Crystal Structure of the Large Type III Glutamine Synthetase from *Bacteroides fragilis*

Jason M. van Rooyen

Thesis Presented for the Degree of
DOCTOR OF PHILOSOPHY
in the Department of Molecular and Cell Biology
UNIVERSITY OF CAPE TOWN

Supervisors:

Associate Professor Trevor Sewell
Associate Professor Val Abratt

March 2010

Abstract

"Crystal Structure of the Large Type III Glutamine Synthetase from *Bacteroides fragilis*"

Jason M. van Rooyen

PhD Thesis, Department of Molecular and Cell Biology, University of Cape Town

March 2010

Glutamine synthetases are one of the most ancient functioning enzymes in existence and these large oligomeric complexes are found in all extant forms of life where they play a critical role in nitrogen metabolism. Over the past five decades, extensive biochemical studies together with structural investigations have helped build a picture of the mechanism of functioning and regulation in the GSI and GSII families. The most divergent GSIII family, however, is poorly characterized and has only recently been recognized. Structural studies, using both cryo-EM and X-ray crystallography, were undertaken on the type III GS, GlnN, from the opportunistic human pathogen, *Bacteroides fragilis*, with a view to better understanding the GSIII family in the light of the known structure function-relationships of the other GS enzymes, and to investigate the potential for the design of selective inhibitors against the divergent family. A low-resolution (16 Å) reconstruction of GlnN was first determined by single particle cryo-EM and image processing. This structure revealed that GlnN was a double-ringed dodecamer with D6 symmetry and the arrangement of active sites within the hexameric rings closely matched the GSI structure. Following the design of a rapid purification protocol and improvements to the stability and solubility of GlnN, conditions were discovered for the production of diffraction quality inhibitor-bound crystals. A second better diffracting crystal form was also produced following proteolytic processing. The crystal structure of GlnN was solved to near atomic resolution (3.0 Å) following phase extension of low-resolution SAD phases, taking into account the cryo-EM structure. The higher resolution of the crystal structure revealed that, surprisingly, the orientation of the hexameric rings in GlnN is inverted in comparison to other families. These results have raised interesting questions surrounding the mechanism and driving forces responsible for the evolution of quaternary structure in the GS enzymes and have suggested that the GSI and GSII structure arose following truncation of a large GSIII-like ancestor. Despite the differences in higher order assembly, the GlnN monomer displayed a high degree of similarity with the GSI and GSII structures in the core active site region, thus, suggesting a conservation of reaction mechanism. Structure-based multiple sequence alignment showed that the residues forming the nucleotide binding pocket are the least conserved in the GS superfamily, and several residue positions, which represent altered modes of ligand binding, were suggested as potential avenues for the design of selective inhibitors against GlnN.

Acknowledgements

I would like to thank my supervisors Assoc. Prof. Trevor Sewell and Prof. Val Abratt for their expert guidance, support, and encouragement. I am grateful to the staff of the EMU for their helpful advice and assistance and to my fellow students for their inspiration and support.

I am indebted to my family for their patience and encouragement. To Megan, thank you for your unquestioning and unwavering support of my work.

Finally, I gratefully acknowledge the financial assistance of the National Research Foundation, the Cape Biotechnology Trust, the Harry Crossley Foundation, K.W. Johnstone, and University of Cape Town.

List of Abbreviations

AF	affinity chromatography	**HA**	heavy atom
AGB	Affi-Gel Blue	**HIC**	hydrophobic interaction chromatography
Ap	ampicillin	**IEX**	ion exchange
ASU	asymmetric unit	**LB**	Luria Broth
ATP	adenosine triphosphate	**MALDI-TOF**	matrix-assisted laser desorption/ionization time-of-flight
BSA	bovine serum albumin	**MetSox**	methionine sulfoximine
C6	6-fold cyclical space group symmetry	**MPD**	(+/-)-2-Methyl-2,4-pentanediol
CC	cross-correlation	**MW**	molecular weight
CCD	charge coupled device	**MWCO**	molecular weight cut-off
CFE	cell free extract	**PDB**	protein databank
CMP	cytidine monophosphate	**PEG**	polyethylene glycol
CS	culture supernatant	**PEI**	polyethylenimine
CTAB	cetyl trimethyl ammonium bromide	**rms**	root mean square
D6	6-fold dihedral space group symmetry	**rpm**	revolutions per minute
DEAE	diethylaminoethyl	**SAD**	single wavelength anomalous dispersion
EDTA	ethylenediamine tetra-acetic acid	**SDS-PAGE**	sodium dodecyl sulphate polyacrylamide gel electophoresis
EM	electron microsocopy/microscope	**SEC**	size exclusion chromatography
FEG	field emission gun emission source	**SPEM**	single particle electron microscopy
FSC	Fourier shell correlation	**SSE**	secondary structure element
GGT	gamma-glutamyl transferase	**TEM**	transmission electron microscopy
GLB	gel-loading buffer	**tris**	tris(hydroxymethyl)aminomethane
GS	glutamine synthetase		

Table of Contents

Chapter 3

"X-ray Crystallographic Studies of GlnN"

Introduction

Glutamine synthetases (GS) are large oligomeric enzymes that play a central role in nitrogen assimilation, catalyzing the condensation of ammonium and glutamate to form glutamine, a precursor for the synthesis of many critical bio-molecules. In enteric bacteria the primary route of nitrogen assimilation under low nitrogen conditions is the dual enzyme system of GS coupled to glutamate synthase [Merrick *et al.* 1995]. In this system the synthesis of glutamine is driven by the energy from the hydrolysis of ATP to ensure efficiency in the absence of high levels of nitrogen. This is the only known route for the biosynthesis of glutamine ([Tyler 1978] cited in [Abrahams 1996]).

The GS enzymes are thought to be one of the most ancient functioning enzymes in existence [Kumada *et al.* 1993] and they are found in all forms of life from primitive to higher organisms [Pesole *et al.* 1991]. This superfamily of ubiquitous enzymes is, therefore, large and evolutionarily diverse. GS enzymes can be divided into three main families (Figure 1), however, the distribution of GS enzymes among the branches of the Tree of Life does not follow these sequence segregations and interestingly most organisms possess multiple functioning enzymes belonging to different families. For instance, enzymes with significant homology to GSI enzymes, which were previously thought to be bacterial in origin, have recently been identified in mammals and plants [Mathis *et al.* 2000; Wyatt *et al.* 2006]. Equally, representatives of the GSII family, which were thought to be exclusive to eukaryotes, have also been found in free-living soil bacteria [Kumada *et al.* 1990]. Confusingly, when an additional GS gene was discovered in the latter *Rhizobiaceae* spp. it was introduced as a separate family. However, because of the similarity to GSIs (Figure 1c) this group is not recognised as a completely separate family by all researchers and is more commonly designated as GlnT. These seeming redundancies hint at specialized roles for the divergent enzymes which appear to interact in complex fashion to mediate the critical process of nitrogen assimilation.

Because of their central importance, GS enzymes are regulated at numerous levels by many different mechanisms [reviewed by Eisenberg *et al.* 2000]. Consequently, the properties of these enzymes, such as catalytic potential, reaction kinetics, pH optima, sensitivity to feedback inhibitors, and conformational state, are modified by a number of effectors and conditions in a complex fashion. Over the past 5 decades, extensive biochemical and more recently structural studies have built a picture of the mechanism of functioning and regulation in the GSI and GSII families (Chapter 3 - Introduction).

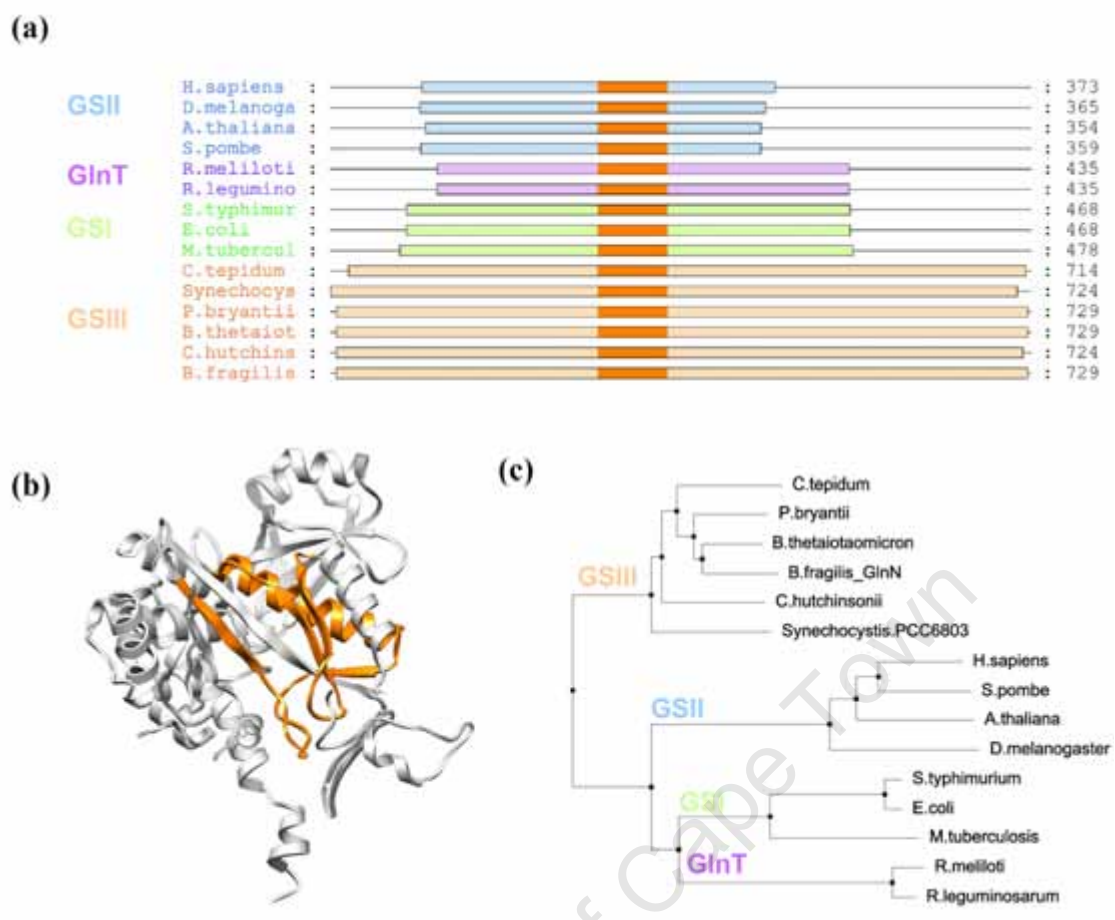


Figure 1: Phylogenetic distribution of the GS superfamily. **(a)** Un-gapped multiple sequence alignment of representative members of the main GS families showing the region of highest sequence conservation in orange. (**b**) Crystal structure of the *S. typhimurium* GSI monomer [Liaw *et al.* 1993b] showing the location of the conserved motif. The protein backbone is shown as a ribbon trace and the view is into the active site with the dodecameric ring interface towards the bottom of the image. **(c)** Phylogenetic neighbour joining tree calculated from the most conserved region (orange) shown in (**a**) & (**b**).

The divergent GSIII family, on the other hand, is the most recent family to be recognized and is, therefore poorly characterized. The first GSIII enzyme to be identified was the GlnN protein from *Bacteroides fragilis* [Southern *et al.* 1986], which is the primary causative agent of abdominal and systemic infections following trauma to, or surgery on, the digestive tract [Gibson *et al.* 1998]. Homologous enzymes have subsequently been found in the anaerobic bacteria: *B. thetaiotaomicron* [Xu *et al.* 2003], *Butyrivibrio fibrisolvens* [Goodman *et al.* 1993], and *Ruminococcus spp.* [Amaya *et al.* 2005]. Sequence analysis has also shown that the evolutionarily distant *Synechocystis spp.* (a photosynthetic blue-green alga) [Reyes *et al.*

1994], as well as *Deinococcus radiodurans* [White *et al.* 1999] possess GS enzymes with significant amino acid homology to *B. fragilis* GlnN. Interestingly, the sequencing of several genomes has recently uncovered the occurrence of these enzymes in single-celled eukaryotes such as algae [Robertson *et al.* 1996] as well as the pathogenic protozoan *Trichomonas vaginalis* [Kinoshita *et al.* 2009]. Together with the occurrence of both GSI and GSIII in *Synechocystis* [Reyes *et al.* 1997], these findings bring the GSIII family in line with the previously noted multiple distributions seen in the other families and suggest a possible pre-prokaroytic origin for divergence of the GS superfamily.

Despite their widespread occurrence, to date, only low-resolution structural information exists to describe these very large (~ 1 MDa) GSIII enzymes [van Rooyen 2004] (Chapter 2 – Introduction) and due to their divergence from the GSI family (only 9% global sequence identity) an understanding of their function in the light of the prior knowledge of GSI and GSII structure-function relationships has not been possible. Therefore, the major aim of this work is to address this lack of high resolution structural information using the complementary structural techniques of cryo-EM and X-ray crystallography.

The application of the **cryo-EM and SPEM imaging techniques** aims to improve upon the previous low-resolution reconstructions, which have already demonstrated the suitability of the large GlnN enzyme from *B. fragili*s to investigation by these methods (Chapter 2 – Introduction). In addition, this technique also has the potential to elucidate the structural basis of the cooperative subunit interactions that are thought to occur upon ligand binding and catalysis in the large oligomeric GS complexes [Wedler *et al.* 1982; Rhee *et al.* 1981].

With the higher resolution afforded by an **atomic resolution X-ray crystal structure**, this work aims to provide significant insights into the structure-function relationships between these divergent GS families and, thereby, improve our understanding of the poorly characterized GSIII enzymes. Comparisons with the previously solved GS structures will also help elucidate the role that quaternary structure has played in the evolution of GS function by revealing which sequence motifs are responsible for the conservation of higher order associations in the superfamily (Chapter 3 – Introduction).

Because of their unique ability to synthesise glutamine under conditions of low nitrogen, GS enzymes are targets for the **rational design of drugs** against commercially and medically

important organisms. Herbicides targeting the GSII enzymes in plants have been commercially available for some time [Obojska, *et al.* 2004] and research is currently underway to design selective inhibitors against the GSI secreted by *Mycobacterium tuberculosis* [Harth *et al.* 1999]. The GlnN enzyme from *B. fragilis,* has been demonstrated to be essential for growth under conditions of low nitrogen [Tumba 2007] and the divergence of this enzyme, therefore, represents a potential opportunity to design new drugs with selective activity against this important opportunistic human pathogen. The recent discovery of a GSIII enzyme in one of the most prevalent protozoan infections in the developed world [Kinoshita *et al.* 2009; Soper 2004] is further motivation for targeting these enzymes. An additional aim of attaining an atomic resolution structure for the *B. fragilis* GlnN enzyme is, therefore, to evaluate the potential for the design of selective inhibitors against these enzymes through a comparison of active site residues and ligand binding interactions in the GSIII and GSII families.

The following chapters describe the efforts to achieve these goals and each is prefaced by a summary which highlights the current state of knowledge surrounding the structural studies of GS proteins whilst demonstrating the applicability of the chosen techniques.

Chapter 1

Expression, Purification, and Crystallization of GlnN

1.1
Summary

GS enzymes have been isolated from numerous different organisms by an array of procedures and together with advances in recombinant technology and sample stabilization these efforts have led to the crystallization of several GSI and GSII enzymes. This chapter describes the isolation of a recombinant GSIII, *B. fragilis* GlnN, for structural characterization by cryo-EM (Chapter 2) and X-ray crystallography (Chapter 3). The quantity and quality of GlnN protein purified using the previously established protocol [van Rooyen 2004] was sufficient for the initial low-resolution cryo-EM investigations (Chapter 2), however, several optimizations were necessary to produce sufficient homogenous and stable material for crystallization.

These investigations suggested that the poor solubility of GlnN was due to the tendency of denatured material to precipitate from solution and solubilizing/stabilizing additives were incorporated to counteract this. It was also discovered that GlnN was sensitive to precipitation by divalent cations and this property was taken advantage of to design a rapid and scalable purification protocol for the production of homogenous GlnN in sufficient quantities for crystallization. This scheme combined the selectivity of precipitation by zinc with the ability of ATP-analogue affinity chromatography to remove denatured protein and the primary contaminating protein GroEL.

These improvements to the sample stability and solubility led to the discovery of conditions for the production of diffraction quality crystals following systematic screenings using commercial sparse matrices. The serendipitous discovery of limited digestion of GlnN near residue 430 by a bacterial protease suggested that this region is a surface exposed loop. Digestion of GlnN with the isolated protease left the complex intact but appeared to modify this surface region sufficiently to yield crystals with higher symmetry packing, and improved growth and diffraction characteristics when grown under the same conditions as the full-length GlnN protein (Chapter 3).

This work has provided a better understanding of the factors limiting the solubility of GlnN and in so doing has led to the design of a rapid purification protocol and the production of diffraction quality crystals (Chapter 3).

1.2
Introduction

"Developments in the Isolation of GS Enzymes for Structural Studies"

1.2.1 Early purifications

Since the first discovery of glutamine synthetase activity in eukaryotic tissue extracts [Krebs *et al.* 1935], these ubiquitous enzymes have been isolated from numerous sources using a variety of purification techniques over the last half-a-century of biochemical, and more recently structural, investigations. The early isolations utilized rudimentary biochemical precipitation techniques such as isoelectric (acidic), ethanol [Speck 1948], ammonium sulphate, nucleic acid [Elliott 1953], and heat-induced precipitations [Elliott 1951] to recover GS activity from diverse sources such as peas [Elliott 1953], sheep's brain [Elliott 1951], and bacteria [Woolfolk *et al.* 1966]. Because the early extractions were primarily from eukaryotic tissues and *Enterobacteriaceae* cultures, the majority of biochemical knowledge concerns the GSI and GSII families. The isolations from the latter group have had the greatest impact on the field by allowing larger quantities of the enzyme to be recovered more easily than from animal or plant tissue. Such systems have also permitted the isolation of GS enzymes with different levels of post-translational regulation, such as adenylylation, through the direct manipulation of the growth conditions [Shapiro *et al.* 1970]. The method of Woolfolk and Stadtman [1966] was first used to isolate the GSI enzyme from *E. coli* and material prepared by this technique was subsequently used in many biochemical investigations [reviewed by Eisenberg *et al.* 2000]. In this approach, a heat denaturation step, taking advantage of the high stability of the complex, followed by acetone and then acidic ammonium sulphate precipitations were used to produce homogenous GSI. Shapiro and Stadtman [1970] subsequently improved the purification protocol by including several additional acidic ammonium sulphate precipitations and a final crystallization step.

1.2.2 Purification improvements

The early purification protocols were adapted and improved upon to isolate enzymes from other sources and to meet the requirements of more extensive characterizations. The advances made to the purification schemes fall into three categories:

Bulk precipitations

Miller *et al.* [1974] made use of the propensity of GSI from *E. coli* to precipitate from solution as ordered tubular aggregates in the presence of divalent cations [as first discovered by Valentine *et al.* 1968] to rapidly isolate pure GSI material. This method, however, did not prove to be universally applicable with only some GSI and no GSII enzymes exhibiting this precipitation behaviour [Denman and Wedler 1984]. Other bulk precipitation manipulations proved to be more useful and were incorporated into protocols as 'capture' or concentration steps along with the earlier acetone, acidic ammonium sulphate, and isoelectic precipitation techniques. Burton and Eisenberg [1980] used PEI precipitation to recover GSI from *E. coli* cell free extract (CFE) and similarly, Streicher *et al.* [1980] discovered that GS could be recovered from CFE with 10% PEG in the presence of 10 mM $MgCl_2$.

Chromatography

The rapid isolation and concentrating abilities of the precipitation techniques were supplemented with the selectivity of column chromatography in the majority of subsequent GS isolations. The first applications of chromatographic separation technology to the isolation of GS enzymes involved the use of anion-exchange, hydroxyapatite, and size exclusion resins. GSII from chicken neural retinas [Sarkar *et al.* 1972], sheep's brain [Rowe *et al.* 1970], and rat liver [Tate *et al.* 1972] were all isolated by chromatography on hydroxyapatite and DEAE columns with either ammonium sulphate or acid precipitation capture and concentration steps. The latter method also made use of size exclusion chromatography (SEC) as a final purification step. The selectivity of the SEC for the large GS enzymes was impressive enough to see its use as the sole chromatographic purification step following PEG and acetone precipitation in the method devised by Streicher *et al.* [1980]. This method was subsequently used to isolate GS enzymes from a large number of sources [reviewed by Rhee *et al.* 1985].

Later biochemical advances saw the inclusion of affinity chromatography (AF) as the principal purification step in the isolation of GS enzymes. The first AF ligand used was anthranilic acid, a known GS inhibitor [Palacios *et al.* 1976]. In this early application, anthranilate-sepharose was used to isolate GSII from *Neurospora crassa* after an initial DEAE ion exchange purification. More recently, this matrix has been used to prepare material from *Zea mays* [Sakakibara *et al.* 1996] which was subsequently crystallized by Unno *et al.* [2006] to yield the first crystal structure of a GSII (Section 3.1). The purification entailed an ammonium sulphate capture step followed by Q-Sepharose anion exchange, anthranilate AF, and a final polishing step of SEC. After this initial application, more widely available dye-based affinity matrices for the purification of nucleotide binding proteins soon found widespread use in the isolation of GS enzymes. The Cibacron Blue 3G-A-agarose (Affi-Gel Blue or AGB) matrix was first used by Jaenicke *et al.* [1977] to isolate GSII from pig's brain and it was subsequently incorporated into many already established purification schemes. Stadtman *et al.* [1980] was the first to show that the GSI from *E. coli* could be purified by this method and Burton *et al.* [1980] utilized this technique as their main purification step in the production of *E. coli* GSI following PEI precipitation. Janson *et al.* [1984] later adapted the Shapiro and Stadtman [1970] purification method to include a final Affi-Gel Blue purification step for the isolation of *Salmonella typhimurium* GSI from an adenylylation deficient mutant. This material was of sufficient quality and quantity to produce the first crystal structure of any GS [Almassy *et al.* 1986] (Section 3.1). Another dye-based affinity resin, Reactive Red 120-agarose, was also used at a later stage for the isolation of GSII from *Rhizobium leguminosarum* [Manco *et al.* 1992] and in one instance, thiopropyl-Sepharose AF was successfully applied to the recovery of GSI from pathogenic *Mycobacterium tuberculosis* [Harth *et al.* 1994].

