close-up of one subunit with residues involved in calcium and ligand binding highlighted and labelled accordingly. Blue represents a positive and red a negative charge.

Table 1.3. Summary of the processing and refinement statistics of the SAP structures solved by X-ray crystallography.

Structure	SAP - Acetate Ion	SAP - MOβDG (Pentamer Form)	SAP - MOβDG (Decamer Form)	SAP - dAMP	SAP - PE	SAP - PC	SAP - GABA
Solved By	MIR	MR	MR	MR	MR	MR	MR
Space Group	$P2_1$ Monoclinic	$P2_1$ Monoclinic	$P2_1$ Monoclinic	$P4_12_12$ Tetragonal	$P2_1$ Monoclinic	$P2_1$ Monoclinic	$P6_1$ Hexagonal
Unit Cell	a=68.9Å b=99.3Å c=96.7Å β=95.9°	a=95.8Å b=70.5Å c=103.4Å β=96.8°	a=118.6Å b=109.1Å c=120.8Å β=95.2°	a=b=190.0Å c=119.7Å	a=67.4Å b=104.9Å c=102.7Å β=95.8°	a=95.6Å b=68.4Å c=102.9Å β=96.4°	a=b=110.5Å c=213.0Å
Pentamers In Asymmetric Unit	1	1	2	1	1	1	1
Solvent Content (%)	52.4	54.8	59.8	71.0	56.6	53.2	58.3
Resolution (Å)	8 - 2.0	20 - 2.2	20 - 2.6	30 - 2.8	20 - 1.88	20 - 2.6	20 - 2.7
Completeness (%)	93.6	95.2	94.3	98.0	98.5	97.4	85.0
R-merge (%)	7.7	5.6	6.2	8.5	9.6	6.3	11.8
R-factor (%)	17.9	18.6	21.9	23.2	20.1	23.8 *	22.3
R-free (%)	-	22.4	24.5	25.2	25.6	28.6 *	27.1
Reference	(Emsley *et al* 1994)	(Thompson 1997)	(Thompson 1997)	(Hohenester *et al* 1997)	(Pye 2000)	(Thompson 2000)	(Pye 2000)

* Denotes partially refined structure

1.6 Human C-Reactive Protein

CRP is an acute phase reactant where circulating levels rise rapidly via a cytokine-mediated response to tissue injury, infection, and inflammation (Gewurz *et al* 1982; Pepys 1996; Pepys & Baltz 1983). This rapid increase along with the identification of several important biological ligands and the failure to find any polymorphism or deficiency in man strongly suggest that CRP has important biological functions (Thompson *et al* 1999). CRP has been implicated in the binding and handling of potential autoantigens, specifically with respect to small nuclear ribonucleoprotein particles (snRNPs) (Du Clos 1989; Pepys *et al* 1994), and is an important defence protein against pneumococcal (Mold *et al* 1981; Szalai *et al* 1995) and *Haemophilus influenzae* infection (Weiser *et al* 1998). CRP is also able to bind damaged plasma membranes (Volanakis & Wirtz 1979), which show an increased proportion of lysophospholipids, and can then activate the classical complement pathway and possibly stimulate tissue factor leading to enhanced tissue damage and scavenging functions (Cermak *et al* 1993; Ramani *et al* 1994). In fact, the increase in circulating CRP levels are prognostic of coronary heart disease (Haverkate *et al* 1997; Koenig *et al* 1999), particularly in myocardial infarction and have been shown to substantially increase infarct size in animal experiments (Pepys 1998). CRP has also recently been proposed to contribute to thrombo-occlusive complications of atheroma (Liuzzo *et al* 1994).

The crystal structure of human CRP was originally solved in 1996 by molecular replacement using human SAP as the search model (Shrive *et al* 1996). The crystals were grown by an unconventional method of solvent diffusion through a Sepharose gel down a gradient of free calcium ions. Due to this method of crystallisation and the large number of crystals required for data collection, the structure was based upon incomplete data with

poor processing statistics and was found to be only partially calcium loaded. The crystal structure of fully calcified CRP was subsequently solved by the same method of molecular replacement (Thompson 1997; Thompson *et al* 1999), however the crystals were grown by the conventional method of vapour diffusion and the structure based upon complete data with good processing statistics from a single crystal. Although the two structures show great similarity there are some minor differences between them thought to be due to errors in the original structure. Thus, the fully calcified structure will be the basis of the native structure described here. The atomic coordinates for both the partially and fully calcified CRP structures are currently available from the Protein Data Bank, entry codes 1GNH and 1B09 respectively.

1.6.1 Protomer Structure

CRP is an oligomeric protein composed of identical 206 residue protomers (molecular weight 23,027Da) coded for by a single gene on chromosome 1 (Whitehead *et al* 1983), which is closely linked with the gene for human SAP (Mantzouranis *et al* 1985). Each protomer contains 14 β-strands, a single α-helix, and three 3_{10} helices (see table 1.5). The β-strands, labelled A to N from the N- to C-terminus, are arranged in two large antiparallel sheets of strands ACHIJLN and BDEFGKM in the form of a flattened β-barrel with jellyroll topology. The arrangement is very similar to SAP, which when superimposed shows a C_α rms fit of 1.3Å (see figure 1.13). The main structural differences occur in the loops between strands C-D (residues 43-48), E-F (68-72), and G-H (85-91) which when omitted from the C_α rms calculation improved the rms fit to 0.83Å (Thompson *et al* 1999). There are also a number of minor differences, which include the reduction of a small β-strand (strand M in SAP), the reduction of the small α-helix (145-149 in SAP) to a 3_{10} helix around residue 149, and the formation of two other 3_{10} helices around residues 44 and 188

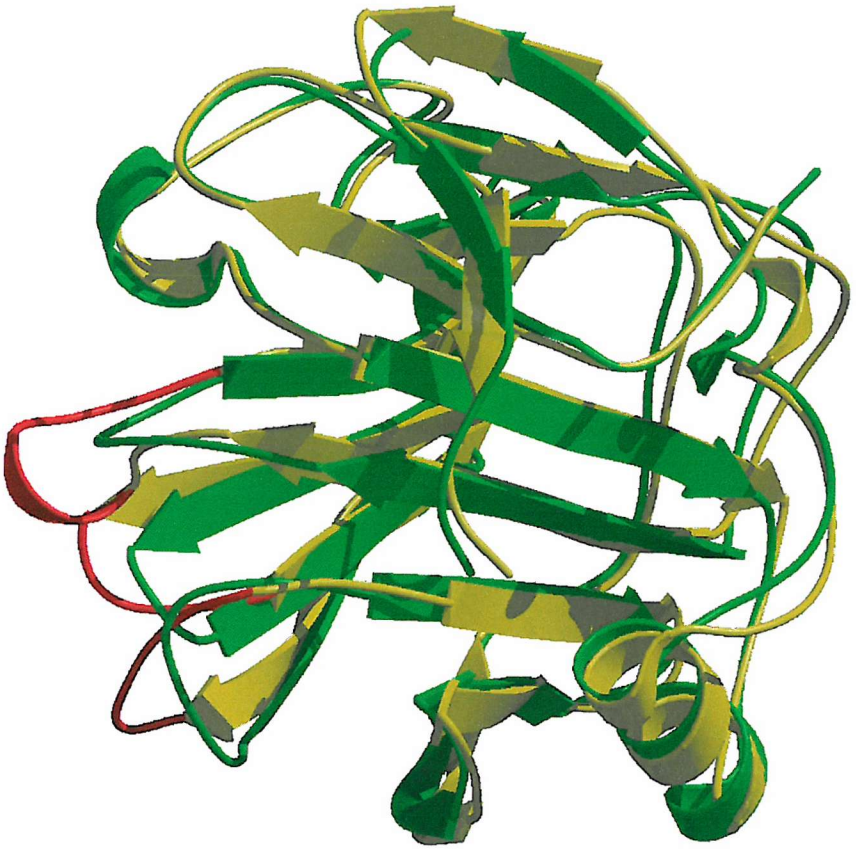


Figure 1.13. Ribbon overlay of a CRP (yellow) and SAP (green) protomer superimposed with the structural elements in the same orientation (prepared using Bobscript and Raster3D). The three most differing regions with respect to CRP are between residues 43-48, 68-72, and 85-91 and are coloured in red (Thompson *et al* 1999).

in CRP. The exposed surfaces of the β-sheets are also similarly arranged to SAP. The flatter sheet (strands ACHIJLN) supports the α-helix that lies over the disulphide bridge between Cys36-Cys97 (Cys36-Cys95 in SAP) of adjacent strands C and H respectively. There is no glycosylation site at the equivalent position to Asn32 on SAP due to the presence of alanine, but there is a marked furrow accentuated in CRP due to smaller side-chains and the reorientation of others. The opposing surface (strands BDEFGKM) contains the double calcium-binding site.

1.6.2 Calcium Binding

The double calcium-binding site is organised in a similar way to SAP, although there are a significant number of differences in calcium coordination responsible for the differences in binding affinity. Calcium I is coordinated by the side-chains of Asp60, Asn61, Glu138, Asp140, and the main chain carbonyl of Gln139. Asp60 provides only a single oxygen ligand to the calcium ion whereas the equivalent residue (Asp58) in SAP provides two. Calcium II is coordinated by the side-chains of Glu138, Asp140, Glu147, and Gln150. Glu147 extends further towards the ion than the equivalent residue (Asp145) in SAP creating an additional ligand. Residues Glu138 and Asp140 bridge the two calcium ions which are held ~3.9Å apart by a total of 4 ligands for calcium II and 5 for calcium I. The two sites are thus differentiated by only a single ligand making the sites of almost equal affinity for calcium, unlike SAP that can readily loose an ion at site II.

When the calcium sites are not occupied, as shown in the original partially calcium loaded structure (Shrive *et al* 1996), residues 140-150 are no longer held at the calcium site but protrude out away from the body of the protein. This opens the loop to proteolytic attack with α-chymotrypsin, resulting in the loss of calcium binding and calcium-dependent binding reactivities (Kinoshita *et al* 1989).

1.6.3 Oligomeric Structure

CRP forms stable oligomeric structures similar to that of SAP where five protomers are arranged in a ring with radial symmetry creating two separate and distinct faces. The major difference arises from sequence and structural deviations that alters the bonding pattern between subunits. Whilst the β-sheets are nearly perpendicular to the radial symmetry in SAP, each subunit is rotated 22° towards the 5-fold axis in CRP (see figure 1.14) (Thompson *et al* 1999). This effectively moves the helices 5Å towards the centre of mass on the A-face, the double calcium-binding sites an equivalent distance away on the opposite B-face, and orientates the 3 most differing loops with respect to SAP and the furrow towards the central pore.

CRP has not been shown to form stable decamers in solution, and to date no ligand-induced decamers have currently been identified.

1.6.4 Aggregation

Unlike SAP, human CRP does not aggregate in the presence of free calcium ions (Baltz *et al* 1982). Although the carboxyl group thought to be involved in the autoaggregation of SAP (Glu167) is maintained, Asp169 in CRP, the shorter side-chain and 22° subunit rotation prevents it from interacting with the bound calcium ions of an adjacent B-face. CRP does however precipitate in the absence of calcium, presumably due to the instability caused by the protruding loops. This was the principle behind the original method of crystallisation (Pepys 2000).

Figure 1.14. Ribbon overlay of a CRP (yellow) and SAP (green) protomer showing the difference in orientation with respect to the five-fold axis of the pentamer (Thompson *et al* 1999) (prepared using Bobscript and Raster3D).

1.6.5 Ligand Binding

CRP has been shown to bind a large number of biological ligands with calcium dependency. CRP was originally discovered due to its ability to bind pneumococcal C-polysaccharide (Tillet & Francis, Jr. 1930). The C-polysaccharide was found to contain appreciable amounts of PC (Tomasz 1967), which is able to inhibit polysaccharide binding better than any other component (Volanakis & Kaplan 1971). Since then PC has been identified as an important biological ligand found in the phospholipids of cell membranes, plasma lipoproteins, and in the complex polysaccharides of some bacteria and plant fungi. In addition to the PC containing ligands, CRP has been shown to bind galactose containing polysaccharides (Soelter & Uhlenbruck 1986), the complement component C1q (Jiang *et al* 1992), snRNPs under physiological conditions (Du Clos 1989; Pepys *et al* 1994), chromatin (Robey *et al* 1984), and PE which binds with an affinity slightly lower than that of PC (K_d 1-2x10^{-5}M (Anderson *et al* 1978)).

The binding of human CRP to chromatin is different to that of SAP, occurring only under low ionic strength conditions *in vitro* (Pepys & Butler 1987). The physiological significance of this is a subject of much debate, although the inability of CRP to bind chromatin under physiological conditions indicates that this is unlikely to occur *in vivo*. The binding of chromatin is mediated through the histones and has been located to histone H1, H2A and to a lesser extent H2B (Du Clos *et al* 1988; Du Clos *et al* 1991). Unlike SAP, CRP is unable to bind DNA or H1 stripped chromatin and has no effect upon SAP binding, although SAP partially inhibits CRP binding to chromatin and H1 histones (Hicks *et al* 1992). C1q binding occurs at the same CLR region as SAP and has been implicated in the activation of the classical complement pathway (Jiang *et al* 1992; Jiang *et al* 1991). *In vitro* studies on the activation of complement have shown that chemically aggregated or

complexed CRP is able to activate the classical complement pathway (Kaplan & Volanakis 1974).

Of the ligands currently identified only PC has currently been co-crystallised with CRP (Thompson 1997; Thompson *et al* 1999).

1.6.6 Ligand-Binding Site

Ligand binding is calcium dependent and like SAP occurs directly through the bound calcium ions located on the B-face. The smaller side-chains of Thr76 and Glu81 (Tyr74 and Lys79 respectively in SAP) and the main chain shift in CRP significantly alter the disposition of potential hydrogen bonding side-chains and create a much larger hydrophobic pocket between the side-chains of Leu64, Phe66, and Thr76. The difference in conformation and location of the site upon the B-face (due to the 22° subunit rotation) is responsible for the different binding properties of CRP and SAP. The most distinct differences are in MOβDG, amyloid fibril, and PC binding where CRP is only able to bind the latter with high affinity.

1.7 Objectives

This work continues the long running collaboration between the groups of Professor Mark Pepys and Professor Steve Wood, and aims to study the molecular basis of a new potential therapeutic treatment for human amyloidosis and the effect of SAP upon fibril formation through the interaction with protein folding intermediates.

Ro 63-8695 is a potent non-toxic inhibitor of SAP and the first compound shown to inhibit and reverse the binding of SAP to amyloid deposits *in vivo*. Ro 63-8695 is currently undergoing human clinical trials under the supervision of Professor Mark Pepys at the

Royal Free Hospital in London. Chapters 3 and 4 describe the structure determinations of SAP with bound Ro 63-8695 and components of it using X-ray crystallography, and discusses the novel binding and structural significance of Ro 63-8695 in relation to the biological and physiological effects of administering the compound.

The selective binding of SAP to the abnormal fibril conformation and the initial reduction in fibril formation suggest that SAP has chaperone like functions. Chapter 5 describes the effects of SAP upon protein folding intermediates via the use of a well-characterised refolding assay, and discusses the implications and the novel site involved.

1.8 References

al Mutlaq, H., Wheeler, J., Robertson, H., Watchorn, C., Morley, A. R. (1993) *Histochem. J.* **25**, 219-227.

Amatayakul-Chantler, S., Dwek, R. A., Tennent, G. A., Pepys, M. B., Rademacher, T. W. (1993) *Eur. J. Biochem.* **214**, 99-110.

Andersen, O., Ravn, K. V., Sørensen, I. J., Jonson, G., Holm Nielsen, E., Svehag, S. E. (1997) *Scand. J. Immunol.* **46**, 331-337.

Anderson, J. K., Stroud, R. M., Volanakis, J. E. (1978) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **37,** 1495.

Arason, G. J. (1996) *Fish Shellfish Immunol.* **6**, 277-289.

Ashall, F. & Goate, A. M. (1994) *Trends Biochem. Sci.* **19**, 42-46.

Ashton, A. W., Boehm, M. K., Gallimore, J. R., Pepys, M. B., Perkins, S. J. (1997) *J. Mol. Biol.* **272**, 408-422.

Assimeh, S. N. & Painter, R. H. (1975a) *J. Immunol.* **115**, 482-487.

Assimeh, S. N. & Painter, R. H. (1975b) *J. Immunol.* **115**, 488-494.

Baltz, M. L., de Beer, F. C., Feinstein, A., Pepys, M. B. (1982) *Biochim. Biophys. Acta* **701,** 229-236.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E. (2000) *Nucleic Acids Res.* **28**, 235-242.

Bharadwaj, D., Mold, C., Markham, E., Du Clos, T. W. (2001) *J. Immunol.* **166**, 6735-6741.

Bickerstaff, M. C. M., Botto, M., Hutchinson, W. L., Herbert, J., Tennent, G. A., Bybee, A., Mitchell, D. A., Cook, H. T., Butler, P. J. G., Walport, M. J., Pepys, M. B. (1999) *Nat. Med.* **5**, 694-697.

Bladen, H. A., Nylen, M. U., Glenner, G. G. (1966) *J. Ultrastruct. Res.* **14**, 449-459.

Blake, C. & Serpell, L. (1996) *Structure* **4**, 989-998.

Blake, C. C. F., Geisow, M. J., Oatley, S. J., Rerat, B., Rerat, C. (1978) *J. Mol. Biol.* **121**, 339-356.

Bonar, L., Cohen, A. S., Skinner, M. M. (1967) *Proc. Soc. Exp. Biol. Med.* **131**, 1373-1375.

Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E., Hawkins, P. N., Dobson, C. M., Radford, S. E., Blake, C. C., Pepys, M. B. (1997) *Nature* **385**, 787-793.

Bottazzi, B., Vouret-Craviari, V., Bastone, A., De Gioia, L., Matteucci, C., Peri, G., Spreafico, F., Pausa, M., D'Ettorre, C., Gianazza, E., Tagliabue, A., Salmona, M., Tedesco, F., Introna, M., Mantovani, A. (1997) *J. Biol. Chem.* **272**, 32817-32823.

Botto, M., Dell'Agnola, C., Bygrave, A. E., Thompson, E. M., Cook, H. T., Petry, F., Loos, M., Pandolfi, P. P., Walport, M. J. (1998) *Nat. Genet.* **19**, 56-59.

Botto, M., Hawkins, P. N., Bickerstaff, M. C. M., Herbert, J., Bygrave, A. E., McBride, A., Hutchinson, W. L., Tennent, G. A., Walport, M. J., Pepys, M. B. (1997) *Nat. Med.* **3**, 855-859.

Breathnach, S. M., Kofler, H., Sepp, N., Ashworth, J., Woodrow, D., Pepys, M. B., Hintner, H. (1989) *J. Exp. Med.* **170,** 1433-1438.

Burlingame, R. W., Volzer, M. A., Harris, J., Du Clos, T. W. (1996) *J. Immunol.* **156**, 4783-4788.

Bush, A. I., Pettingell, W. H., Multhaup, G., Paradis, M., Vonsattel, J. P., Gusella, J. F., Beyreuther, K., Masters, C. L., Tanzi, R. E. (1994) *Science* **265**, 1464-1467.

Butler, P. J. G., Tennent, G. A., Pepys, M. B. (1990) *J. Exp. Med.* **172**, 13-18.

Canet, D., Sunde, M., Last, A. M., Miranker, A., Spencer, A., Robinson, C. V., Dobson, C. M. (1999) *Biochemistry* **38**, 6419-6427.

Cathcart, E. S., Shirahama, T., Cohen, A. S. (1967) *Biochim. Biophys. Acta* **147**, 392-393.

Cermak, J., Key, N. S., Bach, R. R., Balla, J., Jacob, H. S., Vercellotti, G. M. (1993) *Blood* **82**, 513-520.

Coe, J. E., Margossian, S. S., Slayter, H. M. S., Sogn, J. A. (1981) *J. Exp. Med.* **153**, 977-991.

Coe, J. E. & Ross, M. J. (1985) *J. Clin. Invest.* **76**, 66-74.

Cohen, A. S. & Calkins, E. (1959) *Nature* **183**, 1202-1203.

Cohen, A. S., Shirahama, T., Skinner, M. (1982) in *Electron Microsopy of Proteins.* **Vol 3**. Harris, I. R. (ed) Academic Press, London, 165-205.

Creighton, T. E. (1996) in *Mechanisms of Protein Folding*. Pain, R. H. (ed) Oxford University Press Inc., New York, 1-25.

de Beer, F. C., Baltz, M. L., Holford, S., Feinstein, A., Pepys, M. B. (1981) *J. Exp. Med.* **154**, 1134-1139.

de Beer, F. C. & Pepys, M. B. (1982) *J. Immunol. Methods* **50**, 17-31.

Du Clos, T. W. (1989) *J. Immunol.* **143,** 2553-2559.

Du Clos, T. W., Marnell, L., Zlock, L. R., Burlingame, R. W. (1991) *J. Immunol.* **146**, 1220-1225.

Du Clos, T. W., Zlock, L. T., Rubin, R. L. (1988) *J. Immunol.* **141**, 4266-4270.

Dyck, R. F., Evans, D. J., Lockwood, C. M., Rees, A. J., Turner, D., Pepys, M. B. (1980a) *Lancet* **2**, 606-609.

Dyck, R. F., Lockwood, C. M., Kershaw, M., McHugh, N., Duance, V. C., Baltz, M. L., Pepys, M. B. (1980b) *J. Exp. Med.* **152**, 1162-1174.

Eanes, E. D. & Glenner, G. G. (1968) *J. Histochem. Cytochem.* **16**, 673-677.

Emsley, J. (1994) *PhD Thesis University of London.*

Emsley, J., White, H. E., O'Hara, B. P., Oliva, G., Srinivasan, N., Tickle, I. J., Blundell, T. L., Pepys, M. B., Wood, S. P. (1994) *Nature* **367**, 338-345.

Esnouf, R. A. (1997) *J. Mol. Graph. Model.* **15**, 132-134.

Esser, L. (2001) *Personal Communication.*

Evans, S. V. (1993) *J. Mol. Graph.* **11**, 134-138.

Evans, T. C., Jr. & Nelsestuen, G. L. (1995) *Biochemistry* **34**, 10440-10447.

Fraser, P. E., Nguyen, J. T., Chin, D. T., Kirschner, D. A. (1992) *J. Neurochem.* **59**, 1531-1540.

García de Frutos, P. & Dahlbäck, B. (1994) *J. Immunol.* **152**, 2430-2437.

Gauthier, V. J., Tyler, L. N., Mannik, M. (1996) *J. Immunol.* **156**, 1151-1156.

Geddes, A. J., Parker, K. D., Atkins, E. D. T., Beighton, E. (1968) *J. Mol. Biol.* **32**, 343-358.

Gewurz, H., Mold, C., Siegel, J., Fiedel, B. (1982) *Adv. Intern. Med.* **27**, 345-373.

Gewurz, H., Zhang, X. H., Lint, T. F. (1995) *Curr. Opin. Immunol.* **7**, 54-64.

Gillmore, J. D., Hawkins, P. N., Pepys, M. B. (1997) *Br. J. Haematol.* **99**, 245-256.

Glenner, G. G. (1980a) *N. Engl. J. Med.* **302**, 1283-1292.

Glenner, G. G. (1980b) *N. Engl. J. Med.* **302**, 1333-1343.

Glenner, G. G. & Bladen, H. A. (1966) *Science* **154**, 271-272.

Gotschlich, E. C., Lui, T. Y., Oliveira, E. (1982) *Ann. N. Y. Acad. Sci.* **389**, 163-171.

Gupta-Bansal, R. & Brunden, K. R. (1998) *J. Neurochem.* **70**, 292-298.

Gupta-Bansal, R., Frederickson, R. C. A., Brunden, K. R. (1995) *J. Biol. Chem.* **270**, 18666-18671.

Hamazaki, H. (1987) *J. Biol. Chem.* **262**, 1456-1460.

Hamazaki, H. (1989) *Biochim. Biophys. Acta* **998**, 231-235.

Haupt, H., Heimburger, N., Kranz, T., Baudner, S. (1972) *Hoppe-Seyler's Z physiol. Chem.* **353,** 1841-1849.

Haverkate, F., Thompson, S. G., Pyke, S. D. M., Gallimore, J. R., Pepys, M. B. (1997) *Lancet* **349**, 462-466.

Hawkins, P. N. (1994) *Clin. Sci.* **87**, 289-295.

Hawkins, P. N., Lavender, J. P., Pepys, M. B. (1990a) *N. Engl. J. Med.* **323**, 508-513.

Hawkins, P. N., Myers, M. J., Lavender, J. P., Pepys, M. B. (1988) *Lancet* **1**, 1413-1418.

Hawkins, P. N., Tennent, G. A., Woo, P., Pepys, M. B. (1991) *Clin. Exp. Immunol.* **84**, 308-316.

Hawkins, P. N., Wootton, R., Pepys, M. B. (1990b) *J. Clin. Invest.* **86**, 1862-1869.

Hicks, P. S., Saunero-Nava, L., Du Clos, T. W., Mold, C. (1992) *J. Immunol.* **149**, 3689-3694.

Hind, C. R. K., Collins, P. M., Baltz, M. L., Pepys, M. B. (1985) *Biochem. J.* **225**, 107-111.

Hind, C. R. K., Collins, P. M., Caspi, D., Baltz, M. L., Pepys, M. B. (1984a) *Lancet* **2**, 376-378.

Hind, C. R. K., Collins, P. M., Renn, D., Cook, R. B., Caspi, D., Baltz, M. L., Pepys, M. B. (1984b) *J. Exp. Med.* **159,** 1058-1069.

Hintner, H., Booker, J., Ashworth, J., Aubock, J., Pepys, M. B., Breathnach, S. M. (1988) *J. Invest. Dermatol.* **91**, 22-28.

Hohenester, E., Hutchinson, W. L., Pepys, M. B., Wood, S. P. (1997) *J. Mol. Biol.* **269**, 570-578.

Howlett, D. R., George, A. R., Owen, D. E., Ward, R. V., Markwell, R. E. (1999a) *Biochem. J.* **343**, 419-423.

Howlett, D. R., Perry, A. E., Godfrey, F., Swatton, J. E., Jennings, K. H., Spitzfaden, C., Wadsworth, H., Wood, S. J., Markwell, R. E. (1999b) *Biochem. J.* **340**, 283-289.

Hutchinson, W. L., Hohenester, E., Pepys, M. B. (2000) *Mol. Med.* **6**, 482-493.

Janciauskiene, S., García de Frutos, P., Carlemalm, E., Dahlbäck, B., Eriksson, S. (1995) *J. Biol. Chem.* **270**, 26041-26044.

Jarrett, J. T. & Lansbury, P. T., Jr. (1993) *Cell* **73**, 1055-1058.

Jiang, H., Robey, F. A., Gewurz, H. (1992) *J. Exp. Med.* **175**, 1373-1379.

Jiang, H. X., Siegel, J. N., Gewurz, H. (1991) *J. Immunol.* **146**, 2324-2330.

Jiménez, J. L., Guijarro, J. I., Orlova, E., Zurdo, J., Dobson, C. M., Sunde, M., Saibil, H. R. (1999) *EMBO J.* **18**, 815-821.

Johan, K., Westermark, G., Engström, U., Gustavsson, Å., Hultman, P., Westermark, P. (1998) *Proc. Nat. Acad. Sci. (USA)* **95**, 2558-2563.

Kaplan, M. H. & Volanakis, J. E. (1974) *J. Immunol.* **112**, 2135-2147.

Kelly, J. W. (1997) *Structure* **5**, 595-600.

Kinoshita, C. M., Gewurz, A. T., Siegel, J. N., Ying, S. C., Hugli, T. E., Coe, J. E., Gupta, R. K., Huckman, R., Gewurz, H. (1992) *Protein Sci.* **1**, 700-709.

Kinoshita, C. M., Ying, S. C., Hugli, T. E., Siegel, J. N., Potempa, L. A., Jiang, H. J., Houghten, R. A., Gewurz, H. (1989) *Biochemistry* **28**, 9840-9848.

Kisilevsky, R., Lemieux, L. J., Fraser, P. E., Kong, X. Q., Hultin, P. G., Szarek, W. A. (1995) *Nat. Med.* **1**, 143-148.

Kisilevsky, R. & Boudreau, L. (1983) *Lab. Invest.* **48**, 53-59.

Kisilevsky, R. & Fraser, P. (1996) in *The Nature and Origins of Amyloid Fibrils*. (Ciba Foundation Symposium 199) John Wiley & Sons, Chichester, 58-72.

Koenig, W., Sund, M., Frohlich, M., Fischer, H. G., Lowel, H., Doring, A., Hutchinson, W. L., Pepys, M. B. (1999) *Circulation* **99**, 237-242.

Kolatkar, A. R. & Weis, W. I. (1996) *J. Biol. Chem.* **271**, 6679-6685.

Kubak, B. M., Potempa, L. A., Anderson, B., Mahklouf, S., Venegas, M., Gewurz, H., Gewurz, A. T. (1988) *Mol. Immunol.* **25**, 851-858.

Kula, R. W., Engel, W. K., Line, B. R. (1977) *Lancet* **1**, 92-93.

Kuner, P., Bohrmann, B., Tjernberg, L. O., Näslund, J., Huber, G., Celenk, S., Grüninger-Leitch, F., Richards, J. G., Jakob-Rœtne, R., Kemp, J. A., Nordstedt, C. (2000) *J. Biol. Chem.* **275**, 1673-1678.

Lee, G. W., Goodman, A. R., Lee, T. H., Vilcek, J. (1994) *J. Immunol.* **153**, 3700-3707.

Li, X. A., Hatanaka, K., Guo, L., Kitamura, Y., Yamamoto, A. (1994) *Biochim. Biophys. Acta* **1201**, 143-148.

Li, X. A., Yutani, C., Shimokado, K. (1998) *Biochem. Biophys. Res. Commun.* **244**, 249-252.

Liuzzo, G., Biasucci, L. M., Gallimore, J. R., Grillo, R. L., Rebuzzi, A. G., Pepys, M. B., Maseri, A. (1994) *N. Engl. J. Med.* **331**, 417-424.

Loris, R., Hamelryck, T., Bouckaert, J., Wyns, L. (1998) *Biochim. Biophys. Acta* **1383**, 9-36.

Lovat, L. B. (1997) *PhD Thesis University of London.*

Loveless, R. W., Floyd-O'Sullivan, G., Raynes, J. G., Yuen, C. T., Feizi, T. (1992) *EMBO J.* **11**, 813-819.

Mantzouranis, E. C., Dowton, S. B., Whitehead, A. S., Edge, M. D., Bruns, G. A., Colten, H. R. (1985) *J. Biol. Chem.* **260**, 7752-7756.

McCubbin, W. D., Kay, C. M., Narindrasorasak, S., Kisilevsky, R. (1988) *Biochem. J.* **256**, 775-783.

McCutchen, S. L., Colon, W., Kelly, J. W. (1993) *Biochemistry* **32**, 12119-12127.

McCutchen, S. L., Lai, Z., Miroy, G. J., Kelly, J. W., Colon, W. (1995) *Biochemistry* **34**, 13527-13536.

Melvin, T., Kim, Y., Michael, A. F. (1986) *Am. J. Path.* **125**, 460-464.

Merritt, E. A. & Murphy, M. E. P. (1994) *Acta Cryst.* **D50**, 869-873.

Mold, C., Nakayama, S., Holzer, T. J., Gewurz, H., Du Clos, T. W. (1981) *J. Exp. Med.* **154**, 1703-1708.

Naiki, H. & Gejyo, F. (1999) *Methods Enzymol.* **309**, 305-318.

Nelson, S. R., Lyon, M., Gallagher, J. T., Johnson, E. A., Pepys, M. B. (1991a) *Biochem. J.* **275**, 67-73.

Nelson, S. R., Tennent, G. A., Sethi, D., Gower, P. E., Ballardie, F. W., Amatayakul-Chantler, S., Pepys, M. B. (1991b) *Clin. Chim. Acta* **200**, 191-200.

Nicholls, A., Sharp, K. A., Honig, B. (1991) *Proteins Struct. Funct. Genet.* **11**, 281-296.

Niewold, T. A., Hol, P. R., van Andel, A. C. J., Lutz, E. T. G., Gruys, E. (1987) *Lab. Invest.* **56**, 544-549.

Noland, T. D., Friday, B. B., Maulit, M. T., Gerton, G. L. (1994) *J. Biol. Chem.* **269**, 32607-32614.

O'Hara, B. P. (1992) *PhD Thesis University of London.*

Orpiszewski, J. & Benson, M. D. (1999) *J. Mol. Biol.* **289**, 413-428.

Osmand, A. P., Friedenson, B., Gewurz, H., Painter, R. H., Hofmann, T., Shelton, E. (1977) *Proc. Nat. Acad. Sci. (USA)* **74**, 739-743.

Painter, R. H., De Escallón, I., Massey, A., Pinteric, L., Stern, S. B. (1982) *Ann. N. Y. Acad. Sci.* **389**, 199-215.

Parkin, S., Rupp, B., Hope, H. (1996) *Acta Cryst.* **D52,** 1161-1168.

Pepys, M. B. (1986) in *Amyloidosis.* Marrink, J. & van Rijswijk, M.H. (eds) Martinus Nijhoff, Dordrecht, 43-50.

Pepys, M. B. (1996) in *Oxford Textbook of Medicine.* Edition 3. Weatherall, D. J., Ledingham, J. G. G., Warrell, D. A. (eds) Oxford University Press Inc, New York, 1527-1533.

Pepys, M. B. (1998) *Personal Communication.*

Pepys, M. B. (2000) *Personal Communication.*

Pepys, M. B. & Baltz, M. L. (1983) *Adv. Immunol.* **34**, 141-212.

Pepys, M. B., Baltz, M. L., de Beer, F. C., Dyck, R. F., Holford, S., Breathnach, S. M., Black, M. M., Tribe, C. R., Evans, D. J., Feinstein, A. (1982) *Ann. N. Y. Acad. Sci.* **389**, 286-298.

Pepys, M. B., Booth, D. R., Hutchinson, W. L., Gallimore, J. R., Collins, P. M., Hohenester, E. (1997) *Int. J. Exp. Clin. Invest.* **4**, 274-295.

Pepys, M. B., Booth, S. E., Tennent, G. A., Butler, P. J. G., Williams, D. G. (1994) *Clin. Exp. Immunol.* **97**, 152-157.

Pepys, M. B. & Butler, P. J. G. (1987) *Biochem. Biophys. Res. Commun.* **148**, 308-313.

Pepys, M. B., Dash, A. C., Fletcher, T. C., Richardson, N., Munn, E. A., Feinstein, A. (1978) *Nature* **273**, 168-170.

Pepys, M. B., Dash, A. C., Munn, E. A., Feinstein, A., Skinner, M., Cohen, A. S., Gewurz, H., Osmand, A. P., Painter, R. H. (1977) *Lancet* **1**, 1029-1031.

Pepys, M. B., Dyck, R. F., de Beer, F. C., Skinner, M., Cohen, A. S. (1979) *Clin. Exp. Immunol.* **38**, 284-293.

Pepys, M. B., Tennent, G. A., Booth, D. R., Bellotti, V., Lovat, L. B., Tan, S. Y., Persey, M. R., Hutchinson, W. L., Booth, S. E., Madhoo, S., Soutar, A. K., Hawkins, P. N., Van Zyl-Smit, R., Campistol, J. M., Fraser, P. E., Radford, S. E., Robinson, C. V., Sunde, M., Serpell, L. C., Blake, C. C. F. (1996) in *The Nature and Origins of*

Amyloid Fibrils. (Ciba Foundation Symposium 199) John Wiley & Sons, Chichester, 73-89.

Persey, M. R., Lovat, L. B., Apperley, J. F., Madhoo, S., Pepys, M. B., Hawkins, P. N. (1996) *Br. J. Rheumatol.* **35,** 12.

Pontet, M., Engler, R., Jayle, M. F. (1978) *FEBS Lett.* **88**, 172-175.

Potempa, L. A., Kubak, B. M., Gewurz, H. (1985) *J. Biol. Chem.* **260**, 12142-12147.

Prelli, F., Pras, M., Frangione, B. (1985) *J. Biol. Chem.* **260,** 12895-12898.

Puchtler, H., Sweat, F., Levine, M. (1962) *J. Histochem. Cytochem.* **10**, 355-364.

Pye, V (2000) *PhD Thesis University of London.*

Quigley, J., Misquith, S., Surolia, A., Srimal, S., Armstrong, P. (1994) *Biol. Bull.* **187**, 229-230.

Ramani, M., Khechai, F., Ollivier, V., Ternisien, C., Bridey, F., Hakim, J., de Prost, D. (1994) *FEBS Lett.* **356**, 86-88.

Reid, M. S. & Blobel, C. P. (1994) *J. Biol. Chem.* **269**, 32615-32620.

Rini, J. M. (1995) *Annu. Rev. Biophys. Biomol. Struct.* **24**, 551-577.

Robey, F. A., Jones, K. D., Tanaka, T., Lui, T. Y. (1984) *J. Biol. Chem.* **259**, 7311-7316.

Roher, A. E., Lowenson, J. D., Clarke, S., Wolkow, C., Wang, R., Cotter, R. J., Reardon, I. M., Zürcher-Neely, H. A., Heinrikson, R. L., Ball, M. J., Greenberg, B. D. (1993) *J. Biol. Chem.* **268**, 3072-3083.

Rowe, I. F., Jensson, O., Lewis, P. D., Candy, J., Tennent, G. A., Pepys, M. B. (1984) *Neuropathol. Appl. Neurobiol.* **10**, 53-61.

Rubio, N., Sharp, P. M., Rits, M., Zahedi, K., Whitehead, A. S. (1993) *J. Biochem. (Tokyo)* **113**, 277-284.

Schenk, D., Barbour, R., Dunn, W., Gordon, G., Grajeda, H., Guido, T., Hu, K., Huang, J., Johnson-Wood, K., Khan, K., Kholodenko, D., Lee, M., Liao, Z., Lieberburg, I.,

Motter, R., Mutter, L., Soriano, F., Shopp, G., Vasquez, N., Vandevert, C., Walker, S., Wogulis, M., Yednock, T., Games, D., Seubert, P. (1999) *Nature* **400**, 173-177.

Schwalbe, R. A., Dahlbäck, B., Coe, J. E., Nelsestuen, G. L. (1992) *Biochemistry* **31**, 4907-4915.

Schwalbe, R. A., Dahlbäck, B., Nelsestuen, G. L. (1990) *J. Biol. Chem.* **265**, 21749-21757.

Sebastião, M. P., Saraiva, M. J., Damas, A. M. (1998) *J. Biol. Chem.* **273**, 24715-24722.

Seery, L. T., Schoenberg, D. R., Barbaux, S., Sharp, P. M., Whitehead, A. S. (1993) *Proc. R. Soc. Lond. B Biol. Sci.* **253**, 263-270.

Serpell, L. C., Sunde, M., Fraser, P. E., Luther, P. K., Morris, E. P., Sangren, O., Lundgren, E., Blake, C. C. (1995) *J. Mol. Biol.* **254**, 113-118.

Shaffer, L. M., Dority, M. D., Gupta-Bansal, R., Frederickson, R. C. A., Younkin, S. G., Brunden, K. R. (1995) *Neurobiol. Aging* **16,** 737-745.

Shrive, A. K., Cheetham, G. M. T., Holden, D., Myles, D. A. A., Turnell, W. G., Volanakis, J. E., Pepys, M. B., Bloomer, A. C., Greenhough, T. J. (1996) *Nat. Struct. Biol.* **3**, 346-354.

Shrive, A. K., Metcalfe, A. M., Cartwright, J. R., Greenhough, T. J. (1999) *J. Mol. Biol.* **290**, 997-1008.

Sigurdsson, E. M., Scholtzova, H., Mehta, P. D., Frangione, B., Wisniewski, T. (2001) *Am. J. Path.* **159**, 439-447.

Skinner, M., Pepys, M. B., Cohen, A. S., Heller, L. M., Lian, J. B. (1980) in *Amyloid and Amyloidosis*. Glenner, G. G., Pinho e Costa, P., Falcão de Freitas, F. (eds) Excerpta Medica, Amsterdam, 384-391.

Skinner, M., Sipe, J. D., Yood, R. A., Shirahama, T., Cohen, A. S. (1982) *Ann. N. Y. Acad. Sci.* **389**, 190-198.

Snow, A. D. & White, T. N. (1989) *Neurobiol. Aging* **10**, 481-497.

Soelter, J. & Uhlenbruck, G. (1986) *Immunology* **58**, 139-144.

Soto, C., Castano, E. M., Prelli, F., Kumar, R. A., Baumann, M. (1995) *FEBS Lett.* **371**, 110-114.

Sørensen, I. J., Andersen, O., Holm Nielsen, E., Svehag, S. E. (1995) *Scand. J. Immunol.* **41**, 263-267.

Sørensen, I. J., Holm Nielsen, E., Schrøder, L., Voss, A., Horváth, L., Svehag, S. E. (2000) *J. Clin. Immunol.* **20**, 408-415.

Srinivasan, N., White, H. E., Emsley, J., Wood, S. P., Pepys, M. B., Blundell, T. L. (1994) *Structure* **2,** 1017-1027.

Stenstad, T., Magnus, J. H., Husby, G. (1994) *Biochem. J.* **303**, 663-670.

Stenstad, T., Magnus, J. H., Syse, K., Husby, G. (1993) *Clin. Exp. Immunol.* **94**, 189-195.

Sunde, M., Serpell, L. C., Bartlam, M., Fraser, P. E., Pepys, M. B., Blake, C. C. F. (1997) *J. Mol. Biol.* **273**, 729-739.

Swanson, S. J., Christner, R. B., Mortensen, R. F. (1992) *Biochim. Biophys. Acta* **1160**, 309-316.

Szalai, A. J., Briles, D. E., Volanakis, J. E. (1995) *J. Immunol.* **155**, 2557-2563.

Tan, S. Y. & Pepys, M. B. (1994) *Histopathology* **25,** 403-414.

Tennent, G. A., Baltz, M. L., Osborn, G. D., Butler, P. J., Noble, G. E., Hawkins, P. N., Pepys, M. B. (1993a) *Immunology* **80**, 645-651.

Tennent, G. A., Butler, P. J., Hutton, T., Woolfitt, A. R., Harvey, D. J., Rademacher, T. W., Pepys, M. B. (1993b) *Eur. J. Biochem.* **214**, 91-97.

Tennent, G. A., Lovat, L. B., Pepys, M. B. (1995) *Proc. Nat. Acad. Sci. (USA)* **92**, 4299-4303.

Thompson, D. (1997) *PhD Thesis University of London.*

Thompson, D. (2000) *Personal Communication.*

Thompson, D., Pepys, M. B., Wood, S. P. (1999) *Structure* **7**, 169-177.

Tillet, W. S. & Francis, T., Jr. (1930) *J. Exp. Med.* **52**, 561-571.

Tomasz, A. (1967) *Science* **157**, 694-697.

Tsui, C. C., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Barnes, C., Worley, P. F. (1996) *J. Neurosci.* **16**, 2463-2478.

Turnell, W. G., Satchwell, S. C., Travers, A. A. (1988) *FEBS Lett.* **232**, 263-268.

Urry, D. W. (1974) *Adv. Exp. Med. Biol.* **43**, 211-243.

Virchow, R. (1854) *Arch. Path. Anat.* **6,** 416.

Volanakis, J. E. & Kaplan, M. H. (1971) *Proc. Soc. Exp. Biol. Med.* **136**, 612-614.

Volanakis, J. E. & Wirtz, K. W. (1979) *Nature* **281**, 155-157.

Vorum, H., Jacobsen, C., Honoré, B. (2000) *FEBS Lett.* **465**, 129-134.

Weis, W. I., Drickamer, K., Hendrickson, W. A. (1992) *Nature* **360**, 127-134.

Weiser, J. N., Pan, N., McGowan, K. L., Musher, D., Martin, A., Richards, J. (1998) *J. Exp. Med.* **187**, 631-640.

Wheeler, W. C. & Gladstein, D. S. (1994) *J. Hered.* **85**, 417-418.

Whitehead, A. S., Bruns, G. A. P., Markham, A. F., Colten, H. R., Woods, D. E. (1983) *Science* **221**, 69-71.

Wood, S. J., Maleeff, B., Hart, T., Wetzel, R. (1996) *J. Mol. Biol.* **256**, 870-877.

Ying, S. C., Gewurz, A. T., Jiang, H., Gewurz, H. (1993) *J. Immunol.* **150**, 169-176.

Zahedi, K. (1996) *J. Biol. Chem.* **271**, 14897-14902.

Zahedi, K. (1997) *J. Biol. Chem.* **272**, 2143-2148.

CHAPTER 2

CRYSTALLOGRAPHIC METHODS

2.1 *Introduction*

X-ray crystallography is the method by which an increasing number of protein structures have been derived and involves interpreting the way X-rays are diffracted from a well-ordered single crystal. Although the practical aspects involved in crystallography are continually being developed, the theory of diffraction has remained unchanged since the early twentieth century (Rhodes 1993; Stout & Jensen 1968).

2.2 *Diffraction Theory*

Diffraction arises from the interaction of matter with waves. The waves are scattered in all directions by different features, combining in such a way that they can either reinforce or annul one another. The scale of detail of the matter must be close to the wavelength (λ) of the impinging waves. For atomic detail, the wavelength of radiation falls within the X-ray range (λ=0.1-3Å) of electromagnetic radiation. These electromagnetic waves interact with matter through their fluctuating electric field, disturbing the electrons, which absorb the X-rays and reemit them at the same frequency in all directions. Thus, X-rays are scattered by the electron clouds surrounding the atomic nucleus.

When molecules are arranged in a well-ordered array, such as a crystal, the pattern of X-ray scattering form their constituent electrons can be described as if diffracted by a three-dimensional lattice or grating (Laue 1912). The diffracted X-rays can be individually described as if reflected from planes within the crystal lattice (defined by the Miller indices (h,k,l)) by Bragg's law (Bragg 1913), represented by equation 2.1. The Bragg condition is only satisfied when; the angle of incidence (θ) and the spacing between neighbouring

planes (d_{hkl}) are such that the path difference ($2d_{hkl}\sin\theta$) of the reflected X-rays is an integral (n) of the wavelength. Hence, the reflected waves will be in phase and interfere constructively to form a strong reflection. However, where the angle of incidence does not satisfy the Bragg condition, that is to say the path difference is not an integral of λ, the emerging waves are out of phase and interfere destructively.

$$2d_{hkl} \sin \theta = n\lambda \qquad (2.1)$$

The geometry of the Bragg reflections correspond to lattice-points within reciprocal space (defined by the reciprocal space coordinates h,k,l) and is only dependent upon the lattice array, or rather the unit cell geometry of the crystal. The amplitude of each reflection is dependent upon the arrangement of atoms within the unit cell. The amplitude of each reflection is, therefore, the sum of the individual scattered waves and is best described as a Fourier summation. The computed sum for reflection h,k,l is called the structure factor F_{hkl}. Each atom's contribution towards the net magnitude and phase of F_{hkl} ($F_{hkl} = |F_{hkl}|\ e^{2\pi i\alpha_{hkl}}$) is dependent upon the atomic scattering factor (f) of the element and its location within the unit cell (defined by the fractional coordinates x, y, z); this is described for atom j in equation 2.2.

$$F_{hkl} = \sum_{j=1}^{atoms} f_{(j)}\ e^{2\pi i(hx_{(j)} + ky_{(j)} + lz_{(j)})} \qquad (2.2)$$

Alternatively, F_{hkl} can be expressed as the sum of waves scattered by the electron density (ρ) within the unit cell volume (V). This is shown in equation 2.3, where the

electron density is described as the average distribution within a given unit volume centred at (x, y, z).

$$F_{hkl} = \int_V \rho(x, y, z)\, e^{2\pi i(hx + ky + lz)}\, dV \qquad (2.3)$$

Equation 2.3 is, thus, an expression for F_{hkl} in reciprocal space from the electron density within real space, or rather the mathematical relationship between the electron density and the F_{hkl}'s is that of a Fourier transform. Therefore, the electron density can be calculated from a Fourier summation of the structure factors, as shown in equation 2.4. Because F_{hkl} is a periodic function describing a complex wave, in order to calculate $\rho(x,y,z)$ you need a description of the magnitude, frequency, and phase of each structure factor. The magnitude is proportional to the reflection intensity ($|F_{hkl}| \propto (I_{hkl})^{1/2}$ see equation 2.6) and can be directly measured using a detector or X-ray film. The frequency is that of the X-ray source, but the phase angle (α_{hkl}) cannot be directly measured and must be obtained from a different source. This is known as the phase problem!

$$\rho(x, y, z) = {}^{1}/_{V} \sum_h \sum_k \sum_l F_{hkl}\, e^{-2\pi i(hx + ky + lz)} \qquad (2.4)$$

2.3 *Practical Aspects*

Advances in X-ray sources, data collection instruments, computer power, and processing software have facilitated more accurate and higher-resolution structural

determinations for proteins. Increased speed and automation of the computer processing software has also enabled the techniques involved to become more easily available and user friendly. This section deals with some of the techniques used during this work. A brief overview of the steps involved in protein crystallography is shown in figure 2.1.

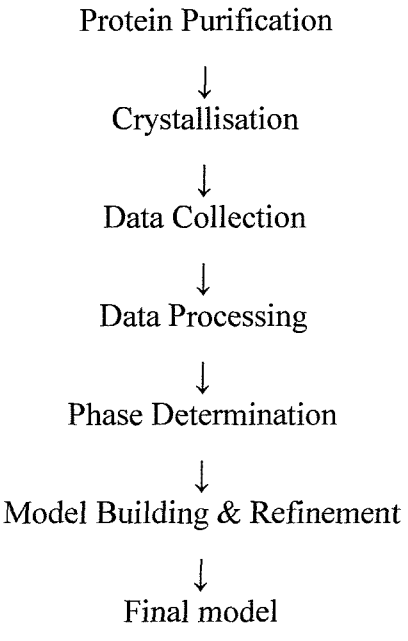


Figure 2.1. Flow diagram showing the steps involved in structural determination by X-ray crystallography, from the initial protein purification down through to the final model.

The three major rate-limiting steps in X-ray crystallography are the purification of sufficient protein, the growth of good quality crystals, and the phase determination of the recorded Bragg reflections. The various techniques involved in obtaining protein for crystallisation are beyond the scope of this work as the protein was purified and provided by Professor Mark Pepys at the Royal Free Hospital in London.

2.4 *Crystallisation*

Due to the large number of variables, locating and optimising the conditions required to achieve good quality crystals, that diffract X-rays to sufficient resolution, is trial and error and can often take a long time. Protein crystallisation can be achieved by minimizing the free energy of the system whilst maintaining the structural integrity of the molecule (McPherson 1990). The entropy of the system is lowered due to the loss of rotational and translational freedom by the formation of weak intermolecular forces between the molecules. This is accomplished by reducing the solubility of the protein until the system is forced to move towards a free energy minimum via either amorphous precipitation or the formation of an ordered crystal lattice. Supersaturation can be achieved by increasing the protein concentration to a point where there are insufficient water molecules to maintain full hydration of the protein. The solubility is also dependent upon the pH, temperature, and salt concentration (Dixon & Webb 1961), and the presence of precipitants like organic solvents, salts, and polymers (Ducruix & Giegé 1992; Green & Hughes 1955; Patel *et al* 1995). The organic solvents and polymers act by solvent exclusion with the former also lowering the dielectric constant of the solution. Salts, however, initially increase the solubility by locating between opposing charges at low concentrations (salting-in) but reduce the solubility at high concentrations by dehydrating the protein (salting-out).

The phase diagram in figure 2.2 shows the complexity of crystallisation, where changes in solution parameters can alter the solubility of the protein from unsaturated to supersaturated. When unsaturated the solubility of the protein is maintained and crystallisation does not occur. Supersaturation is subdivided into regions where

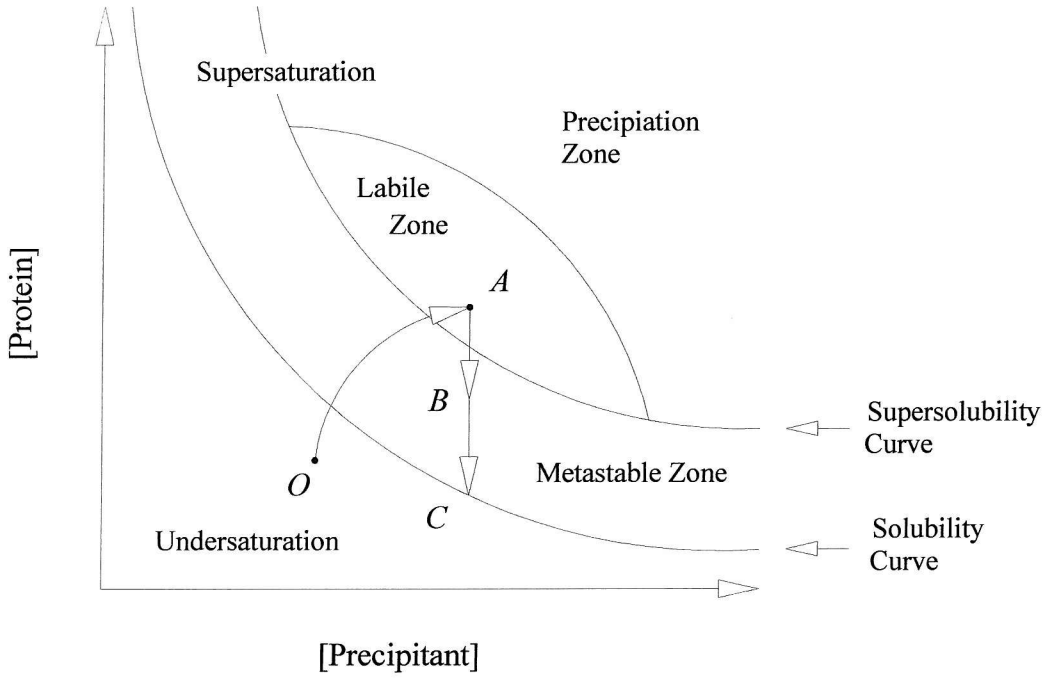


Figure 2.2. Typical two-dimensional phase diagram illustrating the effect of varying the protein and precipitant concentration upon protein solubility. The solubility curve separates the regions that support crystal growth (supersaturated) from those where crystals dissolve (understaurated). The supersolubility curve further divides the supersaturation regions where nucleation and growth compete (labile zone) and where only crystal growth occurs (metastable zone). Higher levels of supersaturation result in precipitation (precipitation zone). For an ideal vapour diffusion experiment a level of supersaturation within the metastable zone needs to be reached for nucleation to occur (O→A), during which the protein is used up and the level of saturation falls to within the metastable zone (A→B), where crystal growth is maintained until the solubility curve is reached (B→C).

precipitation (precipitation zone), crystal nucleation and growth (labile zone), and crystal growth occurs (metastable zone). Spontaneous nucleation only occurs within the labile zone and therefore needs to be entered in order to achieve crystals. More often than not this is either overshot producing an amorphous precipitate (precipitation zone) or undershot where the protein remains in solution (metastable zone). In order to grow large crystals you ideally need to separate the nucleation and growth phases by entering the labile zone for only a short period of time before entering the metastable zone (point A→B). This would facilitate limited nucleation whilst sustaining crystal growth until either the protein is depleted (point B→C) or errors occur upon the crystal surface (Kam *et al* 1978). Saturation can be controlled by several methods such as batch, micro-batch, vapour diffusion, and equilibrium dialysis (Ducruix & Giegé 1992; Weber 1997). Vapour diffusion is the most widespread method used in crystallisation, and was the only technique used within this work.

2.4.1 Vapour Diffusion

Vapour diffusion is a technique in which a solution containing the crystallisation agents (precipitant, salt, buffer) is sealed in a closed system with a small drop of protein solution (typically 2-5µl) mixed with an equal volume of the reservoir solution. The lower precipitant concentration in the drop causes it to equilibrate with the reservoir solution via vapour diffusion. This has the effect of slowly increasing the protein concentration as water molecules diffuse out of the drop. This continues until equilibrium is reached, hopefully after nucleation has been achieved (shown as point O→A in figure 2.2).

The hanging drop method of vapour diffusion can be set up using limbro plates, wherein the drop is placed on a siliconised cover glass that is then inverted and placed over a greased well containing the reservoir solution. This is a very simple and inexpensive

method of crystallisation and allows easy implementation of factorial screening (Ducruix & Giegé 1992).

2.5 Data Collection

In order to record the intensities of the Bragg reflections the crystal must be placed in the direct path of an X-ray beam between the collimator and detector. The crystal is mounted according to the temperature of collection and is attached to a goniometer, which is in turn screwed onto the goniostat. The crystal must then be centred using a TV camera so as to remain within the beam during the experiment.

The data can be collected by either the Weissenberg method (Weissenberg 1924) or precession photography (Buerger 1964) using film, with a diffractometer using a scintillation-counter detector (Abrahams 1964; Arndt & Phillips 1961), or by the oscillation method using an area detector (Arndt & Wonacott 1977). The oscillation method is now the most widely used due to the automation and speed of data collection (see figure 2.3). The intensity of each reflection is recorded upon the detector as the crystal is rotated on the goniostat by a small oscillation angle (typically 0.25-2.0°) during X-ray exposure. The detector is then scanned and erased prior to the next exposure, which continues from the previous oscillation angle. This continues until a full dataset is recorded, where the number of images required is dependent upon both the crystal symmetry and the orientation of the crystal relative to the rotation axis. In order to maximise data quality whilst minimising radiation damage the exposure time, oscillation range, oscillation angle, and crystal-detector distance all need to be optimised (Dauter 1999).