Recombinant technology

In more recent years the application of recombinant DNA technology has not only allowed the production of far larger quantities of material, particularly in the case of eukaryotic GS enzymes, but it has also simplified their purification through the introduction of affinity tags. The first application of this technique to isolation of GS enzymes was the purification of *Phaseolus vulgaris* GSII from an *E. coli* expression host [Betti *et al.* 2002]. The low-resolution structure of this his-tagged enzyme was later solved by cryo-EM and image processing [Llorca *et al.* 2006]. The isolation of GSI from human pathogen *Mycobacterium*

tuberculosis has also been advanced via recombinant technologies. Initially, GSI was isolated directly from pathogen cultures using a combination of ammonium sulphate precipitation, DEAE anion exchange, thiopropyl-Sepharose affinity, and SEC chromatography purifications [Harth *et al.* 1994]. Subsequent cloning of the *M. tuberculosis* GSI into an adenylylation deficient *E. coli* strain made it easier to recover the protein and led to the production of diffraction quality crystals [Gill *et al.* 2002] (Section 3.1). This purification protocol was an adaptation of the earlier Woolfolk *et al.* [1966] method but included a final Affi-Gel Blue affinity purification step [Gill *et al.* 1999] as used in the preparation of material for crystallization of *S. typhimurium* GSI [Janson *et al.* 1984]. The purification process was greatly simplified and the large culture volumes necessary with this method were subsequently avoided when his-tags were introduced into the recombinant *M. tuberculosis* protein [Singh *et al.* 2004; Krajewski *et al.* 2005]. The GSI enzyme from *Corynebacterium glutamicum* [Li *et al.* 2009] has also recently been purified using this technology and so too has the human, dog [Krajewski *et al.* 2008], and *Saccharomyces cerevisia* GSII enzymes [He *et al.* 2009].

1.2.3 GSIII isolations

In comparison to the extensive number of isolations from the GSI and GSII families, only three enzymes have been isolated so far from the most recently discovered family, the GSIII enzymes. The first to be isolated was the GlnN enzyme from *B. fragilis* which was cloned into an *E. coli* expression system by Southern *et al.* [1986]. The *glnN* gene was first identified from a *B. fragilis* genomic library and the resulting clone (pJS139) was found to express the structural gene under the control of its own promoter. The purification protocol was similar to the earlier method devised by Streicher *et al.* [1980] in that it used differential PEG fractionation to first isolate GlnN from CFE and then achieved purification using SEC. During these investigations Southern *et al.* [1986] discovered that the cloned GlnN enzyme displayed some level of nitrogen-induced down regulation during expression. This phenomenon was later attributed to feedback inhibition and not an adenylylation based regulation cascade as seen in *Enterobacteriaceae* [Southern *et al.* 1987]. In addition, the same investigation revealed the presence of an inhibitory phenomenon at work in both the recombinant and wild type native expression systems as evidenced by an over-recovery of enzyme activity during the purification protocol. The purification strategy of Southern *et al.* [1986] was later adapted by van Rooyen [2004] to prepare GSIII material, from the same

construct and expression system, for structural determination by electron microscopy. The primary modifications involved the replacement of the 4 and 6% differential PEG fractionation steps with a single higher percentage (10%) fractionation step and the resolubilization of the resulting precipitated fraction in high salt buffer as described by Streicher *et al.* [1980]. A decade after the first GSIII isolation, the GlnN enzyme from *Synechocystis* sp. PCC 6803 was isolated by Garcia-Dominguez *et al.* [1997]. The two enzymes share 41% sequence identity and are, therefore, expected to possess similar properties but the latter purification was methodologically quite different. It made use of DEAE anion-exchange and Reactive Red 120-agarose AF to purify the enzyme directly from the nitrogen-starved *Synechocystis* cultures and from an *E. coli* expression host. More recently, Amaya *et al.* [2005] have cloned and isolated the GSIII from *Ruminococcus albus* using a 6-histidine fusion tag. After the initial IMAC purification step, the enzyme was further purified by Q-Sepharose anion exchange and SEC.

1.2.4 Problems encountered during isolations

Problems of stability and poor solubility have been encountered throughout the history of GS purifications, with the different GS families exhibiting these characteristics to differing degrees. The GSI enzymes appear to be the most stable as evidenced by their tolerance of high temperatures (63°C), which was taken advantage of during purifications [Woolfolk *et al.* 1966], and their high solubility for crystallization (~100 mg/ml) [Gill *et al.* 1999]. The solubility problems that have been encountered have usually occurred in systems expressing recombinant constructs containing affinity tags in addition to the insoluble truncated lengsin protein from humans [Wyatt *et al.* 2006]. Li *et al.* [2009] showed that only N-terminal 6-histidine tags were soluble as expected from the early crystal structures of GSI which clearly showed that the C-terminus was not solvent accessible [Almassy *et al.* 1986]. The recombinant GSI from *M. tuberculosis* was even robust enough to be recovered, in an active state, from inclusion bodies after *in vitro* refolding [Singh *et al.* 2004]. The GSII family, however, appears to be the least tractable with a tendency towards dissociation under dilute conditions [Denman *et al.* 1984; Manco *et al.* 1992]. This latter property probably explains the great variety of reported quaternary structures for these enzymes as deduced from biochemical [Manco *et al.* 1992; Denman *et al.* 1984], electron microscopy-based [Llorca *et al.* 2006; Tsuprun *et al.* 1987] (Section 2.1), and crystallographic [Unno *et al.* 1996; Krajewski *et al.* 2008] (Section 3.1) investigations. GSII enzymes have also been purified

with both N and C-terminal fusion tags [Kraweski *et al.* 2008], in line with the the solvent accessibility of both regions revealed by the first structure of a GSII [Unno *et al.* 2006] (Section 3.1). In the latter isolation of the dog GSII, sensitivity to cold storage and a tendency to precipitate upon concentration were also noted. As only three enzymes have been isolated so far from the GSIII family, generalizations cannot be made about the solubility of these enzymes. However, the poor recoveries of active material from the purifications of van Rooyen [2004] and Garcia-Dominquez *et al.* [1996] hint at major solubility limitations in the current isolation protocols. Despite an optimum 'temperature of reaction' of around 45°C [Southern *et al.* 1986], the recombinant GSIII from *B. fragilis* could not be fully recovered upon resolubilization of the PEG precipitated fraction and the soluble enzyme fraction displayed a tendency to irreversibly precipitate upon concentration [van Rooyen 2004].

1.2.5 Sample improvements

In response to similar problems, several groups have supplemented the standard isolation and crystallization solutions with solutes and ligands to improve the recovery and stability of purified GS enzymes. The majority of modern purifications have used similar conditions for their isolations: the consensus appears to be either a tris-HCl/imidazole/phosphate buffer system with a pH between 7 – 8, including 5 – 20 mM divalent cations and 50-100 mM salt (usually KCl). The choice of divalent cation species (Mg^{2+} or Mn^{2+}) is usually determined empirically to suit the state of regulation of the isolated enzyme, thus, yielding the highest level of activity. The activity of GS enzymes is modulated by the divalent cation concentration and species, and these cofactors are, therefore, essential for the functioning and stability of the enzymes [reviewed by Eisenberg 2000]. Protease inhibitors, EDTA (~ 0.5 mM), and reducing agents such as 2-mercapto-ethanol (2 mM) are also routinely included. Where specific problems with solubility have been encountered or higher stabilities have been desired, such as in crystallization experiments, several additives have found common use. GSII extractions routinely include 10-20% glycerol to stabilize the oligomeric structure of the enzymes [Llorca *et al.* 2006; Tate *et al.* 1972; Denman *et al.* 1984] and all the GSI isolations carried out with the Streicher method [1980] contained 10% glycerol [reviewed by Rhee *et al.* 1985]. The isolation of his-tagged *P. vulgaris* GSII was similarly stabilized by ethylene glycol [Betti *et al.* 2002] and the direct isolation of *M. tuberculosis* GSI by Harth *et al.* [1994] utilized 10% sorbitol. Garcia-Dominguez *et al.* [1997] also found that glycerol and Mn^{2+} improved the stability of the *Synechocystis* GSIII but not to the desired level. Amaya *et al.*

[2005] in their isolation of the *R. albus* GSIII, however, found no such correlation between divalent cations and stability. It has also been shown by several groups that the binding of ligands and substrates to GS enzymes leads to increased stability of the oliogmeric complex. Maurizi *et al.* [1982b] showed that *E. coli* GSI enzymes, auto-inhibited by MetSox and ATP, were not only stable for six months at 4°C but were also resistant to disruption by several harsh conditions including: high salt, alkali pH, and heating in the presence of chelating agents. The inactivated complex was also more resistant to unfolding by 6 M guanadinium-HCl [Maurizi *et al.* 1982a]. *E. coli* GSI was also protected against proteolysis under non-denaturing conditions by the substrate glutamate and several related feedback inhibitors including CMP [Lei *et al.* 1979]. Denman *et al.* [1984] achieved similar results for the GSII enzymes from pig brain by showing that the substrates glutamine and ADP promoted the reassociation of enzymatic tetramers into higher order octamers. The recent efforts of Krajewki *et al.* [2008] to stabilize the human GSII enzyme have confirmed this result by quantifying the stabilizing properties of several co-factors. Dye-based melting curves were used to investigate the effect of various solutes and ligands on the thermal stability of the purified enzyme. Mn^{2+} and ATP were shown to have the largest effect, extending the melting temperature by 15°C. These discoveries have prompted the inclusion of inhibitors, non-hydrolysable nucleotides and other substrates and ligands in the crystallization solutions of several GS enzymes to try and reduce their conformational heterogeneity (reviewed in Section 3.1).

1.2.6 Summary

Developments in genetic and biochemical technologies have greatly facilitated the recovery of the large GS enzymes from a diverse number of sources. These rapid advances have not, however, been without problems and the stability and solubility of the enzymes has had to be optimized in a case by case manner. In several instances, the attainment of homogenous and sufficiently stable material has resulted in the production of diffraction quality crystals.

1.3
Materials and Methods

1.3.1 Protein expression and cell lysis

Expression was achieved using an auxotrophic *E. coli* strain, YMC11 (glnA-, ntrB-, ntrC-, ApS) [Backman *et al.* 1981] transformed with a pEcoR1-derived low copy number plasmid, pJS139, carrying the structural gene for glutamine synthetase (glnN) from *B. fragilis* BF-1 on an 8.7 kb insert [Southern *et al.* 1986]. The initial conditions for expression were similar to those described by van Rooyen [2004] with the modification that the low nitrogen induction and CTAB treatment steps were omitted. *E. coli* YMC11 (pJS139) was grown with aeration for 6 hrs at 37°C in precultures (5 ml) of Luria Broth [Davis *et al.* 1980] containing ampicillin at 100 µg/ml (LB + ap) following inoculation from single colonies previously plated out on LA (LB + ap). Aliquots (500 µl) were used to inoculate each of three large cultures (2.5 L LB + ap) which were then incubated for approximately 16 hr at 37°C with aeration. Cells were collected by centrifugation for 15 minutes at 7 500 rpm at 4°C, and resuspended in 1/50th the original culture volume in extraction buffer (50 mM imidazole, 50 mM KCl, pH 7.1). Cells were disrupted by sonication for 4 minutes (15 second cycles with 15 seconds cooling periods) using a Misonix sonicator fitted with a microtip operating at a power output of 9 W, and using ice/ethanol slurry to cool the sample. The supernatant was then clarified by centrifugation at 15 000 rpm for 30 minutes. This cell free extract (CFE) was then diluted 1:2 with glycerol, aliquoted into 10 ml Sterilin® plastic tubes and flash-frozen in liquid nitrogen. Frozen CFE samples were then immediately moved to -80°C freezer for long term storage.

1.3.2 Protein purification

Purification optimizations

The following section describes the methods tested during the design of the purification protocol (as summarized in Figure 1 below). The final purification conditions are given below.

All chromatography matrices were from GE Healthcare (formerly Pharmacia/Amersham) and standard reagents and chemicals were sourced from Sigma-Aldrich unless otherwise stated.

Starting material was obtained from rapidly thawed CFE containing 50% glycerol and stored at -80°C. The initial differential PEG precipitation fractionations were performed with PEG 6000 added drop wise, with stirring, from a 50% stock solution. During the first precipitation step, PEG 6000 was added to a final concentration of 6% (w/v) and left for 1 hour at ambient temperature. GlnN remained in solution and was separated from precipitated material by centrifugation at 15 000 rpm for 30 minutes. Additional PEG was then added to bring the final concentration to 9% (w/v) before incubation at ambient temperature for 1 hour followed by storage at 4°C overnight. Thereafter, the precipitated GlnN was recovered by centrifugation at 15 000 rpm at 4°C for 30 min and resuspended in extraction buffer containing 4 M NaCl at half the original precipitation volume as adapted from Streicher *et al.* [1980].

Both the Affi-Gel Blue (ABG) affinity and anion exchange purifications were carried out using the same buffer systems. Sample loading and washing of the 25 ml columns took place in a low salt buffer (50 mM imidazole, 50 mM KCl, and 20 mM $MgCl_2$ pH 7.1). Elution was then achieved over 10 column volumes with a linear gradient between 50 mM and 1M KCl, made up in the low salt buffer. Initially, a Q-Sepharose anion exchange resin (6% cross-linked agarose) was used but subsequent purifications employed an ANX Fast Flow matrix (4% cross-linked agarose) with a bigger MW exclusion limit. The chromatographic purifications were carried out on a Waters Delta Prep 3000 LC machine with a flow rate of 5 ml/min. Samples were loaded from reservoirs kept at a temperature of 0°C but the columns were not refrigerated. The elution of protein in the eluent was monitored by the inline Waters 484 spectrophotometer and fractions were collected automatically by a Gilson 204 fraction collector.

Preparative SEC was carried out on several related matrices, all with different exclusion limits and column volumes: HR Sephacryl 300 (120 ml), 400 (420 ml), and 500 (180 ml). SEC chromatography was carried out in 15 mM imidazole, 100 mM KCl, and 20 mM $MgCl_2$ pH 7.1 at ambient temperature on a Gilson system equipped with a 305 pump and 806 manometric module at the recommended linear flow-rate for the resins. Protein concentration

in the eluent was monitored by the inline Gilson 151 spectrophotometer and fractions were collected using a Gilson 203b unit.

Hydrophobic interaction chromatography was performed on a Butyl-Sepharose matrix equilibrated with a 0.8 M solution of ammonium sulphate in 15 mM imidazole and 20 mM $MgCl_2$ pH 7.1. Elution was achieved with a linear gradient from 100 to 10% of the ammonium sulphate stock solution and samples of GlnN protein were tested for stability in the mobile phase prior to purification. This purification was performed on the Waters chromatography system as described for the IEX and affinity purifications.

Concentration and buffer exchanges of large volumes (up to a volume of 5 ml) were performed in an Amicon (Bioseparations) stirred cell with either a Millipore PM 100 kDa or 10 kDa molecular weight cut-off (MWCO) filter operating at the maximum permitted pressure at 4°C. Smaller volumes were processed in the Micro (3.5 ml) and Nanosep (500 µl) Centricon centrifugal concentrators from Pall Corporation.

Solubility studies

PEG precipitation curves were calculated from semi-pure GlnN, prepared by ANX IEX chromatography, in the presence of several different ions and solutes. Aliquots of GlnN were buffer exchanged and concentrated by ultra-filtration into 15 mM imidazole pH 7.1 using a 100 kDa Nanosep concentrator (Pall Corporation). These samples were then diluted 1:2 with the concentrated stock solution of the solutes being investigated, before the addition of equal volumes of PEG 4000 solutions at varying concentrations. After 30 minutes incubation at ambient temperature, the activity of the GlnN remaining in solution was assayed as described below. Relative solubility was calculated from this activity by normalizing against the activity in the absence of any PEG but in the presence of the solute.

In addition, visual turbidity tests were carried out to rapidly ascertain the solubility of GlnN in the presence of various solutes. These assays also served to determine conditions that prevented the precipitation of GlnN. The precipitation experiments were carried out under Al's oil (Hampton Research) in microbatch crystallization trays (Hampton Research). Aliquots of GlnN (1 µl) were mixed with an equal volume of precipitant already dispensed under the oil and left to equilibrate for 30 min at room temperature. The amount of

precipitation was estimated qualitatively by inspection under a light microscope (see below). A semi-quantitative estimation of such a precipitation experiment was also carried out. Specifically, the ability of low pH to prevent Zn^{2+}-mediated precipitation was tested. Aliquots of GlnN (4 mg/ml), prepared by the Zn^{2+} - PEG/Mg^{2+} method, were diluted 1/10 with 0.2 M citrate/Na_2HPO_4 buffers of varying pHs before the addition of $ZnCl_2$ to 10 mM. The level of precipitation was then ascertained by measuring the turbidity of the solutions at 620 nm in the Titertek MultiScan plus MKII plate reader.

Final protein purifications for cryo-EM

The initial cryo-EM reconstruction (Chapter 2) was carried out using data collected from GlnN prepared in the same manner as described in van Rooyen [2004]. Briefly, in this method GlnN was isolated from CFE by differential PEG fractionation at 11% (w/v PEG 4000) followed by SEC on a 120 ml S300-HR column. Material for the second cryo-EM reconstruction was purified using an improved purification scheme that involved an anion exchange capture step (ANX Fast Flow Sepharose) followed by SEC on a S400-HR column and a final purification by HIC on a butyl-Sepharose column.

Final purification scheme for crystallization

Frozen cell free extract (22 ml) was thawed and diluted 1:2 with extraction buffer (50 mM imidazole, 50 mM KCl, pH 7.1). $ZnCl_2$ was then added to a final concentration of 2 mM and the preparation was left to precipitate for 1 hour at room temperature. Precipitated material was collected by centrifugation at 15 000 rpm at 4°C for 30 min and resuspended, using a Dounce homogenizer, in 20 ml ice-cold wash buffer (15 mM imidazole, 2 mM $MgCl_2$, pH 7.1). Following recollection of the washed material, a final resuspension was performed in 5 ml resuspension buffer (50 mM imidazole, 500 mM KCl, 10% glycerol, pH 7.1) and the sample placed at 4°C.

Insoluble material remaining after 1 h at 4°C was removed by a 30 min centrifugation step at 14 000 rpm in a desktop centrifuge at 4°C. Before loading onto the pre-equilibrated 5 ml Affi-Gel Blue HiTrap affinity column, the supernatant was diluted 1:10 with AGB equilibration buffer (50 mM imidazole pH 7.1, 50 mM KCl, 20 mM $MgCl_2$, 10% glycerol, and 0.02% sodium azide) and filtered through a 0.45 µm filter. During the loading of the column, the

sample was kept on ice but the column was maintained at room temperature. The column was then washed with 5 volumes of equilibration buffer before elution of the bound protein with 10 ml of 10 mM ATP in equilibration buffer. A period of 5 minutes was allowed to elapse between the application of the first column volume of elution buffer and the second. All chromatography steps were carried out at a flow rate of 1 ml/min.

After chromatographic purification, MetSox was then added to the sample to a final concentration of 8 mM and the sample was stored for 16 h at 4°C. This final preparation was concentrated using a Microsep 10 kDa MWCO centrifugal concentrator (Pall Corporation). After concentration to 1 ml, the retentate was transferred to a Nanosep concentrator and a final concentration of 6 mg/ml was achieved. This preparation (105 µl) was divided into 10 µl aliquots and flash-frozen in cryo-vials for storage under liquid nitrogen.

1.3.3 Biochemical assays and biophysical analyses

Protein quantitation

The Bradford assay [Bradford 1976], using bovine serum albumin (BSA) as a standard, was used to determine the protein concentrations. Bradford reagent was purchased from Bio-Rad and the manufacturers instructions were followed for the standard assay as designed for micro-titre plate applications. The results of the assays were read with a Titertek Multiscan plus MKII plate reader fitted with a 595 nm filter and the data were processed in the Genesis Lite software.

Measurement of activity

GS activity was assayed by the γ-glutamyl transferase (GGT) assay as described by Bender *et al.* [1977] and subsequently modified by Southern [1986]. Total assay volumes including reaction mixtures and samples were also reduced from 1.5 ml to 250 µl to allow measurement in the Titertek Multiscan plate reader. Specific activity was expressed as µmoles of glutamyl hydroxamate formed per min, per mg protein and 1 µmole of glutamyl hydroxamate gave 2.916 absorbance units at a wavelength of 540 nm.

SDS-PAGE

Discontinuous SDS-PAGE was carried out according to Laemmli [1970] using a Mini-PROTEAN 3 cell from Bio-Rad. Acrylamide (19:1 monomer:cross-linker) was used to prepare 5% stacking gels and 7.5 - 10% separating gels. Pre-stained molecular weight markers were included in all electrophoretic separations (Fermentas Life Sciences). Gels were stained either with 0.25% Coomassie Brilliant Blue R250 in 45% methanol and 10% acetic acid (destain was identical minus the stain) or by a non-ammonical silver-staining protocol [Blum *et al.* 1987] using reagents from the GE Healthcare Plus One silver-staining kit. Protein purity was estimated by densitometry using the program IMAGEJ [Abramoff *et al.* 2004]. Peak areas, corresponding to bands in the lane under investigation, were calculated after baseline subtraction and the purity of the GlnN band was determined from the ratio of its peak area to the total for all the bands in the lane.

MALDI-TOF identification of proteins

In-gel tryptic digestion and MADLI-TOF mass spectrometric fingerprinting of the resulting peptides were carried out using established protocols [Rosenfeld *et al.*1992]. Mass spectra were collected on an ABI 4800 MALDI TOF/TOF machine in reflection positive mode operating at a 20 kV source voltage and 16 kV grid voltage with a 400 ns delayed extraction time. A default calibration was applied and a scan range of 800 to -4000 m/z was used to collect data with 50 shots per sub-spectrum and a total of 1000 spectra. Data processing was carried out in GPS EXPLORER software from ABI and proteins were identified using the MASCOT web server. Tryptic digestion, sample cleanup, and MALDI-TOF MS analysis were carried out by Dr. Maré Vlok at the Centre for Proteomic & Genomic Research (Cape Town, South Africa).

1.3.4 Non-denaturing proteolysis

Isolation of a bacterial protease

A natural protease-producing bacterial isolate was cultured as a contaminant from a GlnN protein solution stored at 4°C. Cultures were grown on LA plates or in 5 ml volumes of LB at room temperature without aeration. The identity of the bacterium was confirmed by 16s

rRNA gene PCR and sequencing and shown to be *Psueodomonas fluorescens* (Mr L. Motlogelwa, Ms A Thorpe, and Ms C Williams). Protease activity was investigated at different temperatures on casein-agar plates (2% skimmed milk powder).

The culture supernatant (CS) was prepared from a 5 ml culture of *P. fluorescens* grown in LB and incubated at room temperature for two days. After centrifugation of the culture for 15 min at 14 000 rpm in a desktop centrifuge, the CS was stored at -20°C and used for all subsequent proteolysis experiments.

Protease susceptibility assays

Semi-pure GlnN, isolated by the Zn^{2+} precipitation method, was digested with both the *P. fluorescens* protease and trypsin (bovine pancreatic extract) as a reference. Aliquots of GlnN in 15 mM imidazole pH 7.1, at 4 mg/ml, were mixed with equal volumes of different concentrations of each protease and incubated at room temperature for one hour in the case of trypsin and 16 h for *P. fluorescens* protease. The trypsin dilution series was made up in 0.1 M tris-HCl pH 8.0 at a starting concentration of 0.08 mg/ml and each sample represented a 1:5 dilution. The *P. fluorescens* protease dilution series was achieved by serially diluting the CS 1:2 with dH_2O. Digestion was halted by the addition of gel loading buffer (GLB) and the samples were analyzed by SDS-PAGE. An untreated sample of GlnN was incubated at room temperature for 24 hrs to serve as a negative control. Enzyme activity was monitored by the GGT-transferase assay.

Characterization of digested GlnN complex

GlnN protein (10 mg/ml) prepared by the Zn^{2+} - S500 SEC method, was digested by adding an equal volume of *P. fluorescens* CS and incubating the mixture overnight at room temperature. A control was also prepared from undigested GlnN diluted 1:2 with equilibration buffer and left for several hours at room temperature. After clarification, in a microfuge at 14000 rpm for 15 min, these samples were loaded separately onto a Tosoh PWXL-4000 column pre-equilibrated with 0.1 M citric acid / 0.05 M NaH_2PO_4 pH 5.8 and 100 mM KCl at 0.4 ml/min. The column was previously calibrated with the high-molecular weight standards from Bio-Rad, 3% (v/v) acetone for *Vt* determination, and TMV for *Vo* determination. Discontinuous native PAGE [Ornstein *et al.* 1964] with a 2.5% stacking and a 4% acrylamide

separating gel was used. Samples (1 µg) were mixed with GLB (as described above but without SDS and mercapto-ethanol) prior to loading. Samples of pure GlnN were dialyzed for 16 h at 4°C in a mildly denaturing buffer (10 mM EDTA, 50 mM KCl, and 0.1 M Tris pH 8.8) before analysis by native PAGE.

1.3.5 Crystallization of full-length GlnN

Initial screening of crystallization conditions employed commercially available sparse matrix screens from Hampton Research: Crystal Screen; Qiagen (formerly NeXtal Biotechnologies): PEG/ion screen & JCSG+ screen; and Jena Bioscience: JBScreen Mixed. Protein concentrations for crystallization trials were decided upon using solubility screens (ammonium sulphate: 1, 3.2M; MPD: 23, 65%; and PEG 4000: 12%, 30% - adapted from the Qiagen/NeXtal pre-screen). GlnN, stored under liquid nitrogen, was thawed, diluted from 12 mg/ml with AGB equilibration buffer, and clarified by centrifugation at 14 000 rpm in desktop centrifuge just prior to use. Crystallization trials were carried out by mixing equal volumes of protein and precipitant (1 µl each) dispensed under 5 ml Al's oil (Hampton Research) in 72 well microbatch plates (Hampton Research) and stored at ambient temperature (21-23°C). All assessments of crystal growth and manipulations were performed on a Leica MZ6 stereo microscope fitted with a CLS 150X cold light source. Photographs of crystals were captured with a Zeiss Axiocam attached to a Nikon Diaphot inverted fluorescence microscope fitted with a 20x short working distance objective lens.

When it was possible to obtain sufficiently large crystals, the molecular weights of the proteins comprising the crystals were determined by silver-stained SDS-PAGE. Specifically for the identification of the full-length GlnN crystals, five large (~ 0.8 µm) crystals were washed individually and serially in 4 droplets (2 µl) of crystallization solution (Hampton condition 20 diluted 1:2 with H_2O) placed under oil in the microbatch plate. The manipulating loop was cleaned between each transfer to avoid carry-over of any crystallization medium. The washed crystals were finally dissolved in a drop of diluted SDS-PAGE loading buffer (2 µl H_2O + 2 µl GLB) before electrophoretic analysis.

1.3.6 Crystallization of digested GlnN

Proteolysis was achieved by mixing equal volumes of *P. fluorescens* CS with pure concentrated GlnN (6 mg/ml) in AGB AF equilibration buffer. This mixture was then incubated at room temperature overnight. The digested protein was then re-concentrated and the buffer exchanged three times to 1/5th the volume in equilibration buffer (1/125 reduction in solutes below the MWCO) using a 100 kDa MW cut-off Nanosep centrifugal concentrator (Pall Corporation) to give a final concentration of 3.5 mg/ml. Crystallization was achieved using the same conditions that proved successful for the full-length GlnN protein and growth was initiated by cross-seeding with nuclei from crystals of the latter. Further batches of crystals were prepared by streak-seeding equilibrated drops (1.75 mg/ml GlnN) with nuclei from these new crystals. Crystal growth to about 0.1 mm was achieved in several days (2-4).

1.3.7 Electron microscopy of GlnN crystals

Several µl's of crystallization solution were removed from the batch droplets under oil and applied to EM grids coated with a glow-discharged thin carbon support film, before being stained with 2% uranyl acetate solution using the droplet method as described in [van Rooyen 2004].

Electron micrographs were recorded by a Proscan 2048×2048 slow-scan CCD camera, with 14 µm pixels, using a Leo 912 TEM operating at 120 kV with zero-loss energy filtering. Fourier transforms were calculated using a FFT plug-in [Chirokov 2005] in Paint Shop Pro® and patch correlation averages were calculated in 2DX after Fourier filtering [Gipson *et al.* 2007].

1.4
Results

1.4.1 Improvements to the expression conditions

The initial preparations of GlnN for structural characterization by cryo-EM (Chapter 2) were performed using the same purification methods and heterologous *E. coli* YMC11 (pJS139) expression system first devised by Southern *et al.* [1986] and later adapted by van Rooyen [2004]. The decision to retain this system was based on the early successes of the preliminary structural investigations by van Rooyen [2004]. However, these expression protocols were simplified in the current study (Figure 1) after it was discovered that the low nitrogen induction, CTAB treatment, and the β-mercapto-ethanol additive had no impact on the yield or stability of GlnN isolated from the *E. coli* auxotrophic expression host cultured in large volumes of complete medium (data not shown). Isolations from smaller culture volumes with proportionately larger inoculums, however, displayed the same inhibitory phenomenon first identified by Southern *et al.* [1986] and, following the first fractionation step, activity was recovered far in excess of the level present in the CFE. During the isolations it was also noticed that the GlnN protein displayed a weak 280 nm but strong 260 nm absorbance as evidenced by SEC separations monitored at the corresponding wavelengths. All preparations of GlnN for structural studies were isolated from cultures grown in excess complete media to avoid the inhibitory phenomenon.

The assessment of different purification methodologies (see below) necessitated large numbers of cellular extractions with consistent levels of stable GlnN. For this reason, large volumes of CFE were prepared and rapidly frozen in liquid nitrogen after the addition of glycerol (final 50% v/v). In the absence of glycerol, GlnN precipitated out of solution as a fibrous mass during thawing. GlnN activity was recoverable from the cryo-protected frozen material and all subsequent purifications were achieved from this starting material.

1.4.2 Design of a novel purification procedure

Although the quality and the yields of the GlnN material prepared by the previous methods was sufficient for the preliminary low-resolution structural characterizations (Chapter 2),

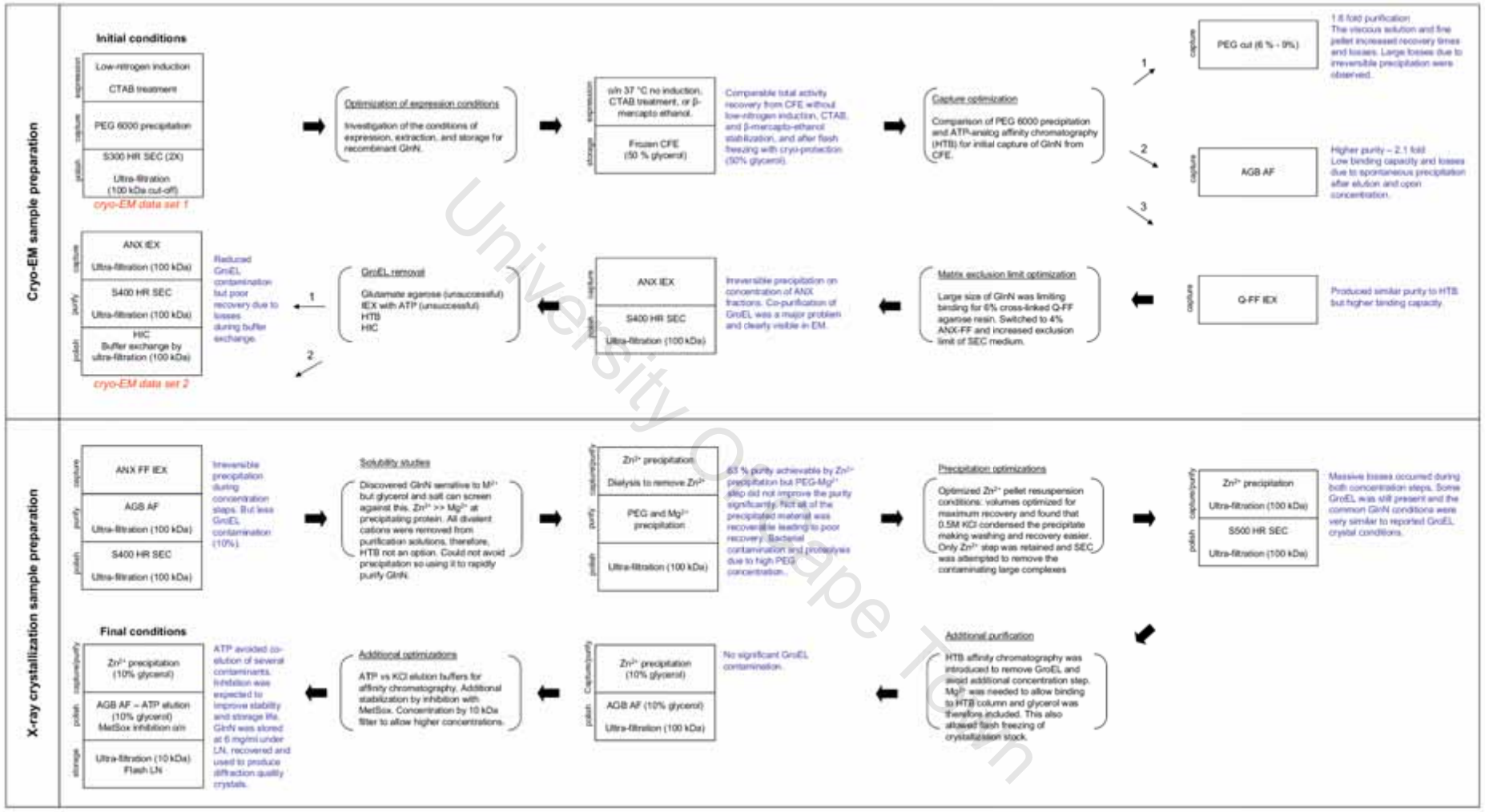


Figure 1: Flow-diagram of purification protocol optimizations. Purification steps are displayed as blocks with sequential steps appearing below each other. Optimizations are shown in parentheses and the resulting observations are highlighted in blue. Where several alternate strategies were tested these are numbered and the flow chart continues from the best/chosen strategy. Details of the optimizations are given in the text and the purification steps are explained in the methods section. The methods used to prepare material for the two cryo-EM studies shown in Chapter 2 are highlighted in red.

improvements were sought to the purification procedures to produce larger quantities of more homogenous material for higher resolution cryo-EM and crystallization trials. A series of optimizations, which are explained in detail in Figure 1, were undertaken to overcome the two major factors limiting the achievement of these goals, namely, limited solubility and limited purity. The two processes of sample improvement and crystallization screening (discussed below) paralleled each other over the course of the experiment with information flowing in both directions in an iterative fashion to inform decisions about the protein purification strategy and crystallization conditions.

Limited solubility

Over the course of this investigation, GlnN, displayed a propensity to precipitate out of solution as previously noted by van Rooyen [2004]. Fibrous material, not dissimilar to that seen upon freeze-thawing, would routinely appear upon concentration by ultra-filtration, and similar aggregates were sometimes visible in IEX, AF, and even more dilute SEC elution fractions. This material was identified as precipitated GlnN by SDS-PAGE following resolubilization in SDS (data not shown). In addition to these "spontaneous" precipitation events, information about the limited solubility of GlnN under more specific conditions emerged from the PEG fractionation steps and the numerous crystallization trials (see below). It was recognized early on that GlnN was very sensitive to precipitation by moderate MW polymers such as PEG and pH values below 5. Precipitations under these conditions were only partially reversible and this lead to significant losses, as previously noted by van Rooyen [2004].

This sensitivity to precipitation by PEG was utilized to investigate possible factors limiting the solubility of GlnN (Figure 2a). Due to the suggested mechanism of precipitation (by reducing excluded volumes and inducing molecular crowding [Ingham 1984]) the results of these PEG precipitation studies mimicked conditions of high protein concentrations, something which was not achievable due to the poor yields. It can be seen from Figure 2a that in the absence of any solutes, except for 15 mM imidazole pH 7.1, no precipitation of GlnN was detected up to a final concentration of 25% PEG 4000. In the presence of KCl, however, GlnN solubility decreased in proportion to the concentration of this salt until no further change was seen above 100 mM.

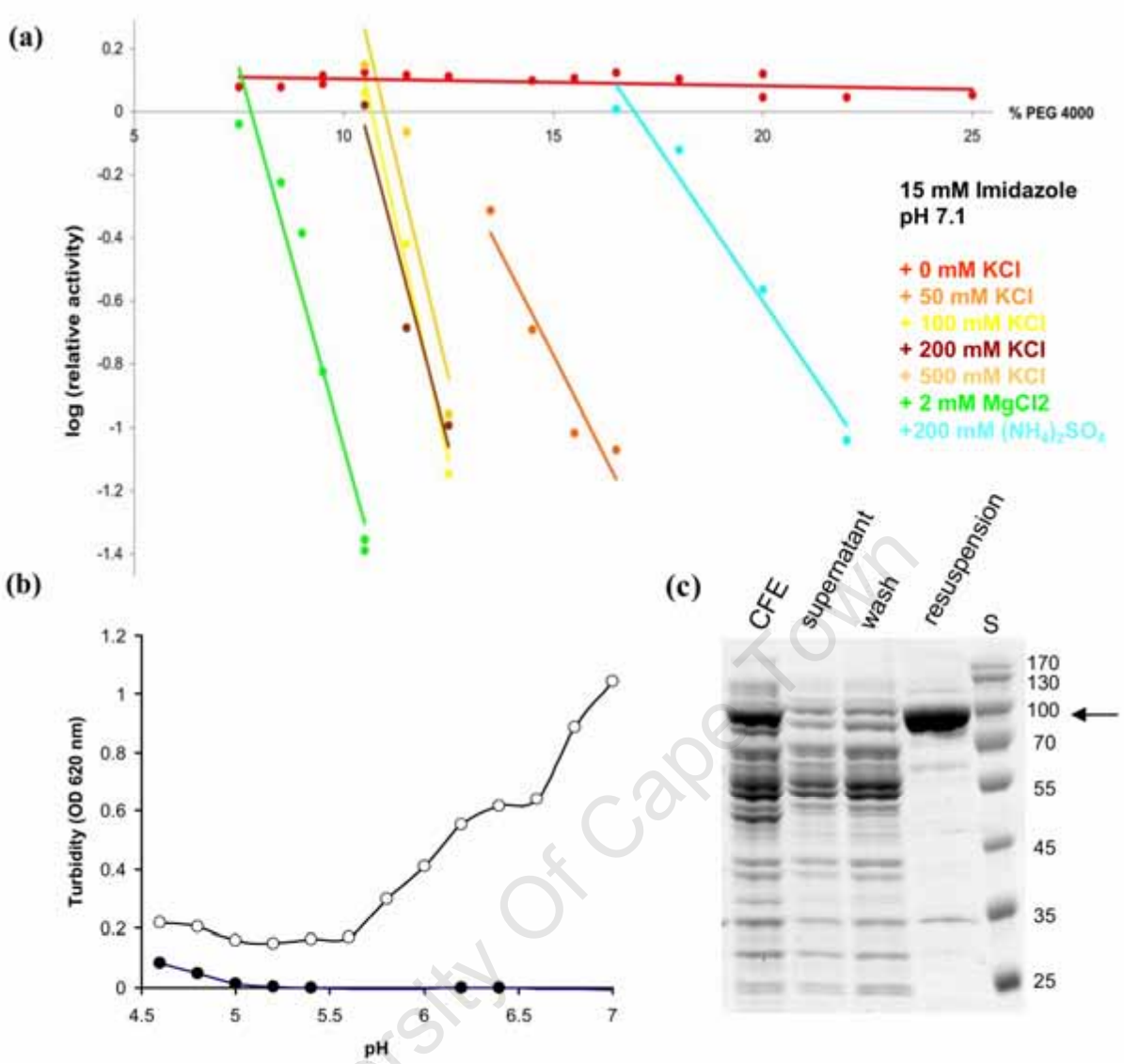


Figure 2: Limited solubility of purified GlnN. **(a)** The effect of different salts on the solubility of GlnN as measured by PEG precipitation curves. All samples contained 15 mM imidazole pH 7.1 supplemented according to the coloured key. These precipitation curves conform to the linear relationship between log (solubility) and the % PEG. The slope represents the rate at which the material precipitates and the Y intercept represents the maximum solubility of the protein in the absence of precipitant. **(b)** Screening of 10 mM $ZnCl_2$ induced precipitation at low pH as measured by the turbidity of the solution (open circles). The filled circles represent the baseline measurements before the addition of any $ZnCl_2$. **(c)** Coomassie Blue-stained 10% SDS-PAGE showing the selective precipitation of GlnN (black arrow) by Zn^{2+} ions. Molecular weight markers were loaded in the lane marked **S** and their weights are given in kDa on the right. All lanes contained 10 µg of protein with the exception of the CFE lanes which contained 20 µg.

Interestingly, equivalent concentrations of $(NH_4)_2SO_4$ had less of a negative effect but still decreased the solubility of GlnN as confirmed by crystallization trials containing PEG 3350 and 0.2 M $(NH_4)_2SO_4$. These results conflicted with the findings of the solubility studies for HIC chromatography (see below), which showed that in the absence of PEG the solubility limit of GlnN was reached in 2.4 M KCl in comparison to 0.8 M $(NH_4)_2SO_4$. The largest decrease in GlnN solubility in the presence of PEG was noted for magnesium even at very low concentrations (Figure 2a). Subsequent investigations revealed that, as in the case of GSI [Miller *et al.* 1974], GlnN, was more sensitive to precipitation by the divalent cation, Zn^{2+}, and complete precipitation was possible with this species even in the absence of any PEG. It was also discovered that moderate salt concentrations, such as found in the crystallization conditions (0.2 M ammonium sulphate), and pH values above 4.6 but below 5.8 could prevent this precipitation by divalent Zn^{2+} (Figure 2b).

Following this discovery, it was decided to utilize the partially reversible precipitation of GlnN by Zn^{2+} to rapidly purify the enzyme as previously described for GSI [Miller *et al.* 1974]. It can be seen from Figure 2c that the precipitation of GlnN by 2 mM Zn^{2+} is highly selective. The precipitation procedure presented here differs from the method of Miller *et al.* [1974] in several regards: (1) only the Zn^{2+} fractionation is used; (2) precipitation takes place at pH 7 and not under acidic conditions; (3) the wash solution contains only 2 mM Mg^{2+} instead of 50 mM; and (4) the resuspension volumes and the salt concentrations were optimized to improve the recovery of GlnN from the Zn^{2+} pellet. During the latter optimizations it was also found that high salt concentrations caused the insoluble material remaining after resuspension of the pellet to condense, thus, making recovery of the soluble supernatant fraction much easier.

Although these findings indicated that GlnN was sensitive to precipitation by divalent cations, they did not explain the spontaneous losses seen in dilute solutions of the semi-pure protein or the losses observed in the absence of divalent cations (data not shown). A search for solubility enhancing additives was, therefore, initiated and the results revealed that the solute with the greatest effect was glycerol. The first indication of this came from crystallization trials setup in the presence of the additive (10% glycerol) in an attempt to reduce nucleation of microcrystals (see below). At this concentration, glycerol prevented all protein precipitation, even at high PEG concentrations, and the additive was, therefore, included in all future purification and crystallization solutions. The cryo-protective abilities of glycerol also

allowed the final concentrated GlnN crystallization stock to be directly flash-frozen in liquid nitrogen for long-term storage.