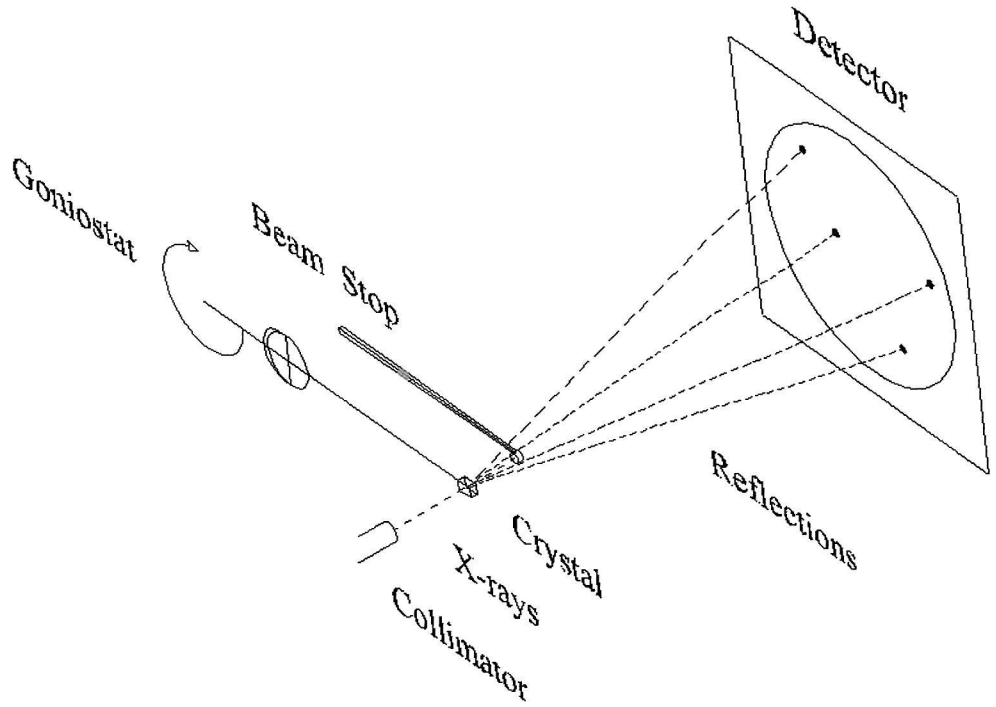


Figure 2.3. Example of the experimental setup for the oscillation method of data collection.

2.5.1 Crystal Mounting

Crystals can be mounted for data collection at either room or cryogenic temperatures. At room temperature the crystal is mounted inside a sealed capillary with a small amount of reservoir solution to prevent the crystal drying out (King 1954). This is achieved by gently drawing the crystal up the capillary and then drawing off the excess solution surrounding it using either a pulled Pasteur pipette or thin slivers of filter paper. A small amount of the reservoir solution is then drawn up from each end before the ends are sealed using heated wax. This method of crystal mounting is generally reserved for weaker

radiation sources where crystal damage is considerably lower due to the intensity of the beam, or when a suitable cryoprotectant cannot be found.

Radiation damage can be minimised by collecting data at cryogenic temperatures (typically 100K) (Garman 1999; Garman & Schneider 1997; Rodgers 1997). The crystal is first soaked into a solution of the reservoir containing a suitable small molecule cryoprotectant. After soaking, the crystal is then mounted in a small fibre loop, slightly bigger than the crystal, attached to a metal or constructed pin. The crystal and the assembly are then flash frozen by either placing in the path of a gaseous nitrogen stream or into liquid ethane (or liquid propane) before being transferred to liquid nitrogen. The crystal requires cooling so that the water contents both inside (27-65% (Matthews 1968)) and surrounding it is vitrified into a non-ordered glassy form (Petsko 1975). This prevents the formation of ice, which strongly diffracts X-rays and damages the crystal due to the 9% increase in volume (Garman & Schneider 1997). Finding a suitable cryoprotectant and the concentration required for vitrification is largely trial and error, although concentrations between 15-25% of glycerol (Garman & Mitchell 1996), ethylene glycol, low molecular weight polyethylene glycol (PEG), 2-methyl-2,4-pentanediol (MPD), or various sugars can often be used to achieve this (Garman & Schneider 1997; Rodgers 1997).

2.5.2 X-ray Sources

The two main X-ray sources used in crystallography are the rotating anode generator and the synchrotron. Rotating anode generators produce X-rays by bombarding a rapidly rotating metal anode (typically 6,000 rpm) with focused electrons emitted from a heated element. This produces a characteristic spectrum depending upon the type of metal used for the anode. In order to interpret the collected reflections a monochromatic source is required. This can be achieved by the use of a filter or a system of mirrors (Pflugrath

1992). For a copper anode, a graphite crystal monochromator can be used to divert the weaker radiation at K_β (λ=1.39Å) producing a beam of the intense radiation at K_α (λ=1.54Å).

Synchrotrons are the most powerful X-rays sources (Lindley 1999). These work by accelerating electrons (or positrons) in a ring to velocities near the speed of light using the energy from radio-frequency transmitters. The ring is maintained using powerful bending magnets that force the electrons to change direction. As this occurs the electrons radiate out synchrotron radiation. Third generation synchrotrons have better focusing mirrors and insertion devices, such as wigglers and undulators, which use strong magnetic fields to make the electrons oscillate. This produces a more intense and focused parallel beam with a tuneable wavelength between 0.3-1.8Å (Mitchell *et al* 1999). Thus, synchrotron sources allow the use of smaller crystals, shorter exposure times, provide higher resolution diffraction, and facilitate multiple wavelength anomalous dispersion or multiple isomorphous replacement phase determination.

2.5.3 Area Detectors

There are several types of detector available for recording the diffraction images. The simplest is photographic film, but automated area detectors such as image plates, charge-coupled devices (CCD), and multi-wire proportional counters have largely replaced film due their speed and ease of use. Image plates and CCD detectors are the most routinely used in protein crystallography and were also the only types used within this work.

The surface of an image plate detector consists of a barium halide phosphor (BaFBr) doped with europium (Eu^{2+}) (Amemiya 1997). During exposure to X-rays the europium is excited into a metastable state (Eu^{3+}), the quantity of which can then be read off by

stimulating the europium into emitting violet luminescence using a red He-Ne laser. The intensity of the emitted light is proportional to the quantity of absorbed photons and is detected by a photomultiplier system. The image plate is then erased by exposure to bright visible light.

The surface of a CCD detector consists of a phosphor faceplate that emits photons when hit by X-rays (Westbrook & Naday 1997). The photons are collected through an optical taper and detected on a CCD array or semiconductor chip module. These allow extremely fast image readouts with high resolution and high count rates.

2.6 Data Processing

A dataset forms a three-dimensional representation of the reflections within reciprocal space. Each diffraction image represents a different curved slice of reciprocal space in a distorted two-dimensional array. The distortion is a factor of the oscillation method of data collection, diffraction geometry, and detector characteristics. Several programs are available to compensate for these distortions, and convert the recorded intensities into structure factors within a reciprocal space coordinate system. The most widely available and most commonly used programs for processing the raw images are Mosflm (Leslie 1997), a component within the Collaborative Computational Project Number 4 (CCP4) suite (Collaborative Computational Project Number 4 1994), and Denzo and Scalepack (Otwinowski & Minor 1997) which make up the HKL suite (Minor 1997). Although Mosflm and Denzo use different procedures and algorithms they both perform the same basic steps as summarised in figure 2.4.

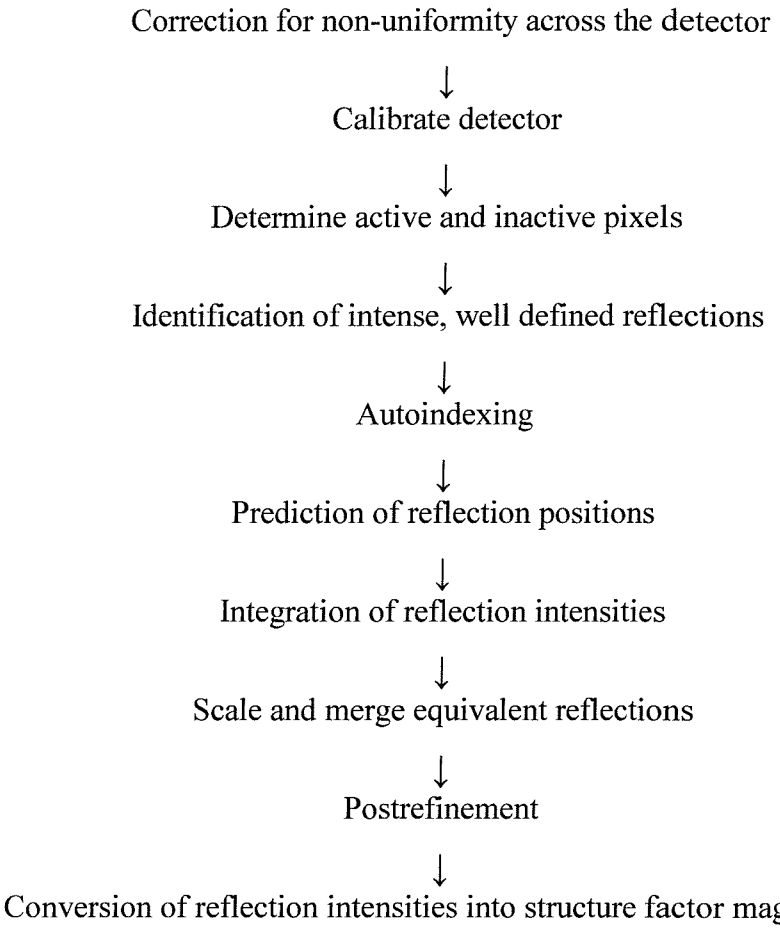


Figure 2.4. Flow diagram showing the basic steps involved in data processing.

Initially the raw images need to be corrected for non-uniformity along the surface of the detector face and calibrated for spatial distortions. The active and inactive pixels are then determined prior to the identification of intense and well-defined reflections. These reflections are used for autoindexing where they are mapped within reciprocal space and the difference vectors searched. This gives an estimate of the orientation matrix and the cell dimensions for each of the 14 Bravais lattices (Henry *et al* 1969). Once an initial estimate has been made, the reflection positions can be predicted and the matrix, cell and

detector parameters refined using the least-squares refinement process. The reflections can then be integrated by summation and/or profile fitting to produce a list of reflections, which are then scaled and merged into a unique ordered set taking into account geometric and absorption factors (see section 2.6.1). Finally, post-refinement can be carried out where the matrix, cell and detector parameters are refined over the whole dataset. The intensities can then be converted into structure factor magnitudes. The resulting file is an ordered list of the indices and structure factors magnitudes for each unique reflection within reciprocal space.

During data processing the relative intensity of symmetry related reflections are used to measure the precision of the data processing. This is quoted as a merging residual index (R) factor, or *R-merge*, and gives an estimate of the disagreement between the weighted mean intensity (I_h) and the measured intensity (I_{hi}) of reflections that according to the space group should be identical. The *R-merge(I)* is expressed as shown in equation 2.5.

$$R\text{-}merge(I) = \frac{\Sigma_h \Sigma_i \left|(I_h - I_{hi})\right|}{\Sigma_h \Sigma_i I_{hi}} \times 100 \qquad (2.5)$$

2.6.1 Data Reduction

The nature of the experiment will inherently introduce geometric and absorption factors into the data. These will have a significant effect upon the location and intensity of each measured reflection, and must therefore be corrected so an accurate quantity can be obtained for $|F_{hkl}|$. This conversion of intensities to observed structure factor magnitudes ($|F_{obs}|$) is often referred to as data reduction (Buerger 1970).

The relation between reflection intensity and $|F_{hkl}|$ can be expressed as follows:

$$I_{hkl} K = |F_{hkl}|^2 L p A \qquad (2.6)$$

Where L is the Lorentz factor, p the polarization factor, A the absorption correction factor, and K a constant comprised of a number of universal and other experimental constants.

2.7 *Solving The Phase Problem*

The phase of each reflection is dependent upon the position of the atoms within the unit cell. In order to determine the unknown structure an initial estimation of these phases is required to calculate an electron density map (equation 2.4). Direct methods, molecular replacement, isomorphous replacement, and anomalous scattering can all be used to achieve the initial phase estimation. Of these, only molecular replacement was used within this work and will therefore be the only method described.

Molecular replacement utilises the phases of an homologous model (phase model) as an initial estimate of the unknown structure. However, the phasing model requires precise orientation and positioning within the unit cell so the phases are as close to those of the unknown structure as possible, effectively superimposing the phasing model onto the unknown structure. Essentially this is a six-dimensional problem that is simplified by separating the orientation (rotation function) and positional (translation function) searches, creating two three-dimensional problems.

Several packages can be used to determine the molecular replacement solution, such as Crystallography & NMR System (CNS) (Brünger *et al* 1998), X-PLOR (Brünger 1992b), and CCP4. There are several programs within the CCP4 suite that can be used,

these include Molrep (Vagin & Teplyakov 1997) and AMoRe (Navaza 1994) which are able to perform both searches, while Polarrfn can be used to determine the self-rotation and TFFC (Driessen *et al* 1991) the translation function.

2.7.1 Rotation Function

There are two types of rotation function, the cross-rotation and self-rotation, both of which can be calculated using the Patterson function (equation 2.7) (Patterson 1935). The cross rotation (R) is used to determine the rotation matrix of the model in relation to the data, and is defined as three Eulerian angles α, β, and γ. The self-rotation determines the rotation matrix between symmetry related molecules within the asymmetric unit. This symmetry is known as non-crystallographic symmetry (NCS), and can be visualised by means of polar angles ϕ, ω, and χ. The Patterson function ($P(u,v,w)$) is used to calculate a Patterson map, which contains peaks corresponding to the intra-molecular and inter-molecular vectors between the atoms within the unit cell. This Fourier summation can be calculated from just the magnitudes as the phases are set to zero. Therefore, as all the vectors are shifted to the origin, the intra-molecular vectors of the Patterson function are dependent upon the orientation of the molecule, whilst the inter-molecular vectors are dependent upon both the orientation and position within the unit cell. Thus the rotation function is determined prior to the translation function.

$$P(u,v,w) = {}^{1}/_{V} \sum_{h} \sum_{k} \sum_{l} |F_{hkl}|^2 \cos 2\pi(hu + kv + lw) \qquad (2.7)$$

Both the cross- and self-rotation functions can be determined by calculating the agreement between the intra-molecular vectors of two Patterson functions as one is rotated

with respect to the other. The product of the two Patterson functions is large when they are in the same orientation. The quantity of agreement was originally formulated as shown in equation 2.8 (Rossmann & Blow 1962), where P_1 and P_2 are the two Patterson maps, C the matrix defining the rotation of x with respect to x′, and U the volume of integration. The cross-rotation is determined by the agreement between the two Patterson functions when one is calculated from the observed data and the other from the model, while the self-rotation is determined when both are calculated from the observed data. The fast Fourier transform (FFT) method is usually used to calculate these rotation functions where the Patterson density is approximated using spherical harmonics with in real space (Crowther 1972).

$$R(C) = \int_U P_1(x)\, P_2(x')\, dU \qquad (2.8)$$

The success of the cross-rotation function is dependent upon the choice of several parameters including the radius of integration, resolution limit, and choice of search model (Cooper 2000).

2.7.2 Translation Function

Once the model has been correctly orientated then the translation function (T) parameters can be determined. The translation function involves calculating the correlation between the inter-molecular vectors of two Patterson functions, one from data and the other from the model, as the model is moved throughout the cell. When the model is correctly positioned the Patterson functions will overlap and the correlation will be high. This was originally proposed by Crowther and Blow and is expressed in equation 2.9 (Crowther &

Blow 1967), where $P_o(u)$ represents the Patterson map calculated from the data and $P_c(u,t)$ represents the Patterson map calculated from the model translated by the vector t within real space, and V the unit cell volume.

$$T(t) = \int_V P_c(u,t)\, P_o(u)\, dV \qquad (2.9)$$

Most computer programs use a modified version of the translation function such as the T_2 function (Harada *et al* 1981). The T_2 function calculates a three-dimensional Patterson overlap excluding all the intra-molecular vectors. The translation function overlaps of T_2 can be represented by equation 2.10, where the $P_m(u)$ represents the sum of the self-Pattersons of n asymmetric units of the model, the other terms are as described for the previous equation. Several programs also use normalised structure factors or negative temperature factors to increase the signal-to-noise ratio of the translation function peaks.

$$T_2(t) = \int_V (P_o(u) - P_m(u))\, (P_c(u,t) - P_m(u))\, dV \qquad (2.10)$$

The translation vector can also be determined by testing the agreement between the data and the model, as the model is moved throughout the cell (Collaborative Computational Project Number 4 1998). The *R-factor* search calculates the agreement between the recorded ($|F_{obs}|$) and the calculated structure factors ($|F_{calc}|$) at each position using an approximate scale factor (k). When the model is in the correct location then the *R-factor* will be low (equation 2.11). Other factors such as the correlation coefficient (CC) can also be used to calculate the agreement (equation 2.12) (Vagin & Teplyakov 1997).

$$R\text{-}factor = \frac{\Sigma \left| |F_{obs}| - k|F_{calc}| \right|}{\Sigma |F_{obs}|} \times 100 \qquad (2.11)$$

$$CC = \frac{\Sigma \left| |F_{obs}|\,|F_{calc}| \right| - \left| \Sigma |F_{obs}| \right| \left| \Sigma |F_{calc}| \right|}{\left[\left| \Sigma |F_{obs}|^2 - \left| \Sigma |F_{obs}| \right|^2 \right| \left| \Sigma |F_{calc}|^2 - \left| \Sigma |F_{calc}| \right|^2 \right| \right]^{1/2}} \qquad (2.12)$$

2.8 *Model Building & Refinement*

The initial solution derived from molecular replacement will inevitably contain a large number of errors. In order to produce a model that adequately explains the experimental data, crystallographic refinement and model building needs to be carried out (Kleywegt & Jones 1997). Refinement programs are used to improve the fit of the observed and calculated structure factor amplitudes, whilst ensuring reasonable stereo-chemical properties. Each atom can be refined about the *x, y, z* coordinates, thermal parameters, and occupancy. Every structure therefore contains a large number of parameters to refine ($5n$), however, for an accurate determination the ratio of observations to variables must be high. The number of observations is dependent upon the resolution and completeness of the data, but can be augmented by the use of chemical information such as bond lengths and angles, while imposing restraints and constraints can reduce the number of parameters (Konnert 1976). Therefore the number of parameters refined can be tailored to the size of the molecule and related to the amount and quality of the experimental data.

Initial refinement is carried out with a small number of degrees of freedom (or parameters), which is gradually increased during each round of refinement as the model becomes closer to the observed data. The model can be refined by the classical method of least squares, whereby the atom positions that minimise the sum of the squares of the differences between the $|F_{obs}|$ and $|F_{calc}|$ (Φ) are selected (see equation 2.13: where each difference is weighted by w_{hkl}) (Rhodes 1993), thus progressing down towards an energy minimum. Most programs use a variation of the least squares method known as maximum likelihood, where errors are treated more vigorously leading to higher probabilities of reaching an energy minimum. By slowly increasing the number of parameters the global minimum is more likely to be achieved, however, the model can get trapped in one of the many local minima. Manual model building is performed between each round of refinement to help cross these energy barriers and converge upon the global minimum. This can also be achieved by refining the model using simulated annealing.

$$\Phi = \sum_{hkl} w_{hkl} \left||F_{obs}| - |F_{calc}|\right|^2_{hkl} \qquad (2.13)$$

The refinement process is monitored by the agreement between the observed structure factor magnitudes and those calculated from the model. This is expressed as the *R-factor* (equation 2.14) and *R-free* values (Brünger 1992a; Brünger 1993). The *R-free* value is similar to the *R-factor* but is calculated from data (5-10%) excluded from the refinement process. Both the *R-factor* and *R-free* values should decrease at similar rates during the refinement process, but if the *R-free* increases then the model is being over-fitted to the data.

$$R\text{-}factor = \frac{\Sigma \left| |F_{obs}| - |F_{calc}| \right|}{\Sigma |F_{obs}|} \times 100 \qquad (2.14)$$

2.8.1 *Crystallographic Refinement*

There are several packages available for performing all types of crystallographic refinement, such as CCP4, CNS, SHELXL (Sheldrick & Schneider 1997), and X-PLOR. Each offers a variety of refinement protocols including rigid body, simulated annealing, positional, conjugate gradient minimization, and temperature (B) factor refinement. The strategy of refinement is a controversial issue and is often a personal affair. A summary of the general procedure is described below:

Initially the model is split-up into domains or rigid-bodies where the rotation and translational position of each is refined. This is known as rigid body refinement and is often performed using low-resolution data to increase the radius of convergence. This is often followed by simulated annealing where either torsion-angle (constrained) or cartesian (restrained) molecular dynamics can be used to allow the model to move as if at high temperature (2,500-5,000K) (Brünger *et al* 1987). The model is then slowly cooled to progressively restrict the molecular movement in order to locate the preferred conformation. This hopefully lifts the model out of a local energy minimum. The energy of the system can then be minimised using positional refinement, and then finally the B-factors can be refined. The type of B-factor refinement used is dependant upon the resolution of the data. The overall B-factors can be refined when the data extends to 3.0Å, the grouped B-factors when the data extends to 2.8Å, and the individual isotropic

temperature factors if the data extends to 2.5Å. Individual anisotropic B-factor refinement should only be carried out when the data is near or better than 1.0Å resolution.

2.8.2 Model Building

Manual model building requires the use of a visualisation program, such as O (Jones *et al* 1991), Quanta (Molecular Simulations 1996), or XtalView (McRee 1999), to simultaneously display the calculated electron density maps and the molecular model. The model can then be manipulated to better fit the electron density. Typically for rebuilding $|F_{obs}|$-$|F_{calc}|$ and $2|F_{obs}|$-$|F_{calc}|$ electron density maps are calculated, unfortunately, during the initial stages of refinement the model may contain a large number of errors, which may seriously affect the quality of the maps. There are several methods available to modify the maps in order to improve their connectivity; these include NCS averaging (Bricogne 1974), solvent flattening and histogram matching (Kam *et al* 1990). These modified electron density maps can be calculated using either CNS or DM (Cowtan 1994).

2.8.2 Quality Control

During refinement the stereo-chemical properties of the model should be continually checked for areas or residues that may require model building. Validation programs like WHAT IF (Vriend 1990) and Procheck (Laskowski *et al* 1993) can be used to identify deviations from the ideal bond lengths and angles (Engh & Huber 1991; Ramachandran *et al* 1963), chirality, planarity (Hooft *et al* 1996), puckering amplitude (Cremer & Pople 1975), inter-atomic distances (Vriend & Sander 1993), and rotamer conformations.

2.9 References

Abrahams, S. C. (1964) *Acta Cryst.* **17**, 1190-1195.

Amemiya, Y. (1997) *Methods Enzymol.* **276**, 233-243.

Arndt, U. W. & Phillips, D. C. (1961) *Acta Cryst.* **14**, 807-818.

Arndt, U. W. & Wonacott, A. J. (1977) *The Rotation Method in Crystallography*. North-Holland, Amsterdam.

Bragg, W. L. (1913) *Proc. Cambr. Phil. Soc.* **17,** 43.

Bricogne, G. (1974) *Acta Cryst.* **A30**, 395-405.

Brünger, A. T. (1992a) *Nature* **355**, 472-475.

Brünger, A. T. (1992b) *X-PLOR: A System for X-ray Crystallography and NMR*. Version 3.1. Yale University Press, New Haven and London.

Brünger, A. T. (1993) *Acta Cryst.* **D49**, 24-36.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., Warren, G. L. (1998) *Acta Cryst.* **D54**, 905-921.

Brünger, A. T., Kuriyan, J., Karplus, M. (1987) *Science* **235**, 458-460.

Buerger, M. J. (1964) *The Precession Method.* Wiley, New York.

Buerger, M. J. (1970) *Contemporary Crystallography*. McGraw-Hill Inc.

Collaborative Computational Project Number 4 (1994) *Acta Cryst.* **D50**, 760-763.

Collaborative Computational Project Number 4 (1998) *The CCP4 Suite- Computer programs for protein crystallography: Overview and manual*. CCLRC Daresbury Labotatory.

Cooper, J. (2000) *Molecular Replacement*. 7th BCA Protein Crystallography Summer School, Bristol.

Cowtan, K. (1994) *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31**, 34-38.

Cremer, D. & Pople, J. A. (1975) *J. Am. Chem. Soc.* **97**, 1354-1358.

Crowther, R. A. (1972) in *The Molecular Replacement Method.* Rossman, M. G. (ed) Breach, New York, 173-178.

Crowther, R. A. & Blow, D. M. (1967) *Acta Cryst.* **23**, 544-548.

Dauter, Z. (1999) *Acta Cryst.* **D55**, 1703-1717.

Dixon, M & Webb, E. C. (1961) *Adv. Protein Chem.* **16,** 197-219.

Driessen, H. P., Bax, B., Slingsby, C., Lindley, P. F., Mahadevan, D., Moss, D. S., Tickle, I. J. (1991) *Acta Cryst.* **B47**, 987-997.

Ducruix, A. & Giegé, R. (1992) *Crystallization of nucleic Acids and Proteins: A Practical Approach.* Oxford University Press.

Engh, R. A. & Huber, R. (1991) *Acta Cryst.* **A47**, 392-400.

Garman, E. (1999) *Acta Cryst.* **D55**, 1641-1653.

Garman, E. F. & Mitchell, E. P. (1996) *J. Appl. Crys.* **29**, 584-587.

Garman, E. F. & Schneider, T. R. (1997) *J. Appl. Crys.* **30**, 211-237.

Green, A. A. & Hughes, W. L. (1955) *Methods Enzymol.* **1**, 67-90.

Harada, Y., Lifchitz, A., Berthou, J., Jolles, P. (1981) *Acta Cryst.* **A37**, 398-406.

Henry, N. F. M., Lonsdale, K., eds (1969) *International Tables for X-ray Crystallography.* **Vol 1**. Reidel NE/Kluwer Academic Publishers, Norwell, Massachusetts.

Hooft, R. W. W., Sander, C., Vriend, G. (1996) *J. Appl. Crys.* **29**, 714-716.

Jones, T. A., Zou, J.-Y., Cowan, S. W., Kjeldgaard, M. (1991) *Acta Cryst.* **A47**, 110-119.

Kam, Y. J., Zhang, K. Y. J., Main, P. (1990) *Acta Cryst.* **A46**, 377-381.

Kam, Z., Shore, H. B., Feher, G. (1978) *J. Mol. Biol.* **123**, 539-555.

King, M. V. (1954) *Acta Cryst.* **7**, 601-602.

Kleywegt, G. J. & Jones, T. A. (1997) *Methods Enzymol.* **277**, 208-230.

Konnert, J. H. (1976) *Acta Cryst.* **A32,** 614-617.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., Thornton, J. M. (1993) *J. Appl. Crys.* **26**, 283-291.

Laue, M. von (1912) *Sitzber. Math. Physik. Kl. Bayer. Akad. Wiss.*, 303.

Leslie, A. G. W. (1997) *Mosflm Users Guide*. MRC Laboratory of Molecular Biology, Cambridge.

Lindley, P. F. (1999) *Acta Cryst.* **D55,** 1654-1662.

Matthews, B. W. (1968) *J. Mol. Biol.* **33**, 491-497.

McPherson, A. (1990) *Eur. J. Biochem.* **189**, 1-23.

McRee, D. E. (1999) *J. Struct. Biol.* **125**, 156-165.

Minor, W. (1997) *The HKL Manual*. Edition 5.

Mitchell, E., Kuhn, P., Garman, E. (1999) *Structure* **7**, R111-121.

Molecular Simulations (1996) *QUANTA96 X-ray Structure Analysis User's Reference*. San Diego.

Navaza, G. (1994) *Acta Cryst.* **A50**, 157-163.

Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276**, 307-326.

Patel, S., Cudney, B., McPherson, A. (1995) *Biochem. Biophys. Res. Commun.* **207**, 819-828.

Patterson, A. L. (1935) *Z. Krist.* **A90,** 517.

Petsko, G. A. (1975) *J. Mol. Biol.* **96**, 381-392.

Pflugrath, J. W. (1992) *Curr. Opin. Struct. Biol.* **2,** 811.

Ramachandran, G. N., Ramakrishnan, C., Sasisekharan, V. (1963) *J. Mol. Biol* **7**, 95-99.

Rhodes, G. (1993) *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models*. Academic Press Inc.

Rodgers, D. W. (1997) *Methods Enzymol.* **276**, 183-203.

Rossmann, M. G. & Blow, D. M. (1962) *Acta Cryst.* **15**, 24-31.

Sheldrick, G. M. & Schneider, T. R. (1997) *Methods Enzymol.* **277**, 319-343.

Stout, G. H. & Jensen, L. H. (1968) *X-Ray Structure Determination: A Practical Guide*. Macmillan Publishing Co. Inc., New York.

Vagin, A. & Teplyakov, A. (1997) *J. Appl. Crys.* **30,** 1022-1025.

Vriend, G. (1990) *J. Mol. Graph.* **8**, 52-56.

Vriend, G. & Sander, C. (1993) *J. Appl. Crys.* **26**, 47-60.

Weber, P. C. (1997) *Methods Enzymol.* **276**, 13-22.

Weissenberg, K. (1924) *Z. Physik.* **23,** 229.

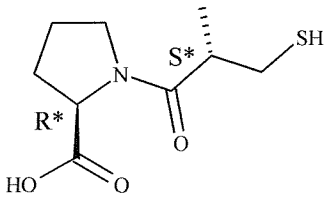
Westbrook, E. W. & Naday, I. (1997) *Methods Enzymol.* **276**, 244-268.

CHAPTER 3

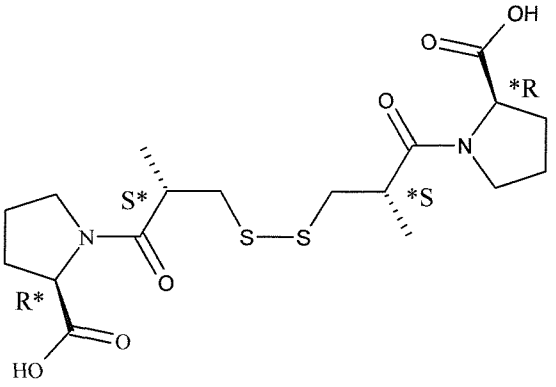
STRUCTURE DETERMINATION OF N-ACETYL-D-PROLINE & Ro 63-8695 BOUND TO SAP

3.1 Introduction

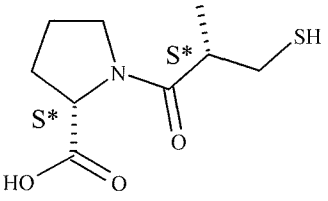
Professor Pepys and his group at the Immunological Medicine unit at the Hammersmith hospital in London (now the Centre for Amyloidosis and Acute Phase Proteins at the Royal Free hospital & University College Medical School in London) were the first to identify SAP as a potential therapeutic target for the treatment of all types of amyloidosis (see section 1.4). They subsequently developed and patented a high throughput assay for inhibitors of SAP binding to Aβ amyloid fibrils immobilised in microtitre plates (Pepys & Blundell 2000). This enabled potential inhibitors to be rapidly identified and quantified, and was used to screen the entire Roche compound library for inhibitors with IC_{50}'s comparable to that of MOβDG (IC_{50} 20mM), a galactoside previously shown to inhibit/dissociate SAP from amyloid fibrils *in vitro* (Hind *et al* 1984). One major hit was identified as (R)-1-[(S)-3-mercapto-2-methylpropionyl]-pyrrolidine-2-carboxylic acid (Ro 15-3479) which is a diastereoisomer of Captopril ((S)-1-[(S)-3-mercapto-2-methylpropionyl]-pyrrolidine-2-carboxylic acid) (see figure 3.1), a drug currently licensed for the treatment of hypertension and heart failure (Brogden *et al* 1988; Gavras 1983). The trial compound was only a crude extract and was found to contain several oxidised disulphide dimers, such as (R)-1-[(S)-3-[(S)-3-[(R)-2-carboxy-pyrrolidin-1-yl]-2-methyl-3-oxo-propyldisulphanyl)-2-methyl-propionyl]-pyrrolidine-2-carboxylic acid (Ro 63-3300) which when individually prepared had an IC_{50} over a thousand-fold lower than that of the monomeric thiol (see figure 3.1). The heightened affinity and palindromic nature of Ro 63-3300 implied the induction of a decameric complex by bridging pairs of SAP pentamers, similar to dAMP (Hohenester *et al* 1997). A screening program of potential head and linking groups was initiated to optimise the solubility, stability, and binding affinity whilst



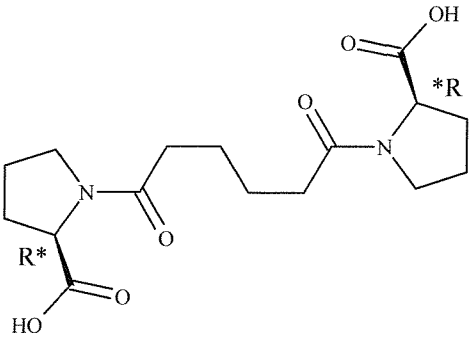
Ro15-3479
IC_{50} 2mM



Ro 63-3300
IC_{50} 1.5μM



Captopril
INACTIVE



Ro 63-8695
IC_{50} 0.95μM

Figure 3.1. Inhibitors of SAP binding to amyloid fibrils (isomer labelled according to IUPAC recommendations (Panico *et al* 1993)). The original screen hit was Ro 15-3479, an epimer of Captopril which itself was inactive. Ro 63-3300, the dimeric form of Ro 15-3479, was found to be a more potent inhibitor, and was used as the initial compound for optimisation. Ro 63-8695 was the optimal compound chosen for clinical testing.

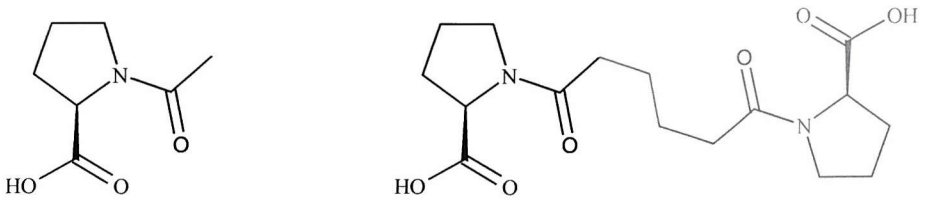
minimizing toxicity *in vivo*. Evaluation of the potential candidates identified (R)-1-[6-[(R)-2-carboxy-pyrrolidin-1-yl]-6-oxo-hexanoyl]-pyrrolidine-2-carboxylic acid (Ro 63-8695) as the optimum compound for clinical trials.

Ro 63-8695 is a bipolar palindrome reminiscent of the original hit (see figure 3.1), where two (R)-pyrrolidine-2-carboxylic acid groups are linked through a peptide bond at the amine nitrogen by a four carbon cross bridge. Ro 63-8695 shows good solubility, high affinity for SAP (IC_{50} 0.95μM), and surprisingly low toxicity in rodents. Binding is calcium-dependent and like many low molecular weight ligands is able to inhibit autoaggregation. Furthermore, as expected, Ro 63-8695 facilitates the formation of a stable decameric complex *in vitro*.

This chapter covers the structure determination of SAP crystallised in the presence of N-acetyl-D-proline ((R)-1-acetyl-pyrrolidine-2-carboxylic acid), a small molecule reminiscent of the head group of Ro 63-8695, and the comparable analysis of the Ro 63-8695 soaked into a previously crystallised form of SAP. The structure of the SAP-Ro 63-8695 complex has also been determined directly, but to lower resolution and is covered in chapter 4.

3.2 SAP Crystallised with N-Acetyl-D-Proline

Initially Ro 63-8695 was unavailable for crystallisation trials although Professor Pepys did make the structure available. The similarity of the cyclic pyrrolidine head group to a proline amino acid in the D-isomer is clearly apparent and the closest resembling molecule, commercially available, was found to be N-acetyl-D-proline (see figure 3.2). The N-acetyl-D-proline molecule contains all the important structural elements of the head group and was therefore used as a representation of Ro 63-8695.



N-acetyl-D-proline Ro 63-8695

Figure 3.2. Comparison of N-acetyl-D-proline with Ro 63-8695.

3.2.1 Crystallisation

The previously solved structures of SAP in the presence of MOβDG (Thompson 1997), PE (Pye 2000), and PC (Thompson 2000) were derived from crystals grown under similar conditions by vapour diffusion at 4°C. All produced large monoclinic ($P2_1$) crystals and diffracted X-rays to a resolution of 2.6Å or higher at room temperature (see table 1.3). These previously refined conditions were encompassed into the initial factorial screens for SAP in the presence of N-acetyl-D-proline. The screens were set up using the vapour diffusion technique of hanging drops under the following conditions: 3µl drop of protein solution containing 10-20mg/ml SAP, 4-8mM N-acetyl-D-proline, 60mM Tris pH 7.2-7.8, 84mM NaCl mixed with an equal volume of reservoir solution containing 60mM Tris pH 7.2-7.8, 84mM NaCl, 8-23% (w/v) polyethylene glycol monomethol ether (PEG MME) 550, 10mM calcium acetate, and allowed to equilibrate against 1ml of the reservoir solution at 4°C. Several crystals grew within 3 weeks, an example of which is shown in figure 3.3. The largest grew to an approximate size of 0.7x0.2x0.2mm under the following conditions: 10mg/ml SAP, 4mM N-acetyl-D-proline, 60mM Tris pH 7.2, 84mM NaCl, 15% (w/v) PEG MME 550, and 10mM calcium acetate. This crystal was consequently used for data collection.

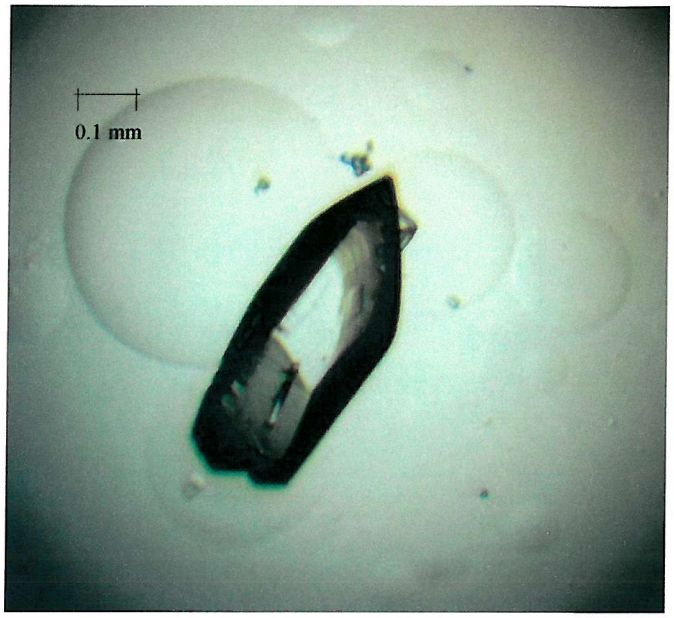


Figure 3.3. Photograph of a SAP crystal grown by vapour diffusion in the presence of N-acetyl-D-proline.

3.2.2 Data Collection & Processing

Diffraction data were collected from a single crystal at room temperature by the oscillation method using a MAR 300mm image plate. These data were collected in-house using an ENRAF NONIUS FR 591 rotating copper anode generator (λ=1.542Å). Each image was recorded during a 30-minute exposure over an oscillation range of 1°, with the detector placed 190mm away from the crystal. The dataset was recorded over a total of 97 images, an example of which is shown in figure 3.4. The data were corrected, processed, scaled, and merged using Mosflm (Leslie 1997) and other programs from the CCP4 suite (Collaborative Computational Project Number 4 1994). The crystal packing was found to be monoclinic with unit cell dimensions of a=96.1Å, b=70.8Å, c=103.6Å, and β=96.8°, and

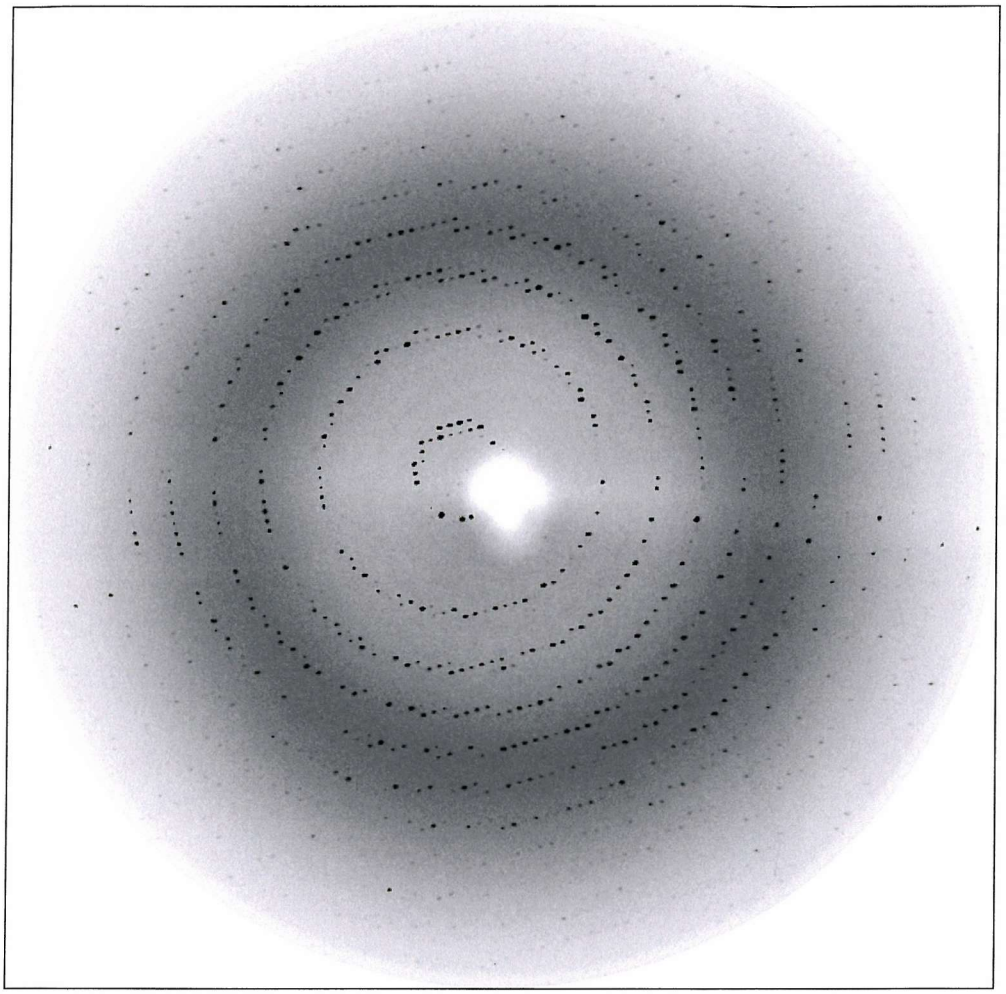


Figure 3.4. A 1° oscillation image collected in-house on a MAR 300mm image plate from a SAP-N-acetyl-D-proline crystal, recorded to a maximum resolution of 2.35Å (prepared using Mosflm).

was processed in the space group P2. Examination of the processed data revealed systematic absences in 0,*k*,0 (where $k \neq 2n$) implying a space group of $P2_1$ with a unique axis along b (see figure 3.5). The data were then reprocessed, scaled, and merged in the space group $P2_1$ to a maximum resolution of 2.4Å. The overall completeness of the data was 97.0%, with an *R-merge(I)* of 9.0%.

3.2.3 Molecular Replacement

A solvent content of 55.3% (V_M=2.61 Å^3/Da (Matthews 1968)) was calculated for the SAP-N-acetyl-D-proline crystal form assuming one SAP pentamer present in the asymmetric unit. This is consistent with isomorphous crystals grown in the presence of 37mM MOβDG (pentamer form) or 50mM PC (see table 1.3). Of these, only the coordinates for the MOβDG form were available at the time. Hence, this previously derived SAP pentamer with the calcium ions, MOβDG, and water molecules omitted was used for the initial phase estimation. A translational search was performed to correct for any minor translational shift due to the slight deviation in unit cell dimensions (a ~0.3Å, b ~0.3Å, and c ~0.2Å). In hindsight this was unnecessary, as a difference map would have clearly highlighted any such phase error, which would have easily been corrected during rigid-body refinement with the pentamer treated as the rigid-body. This would have also been more accurate than the relatively crude translation function.

The translation function was calculated using TFFC (Driessen *et al* 1991). The translational search was performed using reflections within the range of 10 to 2.4Å, and generated 4 large peaks of 49σ (46σ above the next largest peak) at symmetry related locations (see table 3.1). The minor fractional shift in peak 3 (x=0.003, y=0.000, and z=-0.001) was applied to the phasing model, and showed good sensible contacts between symmetry related molecules (see figure 3.6).

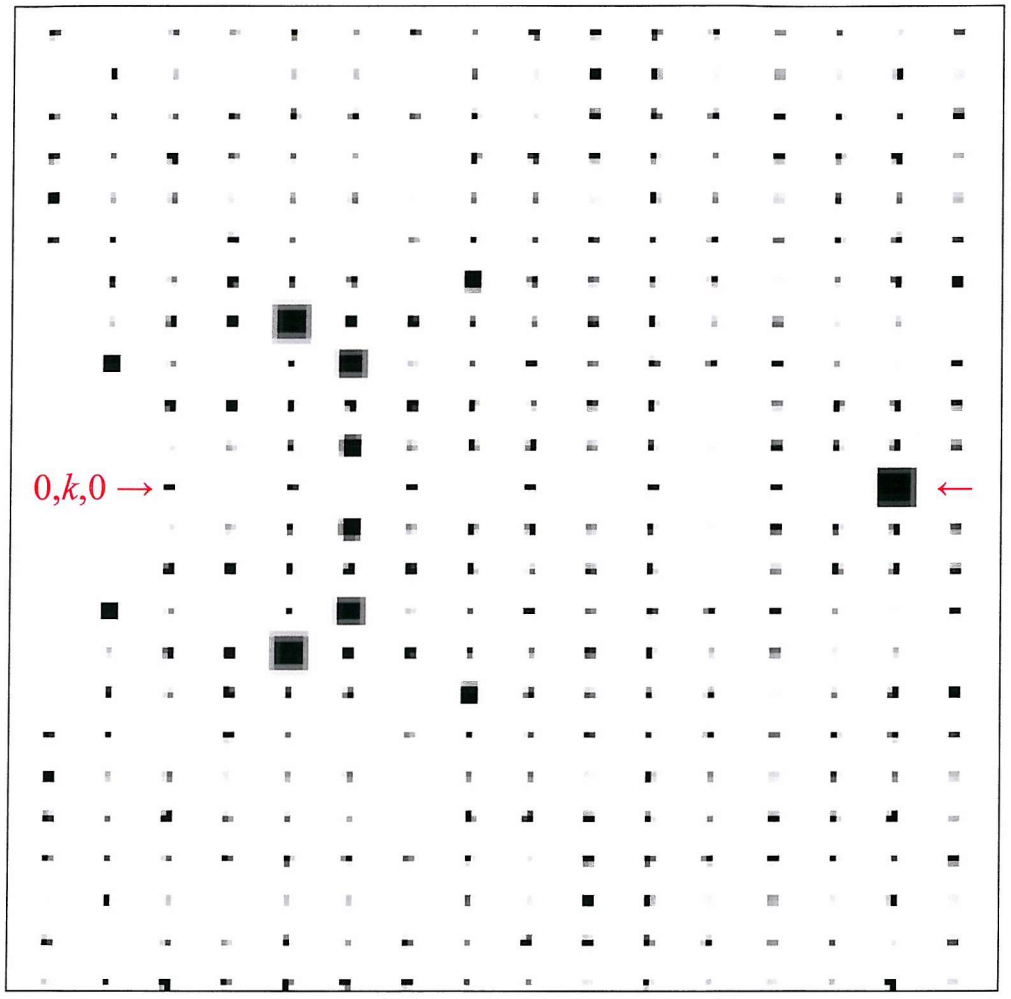


Figure 3.5. A pseudo precession picture displaying systematic absences along axis 0,*k*,0 (*k*≠2n) within the 0,*k*,*l* zone, implying a 2_1 screw axis along b for SAP-N-acetyl-D-proline data processed in P2 (prepared using HKLView (Collaborative Computational Project Number 4 1994)).

Table 3.1. Top 5 translation function peaks calculated with TFFC.

Peak Number	Shift Fraction			Signal/ Noise
	x	*y*	*z*	
1	0.503	0.000	-0.001	49.30
2	0.003	0.000	0.499	49.37
3	0.003	0.000	-0.001	49.24
4	0.503	0.000	0.499	49.23
5	0.621	0.000	0.890	2.81

3.2.4 Refinement & Model Building

Crystallographic refinement was performed using CNS (Brünger *et al* 1998) and X-PLOR (Brünger 1992), with a test set comprising of 10% (CNS) or 5% (X-PLOR) of the reflections randomly flagged to calculate the *R-free* value. The model was refined against the data in 23 stages (stages 1, 4, and 5 performed using X-PLOR). Between each round of refinement the improved model was used to calculate $|F_{obs}|-|F_{calc}|$ and $2|F_{obs}|-|F_{calc}|$ sigmaA-weighted electron density maps for visualisation and manual model building within Quanta (Molecular Simulations 1996) and XtalView (McRee 1999). Also, throughout the refinement bulk solvent correction was applied and the stereo-chemical properties of the model monitored using WHAT IF (Vriend 1990) and Procheck (Laskowski *et al* 1993). A summary of the *R-factor* and *R-free* values during the refinement are shown in figure 3.7.

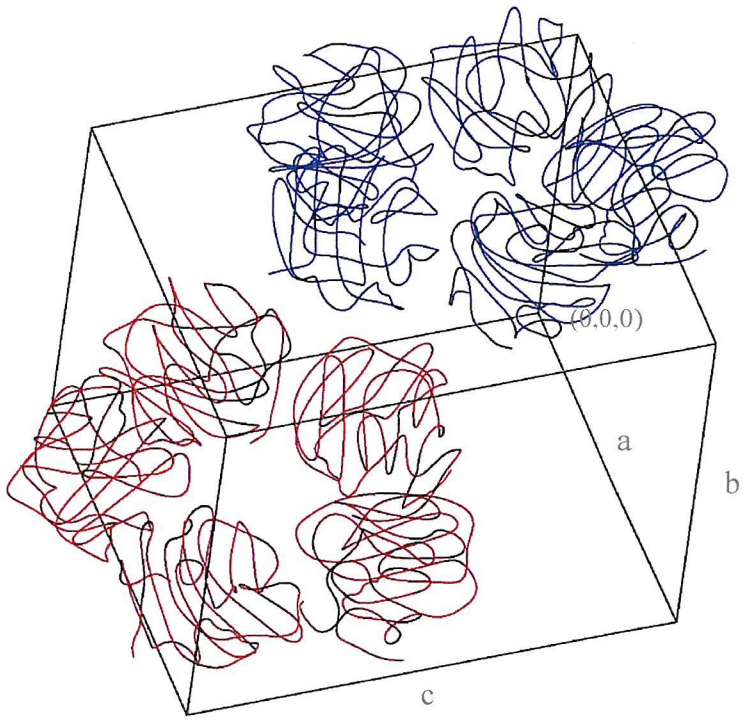
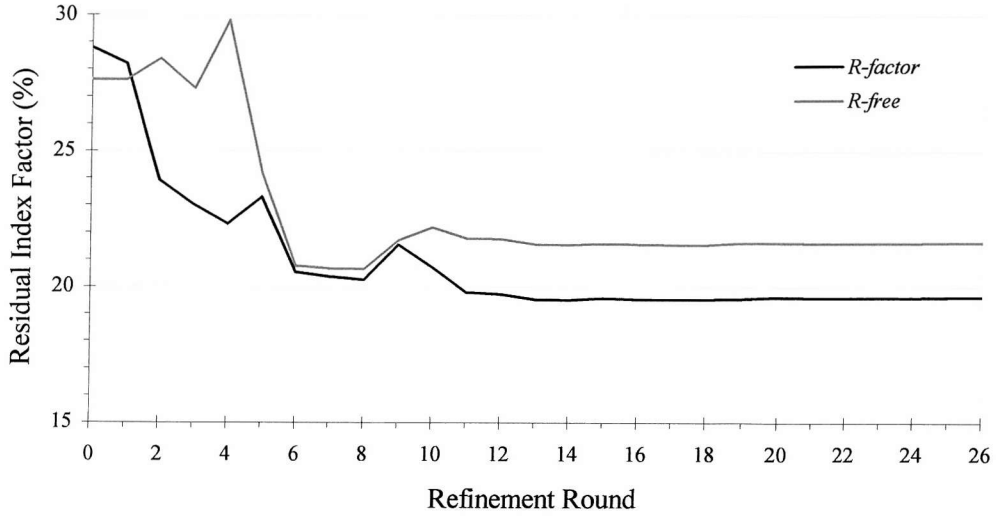


Figure 3.6. The crystal packing of SAP-N-acetyl-D-proline (prepared using **Bobscript** (Esnouf 1997), GL Render (Esser 2001), and Raster3D (Merritt & Murphy 1994)). In the presence of N-acetyl-D-Proline, SAP crystallises in the monoclinic space group $P2_1$ (a=96.1Å, b=70.8Å, c=103.6Å, and β=96.8°) with one SAP pentamer in the asymmetric unit. SAP pentamers are depicted in different colours with the axes and origin of the unit cell labelled accordingly.

Figure 3.7. Summary of the refinement process for SAP-N-acetyl-D-proline.



The model was initially refined by rigid-body refinement with the protomers treated as rigid bodies using reflections within the range 10 to 2.5Å (round 1), reducing the *R-factor* only slightly from 28.8% (*R-free* 27.6%) to 28.2% (*R-free* 27.6%). Examination of the resulting electron density maps revealed well-defined density for two calcium ions and a N-acetyl-D-proline molecule at the ligand-binding site of each subunit (see figure 3.8). In order to prevent protomer deviations in the phasing model biasing the refinement, manual model building was performed using protomer averaged (five-fold) NCS electron density maps. Simulated annealing using torsion angle dynamics (slow cool from 2,500 to 300K in 25K decrements) was then performed with the resulting model (round 2), using reflections within an expanded resolution range of 20 to 2.5Å. Calcium ions were built into the electron density at the double calcium-binding sites before a further round of simulated annealing was performed (round 3). N-acetyl-D-proline molecules, built from a modified coordinate file downloaded from the Cambridge Structural Database (Fletcher *et al* 1996),

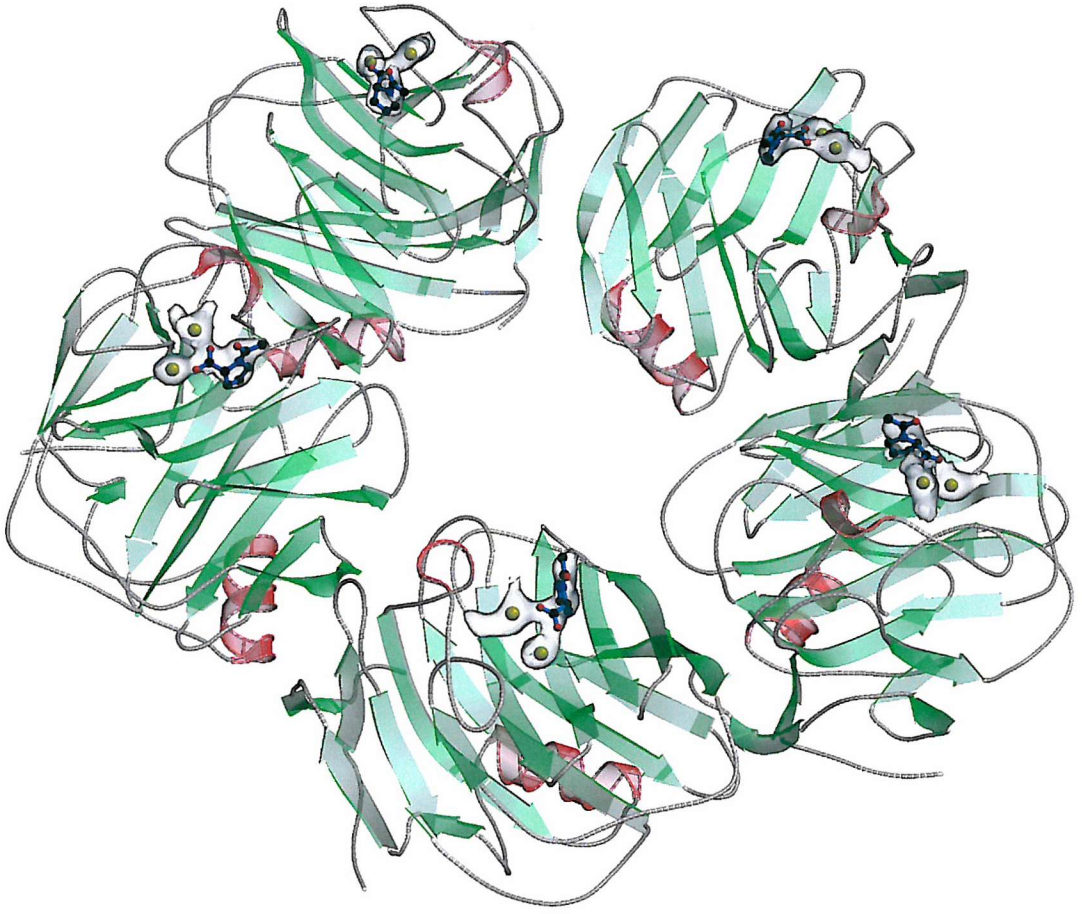


Figure 3.8. Three-dimensional structure of SAP-N-acetyl-D-proline illustrating the electron density for two calcium ions and a N-acetyl-D-proline molecule at the ligand-binding site of each subunit (prepared using Bobscript, GL Render, and Raster3D). The $|F_{obs}|$-$|F_{calc}|$ difference density from the initial electron density map is represented as transparent surfaces (contoured at 2σ) and clearly resembles the final atomic positions of the calcium ions and N-acetyl-D-proline molecules. The β-strands are represented as green arrows, α-helices as red ribbons, loops as gray cords, calcium ions as yellow spheres, and N-acetyl-D-proline molecules as ball-and-stick models.

were then built into electron density before positional and individual isotropic B-factor refinement was performed (rounds 4 and 5 respectively). Isotropic B-factor refinement was then repeated with an expanded resolution range of 30 to 2.4Å (round 6), giving an *R-factor* of 20.5% (*R-free* 20.8%).