Protein purity

The second major problem thought to be hindering the crystallization of GlnN was the inability of the early purification procedures to separate GlnN from a similarly sized contaminating protein complex (Figure 3a). This complex, which co-eluted with GlnN from a number of chromatographic matrices (Figure 3a and Figure 1), was putatively identified as GroEL on the basis of its characteristic geometry, as seen under the EM (Figure 3b), and the MW of the monomer, as given by SDS-PAGE (Figure 3a). The EM analyses also revealed a close association between the GlnN and GroEL complexes purified by the ANX-AGB-SEC method (Figure 3c). The presence of this contaminant, even at trace levels, was particular concerning because the reported crystallization conditions for the native protein from *E. coli* [Bartolucci *et al.* 2005] were very similar to the conditions that produced the most crystallization phenomena (see below). The two most successful methods for removing GroEL were the already proven AGB AF [Burton *et al.* 1980] and hydrophobic interaction (HIC) [Bartolucci *et al.* 2005] separations (Figure 3a). The latter scheme together with a SEC purification step was used to prepare the sample for the second cryo-EM reconstruction discussed in Chapter 2 (Figure 1). However, the large losses observed during the three concentration steps required for this method precluded its usefulness for preparing the large amounts of protein needed for crystallization trials. The AGB affinity step was, therefore, combined with the Zn^{2+} precipitation step to produce the final purification protocol, which showed almost no GroEL contamination compared to the results from SEC (Figure 3a).

Final purification protocol

Several additional enhancements were made to the purity and stability of the GlnN preparations prior to crystallization (Figure 1 and Figure 4a). Firstly, the purity of the fraction eluted from the Affi-Gel Blue affinity matrix was improved by changing from an ionic based elution to a more specific competitive elution with ATP (Figure 4a). This alteration also facilitated the next improvement which was the inhibition of GlnN by the well known GS inhibitor MetSox, the aim of which was the improved stability of the final preparation. The

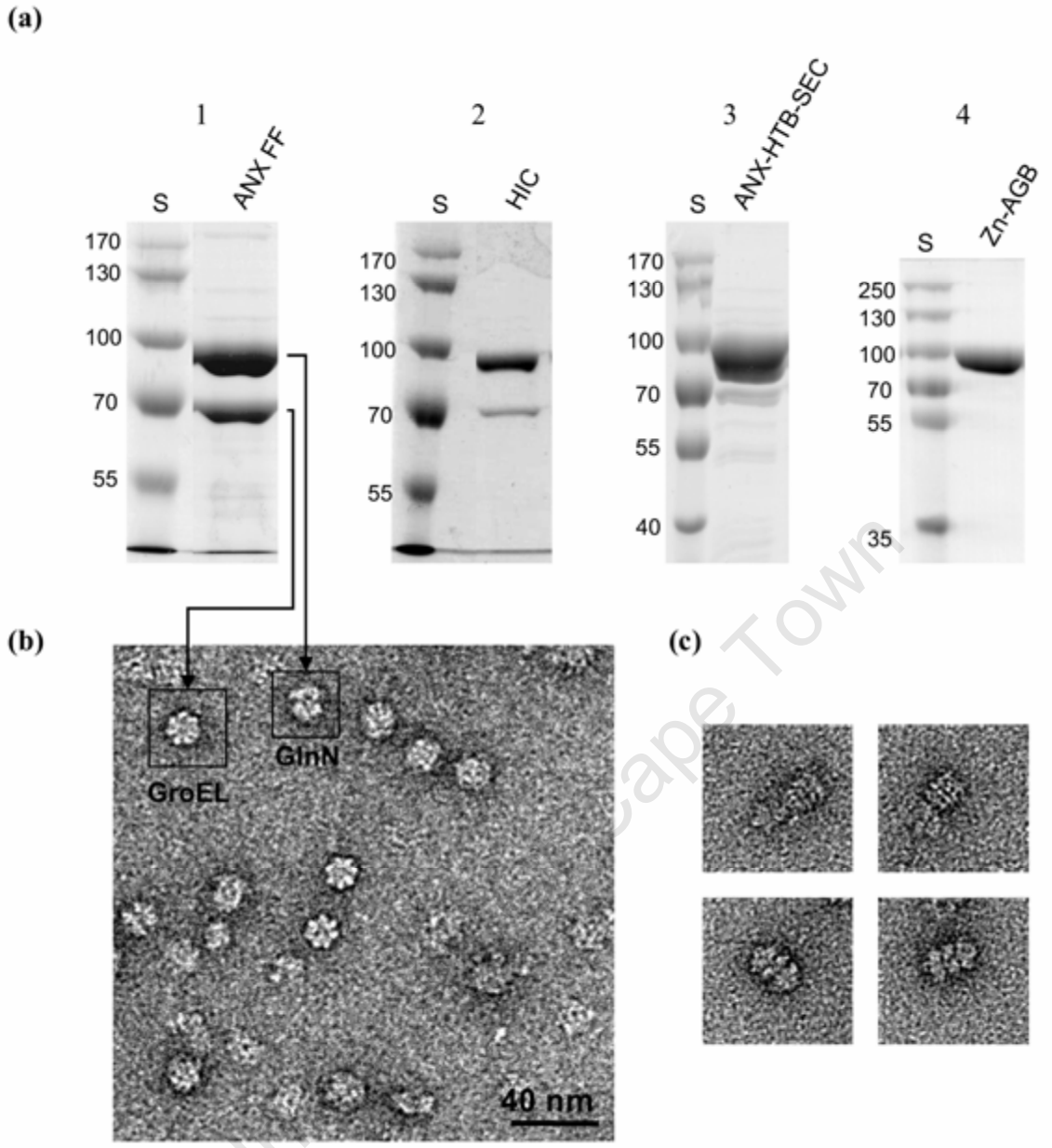


Figure 3: Removal of the primary protein contaminant, a putative GroEL complex, during the development of the purification procedure. **(a)** Coomassie Blue-stained SDS-PAGE gels (1-3: 7.5% and 4: 10%) comparing alternate strategies for the separation of GlnN from GroEL. The lanes labelled **S** contain molecular weight markers and their weights are given in kDa on the left. **(b)** Electron micrograph of negatively stained ANX anion exchange fractions. The GlnN and GroEL complexes are identified and linked to the corresponding bands in (a) by black arrows. **(c)** Close association of GroEL and GlnN complexes in the ANX-AGB-SEC fractions as seen with electron microscopy.

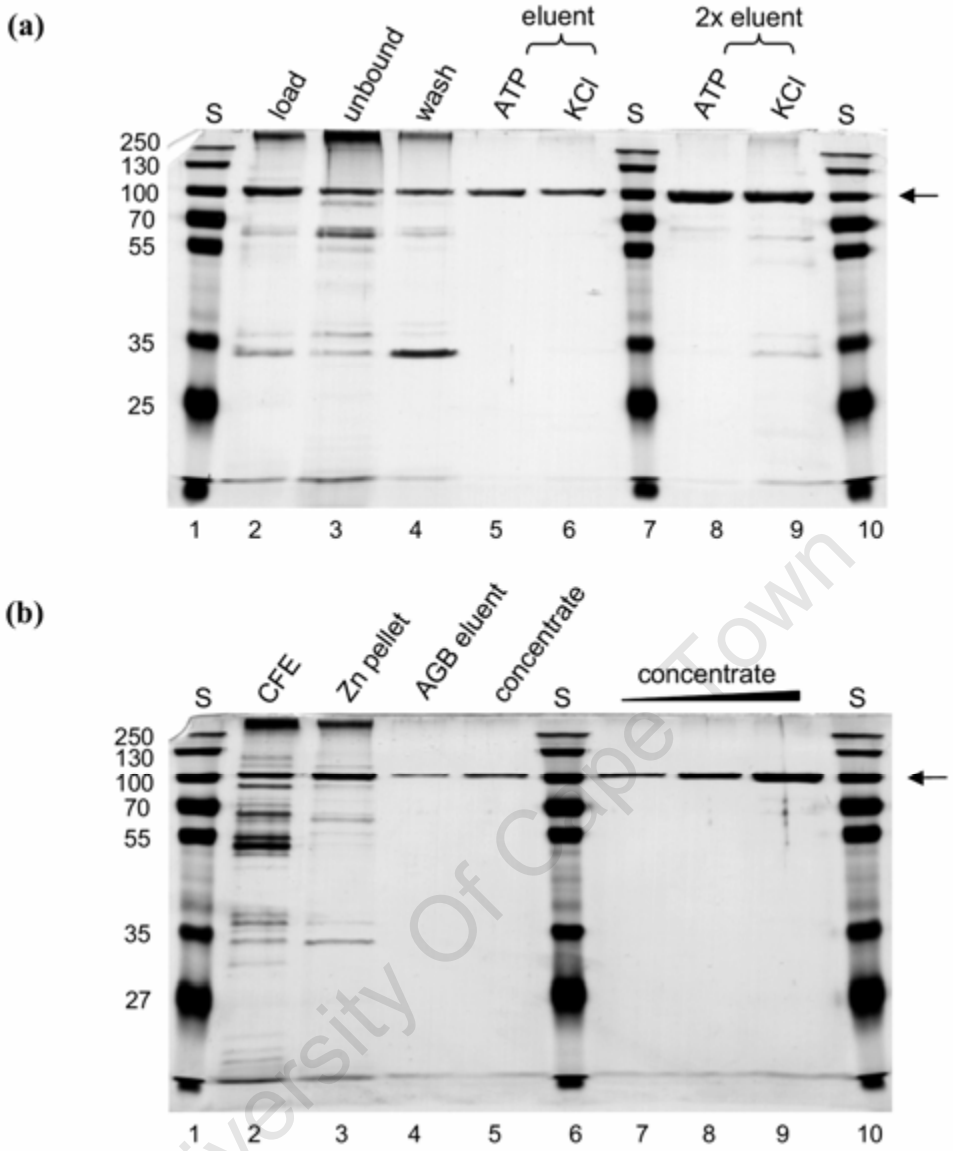


Figure 4: Design of the modified purification protocol. **(a)** Silver-stained 10% SDS-PAGE gel showing the improved purity of GlnN recovered from the Affi-Gel Blue affinity column by elution with ATP instead of KCl. Approximately 0.2 µg were loaded for each sample in the lanes 2-4 and 0.4 µg for lanes 8-9. **(b)** Silver-stained SDS-PAGE gel (10%) summarizing the final purification procedure (lanes 2-5) and assessing the final protein purity (lanes 7-9). Sample loadings in µg: lane 2 - 0.67, 3 - 0.33, 4 - 0.07, 5 - 0.07, 7 - 0.13, 8 - 0.26, and 9 - 0.56. Molecular weight markers were loaded into lanes labelled with **S** and their molecular weights are given in kDa on the left. The colour images of the silver stained gels were split into RGB channels and the grey levels of the blue channel was adjusted for presentation purposes but no bands in the gel image were masked by this alteration.

purity of the affinity fraction was also sufficiently high that the more selective 100 kDa MWCO filter could be substituted with a smaller 10 kDa one, thus, allowing much higher concentrations (6-8 mg/ml) for storage and crystallization to be achieved in less time. SEC of the thawed GlnN stock revealed a monodisperse population of GlnN dodecamers as suggested by the single peak with a molecular weight of 1.3 MDa but also a minor fraction of aggregates eluting at the column void volume (see Section 1.4.4). Samples were, therefore, clarified by centrifugation prior to crystallization.

Table 1: Purification statistics for production of recombinant GlnN from *E. coli* (YMC11).

Purification step	Volume (ml)	Total protein (mg)	Total enzyme activity (U)*	SA (U/mg)	Recovery (%)	Purity (%)‡	
						SA	SDS-PAGE
CFE	**44**	**241**	**129**	**0.534**	**100**	**4.9**	**6**
Zn^{2+} precipitation							
supernatent	44	65.4	2.96	0.0452	2.3		
wash	20	5.25	0.658	0.125	0.51		
AGB AF							
load	**50**	**38.7**	**29.2**	**0.754**	**23**	**6.9**	**63**
unbound	50	4.70	2.95	0.627	2.3		
wash	30	2.50	1.56	0.624	1.2		
eluent - ATP	**9**	**1.59**	**17.5**	**11.0**	**14**	**100**	**100**
eluent - KCl	8	6.47	1.74	0.269	1.4		
MetSox inhibited eluent †	9	1.18	N/A				
Ultra-filtration concentrate	**0.105**	**0.65**					

Values in the table represent the average of at least 2 measurements and all assays were carried out at dilutions within their linear response regions.

* Activity is expressed as μmoles of glutamyl hydroxamate formed per min.

† The pooled AGB ATP eluent was treated with MetSox before incubation o/n at 4°C.

‡% Purity was estimated from the calculated increase in SA during the purification and from densitometric estimations from 10% SDS-PAGE gels.

A summary of the final purification scheme is given in Table 1 and representative samples taken during the purification process are shown in Figure 4b. It can be seen from these results

that pure GlnN can be isolated quickly and easily with this method. The Zn^{2+} precipitation step alone is able to produce 63% pure GlnN at concentrations as high as 6 mg/ml (as measured by densitometry of all protein bands above 25 kDa).

The subsequent affinity purification removes all traces of contaminants. Unfortunately, the yields using this scheme are quite poor due to irreversible precipitation of GlnN. This occurs primarily at the Zn^{2+} precipitation stage (77% loss) and subsequently during AGB AF (additional 9% loss) and protein concentration (additional 8.5% loss). Similarly, it can also be seen in Figure 4a that the purity of the AGB eluted fraction was improved at the cost of recovery, as the fraction eluted with KCl contained a large proportion of GlnN. This disagrees with the results presented in Table 1 which show that the ATP eluted fraction contained 10 times more GlnN.

The final yield could not be measured because GlnN was inactivated by MetSox but significant additional losses were experienced during the final ultra-filtration step which increased the concentration to 6 mg/ml. If it is assumed that the MetSox fraction only contained GlnN, then the final recovery after losses due to concentration was only 5.5%. However, the final concentration was deemed to be sufficient for crystallization with PEG precipitants on the basis of preliminary crystallization pre-screens (see below).

1.4.3 Crystallization of full-length GlnN

Over the course of the investigation, the optimizations to the purification protocol produced several different preparations with varying purities, stabilities, and solute compositions. From the outset, attempts were made to crystallize these preparations and the successes of the optimizations were partly evaluated based on the feedback from the crystallization trials. These preliminary searches were mostly carried out with commercially available sparse matrix screens using the established microbatch [Chayen *et al.* 1990] and modified microbatch techniques [D'Arcy *et al.* 1996], which permitted rapid feedback concerning protein solubility and allowed easy crystal manipulations.

The commercial screens yielded a number of promising leads and these were elaborated upon using custom designed crystallization solutions to explore variables such as protein and precipitant concentrations systematically. For the most part, these "hits" manifested as

showers of micro-crystals which appeared to grow only after the initial precipitation of protein in the crystallization droplet. The PEG 3350/4000 precipitants, in particular, produced the largest number of crystallization phenomena and the most promising leads were followed in parallel in an attempt to improve the crystal growth conditions to allow positive identification by SDS-PAGE or the assessment of diffraction properties. Two forms in particular were routinely observed in crystallization trials setup with GlnN from different preparations and the most promising PEG 3350 conditions, namely, rectangular plates and 6-sided plates, both of which were < 30 µm in length and only about 1 µm thick. Although these micro-crystals proved too small to identify conclusively using SDS-PAGE, preliminary electron microscopy investigations suggested that they were composed of the large dodecameric GlnN complex (Figure 5). Many alternate strategies were subsequently attempted to try and reduce nucleation, thereby improving the size of the crystals, and these included alternate methodologies such as vapour diffusion, capillary counter diffusion, dialysis, and gel-permeation crystallization (data not shown). The most successful approach to increasing crystal size was the serial transfer of crystals into pre-equilibrated and clarified batch crystallization droplets. It was also discovered that 2-Methyl-2, 4-pentanediol (MPD) increased the solubility of GlnN in the presence of PEG and produced slightly thicker micro-crystals. Preliminary X-ray data (very weak and anisotropic diffraction to 7 Å) was collected from one large crystal which grew after several weeks following multiple macro-seeding transfers (data not shown). Molecular replacement searches with the cryo-EM volume (Chapter 2), however, failed to produce a solution for this data and the large size of the crystals could not be repeated. Such difficulties in improving the crystal growth characteristics were one of the primary motivators for the optimizations to the solubility and stability of the GlnN preparations.

Following the improvements to the purity, solubility and stability of the GlnN protein (described above), complete re-screening using a commercial matrix (Hampton Crystal Screen) yielded much improved crystallization conditions. As before, the initial screening results produced a large number of precipitation phenomena, particularly in the PEG conditions, but subsequent optimizations of the protein and precipitant concentrations produced irregular protein crystals after 4 days of growth for a diluted Hampton condition 20: 0.1 M ammonium sulphate, 50 mM sodium acetate trihydrate pH 4.6, and 12.5% PEG 4000. After mapping the meta-stable zone, streak-seeding produced large single regular monoclinic crystals (Figure 6a) and these were conclusively identified as GlnN by SDS-PAGE (Figure

6b). Diffraction quality crystals were routinely produced in this way, following several days of growth, although the quality of the diffraction proved to be quite variable (see Chapter 3).

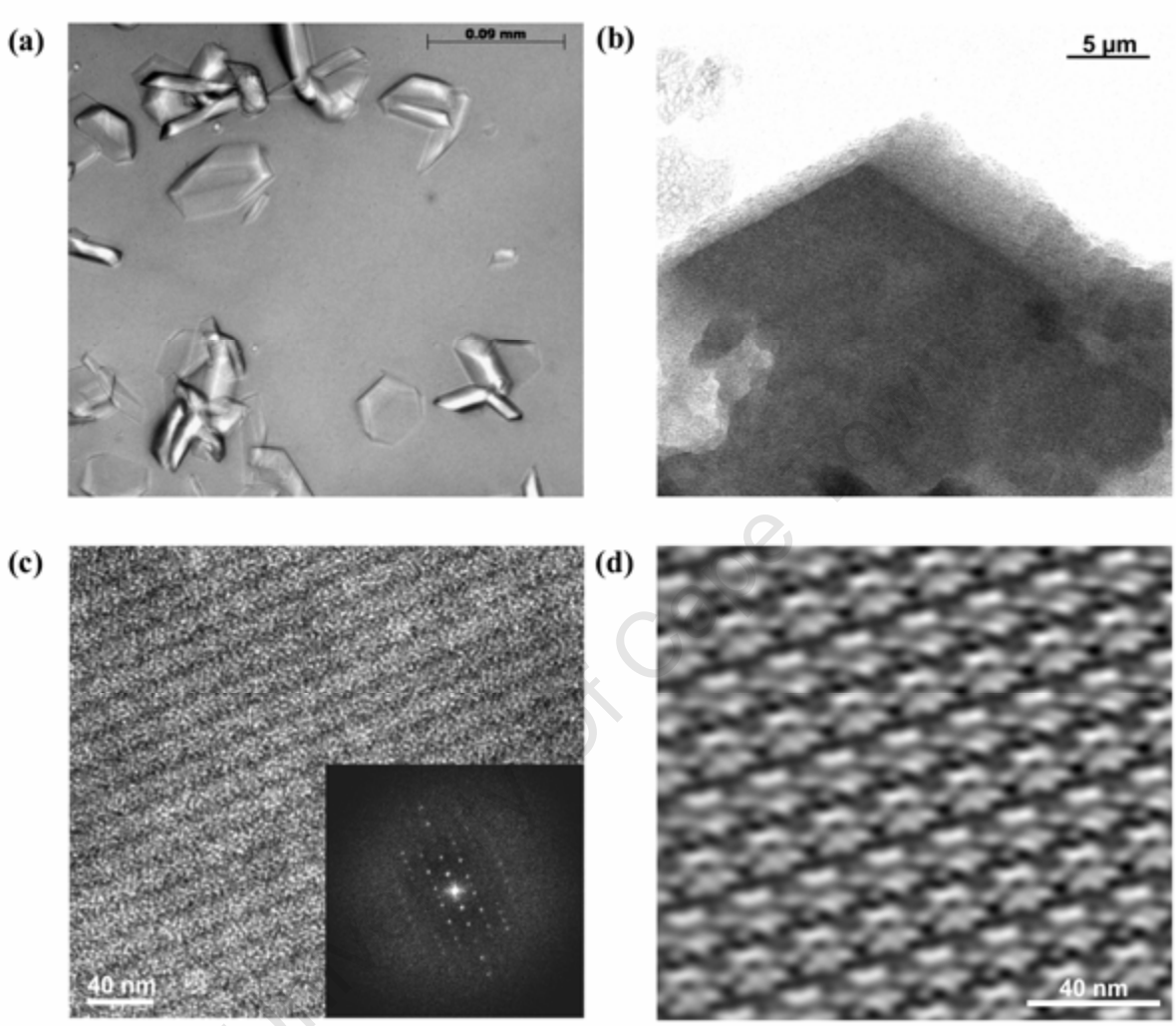


Figure 5: Microscopic investigations of GlnN micro-crystals. **(a)** Photograph of platelet micro-crystals grown using 10% PEG 4000 and 0.1 M $(NH_4)_2SO_4$ in tris pH 8.0. **(b)** Low magnification transmission electron micrograph of negatively stained micro-crystals grown in 12.5% PEG 3350 and 0.1 M $(NH_4)_2SO_4$ in bis-tris pH 5.5. **(c)** Higher magnification electron micrograph of the crystalline lattice from the crystals similar to those in (b) with the corresponding power spectrum shown in the inset. **(d)** Fourier filtered and correlation averaged image of the crystal lattice shown in (c). Grey levels were adjusted for presentation purposes.

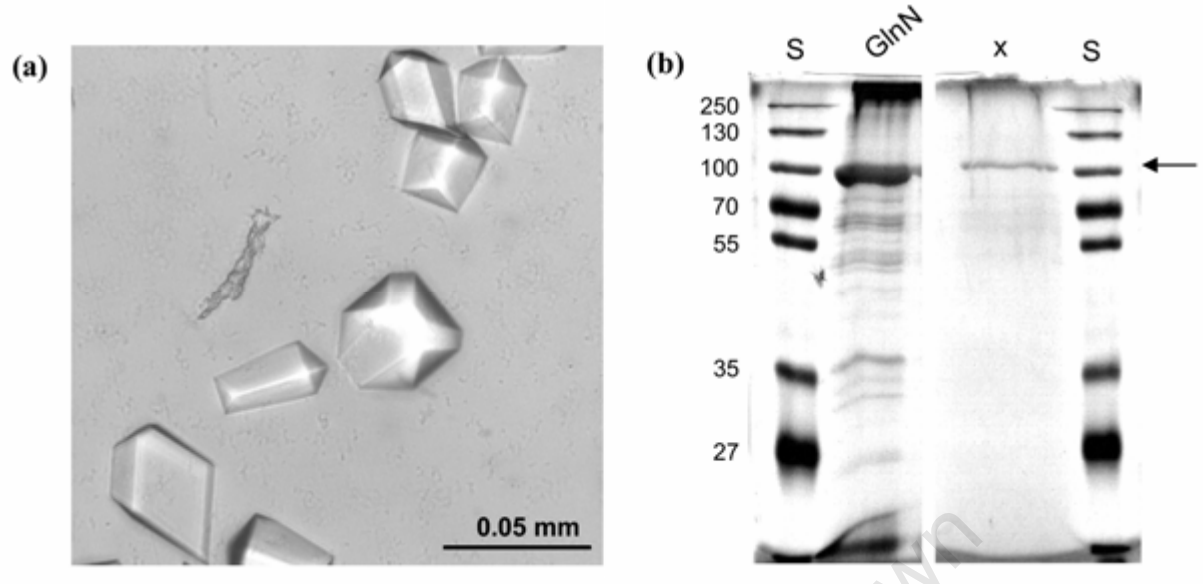


Figure 6: Full-length GlnN crystals. **(a)** Photograph of full-length GlnN crystals grown over several days by streak seeding. **(b)** Silver-stained SDS-PAGE gel (10%) comparing washed and dissolved crystals (**x –** 5 large crystals) with pure GlnN crystallization stock (**GlnN** – 1.9 µg). Molecular weight markers, **S**, were loaded into adjacent lanes and their weights are given in kDa on the left. The grey levels in image (a) were adjusted for presentation. The colour image of the silver stained gel in (b) was split into RGB channels and the blue channel was adjusted for contrast but no bands in the gel image were masked by this alteration.

1.4.4 Crystallization of digested GlnN

During the isolation and crystallization of GlnN it was serendipitously discovered that a bacterial contaminant was secreting a protease capable of cleaving GlnN into two peptides with masses around 48 and 35 kDa. This phenomenon was first identified in crystallization plates and following the isolation of the peptide products using SDS-PAGE, the identity of the larger fragment was confirmed as an N-terminal fragment of GlnN using peptide fingerprinting mass spectroscopy (Figure 7).

To verify the cleavage hypothesis and to test the sensitivity of GlnN to digestion, a trypsin dilution series was used to produce a profile of proteolytic fragments. As predicted, trypsin did produce the correct size fragments but the proteolysis progressed rapidly to completion

(data not shown). The bacterial contaminant, later identified as *Pseudomonas fluorescens* was, therefore, isolated and cultured in order to better characterize its more selective protease. The

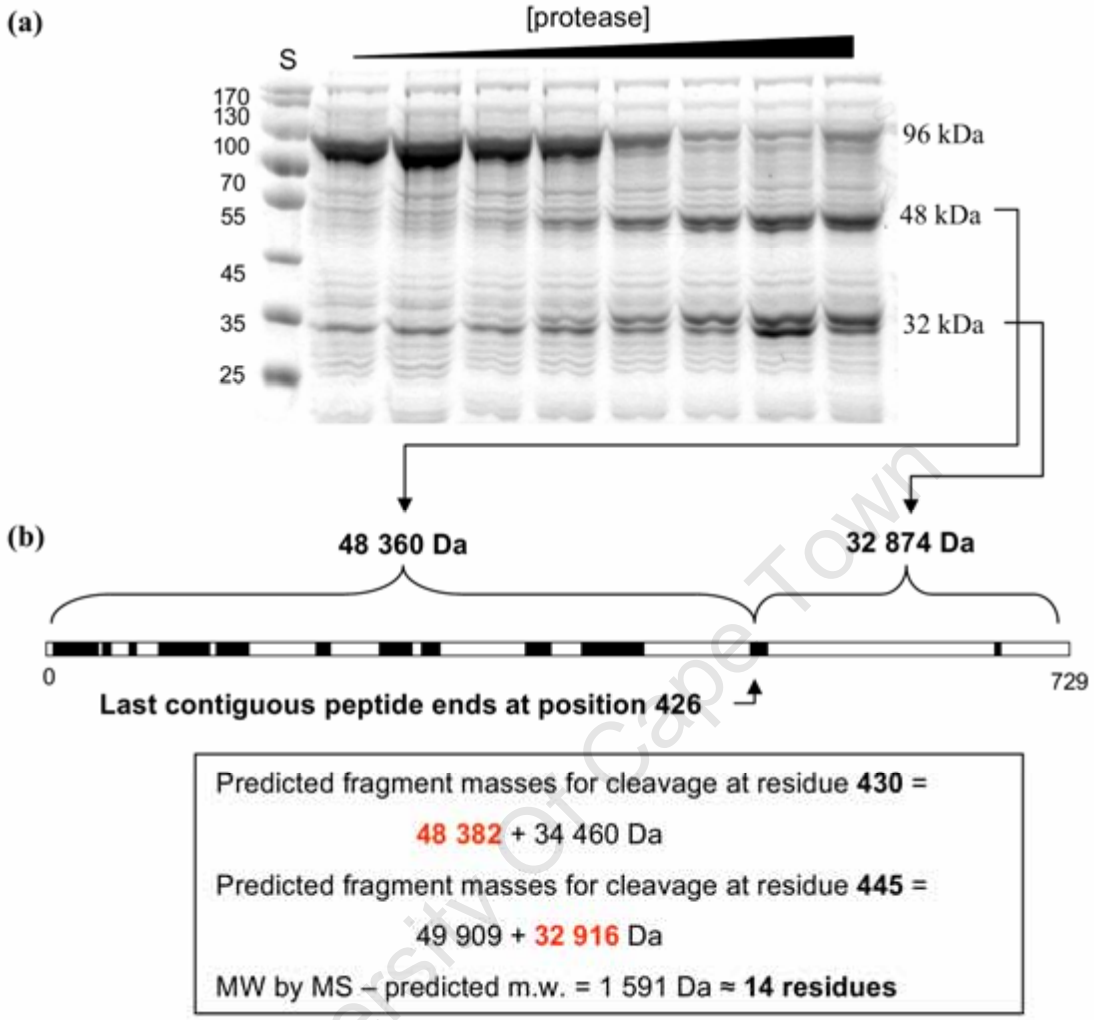


Figure 7: Non-denaturing proteolysis of GlnN by a secreted protease from *P. fluorescens*. **(a)** Coomassie Blue-stained SDS-PAGE gel (10%) of the proteolysis products (48 kDa and 32 kDa) resulting from the addition of different concentrations of *P. fluorescens* culture supernatant to semi-pure GlnN (96 kDa). Molecular weight standards were loaded into the lane marked **S** and their weights are given in kDa on the left. **(b)** A tryptic peptide coverage map from MALDI-TOF MS analysis (following in-gel digestion of the larger proteolytic fragment). The black areas represent tryptic peptide fragments which match predictions from *in silico* digestion of the GlnN sequence. The masses of the two GlnN *P. fluorescens* proteolytic fragments shown above the fragment map were accurately determined by MS without tryptic digestion. The calculation of the predicted cleavage points required to produce the two correctly sized fragments is given in the box below.

culture supernatant was tested for activity and the peptides produced exactly matched those seen in the crystal trials (Figure 7a). Estimates of the proteolytic cleavage site were made from the masses of the intact fragments derived from MADLI-TOF mass spectroscopy (32 874 and 48 360 Da) and the known GlnN sequence (Figure 7b). A cleavage of the GlnN protein at residue 430 would produce a peptide with a mass similar to that of the larger fragment but in order to account for the smaller peptide it must be assumed that a second cleavage occurs at residue 445 and a small 14 residue peptide is lost (Figure 7b).

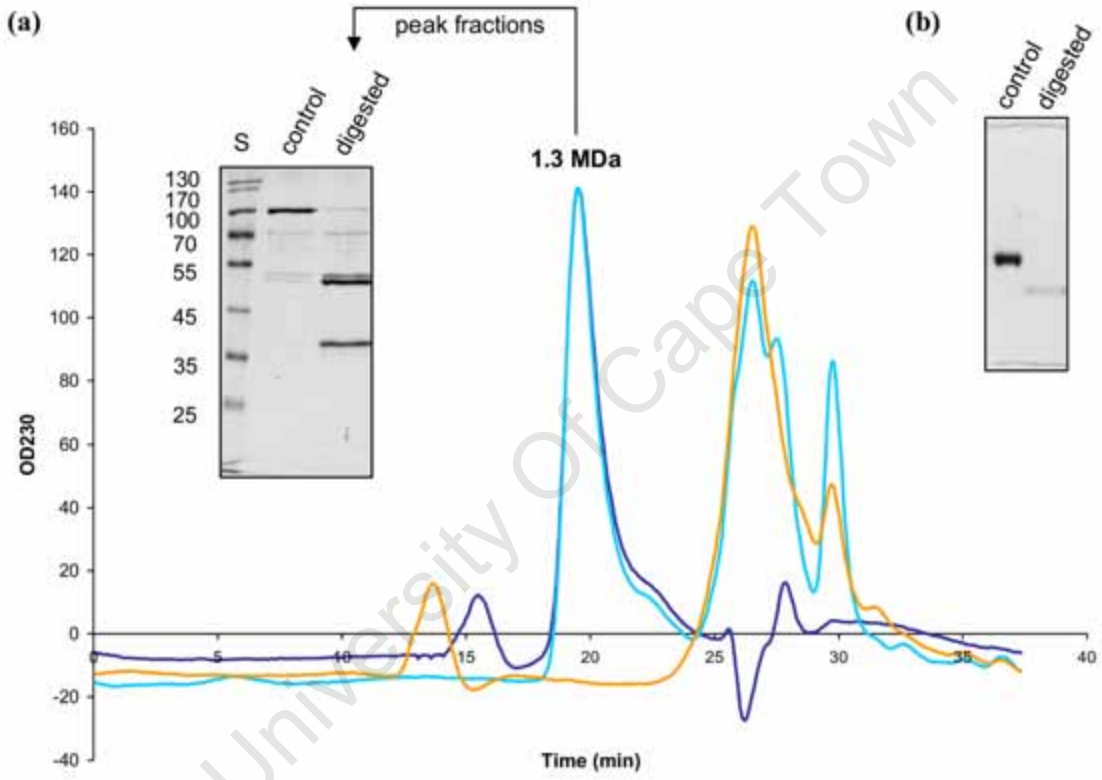


Figure 8: Characterization of higher order structure of the digested GlnN complex. **(a)** SEC chromatographs of digested (cyan) and undigested GlnN (dark blue). The *P. fluorescens* culture supernatant (orange) was analyzed as a reference. The peak fractions (15 µl) from both separations were analyzed by SDS-PAGE (**inset**) to verify the identity of the coincident peaks. Molecular weight markers are labeled with an **S** and their weights are given in kDa on the left. **(b)** Native PAGE analysis of the full-length and digested complex.

It was also shown that samples of GlnN treated with the protease did not progress to further digestion even after prolonged storage (2 weeks) at 4°C. SEC chromatography confirmed this by showing that the complex remained intact after digestion (Figure 8a) and the higher

resolution of native PAGE revealed a slight change in the mass-charge ratio (Figure 8b). The digested but intact complex also co-purified during Zn^{2+} precipitation and AGB AF (data not shown). Crystallization of the digested complex was, therefore, attempted in the hope that the modified GlnN complex might yield higher quality crystals. Crystal trials of digested GlnN were carried out using the same conditions that proved successful for the full-length enzyme and similar protein concentrations. Seedings using nuclei from the full-length crystals were able to produce large individual crystals displaying a clearly different bipyramidal habit after several days of growth (Figure 9). Subsequent streak seeded crystals grew larger and routinely displayed stronger diffraction than crystals of full-length GlnN (see Chapter 3).

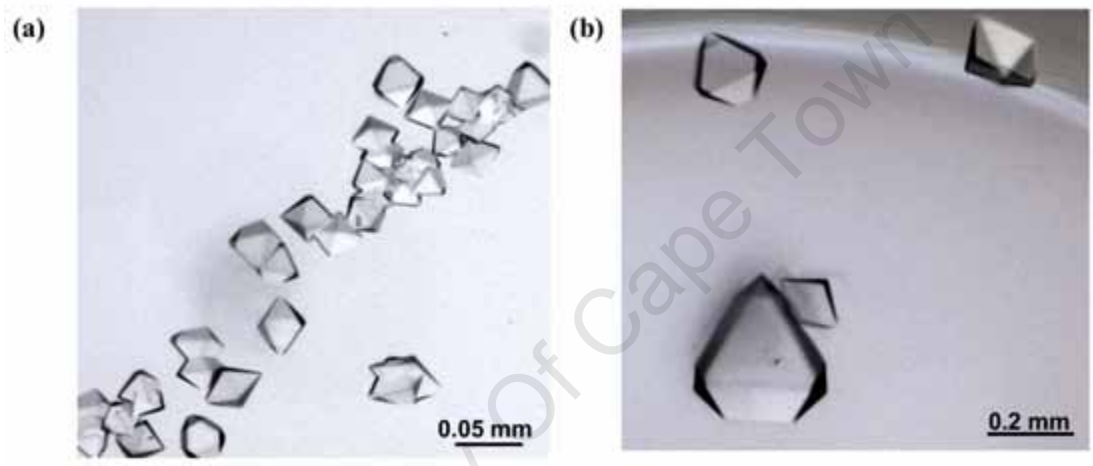


Figure 9: Crystals of digested GlnN. **(a)** Crystals grown in a few days by streak-seeding using identical conditions to the full-length protein. **(b)** Large crystals grown for diffraction analysis by macro-seeding.

1.5
Discussion

1.5.1 GlnN regulation

Previous isolations of the recombinant GlnN enzyme have cited the removal of an inhibitor during the purification process as the reason for an over recovery of enzymatic units [Southern *et al.* 1986]. Elliot [1953] also noticed such an inhibitory phenomenon during the isolation of a GSII from green peas. The current work was no exception and a tentative correlation was noted between the over-recovery of activity and the culture volumes. Southern *et al.* [1986] have previously shown that the recombinant enzyme is sensitive to feedback inhibition by several amino-acids and it is, therefore, possible that an accumulation of metabolic products is responsible for this inhibition. The ability to recover high levels of GGT activity following the first fractionation step further supports this hypothesis. The alternative hypothesis that the nitrogen supply is exhausted in smaller culture volumes and GlnN is then specifically inhibited by regulatory machinery in the *E. coli* expression host cannot explain the rapid reversibility and is, therefore, unlikely. Although the inhibitory phenomenon deserves closer scrutiny, the expression protocols presented in this work have made it possible to isolate material suitable for structural studies whilst allowing quantitative assessment of the purification steps.

1.5.2 GlnN solubility

The sensitivity of GlnN to precipitation by divalent cations is reminiscent of the phenomenon first identified in GSI [Valentine *et al.* 1968]. The mechanism of precipitation in GSI was later explained by Dabrowski *et al.* [1994] as the higher-order face-to-face stacking of GSI dodecamers which is driven by the formation of divalent cation binding sites between symmetry related N-terminal helices in adjacent complexes. These tubular structures then stack side-by-side to form large insoluble aggregates. Interestingly, the same N-terminal residues have been implicated in the formation of aggregates during the ‘salting-out’ of GSI under acidic conditions [Dabrowski *et al.* 1996]. The mechanism of precipitation by divalent cations in GSI does not appear to be conserved in GlnN because of the different pH dependence of the phenomena and the non-crystalline nature of the aggregates discovered in

this study. The possible conservation of this N-terminal motif in the GlnN structure and the implications for the mechanism of precipitation will be revisited in the light of the results of the crystal structure of GlnN presented in Chapter 3.

Although it is tempting to speculate that the precipitation phenomena experienced during this and previous GlnN isolations [van Rooyen 2004] was due to the presence of divalent cations, the continued losses in the absence of the ions suggest otherwise. The high concentrations of the salts $(NH_4)_2SO_4$ (0.8 M) and KCl (2.4 M) required to precipitate GlnN also indicate that at the levels normally present in the purification solutions they cannot be responsible. An alternate mechanism, however, can be suggested on the basis of the surprising results from the PEG solubility studies. These investigations revealed that in the presence of solubility limiting concentrations of PEG, GlnN, was most sensitive to precipitation by $MgCl_2$ followed by KCl and then $(NH_4)_2SO_4$. These results contradict the well established Hofmeister series, which ranks ions according to their ability to precipitate proteins [Hofmeister 1888], and suggest that under conditions of molecular crowding, induced by the high PEG concentrations, the mechanism of precipitation is due to the aggregation of denatured protein molecules. In line with this hypothesis, the more chaotropic salt, KCl reduces the solubility of GlnN substantially more than the cosmotropic $(NH_4)_2SO_4$ and indeed, Mg^{2+} also follows this trend by being both the most denaturing on the Hofmeister scale and the most solubility limiting. It, therefore, seems that both the spontaneous and divalent cation induced precipitation events are brought about by the aggregation of denatured GlnN molecules but in the latter case the denaturation is induced by the metal ions. This would explain the ability of pH values below the pI of acidic GlnN to prevent the aggregation phenomena by altering the affinity of the cations for the protein. Equally, the limited reversibility of the Zn^{2+} precipitations and the enhancements to the solubility of GlnN in the presence of glycerol and the MetSox inhibitor can be explained by this hypothesis. Glycerol is a common additive that is thought to improve the solubility of macromolecules by increasing their stability, thereby, preventing the aggregation that occurs through the exposure of hydrophobic surfaces in denatured proteins [Schein 1990]. Equally, GS inhibitors are known to stabilize the enzymes [Krajewski *et al.* 2008]

1.5.3 Purification protocol

The final purification protocol comprises a quick and effective capture step followed by a selective polishing without any intermediate concentration or buffer exchange manipulations. The combined protocol is able to rapidly produce highly pure GlnN with a total enrichment factor of 22-fold above the level found in the starting material. This high purity, however, comes at the cost of yield. Only 5.5% of the total GlnN present in the initial sample could be recovered due to losses during the Zn^{2+} precipitation (77% loss), AF (additional 9%), and concentration steps (additional 8.5%). The disagreement between the purity estimates from SDS-PAGE (10 fold increase) and from the change in specific activity (1.4 fold increase) suggests that a large proportion of the GlnN in the resuspended Zn^{2+} fraction was soluble but not active. The subsequent AF step, however, appears to remove the majority of inactive material with a resulting disproportionate increase in the specific activity. Nonetheless, the poor yields achieved with this protocol were tolerable considering the rapidity, ease, and scalability of the purification. As demonstrated here, despite only recovering 5.5% of the total enzyme units from a modest 44 ml of CFE, the protocol is able to produce enough pure GlnN enzyme for ~ 300 batch crystallizations (each requiring 1 µl @ 2 mg/ml).

1.5.4 GroEL contamination

The failure of the majority of the tested chromatographic and precipitation techniques to separate GlnN and the putative GroEL contaminant, together with the results from electron microscopy, suggests a specific association between the molecules. Similar associations have been noted before during the isolations of other GS enzymes. Burton *et al.* [1980] identified GroE as the primary contaminant during the isolation of GSI from *E. coli* and Tsuprun *et al.* [1992] carried out a detailed study of the interaction between a putative GroEL protein and the GSII from pea leaves using electron microscopy. It also appears that this problem is not specific to GS proteins and GroEL is widely reported to associate with recombinant proteins and in particular those containing affinity or solubilizing fusion domains [Thain *et al.* 1996]. As such constructs often display poor folding, it seems likely that these associations are driven by specific interactions of the chaperone GroEL with exposed hydrophobic regions of the recombinant proteins. Considering GlnN's hydrophobic surface character (demonstrated during HIC chromatography) and its tendency to aggregate, it seems more than likely that partially denatured GlnN is a substrate of GroEL.

1.5.5 Crystallization

Considering the numerous early crystallization phenomena, the sensitivity of GlnN to precipitation by low concentrations of PEG was initially anticipated as an advantage because of the limited supply of material from the original purification procedures. In retrospect, however, the limited solubility of GlnN was a major barrier to attaining suitable crystals for structural analysis. The growth of micro-crystals from precipitate and the lengthy macro-seeding equilibrations needed to enlarge them suggests that the biphastic nature of the early crystallization results were due to the initial precipitation of GlnN from solution and the subsequent concentration of the remaining soluble protein. The limited dimensions of the micro-crystals can also be explained by the resulting low protein concentrations and multiple nucleation sites. The subsequent application of AF and the inclusion of additives such as glycerol and GS inhibitors, however, appear to have acted in concert to decreasing the conformational heterogeneity of GlnN whilst increasing its solubility sufficiently to allow the growth of larger ordered crystals.

These successes cannot solely be attributed to changes in the sample as the final crystallization conditions also underwent optimization. The successful conditions (50 mM ammonium sulphate, 25 mM sodium acetate trihydrate pH 4.6, and 6.25% PEG 4000) were similar in composition to those that had produced the initial microcrystalline phenomena (0.2 M ammonium sulphate, 0.1 M tris-HCl pH 8.0, and > 8% PEG 4000) except for a change in pH from 8.0 to 4.6, which as shown earlier, is known to reduce the solubility of GlnN. The precipitant concentration also proved to be very important as no crystal growth was observed until the PEG concentration was reduced to 6.25%. Such a dilution is also predicted to have effected the pH of the final crystallization conditions as the concentration of 50 mM sodium acetate trihydrate pH 4.6 from the precipitant was changed relative to the 50 mM imidazole at pH 7.1 in the sample buffer. It, therefore, seems that the conditions which ultimately led to the crystallization of GlnN were achieved by a fine balance between solubility enhancing and limiting conditions, with the former reducing the conformational heterogeneity sufficiently to permit ordered aggregation under supersaturating conditions induced by the latter.