During subsequent rounds of refinement solvent molecules and sugar residues were built into electron density between successive rounds of individual isotropic B-factor refinement. Due to the resolution limit only well-defined water molecules were included, requiring electron density over 3σ in the $|F_{obs}|-|F_{calc}|$ map, over 2σ in the $2|F_{obs}|-|F_{calc}|$ map, correct hydrogen bonding parameters, and a B-factor below 50Å^2. 66 water molecules were initially accepted after two successive rounds of model building and refinement (rounds 7 and 8). The first sugar residue of the glycosylation at Asn32 was then built into the electron density on each subunit, causing a slight increase in the *R-factor* and *R-free* values. The preceding round of individual isotropic B-factor refinement (round 9) was unable to refine the sugar residues with sufficient vigour, reducing the *R-factor* and *R-free* only slightly. Hence, the sugar residues and local surrounding areas (within 5Å) were subjected to positional refinement prior to further isotropic B-factor refinement (round 10). A further 33 water molecules were added in successive rounds of model building and refinement (rounds 11 to 18), with a second sugar residue added between rounds 17 and 18 to one subunit. The other remaining sugar residues were too poorly defined to be built in with sufficient accuracy (see figure 3.9). During the latter stages of the refinement process the ring pucker of the N-acetyl-D-proline molecules began to deviate from the original β/γ-carbon 'up' configuration to an unusual α-carbon envelope (Cremer & Pople 1975; Milner-White *et al* 1992). This cannot be justified by the resolution of the data and the pucker was normalised during the final rounds of refinement (rounds 19 to 23).

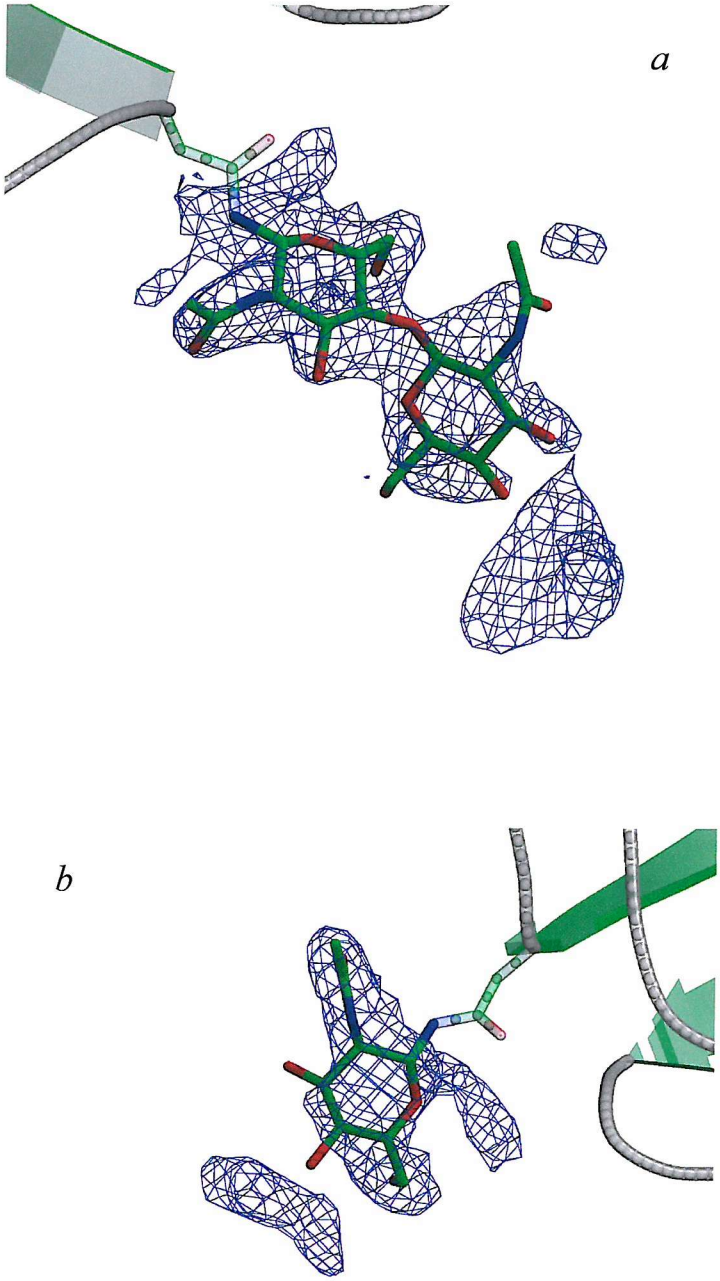


Figure 3.9. $|F_{obs}|$-$|F_{calc}|$ difference density for the glycosylation at Asn32 on subunit A (*a*) and D (*b*) (prepared using Bobscript, GL Render, and Raster3D). Electron density contoured at 2σ (depicted in blue) was sufficient for 2 sugar residues to be built in subunit A and 1 in subunit D (subunits labelled clockwise from A to E when viewed from the B-face). The final positions of these sugar residues in the fully refined structure are also shown. Secondary structural elements are represented as previously described with Asn32 and N-acetylglucosamine sugar residues shown as stick models.

The final round of individual isotropic B-factor refinement gave an *R-factor* of 19.6% (*R-free* 21.6%) with an average B-factor of 21.6Å². The resulting model showed good stereo-chemical properties with no residues within the disallowed regions of the Ramachandran plot (see figure 3.10). The overall statistics for the fully refined structure are summarised in table 3.2.

3.2.5 Protomer Structure of SAP

A detailed secondary structure assignment was performed manually according to published descriptors (described in detail in the appendix) and is summarised in tables 3.3 to 3.6. The structural assignment highlighted 48% (98 residues) of all residues in β-strand structures, 7% (14 residues) in helical structures, and 57% (117 residues) non-repetitive structures that includes 22% (44 residues) random coil. The strand residues are partitioned into 15 β-strands, labelled A to O from the N- to C-terminus, arranged in 2 large antiparallel sheets in the form of a flattened β-barrel with jellyroll topology (described previously in section 1.5.1). The tertiary structure is stabilised by the hydrophobic core created by the packing of hydrophobic and aromatic side chains between the adjacent faces of the β-sheets. The stability of the β-strands is clearly highlighted by the low thermal motion of the residues involved, average B-factor 17.4Å² (see figure 3.11). The most stable region of the tertiary structure is that of the β-meander (strands EFGHIJ) that comprises 39% of the total residues (63 residues) with an average B-factor of 17.3Å². This is primarily due to the hydrophobic core and tight 4:6 and 2:2 β-hairpins between adjacent strands (see table 3.5). The other strands (ABCDKLMNO) display the classic jellyroll topology, with B and N showing slightly higher thermal motion than the others (21.4Å² and 22.1Å² respectively) due to solvent exposure of the N/O end of the hydrophobic core.

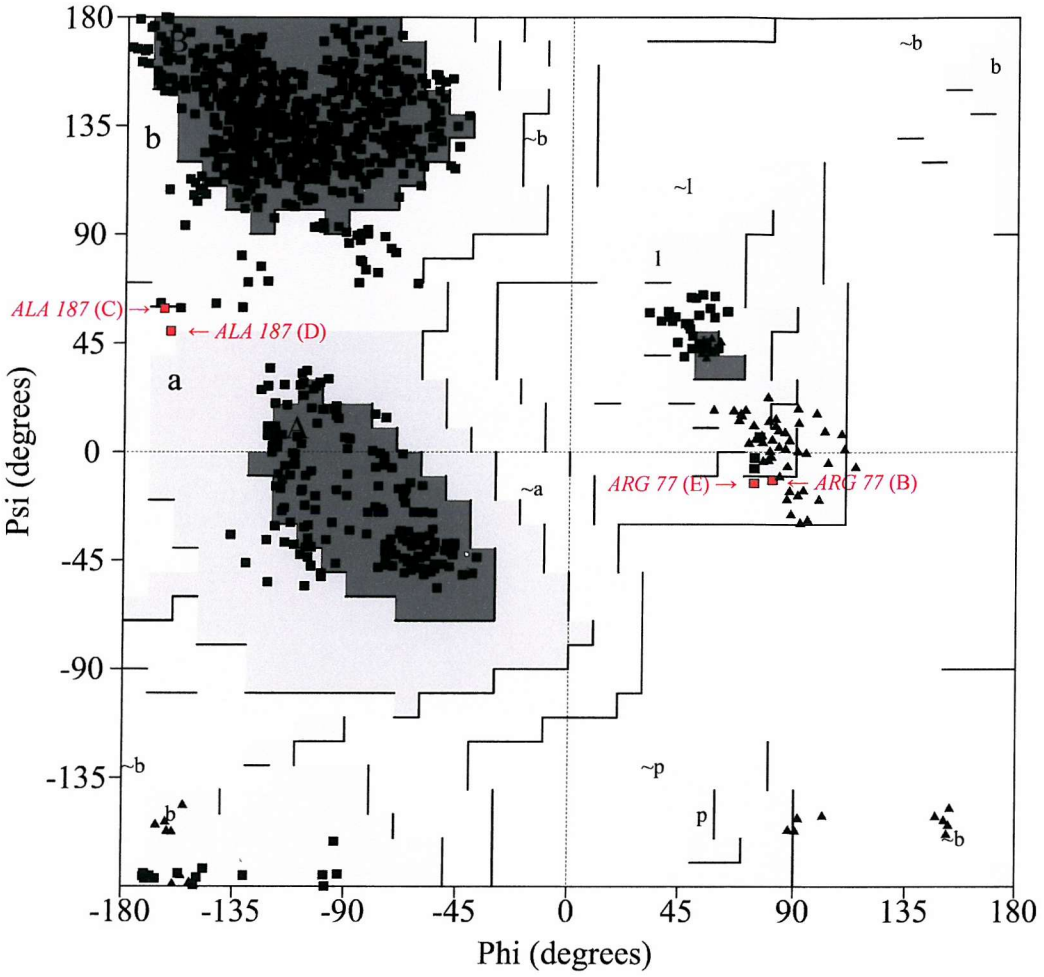


Figure 3.10. Ramachandran plot for the final SAP-N-acetyl-D-proline structure (prepared using Procheck). Statistics for non-glycine and non-proline residues (875 residues) indicate 86.1% of residues in most favoured regions [A, B, L], 13.5% residues in additional allowed regions [a, b, l, p], 0.5% residues in generously allowed regions [~a, ~b, ~l, ~p], and 0.0% residues in disallowed regions. Glycine residues are represented as black triangles, with all other residues as black squares with the exception of the 4 residues within the generously allowed region (Ala187 in subunit C & D, and Arg77 in subunits B & E) which are highlighted in red.

Table 3.2. Summary of the processing and model statistics for the SAP N-acetyl-D-proline structure.

Space Group	$P2_1$ (unique axis along b)
Unit Cell	a=96.1Å b=70.8Å c=103.6Å
	β=96.8°
Unit Cell Volume (Å³)	700,040
Resolution Range (Å)	30 - 2.4
Total Number of Reflections	108,982
Unique Number of Reflections	52,618
Multiplicity	2.1
Mosaicity (°)	0.3
Completeness (%)	97.0 (97.0)
R-merge(I) (%)	9.0 (30.0)
Average I/σ(I)	6.6 (3.1)
Molecules in Asymmetric Unit	1 SAP Pentamer
	10 Calcium Ions
	5 N-acetyl-D-proline Molecules
	6 Sugar Residues
	99 Water Molecules
Solvent Content (%)	55.3
Model *R-factor* (%)	19.6
Model *R-free* (%)	21.6
rms Bond Lengths (Å)	0.010
rms Bond Angles (°)	1.57

Figures in parentheses denotes the highest resolution shell statistics between 2.53-2.40Å

The helical and the majority of the non-repetitive structures are contained within the loop regions connecting the β-strands together (see table 3.3 and figure 3.12). The loop regions show increased thermal motion (see figure 3.11) with the highest thermal disorder between strands B-C, D-E, J-K, K-L, and L-M, all of which are located upon the exposed surface of the protein. Surprisingly the tight 2:2 β-hairpin between strands D and E shows significant thermal disorder, probably caused by the thermal instability of the nearby loops (between stands J-K and K-L) and/or the presence of the proceeding β-bulge (see table 3.6). The 10-residue α-helix (residues 165-174) is contained within the loop region between strands L and M, and lies on the exposed surface of flatter sheet (strands ACHIJLMO) directly above the disulphide bridge between Cys36 and Cys95. The helix interacts with underlying strands (C and H) through hydrophobic interactions with Pro166, Ile169, and Ala172, and is anchored to strand C though a hydrogen bond between the side chain of Arg38 and the main chain carbonyl of Tyr173. The remaining helical residues are contained within a short 3_{10} helix (residues 145-148) in the loop region between strands K and L. The 3_{10} helix and preceding turns are held against the surface of the buckled sheet (strands BDEFGKN) in close proximity to the double calcium-binding site (see figure 3.12). Despite calcium coordination (discussed in detail in section 3.2.8) this loop region is stabilised by internal hydrogen bonds surrounding the helical and γ-turn structures, and by 10 or 11 hydrogen bonds with the surrounding regions 14-18, 45-51, and 128-130. Interestingly the random coil region containing the site of proteolysis in the absence of calcium (residues 142-144 (see section 1.5.2)), located between the 3_{10} helix and a γ-turn structure, is not involved in the formation of hydrogen bonds with surrounding residues.

The remainder of the non-repetitive structures are present in the form of β-turns and β-bulges within the β-strands (see table 3.3 and 3.6). The disposition of these non-repetitive structures upon the 'binding' face (see figure 3.12), are at least in part responsible

Table 3.3. Secondary structural elements of SAP listed in order from N- to C-terminus.

Residue Range	Sequence	Secondary Structure
1 - 3	HTD	Random Coil
4 - 7	LSGK	β-Turn Type II *
7 - 11	KVFVF	β-Strand A
12 - 14	PRE	Random Coil
15 - 18	SVTD	β-Turn Type IV
19 - 23	HVNLI	β-Strand B
24 - 26	TPL	γ-Turn Inverse
27 - 31	EKPLQ	Random Coil
32 - 40	NFTLCFRAY	β-Strand C
41 - 46	SDLSRA	Random Coil
47 - 54	YSLFSYNT	β-Strand D
48 - 51	SLFS	β-Turn Type VIII
54 - 57	TQGR	β-Turn Type II *
57 - 67	RDNELLVYKER	β-Strand E
57 - 60	RDNE	β-Turn Type IV
59 - 62	NELL	β-Turn Type VIII
67 - 70	RVGE	β-Turn Type II *
70 - 75	EYSLYI	β-Strand F
75 - 78	IGRH	β-Turn Type I′ *
78 - 84	HKVTSKV	β-Strand G
85 - 86	IE	Random Coil
87 - 90	KFPA	β-Turn Type VIb
91	P	Random Coil
92 - 99	VHICVSWE	β-Strand H
99 - 102	ESSS	β-Turn Type III

100 - 103	SSSG	β-Turn Type I
104 - 109	IAEFWI	β-Strand I
108 - 111	WING	β-Turn Type IV
109 - 112	INGT	β-Turn Type I′ *
112 - 119	TPLVKKGL	β-Strand J
120 - 123	RQGY	β-Turn Type II *
124 - 129	FVEAQP	Random Coil
130 - 133	KIVL	β-Strand K
132 - 135	VLGQ	β-Turn Type IV
136 - 138	EQD	Random coil
139 - 141	SYG	γ-Turn Inverse
142 - 144	GKF	Random Coil
145 - 148	DRSQ	3_{10} Helix
149 - 151	SFV	Random Coil
152 - 160	GEIGDLYMW	β-Strand L
161 - 164	DSVL	Random Coil
165 - 174	PPENILSAYQ	α-Helix
175 - 179	GTPLP	Random Coil
180 - 183	ANIL	β-Turn Type IV
181 - 184	NILD	β-Turn Type IV
183 - 184	LD	β-Strand M
184 - 187	DWQA	β-Turn Type I *
185 - 188	WQAL	β-Turn Type IV
189 - 193	NYEIR	β-Strand N
194 - 196	GYV	γ-Turn Inverse
197 - 201	IIKPL	β-Strand O
201 - 204	LVWV	β-Turn Type IV

* β-turns with a hydrogen bond between residue *i* and *i*+3

Table 3.4. Phi and psi dihedral angles for the turn structures in subunit A of SAP listed in order from N- to C-terminus (subunits labelled clockwise from A to E when viewed from the B-face).

Residue Range	Turn Type	Phi *i*+1 (°)	Psi *i*+1 (°)	Phi *i*+2 (°)	Psi *i*+2 (°)
4 - 7	β-Turn Type II *	-53.9	141.3	68.2	12.8
15 - 18	β-Turn Type IV	-120.7	8.3	-118.8	-30.6
24 - 26	γ-Turn Inverse	-84.5	78.7	-	-
48 - 51	β-Turn Type VIII	-98.1	-36.6	-149.9	113.9
54 - 57	β-Turn Type II *	-47.5	117.2	58.3	17.2
57 - 60	β-Turn Type IV	-83.2	154.8	45.7	49.5
59 - 62	β-Turn Type VIII	-75.0	-36.8	-151.6	103.6
67 - 70	β-Turn Type II *	-61.3	124.6	81.9	13.5
75 - 78	β-Turn Type I′ *	60.8	45.4	74.0	-2.4
87 - 90	β-Turn Type VIb	-142.3	121.5	-93.5	153.4
99 - 102	β-Turn Type III	-60.9	-43.5	-54.6	-39.7
100 - 103	β-Turn Type I	-54.6	-39.7	-106.8	-22.2
108 - 111	β-Turn Type IV	-114.0	117.0	54.3	42.9
109 - 112	β-Turn Type I′ *	54.3	42.9	76.7	6.6
120 - 123	β-Turn Type II *	-63.0	133.4	86.5	-6.0
132 - 135	β-Turn Type IV	-114.0	134.9	99.1	-19.9
139 - 141	γ-Turn Inverse	-75.4	89.1	-	-
180 - 183	β-Turn Type IV	-104.4	-8.0	-100.1	-50.4
181 - 184	β-Turn Type IV	-100.1	-50.4	-137.5	127.7
184 - 187	β-Turn Type I *	-55.5	-25.4	-91.0	5.9
185 - 188	β-Turn Type IV	-91.0	5.9	-166.1	61.5
194 - 196	γ-Turn Inverse	-91.2	90.8	-	-
201 - 204	β-Turn Type IV	-114.9	-5.8	-109.0	27.6

* β-turns with a hydrogen bond between residue *i* and *i*+3

Table 3.5. β-hairpins structures of SAP listed in order from N- to C-terminus.

Residue Range	Between Strands	Hairpin Class	Secondary Structural composition
54 - 57	D-E	2 : 2	β-Turn Type II
67 - 70	E-F	2 : 2	β-Turn Type II
75 - 78	F-G	2 : 2	β-Turn Type I′
99 - 104	H-I	4 : 6	β-Turn Type III & I, & G1 β-Bulge
109 - 112	I-J	2 : 2	β-Turn Type I′
159 - 183	L-M	23 : 23	α-Helix & β-Turn Type IV

Table 3.6. β-bulge structures of SAP listed in order from N- to C-terminus of residue X (see appendix).

Residue				Bulge Type
X	1	2	3	
37	156	157	-	Antiparallel Wide
53	58	59	60	Antiparallel Special
63	49	50	-	Antiparallel Classic
70	66	67	-	Antiparallel Classic
99	103	104	-	Antiparallel G1
159	182	183	-	Antiparallel Classic

Figure 3.11 Comparison of average isotropic B-factors for the main chain atoms of different SAP subunits (subunits labelled clockwise from A to E when viewed from the B-face).

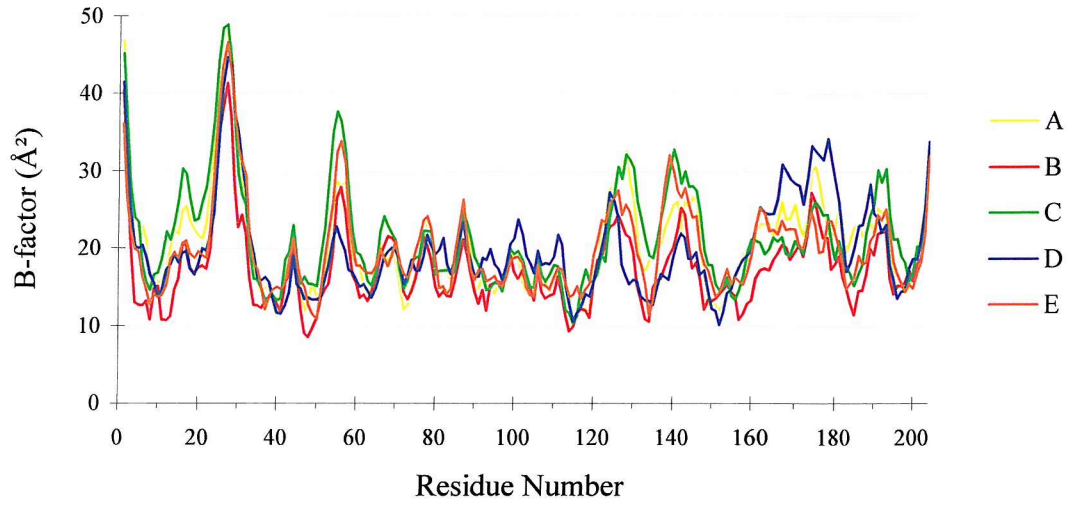
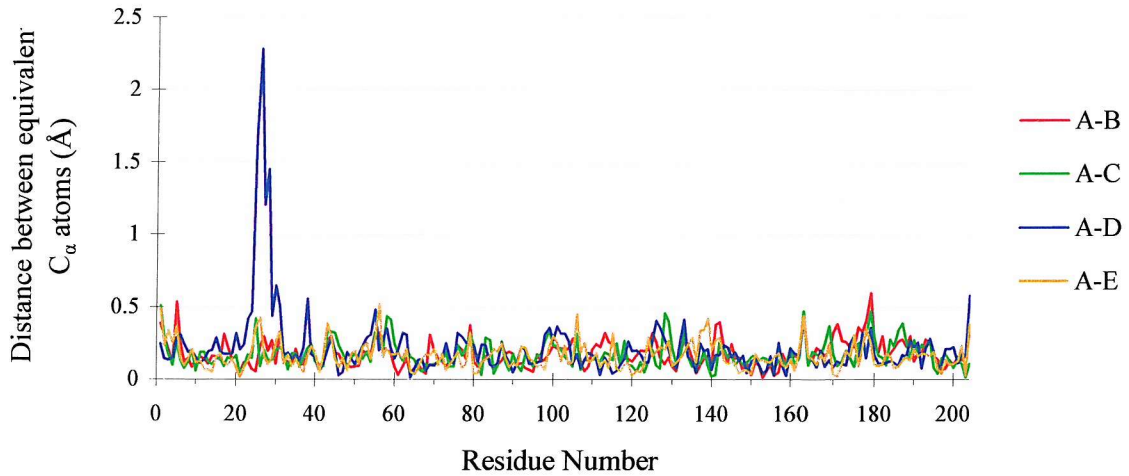


Figure 3.13. C_α distances between superimposed subunits of SAP (subunits labelled clockwise from A to E when viewed from the B-face).



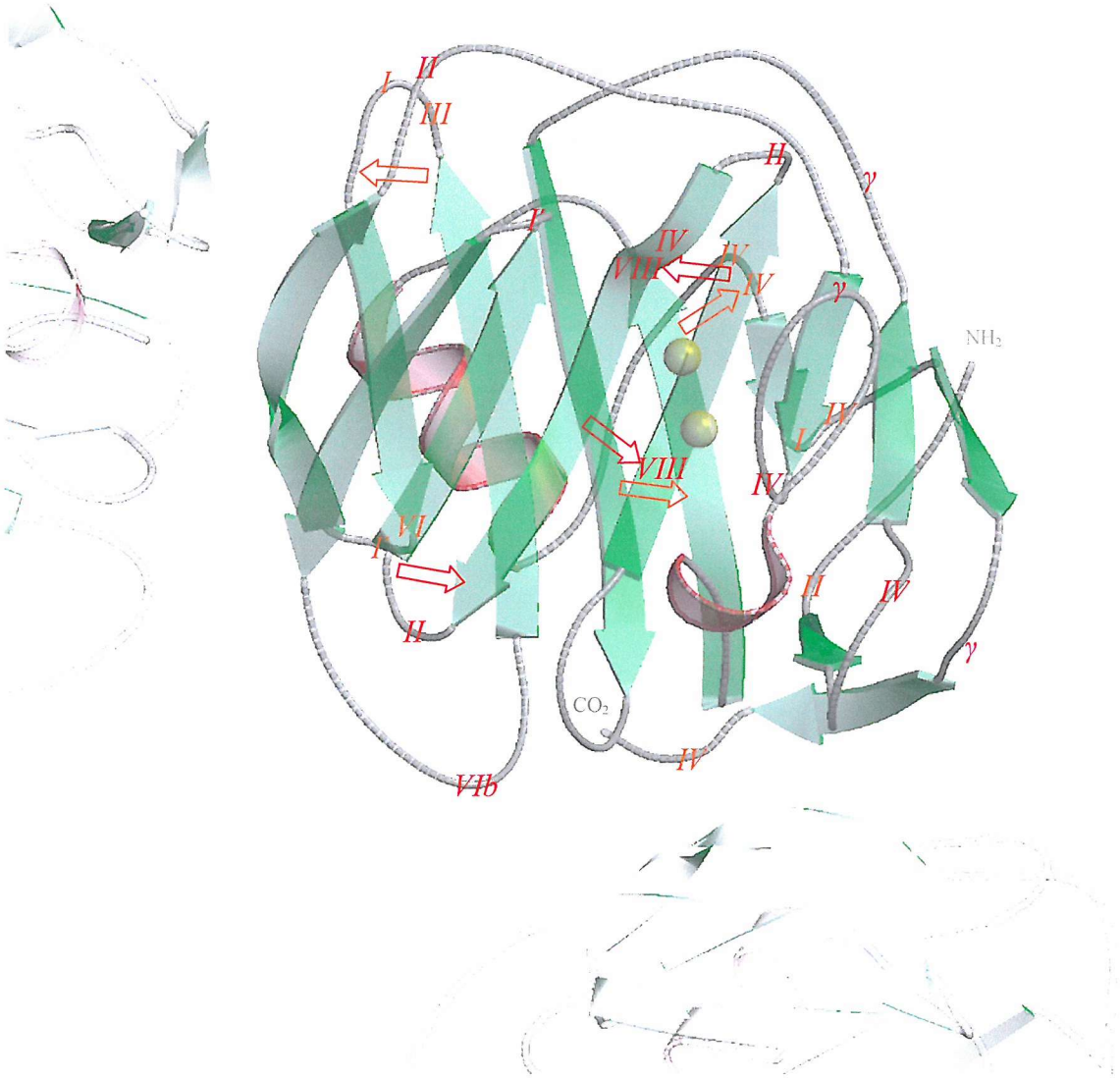


Figure 3.12. Three-dimensional structure of subunit A from the final SAP-N-acetyl-D-proline model, viewed from the double calcium-binding site (prepared using Bobscript, GL Render, and Raster3D). Secondary structural elements are represented as previously described with the disposition of non-repetitive structures labelled in red (upper sheet) and orange (lower sheet). The β-turns are labelled according to their type, inverse γ-turns as 'γ', and β-bulges as arrows from residue X to the opposing strand.

for the buckling of the sheet. Those located at the terminal regions of the central strands also enable the coordination of side chains involved in the ligand-binding site.

There is a remarkable amount of similarity between the 5 subunits, which when superimposed show an average C_α rms fit of 0.14Å. Examination of the protomers reveals one major deviation within the loop region between strands B and C (residues 23-31) of subunit D (subunits distinguished by labelling clockwise from A to E when viewed from the B-face) (see figure 3.13 and figure 3.14), which when omitted from the C_α rms calculation improves the fit to 0.10Å. Not only does the loop show a large deviation in C_α distance but also shows a deviation in secondary structure with the reduction of an inverse γ-turn (residues 24-26) to a random coil structure. There is also a minor deviation in secondary structure where the main chain dihedrals are slightly altered from a type IV to a type VIII β-turn (residues 181-184 in subunit D). These deviations are directly and indirectly due to crystal packing interaction described in section 3.2.7.

3.2.6 Pentamer Structure of SAP

Five protomers are arranged in a ring with five-fold radial symmetry creating a pentameric structure with two separate and distinct faces (described previously in section 1.5.3), burying 21.5% (1,751Å²) of the protomers surface area. The buried regions between neighbouring subunits are comprised of residues 10-12, 40-42, 88-89, 151-153, and 195-203, and residues 82-85, 99, 102-108, 113-121, and 166. There are an extensive number of interactions between these surfaces (see table 3.7 and figure 3.15), including three salt bridges between Asp42-Lys117, Glu153-Lys166, and Lys199-Glu99, and three hydrophobic interactions between the side chains of Tyr40-Val115, Ile197-Ile104, and Val202 with Trp108, Pro113, and Pro166. There are also four direct hydrogen bonds formed through the main chain and side chain interactions of Pro12-Gly118, Tyr40-Pro113,

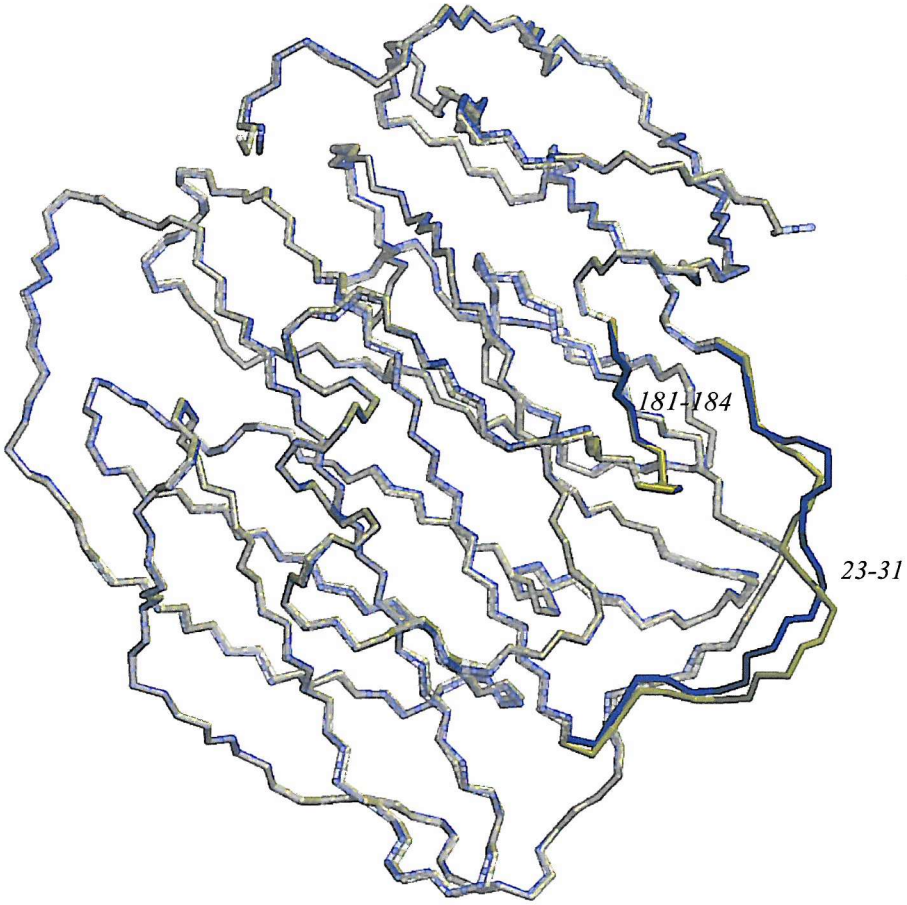


Figure 3.14. Superposition of subunits A (yellow) and D (blue) of SAP (prepared using Bobscript, GL Render, and Raster3D). The rms deviation is 0.24Å for equivalent C_α atoms. The molecular regions displaying the largest C_α and structural deviations are highlighted and labelled according to the residues involved; when omitted from the C_α rms calculation the fit is improved to 0.10Å.

Table 3.7. Inter-subunit contacts of SAP listed in order from N- to C-terminus of subunit *i* (see figure 3.15).

Residue/s (subunit *i*)	Residue/s (subunit *i*+1)	Interaction
Pro12	Gly118	Hydrogen Bond
Tyr40	Pro113	Hydrogen Bond
Tyr40	Val115	Hydrophobic
Asp42	Lys83	Hydrogen Bond
Asp42	Lys83 Val115	Hydrogen Bond Via Water
Asp42	Lys117	Salt Bridge
Glu153	Lys116 Leu114	Hydrogen Bond Via Water
Glu153 Pro200 Val202	Lys116	Hydrogen Bond Via Water
Glu153	Lys116	Salt Bridge
Tyr195	Ser102	Hydrogen Bond
Ile197	Ile104	Hydrophobic
Lys199	Glu99	Salt Bridge
Val202	Trp108 Pro113 Pro166	Hydrophobic

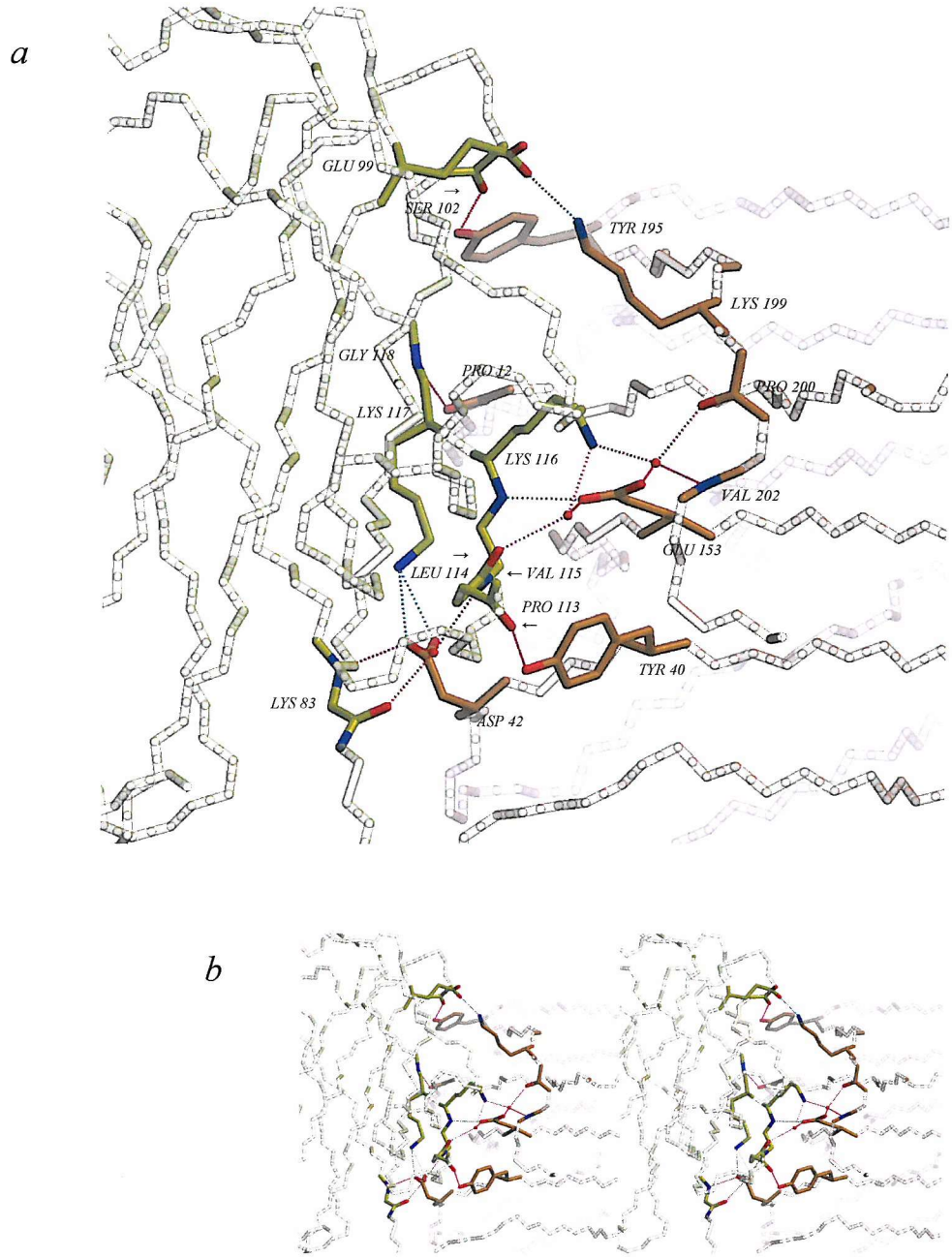


Figure 3.15. Three-dimensional structure of the interface between subunits A (yellow) and E (orange) from the SAP-N-acetyl-D-proline model viewed as an orthogonal projection (*a*) and stereoview (*b*) (prepared using Bobscript and Raster3D). The main chain is represented as a C_α trace with the interacting residues shown as stick models and coordinated water molecules as red spheres. Interactions are highlighted as blue dotted lines for salt bridges and red dotted lines for hydrogen bonds.

Asp42-Lys83, and Tyr195-Ser102, and several indirect hydrogen bonds formed through the coordination of three water molecules between the buried surfaces.

3.2.7 Packing Interactions

Crystallographic packing of the monoclinic SAP-N-acetyl-D-proline crystal form occurs in parallel planes with pentamers orientated in the same direction with their five-fold radial symmetry perpendicular. This enables each pentamer to interact with 12 surrounding pentamers, of which 10 form solvent excluded contacts (see table 3.8). Interactions within each plane occur between the loop regions of subunits B and D, where the flexibility and deformation of the latter (residues 23-31) enables the formation of two salt bridges (Glu27-Arg77 and Glu27-Arg120) and a hydrogen bond (Glu27-Arg57). Interactions between stacking planes occur between the external loop regions of subunits C, D, and E, and the opposing faces of subunits A and B. Contacts formed through interactions with exterior loop regions involve the formation of a salt bridge (Asp184-Arg57 or Asp58-Glu167) and in one of the cases a hydrogen bond (His78-Thr176). The remaining contacts involve packing against or near side chains involved in the double calcium-binding site. These hydrophobic/solvent excluded interactions are not within the bonding distance (over 3.5Å apart) and do not affect the calcium coordination or ligand-binding site. Further interactions are likely to occur within the porous regions where the mobile oligosaccharides are likely to reside.

3.2.8 Double Calcium-Binding Site

As with previous SAP structures, the double calcium-binding site is formed by the coordination of six localised protein residues on the B-face of each subunit (described previously in section 1.5.2). The residues involved are located within 3 short structural

Table 3.8. Packing contacts between adjacent SAP pentamers (subunits labelled clockwise from A to E when viewed from the B-face).

Subunit Contacts (*i* - *i*+1)	Residues (subunit *i*)	Residues (subunit *i*+1)	Interactions (inter-atomic distance < 3.2Å)
A-A	58, 138-139	178-179	-
B-B B-C	146-147, 176 178	3, 5, 17, 186 78	- -
B-D	57, 77, 120-124	25-27	Arg57-Glu27† Arg77-Glu27* Arg120-Glu27*
C-D	179, 184, 186	55-57	Asp184-Arg57*
D-E E-E	16-17 56-58, 77-78,120	1-3, 186-187 167-168, 176-178	Asp58-Glu167* His78-Thr176†

* Denotes salt bridge
† Denotes hydrogen bond

regions, the β-bulge region (residues 58-60) within strand E, and in a random coil (residues 136-138) and 3_{10} helical region (residues 145-148) between strands K and L. Coordination of the calcium sites is well defined with low thermal motion of the bound ions, average B-factors 19.7Å² (calcium I) and 20.9Å² (calcium II). The average distances for the calcium-ligand interactions are between 2.35 to 2.64Å (see table 3.9), which when compared to other reported metal-ligand interactions (Harding 1999) are within the correct range for the resolution of the data. However the strength of ligation may vary slightly with the inter-atomic distance. The calcium coordination positions the two ions 3.9 to 4.1Å apart, and they are bridged by the carboxyls of Glu136, Asp138, and N-acetyl-D-proline.

Coordination of the double calcium-binding site is summarised in figure 3.16 and figure 3.17.

Calcium site I shows seven ligands in a pentagonal bi-pyramidal arrangement, receiving six of the ligands from the surrounding protein residues (see figure 3.16). The side chains of Asn59, Glu136, Asp138, and the main chain carbonyl of Gln137 each provide one ligand while the carboxyl of Asp58 provides two. The coordination sphere is completed by a carboxyl oxygen of N-acetyl-D-proline, which also completes the pentagonal arrangement of the ligands provided by Asp58, Glu136, and Gln137. Calcium site II is reported to have six ligands in an octahedral coordination arrangement, with only three of the ligands provided by the protein (Emsley *et al* 1994). However, examination of the protein structure reveals a fourth protein ligand (see figure 3.16 and 3.17). The side chains of Glu136, Asp138, and Gln148 provide the three-reported protein ligands, while the second carboxyl oxygen of Asp138 provides the fourth. Although the average inter-atomic distance of the second Asp138 ligand is within the higher range (2.6Å) it is comparable to that of Asp58 (see table 3.9), the only other residue providing a symmetrical bidentate coordination. The coordination sphere of site II is completed by two water molecules and a carboxyl oxygen of N-acetyl-D-proline. The water molecules, designated I and II, are stabilised by hydrogen bonding with surrounding residues. Water I is stabilised by the main chain carbonyls of Lys143 and Glu136, while water II is stabilised by the side chain of Asp145 (see figure 3.16). This effectively anchors the solvent molecules to the body of the protein, reducing the degree of freedom whilst orientating the oxygen atom towards the calcium ion. This is reflected in the low thermal motion of the 2 water molecules, average B-factors 16.5Å^2 (water I) and 9.3Å^2 (water II). Glu136, Asp138, Gln148, and water II provide the pentagonal arrangement surrounding calcium II.

Table 3.9. Inter-atomic distances for the coordination of the double calcium-binding site (averages taken from the distances of all five subunits).

Calcium Ion	Ligand	Ligating Oxygen	Average Inter-atomic Distance (Å)
CaI	Asp58	OD1	2.61
	Asp58	OD2	2.73
	Asn59	OD1	2.43
	Glu136	OE2	2.44
	Gln137	O	2.45
	Asp138	OD1	2.39
	N-acetyl-D-Proline	O1	2.41
CaII	Glu136	OE1	2.35
	Asp138*	OD1	2.64
	Asp138	OD2	2.57
	Gln148	OE1	2.45
	H_2O I	-	2.47
	H_2O II	-	2.42
	N-acetyl-D-Proline	O2	2.49

* Ligand previously unreported

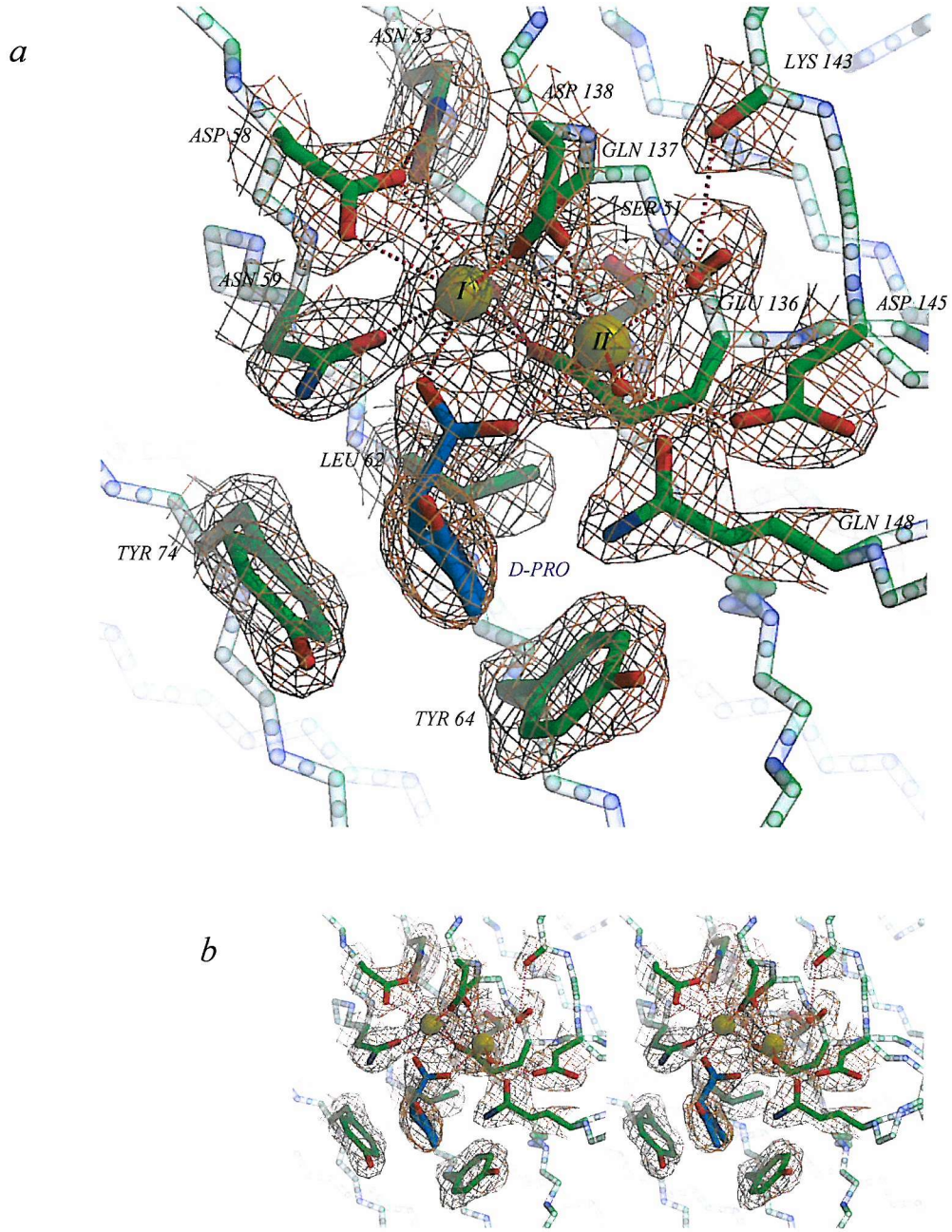


Figure 3.16. Three-dimensional structure of the double calcium-binding site in subunit A of the SAP-N-acetyl-D-proline model viewed as an orthogonal projection (*a*) and stereoview (*b*) (prepared using Bobscript and Raster3D). The $2|F_{obs}|-|F_{calc}|$ electron density surrounding the site is well defined at 1.3σ (depicted in orange) with the heptagonal coordination of both calcium sites and the stabilising hydrogen bonds clearly visible (represented by dotted lines). The binding of N-acetyl-D-proline (D-PRO) can also be seen bridging the two calcium ions and orientating the pyrrolidine ring within the adjacent hydrophobic pocket. The main chain is represented as a C_α trace, calcium ions as yellow spheres, water molecules as red spheres, with N-acetyl-D-proline (blue) and residues involved (green) shown as stick models and labelled accordingly. The previously unreported Asp138 ligand is highlighted as a blue dotted line.

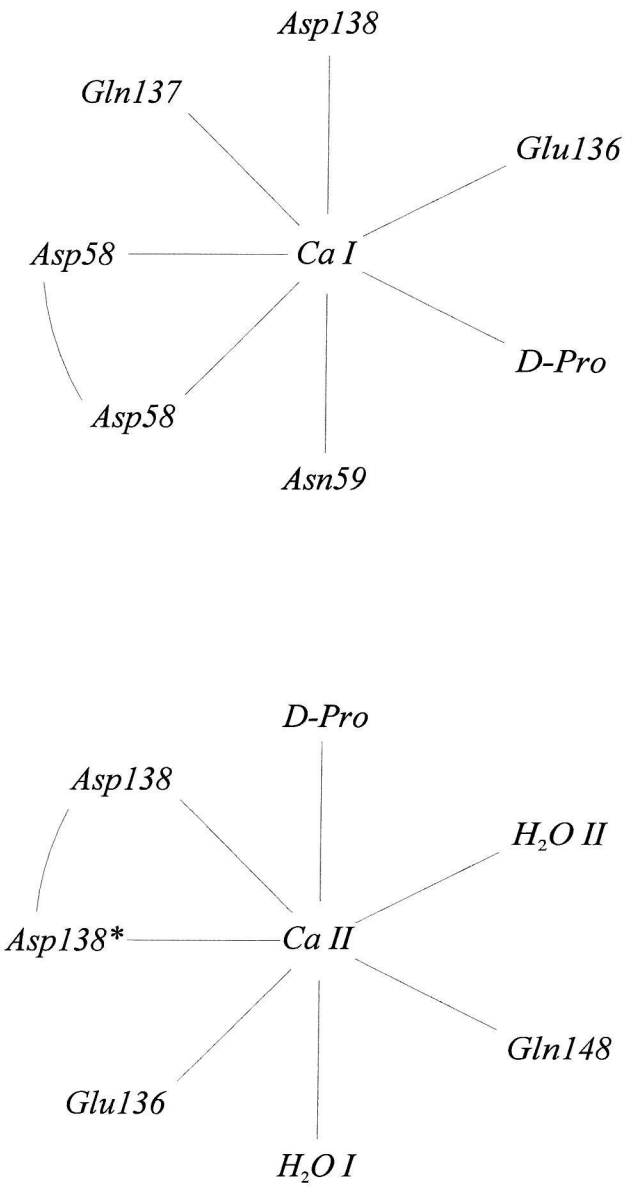


Figure 3.17. Coordination geometry of the two calcium-binding sites in SAP (* denotes the previously unreported ligand).

The double calcium-binding site is maintained by the disposition and small degree of freedom of the protein ligands, with the coordination geometry further stabilised by hydrogen bonding with surrounding residues. The disposition of the Asp58 is stabilised by hydrogen bonding to the amide nitrogen of Asn53, and Glu136 by coordinating through an underlying water molecule to the side chains of Ser51 and Asn53 (see figure 3.16). This creates further asymmetry between the two sites.

3.2.9 N-Acetyl-D-Proline Binding

Each N-acetyl-D-proline molecule is well defined with good electron density and low thermal disorder (average B-factor 24.9Å²). N-acetyl-D-proline is bound through the carboxyl oxygens, which bridge the bound calcium ions through two solvent exposed ligation sites. The carboxylate also receives two hydrogen bonds from the amide nitrogens of Asn59 and Gln148, both of which are involved in calcium coordination. This effectively anchors the molecule to the double calcium-binding site, with the geometry directing the pyrrolidine ring into the adjacent hydrophobic pocket (see figure 3.16 and 3.18). The pyrrolidine ring wedges tightly within the pocket burying only 0.9% (71.9Å²) of the protomers surface area, leaving 51% (74.9Å²) of the N-acetyl-D-proline molecule exposed. The D-isomer configuration enables the pyrrolidine ring to fit into the pocket with the face of the ring packing tightly against the side chain of Tyr74 (C_γ 3.7Å away), while the γ- and δ-carbons pack against Tyr64 (C_γ 3.7Å away), and the β- and γ-carbons against the underlying Leu62 (C_β 3.7Å away) (see figure 3.18). Packing is very snug with the inter-atomic distances (3.5-4.0Å) close to the Van der Waals' contact distances (3.4Å between carbon atoms (Nyburg & Faerman 1985)). Interestingly, anchoring through the carboxyl group allows a small degree of freedom in the insertion of the pyrrolidine ring, which is apparent between subunits A and D, where a 0.4Å deeper insertion in D effectively reduces

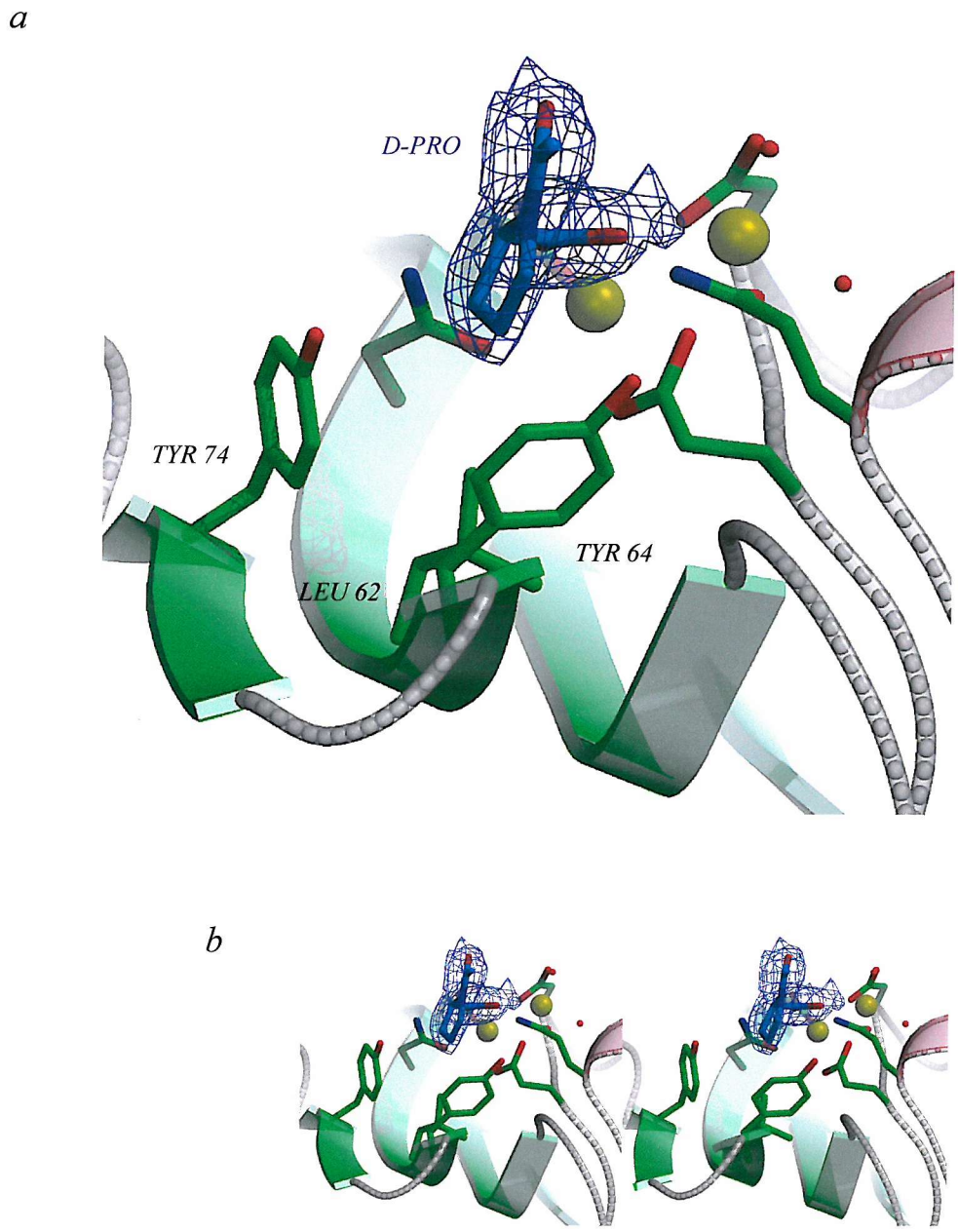


Figure 3.18. Three-dimensional structure of N-acetyl-D-proline binding to SAP viewed as an orthogonal projection (*a*) and stereoview (*b*) (prepared using Bobscript and Raster3D). The $|F_{obs}|$-$|F_{calc}|$ omit map contoured at 3σ (represented in blue) and the atomic positions for the N-acetyl-D-proline molecule shows the binding through the carboxyl group and the positioning of the pyrrolidine ring into the adjacent hydrophobic pocket (created by the side chains of Leu62, Tyr64, and Tyr74). Secondary structural elements are represented as previously described with N-acetyl-D-proline (blue) and important side chains (green) shown as stick models, and water molecules as red spheres.

the thermal motion of N-acetyl-D-proline from 29.4 to 19.3Å^2. The packing of the pyrrolidine ring within the hydrophobic pocket directs the peptide bond away from the body of the protein almost perpendicular to the five-fold radial symmetry. However, there is no discernable distinction between the cis/trans isomeric states within the difference map.

3.3 SAP Soaked with Ro 63-8695

Once Ro 63-8695 was made available then the molecule was soaked into a previously crystallised form of SAP grown in the presence of PC. The low affinity for PC binding and the large structural difference with that of Ro 63-8695 made this the ideal candidate for soaking.