1.5.6 Proteolysis

Proteolytic treatment under non-denaturing conditions has become a standard strategy for modifying crystallization behaviour [Newman 2006]. By removing the intrinsically disordered surface loops that contribute significantly towards conformational heterogeneity the likelihood of a protein crystallizing can be increased. The serendipitous discovery of the proteolytic processing of GlnN by an extracelluar protease from *P. fluorescens* presented the opportunity to crystallize such a modified form of the GlnN protein. The effects of digestion on the complex were subtle enough, as demonstrated by the results of the SEC, to still yield crystals under the same conditions as the full-length protein but altered the molecular packing sufficiently to produce a new crystal form with higher symmetry, improved growth characteristics, and stronger diffraction. It is also well known that proteolytic sensitive sites are generally solvent exposed, extended loops, and that their mobility is the primary determinant of protease susceptibility [reviewed by Hubbard 1998]. It, therefore, seems likely that the predicted cleavage site near residue 430 shares these characteristics.

1.5.7 Conclusions

The work presented in this chapter has led to the production of GlnN for structural characterization by cryo-EM (Chapter 2) and X-ray crystallography (Chapter 3). The latter achievement was accomplished through the design of a rapid protocol for the production of sufficient quantities of structurally homogenous and stable GlnN protein.

Chapter 2

Single Particle Cryo-Electron Microscopy Reconstructions of GlnN

2.1
Summary

Over the last four decades, electron microscopy has made a significant contribution to our understanding of the higher order structure of GS enzymes and it has been revealed that in general they are all large double-ringed structures composed of subunits arranged with dihedral point group symmetry.

This chapter describes the reconstruction of frozen-hydrated GlnN using SPEM image processing techniques. The primary aim of which, was to determine the true quaternary structure of the complex in solution and to extend the current understanding of GSIII structure. Reconstructions of GlnN auto-inhibited by MetSox and ATP were also performed to determine whether any gross structural changes occur in the complex upon ligand binding, in line with the suggestion of allostery in other GS enzymes.

The SPEM reconstruction of the apo-GlnN enzyme (16 Å) displayed clear D6 symmetry and provided novel insights into the nature of the interactions between the subunits in opposing rings, including previously unseen internal density lining the central solvent filled cavity. Comparisons of the cryo-EM volume with the previously solved GSI structure reiterated the surprising similarity of the active site arrangements in the two structures and identified which regions were the least conserved: the domed caps at either end of the complex, the tip of the 'pinwheel' arms, and the inter-ring interfaces. No large scale domain motions could be detected upon ligand binding in the MetSox-inhibited reconstruction, which was solved to the same resolution as the apo-GlnN reconstruction. Analyses and simulations suggested that particle heterogeneity was most likely responsible for the limited resolution of the reconstructions.

This work has provided the highest resolution structural information available at the time for the divergent GSIII enzymes and in so doing has not only revealed the true quaternary structure of the GlnN complex in solution and showed that no large scale domain motions accompany auto-inhibition by MetSox but also provided essential information for the solution of the higher resolution GlnN crystal structure presented in Chapter 3.

2.2
Introduction

"The Role of Electron Microscopy and Image Processing in the Structure Determination of GS Enzymes"

Since the first application of EM to studying macromolecules, it was clear that the technique had the potential to provide detailed information about the quaternary structure of proteins [reviewed Haschemeyer *et al.* 1974]. The large size and symmetry of the GS enzymes has made this superfamily a prime subject for structural analysis by EM and these efforts have paralleled developments in image processing and 3D reconstruction. GS has often been chosen as suitable sample for developmental image processing studies because of the ease of interpretation of its projections, which results from its regular structure and preferential attachment to the carbon substrate.

2.2.1 GSI family

Prior to the application of EM, the only information about the quaternary structure of GS enzymes came from biochemical studies such as ligand titration experiments and biophysical studies such as SEC, ultra-centrifugation, and electrophoretic separations. These results were often conflicting and difficult to interpret. For example, at the resolution of density-gradient ultra-centrifugation results, the GSI from *E. coli* was thought to be composed of 12-14 identical subunits but a precise prediction was not possible [Woolfolk *et al.* 1966]. The early investigations also revealed that GSI could exist in two distinct states: the native 'taut' enzyme and a 'relaxed' enzyme produced by treatment of the former complex with EDTA at raised pH (~8) [Woolfolk *et al.* 1966; Woolfolk *et al.* 1967b]. These two states were characterized by differing levels of activity, metal ion occupancy, and hydrodynamic volumes. The first application of electron microscopy to the study of a GS structure sought to confirm the predicted quaternary structure and to discern the conformational changes that accompany the inter conversion of the relaxed and taut states [Valentine *et al.* 1968].

Direct visualization of negatively stained GSI particles from *E. coli* confirmed the dodecameric composition of the enzyme and for the first time revealed the arrangement of these subunits in two symmetrically opposing and eclipsed hexagonal rings (point group symmetry 622 i.e. dihedral symmetry D6) [Valentine *et al.* 1968]. Each ring was suggested to be formed by heterologous interactions between adjacent subunits and the two hexameric rings were proposed to associate via isologous interactions. It was noted by these authors that these isologous interactions, form a closed oligomeric arrangement, and prevent the GS molecules from stacking into long polymers, which would occur with C6 symmetry. This early investigation used optical filtering to average distinct molecular views, thus, reducing noise and enhancing structural features sufficiently to measure the molecular dimensions and describe the subunit shape. These findings were found to agree with the previous hydrodynamic data [Woolfolk *et al.* 1966] but did not reveal any distinction between the molecular arrangement of the 'taut' and 'relaxed' forms, despite sedimentation data to the contrary [Ginsburg 1972]. Nonetheless, the direct visualization of the 'tightened' form did explain the precipitation of GSI by divalent cations by revealing the unexpected face-to-face stacking of GS molecules and the large-paracrystalline aggregates formed by further lateral associations of these tubes. Another important finding from this early study was the fact that isologous contacts between subunits within a ring were weaker than the heterologous contacts between rings, as revealed by the dissociation kinetics of the relaxed enzyme complex. It was also shown that oligomers could be re-associated by the addition of Mn^{2+} but in some instances the recovery was only transient as visibly less-defined oligomers soon broke-down and aggregated. Haschemeyer *et al.* [1982] later went on to visualize the dissociation intermediates of partially inhibited GSI directly and reached the same conclusion. Importantly, these oligomers, even partially saturated with MetSox inhibitor, all showed catalytic activity. The bonding energies involved in oligomerization would later be quantified according to changes in solvation energy on oligomerization [Almassy *et al.* 1986].

The discovery of ordered aggregates of GSI opened the door for higher resolution structural studies using 3D helical reconstruction techniques which were just becoming available [De Rosier *et al.* 1968], but only applied to GS structure seven years later. This technique involves the reconstruction of the object from the optical diffraction pattern, derived from electron micrographs, using Fourier-Bessel techniques. Frey *et al.* [1975] described the first 3D low resolution reconstruction of *E. coli* GSI from electron micrographs of negatively stained molecules arranged in 7-stranded cables which were formed by treating the relaxed enzyme

with Co^{2+}. The resulting three-dimensional reconstruction benefited from the improved signal-to-noise ratio achievable with crystalline samples and had a resolution of 30-50 Å. These efforts revealed the GSI subunit arrangement in higher detail and showed that each dodecamer was related to its neighbour by a rotation of 7.5°.

With the introduction of the single particle electron microscopy method (SPEM), higher resolution reconstructions were no longer constrained to periodic arrays. This approach reduces the noise component of the signal emanating from sources such as the structural variations in the particles and staining, and also the contribution from electron shot noise, by deriving structural information from a large number of individual particles through computer averaging. In one of the earliest implementations of SPEM, Frank *et al.* [1978] were able to reconstruct, to a low resolution, the structure of GS molecules dispersed in negative stain. As with all SPEM studies it was assumed that all of these molecules interacted with the support film in an identical manner and, therefore, only a 2D top-view was reconstructed. This effort used ten times as many particle images (50) as in previous studies, 300 in total if 6-fold averaging was counted, and resulted in several novel insights. The boundaries of sub-domains within the GSI subunits were identified and the reconstruction was found to possess distinct handedness, which was attributed to a non-uniformity of staining.

Another development that significantly advanced the SPEM techniques was the introduction of low-dose data collection practices. By using very low doses to record images of the particles this technique minimizes the very significant noise component arising from structural variation due to radiation damage. The lower doses, however, necessitate larger data sets to recover the lower signal. There have been two examples of EM investigations of GSI using the low-dose technique. These studies demonstrated the usefulness of the low-dose and associated image processing techniques for extracting higher resolution information, and in so doing, revealed further details about the GSI structure. The first study by Kessel *et al.* [1980] utilized correlation averaging to reveal the subdivision of the GSI subunits into two stain-excluding domains. This reconstruction was performed from 45 images which were further 6-fold symmeterized, thus, increasing number of asymmetric units to 270. The 2D reconstructions also displayed clear handedness and the previously suggested cause of partial staining was confirmed by tilt experiments. The later study by Kunath *et al* [1984] visualized the GSI from *S. typhimurium* and produced the highest resolution structural data at that stage (better than 20 Å) by averaging 32 particles with 6-fold symmetrization, filtering, and finally

averaging of five 2D reconstructions. At the time of publication the authors noted similarities between their reconstruction and preliminary 10 Å data from X-ray investigations of GS crystals by Almassy *et al.* [1986]. In both of the low-dose experiments, the segregation of the GS subunit into two distinct stain-excluding domains agreed with previous biochemical evidence from limited proteolysis experiments which produced two similarly proportioned cleavage products [Lei *et al.* 1979]. However, this interpretation was shown to be false after the crystal structure of GSI became available [Almassy *et al.* 1986] (Chapter 3 – Section 3.1).

Members of the GlnT family have also been investigated by EM and the results have tended to confirm the close phylogentic relationship (20% sequence identity) identified between the subgroup and the remaining GSI enzymes. The GS from *Rhizobium lupine*, with a molecular mass of 600 kDa, was revealed by negative stain EM and correlation averaging to also be a large dodecameric complex (13-14 nm diameter & 9 nm height) composed of two hexagonal rings arranged with D6 symmetry, distinct left-handedness, and a bi-lobed sub-domain organisation [Tsuprun, *et al.* 1987]. More recently, a truncated form of the lengsin protein from the mouse eye lens has been solved to low-resolution using cryo-EM [Wyatt *et al.* 2006]. The lengsins are an ancient family which resemble the GlnT proteins and they appear to have been recruited for non-metabolic functions in the lens tissues of higher animals. This discovery also represents the first detection of a GSI-like protein in a eukaryote. The low-resolution volume was obtained from 3000 frozen-hydrated particles using a multivariate statistical analysis [van Heel *et al.* 1996] and angular reconstitution [van Heel 1987] approach to assign particle views to classes and to determine their respective orientations for back-projection in three-dimensions. Although the reconstruction displayed similar overall quaternary structure features to the GlnT and GSI reconstructions, significant remodelling of the individual subunit's tertiary structure was necessary to interpret the volume in terms of homology models based on the GSI crystal structure. The largest difference between the lengsin and GSI structures was found in the region connecting the two hexameric rings. In lengsin, the interconnections were found to be less extensive and situated more towards the outer circumference of the interface than in GSI.

2.2.2 GSII family

Much contention concerning quaternary structure occurs in the GSII family. Not only have the results of biochemical techniques such as SEC, centrifugation, and gel electrophoresis

often been ambiguous due to experimental and theoretical difficulties associated with the techniques [Gouaux *et al.* 1994; Haschemeyer *et al.* 1974] but the GSII family also appears to occur in a range of oligomeric states [reviewed by Eisenberg *et al.* 2000]. Due to this apparent instability and the additional difficulties associated with partial staining and the interpretation of the resulting projections, the determination of the GSII structure has not been as straightforward as GSI and several different models for the quaternary structure of this family have been proposed without consensus (see below). This has occurred despite the first investigations preceding those of the GSI family.

In addition, because the GSII family is predominantly found in higher eukaryotes, the numerous tissue and compartment specific isozymes and their splice variants make generalizations about quaternary structure difficult. This problem is exemplified by the plant GS enzymes which have been the subject of considerable scrutiny by electron microscopy. These GSII enzymes exist in different sub-cellular compartments and occur in many different isoforms with different sizes, stabilities, and catalytic potentials. For instance, whereas the plastid GSII is encoded by a single gene, the cytosolic enzymes are encoded by multiple genes and there are at least five of these in the in *Zea mays* [Sakakibara *et al.* 1996], three in *Phaseolus vulgaris* [Bennett *et al.* 1989], four in pea plants [Pushkin *et al.* 1985], and two in *Medicago truncatula* [Carvalho *et al.* 1997].

The first GSII structure to be revealed by EM studies was the eclipsed cubic octamer visualized by Haschemeyer *et al.* [1966] during investigates of the enzyme from ovine brain. Similar interpretations have also been reported for GSII enzymes isolated from a diverse assortment of tissues: soya bean root nodule cytosol [McParland *et al.* 1976], green algae [Rasulov *et al.* 1986], pea seedling [Pushkin *et al.* 1985], rat liver [Tate *et al.* 1972], chicken neural retina [Sakar *et al.* 1972], hamster liver [Tiemeier *et al.* 1972], and pig brain [Wilk *et al.* 1969]. In comparison, the studies of GSII enzymes from higher plants, their seeds, and prokaryotes have uncovered an eclipsed octameric quaternary structure [Tsuprun *et al.* 1980; Tsuprun *et al.* 1987]. Thus, despite the disparate results it appears from these early microscopy studies that all GSIIs from both eukaryotes and prokaryotes occur as octamers with dihedral four-fold symmetry. However, two alternate conformations have been proposed: an eclipsed and a partially eclipsed octamer. The segregation of these structures into these differing arrangements does not appear to follow species lines.

The proposed cubic octamer arrangement for the GSII family has important implications for the functioning and regulation of the enzymes according to the 'half-site hypothesis' [Eisenberg *et al.* 2000]. Because the active sites in GS enzymes are formed between the N and C-terminal domains of adjacent subunits, the maximum number of functional sites can only be formed in closed ring structures. Tetrameric rings, as proposed by the octameric model, are, therefore, predicted to have only half the number of possible actives sites (four) because of the acute angle formed by the adjacent subunits. The results of different ligand binding experiments both confirm and disprove this hypothesis [Tate *et al.* 1972; Wilk *et al.* 1969] and results supporting more than four active sites have been explained by supposed domain flexibility.

In stark contrast to the model held over the previous thirty years [Haschemeyer *et al.* 1974; Boksha *et al.* 2002], the first applications of modern 3D SPEM techniques to the structure of the GSII from human brain have proposed a tetradecameric arrangement of subunits [Kiang 2001]. A population of dissociation intermediates was also observed: tetradecamers, planar heptamers, and completely dissociated monomers. The reconstructed tetradecamer was composed of two eclipsed heptameric rings and, thus, possessed D7 symmetry. Preliminary investigations into the quaternary structure of alfalfa GS using EM and crystallography appear to have confirmed this tetradecameric arrangement of subunits in GSII [Pfluegl *et al.* 2002].

The most recent and comprehensive 3D SPEM reconstruction of a GSII to date (5 578 particles), however, has suggested another configuration. In line with the early investigations and the results of SEC and native PAGE data, Llorca *et al.* [2006] solved the low resolution structure of the GSII from *P. vulgaris* revealing an octameric quaternary structure. The arrangement of the tetrameric rings in the 100 Å cubic structure, however, was found to be offset by 90° and the final reconstruction, therefore, only possessed two-fold symmetry.

Finally, the recent determination of several GSII crystal structures at atomic resolution (Section 3.1) has produced conflicting evidence for the quaternary structure of enzymes from this family. Enzymes from humans, dogs, and maize leaves were all crystallized as decamers built from partially eclipsed and opposing pentameric rings. It, therefore, appears that the various GSII enzymes either exist in different oligomeric states or incorrect interpretations, resulting from partial staining artefacts and the instability of the complexes, are to blame for the lack of consensus.

2.2.3 GSIII family

Preliminary EM [Southern *et al.* 1986] and stoichiometric data from biochemical studies, including calibrated gel filtration and native PAGE analyses, have previously identified *B. fragilis* GlnN as a hexamer [Southern *et al.* 1987]. The GSIII from *Synechocystis* sp PCC 6803 [Garcia-Dominguez *et al.* 1997] is also thought to be hexameric, whereas, the enzyme from *R. albus* [Amaya *et al.* 2005] is thought to be dodecameric. Both of the latter determinations were made on the basis of calibrated SEC studies.

The most comprehensive analysis of the quaternary structure of a GSIII was carried out recently by van Rooyen [2004]. Negative stain EM and projection matching image processing techniques were used to reconstruct the 3D structure of recombinant *B. fragilis* GlnN from ~ 15 000 particles views (Figure 1). In contrast to the above mentioned observations, which identified GlnN as a hexamer, this work revealed a dodecameric structure, with subunits arranged in two opposing hexagonal rings. However, the opposing rings differed in diameter and this asymmetry was thought to arise from partial staining artifacts due to the large size of the GlnN complex and the limited stain depth. It was also discovered during the reconstruction process that the GlnN particles adopted a preferred orientation on the carbon support film, with the majority of particles attaching via the largest flat region of the molecule.

The handedness of the reconstruction was not determined but a strong agreement between the overall arrangements of subunits in the GlnN and GSI complexes was noted despite a large size difference between the two (16.5 vs 11.0 nm length). Based on the assumption of a common handedness for the GlnN reconstruction, “pinwheel” views of negatively stained particles were interpreted as arising from hexameric dissociation intermediates of the dodecameric GlnN complex (Figure 1 – inset). The coincidence in the positions of the indentations between the arms in GlnN hexameric rings and the GSI active sites was taken as evidence for the conservation of the β-barrel active site fold and heterologous intra-ring contacts in GlnN.

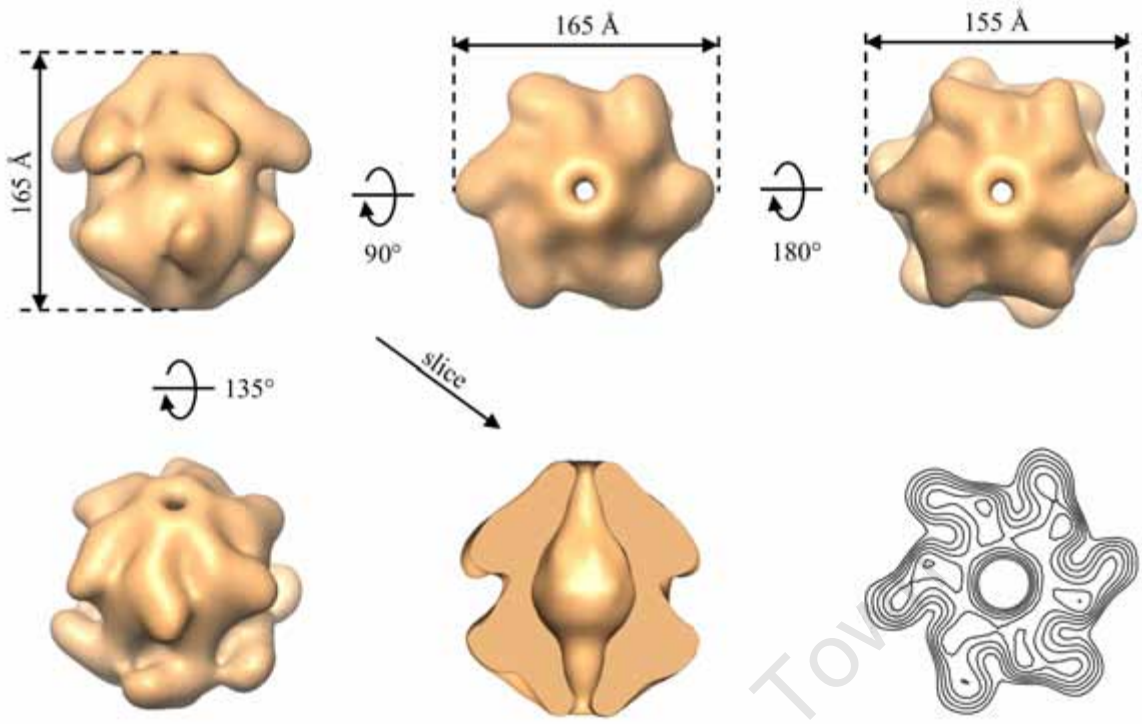


Figure 1: Molecular envelope of GlnN derived by negative stain SPEM [van Rooyen 2004]. The reconstructed volume is shown as a surface representation with the density contoured to enclose the volume corresponding to 12 subunits - 1 360 nm^3. The handedness of the reconstruction is represented here to match the GSI structure. Inset – 2D projection of averaged and 6-fold symmeterized views (150) of the hexameric particle observed in negative stain EM [van Rooyen 2004]. Density contours are from -3.24 - 2.48 in steps of 0.95.

2.2.4 Summary

It is clear that EM and image processing techniques have played a crucial role in elucidating the structure of the GS enzymes. Although the insights derived from X-ray crystallography studies have exceeded those from EM due their higher resolution, such atomic resolution structures only exist for the GSI and GSII families and EM has, therefore, provided the first structural information for a number of enzymes. All GS structures solved to date are double-ringed structures with dihedral point group symmetry.

2.3
Materials and Methods

2.3.1 Data collection

Apo-GlnN

Recombinant *B. fragilis* GlnN was isolated from a GS deficient *E. coli* expression host and purified as described in Section 1.3.2. Stock material was stored at a concentration of ~1 mg/ml in a low-salt buffer (10 mM imidazole-HCl, 10mM $MnCl_2$, pH 7.15) at 4°C. A 25% dilution of this solution (3.5 µl) was applied to Quantifoil® R2/2 300 holy carbon grids held within a humidified chamber (~80% RH). These grids were previously glow discharged in air for 20 seconds at 20 µA. After 10 seconds incubation, the sample was blotted for 5 seconds with filter paper before being rapidly plunged into liquid ethane cryogen using the custom built gravity driven plunging apparatus. The humidity levels, adsorption, and blotting times, required for optimal particle concentration and the ice thickness were optimized empirically.

Electron micrographs (53) of frozen-hydrated GlnN were taken on a Tecnai F20 (by Assoc. Prof. Trevor Sewell during a visit to Birkbeck college, London) operating at 200 kV under standard conditions of minimum-dose at a magnification of 50 000 times. During the focussing stage of this procedure, astigmatism and defocus were corrected by inspection of the real-time power spectra calculated from the images. The micrographs were digitized using an Ilford Leafscan and converted to MRC format before being downsized by a factor of 2 using the IMAGE2000 [Crowther *et al.* 1996] program, LABEL, to average adjacent pixels giving a final sampling of 4 Å per pixel at the specimen level.

MetSox-inhibited GlnN

Purified GlnN, prepared as described in Section 1.3.2, was inhibited with the ligands MetSox and ATP before collection of a complete data set under cryogenic conditions. Specifically, pure GlnN was buffer exchanged and concentrated by ultra-filtration to 0.33 mg/ml in 50 mM imidazole, 20 mM $MgCl_2$, and 50 mM KCl pH 7.15. MetSox and ATP were then added to a final concentration of 5 mM and the preparation was incubated at 37°C for 30 minutes.

Activity assays (Section 1.3.3) confirmed that inhibition took place and lasted the duration of the experiment.

The preparation of the vitrified samples on glow-discharged Quantifoil grids® was identical to the method used in the initial apo-GlnN cryo-EM data collection (see above). Data (77 micrographs) were collected on a Tecnai F20 (Laboratory of Molecular Biology, Cambridge, England) operating at 200 kV under standard conditions of minimum-dose imaging at a magnification of 50 000 times. The micrographs were digitized using a Zeiss SCAi and converted to MRC format before being downsized by a factor of 2 using the IMAGE2000 program [Crowther *et al.* 1996], LABEL giving a final sampling of 2.96 Å per pixel at the specimen level.

2.3.2 Image processing and reconstruction

The majority of image processing steps and all reconstruction steps were accomplished using the program, SPIDER [Frank *et al.* 1996]. Average periodograms were calculated from the micrographs, using standard techniques and scripts provided in the latter program, in order to screen for astigmatism, drift, and incorrect defocus ranges. These scripts were modified to only include power spectrum patches that had a variance less than the average value for all the patches in the final periodogram, thus, excluding regions containing ice particles and the micrograph borders and labels.

Initial apo-GlnN cryo-EM reconstruction

GlnN particles (15 688) were interactively picked from micrographs displayed in the MRC program, Ximdisp [Smith 1999]. SPIDER was then used to window these particles, placing them in 90×90 image boxes which were twice the particle diameter. CTF correction in the form of phase-flipping was applied to the 16 defocus groups based on the CTF parameters determined using the MRC program CTFFIND2 [Crowther *et al.* 1996]. Images then were band-pass filtered according to standard single particle procedures using an inverse top-hat high-pass filter function (1/200 Å) and a Gaussian low-pass function (1/8 Å). Images were normalized to zero average density and a standard deviation of one after masking with a soft circular mask. Images displaying dust or inconsistent grey levels were detected from visual inspection of all particles in WEB [Frank *et al.* 1996] and deleted at this point.

All 3D reconstructions were obtained using an iterative projection matching reconstruction strategy [Penczek *et al.* 1992] (Figure 3 below) consisting of the following main steps:

(1) **Generation of reference views**. Two independent 3D reconstructions were calculated from different start models: (a) The first was an *ab initio* model derived by simultaneous minimization based on common lines: particle images (660) from the highest defocus group (3.47 µm underfocus) were centered by reference free alignment [Penczek *et al.* 1992], and classified by rotationally invariant K-means clustering [Penczek *et al.* 1996], with class membership refined iteratively using multi-reference alignment [Joyeux, 2002]. The angular relationships between the "best" 20 class averages were determined by a simultaneous minimization program based on common lines [Penczek *et al.* 1996] and a 3D model was produced by back-projection in Fourier space. (b) The previously determined negative stain reconstruction served as the second starting model [van Rooyen 2004].

Following the generation of the reference volumes, templates (46) were derived by projection of model volumes in quasi-evenly spaced directions (15°) covering 3D space and preprocessed as described above. This sampling of the projection directions was chosen to adequately represent the theoretically attainable resolution but the number of template was increased in subsequent optimizations (see below). Additionally, 'trap class' templates i.e. characteristic views of contaminating particles (putative GroEL) were used to exclude these particles from the reconstruction.

(2) **Classification by alignment**. Particles were aligned against the reference templates using an angular refinement approach [Penczek *et al.* 1992]. During the alignment procedure, a threshold cross-correlation score was imposed to exclude 30% of the worst aligning particles from the reconstruction and empty classes were replaced by their template images.

(3) **3D reconstruction**. Back-projection of class averages by interpolation in Fourier space, was used to reconstruct the volumes and D6 symmetry was enforced during the reconstructions. In the final reconstruction of the model (see below) the template

projection directions were limited to the asymmetric unit only and the alignment algorithm employed a mirror-check to speed up processing.

In subsequent rounds of the reconstruction process the second and third steps were iterated, with the volume from the previous round serving as the new reference, until the algorithm converged (as judged by the average cross-correlation score between all images and the templates derived from the reference volume).

After convergence of the final reconstruction, the structures were aligned and the Fourier shell correlation (FSC) between the volumes was calculated [van Heel *et al.* 1986]. The quoted resolutions in this thesis correspond to a FSC value of 0.5 [Saxton *et al.* 1986; van Heel *et al.* 1986]. The final reconstruction was then low-pass filtered to the given resolution using a Fermi low-pass filter with a temperature factor of 0.025. No B-factor correction was applied, but the higher frequencies were boosted relative to the lower frequencies by applying a Gaussian high-pass filter with a fall-off of 50 Å.

The cryo-EM reconstruction of dodecameric GlnN has been submitted to the EM databank [Henrick *et al.* 2003; http://www.ebi.ac.uk/pdbe/emdb/] under the entry code: EMD-1204.

MetSox-inhibited reconstruction

Particles (12 304) were automatically picked from the (54) micrographs remaining after CTF screening using the program SIGNATURE [Chen *et al.* 2007] and windowed (140 x 140 pixels) in SPIDER [Frank *et al.* 1996] before pre-processing as described above. A total average of several manually picked particles served as the reference for the automatic picking algorithm. CTF correction in the form of phase-flipping was applied in SPIDER after estimating the CTF parameters with CTFFIND2 [Crowther *et al.* 1996] and dividing the data into ten defocus groups.

The iterative projection matching algorithm was identical to the initial apo-GlnN reconstruction algorithm except that 120 templates were generated from the first cryo-EM model, which corresponded to an angular spacing of 5° in the asymmetric unit of the constrained D6 symmetry. A CC score threshold was applied, retaining 80% of the data set during the reference based alignment.

2.3.3 Alternate reconstruction algorithm

A completely different reconstruction approach utilizing angular reconstitution instead of projection matching was also attempted in IMAGIC [van Heel *et al.* 1996]. Briefly, the initial 15 Å cryo-EM model was band-pass filtered from the Nyquist limit to 40 Å. The volume was then masked in 3D with a binary mask volume created by several rounds of Gaussian filtering, binarization, dilation, and finally Gaussian smoothing again. Reference projections (19) were then generated from this volume at evenly spaced divisions (15°) of the D6 asymmetric unit. Multi-reference alignment was then carried out against these references and the best 20% of the aligned images were selected based on their CC score. Multivariate statistical analysis was then carried out and these images were classified into 50 classes on the basis of all 69 eigen-factors. Poor class-sums were then deleted before angular reconstitution and back-projection was used to generate a start model after anchor set refinement of the original reference projections. This model then served as the reference for the next round of anchor set refinement after re-projection. This procedure was iterated until an acceptable agreement between the projections of the model and the class-sums was achieved. After convergence, the number of reference images was increased to 70 (7°) and the 19 (15°) anchor set templates were regenerated and the reconstruction procedure iterated.

2.3.4 Angular coverage assessment and simulations

Tilted data

A series of micrographs were collected at a 20° tilt, on the Tecnai F20 (now installed locally at the University of Cape Town), to provide additional angular information for the 3D reconstruction. A 1/10 dilution of inhibited GlnN at 2 mg/ml in 0.2 M citrate/Na_2HPO_4 pH 5.8 buffer was investigated under cryogenic conditions. Two pairs of micrographs were collected, under conditions of minimum-dose, for each ice covered hole in the Quantifoil® grids. The first was collected at -2.5 µm defocus and 20° tilt and the second was collected at -3.0 µm (to account for the 0.5 µm offset between the two tilts) and 0° tilt. The relative intensity derived ice thickness of all the collected micrographs was 0.02-0.03. Micrographs were digitized on an Ilford Leafscan scanner before being downsized by a factor of 2 to give a final sampling of 4 Å/pixel at the specimen level (as described above). Averaging of adjacent pixels and bandpass filtering between 200 Å and 16 Å were carried out in IMAGEJ [Abramoff *et al.*

2004]. Particles were picked automatically from these modified micrographs using SIGNATURE [Chen *et al.* 2007] with a total average of several manually picked particles serving as the reference template. Windowed (60 × 60) images were then normalized in SIGNATURE. In total, 14 175 and 17 128 particles were extracted from the tilted (24) and untilted (14) micrographs respectively. No CTF correction was applied to these data but the templates were filtered to 16 Å (the first zero of the CTF was at 25 Å). A single round of projection matching was then used to determine the orientations of the images from the tilted and untilted data. CC cut-off scores were adjusted to reject 15% of the particles.

Resolution limitation simulations

The *S. typhimurium* GSI crystal structure (pdbid:*2gls*) [Yamashita *et al.* 1989] was imported into SPIDER at a sampling of 4 Å/pixel and projections (27) were then made at evenly spaced increments (10°) within the asymmetric unit of the D6 symmetry group. A series of volumes were then reconstructed by back-projection in Fourier space using different subsets of these projections (with D6 symmetry imposed). The resolution of the resulting volumes was assessed by calculating the FSC between themselves and the imported 4 Å/pixel *2gls* volume.

2.3.5 Data visualization

Reconstructed volumes were visualized using UCSF CHIMERA [Pettersen *et al.* 2004] and all renderings of volumes for presentation in this these were performed in this program unless otherwise stated. The threshold for contouring the density for presentation was decided based upon the volume enclosed by the surface, as calculated in CHIMERA. This value (1 364 nm^3) was determined from the predicted molecular weight of GSIII (82.7 kDa x 12) and an average value for protein density of 0.73 $Da/Å^3$ [Quillin *et al.* 2000]. All particle images were visualized in WEB [Frank *et al.* 1996].

2.4
Results

2.4.1 Structure solution

Data collection and reconstruction

Initial inspection of the grids containing frozen-hydrated GlnN, prepared under optimized conditions, revealed a homogenous population of particles with an average diameter of 185 Å and features which matched the previously determined negative stain GlnN structure (Figure 2). The statistics for the data collected on the FEG equipped microscope are shown in Table 1.

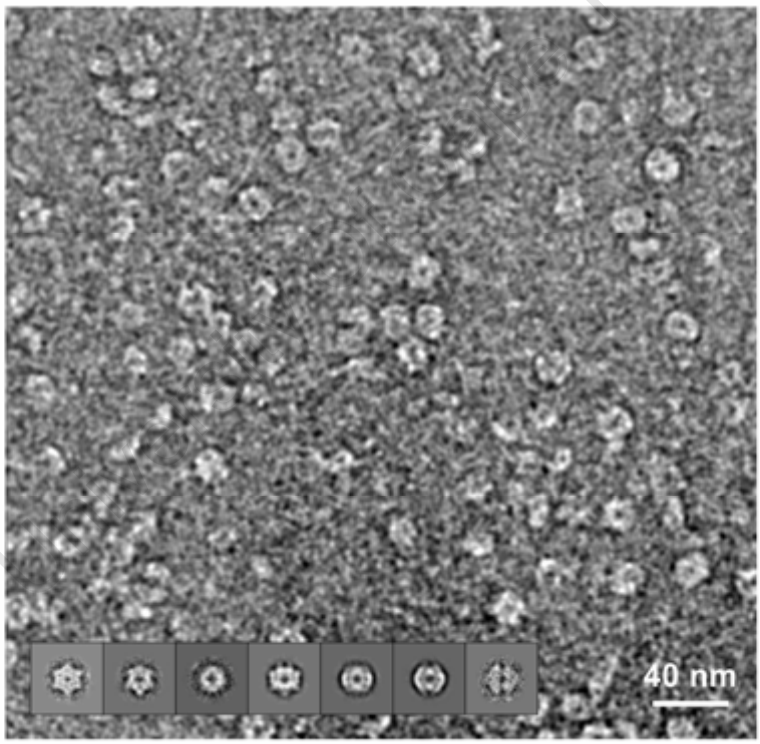


Figure 2: Frozen-hydrated GlnN single particles (negative contrast). This representative sampling of the GlnN cryo dataset comes from the highest defocus micrograph (-3.4 µm). The image has been reduced from the original sampling of 4 Å/pixel, low-pass filtered, and background-subtracted prior to grey level adjustment for presentation purposes. **Inset** – representative class averages generated by projection matching during the reconstruction process.

Table 1: Data collection statistics for the first GlnN cryo-EM reconstruction.

Defocus Group	Average Defocus (µm)	First 0 (Å)	No. Micrographs	No. Particles
1	1.01	16	1	131
2	1.32	18	2	90
3	1.46	18	2	135
4	1.67	20	2	728
5	1.83	20	5	1359
6	1.96	21	3	597
7	2.10	22	4	985
8	2.23	23	3	847
9	2.35	23	3	1024
10	2.46	24	3	1271
11	2.59	24	9	2975
12	2.71	25	3	1261
13	2.94	26	3	852
14	3.12	27	5	1154
15	3.18	28	3	1667
16	3.44	28	2	660

Angular refinement based reconstructions, starting from two independent reference structures, converged to similar volumes (as judged by the FSC shown below) (Figure 3). One starting model (a) was derived from the previously determined negative stain model [van Rooyen 2004] and the other (b) was derived *ab inito* by classification and common-lines based simultaneous minimization. The reason for calculating two independent reconstructions from different models was to avoid the problem of extension of resolution due to the correlation of noise components resulting from alignments against a common reference [Grigorieff 2000]. Symmetry was enforced during these reconstructions after it was determined that in the absence of any constraints the reconstruction from the *ab initio* reference converged to a volume displaying clear D6 symmetry, as determined by self-rotation function calculations (Figure 4). In addition, multi-variate statistical analysis of centred but not rotationally aligned particles produced eigen-images with clear 4-fold symmetry, in agreement with the presence of dihedral symmetry axes (see below).

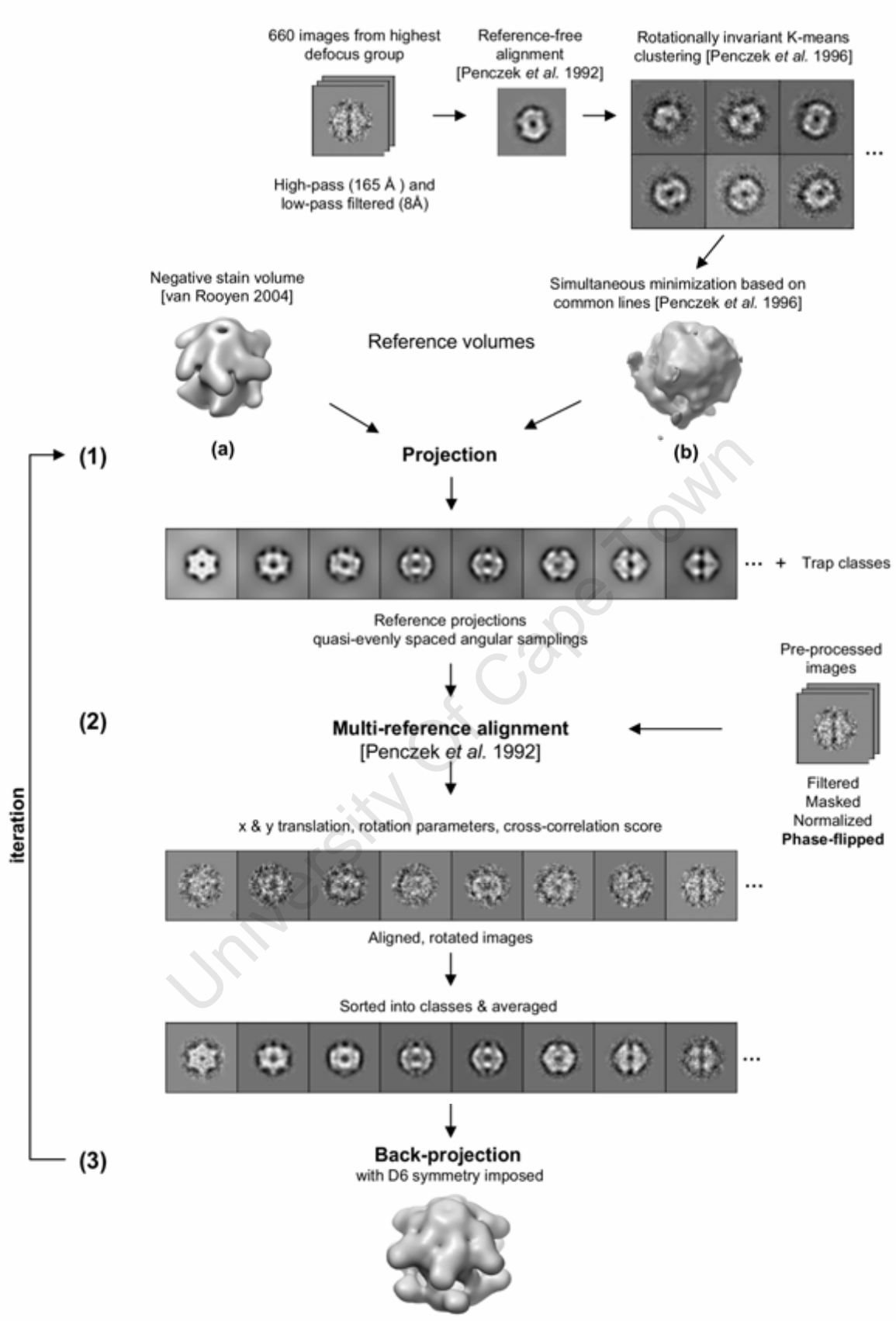


Figure 3: Summary of the iterative projection-matching SPEM reconstruction process. Details are given in Section 2.3.2.

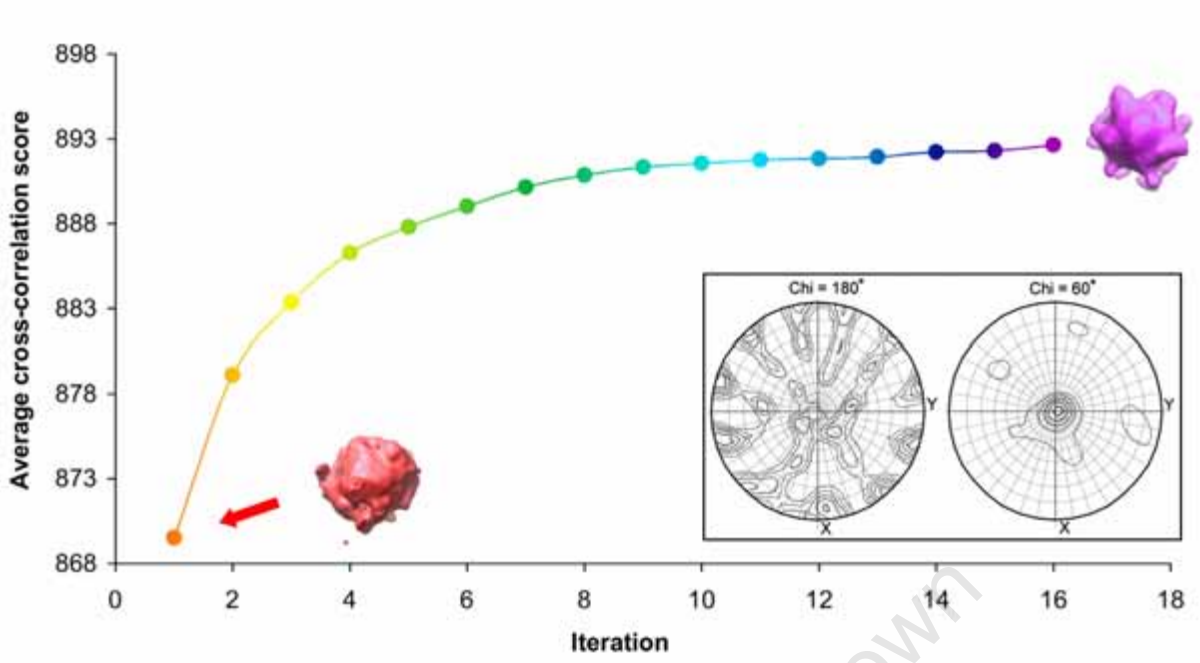


Figure 4: Convergence of the reconstruction from an *ab intio* volume (red) to a structure displaying clear D6 symmetry (purple). Progress of the reconstruction was monitored by the average CC score between the images and reference projections. **Inset** - Self-rotation function (theta, phi, chi) calculated from the unsymmetrized cryo-EM reconstruction (purple) in MOLREP [Vagin *et al.* 1997]. The model was orientated with the putative 6-fold axis along phi=0, theta=0; theta is plotted as the radius and phi is plotted as the azimuthal angle of the polar plot.

Resolution and validity of the reconstruction

Fourier shell correlation of the two converged volumes (from the two different starting models) showed that they were equivalent to a resolution of 16.4 Å as measured with the conservative 0.5 FSC measure (Figure 5a) [Saxton *et al.* 1986; van Heel *et al.* 1986]. The self consistency of the reconstruction was also evident from a comparison of the characteristic views, generated by multi-reference alignment and averaging, with re-projections of the model (Figure 5b).