3.3.1 Crystallisation

Crystals were grown in the presence of PC by the vapour diffusion technique of hanging drops using the conditions previously determined (Thompson 2000): 3µl drop of protein solution containing 12mg/ml SAP, 50mM PC, 60mM Tris pH 8.0, 84mM NaCl mixed with an equal volume of reservoir solution containing 60mM Tris pH 8.0, 84mM NaCl, 8% (w/v) PEG MME 550, 10mM calcium acetate, and allowed to equilibrate against 1ml of the reservoir solution at 4°C. Several crystals grew to an approximate size of 0.5x0.3x0.2mm within 3 weeks.

3.3.2 Data Collection & Processing

A crystal was transferred to a small drop of the reservoir solution containing 10mM Ro 63-8695, and allowed to equilibrate for 30 minutes at room temperature prior to mounting for data collection. Diffraction data were collected from a single crystal at room

temperature by the oscillation method using a MAR 300mm image plate. These data were collected in-house using an ENRAF NONIUS FR 591 rotating copper anode generator (λ=1.542Å). Each image was recorded during a 20-minute exposure over an oscillation range of 1°, with the detector placed 210mm away from the crystal. The dataset was recorded over a total of 97 images, an example of which is shown in figure 3.19. The data were corrected, processed, scaled, and merged using Denzo and Scalepack (Minor 1997), and programs from the CCP4 suite. The crystal packing was found to be monoclinic with unit cell dimensions of a=96.1Å, b=70.9Å, c=103.7Å, and β=96.8°, and was processed in the space group P2. Examination of the processed data revealed systematic absences in 0,*k*,0 (where $k \neq 2n$) implying a space group of $P2_1$ with a unique axis along b (see figure 3.20). The data were then reprocessed, scaled, and merged in the space group $P2_1$ to a maximum resolution of 2.55Å. The overall completeness of the data was 84.3%, with an *R-merge(I)* of 13.3%.

3.3.3 Molecular Replacement

A solvent content of 55.4% (V_M=2.61 Å^3/Da) was calculated for the SAP-Ro 63-8695 soaked crystal assuming one SAP pentamer is present in the asymmetric unit. This is consistent with the previously solved crystal structure of SAP in the presence of PC (see table 1.3). Unfortunately coordinates for the SAP-PC crystal form were unavailable at the time, although coordinates for isomorphous crystal forms grown in the presence of 37mM MOβDG (pentamer form) or 4mM N-acetyl-D-proline (see section 3.2) were available a non-isomorphous form was chosen to enable molecular replacement to be performed.

The initial phases for the SAP-Ro 63-8695 soaked crystal were obtained by molecular replacement using Molrep (Vagin & Teplyakov 1997). The originally derived SAP pentamer refined to 2.0Å resolution (Emsley *et al* 1994) (Protein Data Bank

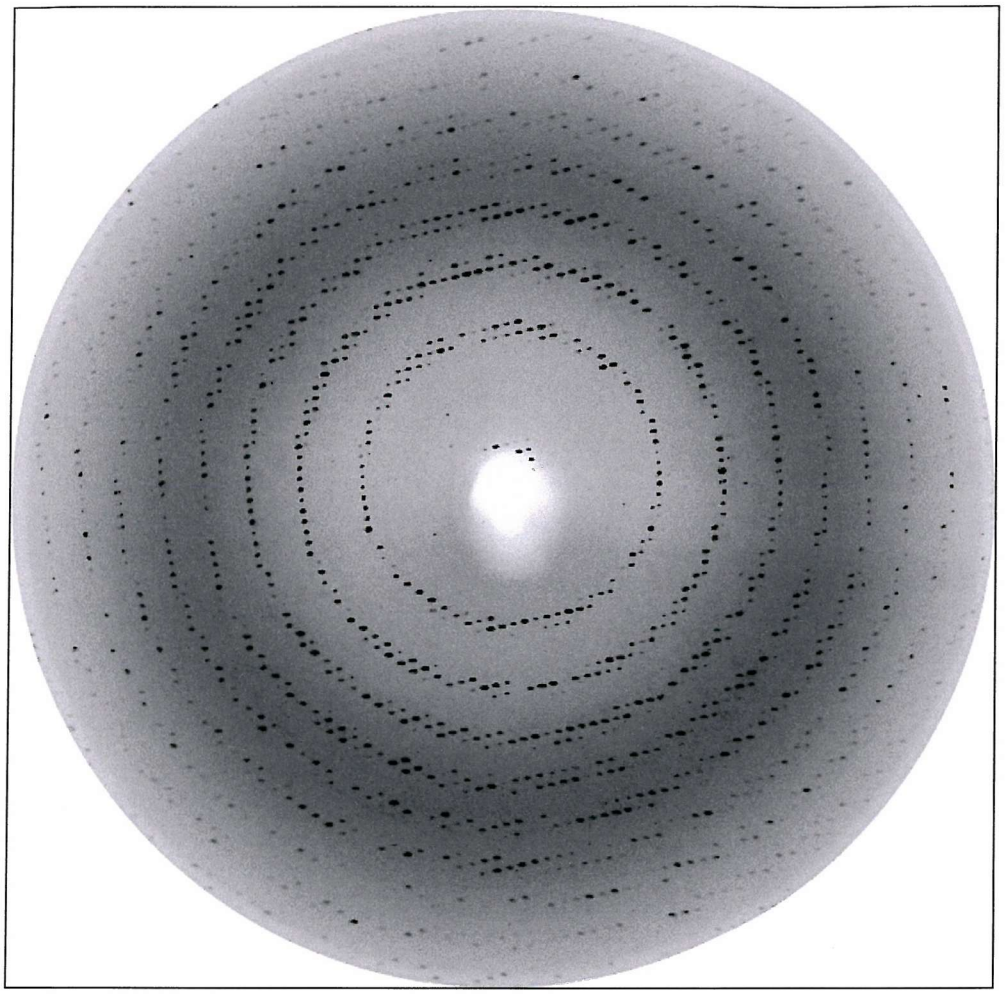


Figure 3.19. A 1° oscillation image collected in-house on a MAR 300mm image plate from a SAP-PC crystal soaked with Ro 63-8695, recorded to a maximum resolution of 2.53Å (prepared using Mosflm).

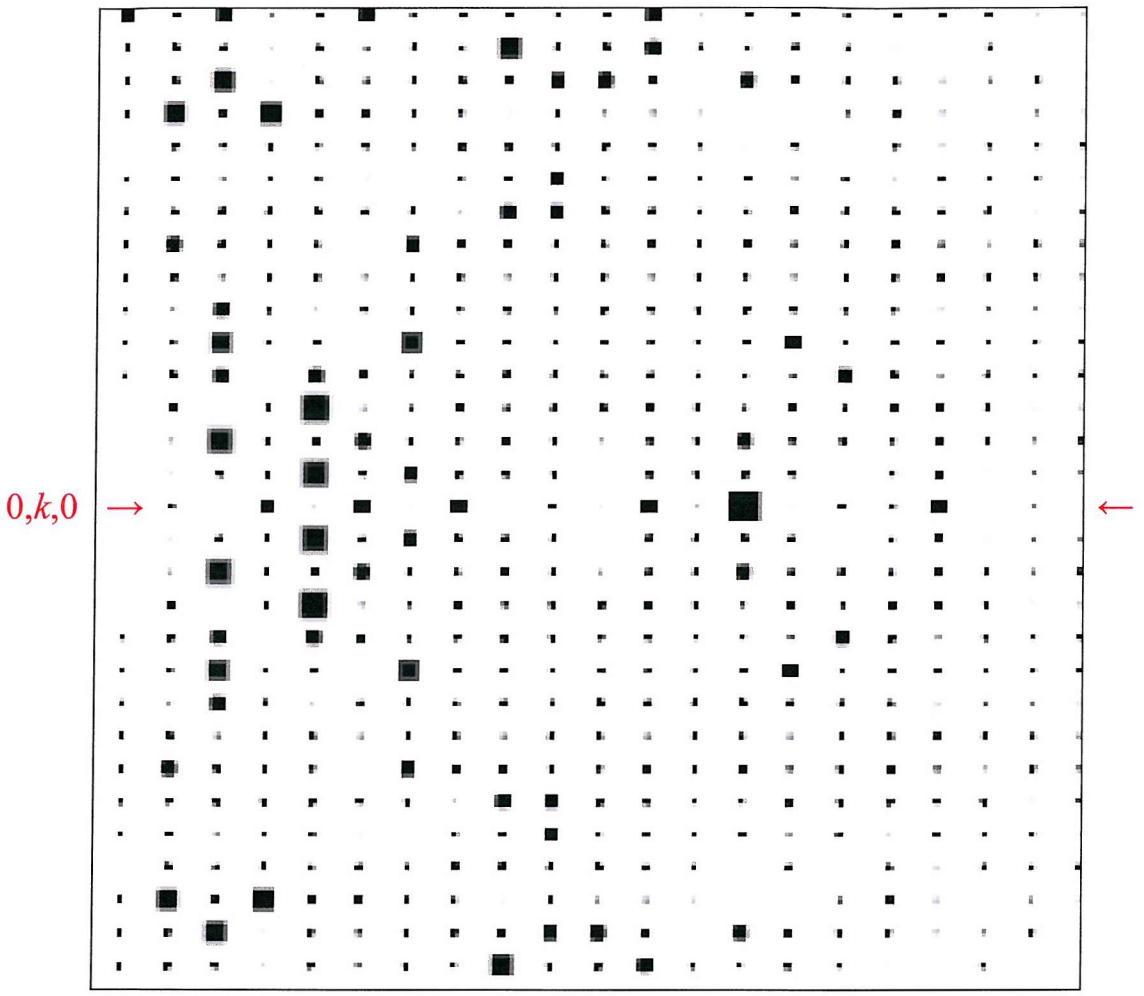


Figure 3.20. A pseudo precession picture displaying systematic absences along axis 0,*k*,0 (*k*≠2n) within the 0,*k*,*l* zone, implying a 2_1 screw axis along b for SAP-Ro 63-8695 soaked data processed in P2 (prepared using HKLView).

(Berman *et al* 2000) entry code 1SAC) was used as the search model with the calcium ions and acetate molecules omitted. The rotation function was using reflections within the range of 30 to 2.6Å with a radius of integration of 42.55Å. Several orientations produced peaks, with the two largest corresponding to symmetry related orientations (see table 3.10). The translation function was performed with the best 4 orientations, using reflections within the range of 30 to 2.6Å. The top two orientations both produced a large peak, but within different asymmetric units. The second orientation (α=172.73°, β=72.61°, and γ=151.45°) produced a slightly larger peak of 31σ (27σ above the next largest) corresponding to a fractional shift of x=0.228, y=0.000, and z=0.426 (see table 3.11). This correctly orientated and translated phasing model gave a resultant *R-factor* of 35.7%, correlation coefficient of 76.4%, and showed good sensible packing contacts between symmetry related molecules (see figure 3.21).

3.3.4 Refinement & Model Building

Crystallographic refinement was performed using CNS, with a test set comprising of 10% of the reflections randomly flagged to calculate the *R-free* value. The model was refined against the data in 9 stages. Between each round of refinement the improved model was used to calculate $|F_{obs}|$-$|F_{calc}|$ and $2|F_{obs}|$-$|F_{calc}|$ sigmaA-weighted electron density maps for visualisation and manual model building within Quanta and XtalView. Also, throughout the refinement bulk solvent correction was applied and the stereo-chemical properties of the model monitored using WHAT IF and Procheck. A summary of the *R-factor* and *R-free* values during the refinement are shown in figure 3.22.

Table 3.10. Top 5 rotation function peaks calculated with Molrep.

Peak Number	Rotation			RF/σ
	α (°)	β (°)	γ (°)	
1	2.87	35.33	335.15	8.28
2	172.73	72.61	151.45	8.27
3	184.73	90.00	158.04	8.07
4	178.96	90.00	142.74	7.92
5	215.20	90.00	192.36	7.42

Table 3.11. Top 5 translation function peaks calculated with Molrep, using the second orientation from the rotation function search (α=172.73°, β=72.61°, and γ=151.45°).

Peak Number	Shift Fraction			Dens/σ
	x	*y*	*z*	
1	0.228	0.000	0.426	30.83
2	0.236	0.000	0.405	3.57
3	0.221	0.000	0.448	3.30
4	0.465	0.000	0.478	3.01
5	0.037	0.000	0.071	2.73

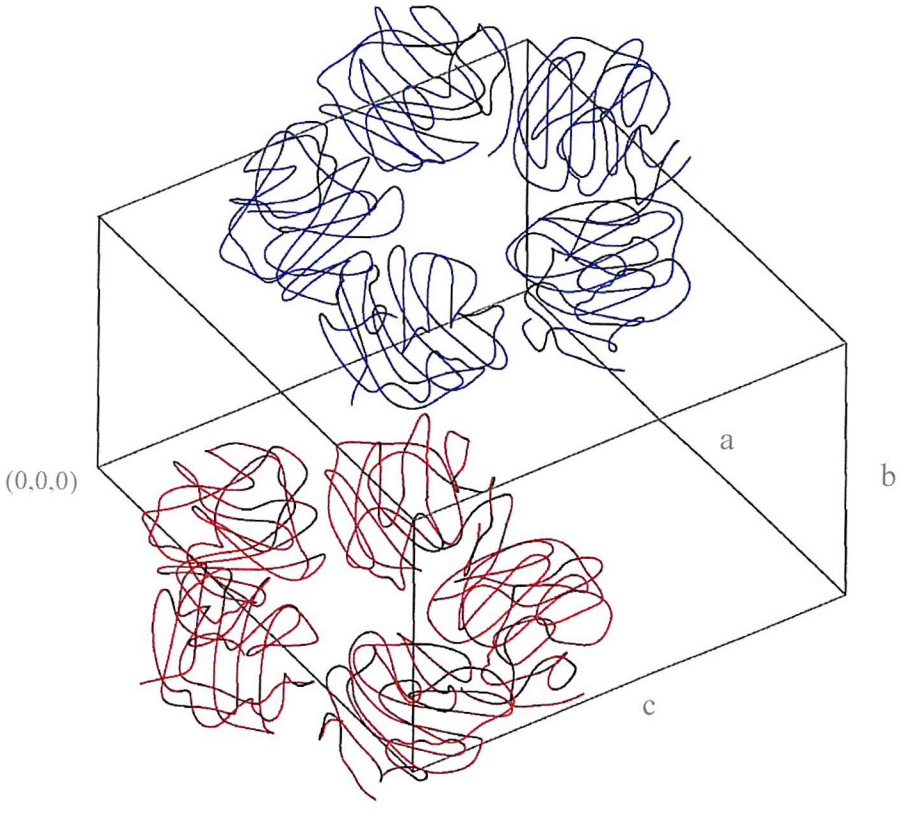
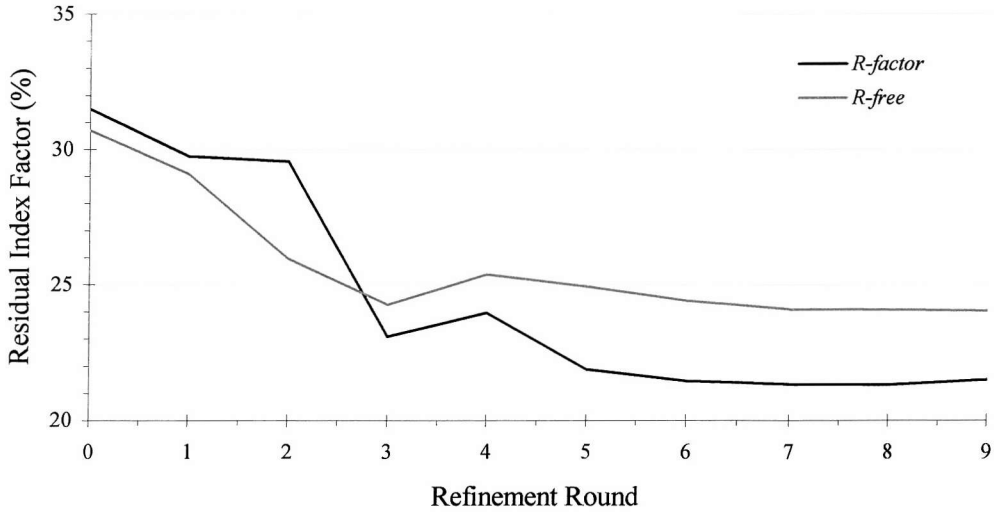


Figure 3.21. The crystal packing of SAP-PC soaked with Ro 63-8695 (prepared using Bobscript, GL Render, and Raster3D). In the presence of PC, SAP crystallises in the monoclinic space group $P2_1$ (a=96.1Å, b=70.9Å, c=103.7Å, and β=96.8°) with one SAP pentamer in the asymmetric unit. SAP pentamers are depicted in different colours with the axes and origin of the unit cell labelled accordingly.

Figure 3.22. Summary of the refinement process for SAP soaked with Ro 63-8695.



The model was initially refined by several cycles of rigid-body refinement with the pentamer followed by the protomers treated as the rigid bodies (round 1), during which, the resolution was progressively increased from 6 to 2.55Å and 3 to 2.55Å respectively. This reduced the *R-factor* slightly from 31.5% (*R-free* 30.7%) to 29.7% (*R-free* 29.1%). Examination of the resulting electron density maps revealed well-defined density for two calcium ions and a ligand clearly resembling the head group of Ro 63-8695, rather than that of PC (see figure 3.23), at the ligand-binding site of each subunit. In order to prevent protomer deviations in the phasing model biasing the refinement, one subunit was superimposed onto the other four. Simulated annealing was then performed with torsion angle dynamics (slow cool from 5,000 to 500K in 50K decrements) and high weighted protomer (5-fold) NCS restraints (round 2), using reflections within the range of 30 to 2.55Å. At this point the side chains of 11 solvent exposed residues were found to have indiscernible orientations, and were mutated to alanine. The NCS restraints were then

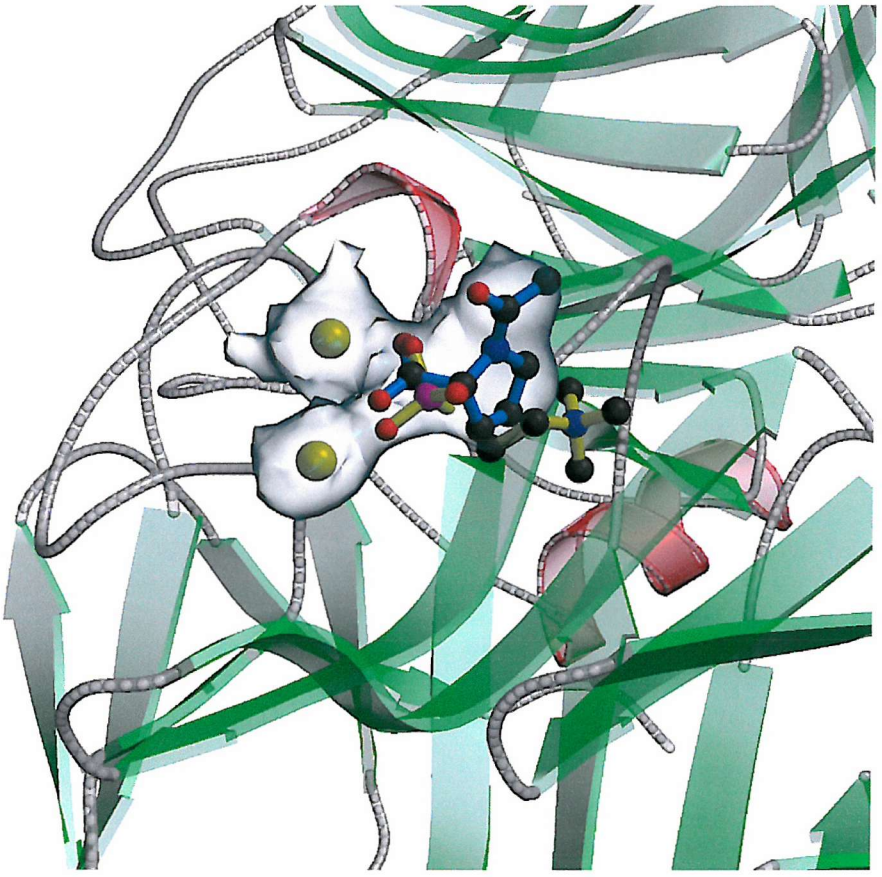


Figure 3.23. Three-dimensional structure of SAP-PC soaked with Ro 63-8695 illustrating the electron density for two calcium ions and the replacement of PC (yellow) with Ro 63-8695 (blue) at the ligand-binding site of subunit D (prepared using Bobscript and Raster3D). The $|F_{obs}|$-$|F_{calc}|$ difference density is represented as a transparent surface (contoured at 2σ) and clearly resembles the atomic positions of the head group of Ro 63-8695 over that of PC. Secondary structural elements are represented as previously described with the head group of Ro 63-8695 and PC shown as ball-and-stick models.

relaxed and individual isotropic B-factor refinement performed (round 3). The resulting electron density maps showing better-pronounced density for five of the mutated side chains, which were subsequently mutated back and built into the electron density. The model was then refined by isotropic B-factor refinement (rounds 4), followed by positional and isotropic B-factor refinement (round 5). Between subsequent rounds of isotropic B-factor refinement the calcium ions were built into density at the double calcium-binding sites (rounds 5-6), followed by the head group of each Ro 63-8695 molecule (rounds 6-7) and the remaining 6 residues mutated back (rounds 6-9). Due to the resolution of the data and the low observation/parameter ratio (1.2:1) the refinement was stopped prior to the inclusion of water molecules and sugar residues.

The final round of individual isotropic B-factor refinement gave an *R-factor* of 21.5% (*R-free* 24.0%) with an average B-factor of 33.9Å^2. The resulting model showed good stereo-chemical properties with no residues within the disallowed regions of the Ramachandran plot (see figure 3.24). The overall statistics for the fully refined structure are summarised in table 3.12.

3.3.5 Structure Analysis of SAP Soaked with Ro 63-8695

The structure of each SAP protomer is well defined with a good B-factor distribution (see figure 3.25) (subunits distinguished by labelling clockwise from A to E when viewed from the B-face). There are only minor structural deviations located within some of the protomers when compared to that of SAP-N-acetyl-D-proline (see table 3.3). These are localised to weak β-turn regions where a small change in one of the dihedral angles is sufficient to alter a type VIII β-turn (residues 48-51) to a type IV in four of the subunits (A, C, D, and E), and a type IV (residues 57-60) to a type II in two of the subunits (A and B). A third β-turn (residues 99-102) is not altered to a different type, but is

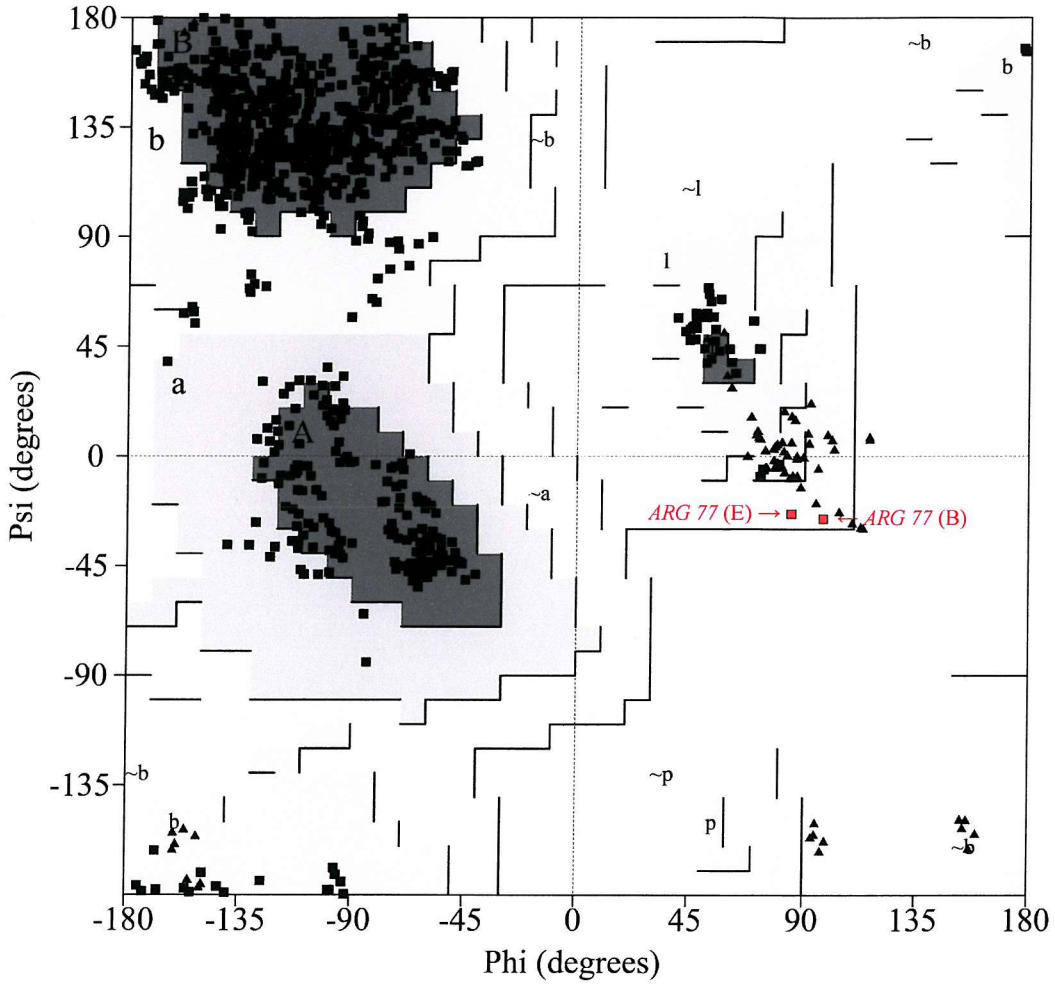


Figure 3.24. Ramachandran plot for the final SAP-Ro 63-8595 soaked structure (prepared using Procheck). Statistics for non-glycine and non-proline residues (875 residues) indicate 83.5% of residues in most favoured regions [A, B, L], 16.2% residues in additional allowed regions [a, b, l, p], 0.2% residues in generously allowed regions [~a, ~b, ~l, ~p], and 0.0% residues in disallowed regions. Glycine residues are represented as black triangles, with all other residues as black squares with the exception of the 2 residues within the generously allowed region (Arg77 in subunits B & E) which are highlighted in red.

Table 3.12. Summary of the processing and model statistics for the SAP-Ro 63-8695 soaked structure.

Space Group	$P2_1$ (unique axis along b)
Unit Cell	a=96.1Å b=70.9Å c=103.7Å
	β=96.8°
Unit Cell Volume (Å^3)	701,588
Resolution Range (Å)	30 - 2.55
Total Number of Reflections	62,927
Unique Number of Reflections	38,346
Multiplicity	1.6
Mosaicity (°)	0.7
Completeness (%)	84.3 (79.3)
R-merge(I) (%)	13.3 (30.4)
I/σ(I) (%>2.0)	81.2 (51.7)
Molecules in Asymmetric Unit	1 SAP Pentamer
	10 Calcium Ions
	5 Ro 63-8695 Molecules
Solvent Content (%)	55.4
Model *R-factor* (%)	21.5
Model *R-free* (%)	24.0
rms Bond Lengths (Å)	0.010
rms Bond Angles (°)	1.51

Figures in parentheses denotes the highest resolution shell statistics between 2.64-2.55Å

stabilised in all of the subunits by the formation of a hydrogen bond between residues *i* and *i*+3. The B-factor distribution for the main chain is similar to that of the SAP-N-acetyl-D-proline (see figure 3.11 and figure 3.25). However, due to the resolution and quality of the data the average thermal disorder is increased by 12.3Å². This is reflected in the β-strands where the average B-factor is 30.5Å². As with SAP-N-acetyl-D-proline, the subunits show remarkable similarity which when superimposed have average C_α rms fit of 0.14Å, with only one major deviation between strands B and C in subunit D (see figure 3.26). The deviation is localised to residues 25-28, which when omitted from the C_α rms calculation improves the fit to 0.11Å. But unlike SAP-N-acetyl-D-proline, the inverse γ-turn (residues 24-26) located within the deviating region is maintained as is the nearby type IV β-turn (residues 181-184).

The pentameric assembly is stabilised by the same direct interactions as previously described for SAP-N-acetyl-D-proline (see table 3.7). Although water molecules have not been included within the refinement, electron density is present at locations corresponding to the three coordinating water molecules. The crystal packing is also similar, involving the same residues and the majority of the same interactions (see table 3.8). Although additional salt bridges are formed between Glu126-Lys28 (subunit B-D contact) and Asp184-Arg57 (subunit C-D contact), while the His78-Thr176 hydrogen bond is replaced by an Arg120-Thr176 (subunit E-E contact). Despite the use of different phasing models the similarity between equivalent subunits of N-acetyl-D-proline and Ro 63-8695 soaked pentamers is very high, showing only minor variations in structural and packing interaction with a C_α rms fit of 0.11Å when the pentamers are superimposed (see figure 3.27 and 3.28).

Calcium coordination is well defined with low thermal motion of the bound ions, average B-factors 40.4Å² (calcium I) and 44.8Å² (calcium II). The calcium ions are positioned 4.0 to 4.1Å apart with average calcium-ligand interactions between 2.34 to

Figure 3.25 Comparison of average isotropic B-factors for the main chain atoms of different SAP subunits (subunits labelled clockwise from A to E when viewed from the B-face).

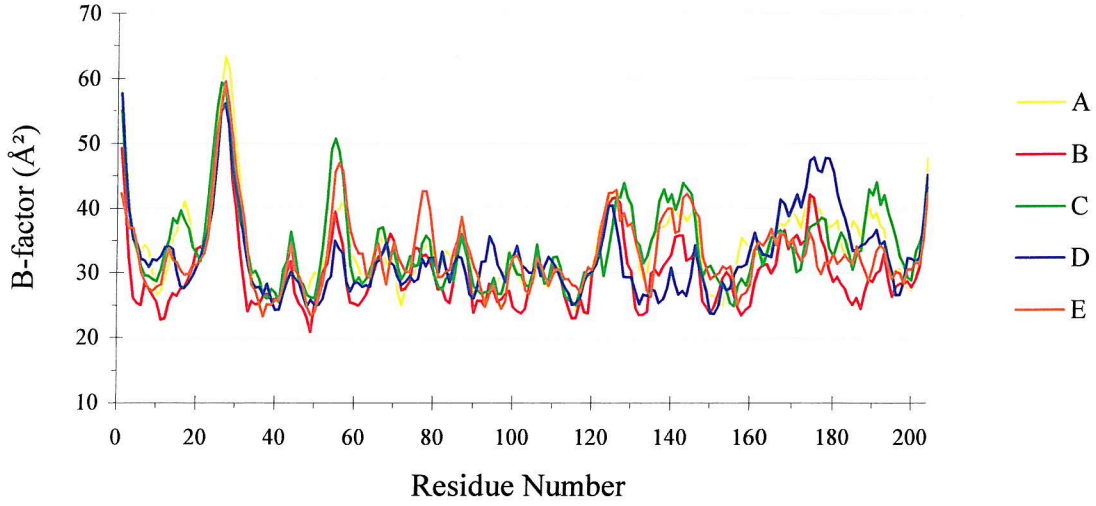
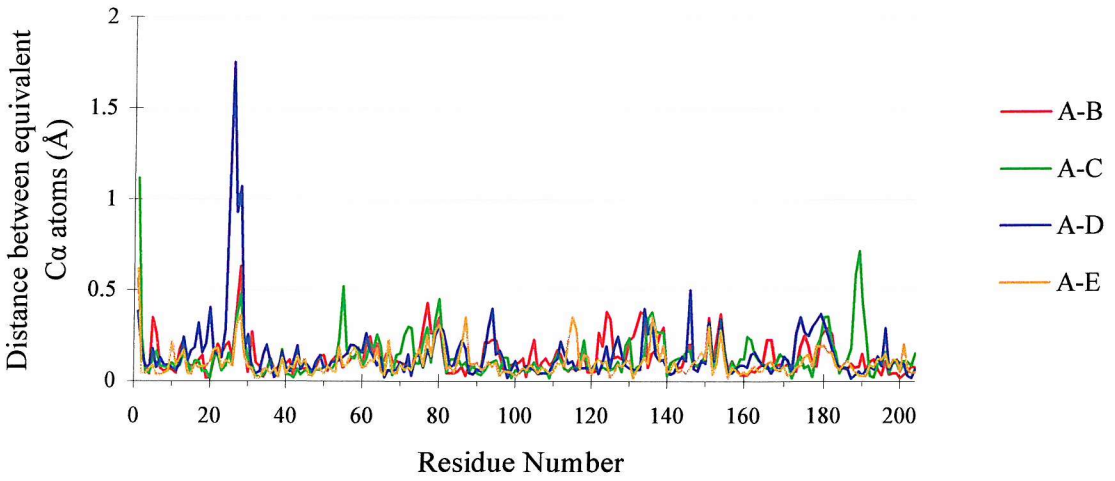


Figure 3.26. C_α distances between superimposed subunits of SAP (subunits labelled clockwise from A to E when viewed from the B-face).



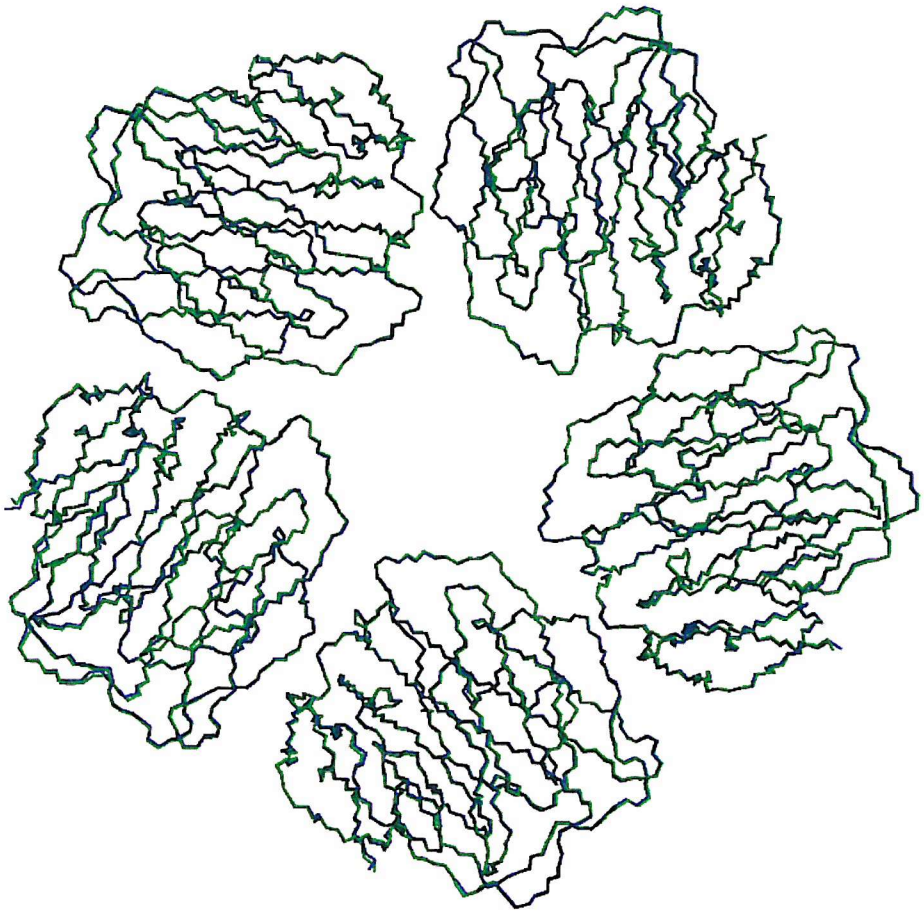
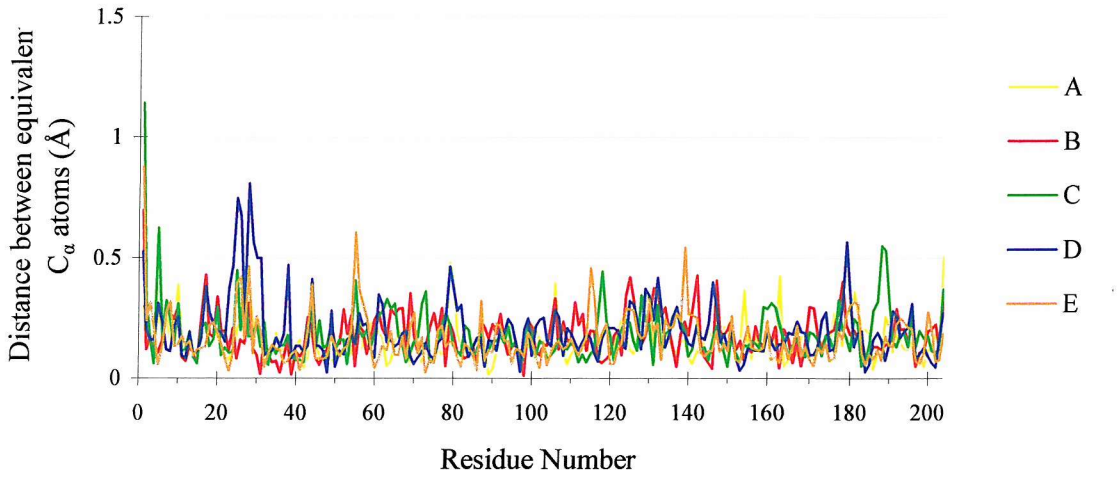


Figure 3.27 Superposition of SAP pentamers derived from the Ro 63-8695 soaked (green) and N-acetyl-D-proline (blue) crystal forms (prepared using Bobscript and Raster3D). The rms deviation is 0.11Å for equivalent C_α atoms when the equivalent subunits are superimposed.

Figure 3.28. C_α distances between equivalent subunits of superimposed Ro 63-8695 soaked and N-acetyl-D-proline pentamers (subunits labelled clockwise from A to E when viewed from the B-face).



2.65Å (see table 3.13). Although not included within the refinement, electron density is present at the corresponding positions for the two water molecules at calcium site II (see figure 3.29). As with N-acetyl-D-proline the carboxyl of Ro 63-8695 completes the coordination by bridging the two calcium ions, forming hydrogen bonds with Asn59 and Gln148 and orientating the pyrrolidine ring into the adjacent hydrophobic pocket. Electron density for the head group is well defined up to and including the peptide bond (average B-factor 45.0Å²). The Ro 63-8695 molecules bound to subunits A, B, and C have indistinguishable cis/trans isomeric states. However, in subunits D and E the trans-isomer appears to be dominant with weak density present for the first carbon atom of the linker (see figure 3.29), although a water molecule hydrogen bonding to the carbonyl in the cis-isoform could also occupy the same space and should therefore not be discounted. The binding and insertion of the (R)-pyrrolidine-2-carboxylic acid group is identical to that of

Table 3.13. Inter-atomic distances for the coordination of the double calcium-binding site (averages taken from the distances of all five subunits).

Calcium Ion	Ligand	Ligating Oxygen	Average Inter-atomic Distance (Å)
CaI	Asp58	OD1	2.50
	Asp58	OD2	2.58
	Asn59	OD1	2.39
	Glu136	OE2	2.39
	Gln137	O	2.57
	Asp138	OD1	2.44
	Ro 63-8695	O1	2.50
CaII	Glu136	OE1	2.34
	Asp138	OD1	2.65
	Asp138	OD2	2.63
	Gln148	OE1	2.57
	Ro 63-8695	O2	2.48

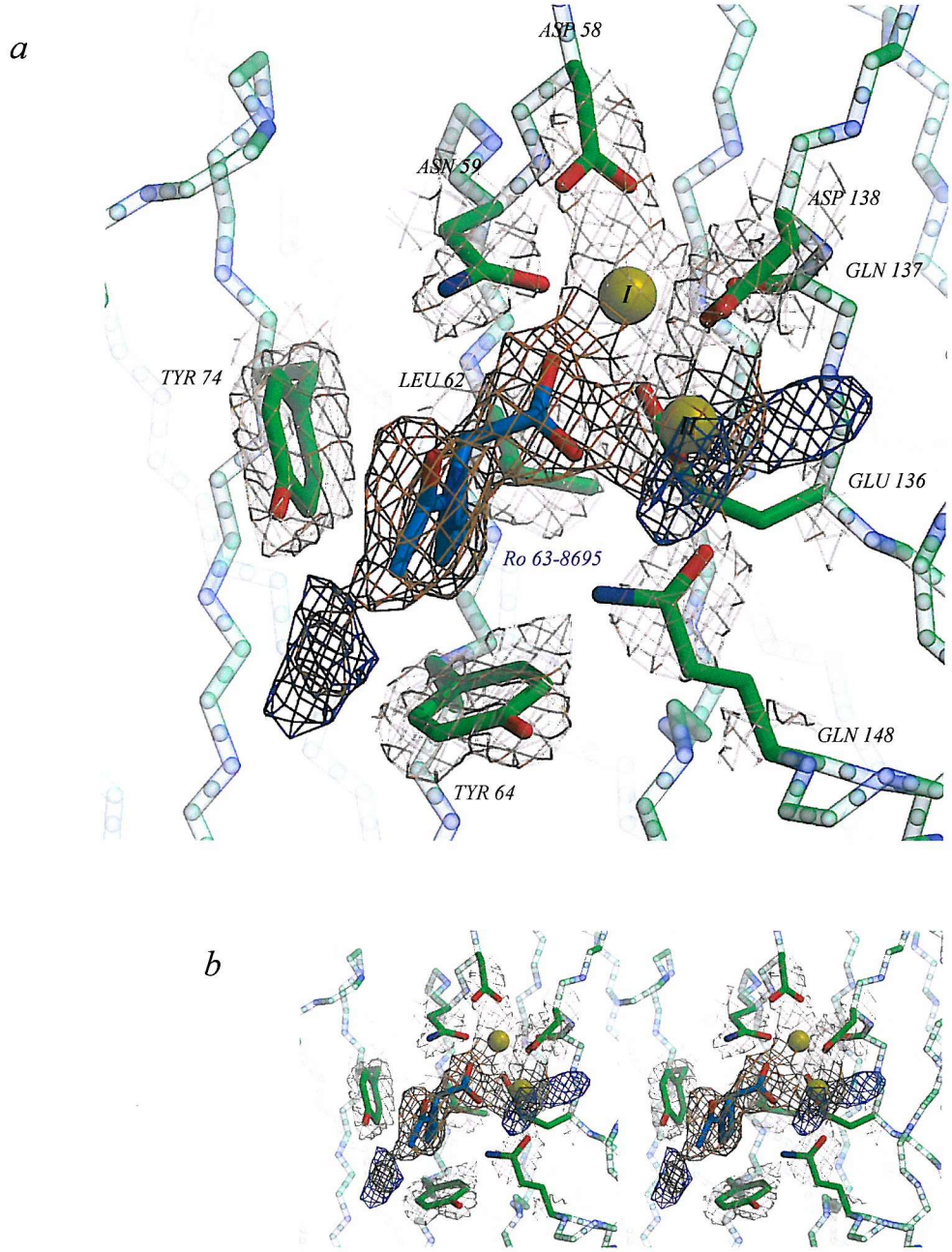


Figure 3.29 Three-dimensional structure of the double calcium-binding site in subunit D of the SAP Ro 63-8695 soaked model viewed as an orthogonal projection (*a*) and stereoview (*b*) (prepared using Bobscript and Raster3D). The $2|F_{obs}|-|F_{calc}|$ electron density surrounding the site is well defined at 1.3σ (depicted in orange) and shows weak but continuous density for the first carbon of the linker of the trans-isomer. The $|F_{obs}|-|F_{calc}|$ Difference density contoured at 2σ (depicted in blue) is also present for the linker carbon as well as the ligating water molecules at calcium site II. The main chain is represented as a C_α trace, calcium ions as yellow spheres, with the Ro 63-8695 (blue) and important side chains (green) shown as stick models and labelled accordingly.

N-acetyl-D-proline (see figure 3.30). Although the insertion distance does appears to be on average 0.2Å smaller (C_β 3.9Å away from Leu62) this is well within the experimental error of this resolution.

3.4 Structure Comparison with 1SAC

There is a remarkably high degree of similarity in tertiary and secondary structure between the SAP protomers from the isomorphous crystals previously presented in this chapter. This similarity also extends to the original 2.0Å structure (1SAC) that shows a comparable C_α rms fit of 0.11Å between the subunits, and a fit of 0.14Å and 0.13Å when compared to equivalent subunits from N-acetyl-D-proline and Ro 63-8695 soaked pentamers respectively. The majority of the structural deviations between these non-isomorphous crystal forms are due to the differing crystal contact regions and solvent constituents (particularly the acidic pH of 5.5), with the only major main chain deviation occurring within the loop region between strands B and C of subunit D. Examination of the secondary structure of 1SAC reveals only minor differences within weak β-turn and γ-turn structures in some of the subunits when compared to subunit A of the SAP-N-acetyl-D-proline structure (see table 3.14).

Interestingly, the pentameric assembly of 1SAC shows a minor deviation in subunit arrangement with the C_α rms fit increased to 0.22Å when the pentamer as a whole is superimposed onto the N-acetyl-D-proline or Ro 63-8695 soaked pentamers. This reduced fit is localised to subunits A, B and C, indicating that a small degree of freedom (~1-2° rotation) is present between the subunits without any loss or disruption to the subunit interactions.

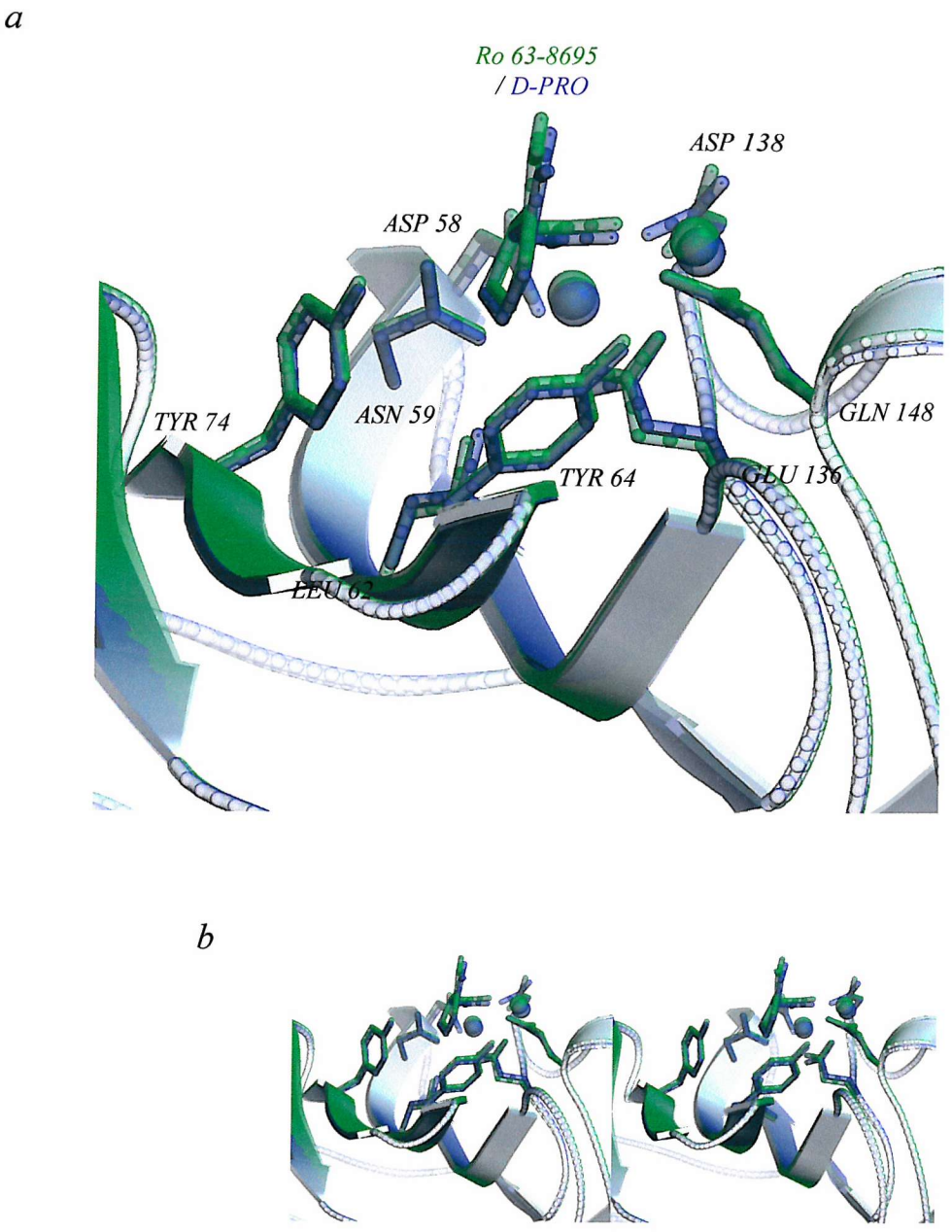


Figure 3.30 Superposition of the ligand-binding sites from subunit D of the Ro 63-8695 soaked (green) and N-acetyl-D-proline (blue) crystal forms viewed as an orthogonal projection (*a*) and stereoview (*b*) (prepared using Bobscript and Raster3D). Secondary structural elements are represented as previously described with calcium ions as spheres, important residues and molecules shown as stick models and labelled accordingly.

Table 3.14. Structural differences between subunit A of SAP-N-acetyl-D-proline (see table 3.3) and the original 2.0Å SAP structure (1SAC) (subunits labelled clockwise from A to E when viewed from the B-face).

N-acetyl-D-proline	1SAC	Residue Range	Subunit (1SAC)
γ-Turn Inverse	Random Coil	24 - 26	D
β-Turn Type VIII	β-Turn Type IV	48 - 51	B, C, & D
β-Turn Type VIII	β-Turn Type IV	59 - 62	D
β-Turn Type IV	β-Turn Type VIII	181 - 184	B & E
γ-Turn Inverse	Random Coil	194 - 196	C, D, & E

Contrary to published descriptors (see section 1.5.2) the double calcium-binding sites of 1SAC show heptagonal coordination geometries similar to those previously presented in this chapter. Calcium-ligand interactions are also within the same range (2.44 to 2.78Å) with the previously unreported Asp138 ligand an average 2.7Å away from calcium II. Although similarities in solvent coordination cannot be examined due to the absence of water molecules in the 1SAC coordinate file, the coordination geometry of the acetate ion is identical to the carboxyls of N-acetyl-D-proline and Ro 63-8695. The pyrrolidine rings show no significant disruption to the surrounding tyrosine side chains when compared to that of 1SAC.

3.5 Discussion

Although solved in 1994, this is the first time a detailed secondary structure assignment has been performed upon SAP. This has shown both the tertiary and secondary

structure to be extremely stable. Although exterior loop regions do show a small degree thermal disorder, their turn regions are well maintained in the absence of external interactions with an overall C_α rms fit of ~0.11Å when crystal contact deviations are taken into account. Furthermore, the small degree of flexibility coupled with the protruding polar side chains of Glu27, Lys28, Arg57, Arg77, Lys143, and Arg146 within these loop regions may facilitate the binding of larger biological ligands in a manner similar to the crystal packing interactions. In fact the identification of protruding arginine residues (see figure 3.31) strongly indicates that DNA binding maybe comparative to the histone octamer, where DNA is bound through the phosphate backbone via the insertion of arginine residues into the minor grove (Harp *et al* 2000) (see figure 3.32). The spatial arrangement of Arg57 and Arg77 upon the exterior rim of the B-face of SAP suggests that a pair of neighbouring subunits may bind two turns of DNA. Although the equivalent arginine residues are positioned ~60Å (radial distance) apart, Arg57 and Arg77 are positioned within 8Å (C_α distance) of each other creating the 68Å distance required for two full turns of DNA. Arg146, however, is positioned near the subunit interface and appears to fit the suggestion that DNA binding may occur at the basic groves between the subunits (Hohenester *et al* 1997; Pepys *et al* 1997).

Careful examination of the double calcium-binding sites from the crystal structures presented previously has identified a fourth protein ligand at calcium II giving the site a heptagonal coordination. Although previously overlooked in the original structure, this does not significantly alter the asymmetry between the two sites. However, the increased stability of three calcium I ligands, via the neighbouring Asn53 and the coordination of an underlying water molecule, suggest that the side chain dispositions of site I may be at least partially maintained in the absence of calcium, while site II will be almost completely disrupted by the destabilisation of the loop region between strands K and L. Therefore

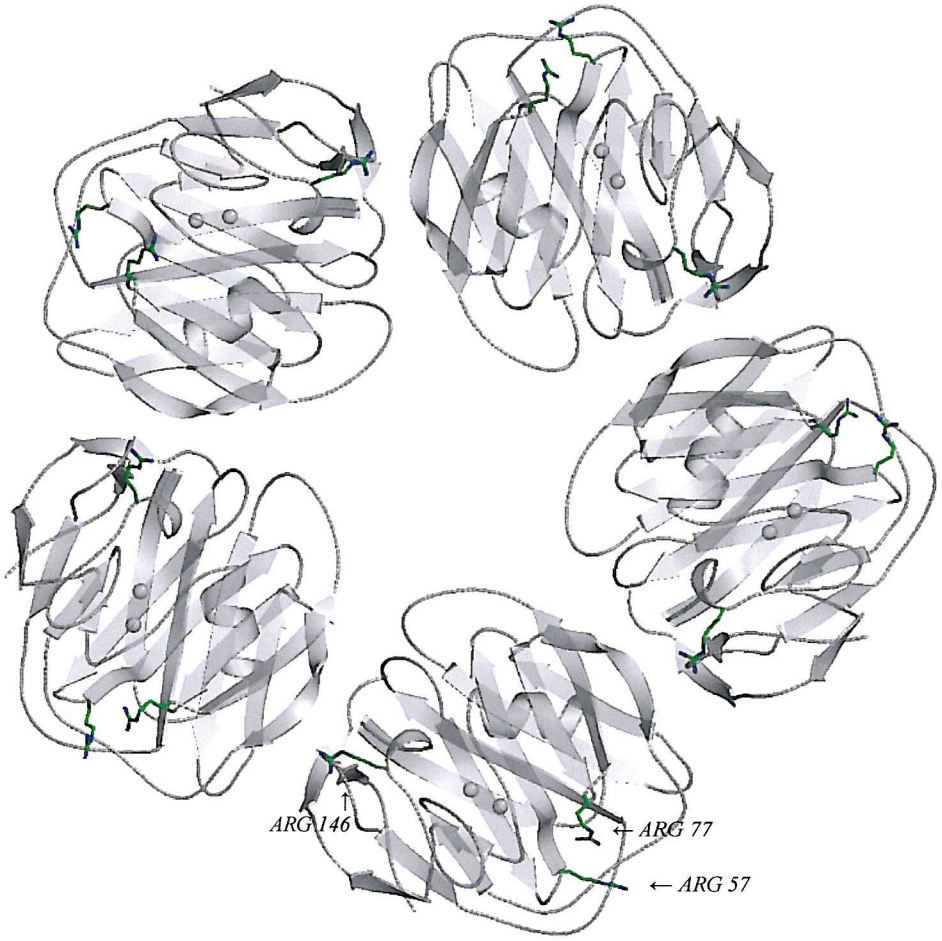


Figure 3.31. Three-dimensional structure of SAP illustrating the spatial arrangement of arginine residues potentially involved in DNA binding (prepared using Bobscript and Raster3D). Arg57, Arg77, and Arg146 are located upon the B-face suggesting DNA binding via either the groves created between the subunits with Arg146 inserting into the minor grove of DNA, or around the exterior of the pentamer involving an Arg57 and Arg77 from neighbouring subunits. Secondary structural elements are represented in grey as previously described with calcium ions as spheres and important arginine residues as green stick models.

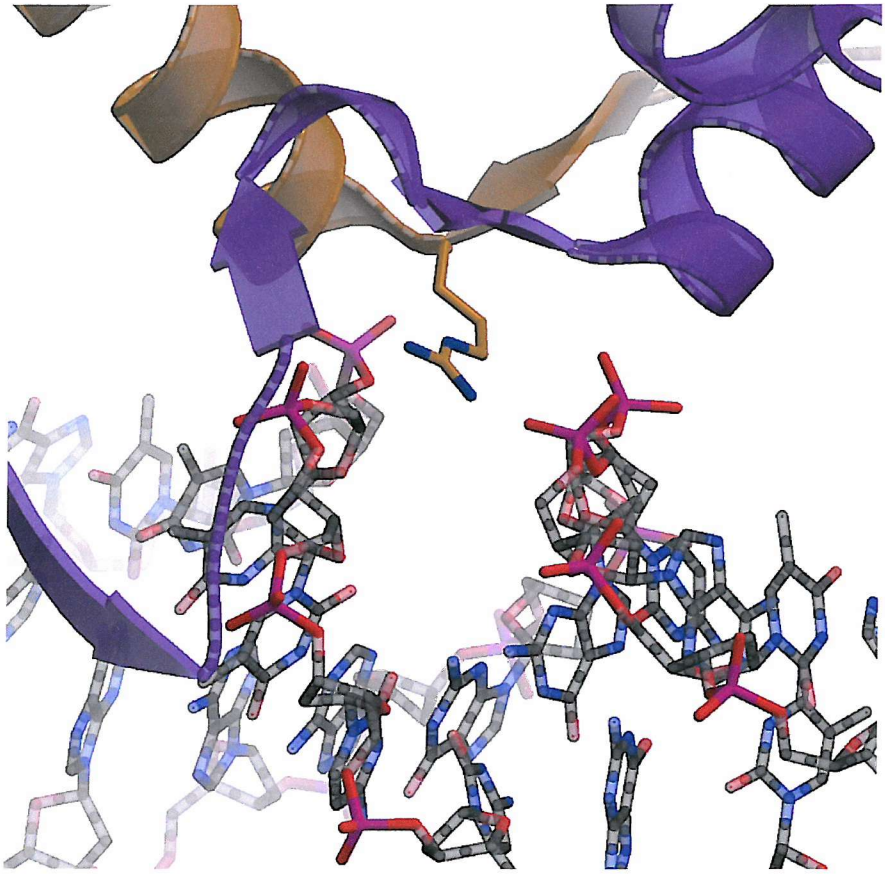


Figure 3.32. Three-dimensional structure of histone H3 (orange) and H4 (purple) interacting with DNA, illustrating the insertion of an arginine residue into the minor groove of the DNA backbone (prepared using Bobscript and Raster3D). Secondary structural elements are represented as previously described with DNA and the important arginine side chain represented as stick models.

calcium binding at site I will facilitate the creation of site II via calcium-coordinated stabilisation of the adjacent loop region.

The novel binding of the Ro 63-8695 head group to SAP via calcium ligation and pyrrolidine ring insertion has recently been measured by isothermal colorimetry. N-acetyl-D-proline has a dissociation constant K_d of 15μM compared with 35μM for PE and 50μM for MOβDG (Kolstoe 2001). The binding of PE and MOβDG is driven by the enthalpy of calcium ligation and hydrogen bonding, but also contains a significant entropy contribution from the burying of the hydrophobic pocket and expulsion of water molecules (71.9Å²). Interestingly, although the Ro 63-8695 complex contains only two hydrogen bonds, the binding affinity for the head group is slightly higher. This is primarily due to the larger entropy contribution from the insertion of the pyrrolidine ring into the hydrophobic pocket, burying 143.8Å² (100.5J/Å² (Chothia 1975)) of the accumulative hydrophobic surface area. There are also a large number of 'weak' Van der Waals' attractions (1.52-0.19J/mol (Jordan 1973)) between carbon and hydrogen atoms from the pyrrolidine ring and the surrounding tyrosine and leucine side chains. Furthermore in addition to the loss of rotational and translational freedom of Ro 63-8695 during binding (290-420kJ/degree of freedom/mol (Chothia & Janin 1975)), there is only a small loss in conformational entropy due to the rigidity of the head group. The only bond rotation restricted during binding is about the carboxyl to C_α atom.

Hydrophobic binding of the pyrrolidine ring is highly specific, with the pocket only able to accommodate the R-isomer of Ro 63-8695 (or D-isomer of N-acetyl-D-proline). In fact, the binding of a pyrrolidine ring strongly suggests that a correctly orientated proline residue maybe involved in protein recognition at the ligand-binding site, providing the region is sufficiently protruding away from the body of the protein. Protruding regions are generally comprised of non-repetitive structures, in particular, β-turns which favour

glycine, proline, asparagine, and aspartic acid residues (see table 3.15), with differing positional preferences depending upon the turn types (see table 3.16 for proline, and table 3.17 for aspartic acid) (Hutchinson & Thornton 1994). Model building of β-turns containing proline at the preferential *i*+1 position in the common type I, II, and VIII β-turns showed that the pyrrolidine ring could be easily docked into the hydrophobic pocket with the C_γ directed down towards Leu62 (see figure 3.33). Although stabilised through hydrogen bonding of the main chain with the surrounding side chains of Asn59, Tyr74, and/or Gln148 (depending upon the direction of the chain), this would result in the loss of two calcium ligands, thus altering the coordination of both ions to a hexagonal arrangement. Insertion of the pyrrolidine ring when proline is located at the *i* or *i*+3 position is possible although highly dependent upon the dihedral angles and side chain size of the preceding (*i*-1) or proceeding (*i*+4) residue respectively. Rare type VI β-turns requiring a cis-proline at the *i*+2 position can also be docked with the formation of similar hydrogen bonds. In contrast to proline, β-turns containing aspartic acid residues can be docked via ligation of the double calcium-binding site (see figure 3.33). The protruding carboxyl group of aspartic acid allows more flexibility in the positional arrangement, due to the location of the main chain further above the ligand-binding site. In the majority of cases, the main chain carbonyl groups are within hydrogen bonding distance of the Asn59 and/or Gln148 amino groups. Positional potentials for proline and aspartic acid residues suggest a significant number of type I and type VIII β-turns will contain both proline and aspartic acid residues enabling both pyrrolidine insertion and calcium ligation (see table 3.16 and 3.17). For type I β-turns with proline at the *i*+1 position, pyrrolidine insertion and calcium ligation is only achievable when aspartic acid precedes the proline (see figure 3.34). Interestingly asparagine, aspartic acid, threonine, and serine are favoured at the proceeding *i*+2 position, which would facilitate stabilisation of the type I β-turn via

Table 3.15. Overall β-turn potentials for each residue (Hutchinson & Thornton 1994).

Residue	β-Turn Potential	Residue	β-Turn Potential	Residue	β-Turn Potential
Ile	0.59	Gly	1.48 *	Glu	1.01
Phe	0.89	Cys	1.08	Asn	1.44 *
Val	0.70	Tyr	0.92	Gln	0.94
Leu	0.66	Pro	1.38 *	Asp	1.41 *
Trp	0.70	Thr	1.00	Lys	1.01
Met	0.57	Ser	1.15	Arg	0.82
Ala	0.83	His	1.07		

* Denotes significantly high potentials

hydrogen bonding with the surrounding side chains. For example, the hydroxyl group from a threonine or serine residue would lie within hydrogen bonding distance of the amino group of Gln148 (see figure 3.34). Due to the dihedral angles of the type VIII β-turn both pyrrolidine insertion and calcium ligation is not achievable at the same time. These results suggest that proline and aspartic acid containing β-turns maybe involved in the recognition of protein structures by SAP, particularly with respect to amyloid fibrils that are believed to contain a large number of exposed turns and loop structures along the axis.

Table 3.16. Type-dependent positional potentials for proline residues at each of the 4 positions (*i*, *i*+1, *i*+2, and *i*+3) of β-turns (Hutchinson & Thornton 1994).

β-Turn Type	Proline Position			
	i	*i*+1	*i*+2	*i*+3
I	1.31 *	3.49 *	0.21	0.03
I′	0.50	0.00	0.00	0.00
II	1.99 *	4.91 *	0.00	0.00
II′	0.48	0.00	1.19	0.00
VIII	2.74 *	2.09 *	0.00	4.06 *

* Denotes significantly high potentials

Table 3.17. Type-dependent positional potentials for aspartic acid residues at each of the 4 positions (*i*, *i*+1, *i*+2, and *i*+3) of β-turns (Hutchinson & Thornton 1994).

β-Turn Type	Aspartic Acid Position			
	i	*i*+1	*i*+2	*i*+3
I	2.51 *	1.20	2.54 *	1.11
I′	0.94	2.58 *	0.54	0.82
II	0.34	1.02	0.51	0.89
II′	1.16	0.77	1.93 *	0.38
VIII	0.63	1.96 *	1.75 *	0.80

* Denotes significantly high potentials

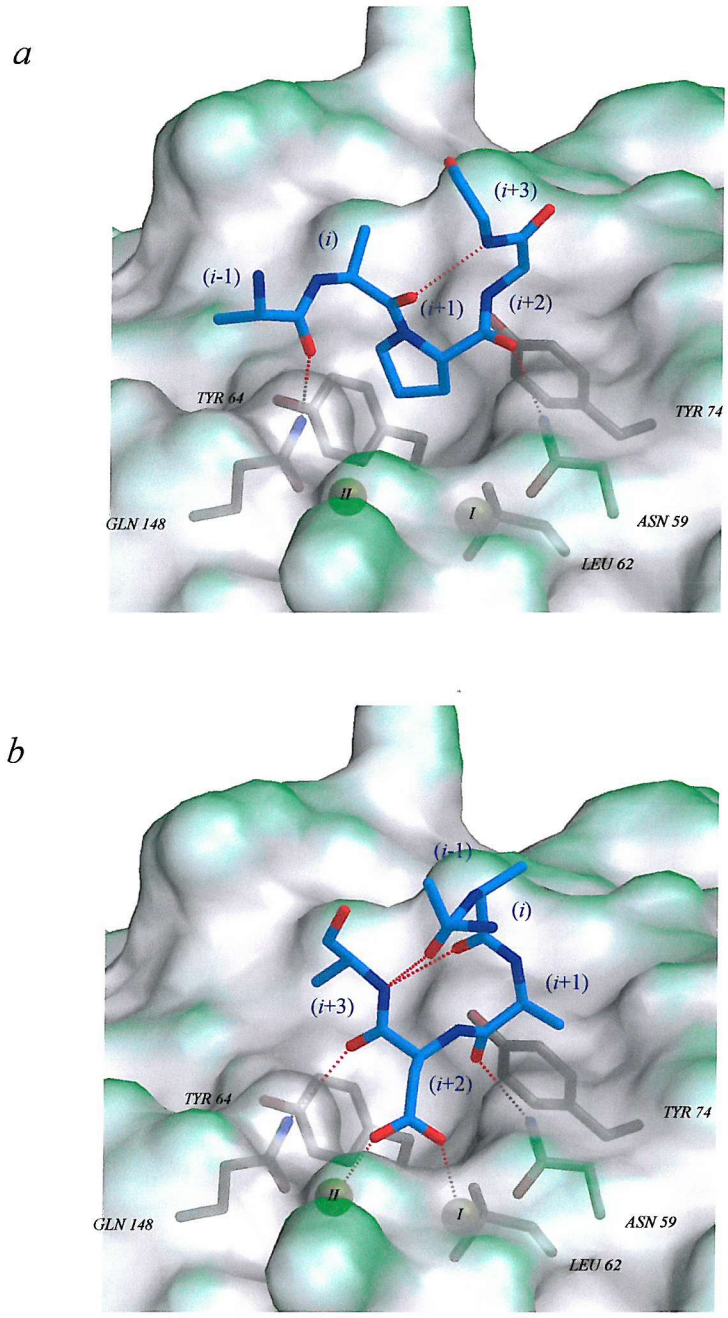


Figure 3.33 Examples of type I β-turns containing proline (*a*) or aspartic acid (*b*) residues modelled into the ligand-binding site of SAP (prepared using Bobscript, Grasp (Nicholls *et al* 1991), GL Render, and Raster3D). The type I β-turn containing proline at the *i*+1 position is docked via insertion of the pyrrolidine ring into the hydrophobic pocket, while the type I β-turn with aspartic acid at *i*+2 is docked at the double calcium-binding site via the carboxyl group. Both loops are stabilised by hydrogen bonding through the main chain carbonyls with Asn59 and Gln148. Surface contouring of SAP is shown with the underlying calcium ions represented as yellow spheres and important residues as stick models. The β-turns are shown as blue stick models with the interactions represented as red dotted lines.

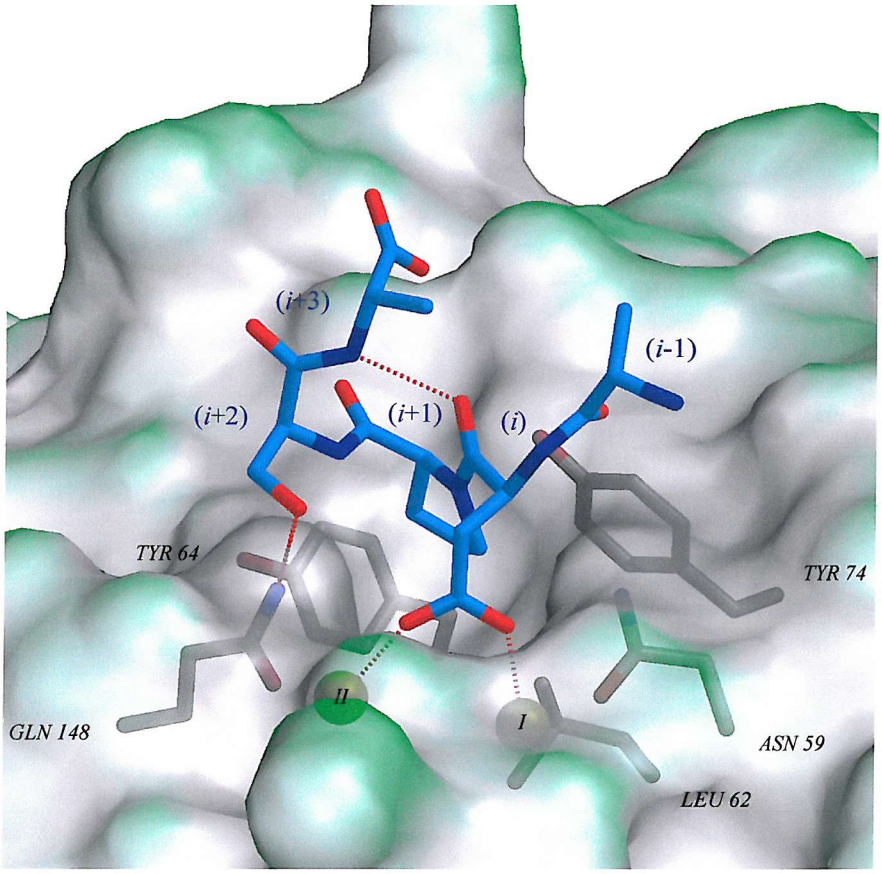


Figure 3.34 Model of an Asp(*i*)–Pro(*i*+1)-Ser(*i*+2) type I β-turn docked into the ligand-binding site of SAP (prepared using Bobscript, Grasp, GL Render, and Raster3D). The dihedral angles of the type I β-turn enable the ligand-binding site to accommodate both calcium ligation, via the aspartic acid carboxyl (*i*), and insertion of proceeding pyrrolidine ring (*i*+1) into the hydrophobic pocket. Although the main chain carbonyls are directed away from Asn59 and Gln148, a serine residue at the *i*+2 position is within hydrogen bonding distance of the amino group of Gln148. Structural elements are represented as previously described for figure 3.33.

3.6 References

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E. (2000) *Nucleic Acids Res.* **28**, 235-242.

Brogden, R. N., Todd, P. A., Sorkin, E. M. (1988) *Drugs* **36**, 540-600.

Brünger, A. T. (1992) *X-PLOR: A System for X-ray Crystallography and NMR*. Version 3.1. Yale University Press, New Haven and London.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., Warren, G. L. (1998) *Acta Cryst.* **D54**, 905-921.

Chothia, C. (1975) *Nature* **254**, 304-308.

Chothia, C. & Janin, J. (1975) *Nature* **256**, 705-708.

Collaborative Computational Project Number 4 (1994) *Acta Cryst.* **D50**, 760-763.

Cremer, D. & Pople, J. A. (1975) *J. Am. Chem. Soc.* **97**, 1354-1358.

Driessen, H. P., Bax, B., Slingsby, C., Lindley, P. F., Mahadevan, D., Moss, D. S., Tickle, I. J. (1991) *Acta Cryst.* **B47**, 987-997.

Emsley, J., White, H. E., O'Hara, B. P., Oliva, G., Srinivasan, N., Tickle, I. J., Blundell, T. L., Pepys, M. B., Wood, S. P. (1994) *Nature* **367**, 338-345.

Esnouf, R. A. (1997) *J. Mol. Graph. Model.* **15**, 132-134.

Esser, L. (2001) *Personal Communication.*

Fletcher, D. A., McMeeking, R. F., Parkin, D. J. (1996) *Chem. Info.* **36,** 746-749.

Gavras, H. (1983) *J. Am. Coll. Cardiol.* **1**, 518-520.

Harding, M. M. (1999) *Acta Cryst.* **D55,** 1432-1443.

Harp, J. M., Hanson, B. L., Timm, D. E., Bunick, G. J. (2000) *Acta Cryst.* **D56**, 1513-1534.

Hind, C. R. K., Collins, P. M., Caspi, D., Baltz, M. L., Pepys, M. B. (1984) *Lancet* **2**, 376-378.

Hohenester, E., Hutchinson, W. L., Pepys, M. B., Wood, S. P. (1997) *J. Mol. Biol.* **269**, 570-578.

Hutchinson, E. G. & Thornton, J. M. (1994) *Protein Sci.* **3**, 2207-2216.

Jordan, F. (1973) *J. Theor. Biol.* **30,** 621-630.

Kolstoe, S. (2001) *Personal Communication.*

Laskowski, R. A., MacArthur, M. W., Moss, D. S., Thornton, J. M. (1993) *J. Appl. Crys.* **26**, 283-291.

Leslie, A. G. W. (1997) *Mosflm Users Guide*. MRC Laboratory of Molecular Biology, Cambridge.

Matthews, B. W. (1968) *J. Mol. Biol.* **33**, 491-497.

McRee, D. E. (1999) *J. Struct. Biol.* **125**, 156-165.

Merritt, E. A. & Murphy, M. E. P. (1994) *Acta Cryst.* **D50**, 869-873.

Milner-White, E. J., Bell, L. H., Maccallum, P. H. (1992) *J. Mol. Biol.* **228**, 725-734.

Minor, W. (1997) *The HKL Manual*. Edition 5.

Molecular Simulations (1996) *QUANTA96 X-ray Structure Analysis User's Reference*. San Diego.

Nicholls, A., Sharp, K. A., Honig, B. (1991) *Proteins Struct. Funct. Genet.* **11**, 281-296.

Nyburg, S. C. & Faerman, C. H. (1985) *Acta Cryst.* **B41**, 274-279.

Panico, R., Powell, W. H., Richer, J. C., eds (1993) *A Guide to IUPAC Nomenclature of Organic Compounds, Recommendations 1993*. Blackwell Scientific Publications.

Pepys, M. B. & Blundell, T. L. (2000) *Patent Number US6126918: Screening assays to identify therapeutic agents for amyloidosis*. Imperial College Innovations Limited, London.

Pepys, M. B., Booth, D. R., Hutchinson, W. L., Gallimore, J. R., Collins, P. M., Hohenester, E. (1997) *Int. J. Exp. Clin. Invest.* **4**, 274-295.

Pye, V (2000) *PhD Thesis University of London.*

Thompson, D. (1997) *PhD Thesis University of London.*

Thompson, D. (2000) *Personal Communication.*

Vagin, A. & Teplyakov, A. (1997) *J. Appl. Crys.* **30,** 1022-1025.

Vriend, G. (1990) *J. Mol. Graph.* **8**, 52-56.

CHAPTER 4

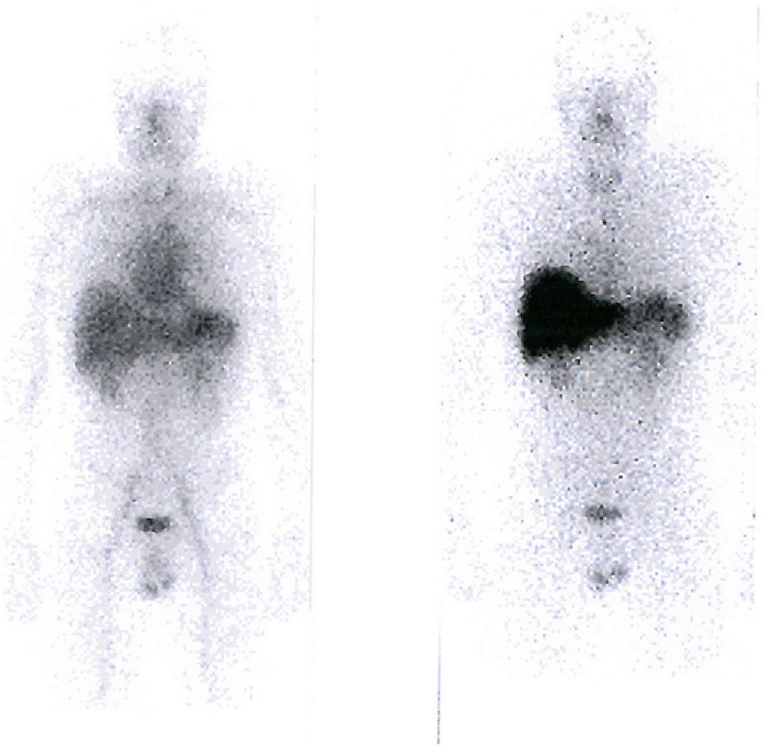
STRUCTURE DETERMINATION OF THE SAP-Ro 63-8695 COMPLEX

4.1 Introduction

Human clinical trials have found that Ro 63-8695 causes rapid clearance of SAP by the liver, producing a remarkable depletion of circulating SAP levels (Pepys 2001). Administration is performed by intravenous infusion, and SAP depletion is complete within 48 hours irrespective of the amyloid load. This promotes the redistribution of SAP from the amyloid deposits into the circulation, where it is immediately targeted by the drug and cleared by the liver (see figure 4.1). If drug infusion is stopped then plasma levels rapidly return to normal in patients with low amyloid load, however patients with higher amyloid load show a significant delay due to redistribution of SAP into the amyloid deposits. Because of this novel mechanism within the bloodstream, Ro 63-8695 does not require access to the tissues in order to exert its biological effect, and therefore maybe applicable to all forms of amyloidosis where the deposits are the primary cause of the disease.

Under physiological conditions *in vitro* Ro 63-8695 binds with remarkably high affinity to SAP (K_d 10nM (Kolstoe 2001)) inducing the formation of a decameric complex. The complex is sufficiently stable to be gel filtered with all of the SAP existing in the decamer form at ratios between equimolar and 100-fold molar excess of Ro 63-8695 over SAP protomers. Decamers are also produced upon addition of equimolar or greater excess of Ro 63-8695 to whole serum. Administration levels required for SAP depletion are sufficient to generate some decameric complexes, which are believed to be responsible for SAP clearance.

The previous chapter has elucidated the binding of Ro 63-8695 through the bound calcium ions and the ligand-binding site upon the exposed B-face of pentameric SAP. This



0 6

Time after starting Ro 63-8695 infusion (hours)

Figure 4.1. Whole body ^{123}I-SAP scintigraphy before and after 6 hours of Ro 63-8695 infusion (Pepys 2001). The patient shows modest load of AL amyloid in the spleen and kidneys, with notable blood pool background of tracer in the heart and circulation before administration of the drug. After 6 hours of administration the blood pool background is completely absent, the intensity of SAP signal from the amyloidotic spleen and kidneys is significantly reduced, and the liver has taken up the tracer.

chapter covers the structure determination of the SAP-Ro 63-8695 decamer complex produced by co-crystallisation to a resolution of 3.2Å.

4.2 Crystallisation & Crystal Screening

The previously solved structures of SAP in the presence of MOβDG (Thompson 1997), PE (Pye 2000), PC (Thompson 2000), and GABA (Pye 2000) were all derived from crystals grown under similar conditions by vapour diffusion at 4°C. These previously refined conditions were encompassed into the initial factorial screens for SAP in the presence of at least 10-fold molar excess Ro 63-8695. The screens were set-up using the vapour diffusion technique of hanging drops under the following conditions: 3μl drop of protein solution containing 10-20mg/ml SAP, 4-8mM Ro 63-8695, 60mM Tris pH 7.2-7.8, 84mM NaCl mixed with an equal volume of reservoir solution containing 60mM Tris pH 7.2-7.8, 84mM NaCl, 8-23% (w/v) PEG MME 550, 10mM Ca acetate, and allowed to equilibrate against 1ml of the reservoir solution at 4°C. Crystals grew within 3 weeks under the conditions containing 20 mg/ml SAP, 8mM Ro 63-8695, and 21-23% PEG MME 550, and continued to grow for a further 4 weeks with crystals eventually appearing at the lower PEG concentrations. Crystals that grew under conditions with limited nucleation but sustained growth, typically at pH 7.4-7.6, grew to a size of 0.4x0.1x0.05mm and appeared perfectly symmetrical, an example of which is shown in figure 4.2. Those crystals that grew under the conditions where nucleation was sustained for longer grew to a size of 0.2x0.2x0.1mm, and appear as shorter and thicker crystals but still contained the same hexagonal and rhomboidal characteristics (see figure 4.3).

Manipulation and mounting of the crystals was fraught with many technical difficulties as they were extremely brittle and very sensitive to temperature variations. In

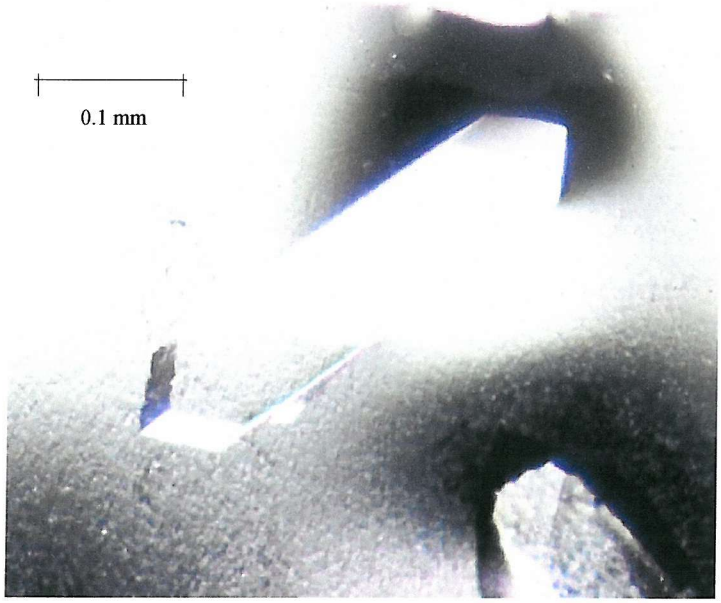


Figure 4.2. Photograph of a symmetrical SAP crystal grown by vapour diffusion in the presence of Ro 63-8695.

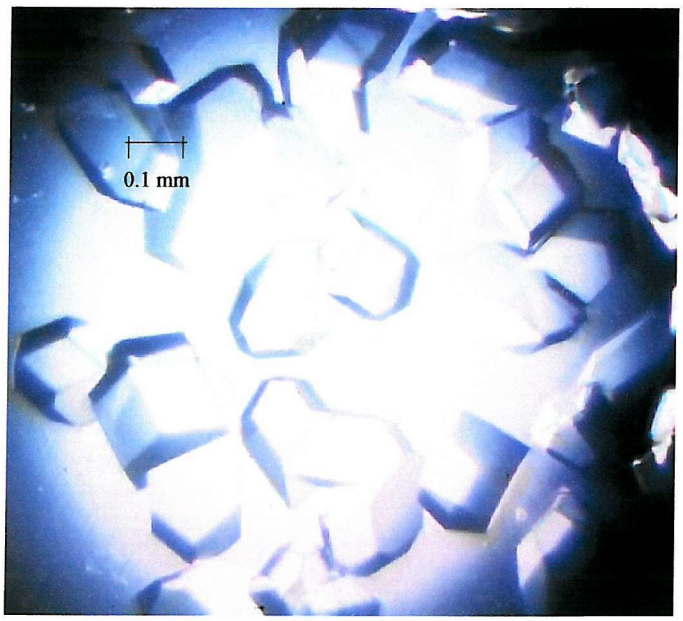


Figure 4.3. Photograph of the thicker SAP crystals grown in the same drop by vapour diffusion in the presence of Ro 63-8695.

fact the crystals degraded and dissolved within days at room temperature. However, a single diffraction image was recorded from a single crystal mounted at room temperature, prior to dissolving, by the oscillation method using a MAR 300mm image plate. These data were collected in-house using an ENRAF NONIUS FR 591 rotating copper anode generator (λ=1.542Å). The image was recorded during a 30-minute exposure over an oscillation range of 0.5°, with the detector placed 300mm away from the crystal. The image showed clear diffraction to a maximum resolution of 3.3Å.

The temperature sensitivity of the crystals was eliminated by the use of cryogenic temperatures during data collection. PEG concentrations used in the crystallisation were found to be insufficient for vitrification and required either increasing or the use of an additional cryoprotecting agent. 10% (v/v) glycerol added to the reservoir solution was found to be sufficient for vitrification, but the crystals rapidly degraded and dissolved upon transfer into the cryoprotectant solution. The rate of degradation was significantly reduced by the addition of 8mM Ro 63-8695 and by the maintenance of the crystal temperature at 4°C. The cryoprotectant was soaked in by sequentially transferring the crystal into solutions containing increasing concentrations of glycerol in 5% increments, allowing equilibration for 30 seconds in each before mounting in a fibre loop and flash freezing in liquid ethane. Several datasets were recorded in-house from single crystals at cryogenic temperatures (using an gaseous nitrogen stream at 100K) by the oscillation method using a MAR 300mm image plate. Each image was recorded during a 45-minute exposure over an oscillation range of 0.5°, with the detector placed 370mm away from the crystal. An example of one of the diffraction images is shown in figure 4.4. Although the crystals sustained no visible radiation damage during data collection, the diffraction was found to be highly anisotropic with cryogenic freezing having a detrimental effect upon both resolution (<4Å) and diffraction quality. All of the crystals produced incomplete unprocessable

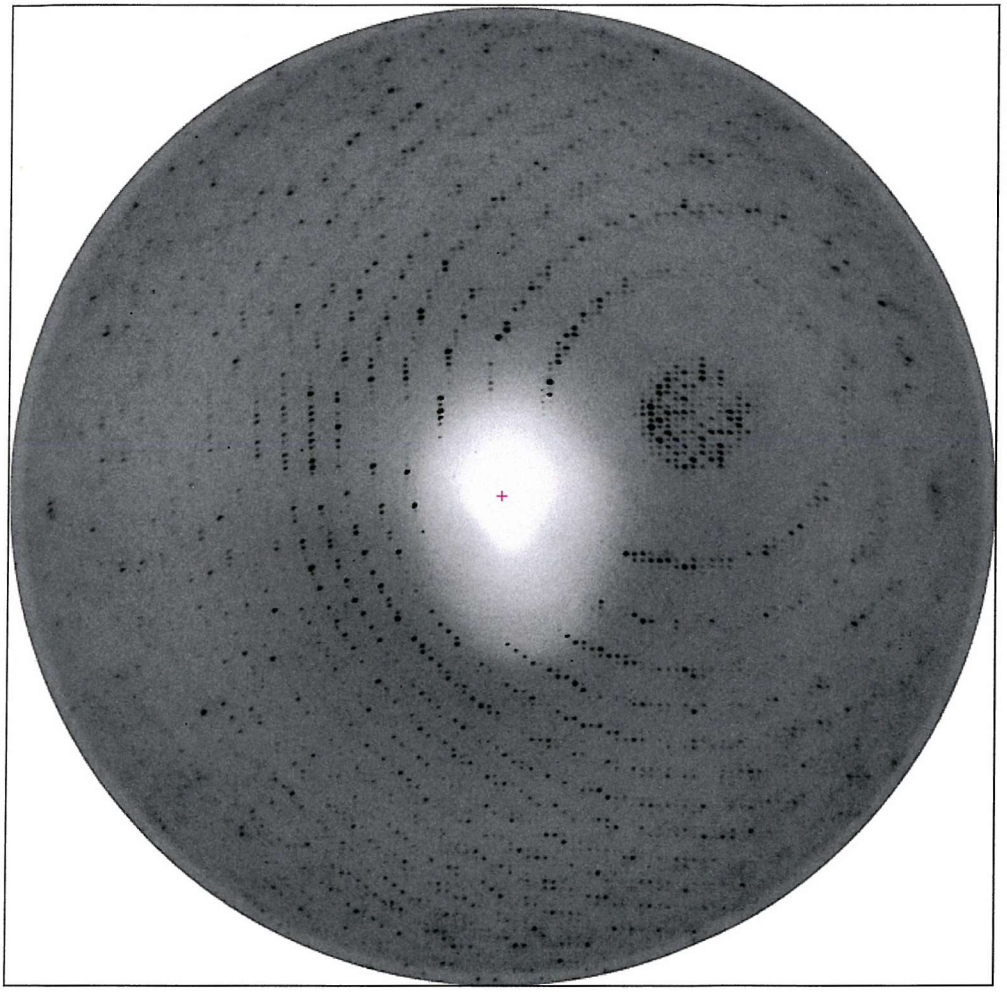


Figure 4.4. A 0.5˚ oscillation image collected in-house on a MAR 300mm image plate from a SAP-Ro 63-8695 crystal at cryogenic temperatures, recorded to a maximum resolution of 4.01Å (prepared using Mosflm (Leslie 1997)).

datasets, however autoindexing using Mosflm (Leslie 1997) did give an estimation of the unit cell dimensions, where a=232Å, b=232Å, c=281Å, α=90°, β=90°, and γ=90°.

The initial crystallisation conditions were refined by factorial screening to minimise nucleation whilst increasing the PEG concentration to enable direct vitrification. The screens were set-up using the same technique of vapour diffusion as previously described, but with a protein solution containing 18-24mg/ml SAP, 7-10mM Ro 63-8695, 60mM Tris pH 7.4-7.6, 84mM NaCl and a reservoir solution containing 60mM Tris pH 7.4-7.6, 84mM NaCl, 20-28% (w/v) PEG MME 550, 10mM Ca acetate. Crystals grew to an average size of 0.5x0.2x0.2mm within 7 weeks, with one reaching a size of 0.85x0.3x0.2mm. Unfortunately even those crystals grown at the highest PEG concentrations (28%) still required an additional cryoprotecting agent. Therefore, the cryogenic freezing protocol was also refined to minimise both crystal manipulation and degradation during soaking. The direct transfer of crystals into the final cryoprotectant solution for a period of only 2-10 seconds, before flash freezing in either a gaseous nitrogen stream (100K) or in liquid ethane, was found to significantly reduce the visible degradation of the crystals. But, the shorter soak time required the use of a more concentrated cryoprotectant solution. Several different cryoprotecting agents including PEG 400, PEG MME 550, ethylene glycol, glucose, glycerol, and sucrose were screened, but only 20% (v/v) glycerol or 34% (w/v) PEG 550 was found to sufficiently vitrify the water content without significantly affecting crystal diffraction.

A total of 173 individual crystals were screened in-house and the least anisotropic and best diffracting crystals were taken to the European Synchrotron Radiation Facility (ESRF) in Grenoble (France), and/or the European Molecular Biology Laboratory (EMBL) in Hamburg (Germany) for subsequent screening and data collection. The majority of the crystals diffracted to a resolution of ~4.0Å with varying degrees of anisotropy. Several

4.0Å datasets were recorded but were either unprocessable or unsolvable by molecular replacement. However, one perfectly symmetrical crystal (number 173) was found to diffract to maximum resolution of 2.9Å with only a small degree of anisotropy. Data collection, processing, molecular replacement, and refinement of this dataset are described in detail in section 4.3.

4.3 Structure Determination

4.3.1 Crystallisation Condition

The crystal was grown by the vapour diffusion technique of hanging drop under the following conditions: 3µl drop of protein solution containing 18mg/ml SAP, 7mM Ro 63-8695, 60mM Tris pH 7.6, 84mM NaCl mixed with an equal volume of reservoir solution containing 60mM Tris pH 7.6, 84mM NaCl, 25% (w/v) PEG MME 550, 10mM Ca acetate, and allowed to equilibrate against 1ml of the reservoir solution at 4°C.

4.3.2 Data Collection & Processing

The crystal was transferred to a small drop of the reservoir solution containing 21% (v/v) glycerol and 7mM Ro 63-8695, and allowed to equilibrate for ~5 seconds at 4°C prior to mounting in a fibre loop and flash freezing in a gaseous nitrogen stream (100K). Diffraction data were collected from the crystal at cryogenic temperatures (using a gaseous nitrogen stream at 100K) by the oscillation method using a MarCCD 165 detector. These data were collected at the ESRF on station ID14-3, with the X-ray beam tuned to a wavelength of 0.931Å. Each image was recorded during three 60-second passes over an oscillation range of 0.5°, with the detector placed 200mm away from the crystal. The

dataset was recorded over a total of 100 images, an example of which is shown in figure 4.5 and 4.6.

The data were originally corrected, processed, scaled, and merged using Mosflm and other programs from the CCP4 suite (Collaborative Computational Project Number 4 1994). The crystal packing was found to be tetragonal with unit cell dimensions of a=b=229.1Å and c=279.6Å, and was processed in the space group P4. The final processing statistics for the data (20-3.3Å) were a poor representation of the quality of diffraction with an overall *R-merge(I)* of 10.4% and I/σ(I) of 6.4. Examination of the data processing revealed that all of the individual reflections were only partially recorded (see figure 4.6) extending over three or more consecutive images. This led to poorly refined reflection positions and detector parameters during post-refinement, as only fully recorded or partially recorded reflections extending over only two images are used (Leslie 1997), resulting in insufficient data to accurately refine the parameters. The raw images were subsequently reprocessed using Denzo and Scalepack (Minor 1997), and programs from the CCP4 suite. Denzo uses a different procedure and algorithms for refining the reflection positions and detector parameters. Each image is individually refined prior to post-refinement where all of the images are refined together, but more importantly contains an additional algorithm for the refinement of partial reflections during both procedures (Otwinowski & Minor 1997). This resulted in a better interpretation of the diffraction data. The crystal packing was again found to be tetragonal but with slightly different unit cell dimensions of a=b=230.9Å and c=281.4Å, and was processed in space group P4. Examination of the processed data revealed a missing section along axis 0,0,*l* (see figure 4.7), which was present in a previous 3.9Å dataset revealing systematic absences in 0,0,*l* (where $l \neq 4n$) (see figure 4.8) implying a $P4_1$ or $P4_3$ screw axis. Data along axes *h*,0,0 and 0,*k*,0 both revealed similar systematic absences (where $h/k \neq 2n$) (see figure 4.9) implying a space group of

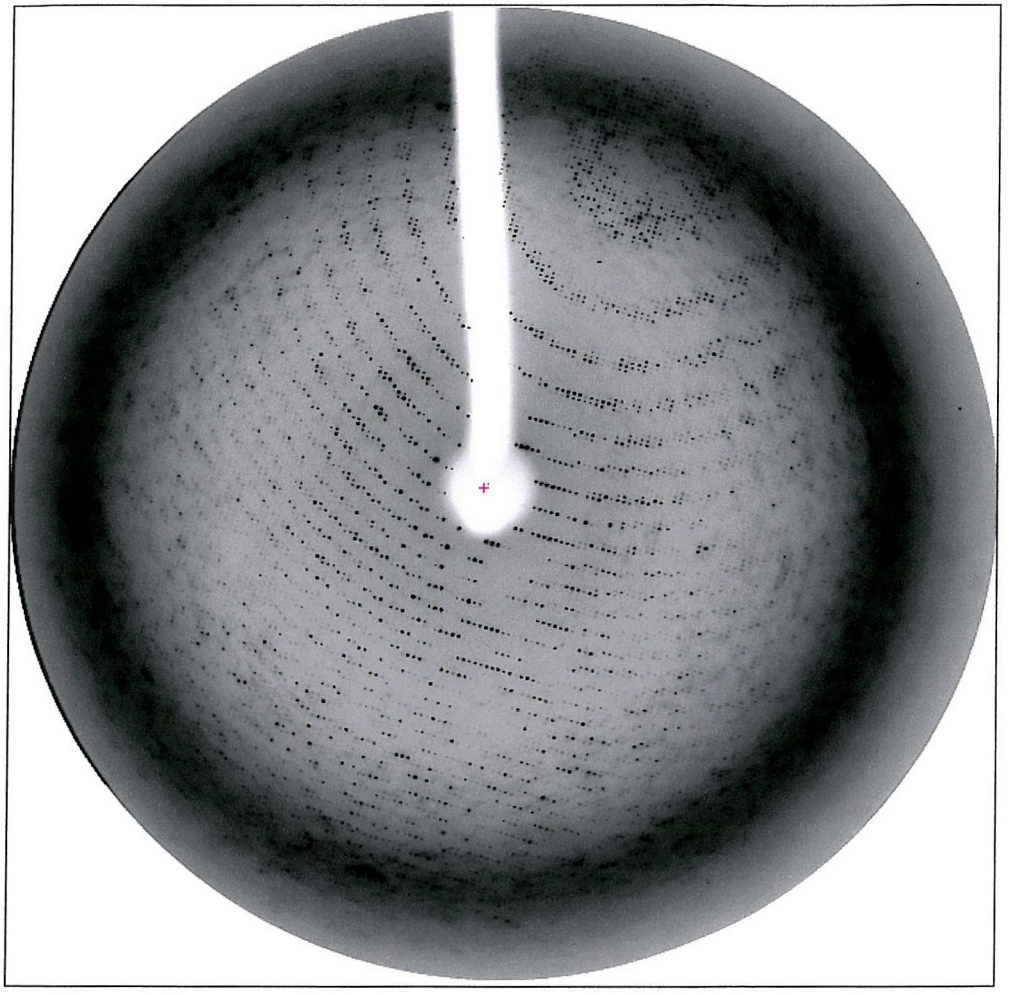


Figure 4.5. A 0.5° oscillation image collected at the ESRF (station ID14-3) on a MarCCD 165 detector from a SAP-Ro 63-8695 crystal at cryogenic temperatures, recorded to a maximum resolution of 2.8Å (prepared using Mosflm). A close-up of the top right section can be seen in figure 4.6.

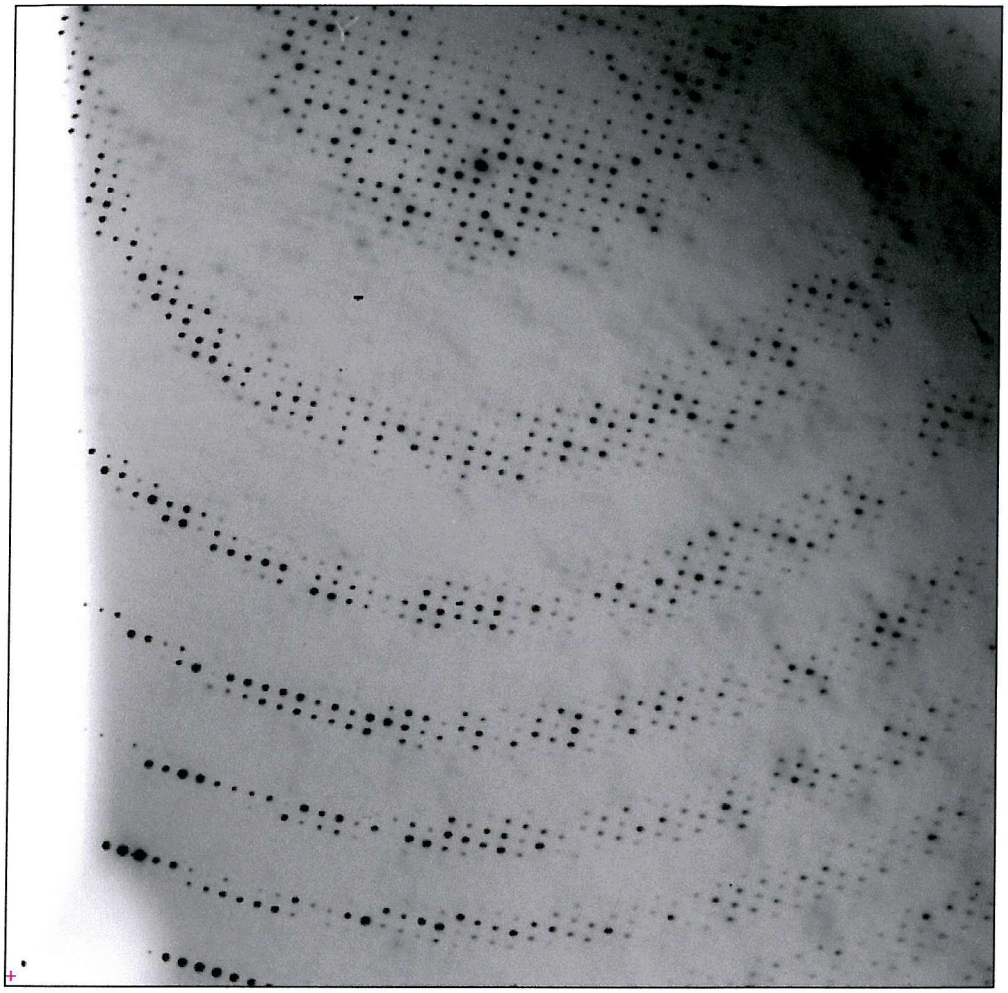


Figure 4.6. Close-up section of a 0.5° oscillation image collected at the ESRF (station ID14-3) on a MarCCD 165 detector from a SAP-Ro 63-8695 crystal at cryogenic temperatures (prepared using Mosflm).

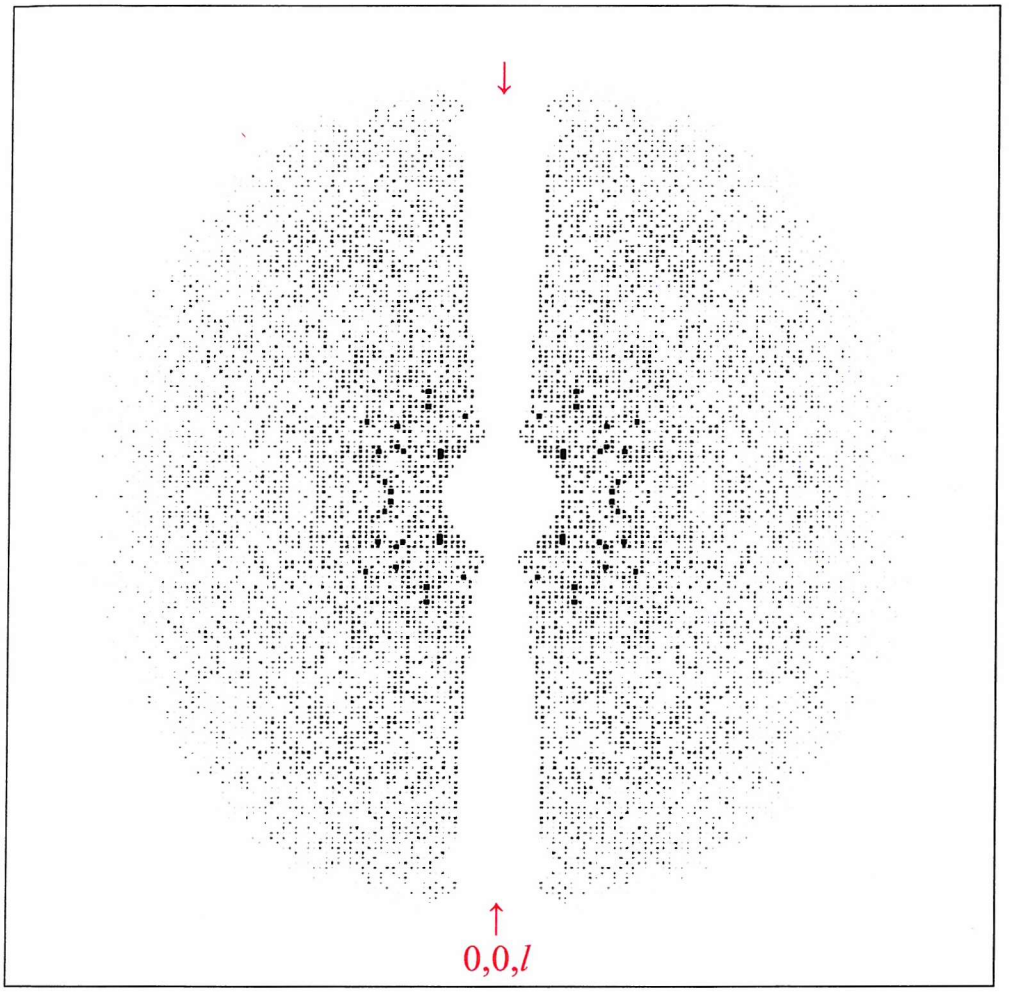


Figure 4.7. A pseudo precession picture displaying missing data along axis 0,0,*l*, within the 0,*k*,*l* zone, for SAP-Ro 63-8695 data processed in P4 (prepared using HKLView (Collaborative Computational Project Number 4 1994))

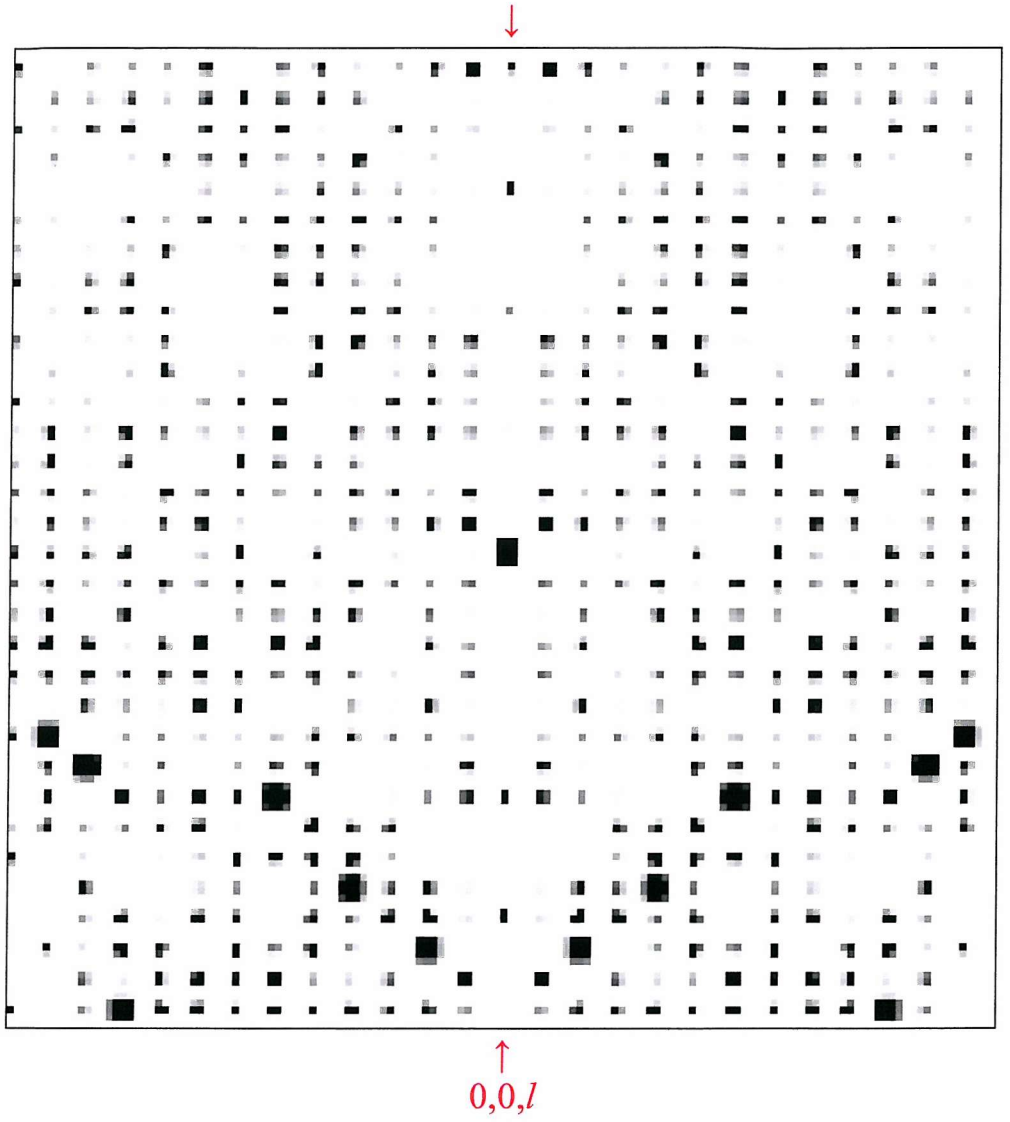


Figure 4.8. A pseudo precession picture displaying systematic absences along axis 0,0,*l* (*l*≠4n), within the 0,*k*,*l* zone, implying a 4_1 or 4_3 screw axis for SAP-Ro 63-8695 data collected at the ESRF (station BM-14) on a MarCCD detector, processed in P4 to a maximum resolution of 3.9Å using Mosflm (prepared using HKLView).

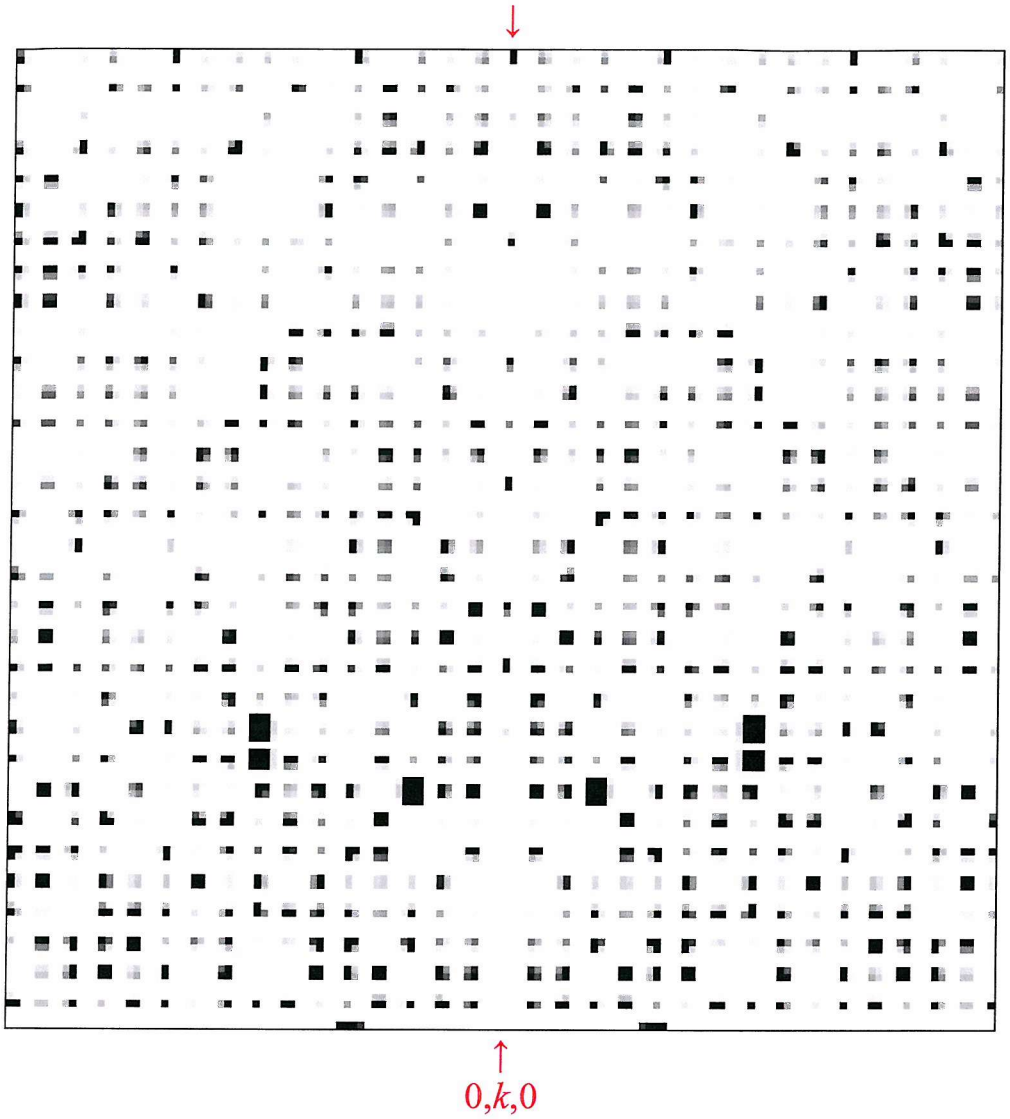


Figure 4.9. A pseudo precession picture displaying systematic absences along axis 0,*k*,0 (*k*≠2n), within the *h*,*k*,0 zone, implying $P4_12_12$ or $P4_32_12$ screw axes for SAP-Ro 63-8695 data processed in P4 (prepared using HKLView).

either $P4_12_12$ or $P4_32_12$. These two space groups can only be differentiated using the translation function, and therefore the data was reprocessed, scaled, and merged in the space group $P4_32_12$ to a maximum resolution of 3.2Å. The overall completeness of the data was 95.7% and had an *R-merge(I)* of 8.6%.

4.3.3 Molecular Replacement

The large unit cell indicates the presence of several molecules within the asymmetric unit. Although crystals can contain between 27 to 95% solvent (Matthews 1968), protein crystals generally contain between 40 to 60% (Crick & Kendrew 1957). An approximation of the solvent content for an increasing number of SAP pentamers in the asymmetric unit was calculated and is shown in table 4.1. The fragile nature, temperature sensitivity, and relatively large amount of cryoprotectant required implies a higher than usual solvent content. However the high stability of the SAP-Ro 63-8695 complex in solution and the presence of 10-fold molar excess Ro 63-8695, would suggest that all of SAP is in the decameric form. Therefore, it seemed most likely that each asymmetric unit contained 4 SAP pentamers, giving the crystal a solvent content of 67%.

Table 4.1. Estimated solvent content for the SAP-Ro 63-8695 crystal form (V_M=2.61 Å^3/Da (Matthews 1968)).

Pentamers In Asymmetric Unit	Solvent Content
1	92%
2	83%
3	75%
4	67%
5	58%
6	50%

The self-rotation function was initially calculated to determine the direction and nature of any non-crystallographic symmetry related elements. This was performed within reciprocal space applying Kabsch's fast polar rotation function to calculate the Patterson overlap using Polarrfn (Collaborative Computational Project Number 4 1994). Reflections within the range of 10 to 3.2Å were included in the calculation with a radius of integration of 40Å. Stereographic projections of the self-rotation function in spherical polar angles clearly showed non-crystallographic peaks present in the χ 75° and 180° sections corresponding to a single 5-fold axis parallel to c, and several 2-fold axes within the a-b plane (see figure 4.10). This is consistent with the formation of a decamer via pentamer face stacking, generating two-fold axes perpendicular to their five-fold radial symmetry.

Molecular replacement was then performed using Molrep (Vagin & Teplyakov 1997). The originally derived SAP pentamer refined to 2.0Å resolution (Emsley *et al* 1994) (Protein Data Bank (Berman *et al* 2000) entry code 1SAC) was used as the search model with the calcium ions and acetate molecules omitted. The rotation function was calculated using reflections within the range of 25 to 3.2Å with a radius of integration of 30Å. Several orientations produced peaks, although relatively weak (see table 4.2). The translation function was performed with the top 10 orientations, using reflections within the range of 25 to 3.2Å. Due to the uncertainty of the space group this was calculated for both enantiomorphs. The translation function was also performed manually as automated searches positioned NCS related molecules at incorrect locations causing steric clashes. The translation function was first calculated in the space group $P4_32_12$, where orientation 2 from the rotation function list (α=51.21°, β=70.49°, and γ=8.17°) produced the largest peak of 63σ (29σ above the next largest) corresponding to a fractional shift of x=0.212, y=0.488, and z=0.151. This correctly orientated and translated molecule was fixed and a second molecule searched relative to the same origin. Orientation 4 from the list (α=29.67°,

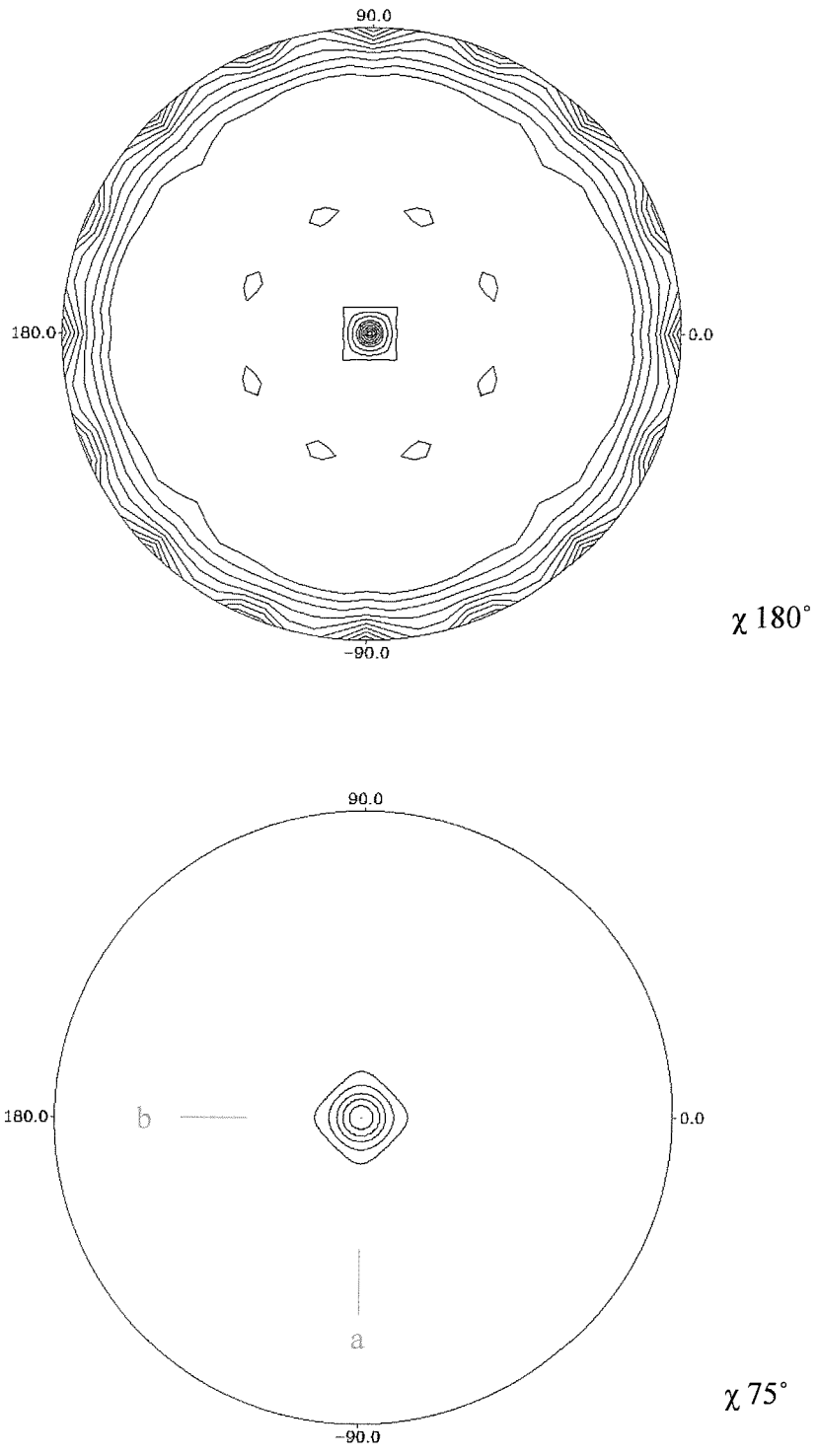


Figure 4.10. Stereographic projections (viewed down the c-axis) of the self-rotation function in spherical polar angles at χ=180° and χ=75° (prepared using Polarrfn). Non-crystallographic peaks can be seen for the spherical polar coordinates of ϕ=±23°/ ±67°/ ±113°/ ±157°, ω=90°, χ=180°, and ϕ=0°, ω=0°, χ=75°.

Table 4.2. Top 10 rotation function peaks calculated with Molrep.

Peak Number	Rotation α (°)	β (°)	γ (°)	RF/σ
1	6.05	70.59	7.49	3.28
2	51.21	70.49	8.17	3.12
3	74.49	82.46	113.38	3.07
4	29.67	81.58	114.11	3.04
5	11.93	37.03	68.25	3.01
6	55.12	37.62	69.29	2.97
7	60.90	65.37	338.00	2.93
8	67.59	43.14	274.81	2.92
9	22.63	43.16	275.17	2.91
10	25.39	10.86	100.00	2.90

Table 4.3. The 4 molecular replacement solutions calculated with Molrep in space group $P4_32_12$.

Solution	Rotation (°)			Shift Fraction			Peak	*R-factor*	CC
	α	β	γ	*x*	*y*	*z*			
1	51.21	70.49	8.17	0.212	0.488	0.151	63.40	47.2%	46.1%
2	29.67	81.58	114.11	0.267	0.968	0.146	131.84	33.0%	75.3%
3	6.05	70.59	7.49	0.640	0.842	0.313	29.03	44.8%	52.3%
4	74.49	82.46	113.38	0.859	0.309	0.986	67.98	51.0%	43.8%

β=81.58°, and γ=114.11°) produced a significantly larger peak of 132σ (95σ above the next largest) at a fractional shift of *x*=0.267, *y*=0.968, and *z*=0.146. This molecule was then fixed and the third searched, in which orientation 1 (α=6.05°, β=70.59°, and γ=7.49°) producing only a weak peak of 29σ (11σ above the next largest) at a fractional shift of *x*=0.640, *y*=0.842, and *z*=0.313. This was then fixed and a fourth molecule searched, where orientation 3 (α=74.49°, β=82.46°, and γ=113.38°) was found to produce a larger 67σ (50σ above the next largest) peak than the previous solution at fractional shift of *x*=0.859, *y*=0.309, and *z*=0.986. Subsequent translational searches for further molecules within the asymmetric unit were carried out with the remaining six orientations, all of which positioned the search model at crystallographic symmetry related locations to one of the previously fixed molecules.

The translation function was subsequently repeated in the alternative space group ($P4_12_12$) where orientation 2 (α=51.21°, β=70.49°, and γ=8.17°) produced an insignificant peak of 30σ (0.32σ above the next largest) at a fractional shift of *x*=0.212, *y*=0.488, and *z*=0.401. Therefore, the correct space group was $P4_32_12$.

The 4 solutions produced in the correct space group (summarised in table 4.3) gave a resultant *R-factor* of 51.0%, correlation coefficient of 43.8%, and showed good sensible packing contacts between symmetry related molecules (see figure 4.11). Upon closer examination, the four SAP pentamers were found to be orientated as two decamers of solutions 1/2 and 3/4. However, the five and two-fold NCS axes do not appear to lie as predicted by the self-rotation function. Crystallographic refinement was performed upon the correctly orientated and translated phasing models, initially starting with an *R-factor* of 54.1% (*R-free* 54.4%). Subsequent crystallographic refinement imposing high weighted (wa=300) protomer (20-fold) NCS restraints, and manual model building was unable to reduce the *R-factor* below 42.6% (*R-free* 43.6%), which is too close to the random figure of

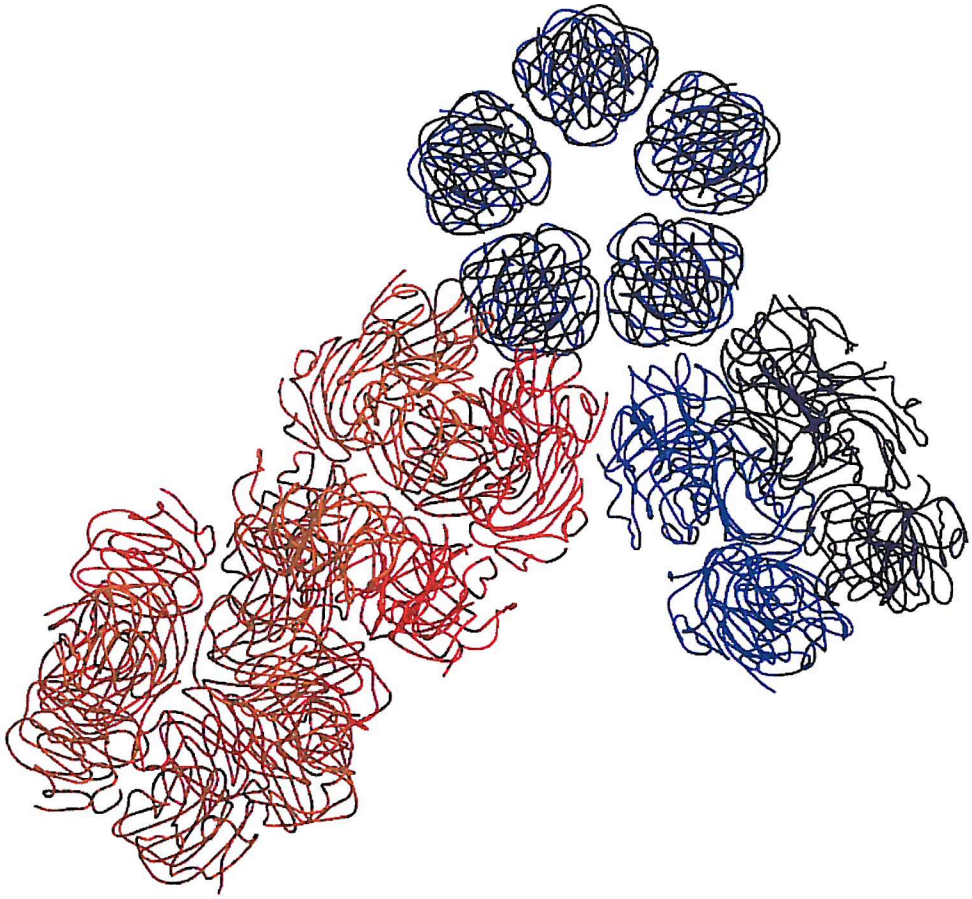


Figure 4.11. The crystal packing of the four molecular replacement solutions calculated using Molrep (prepared using Molscript (Kraulis 1991) and Raster3D (Merritt & Murphy 1994)). SAP pentamers are depicted in different colours according to the solutions in table 4.3: solution 1 in red, solution 2 in orange, solution 3 in blue, and solution 4 in indigo.

60%. Furthermore, the electron density maps showed poor connectivity even after 20-fold NCS averaging.

The translation function for the four solutions produced by Molrep was checked by applying a different modification of the T_2 function using TFFC (Driessen *et al* 1991). The translational search was performed using reflections within the range of 25 to 3.5Å. Firstly, each solution was individually translated, and with the exception of a small translation vector of solution 3 (x=0.001, y=0.000, and z=0.001), all gave the largest peaks in their current and symmetry related locations. Solutions 1 and 2 gave the largest peaks of 80σ and 79σ (both 42σ above the next largest) respectively, whilst 3 and 4 gave weaker peaks of 54σ (28σ above the next largest) and 66σ (33σ above the next largest) respectively. The molecules were then translated with respect to the same origin. If solution 1 was fixed, then solution 2 produced a large peak of 113σ (102σ above the next largest) at its current location. However when solutions 3 and/or 4 were searched relative to the origin defined by fixing solutions 1 and/or 2, the largest peaks were produced at a location half a unit cell along the c-axis at a symmetry related position to the fixed solutions respectively. The reverse was also true, where the origin was found to be the same for solutions 3 and 4, and when fixed, solutions 1 and/or 2 were translated half a unit cell along the c-axis to a symmetry related position. It seems likely that solution 1 and 2 are in fact the same as 3 and 4 respectively, differing in the location of the origin along c, and there are only 2 pentamers in the asymmetric unit. This ambiguity may have been exacerbated by the powerful packing function within Molrep.

Very few protein crystals contain such a high solvent content (83%) as implied by TFFC. Therefore molecular replacement was repeated using AMoRe (Navaza 1994), which has been used to solve a number of difficult structures. The rotation function was calculated using reflections within the range 25 to 3.2Å with a radius of integration of 30Å.

The translation function was then performed with the top 10 orientations before refining each by rigid-body refinement prior to fixing the best solution and searching for the next. AMoRe allows the translation search to be calculated by applying one of four different fast translation functions; the method of Crowther and Blow, phased translation, Harada-Lifchitz, and the correlation coefficient search. The differing translation functions all produced the same two solutions. Those calculated using the method of Crowther and Blow are shown in table 4.4, and gave an *R-factor* of 35.4% and correlation coefficient of 71.4% comparable to that of the first two solutions produced by Molrep (see table 4.3). As with TFFC, subsequent translational searches positioned the search model at symmetry related locations to the two fixed solutions.

Table 4.4. The 2 molecular replacement solutions calculated with AMoRe, using the method of Crowther and Blow.

Solution	Rotation (°)			Shift Fraction			*R-factor*	CC
	α	β	γ	*x*	*y*	*z*		
1	67.99	55.27	302.64	0.200	0.488	0.458	48.7%	43.6%
2	23.00	54.76	158.67	0.632	0.850	0.119	35.4%	71.4%

Electron density maps were calculated for the first two solutions produced by Molrep to verify the presence of only two molecules in the asymmetric unit. Both $|F_{obs}|$-$|F_{calc}|$ and $2|F_{obs}|$-$|F_{calc}|$ electron density maps showed clear density surrounding the two solutions, with only background noise present at the locations corresponding to the two missing solutions (see figure 4.12). There are, therefore, only two pentameric molecules in the asymmetric unit giving the crystal a very high solvent content of 83%. These results

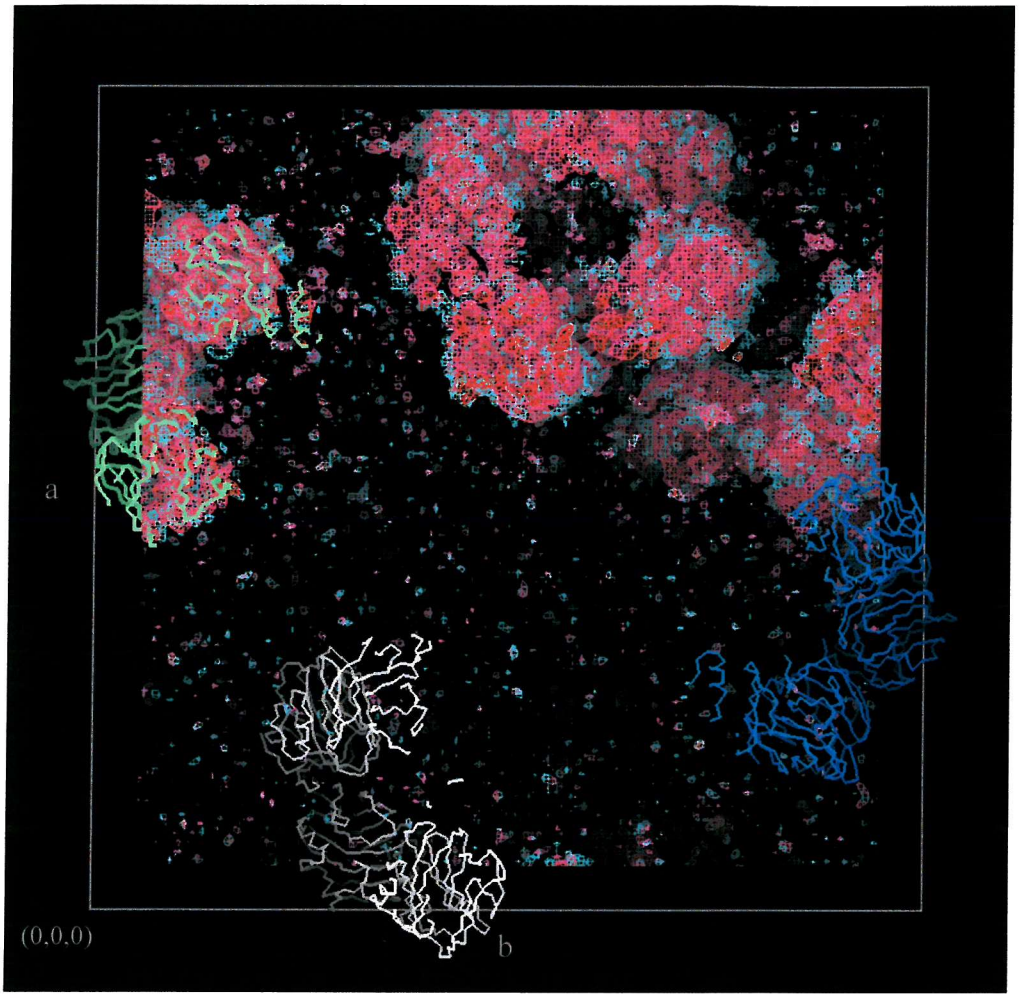


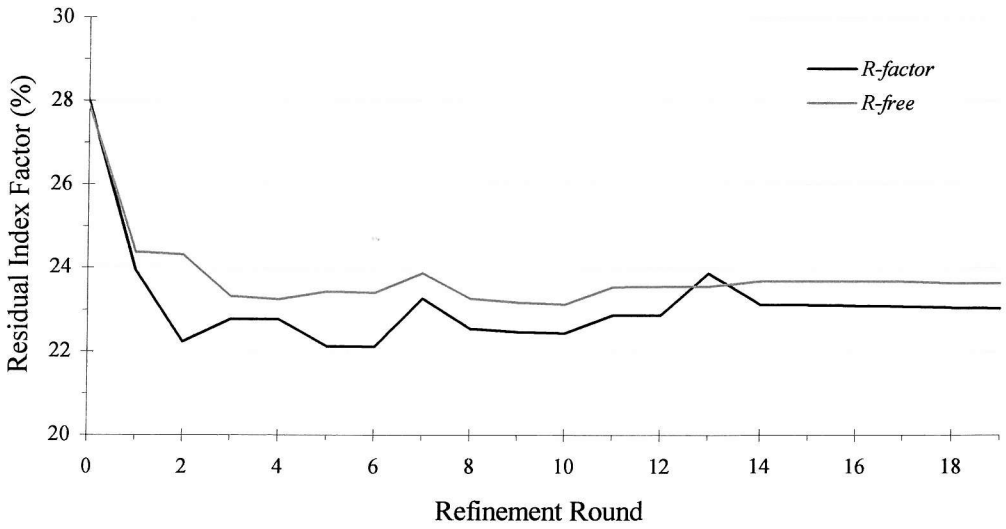
Figure 4.12. Electron density maps calculated for molecular replacement solutions 1 and 2 (calculated using Molrep) contoured over one asymmetric unit (prepared using **Quanta** (Molecular Simulations 1996)). Electron density is clearly present surrounding solutions 1 and 2 (green) with only background noise present at the locations for the missing solutions 3 (white) and 4 (blue). The $2|F_{obs}|-|F_{calc}|$ electron density is contoured at 1σ (depicted in pink) and 1.3σ (depicted in orange), while the $|F_{obs}|-|F_{calc}|$ electron density is contoured at 2σ (depicted in cyan). SAP pentamers are depicted as C_α traces in different colours corresponding to the solution numbers, with the axes and origin of the unit cell labelled accordingly.

agree with the self-rotation function. The crystal packing of solutions 1 and 2 is shown in figure 4.13.

4.3.4 Refinement & Model Building

Crystallographic refinement was performed using CNS (Brünger *et al* 1998), with a test set comprising of 10% of the reflections randomly flagged to calculate the *R-free* value. The model was refined against the data in 19 stages. Between each round of refinement the improved model was used to calculate $|F_{obs}|$-$|F_{calc}|$ and $2|F_{obs}|$-$|F_{calc}|$ sigmaA-weighted electron density maps for visualisation and manual model building within Quanta and XtalView (McRee 1999). Also, throughout the refinement, bulk solvent correction and an overall anisotropic B-factor correction was applied. The stereo-chemical properties of the model were monitored using WHAT IF (Vriend 1990) and Procheck (Laskowski *et al* 1993). A summary of the *R-factor* and *R-free* values during the refinement are shown in figure 4.14.

Figure 4.14. Summary of the refinement process for SAP-Ro 63-8695.



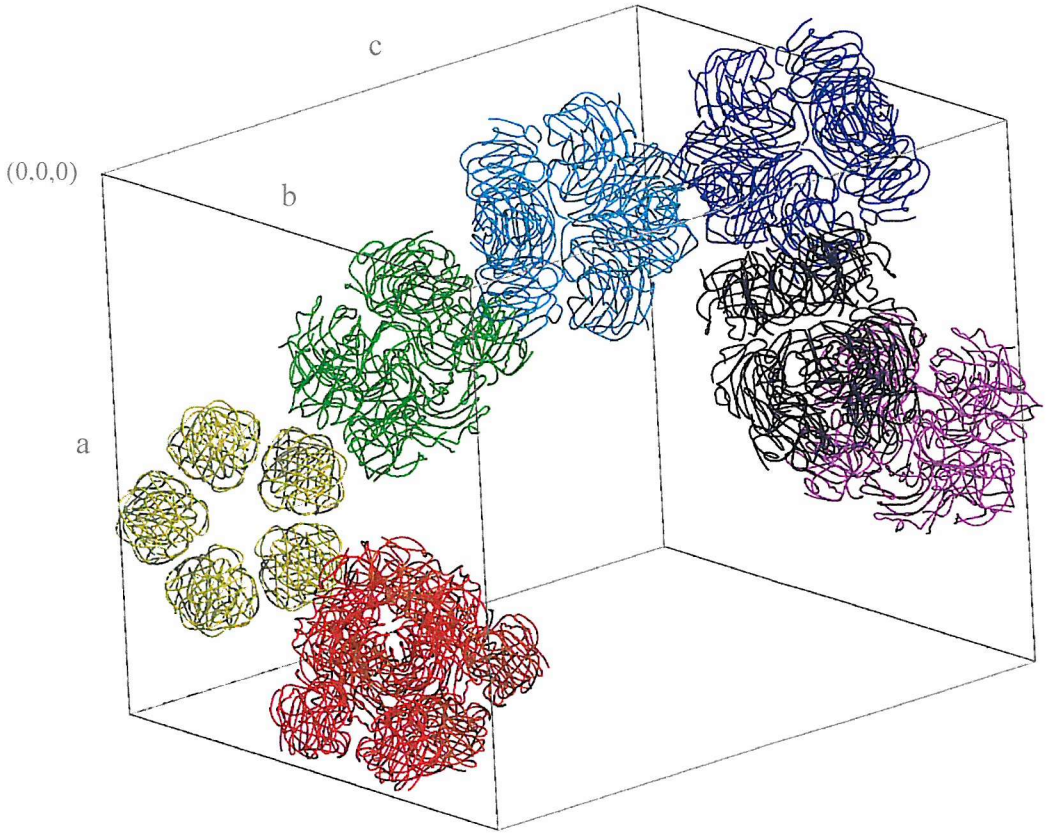


Figure 4.13. The crystal packing of SAP-Ro 63-8695 (prepared using Molscript, GL Render (Esser 2001), and Raster3D). In the presence of Ro 63-8695, SAP crystallises in the tetragonal space group $P4_32_12$ (a=b=230.9Å and c=281.4Å) with two SAP pentamers in the asymmetric unit. The SAP pentamers are arranged in pairs and are depicted in different colours with the axes and origin of the unit cell labelled accordingly.

The model was initially refined by several cycles of rigid-body refinement with the pentamers followed by the protomers treated as the rigid bodies, during which the resolution was progressively increased from 6 to 3.5Å, including only those reflection with I/σ(I) over 2.0. This was followed by 3 cycles of simulated annealing, using torsion angle dynamics at different temperatures (slow cool from 5,000 to 500K in 50K decrements, and from 2,500 to 500K in 25K decrements), using reflections within the range of 25 to 3.5Å with I/σ(I) over 2.0 (round 1). Examination of the resulting electron density maps revealed well-defined density for five ligands linking the two SAP pentamers together through adjacent double calcium-binding sites (see figure 4.15). High weighted (5-fold) NCS restraints (wa=300) were applied to dimers of adjacent subunits, significantly improving the observation to parameter ratio from 2.4:1 to 12:1 (calculated assuming constraints). Five-fold averaged electron density maps were subsequently generated for manual model building, with one subunit from each pentamer altered to better fit the electron density. Seven residues, four in one and three in the adjacent subunit, were found to have indiscernible orientations and were mutated to alanine. The altered subunit from each pentamer was then superimposed onto the other four prior to further refinement using reflections within an increased range of 25 to 3.2Å with I/σ(I) over 2.0. The resulting model was refined by simulated annealing (5,000K), and overall B-factor refinement for each protomer domain (round 2). No discernable differences were found between the adjacent subunits, enabling the NCS restraints to be tightened to the individual protomers (10-fold). 10-fold averaged electron density maps were then generated for manual model building, with 3 residues having indiscernible orientations. One modified subunit was then superimposed onto the other nine prior to three cycles of simulated annealing (2,500K) and overall B-factor (protomer domains) refinement (round 3). Further simulated annealing (2,500K), followed by positional and overall B-factor (protomer domains) refinement was

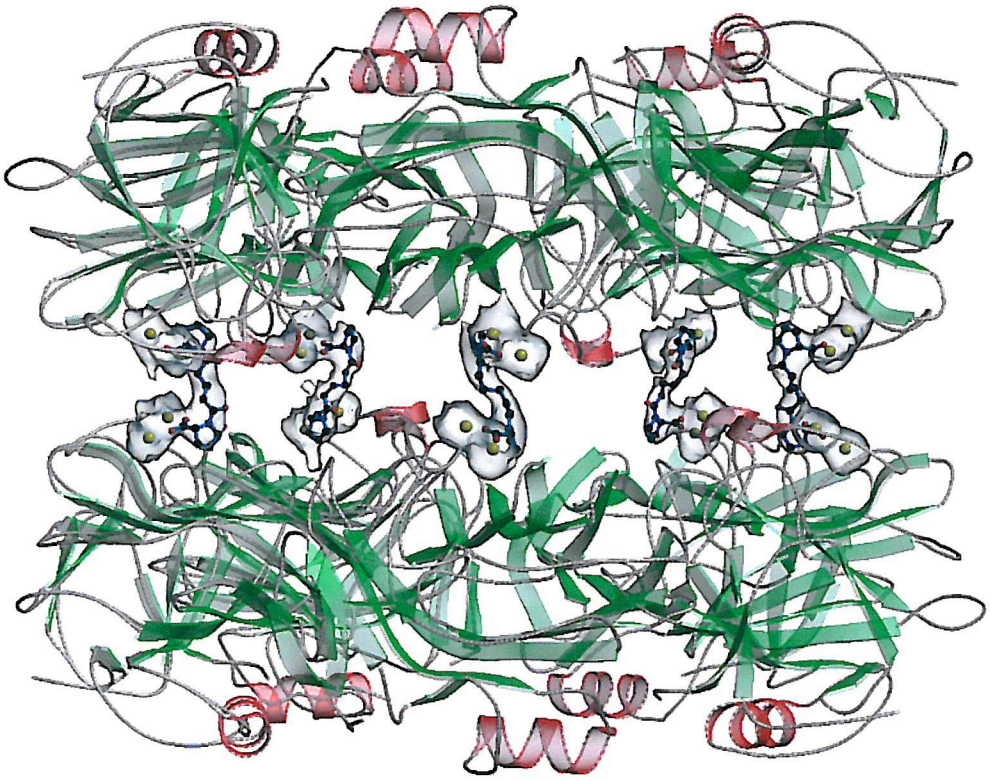


Figure 4.15. Three-dimensional structure of the SAP-Ro 63-8695 complex illustrating the electron density for the calcium ions and Ro 63-8695 molecules linking the B-faces of two adjacent SAP pentamers (prepared using Bobscript (Esnouf 1997), GL Render, and Raster3D). The $|F_{obs}|$-$|F_{calc}|$ difference density for the calcium ions and Ro 63-8695 molecules omitted from the fully refined model is represented as transparent surfaces (contoured at 3σ) over the final positions of the atoms. The β-strands are represented as green arrows, α-helices as red ribbons, loops as gray cords, calcium ions as yellow spheres, and Ro 63-8695 molecules as ball-and-stick models.

then performed (round 4) prior to the inclusion of weak reflections with I/σ(I) below 2. The disulphide bridge between Cys36-Cys95 was broken during subsequent model building, where both 10-fold averaged and unaveraged electron density map showed strong negative and positive difference density for Cys36 pointing away from Cys95 in each of the subunits (see figure 4.16). Two consecutive rounds of refinement were then performed using simulated annealing (2,500K), positional, overall B-factor refinement for each protomer domain, and grouped B-factor refinement for residue side chain and main chain groups (rounds 5-6).

At this point the model was sufficiently refined for the calcium ions, Ro 63-8695 molecules, and sugar residues to be built into the electron density. 10-fold averaged electron density maps were generated and showed good density for one Ro 63-8695 molecule spanning each pair of adjacent double calcium-binding sites. However, there was insufficient resolution to individually refine the calcium ions and Ro 63-8695 molecules separately. The higher resolution SAP-N-acetyl-D-proline structure (see section 3.2) provided an initial estimate of their coordinate positions. The superpositions of the calcium ions and N-acetyl-D-proline molecules fitted the density well (see figure 4.17), with the calcium ions and coordinating side chains requiring only minor changes. The alkyl linker bridging the adjacent N-acetyl-D-proline molecules was then built in using five-fold averaged electron density maps and ideal bond lengths and angles (Engh & Huber 1991). Although all possible combinations were separately refined, the cis-cis isomer of the peptide bond preceding the head group of Ro 63-8695 was chosen over the trans-trans and trans-cis isomers as the optimum configuration (discussed in detail in section 4.4). Finally, the first two sugar residues of the glycosylation at Asn32 were built onto each protomer (see figure 4.18). Calcium ions and sugar residues were added to the high weighted protomer (10-fold) NCS restraints, while separate high weighted NCS restraints (wa=300)

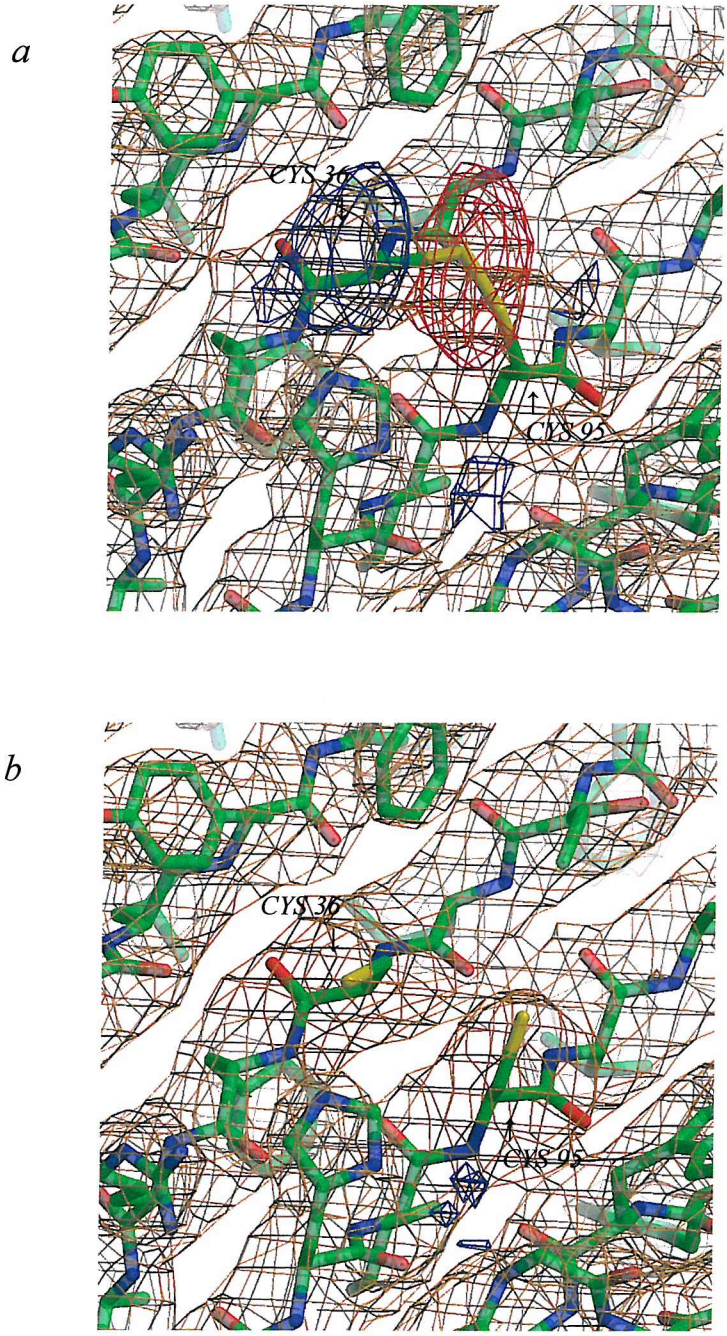


Figure 4.16. Confirmation of the broken disulphide bridge between Cys36-Cys95 (prepared using Bobscript and Raster3D). The 10-fold averaged $|F_{obs}|-|F_{calc}|$ electron density shows strong negative density (depicted in red and contoured at -3σ) surrounding the sulphur atoms and strong positive density (depicted in blue and contoured at 3σ) adjacent to the disulphide bridge, suggesting the Cys95 sulphur is swung slightly away with Cys36 orientated in the opposite direction (*a*). The final model shows no surrounding density for the broken disulphide and/or the reoriented cysteine side chains (*b*). 10-fold averaged $2|F_{obs}|-|F_{calc}|$ density is contoured at 1.3σ (depicted in orange), with the protein shown as stick models.

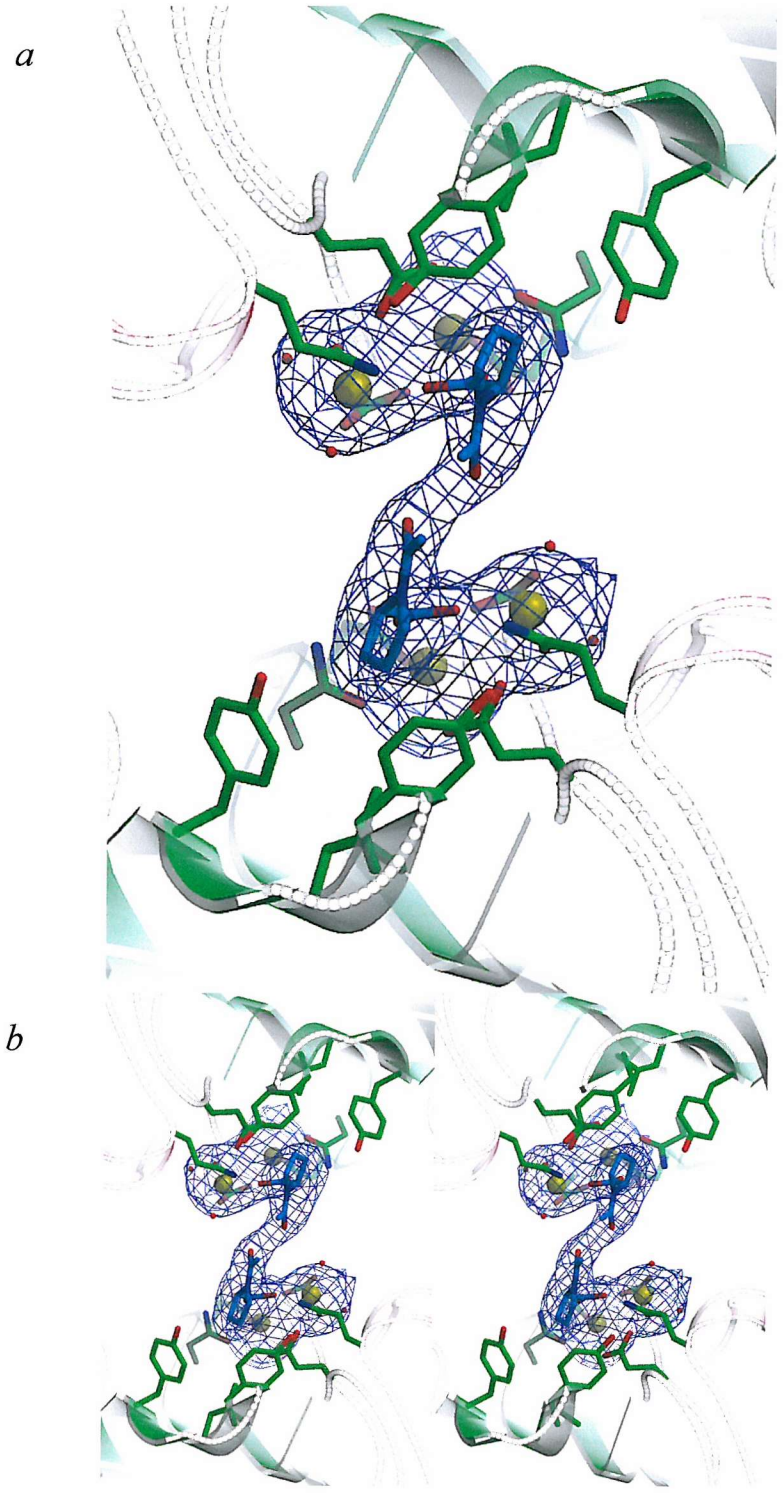


Figure 4.17. Superposition of N-acetyl-D-proline molecules, calcium ions, and water molecules from subunit A of the SAP-N-acetyl-D-proline structure used as an initial estimation of Ro 63-8695 coordinates viewed as an orthogonal projection (*a*) and stereoview (*b*) (prepared using Bobscript and Raster3D). The 10-fold averaged $|F_{obs}|$-$|F_{calc}|$ electron density is contoured at 3σ (depicted in blue), secondary structural elements are represented as previously described with N-acetyl-D-proline and important side chains shown as stick models, and water molecules as red spheres.

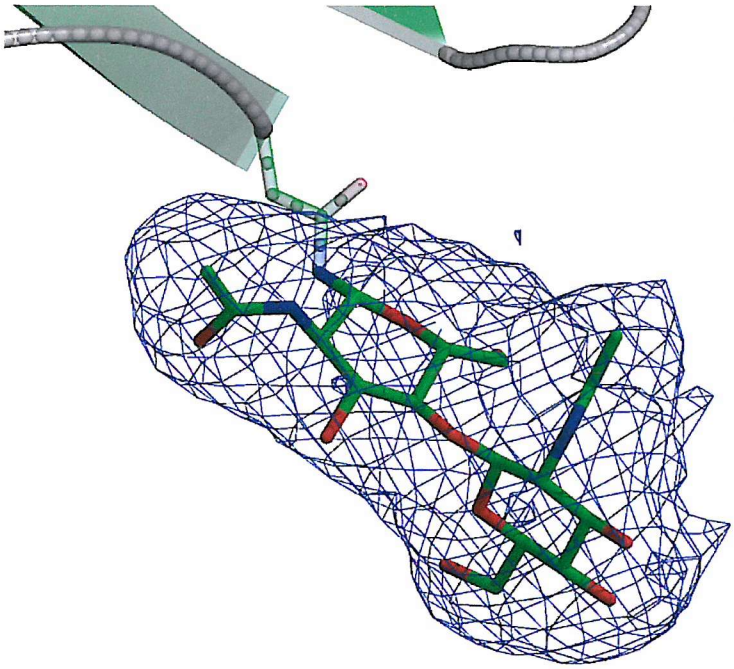


Figure 4.18. Ten-fold averaged $|F_{obs}|$-$|F_{calc}|$ electron density for the glycosylation at Asn32 (prepared using Bobscript and Raster3D). Electron density contoured at 3σ (depicted in blue) was sufficient for the first two sugar residues to be built in. The final positions of these sugar residues in the fully refined structure are also shown. Secondary structural elements are represented as previously described with Asn32 and N-acetylglucosamine sugar residues shown as stick models.

were applied to the five Ro 63-8695 molecules. Subsequent rounds of grouped B-factor refinement were performed with a 100Å² cut-off, using the side chains, main chains, calcium ions, sugar residues, and Ro 63-8695 molecules as separate groups (rounds 7-19). The Ro 63-8695 NCS restraints were later relaxed between rounds 9 and 10 as variations in the unaveraged electron density maps became apparent. During the later stages of the refinement process the calcium ions began to deviate from the optimum coordination. Hence, water molecules were subsequently built into weak $|F_{obs}|$-$|F_{calc}|$ density during model building to complete the coordination of calcium site II before the final rounds of grouped B-factor refinement were performed (round 18-19).

The final round of grouped B-factor refinement gave a *R-factor* of 23.0% (*R-free* 23.6%) with an average B-factor of 79.0Å² (100Å² cut-off). High weighted protomer (10-fold) NCS restraints were maintained giving an observation to parameter ratio of 22:1 (calculated assuming constraints). The resulting model contains zero occupancy for the side chains of Glu27, Lys28, and Lys143, and shows good stereo-chemical properties with no regions within the disallowed regions of the Ramachandran plot (see figure 4.19). The overall statistics for the fully refined structure are summarised in table 4.5.

4.4 Structure Analysis of the SAP-Ro 63-8695 Complex

The structure of the SAP-Ro 63-8695 complex is well defined with all ten protomers virtually identical (C_α rms fit of 0.001Å) due to the high weighted protomer (10-fold) NCS restraints. Despite the resolution, there are only minor structural deviations when compared to that of SAP-N-acetyl-D-proline. These are localised to three turn regions within the exposed loops, resulting in the loss of the γ-turns between residues 24-26 and 139-141, and the loss of a stabilising hydrogen bond between residues *i* and *i*+3 of a

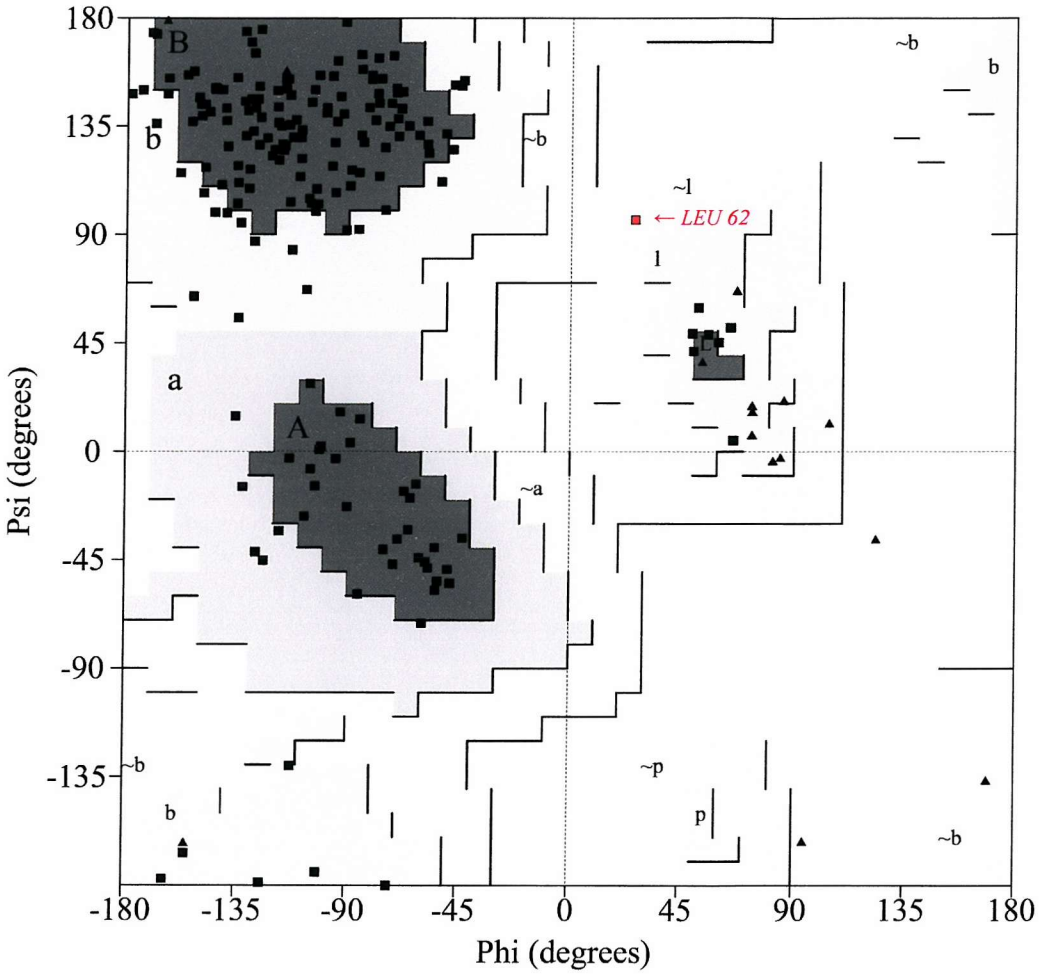


Figure 4.19. Ramachandran plot for one subunit from the final SAP-Ro 63-8695 structure (prepared using Procheck). Statistics for non-glycine and non-proline residues (175 residues) indicate 78.3% of residues in most favoured regions [A, B, L], 21.1% residues in additional allowed regions [a, b, l, p], 0.6% residues in generously allowed regions [~a, ~b, ~l, ~p], and 0.0% residues in disallowed regions. Glycine residues are represented as black triangles, with all other residues as black squares with the exception of the 1 residue within the generously allowed region (Leu62) which is highlighted in red.

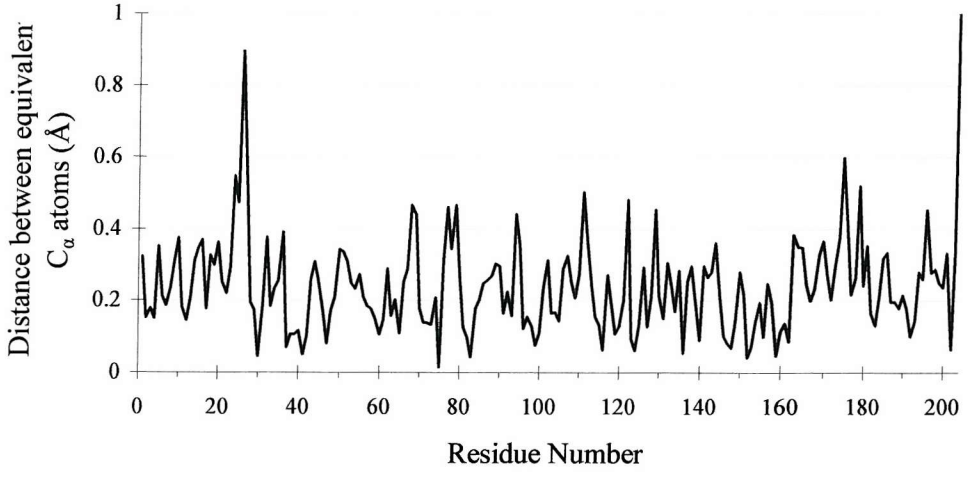
Table 4.5. Summary of the processing and model statistics for the SAP-Ro 63-8695 structure.

Space Group	$P4_32_12$
Unit Cell	a=b=230.9Å c=281.4Å
Unit Cell Volume (Å^3)	14,997,799
Resolution Range (Å)	25 - 3.2
Total Number of Reflections	442,285
Unique Number of Reflections	119,477
Multiplicity	3.7
Mosaicity (°)	0.5
Completeness (%)	95.7 (85.9)
R-merge(I) (%)	8.6 (77.9)
I/σ(I) (%>2.0)	70.4 (9.1)
Molecules in Asymmetric Unit	2 SAP Pentamers
	20 Calcium Ions
	5 Ro 63-8695 Molecules (**cis-cis**)
	20 Sugar Residues
	20 Water Molecules
Solvent Content (%)	83.3
Model *R-factor* (%)	23.0
Model *R-free* (%)	23.6
rms Bond Lengths (Å)	0.010
rms Bond Angles (°)	1.60

Figures in parentheses denotes the highest resolution shell statistics between 3.31-3.20Å

type I′ β-turn (residues 75-78). There is also a clear sign of radiation damage with cleavage of the disulphide bond between Cys36-Cys95 (Burmeister 2000; Ravelli & McSweeney 2000), although this does not cause any structural destabilisation of the β-strands involved (C and H). The similarity with subunit A from SAP-N-acetyl-D-proline shows the protein structure to be comparable to the higher resolution structures presented previously (see chapter 3) with an average C_α rms fit of 0.14Å. The only significant main chain deviation occurs between strands B and C (see figure 4.20), resulting from the loss of the γ-turn between residues 24-26.

Figure 4.20. C_α distances between a SAP subunit from the SAP-Ro 63-8695 complex superimposed with subunit A from the SAP-N-acetyl-D-proline structure.



The pentameric assembly of SAP is stabilised by similar interactions to those described previously for SAP-N-acetyl-D-proline (see section 3.2.6). All of the salt bridges and direct hydrogen bonds are present. However, due to the resolution and absence of

water molecules, an additional salt bridge is formed between the side chains of Glu153-Lys116. The SAP pentamers are stacked in pairs via the B-faces creating two-fold symmetry between the adjacent pentamers and a common five-fold radial symmetry perpendicular (see figure 4.21). The pentamers are reversibly cross-linked by 5 Ro 63-8695 molecules creating a stable 75x95x95Å decameric complex. There are no close contacts between the two pentamers as a clear solvent divide separates them. The closest point is between the main chain carbonyls of adjacent Asp138 residues (6.4Å apart). However if fully extended, the indiscernible side chain of Lys143 is capable of bridging the solvent divide at the exterior edge (within 2Å), although there are no groups within hydrogen-bonding or salt-bridging distance.

The decamer complex is generated by a crystallographic dyad, and interacts with three surrounding decamers (subunits distinguished by labelling adjacent pentamers 1 and 2, with dimers of adjacent subunits labelled clockwise from A to E when viewed down the common 5-fold axis from the A-face of pentamer 2). Packing interactions occur between the exterior loop regions of subunits B-B, and C-E (see table 4.6). Four salt bridges are formed between the adjacent B subunits, with the propensity to form two additional salt bridges between Arg57-Glu126 via the reorientation of Arg57. There is also strong density for two objects linking adjacent 55-56 and 139-142 residues. Each appears to be of the correct size and shape for an individual sugar residue (see figure 4.22). Although there is no preceding density and they are too far apart to be consecutive, there is a sufficient gap between adjacent decamers for a preceding oligosaccharide to egress. Stronger packing interactions are formed between subunits C and E creating four hydrogen bonds and two salt bridges. There is also the propensity to form four additional hydrogen bonds between Gln55-Gly141 and Gln128-Gly194 through the reorientation of Gln55 and Gln128. The packing interactions are likely to cause minor structural variances between the subunits

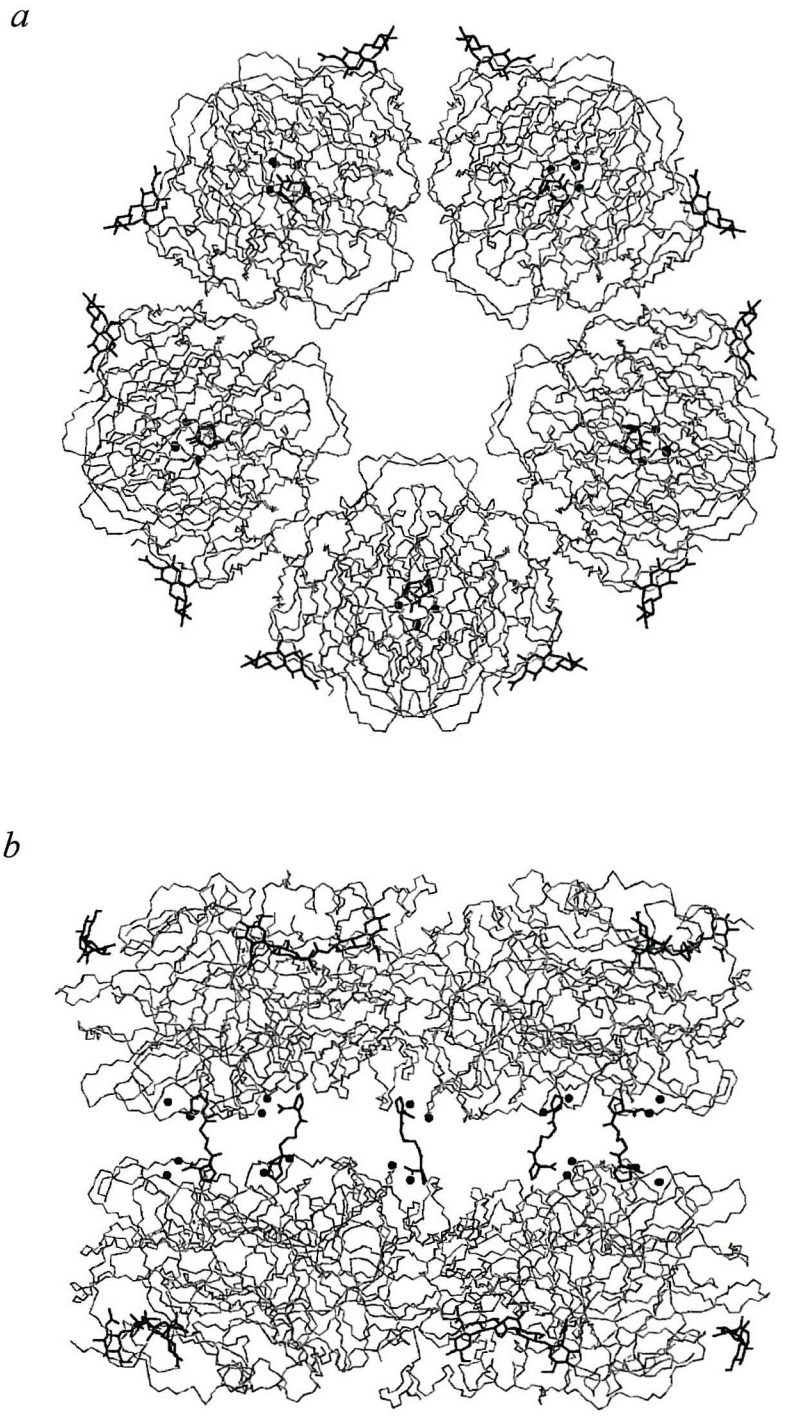


Figure 4.21. Three-dimensional structure of the SAP-Ro 63-8695 complex illustrating the common five-fold radial symmetry (*a*) and two-fold symmetry between adjacent pentamers (*b*) (prepared using Bobscript and Raster3D). The SAP pentamers are represented as C_α traces shown as grey stick models, Ro 63-8695 molecules and sugar residues as black stick models, and calcium ions as black spheres.

Table 4.6. Packing contacts between neighbouring SAP decamers (subunits distinguished by labelling adjacent pentamers 1 & 2, with dimers of adjacent subunits labelled clockwise from A to E when viewed down the common 5-fold axis from the A-face of pentamer 2)

Subunit Contacts (*i* - *i*+1)	Residues (subunit *i*)	Residues (subunit *i*+1)	Interactions (inter-atomic distance < 3.2Å)
B1-B1	139, 143, 191-193	55, 191-193	Glu191-Arg193* Arg193-Glu191*
B1-B2	55-57, 126	55-57, 126	Arg57-Glu126** Glu126-Arg57**
B2-B2	55, 191-193	139, 143, 191-193	Glu191-Arg193* Arg193-Glu191*
C1-E1	27, 55, 128-130, 140	17-19, 130, 140-142, 191-194	Gln55-Gly141†† Gln128-Gly194†† Lys130-Glu191* Tyr140-Arg193†
C2-E1	55-56	55-56	Gln55-Gly56† Gly56-Gln55†
C2-E2	17-19, 130, 140-142, 191-194	27, 55, 128-130, 140	Gly141-Gln55†† Glu191-Lys130* Arg193-Tyr140† Gly194-Gln128††

* Denotes salt bridge
** Denotes possible salt bridge
† Denotes hydrogen bond
†† Denotes possible hydrogen bond

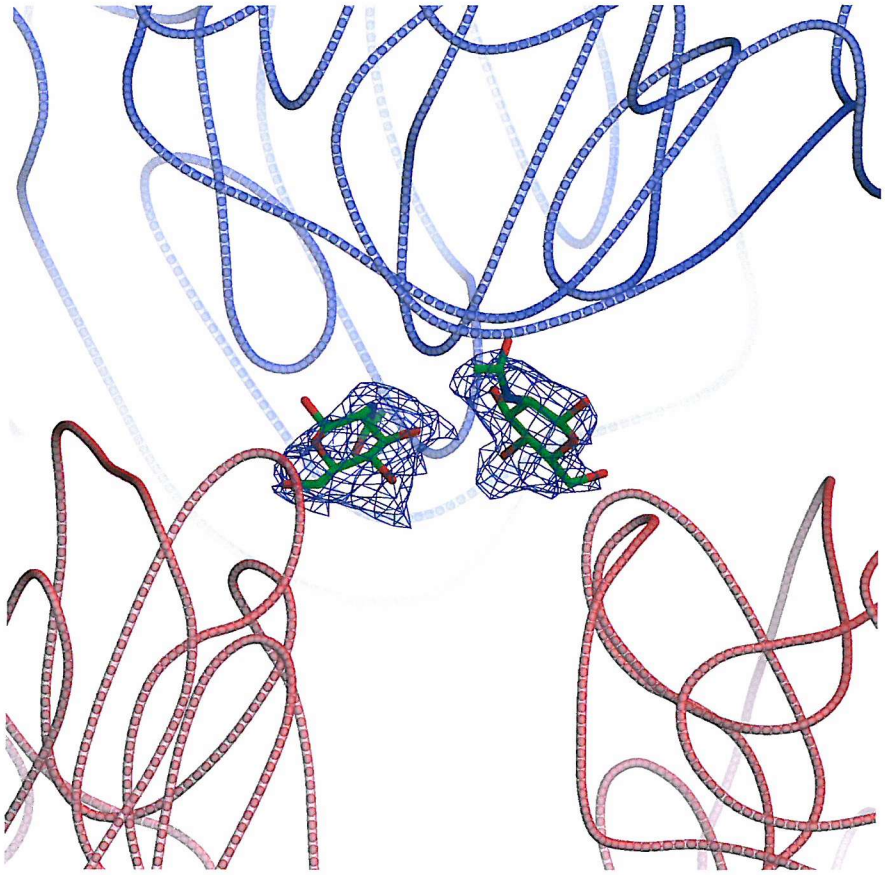


Figure 4.22. Packing contact between the B subunits of neighbouring SAP decamers showing sufficient density for two sugar residues between adjacent 55-56 and 139-142 residues (prepared using Bobscript sand Raster3D). The 10-fold averaged $|F_{obs}|$-$|F_{calc}|$ electron density is contoured at 3σ (depicted in blue), with the neighbouring decamers represented as red and blue cords and the potential N-acetylglucosamine sugar residues shown as stick models.

introducing discrepancies that, due to the tight NCS restraints, will cause minor structural errors. Interestingly the residues with no discernable orientation (Glu27, Lys28, and Lys143) and those within the lost γ-turn structures are either nearby or involved in the packing interactions.

The double calcium-binding site shows seven ligands for both calcium ions, with the coordination spheres completed by a carboxylate from Ro 63-8695 and two water molecules. This is the same as described previously for the higher resolution SAP-N-acetyl-D-proline structure (see section 3.2.8). The calcium ions are positioned 4.0Å apart with the calcium-ligand interactions an average of 2.48Å. In the decamer complex the adjacent double calcium-binding sites are held an average of 9.4Å apart and are bridged by 5 Ro 63-8695 molecules burying only 1.3% (420Å^2) of each proteins surface area. Binding of Ro 63-8695 shows a very close superposition with that of N-acetyl-D-proline, where each carboxyl bridges the two bound calcium ions forming hydrogen bonds with Asn59 and Gln148, orientating the pyrrolidine ring into the adjacent hydrophobic pocket (see figure 4.23). Packing of the pyrrolidine ring is equally as tight against Tyr74 (C_γ 3.6Å away) and Tyr64 (C_γ 3.7Å away), although the insertion distance is slightly greater with Leu62 pushed further towards the protomers centre of mass (C_β 4.0Å away). The adjacent binding sites are related by a two-fold axis, which is accommodated by the alkyl linker. This adopts a kinked rotamer, displacing the rigid head groups (1-acetyl-pyrrolidine-2-carboxylic acid) from the common long axis and facilitating their approach to the adjacent binding sites.

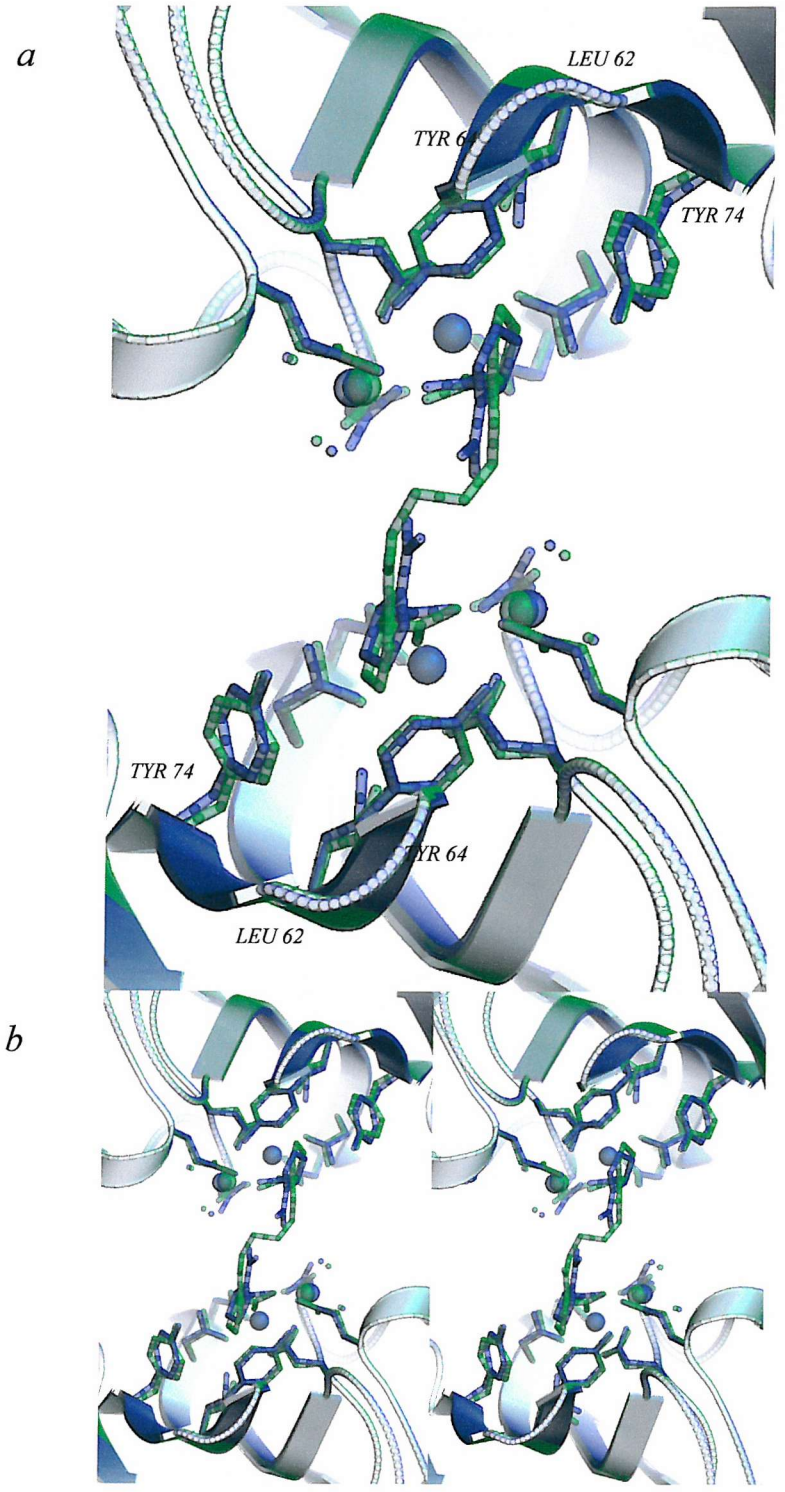


Figure 4.23. Superposition of the ligand-binding site from subunit A of the SAP-N-acetyl-D-proline crystal structure (blue) with the two-fold axis related binding sites of the Ro 63-8695 complex (green) viewed as an orthogonal projection (*a*) and stereoview (*b*) (prepared using Bobscript and Raster3D). Secondary structural elements are represented as previously described with calcium ions as large spheres, water molecules as small spheres, and important residues and molecules shown as stick models.

The peptide bonds allow the Ro 63-8695 to adopt three different isomeric conformations: the trans-trans, trans-cis, and cis-cis forms (see figure 4.24). Each conformer can be built in following a similar path between the head groups with the central carbons (C_2-C_3) adopting anti (staggered) substituents in the trans-trans, compared to eclipsed substituents in both the trans-cis and cis-cis isomers. The peptide conformation will inherently dictate the direction of the carbonyl group and the alkyl linker. In the trans-trans conformer the carbonyl oxygens are directed towards their adjacent double calcium-binding site, within 3.9Å of water II, forcing all 4 carbons of the linker out away from the electron density. In the cis-cis conformer the carbonyls are directed away from the double calcium-binding sites allowing the linker to closely follow the direction of the density. The trans-cis conformer contains both components, requiring a small rotational shift in the head groups to enable the linker to transverse the peptide bonds. Not only does this force two of the linker carbons (C_1-C_2 trans side) and the cis carbonyl out away from the electron density, but also causes a steric clash where the carbonyl oxygens and linker carbons are closer than the Van der Waals' contact distance (3.22Å between carbon and oxygen atoms (Nyburg & Faerman 1985)). Examination of the electron density maps from their separately refined structures confirms the deviation of the alkyl linker in the trans-trans and trans-cis isomers with strong negative difference density for the protruding alkyl atoms and equally strong positive density for the correct path (see figure 4.24*c*). Electron density maps generated from the cis-cis conformation shows the linker following the $2|F_{obs}|-|F_{calc}|$ density with occasional difference density nearby.

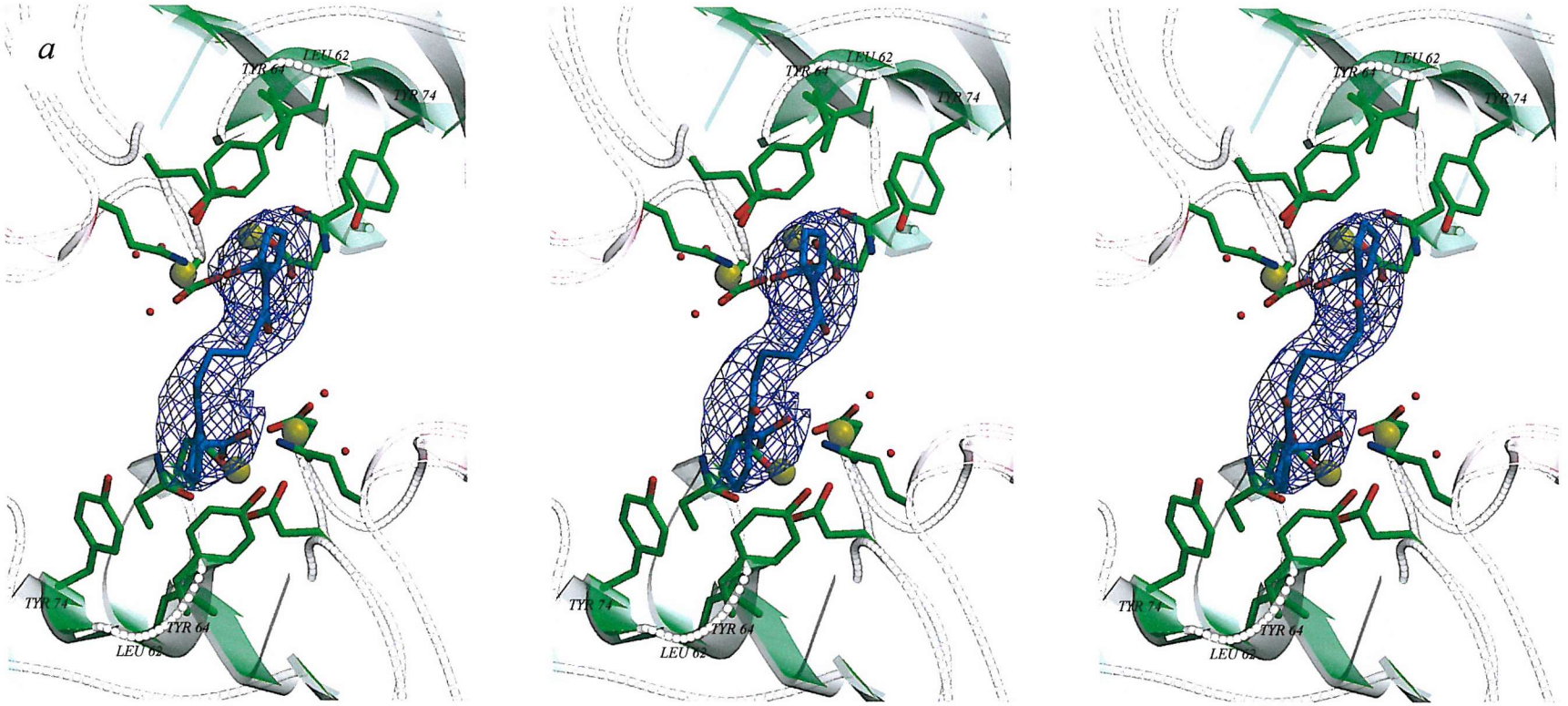
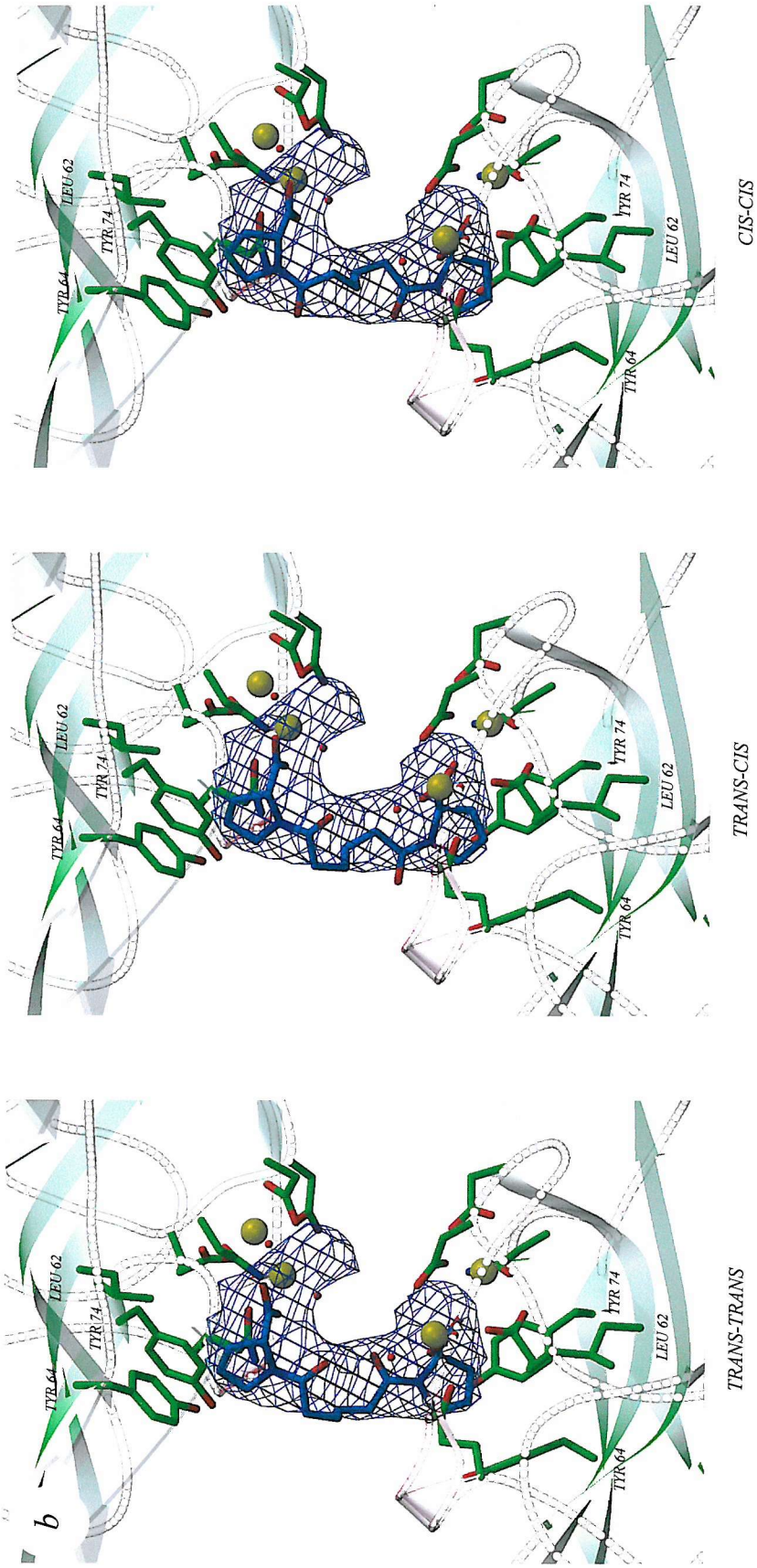
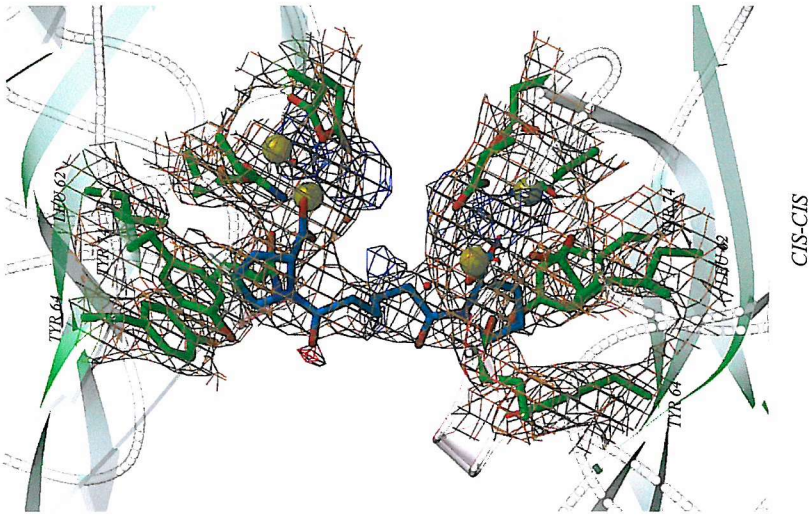
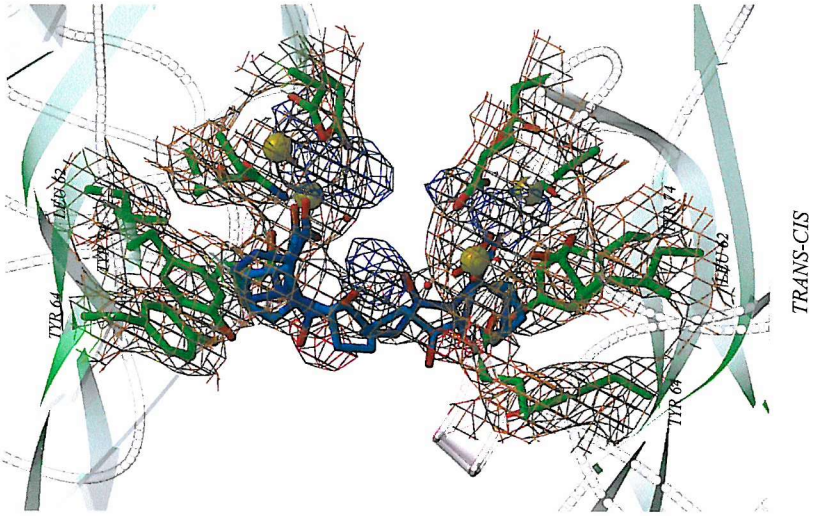


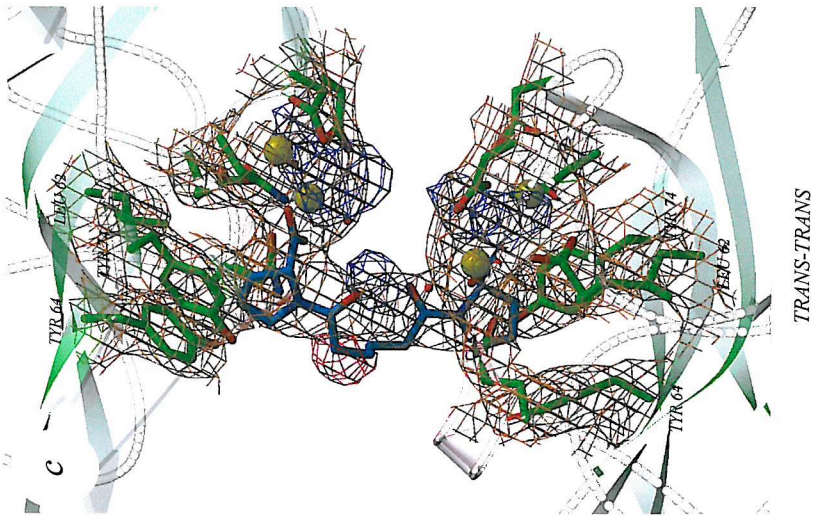
Figure 4.24. Comparison of the trans-trans, trans-cis, and cis-cis conformations of Ro 63-8695 binding between the two-fold axis related SAP subunits of the SAP-Ro 63-8695 complex to the experimental data (prepared using Bobscript and Raster3D). *a* & *b* show the agreement with $|F_{obs}|$-$|F_{calc}|$ annealed (slow cool from 1,000K in 25K decrements) omit map contoured at 3σ (depicted in blue) viewed from two views perpendicular to one another. *c* shows the agreement with $2|F_{obs}|$-$|F_{calc}|$ density contoured to 1.3σ (depicted in orange) and $|F_{obs}|$-$|F_{calc}|$ density contoured to 3σ (depicted in blue) and -3σ (depicted in red), two molecules for the trans-cis conformation are shown in alternative orientations refined with half occupancy; it should be noted that the strong positive density located between the calcium ions is due to averaging of the calcium coordinates and is not present in the ten-fold averaged $|F_{obs}|$-$|F_{calc}|$ map. The binding of the Ro 63-8695 head groups to the adjacent double calcium-binding sites and insertion of the pyrrolidine ring can be seen with kinking of the alkyl linkers. The cis-cis conformer shows the best interpretation of the data, with the alkyl linker clearly following the electron density without the linker atoms protruding out from the electron density. Secondary structural elements are represented as previously described with Ro 63-8695 molecules shown as stick models and water molecules as red spheres.


LEU 62
TYR 74
TYR 64
TYR 74
LEU 62
TYR 64
CIS-CIS
LEU 62
TYR 74
TYR 64
TYR 74
LEU 62
TYR 64
TRANS-CIS
LEU 62
TYR 74
TYR 64
TYR 74
LEU 62
TYR 64
TRANS-TRANS

b


TYR 64
LEU 62
TYR 64
CIS-CIS


TYR 64
LEU 62
TYR 64
TRANS-CIS


TYR 64
LEU 62
TYR 64
TRANS-TRANS

4.5 *Isomeric State of Ro 63-8695*

The occurrence of the different Ro 63-8695 conformers in solution is dependent upon the stability of the trans and cis peptide bonds. This is predominantly determined by the minimum repulsion energy. Quantum-mechanical calculations have predicted the difference between the most stable trans and cis conformations of proline to be only 0.5kcal/mole in favour of the trans isomer (Maigret *et al* 1970). The energy barrier for cis-trans isomerisation has also been calculated and is 16.2kcal/mole (ω=180°→90°→0°). On the basis of the potential energy difference, the probability of occurrence is 60-70% for the trans and 30-40% for the cis peptide bond (Ramachandran & Mitra 1976). This has been found to be the case in small proline-containing molecules using nuclear magnetic resonance (NMR) techniques (Deber *et al* 1970).

One and two-dimensional NMR techniques were utilised to elucidate the conformational proportions of Ro 63-8695 in deuterated water (D_2O). Although there are only small conformational distinctions between the three isomers, the shielding differences will significantly separate their chemical shifts (δ). A one-dimensional ^1H NMR spectrum for Ro 63-8695 was initially measured (see figure 4.25) and showed 15 individual resonance peaks producing a pattern characteristic of previously defined chemical shifts for proline ($C_\alpha H$=4.42, $C_\beta H_2$=2.29 & 1.94, $C_\gamma H_2$=2.02, 2.02, and $C_\delta H_2$=3.63, 3.63 parts per million (ppm) (Wishart *et al* 1995)) and the alkyl groups (CH_2=1.2-2.2 ppm (Van Holde *et al* 1998)). Although the characteristic intense and neighbouring weaker peaks would suggest a disproportionate distribution of these conformers, the ^1H NMR spectrum was found to be too complicated for the assignment of individual resonating protons (22 per conformer) via simple integration. Thus, a two-dimensional Fourier transform total

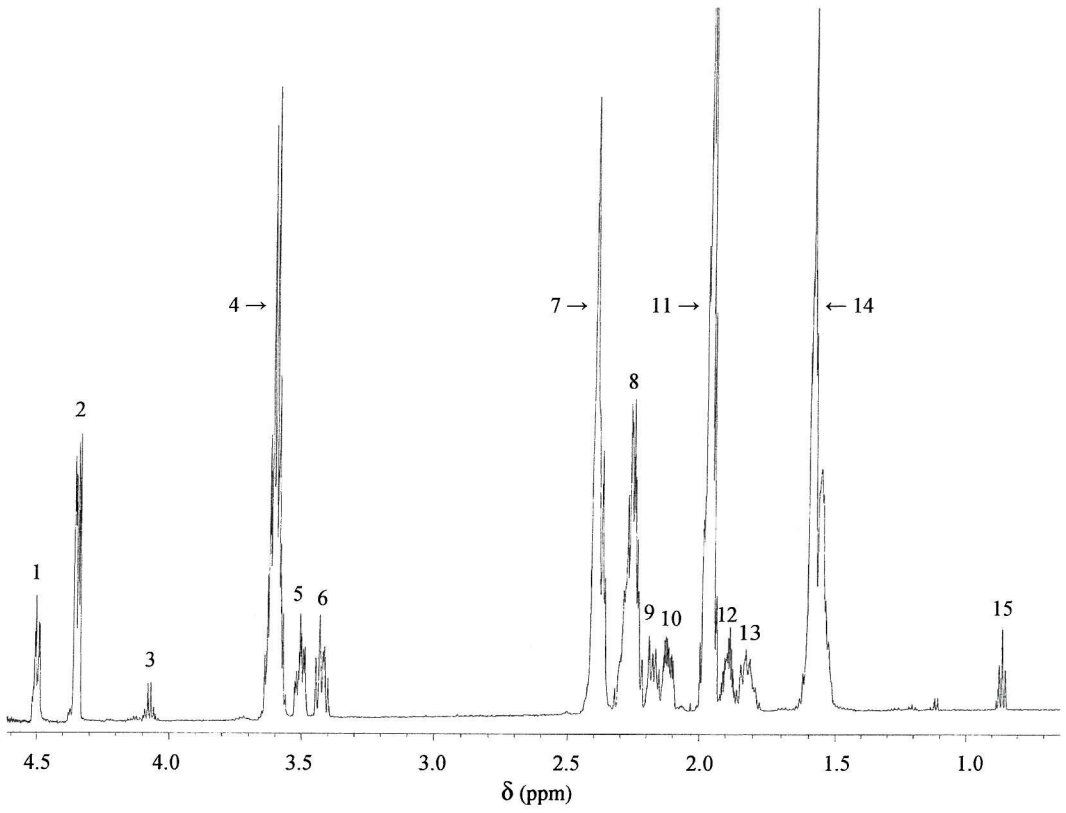


Figure 4.25. One-dimensional ^{1}H NMR spectrum of Ro 63-8695 at 22°C recorded in-house on a Varian Inova-600MHz instrument. 1.7mM Ro 63-8695 was dissolved in 100% D_2O, and data acquired with a spectral width of 8,000Hz and an acquisition time of 2 seconds. 32 repetitions were recorded with a relaxation time of 1.2 seconds between them. Assignment and integration of the peaks (labelled 1 to 15) is shown in table 4.7.

correlation spectroscopy (TOCSY) spectrum was recorded to elucidate through-bond interactions and enable an accurate assignment. Cross-peaks corresponding to the pyrrolidine protons show a clear separation between the chemical shifts for the trans and cis isomers (see figure 4.26), but with no distinction between the different conformers. The trans protons resonate at the same chemical shifts to those previously defined for proline. Although the chemical shifts for cis protons are similar, there is a clear separation of the C_β, C_γ, and C_δ protons. Detailed examination of the 1.4-2.6 ppm section showed separate cross-peaks corresponding to the alkyl protons from the 3 different conformers (see figure 4.27). A two-dimensional Fourier-transform rotating frame Overhauser effect spectroscopy (ROESY) spectrum was recorded to elucidate through-space interactions (<5Å) and confirm the assignment (see figure 4.28). Chemical shifts for the trans-trans conformer were confirmed by the through-space correlations between the C_1-C_δ and C_2-C_δ protons. Unfortunately equivalent C_1H-$C_\alpha H$ and C_4H-$C_\alpha H$ correlations that might be expected for the cis-cis and trans-cis conformers respectively are hidden under the stronger cross-peaks for the pyrrolidine ring.

The full spectral assignment for each of the 66 individual protons is given in table 4.7. Integration of the individual peaks shows good correlation with a slight loss in signal intensity of the C_α proton from the cis (peak 1) and trans (peak 2) isomers due to the partial saturation from the D_2O irradiation. Due to the lack of known variables the proportions of the individual conformers cannot be accurately calculated, however the overall proportions of trans and cis isomers can. Comparison of the individual C_δ integrals (peak 4 with peaks 5 and 6) shows an overall trans population of 75% and an overall cis population of 25% for the head groups of Ro 63-8695. Therefore, assuming no advantageous or disadvantageous interactions occur between the head groups, probability calculations dictate that 56% of

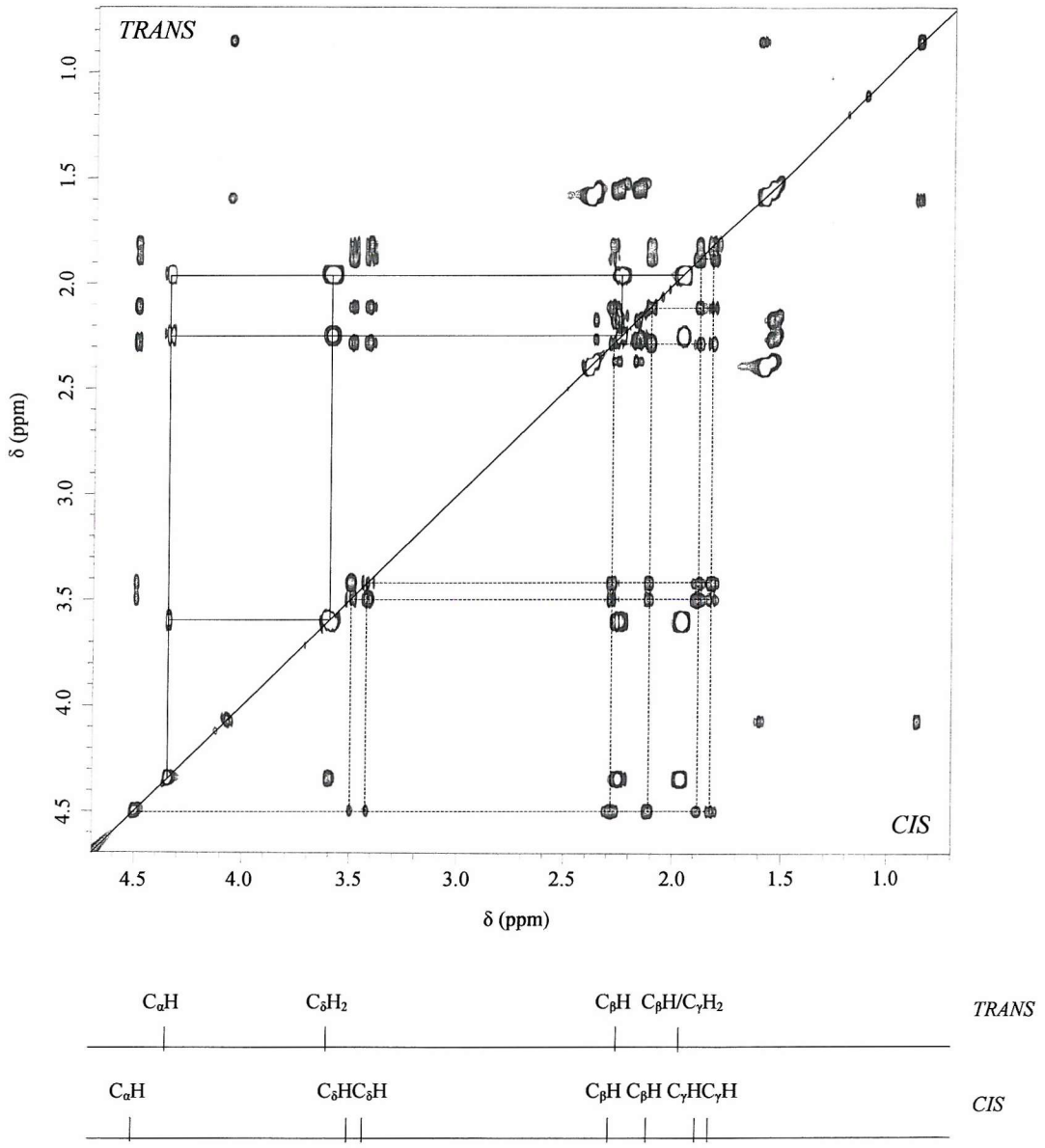


Figure 4.26. Two-dimensional ^{1}H Fourier transform TOCSY spectrum of Ro 63-8695 recorded in-house on a Varian Inova-600MHz instrument at 22°C. 1.7mM Ro 63-8695 was dissolved in 100% D_2O, with data acquired using 512 complex points in t_1, 4096 complex points in t_2, and a spectral width of 6,000Hz. The acquisition time was 341ms, relaxation time 2.5 seconds, spin-locking mixing time 50ms, with each increment scanned 16 times. The assignment of the pyrrolidine ring protons is given for the trans (solid lines) and cis (broken lines) isomers.

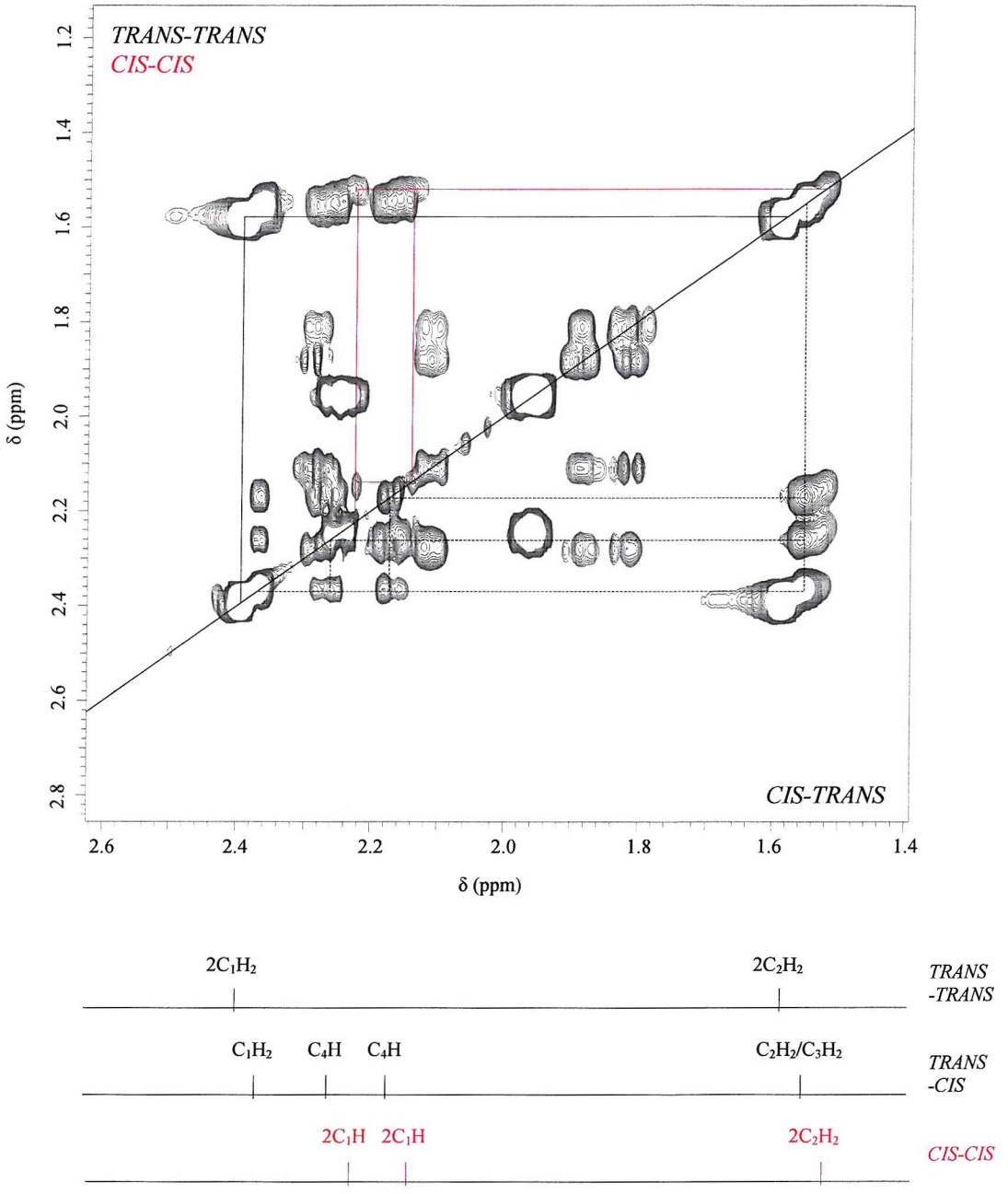


Figure 4.27 A portion of the two-dimensional ^1H Fourier transform TOCSY spectrum of Ro 63-8695 recorded in-house on a Varian Inova-600MHz instrument at 22°C. 7mM Ro 63-8695 was dissolved in 100% D_2O, with data acquired using 512 complex points in t_1, 4096 complex points in t_2, and a spectral width of 6,000Hz. The acquisition time was 341ms, relaxation time 2.5 seconds, spin-locking mixing time 50ms, with each increment scanned 16 times. The assignment of the alkyl linker protons is given for the trans-trans (solid lines), trans-cis (broken lines), and cis-cis (red lines) conformers.

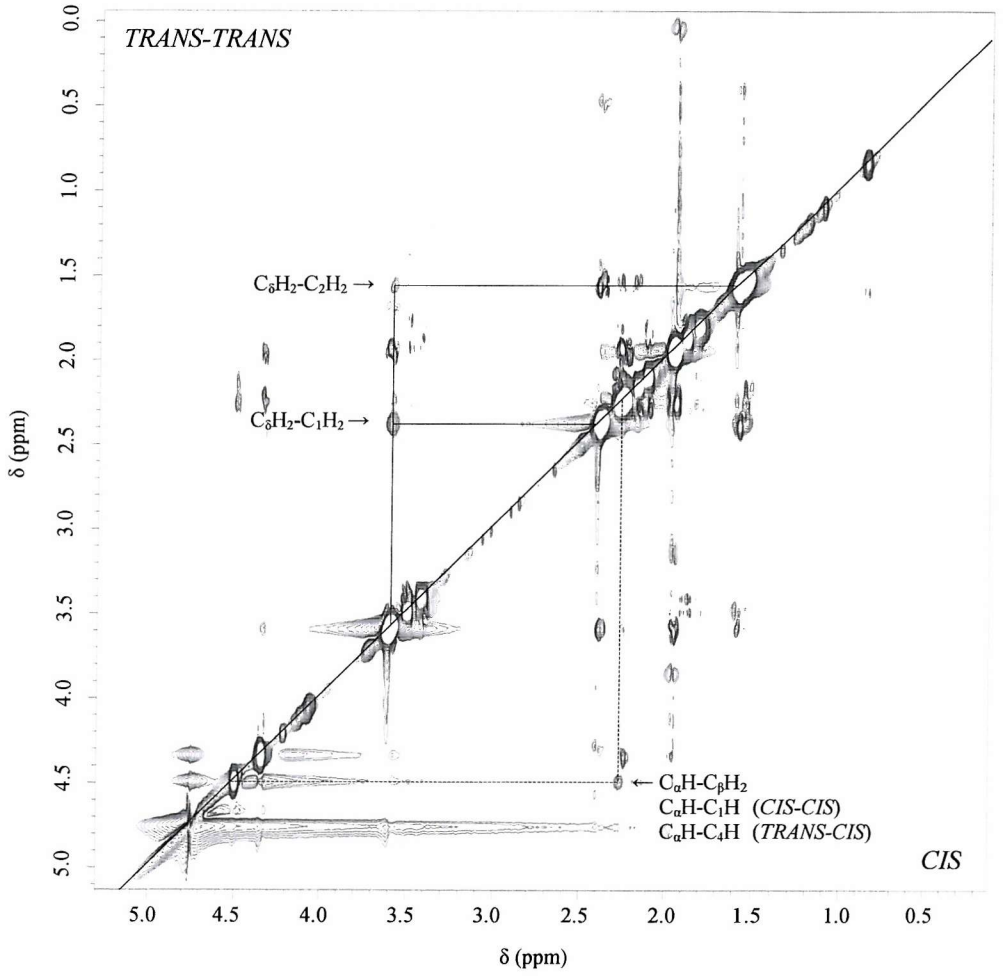


Figure 4.28. Two-dimensional ^1H Fourier transform ROESY spectrum of Ro 63-8695 recorded in-house on a Varian Inova-600MHz instrument at 22°C. 1.7mM Ro 63-8695 was dissolved in 100% D_2O, with data acquired using 512 complex points in t_1, 4096 complex points in t_2, and a spectral width of 6,000Hz. The acquisition time was 341ms, relaxation time 1.5 seconds, spin-locking mixing time 75ms, with each increment scanned 48 times. Cross-peaks are present connecting the $C_\delta H_2$-C_1H_2 and $C_\delta H_2$-C_2H_2 for the trans-trans conformer (solid lines), but the equivalent $C_\alpha H$-C_4H and $C_\alpha H$-C_1H cross-peaks for the trans-cis and cis-cis conformers are hidden under the strong $C_\alpha H$-$C_\beta H_2$ cross-peak (broken lines) for the cis head group.

Table 4.7. Resonance assignment and integration of the 15 individual peaks from the one-dimensional ^1H NMR spectrum for Ro 63-8695 shown in figure 4.25.

Peak Number	Proton Resonance Contribution				Relative Integration
	Trans-Trans	Trans-Cis	Cis-Cis	Unknown	
1		$C_\alpha H$	$2C_\alpha H$		0.13
2	$2C_\alpha H$	$C_\alpha H$			0.37
3				*x*	-
4	$2C_\delta H_2$	$C_\delta H_2$			1.00
5		$C_\delta H$	$2C_\delta H$		0.18
6		$C_\delta H$	$2C_\delta H$		0.15
7	$2C_1H_2$	C_1H_2			1.04
8	$2C_\beta H$	$C_\beta H$, $C_\beta H$, C_4H	$2C_\beta H$ $2C_1H$		0.82
9		C_4H	$2C_1H$		0.15
10		$C_\beta H$	$2C_\beta H$		0.17
11	$2C_\beta H$, $2C_\gamma H_2$	$C_\beta H$, $C_\gamma H_2$			1.47
12		$C_\gamma H$	$2C_\gamma H$		0.32
13		$C_\gamma H$	$2C_\gamma H$		
14	$2C_2H_2$	C_2H_2, C_3H_2	$2C_2H_2$	*x*	1.37
15				*x*	-

Ro 63-8695 will be in the trans-trans conformer, 38% in the trans-cis conformer, and only 6% in the cis-cis conformer.

Consequently a one-dimensional ^1H NMR spectrum was measured for the head group of Ro 63-8695 (N-acetyl-D-proline) to elucidate the effect of the linker upon cis-trans isomerisation (see figure 4.29). The spectrum showed 10 resonance peaks similar to those for the pyrrolidine rings of Ro 63-8695. Two-dimensional Fourier-transform TOCSY spectrum also showed similar through-bond interactions and clearly elucidating the pyrrolidine and terminal methyl protons (see figure 4.30). Integration of the individual peaks (see table 4.8) shows good correlation with a slight loss in signal intensity of the C_α protons (peaks 1 and 2) due to partial saturation from the D_2O irradiation, and also of the C_1 (peaks 7 and 8) protons due to solvent exchange with D_2O. Comparison of the $C_\delta H$ integrals (peak 3 with peaks 4 and 5) shows an overall trans proportion of 68% and an overall cis proportion of 32%. The cis proportion is significantly higher than that found in Ro 63-8695. Thus, there appears to be some disproportionate effect stabilising the trans isomer and/or destabilising the cis isomer in solution, resulting in a 10% increase in the trans-trans, while reducing both the trans-cis and cis-cis populations by 6% and 4% respectively.

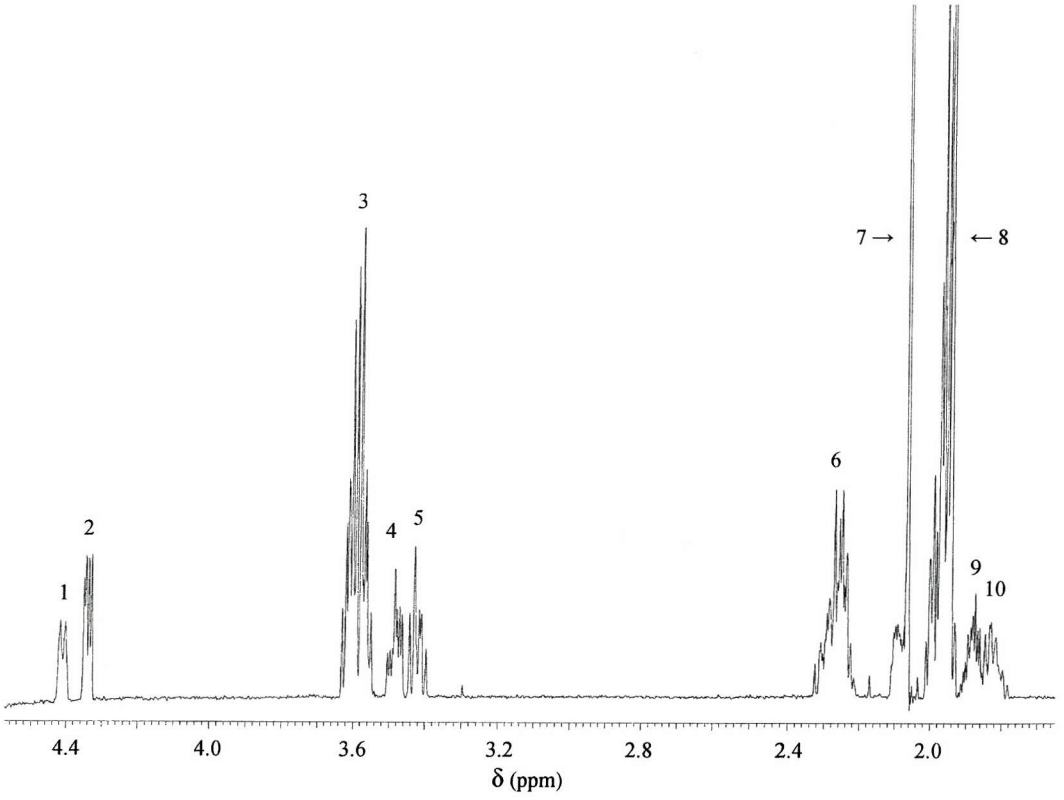


Figure 4.29. One-dimensional ^{1}H NMR spectrum of N-acetyl-D-proline at 22°C recorded in-house on a Varian Inova-600MHz instrument. 2.5mM N-acetyl-D-proline was dissolved in 100% D_2O, and data acquired with a spectral width of 8,000Hz and an acquisition time of 2 seconds. 32 repetitions were recorded with a relaxation time of 1.2 seconds between them. Assignment and integration of the peaks (labelled 1 to 10) is shown in table 4.8.

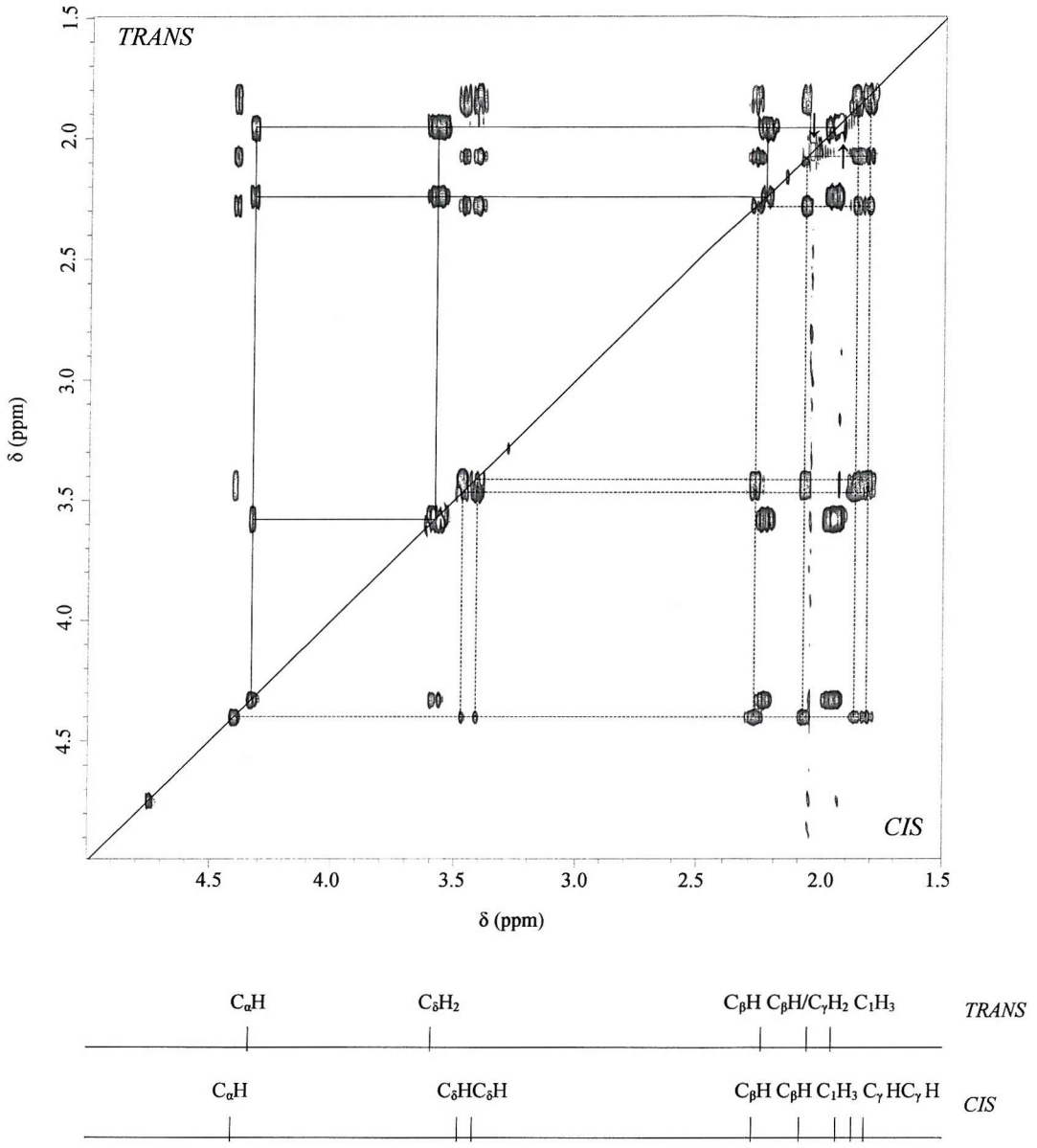


Figure 4.30. Two-dimensional ^{1}H Fourier transform TOCSY spectrum of N-acetyl-D-proline recorded in-house on a Varian Inova-600MHz instrument at 22°C. 2.5mM N-acetyl-D-proline was dissolved in 100% D_2O, with data acquired using 512 complex points in t_1, 4096 complex points in t_2, and a spectral width of 6,000Hz. The acquisition time was 341ms, relaxation time 2.5 seconds, spin-locking mixing time 50ms, with each increment scanned 16 times. The assignment for the pyrrolidine ring protons is specified for trans (solid lines) and cis (broken lines) isomers with the terminal methyl groups highlighted (arrow).

Table 4.8. Resonance assignment and integration of the 15 individual peaks from the one-dimensional ^1H NMR spectrum for N-acetyl-D-proline shown in figure 4.29.

Peak Number	Proton Resonance Contribution		Relative Integration
	Trans	Cis	
1		$C_\alpha H$	0.27
2	$C_\alpha H$		0.46
3	$C_\delta H_2$		2.00
4		$C_\delta H$	0.94
5		$C_\delta H$	
6	$C_\beta H$	$C_\beta H$	1.48
7	$C_1 H_3$	$C_\beta H$	2.69
8	$C_\beta H$, $C_\gamma H_2$	$C_1 H_3$	3.64
9		$C_\gamma H$	0.95
10		$C_\gamma H$	

4.6 Discussion

The ligand-induced decamerisation of SAP by Ro 63-8695 utilises the high cooperativity of multi-site binding, increasing the affinity constant considerably from 15μM for the head group ((R)-1-acetyl-pyrrolidine-2-carboxylic acid) to 10nM for the whole molecule. Initially SAP would probably require the binding of two or three Ro 63-8695 molecules to induce decamerisation. However, once the partial complex is formed the remaining adjacent binding sites will be optimally positioned and rapidly filled, producing a remarkably stable fully filled complex. In limiting drug concentrations this cooperativity would be expected to lead to preferential decamer formation and in consequence a residual pool of ligand free pentamers. In contrast, in excess drug there may be no decamers as fully ligated pentamers cannot form decamers.

Despite the extensive structural and sequence similarity between human SAP and human CRP, Ro 63-8695 does not induce CRP decamerisation (Pepys 2001). Ro 63-8695 is not capable of compensating for the increased distance between adjacent binding sites generated by the 22° subunit rotation (see section 1.6.3). The separation distance between CRP pentamers packed close enough to enable drug binding would be insufficient to prevent a steric clash between adjacent side chains of Asp70 (C_β <2Å apart) (see figure 4.31). This is primarily due to the subunit rotation orienting the β-hairpin between strands E and F (residues 68-72) away from the body of the protein in CRP. There are also significant differences in the binding of the head group. Tight packing of the pyrrolidine ring is lost due to the much larger hydrophobic pocket of CRP. This reduces the affinity of the head group to above the millimolar range (Kolstoe 2001). However, CRP decamerisation may be achievable by utilising the same high cooperativity as Ro 63-8695,

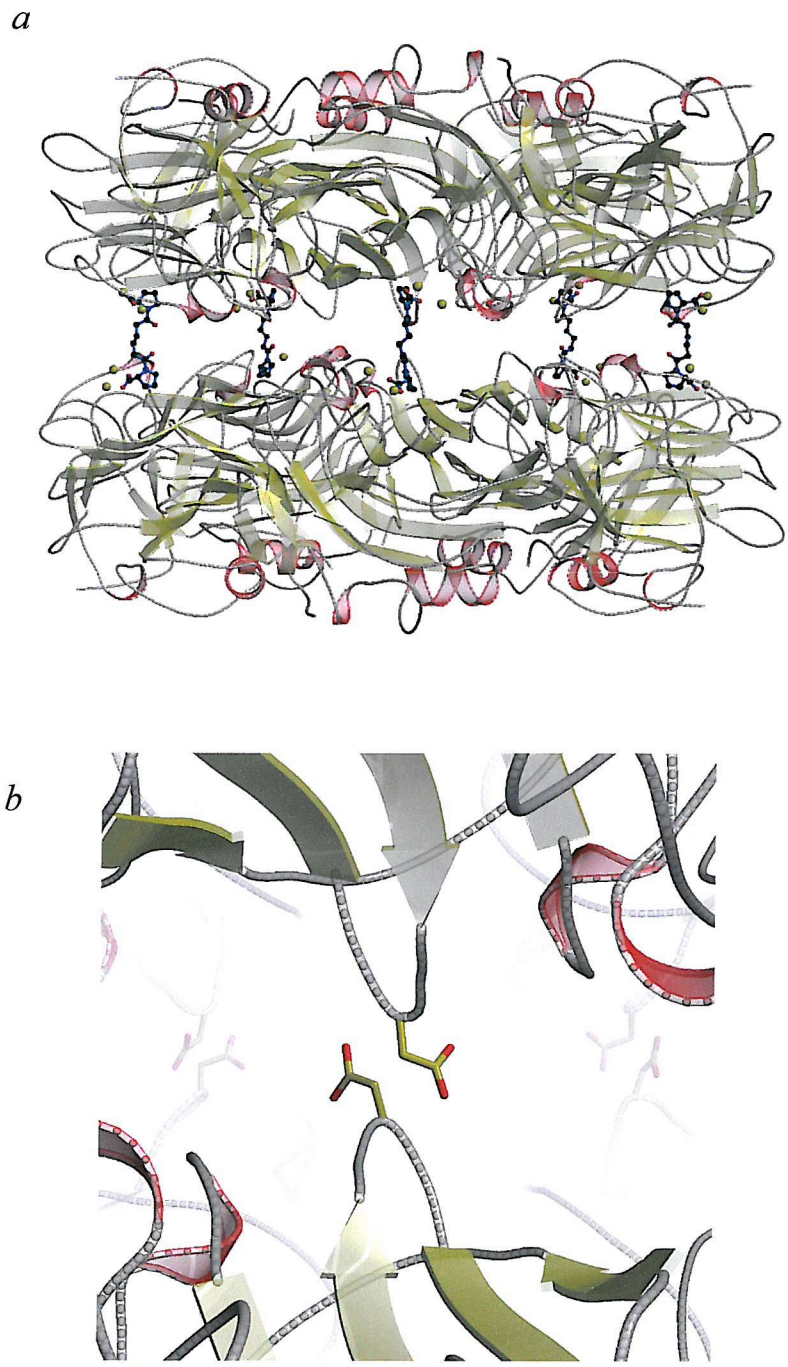


Figure 4.31. Three-dimensional structure of the theoretical CRP-Ro 63-8695 complex (prepared using Bobscript and Raster3D). The Ro 63-8695 molecules are capable of bridging the adjacent pentamers via the terminal carboxyl groups binding to the adjacent calcium ions and insertion of the pyrrolidine ring into the large hydrophobic pocket (*a*). However the separation distance required causes a steric clash between adjacent side chains of Asp70 (*b*). The β-strands are represented as yellow arrows, α-helices as red ribbons, loops as gray cords, calcium ions as yellow spheres, Ro 63-8695 molecules as ball-and-stick models, and Asp70 as stick models.

via replacement of the head group with a higher affinity ligand such as PC and by increasing the separation distance between terminal carboxyl/phosphate groups by 3Å (Thompson 2001). This may also facilitate the design of a novel family of Ro 63-8695 derivatives to facilitate the depletion of other plasma proteins via coupling to SAP.

The 2.8Å structure of SAP in the presence of dAMP (Hohenester *et al* 1997) (Protein Data Bank entry code 1LGN) reveals the formation of a similar decamer to that of SAP-Ro 63-8695, where all of the interactions are mediated through the bound ligands rather than by directly linking the adjacent pentamers together. The decamer is stabilised through base-stacking of adjacent dAMP molecules (see figure 4.32). This creates a slightly wider solvent divide (7.4Å between adjacent Asp138 residues) than that of Ro 63-8695. Stacking of the two-purine bases is relatively weak (estimated free energy of 2-6kcal/mol (Hohenester *et al* 1997)), however cooperative binding through five bound ligands simultaneously is sufficient to create a stable decamer complex.

The binding of Ro 63-8695 to SAP raises an interesting quandary. Only the cis-cis isomer appears to be bound in the decameric complex, although there appears to be no clear preference in the higher resolution structures of bound head groups (presented previously in chapter 3). Even though it is the least stable isomer in free solution, it is feasible that only the cis-cis isomer is capable of decamerisation. Given a population of only 6% in free solution, the cis-cis isomer would have an estimated dissociation constant of ~0.6nM. However, this may be a huge underestimate considering the potential inhibitory effect of non-decamerisable isomers that make up 94% of the population. If the cis-cis isomer were the only form capable of decamerising SAP, then this would limit the rate of decamerisation to the rate of isomerisation which is usually rather slow (hours) unless catalysed by an enzymatic site. However, under the crystallisation conditions for the SAP-Ro 63-8695 complex there are more than enough cis-cis isomers to saturate the protein.

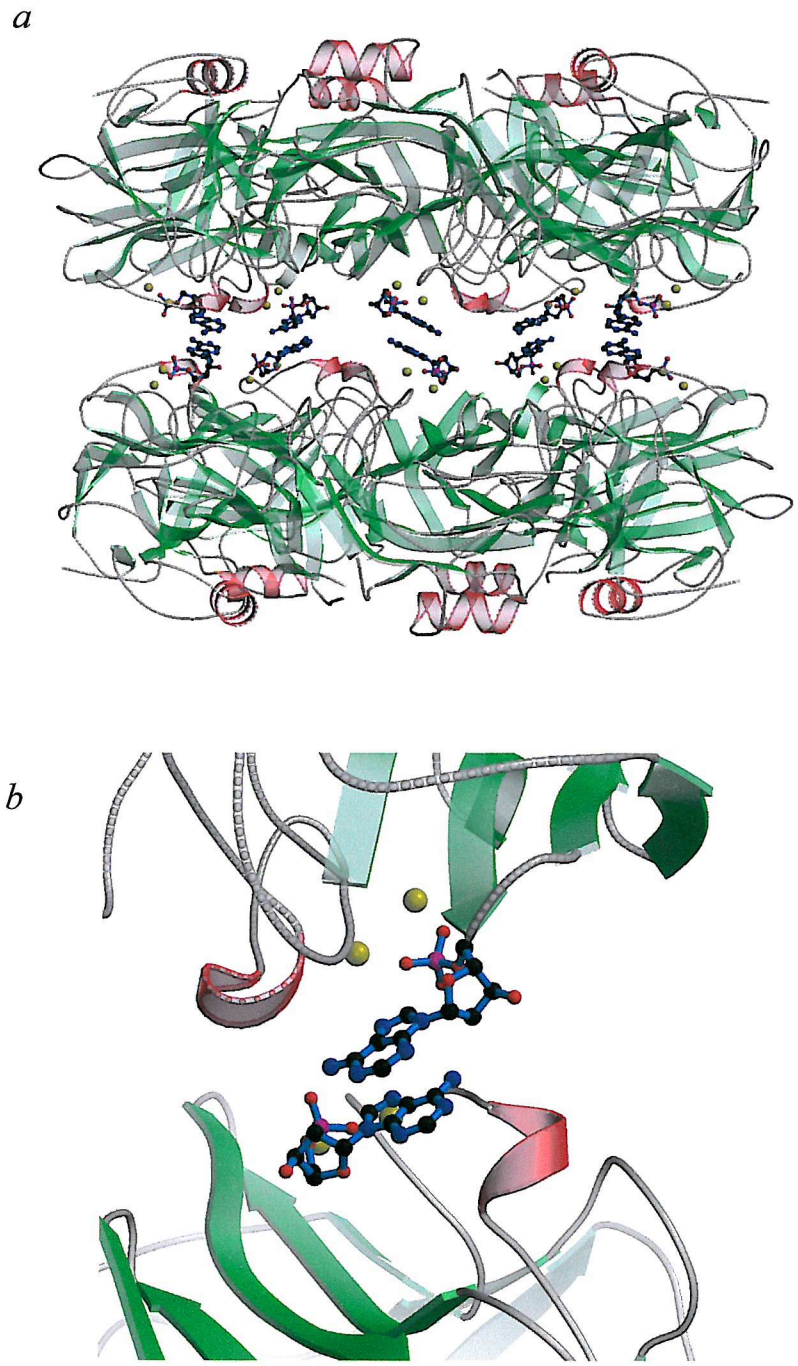


Figure 4.32. Three-dimensional structure of the SAP-dAMP complex illustrating the interactions between adjacent ligands (*a*, close-up shown in *b*) (prepared using Bobscript and Raster3D). The decamer is formed through base-stacking of adjacent dAMP molecules where the only interactions between the adjacent pentamers are mediated through the ligands. Each dAMP molecule is bound through the phosphate group bridging the two calcium ions and the formation of five hydrogen bonds between the protein and the ligand (not shown). The β-strands are represented as green arrows, α-helices as red ribbons, loops as gray cords, calcium ions as yellow spheres, and dAMP molecules as ball-and-stick models.

Interestingly, the binding of the Ro 63-8695 head group to SAP resembles that of a series of proline containing peptides binding to cyclophilin A (Konno *et al* 1996; Zhao & Ke 1996), a well-known proline isomerase. Therefore, SAP might have some cis-trans isomerisation activity possibly by lowering the energy barrier to rotation about the peptide bond via protonation of the pyrrolidine nitrogen by Gln148 (Kofron *et al* 1991), polarizing the carbonyl (Ke *et al* 1993), and desolvation of the amide bonds (Scheiner & Kern 1977). Furthermore, binding the trans isomer may be slightly destabilised via carboxyl-carbonyl repulsion, facilitating the isomerisation to the more stable cis isomer. This may significantly increase the rate of isomerisation of the bound head group, but the peptide bond of the unbound head group would also need to reach the cis conformation to permit decamerisation (see figure 4.33). Therefore decamerisation may be dependent upon both the rate of spontaneous and/or assisted isomerisation. This would essentially make the more abundant trans-trans and trans-cis isomers pro-versions of the cis-cis. Higher-resolution data is required to fully substantiate the presence of only the cis-cis isomer in the decamer complex.

In order to isolate the optimum configuration of Ro 63-8695 a number of compounds have been designed with locked conformations. These are fixed with the least chemical and structural disruption via replacement of the peptide bonds with carbon-carbon double bonds, creating non-interchangeable Z-Z, Z-E, and E-E isomers (see figure 4.34). However, this does result in the reduction of the carbonyl oxygen to a hydroxyl group, pyrrolidine to a cyclopentane ring, and a slight alteration in geometry. The introduction of the carbon-carbon double bonds also facilitates the removal of the carbonyl/hydroxyl groups, allowing the repulsive effects to be studied. These six compounds are currently being prepared for comparative analyses. This rational design process may enable the production of ligands that bind SAP more that 10 times tighter that Ro 63-8695 and allow

lower doses, fewer doses or merely provide a better platform for manipulating the pharmacodynamics by subsequent modification.

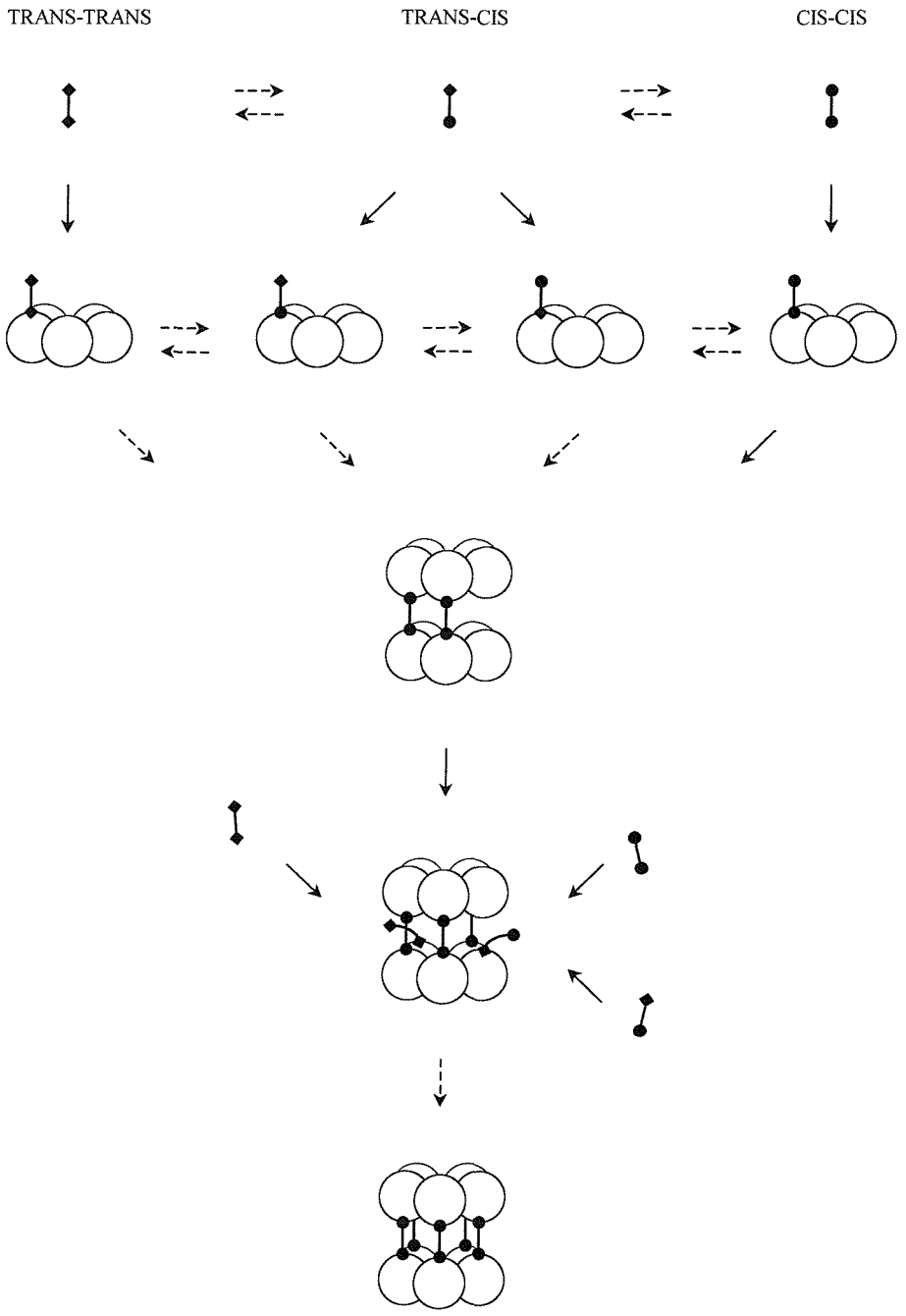
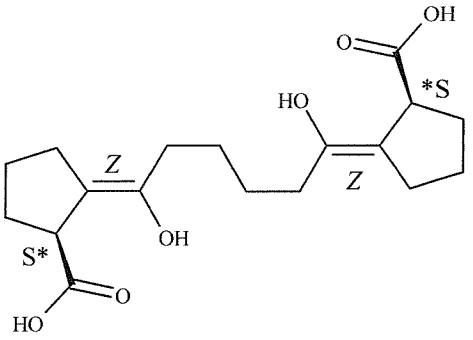
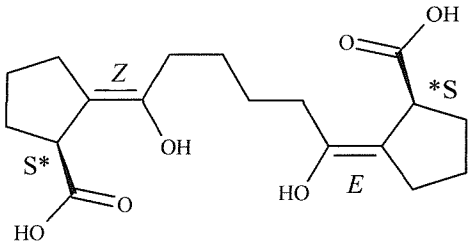


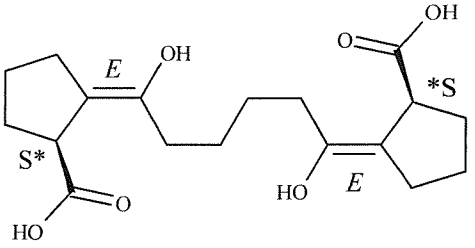
Figure 4.33. Proposed mechanism for the formation of the cis-cis SAP-Ro 63-8695 complex. The interchangeable Ro 63-8695 isomers bind SAP with equal affinity through one head group. Isomerisation to the cis peptide in the bound head group may occur at a fast rate while the unbound at a slow rate. Once a sufficient number of cis-cis isomers are formed or bound, a second pentamer associates. The remaining empty adjacent binding sites are optimally positioned for the cis-cis isomers, however trans-cis and trans-trans isomers can be partially bound through one head group until isomerised to cis-cis. The Ro 63-8695 isomers are represented as dumbbells with diamond (trans) and/or circle (cis) ends for the different head groups, isomerisation is represented as a dashed arrow.



Z-Z Isomer



Z-E Isomer



E-E Isomer

Figure 4.34. Locked geometrical isomers based upon the interchangeable conformations of Ro 63-8695 (isomers labelled according to IUPAC recommendations (Panico *et al* 1993)). The isomers are created by replacing the peptide bonds of Ro 63-8695 with carbon-carbon double bonds. The Z-Z isomer ((1S,2Z)-2-[(Z)-6-[(S)-2-carboxy-cyclopentylidene]-1,6-dihydroxyhexylidene]cyclopentanecarboxylic acid) represents the fixed trans-trans-Ro 63-8695 conformation, the Z-E isomer ((1S,2Z)-2-[(E)-6-[(S)-2-carboxy-cyclopentylidene]-1,6-dihydroxyhexylidene]cyclopentanecarboxylic acid) the fixed trans-cis-Ro 63-8695 conformation, and the E-E isomer ((1S,2E)-2-[(E)-6-[(S)-2-carboxy-cyclopentylidene]-1,6-dihydroxyhexylidene]cyclopentanecarboxylic acid) the cis-cis conformation.

4.7 References

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E. (2000) *Nucleic Acids Res.* **28**, 235-242.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., Warren, G. L. (1998) *Acta Cryst.* **D54**, 905-921.

Burmeister, W. P. (2000) *Acta Cryst.* **D56**, 328-341.

Collaborative Computational Project Number 4 (1994) *Acta Cryst.* **D50**, 760-763.

Crick, F. H. C. & Kendrew, J. C. (1957) *Adv. Protein Chem.* **12,** 134.

Deber, C. M., Bovey, F. A., Carver, J. P., Blout, E. R. (1970) *J. Am. Chem. Soc.* **92**, 6191-6198.

Driessen, H. P., Bax, B., Slingsby, C., lindley, P. F., Mahadevan, D., Moss, D. S., Tickle, I. J. (1991) *Acta Cryst.* **B47**, 987-997.

Emsley, J., White, H. E., O'Hara, B. P., Oliva, G., Srinivasan, N., Tickle, I. J., Blundell, T. L., Pepys, M. B., Wood, S. P. (1994) *Nature* **367**, 338-345.

Engh, R. A. & Huber, R. (1991) *Acta Cryst.* **A47**, 392-400.

Esnouf, R. A. (1997) *J. Mol. Graph. Model.* **15**, 132-134.

Esser, L. (2001) *Personal Communication.*

Hohenester, E., Hutchinson, W. L., Pepys, M. B., Wood, S. P. (1997) *J. Mol. Biol.* **269**, 570-578.

Ke, H., Mayrose, D., Cao, W. (1993) *Proc. Nat. Acad. Sci. (USA)* **90**, 3324-3328.

Kofron, J. L., Kuzmic, P., Kishore, V., Colón-Bonilla, E., Rich, D. H. (1991) *Biochemistry* **30**, 6127-6134.

Kolstoe, S. (2001) *Personal Communication.*

Konno, M., Ito, M., Hayano, T., Takahashi, N. (1996) *J. Mol. Biol.* **256**, 897-908.
26, 283-291.

Leslie, A. G. W. (1997) *Mosflm Users Guide*. MRC Laboratory of Molecular Biology, Cambridge.

Maigret, B., Perahia, D., Pullman, B. (1970) *J. Theor. Biol.* **29**, 275-291.

Matthews, B. W. (1968) *J. Mol. Biol.* **33**, 491-497.

McRee, D. E. (1999) *J. Struct. Biol.* **125**, 156-165.

Merritt, E. A. & Murphy, M. E. P. (1994) *Acta Cryst.* **D50**, 869-873.

Minor, W. (1997) *The HKL Manual*. Edition 5.

Molecular Simulations (1996) *QUANTA96 X-ray Structure Analysis User's Reference*. San Diego.

Navaza, G. (1994) *Acta Cryst.* **A50**, 157-163.

Nyburg, S. C. & Faerman, C. H. (1985) *Acta Cryst.* **B41**, 274-279.
Diego.

Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276**, 307-326.

Panico, R., Powell, W. H., Richer, J. C., eds (1993) *A Guide to IUPAC Nomenclature of Organic Compounds, Recommendations 1993*. Blackwell Scientific Publications.

Pepys, M. B. (2001) *Personal Communication.*

Pye, V (2000) *PhD Thesis University of London.*

Ramachandran, G. N. & Mitra, A. K. (1976) *J. Mol. Biol.* **107**, 85-92.

Ravelli, R. B. G. & McSweeney, S. M. (2000) *Structure* **8**, 315-328.

Scheiner, S. & Kern, C. W. (1977) *J. Am. Chem. Soc.* **99**, 7042-7050.

Thompson, D. (1997) *PhD Thesis University of London.*

Thompson, D. (2000) *Personal Communication.*

Thompson, D. (2001) *Personal Communication.*

Vagin, A. & Teplyakov, A. (1997) *J. Appl. Crys.* **30**, 1022-1025.

Van Holde, K. E., Johnson, W. C., Ho, P. S. (1998) *Principles of Physical Biochemistry.* Prentice-Hall Inc., New Jersey.

Vriend, G. (1990) *J. Mol. Graph.* **8**, 52-56.

Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., Sykes, B. D. (1995) *J. Biomol. NMR* **5**, 67-81.

Zhao, Y. & Ke, H. (1996) *Biochemistry* **35**, 7362-7368.

CHAPTER 5

MOLECULAR CHAPERONE PROPERTIES OF SERUM AMYLOID P COMPONENT

5.1 Introduction

SAP undergoes specific calcium-dependent binding to all types of amyloid fibrils, recognizing and binding to the abnormal protein structure (Pepys *et al* 1997). Binding of SAP stabilises amyloid fibrils *in vitro* (Tennent *et al* 1995), reducing and delaying amyloid deposition in SAP knockout mice (Botto *et al* 1997). The calcium-dependence of SAP binding to the amyloid fibrils (Pepys *et al* 1979) and the fact that MOβDG binding to the double calcium site displaces SAP from amyloid (Hind *et al* 1984), indicate that this is the amyloid recognition site. This capacity of SAP to recognize misfolded proteins of the fibrils resembles the behavior of molecular chaperones and suggests that SAP may affect protein folding.

Molecular chaperones have been shown to bind denatured and partially folded intermediates, stabilising and preventing misfolding and aggregation (Thomas *et al* 1995). An ideal substrate for the study of these is a protein whose refolding mechanism has previously been derived and whose activity is easily and directly measurable. Lactate dehydrogenase (LDH) catalyses the reduction of pyruvate to lactate (see equation 5.1), and is one the most widely used proteins for these studies. To simplify the system, active dimeric LDH from *Bacillus stearothermophilus* is used in preference over tetrameric forms, reducing the number of concentration dependent rate constants (see figure 5.1). The folding and assembly of LDH is also directly measurable via the enzymatic oxidation of NADH at 340nm.

$$\text{Pyruvate} + \text{NADH} + \text{H}^+ \leftrightarrow \text{L-lactate} + \text{NAD}^+ \qquad (5.1)$$

$$2U_1 \xrightarrow[k_1]{} 2U_2 \xrightarrow[k_2]{} 2U_3 \xrightarrow[k_3]{} D$$

Figure 5.1. Refolding scheme of LDH from *B. stearothermophilus*, rate constants are: k_1=3x10^{-3}s^{-1}, k_2=15x10^{-3}s^{-1}, and k_3=1x10^{6}M^{-1}s^{-1} (D represents the active dimer and U_n an unfolded state where n = 1-3) (Badcoe *et al* 1991).

This chapter examines the effects of SAP upon protein folding intermediates of LDH during renaturation and agitation induced inactivation.

5.2 *Methods*

Human SAP and human CRP were purified as described previously (Hawkins *et al* 1991; Vigushin *et al* 1993). SAP was stored at 56mg/ml in 10mM Tris/HCl buffer pH 8.0 containing 140mM NaCl and 10mM EDTA. CRP was stored at 4.4mg/ml in 10mM Tris/HCl buffer pH 8.0 containing 140mM NaCl and 2mM $CaCl_2$. LDH from *B. stearothermophilus* was purchased from Porton Products, or Sigma, UK. SAP was immobilized on cyanogen bromide activated Sepharose as described previously (Pepys *et al* 1982) to provide ~1mg SAP/ml of settled matrix. LDH was denatured by incubation for at least 2 hours at room temperature in 4.2M guanidine hydrochloride in 50mM MOPS/KOH pH 7.0, 10mM β-mercaptoethanol. Refolding was initiated at 25°C by a 200-fold dilution of the denatured LDH to a final protomer concentration of between 40nM and 70nM in the same buffer but without guanidine and β-mercaptoethanol. The effects were tested in the presence during refolding of SAP, CRP, with or without calcium ions, sodium ions, MOβDG, or dAMP. In pilot experiments the refolding cocktail included the LDH assay substrates and activity was monitored continuously. In follow-up experiments LDH

activity was assayed at discrete time intervals by diluting 900µl aliquots of the refolding mixture to a final volume of 1ml containing 10mM pyruvate, 0.2mM β-NADH, and 50mM MOPS/KOH pH 7.0. The absorption was then monitored at 340nm for 60 seconds at 25°C. Isolated human SAP rapidly autoaggregates in the presence of calcium ions and this was monitored by light scattering at an excitation and emission wavelength of 500nm. The concentration of residual soluble SAP was determined by absorbance at 280nm (*E (1%, 1cm)* = 17.1).

To induce inactivation by agitation, 500µl of 50nM LDH in MOPS/KOH buffer pH 7.0, 150mM NaCl, was rotated on a blood tube mixer at 20 rpm (revolutions per minute) in a 1.5ml plastic microfuge tube. Aliquots (10µl) were removed at intervals for examination of the remaining LDH activity as described above. Inactivation was also achieved by continuous stirring at ~5 rps (revolutions per second) with 6ml of 16nM LDH in MOPS/KOH buffer pH 7.0, 140mM NaCl, in a small siliconised glass vessel, where 250µl aliquots were removed at intervals for assaying. SAP, or CRP, with or without calcium ions and MOβDG, was included in the sample at the start of mixing.

5.3 Results

LDH from *B. stearothermophilus* was denatured by 4.2M guanidine hydrochloride to a state showing no secondary structure (Smith *et al* 1991). Only 3-6% of active dimeric enzyme was slowly regained by diluting out the denaturant to a residual concentration (0.02M guanidine) insufficient to destabilize the enzyme (Smith *et al* 1991). The recovery rose to about 8% if β-mercaptoethanol was included in the refolding step, but was usually omitted to avoid reduction of the single disulphide bridge within each SAP protomer. Addition of SAP to the refolding solution markedly enhanced the yield of active LDH,

irrespective of β-mercaptoethanol. A typical experiment, shown in figure 5.2, included 400nM SAP (with respect to pentamer), 3mM calcium ions, and 40nM LDH. In the same conditions, inclusion of 2mM MOβDG, which blocks the calcium-dependent ligand-binding site of SAP and inhibits the calcium-dependent aggregation of SAP, did not reduce the SAP induced enhancement of LDH refolding as expected, but in fact significantly increased the yield when added prior to SAP aggregation.

No change in the refolding yield was observed when SAP was replaced with CRP at 0.4μM or 2.0μM with respect to pentamers, representing a 10-40 fold excess over LDH subunits (figure 5.2). This powerful control demonstrates the highly specific nature of SAP enhanced refolding of LDH, since CRP and SAP share high sequence and structural identity, both binding calcium ions in a similar way (see sections 1.2, 1.5, and 1.6). Although in contrast to human SAP, CRP does not autoaggregate in the presence of calcium (see section 1.6.4) (Baltz *et al* 1982).

Increasing the concentration of SAP to 2μM increased the refolding yield but this did not exceed the value of ~25% shown with MOβDG in figure 5.2. The effective concentration and form of SAP in these experiments was difficult to quantify due to the severe calcium-dependent aggregation of SAP. Light scattering measurements showed that even at SAP concentrations as low as 100nM some aggregation occurred immediately after mixing with 3mM calcium ions (figure 5.3A). At 2μM SAP only an estimated 10% remained in solution. These results are likely to change with time after mixing and suggest that the protein does not exhibit a simple solubility limit.

The calcium-dependent ligand dAMP is also able to prevent SAP aggregation, but when at appropriate molar excess induces the formation of a decamer complex through stacking of adjacent B-faces (see figure 4.32) (Hohenester *et al* 1997). Under these conditions with 4μM SAP, 5mM dAMP, and 7mM calcium in the refolding mixture

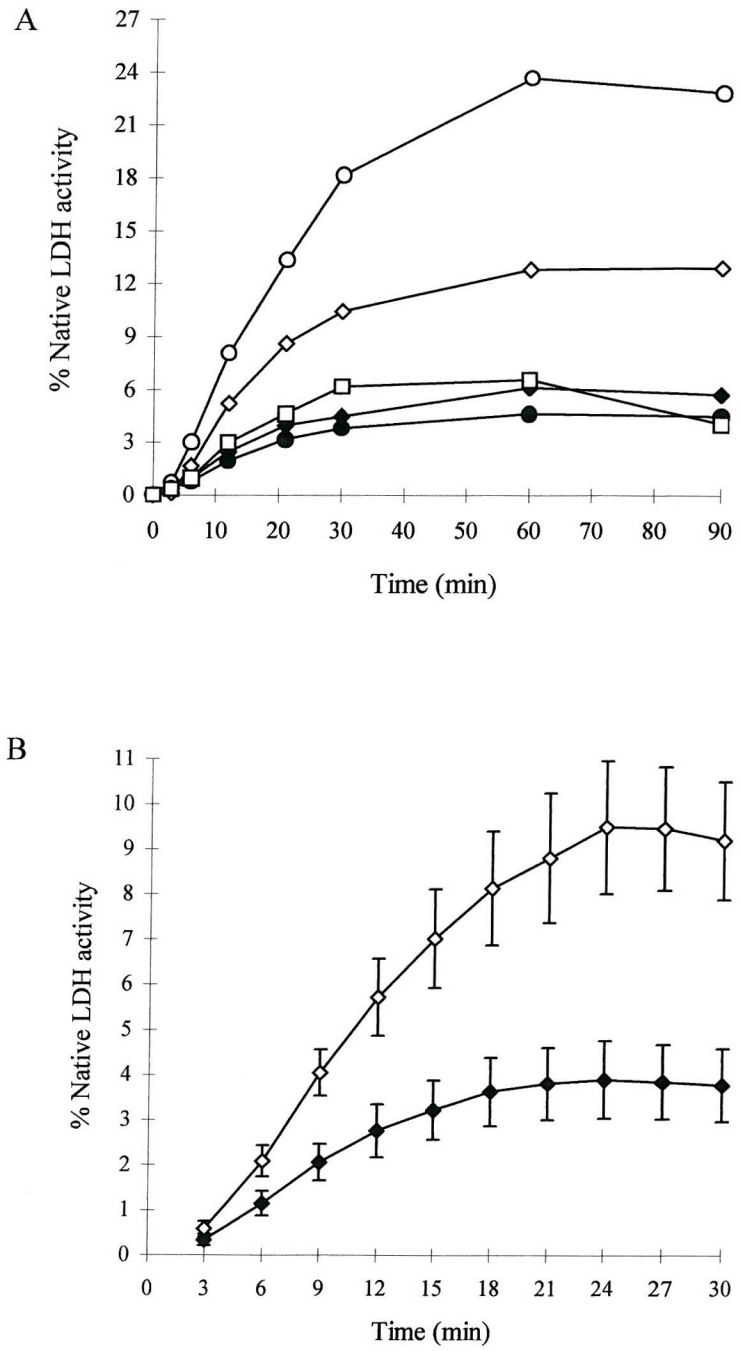
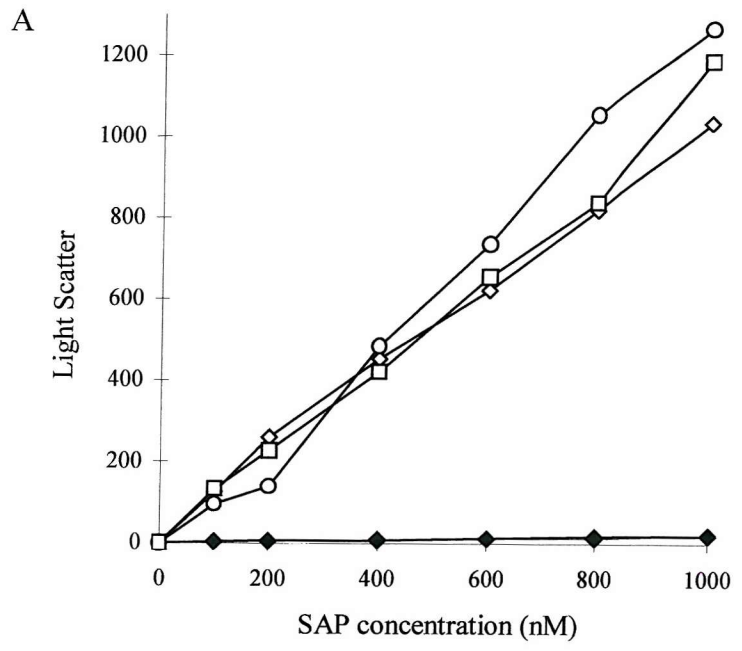


Figure 5.2. Reactivation of chemically denatured LDH (40nM) in the absence (♦) or presence (◊) of 400nM SAP, 400nM CRP (□), 2mM MOβDG (●), or SAP and MOβDG (○) in 50 mM MOPS-KOH pH 7.0 buffer containing (A) 3mM $CaCl_2$ or (B) 150mM NaCl and 7mM $CaCl_2$ (n=5).



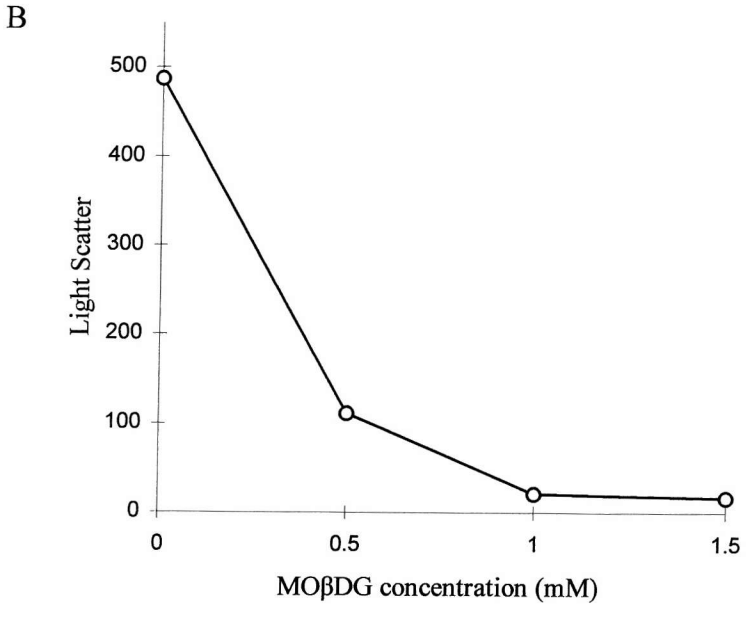


Figure 5.3. Concentration dependence of SAP aggregation induced by calcium ions (A) and inhibited by MOβDG (B) measured by light scattering at an excitation and emission wavelength of 500nm, at room temperature in 50mM MOPS-KOH pH 7.0 with 2mM EDTA (♦), 3mM $CaCl_2$ (◊), 7mM $CaCl_2$ (□), or 3mM $CaCl_2$ and 140mM NaCl (○).

together with LDH and substrates (figure 5.4), there was no difference in enhancement of refolding compared to control mixtures without the ligand. This was also the case when the activity with dAMP was assayed separately after refolding (see table 5.1). The enhancement of LDH refolding was also observed in the absence of calcium ions (figure 5.5) where SAP is highly soluble, but even at SAP concentrations as high as 4.5μM the maximal enhancement of ~25% was never exceeded. Furthermore, aggregation can be prevented by covalently attaching SAP to Sepharose beads without affecting its ability to increase the yield of refolding LDH in the presence of calcium ions.

The calcium-induced aggregation of SAP is apparently due to the binding of protruding carboxylate ligands on the A-face by the double calcium-binding sites on the B-face, leading to A-B face stacking of the pentamers (see section 1.5.4). The MOβDG concentration used here is sufficient to completely inhibit aggregation (see figure 5.3B), probably by occupying these sites and maintaining SAP in solution as free individual pentamers (Hutchinson *et al* 1999). Presumably the enhancement of SAP induced refolding yield produced by the addition of MOβDG (figure 5.2A) results from increased solubility of SAP, making more of SAP readily available for interaction with LDH folding intermediates. It is conceivable that an inhibitory action of MOβDG could be masked by its capacity to increase the concentration of soluble SAP in the presence of calcium. If the calcium sites in SAP do contribute to the refolding enhancement then given the K_d for the SAP-MOβDG interaction of 50μM (Kolstoe 2001), in a mixture of 400nM SAP/3mM calcium/2mM MOβDG (figure 5.2), only 2.5% of the calcium sites in SAP (10nM) would be available and thus attributable to the effect on LDH refolding. It is highly unlikely that the calcium sites would have such potency, as the 200nM soluble SAP that remains on mixing 2μM SAP with 3mM calcium is less effective in enhancing LDH refolding. Furthermore, in the absence of calcium SAP enhances the refolding yield of LDH to a

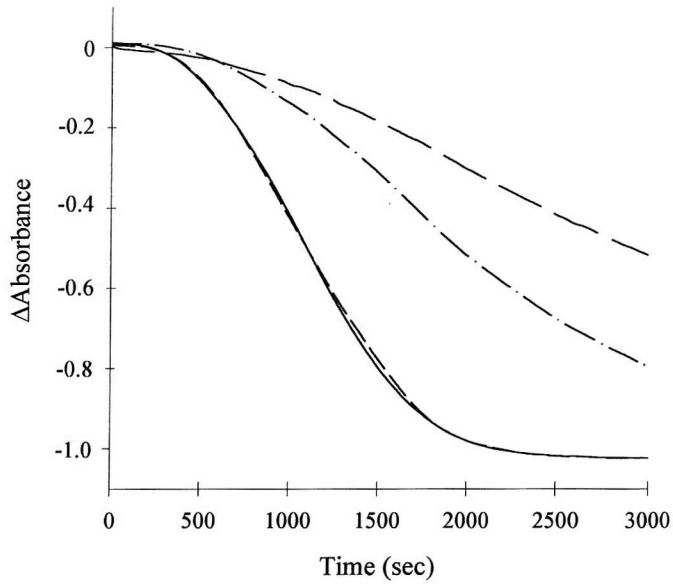


Figure 5.4. Reactivation of chemically denatured LDH (70nM) in the absence (dashed line, long dashes) or presence (solid line) of 4μM SAP, 5mM dAMP (dot dashed line), or SAP and dAMP (dashed line, short dashes) in 50mM MOPS-KOH pH 7.0, 150mM NaCl, 7mM $CaCl_2$ buffer. Activity regain was monitored continuously by absorbance change at 340nm at room temperature.

Table 5.1. Effect of dAMP on reactivation of LDH after guanidine dilution in the presence of SAP. LDH (50nM) was refolded from guanidine in the presence of 3mM $CaCl_2$, 150mM NaCl, 500nM SAP and 5mM dAMP and restored activity determined (n=3).

Reagents present during refolding	% initial LDH activity Mean (SD)
LDH alone	2.28 (0.78)
LDH + SAP	8.61 (0.45)
LDH + SAP + dAMP	9.33 (0.21)
LDH + dAMP	2.24 (0.56)

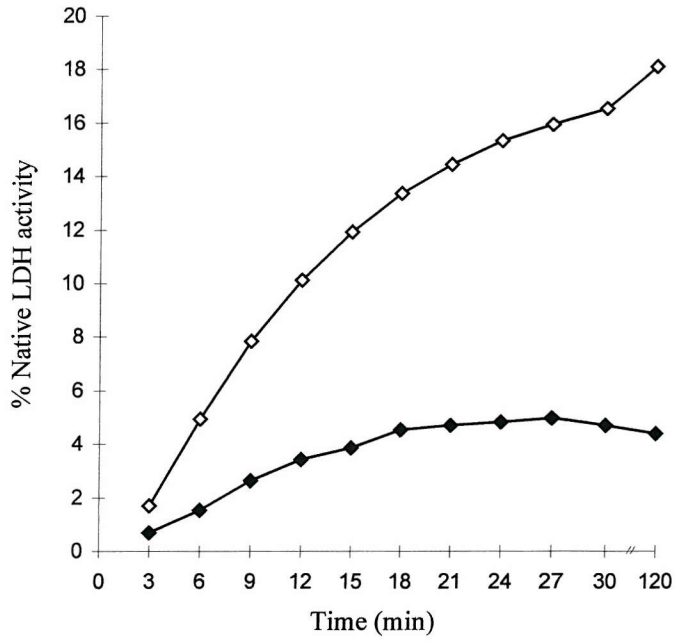


Figure 5.5. Reactivation of chemically denatured LDH (40nM) in the presence (◊) or absence (♦) of 400nM SAP in 50mM MOPS-KOH pH 7.0.

similar extent to calcified SAP at the same concentration in the presence of MOβDG. The effect of MOβDG on SAP solubility therefore seems to provide a good explanation for the apparent increase in activity.

Gentle agitation of LDH at 50nM in the presence of 3mM calcium ions for 30 minutes using the blood tube mixer eliminated about 75% of the enzyme activity (see table 5.2). It is likely that physical forces such as shear at the air/liquid or liquid/container interface, foaming or cavitations are responsible for this loss of activity. Inclusion of β-mercaptoethanol was unable to protect the LDH activity under these conditions, suggesting that cysteine oxidation is not the cause. However, in the presence of ~4μM immobilized SAP or 500nM free SAP (with respect to pentamer) only 25% of the LDH activity was lost upon agitation irrespective of the presence or absence of 2mM MOβDG. Similar but less

Table 5.2. Effect of immobilized SAP on loss of LDH activity during mixing (20 rpm). LDH (50nM) was mixed for 30min in the presence of 3mM $CaCl_2$, 150mM NaCl, ~4μM SAP immobilized covalently on Sepharose and 2mM MOβDG, then assayed for residual activity.

Reagents present during refolding	Residual LDH activity (%)
LDH without mixing	96.8
LDH alone	26.0
LDH + Sepharose-SAP	74.9
LDH +Sepharose-SAP + MOβDG	65.4
LDH + MOβDG	26.9

Table 5.3. Effect of SAP on loss of LDH activity during gentle stirring (5 rps). LDH (16nM) was mixed for 60min in the presence of 3mM $CaCl_2$, 140mM NaCl, 400nM SAP, and 2mM MOβDG, then assayed for residual activity.

Reagents present during refolding	Residual LDH activity (%)
LDH alone	85
LDH + SAP	100
LDH + SAP + MOβDG	100
LDH + MOβDG	79

dramatic inactivation also took place in gently stirred solutions, but with full activity maintained over 60 minutes in the presence of SAP, again irrespective of MOβDG (see table 5.3).

5.4 Discussion

Molecular chaperones function by binding to aggregation-prone folding intermediates. Presumably SAP enhances the efficiency of LDH refolding by similarly interacting with intermediates on the refolding landscape to increase the flux through productive routes (Dill & Chan 1997). The effect of SAP is independent of ATPase activity and/or the binding of co-chaperones. This passive effect is also saturable, with a maximum of 25% activity regained under the present conditions. This may reflect the differential rates of partitioning of intermediates between states that interact with SAP and others. A small proportion of LDH clearly proceeds directly to the native state but the majority goes to an inactive misfolded product. SAP might bind to a misfolded state and enhance LDH folding by removing species capable of forming an unproductive dimer or act in a more direct way by enabling the formation of productive species. The parallel drawn between recognition of the protein misfold of amyloid and protein folding intermediates would favor models in which SAP interacts with misfolding species. However, the site on SAP most likely to be involved in amyloid recognition can be blocked by MOβDG without any inhibitory effect upon the folding enhancement of LDH. The site can also be blocked by formation of B-B face decamers by interaction with dAMP without affecting the LDH refolding phenomenon. Furthermore, the refolding reaction can proceed in the absence of calcium, which is absolutely required for amyloid binding. This implies that the site is not involved in the folding enhancement process, and presumably that the A-face and/or

exterior 'edge' regions of the SAP pentamer are involved. Both of which show significant sequence differences with human CRP, which shows no such refolding enhancement (see figure 1.1).

It has become clear that even proteins unrelated to *in vivo* amyloidogenesis can be induced by suitable denaturation protocols *in vitro* to adopt the amyloid cross-β fold and to aggregate as typical amyloid fibrils (Canet *et al* 1999; Chiti *et al* 1999; Jiménez *et al* 1999; Litvinovich *et al* 1998). This highlights the mysterious fact that only 17 to 18 diverse and unrelated proteins are known to form amyloid fibrils *in vivo* (Tan & Pepys 1994). Here for LDH, only gentle agitation is sufficient to initiate unfolding reactions that must be a general feature in the production of amyloidogenic precursors. Although LDH is not itself an amyloidogenic protein, the observations are reminiscent of results obtained with Aβ1-42, the major amyloid fibril protein in Alzheimer's disease (Janciauskiene *et al* 1995). This spontaneously forms amyloid fibers in physiological, non-denaturing, solvent conditions *in vitro*. SAP binds the mature fibers in a calcium-dependent reaction but SAP is also reported to retard fiber formation in a calcium independent reaction (Janciauskiene *et al* 1995). Since binding of SAP stabilizes amyloid fibrils *in vitro* (Tennent *et al* 1995) and participates in pathogenesis of amyloidosis *in vivo* (Botto *et al* 1997), it is possible that misfolding amyloidogenic precursor proteins displaying the motif recognized by the calcium-dependent binding site of SAP may contribute to their clinical amyloidogenicity. In contrast, other proteins, not recognized by SAP in this way, may undergo productive folding interactions with SAP, or at least not be trapped or enabled by calcium-dependent SAP binding to persist in the pathological conformation that causes amyloidosis.

An alternative scenario is suggested by the protection conferred by SAP against agitation-induced denaturation of LDH. Here there is gradual production of misfolded species and mere binding of SAP to these would not be expected to prevent overall loss of

enzyme activity unless such loss depended on a nucleated aggregation process, as occurs in experimental amyloid fibrillogenesis *in vitro* (see section 1.4.3) (Canet *et al* 1999). SAP may therefore have a surveillance role *in vivo*, binding to misfolded species and preventing the seeding of larger aggregates. The primary driving force towards clinical amyloidosis would then be the inherent instability that is now recognized as a common feature of amyloidogenic precursor proteins (Booth *et al* 1997; Kelly 1996). This may first overwhelm the normal productive folding role of SAP, and subsequent calcium-dependent binding of SAP to mature fibrils could then stabilize and protect them from degradation (Tennent *et al* 1995). Whatever the underlying mechanism these observations support the therapy for amyloidosis with Ro 63-8695 currently being evaluated.

5.5 *References*

Badcoe, I. G., Smith, C. J., Wood, S., Halsall, D. J., Holbrook, J. J., Lund, P., Clarke, A. R. (1991) *Biochemistry* **30**, 9195-9200.

Baltz, M. L., de Beer, F. C., Feinstein, A., Pepys, M. B. (1982) *Biochim. Biophys. Acta* **701**, 229-236.

Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E., Hawkins, P. N., Dobson, C. M., Radford, S. E., Blake, C. C., Pepys, M. B. (1997) *Nature* **385**, 787-793.

Botto, M., Hawkins, P. N., Bickerstaff, M. C. M., Herbert, J., Bygrave, A. E., McBride, A., Hutchinson, W. L., Tennent, G. A., Walport, M. J., Pepys, M. B. (1997) *Nat. Med.* **3**, 855-859.

Canet, D., Sunde, M., Last, A. M., Miranker, A., Spencer, A., Robinson, C. V., Dobson, C. M. (1999) *Biochemistry* **38**, 6419-6427.

Chiti, F., Webster, P., Taddei, N., Clark, A., Stepani, M., Ramponi, G., Dobson, C. M. (1999) *Proc. Nat. Acad. Sci. (USA)* **96**, 3590-3594.

Dill, K. A. & Chan, H. S. (1997) *Nat. Struct. Biol.* **4**, 10-19.

Hawkins, P. N., Tennent, G. A., Woo, P., Pepys, M. B. (1991) *Clin. Exp. Immunol.* **84**, 308-316.

Hind, C. R. K., Collins, P. M., Caspi, D., Baltz, M. L., Pepys, M. B. (1984) *Lancet* **2**, 376-378.

Hohenester, E., Hutchinson, W. L., Pepys, M. B., Wood, S. P. (1997) *J. Mol. Biol.* **269**, 570-578.

Hutchinson, W. L., Hohenester, E., Perkins, S. J., Herbert, J., Pepys, M. B. (1999) in *Amyloid and Amyloidosis 1998*. Kyle, R. A., Gertz, M. A. (eds) Parthenon Publishing, Pearl River, New York, 32-34.

Janciauskiene, S., García de Frutos, P., Carlemalm, E., Dahlbäck, B., Eriksson, S. (1995) *J. Biol. Chem.* **270**, 26041-26044.

Jiménez, J. L., Guijarro, J. I., Orlova, E., Zurdo, J., Dobson, C. M., Sunde, M., Saibil, H. R. (1999) *EMBO J.* **18,** 815-821.

Kelly, J. W. (1996) *Curr. Opin. Struct. Biol.* **6**, 11-17.

Kolstoe, S. (2001) *Personal Communication.*

Litvinovich, S. V., Brew, S. A., Aota, S., Akiyama, S. K., Haudenschild, C., Ingham, K. C. (1998) *J. Mol. Biol.* **280**, 245-258.

Pepys, M. B., Booth, D. R., Hutchinson, W. L., Gallimore, J. R., Collins, P. M., Hohenester, E. (1997) *Int. J. Exp. Clin. Invest.* **4**, 274-295.

Pepys, M. B., de Beer, F. C., Dyck, R. F., Hind, C., Lantain, I., Fagan, E. A., Maton, P. N., Starke, I., Fox, K., Allan, R., Hodgson, H., Chadwick, V. S., Hughes, G. R. V., Goldinam, J., Catolvsky, D., Coalton, D., Krikler, D., Masen, A., Mallya, R. K., Demy, H., Hamilton, E. D. B., Mace, D. E. N. (1982) *Ann. N. Y. Acad. Sci.* **389**, 459-460.

Pepys, M. B., Dyck, R. F., de Beer, F. C., Skinner, M., Cohen, A. S. (1979) *Clin. Exp. Immunol.* **38**, 284-293.

Smith, C. J., Clarke, A. R., Chia, W. N., Irons, L. I., Atkinson, T., Holbrook, J. J. (1991) *Biochemistry* **30**, 1028-1036.

Tan, S. Y. & Pepys, M. B. (1994) *Histopathology* **25,** 403-414.

Tennent, G. A., Lovat, L. B., Pepys, M. B. (1995) *Proc. Nat. Acad. Sci. (USA)* **92**, 4299-4303.

Thomas, P. J., Qu, B. H., Pedersen, P. L. (1995) *Trends Biochem. Sci.* **20**, 456-459.

Vigushin, D. M., Pepys, M. B., Hawkins, P. N. (1993) *J. Clin. Invest.* **91**, 1351-1357.

APPENDIX

β-Bulges

A β-bulge is a region of irregularity in a β-sheet where the hydrogen-bonding pattern is disrupted. This usually involves two or more residues in the bulged strand (labelled numerically from the N-terminal region) opposite to a single residue in the neighbouring strand (residue X). The bulges are classified as parallel or antiparallel depending upon the orientation of the β-strands involved. Each category is further subdivided into classic, wide, G1, bent, and special types depending upon the hydrogen bonding pattern and the number of residues involved. In antiparallel β-sheets the classic bulges occur when an extra residue is located between the narrowly spaced hydrogen bonds, whereas the wide bulges occur when the residue is located between the widely spaced hydrogen bonds. In parallel β-sheets the classic bulges occur when residue 2 forms a hydrogen bond with residue X, whereas the wide bulges occur when no hydrogen bonds are formed between residues of the bulge. G1 β-bulges only occur in antiparallel strands where residue 1 is in a left-handed α-helical configuration (normally a glycine); these are usually located at the terminal ends of the β-strands. Bent bulges occur when an extra residue is located upon both β-strands. Special β-bulges involve a larger insertion of up to three residues in one strand (Chan *et al* 1993; Richardson *et al* 1978).

β-Hairpins

A β-hairpin is defined for the loop region adjoining two antiparallel β-strands that are hydrogen bonded together by at least one bridge. The hairpins are classified according two numbers X:Y, which designate the number of residues within the loop via different IUPAC strand conventions (IUPAC 1970). This is indicative of the hydrogen bonding and dihedral angles for the β-strands, where X is the number of residues excluded by the

hydrogen-bonding pattern (IUPAC convention Rule 6.3), and Y the number excluded by the dihedral angles (IUPAC convention Rule 6.2) or not involved in the formation of two hydrogen bonds (Milner-White & Poet 1986). Hence, if a double hydrogen bond is formed at the terminal residues then X=Y, but if only a single hydrogen bond is formed then Y=X+2 (Sibanda 1986; Sibanda *et al* 1989).

β-Sheets

β-sheets are comprised of a number of extended lengths (β-strands) of a polypeptide chain running alongside each other. The extended conformation (phi and psi dihedral angles centred at -120° and 140° respectively) of the polypeptide chain positions the adjacent residues approximately 3.5Å apart (axial distance). The β-strands are aligned either in the same (parallel) or opposite direction (antiparallel), stabilised by hydrogen bonding through the main chain carbonyl and imino groups of adjacent strands. Antiparallel β-sheets have alternating narrowly spaced and widely spaced hydrogen bonding pairs, while parallel β-sheets have equally spaced hydrogen bonds.

β-Turns

A β-turn is defined for four consecutive residues (donated *i*, *i* +1, *i* +2, and *i* +3 from N- to C-terminus) when the distance between C_α atoms of residues *i* and *i* +3 are within 7Å and the central two residues (*i* +1 and *i* +2) are in a non-helical conformation. The β-turns are assigned to one of 14 different classes according to the phi and psi angles of residues *i*+1 and *i*+2. The ideal dihedral angles for each β-turn type are shown in table 6.1. The phi and psi angles can vary by up to 30° from the ideal values with one allowed the added flexibility of up to 40°. The type VI β-turns are subject to an additional condition whereby

residue *i*+2 is a cis-proline. Turns that do not fit the dihedral criteria are classified as type IV (Chou & Fasman 1977; Hutchinson & Thornton 1994; Wilmot & Thornton 1988).

Table 6.1. Assignment of β-turns classes according to the phi and psi angles of residues *i*+1 and *i*+2.

Type	Phi *i*+1 (°)	Psi *i*+1 (°)	Phi *i*+2 (°)	Psi *i*+2 (°)
I	-60	-30	-90	0
I′	60	30	90	0
II	-60	120	80	0
II′	60	-120	-80	0
III	-60	-30	-60	-30
III′	60	30	60	30
IV	Turns excluded from all other categories			
V	-80	80	80	-80
V′	80	-80	-80	80
VIa1 *	-60	120	-90	0
VIa2 *	-120	120	-60	0
VIb *	-135	135	-75	160
VII	-	180	<60	-
	-	<60	180	-
VIII	-60	-30	-120	120

* Cis-proline at the *i*+2 position

γ-Turns

A γ-turn is defined for three consecutive residues (donated *i*, *i* +1, and *i* +2 from N- to C-terminus) when a hydrogen bond exists between residues *i* and *i*+2. The γ-turns are classified as classic or inverse depending upon the phi and psi angles of residue *i*+1. The ideal dihedral angles for the 2 classes are shown in the table below, and are allowed to vary by up to 40° (Milner-White *et al* 1988; Rose *et al* 1985).

Table 6.2. Assignment of γ-turns classes according to the phi and psi angles of residue *i*+1.

Type	Phi *i*+1 (°)	Psi *i*+1 (°)
Classic	75	-64
Inverse	-79	69

Helices

Helices are structures in which a number of residues rotate and rise in a repeating manor along an axis. Each residue is orientated with their carbonyl groups pointing towards the C-terminal and the imino groups towards the N-terminal end. This enables the helix to be stabilised by hydrogen bonding throughout its length, creating a dipole moment aligned with the helical axis. Due to the orientation of the main chain atoms, all of the C_β carbons are directed towards the N-terminus of the helix. The hydrogen-bonding pattern, number of residues per turn, and the direction of the helices define the type of helical structure. Helices can be right-handed or left-handed, although left-handed conformations

only occur in very short sequences (phi and psi dihedral angles centred at 50° and 60° respectively). There are two type of right-handed helices the α-helix and 3_{10}-helix. All large helices (over 6-7 residues) are of the α-helical type, where consecutive residues have phi and psi angles centred at -60° and -50° respectively. This creates a rise of 1.5Å per residue, with 3.6 residues per turn. Hydrogen bonding is formed through the carbonyl of residue *n* with the amide nitrogen of residue *n*+4 creating a hydrogen-bonded ring of 13 atoms. Consecutive residues of the 3_{10} helices have phi and psi angles centred at -70° and -5° respectively. This gives a rise of 2Å per residue, with 3 residues per turn. These have only 10 atoms in the hydrogen-bonded ring formed through the carbonyl of residue *n* with the imino nitrogen of residue *n*+3.

References

Chan, A. W. E., Hutchinson, E. G., Harris, D., Thornton, J. M. (1993) *Protein Sci.* **2**, 1574-1590.

Chou, P. Y. & Fasman, G. D. (1977) *J. Mol. Biol.* **115**, 135-175.

Hutchinson, E. G. & Thornton, J. M. (1994) *Protein Sci.* **3**, 2207-2216.

IUPAC (1970) *J. Mol. Biol.* **52**, 1-17.

Milner-White, E. J. & Poet, R. (1986) *Biochem. J.* **240**, 289-292.

Milner-White, E. J., Ross, B. M., Ismail, R., Belhadj-Mastefa, K., Poet, R. (1988) *J. Mol. Biol.* **204**, 777-782.

Richardson, J. S., Getzoff, E. D., Richardson, D. C. (1978) *Prot. Nat. Acad. Sci. (USA)* **75**, 2574-2578.

Rose, G. D., Gierasch, L. M., Smith, J. A. (1985) *Adv. Protein Chem.* **37**, 1-109.

Sibanda, B. L. (1986) *PhD Thesis University of London.*

Sibanda, B. L., Blundell, T. L., Thornton, J. M. (1989) *J. Mol. Biol.* **206**, 759-777.

Wilmot, C. M. & Thornton, J. M. (1988) *J. Mol. Biol.* **203**, 221-232.