To avoid the possibility of circularity, a statistical classification based reconstruction strategy was also implemented (Figure 5c) but this algorithm converged to the same volume obtained with the projection matching approach, thus, further confirming the validity of the reconstruction.

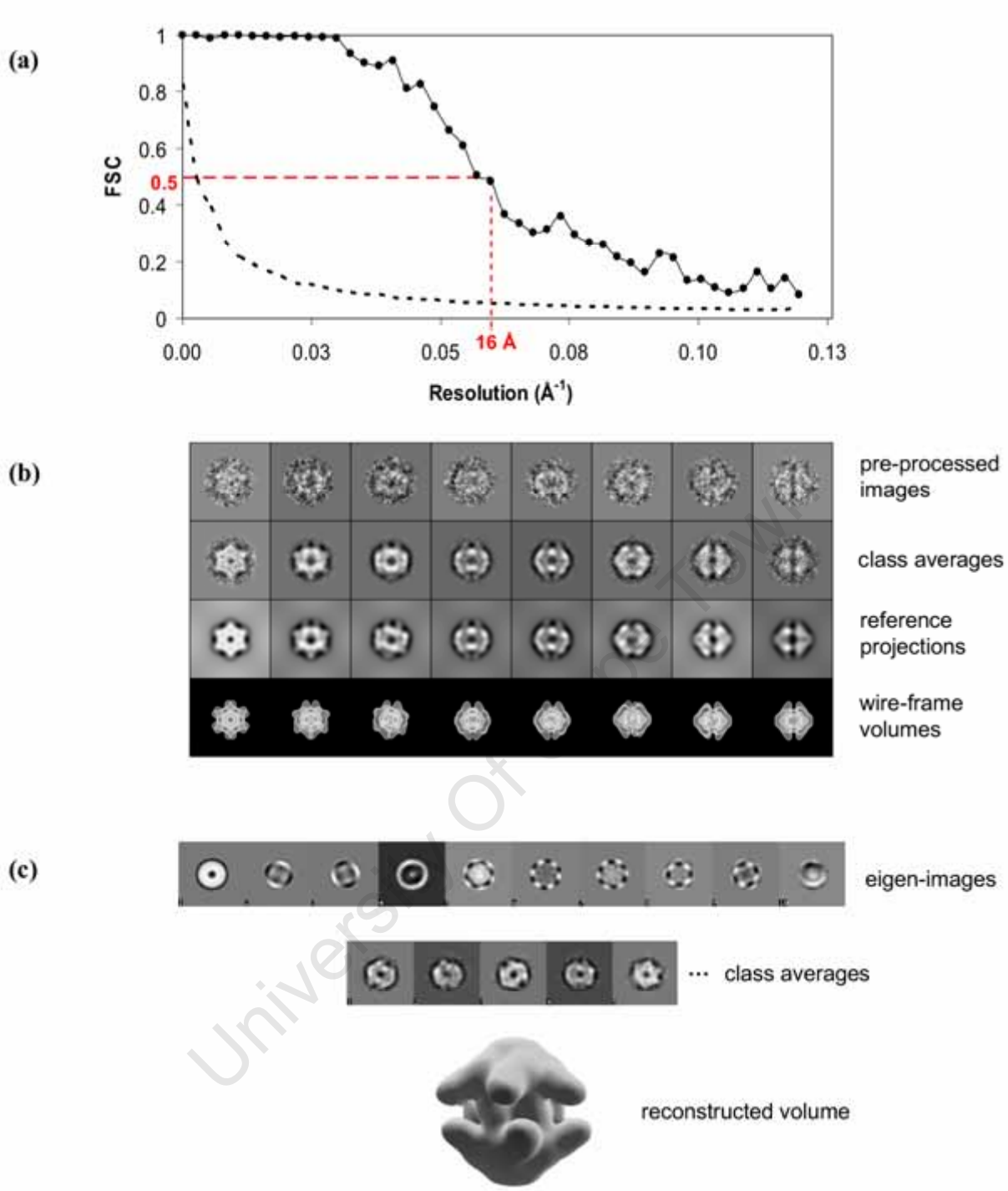


Figure 5: Validity of the cryo-EM reconstruction. **(a)** Fourier shell correlation of reconstructions derived from different starting models using the same image dataset. Black dotted line = 3σ noise fall-off. **(b)** Self-consistency of the reconstruction. **(c)** Alternate multi-variate statistical analysis and angular reconstitution based reconstruction algorithm. The final converged volume is shown alongside representative class averages and the first 10 eigen-images used to classify the data set. Details are given in Section 2.3.3.

2.4.2 Structural insights

The results of the cryo-EM reconstruction revealed a similar structure to the previously determined negative stain model [van Rooyen 2004] but the subunits, the active site folds, and the connecting density between the rings was better defined in the higher resolution volume (Figure 6). The latter interface is formed by columns of density that are connected in an equatorial direction by pairs of stalk-like processes which form a double ring when viewed from the inside of the reconstruction volume (Figure 6 – cut-away view). This ring forms the walls of the central cavity and at lower contour levels these regions are continuous with additional rings of density at either end of the cavity.

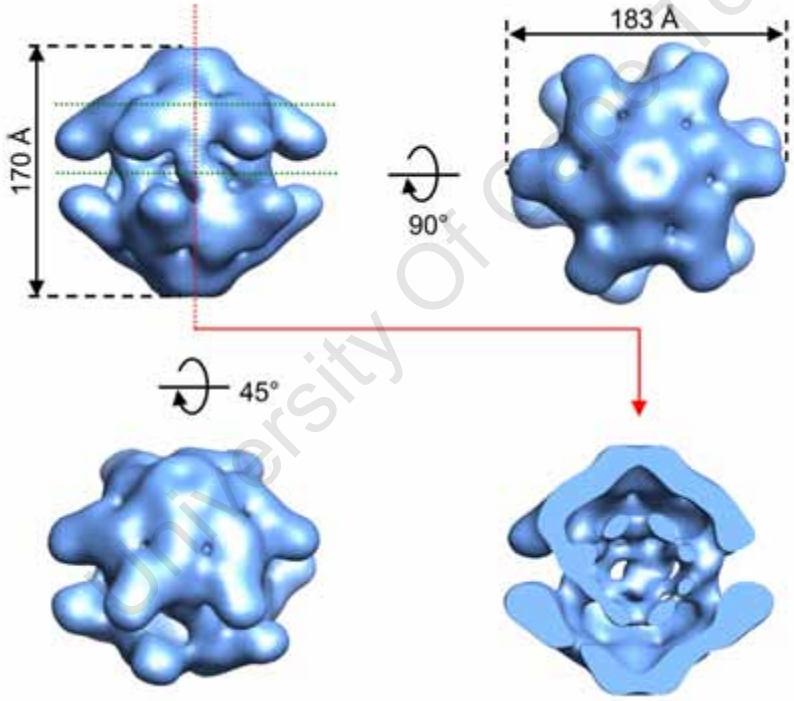


Figure 6: Low-resolution structure of GlnN derived by cryo-SPEM. The final low-pass filtered volume is shown as a surface representation with the density contoured to enclose the volume corresponding to 12 subunits - 1 360 nm^3. A transverse cut-away view (red arrow) is included to show the internal density of the cryo-EM reconstruction.

The GlnN structure was considerably larger than the *S. typhimurium* GSI dodecamer (Figure 7a): 65% longer in the six-fold axial direction (17.0 vs. 10.3 nm) and on average 24% wider (17.8 vs. 14.3 nm). The primary regions responsible for the differences are the large domed

caps at either end of six-fold axis of the molecule, the longer “pinwheel” arms, and the inter-ring connecting density. The latter connections are located towards the circumference of the hollow GlnN structure in comparison to the shorter and more extensive inter-ring connections in GSI (Figure 7a – cut-away view). Although the handedness of the GlnN reconstruction was unknown, a comparison of the superimposed structures, shown with the same handedness, confirmed the previously noted remarkable coincidence of the active site β-barrel folds of GSI and the indentations between the pinwheel arms in the GlnN volume (Figure 7b - left). The correspondence between these features is independent of the choice of hand for the GlnN reconstruction but the agreement for the remaining areas making up the ‘pinwheel’ arms is much poorer if GlnN is chosen to have the opposite hand to GSI (Figure 7b - right).

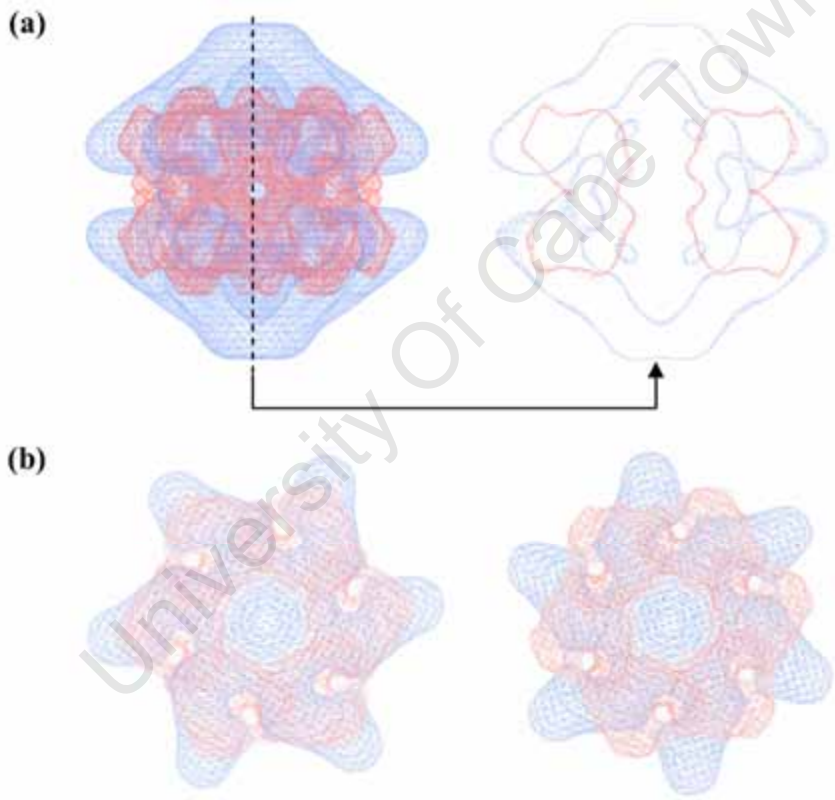


Figure 7: Comparison of the GlnN (blue) and GSI (red) structures at low-resolution. **(a)** Wire-frame representations of the superimposed structures with a transverse cut-away view (black arrow). The GSI volume was generated from the *S. typhimurium* crystal structure [pdbid: *2gls* - Yamashita *et al.* 1989] by low-pass filtering to 10 Å. Structures are contoured to enclose the volume calculated for 12 subunits of: GlnN (82.7 kDa) & GSI (51.6 kDa). **(b)** Superposition of a single hexameric GSI ring with the GlnN volume represented with the same handedness (left) and the opposite handedness (right). Structures were contoured at a higher threshold to emphasize the active site regions.

2.4.3 Reconstruction of MetSox-inhibited GlnN

In order to investigate whether any domain rearrangements take place upon ligand binding, in line with the suggested cooperative modulation of activity in other GS enzymes [Wedler *et al.* 1982; Rhee *et al.* 1981; Shrake *et al.* 1980], the structure of GlnN auto-inhibited by MetSox and ATP was solved by SPEM. This reconstruction also included several improvements to the sample preparation, data collection, and image processing algorithms, which together with the stabilizing potential of the ligands were expected to improve the resolution of resulting GlnN model. However, at the achievable resolution achieved there was no difference between apo and MetSox – inhibited GlnN.

2.4.4 Investigation of resolution limiting factors

The resolutions of both the apo and MetSox-inhibited GlnN cryo-EM reconstructions were estimated at ~ 16 Å despite the fact that oscillations in the measured CTFs of the data could be seen extending past 8 Å (Figure 8). These results suggest that some aspect of the data processing or intrinsic sample characteristic was limiting the resolution of the final reconstruction.

Defocus coverage

The defocus coverage for the MetSox-inhibited data set suggested a possible reason for the lower than expected resolution of the later reconstruction. It can be seen from Figure 9 that the defocus values of these data were not as broad as in the first reconstruction and the average defocus value was quite high at ~ -3.4 µm.

Angular coverage

A persistent difficulty noted during the reconstructions was due to the preferred orientation of the GlnN particles in the ice which caused an uneven sampling of projection angles (Figure 10a). This problem was more acute in the MetSox reconstruction (Figure 10b). The most populous projection direction common to all the reconstructions was the theta = 70° and phi = 40° view which contained 2% of all the images in the apo and 5% in the MetSox-inhibited reconstruction (Figure 10b).

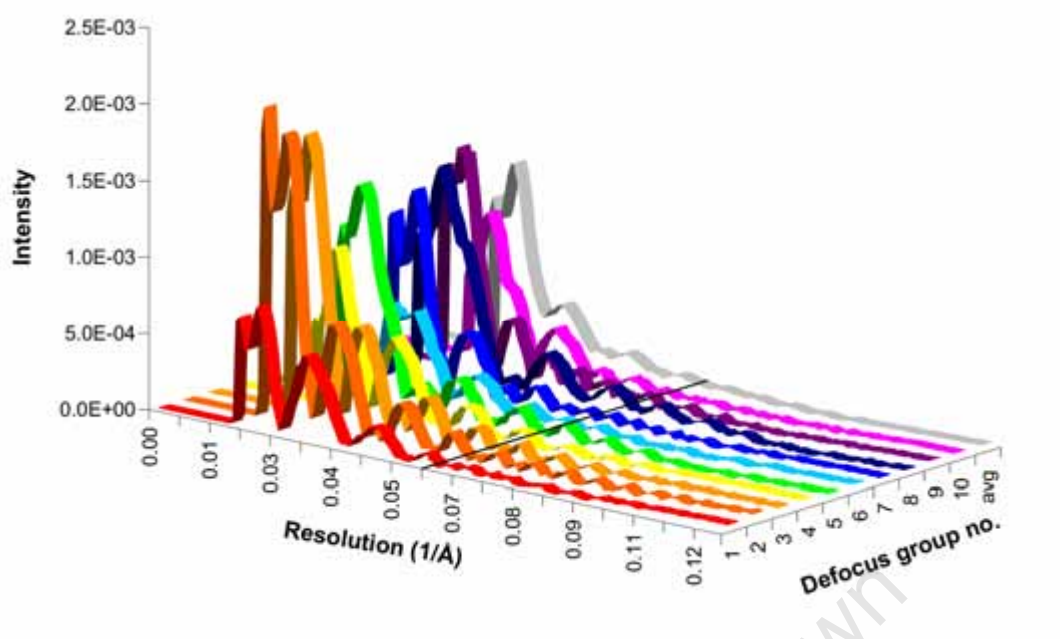


Figure 8: Power spectra for the MetSox-inhibited data set at different defocii. The traces shown are background-subtracted, phase-flipped, radially-averaged periodograms calculated in SPIDER [Frank *et al.* 1996]. The average power spectrum is shown in the last position along the "defocus group" axis. The line at 0.06 Å running parallel to 3rd axis represents the calculated resolution of the reconstructions.

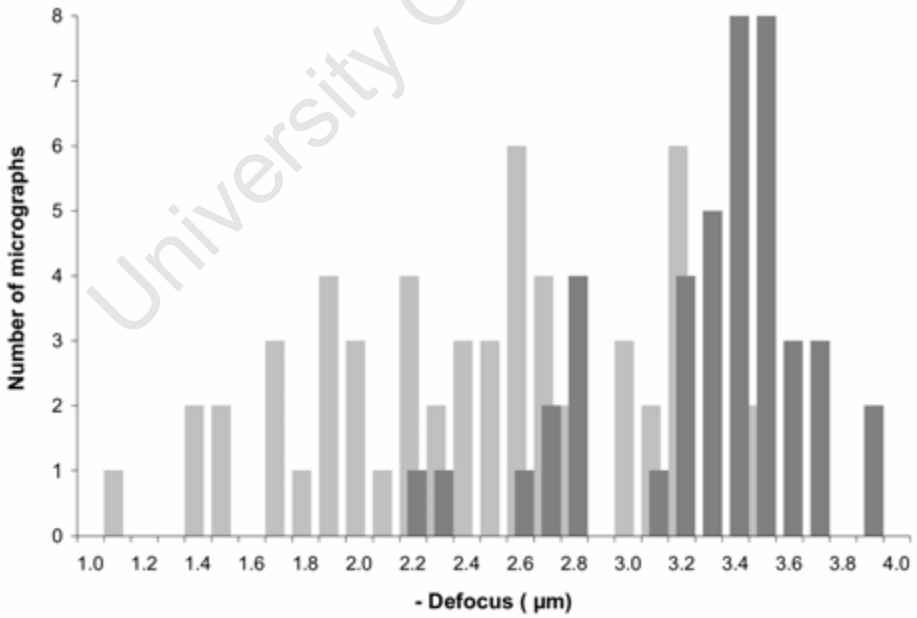


Figure 9: Histogram of the micrograph defocus ranges in the apo (light grey) and the MetSox-inhibited (dark grey) reconstructions.

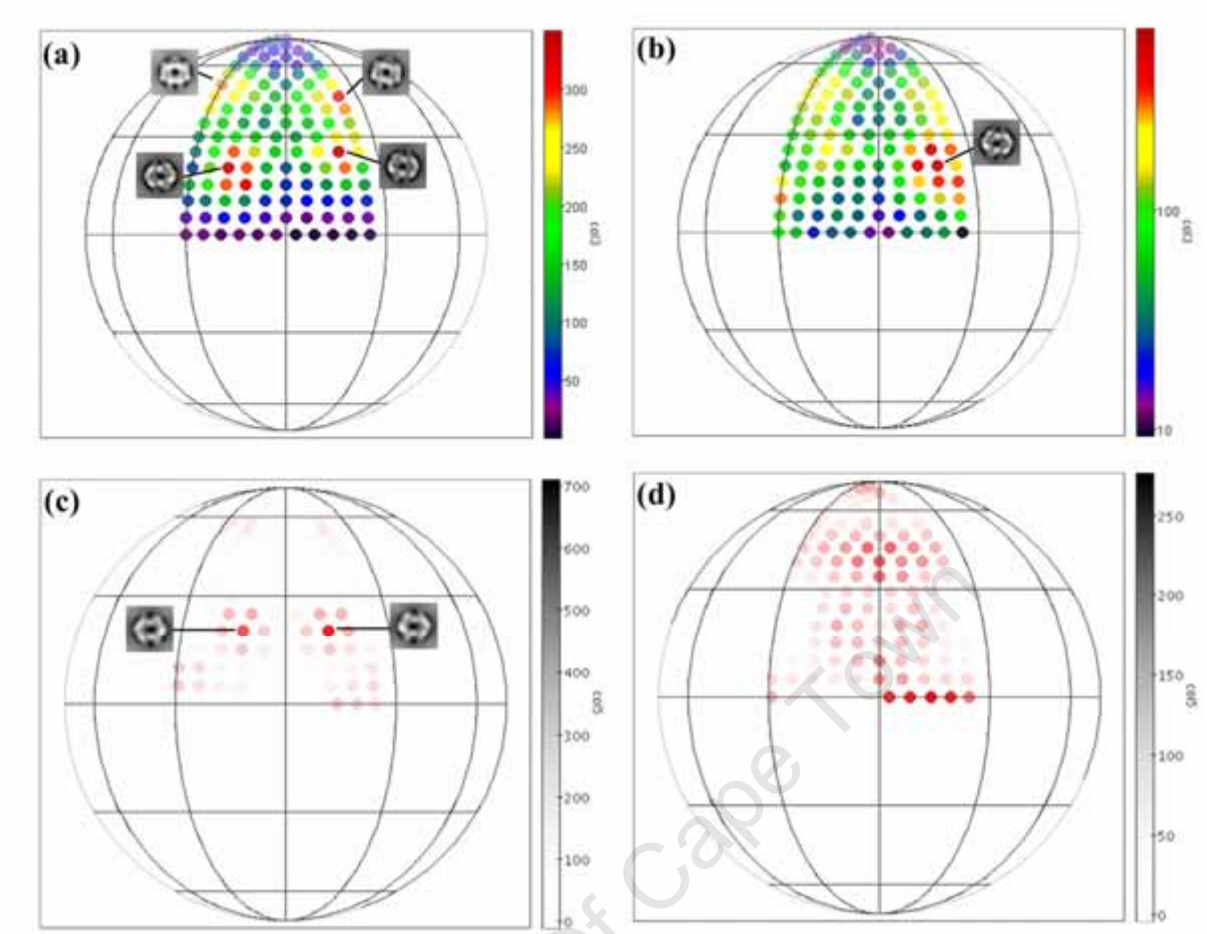


Figure 10: Angular coverage analysis. **(a)** A spherical plot of the image distribution in the apo-GlnN reconstruction. Circles mark the position of reference projection directions (quasi evenly spaced 5° increments in the asymmetric unit) and their colour reflects the number of images in the corresponding class, as given in the associated key. Projections of the final converged volume corresponding to the most populous classes are shown. **(b)** Same as in (a) but the data reflects the distribution of views in the MetSox-inhibited reconstruction. **(c)** Angular distribution of views in the 0° images from the tilt-pair data reconstruction. The data are all coloured in red and their intensity is scaled according to the number of images within each class. **(d)** Angular distribution of views in the 20° images from the same tilt pair reconstruction as in (c). The spherical plots were generated in TOPCAT.

In order to investigate the orientation bias further and to rule out the possibility that misalignment of common particle views was causing the reconstruction to converge on an incorrect volume i.e. a local minimum, data were collected from tilted micrographs of apo-GlnN. The angular assignments from the reconstruction of this 20° tilted data confirmed the presence of an orientation bias in the GlnN samples. It can be seen from a comparison of

Figure 10c and 10d that recorded particle views belonging to the tilted data were far more evenly distributed with a maximum of 250 images per class as compared to the 700 seen in the most populous classes of the untilted data. This results from the distribution of views in the tilted data being spread out along conical projections lying at 20° from their directions in the untilted data, as expected from the random conical tilt projection geometry of a particle adopting a preferred orientation.

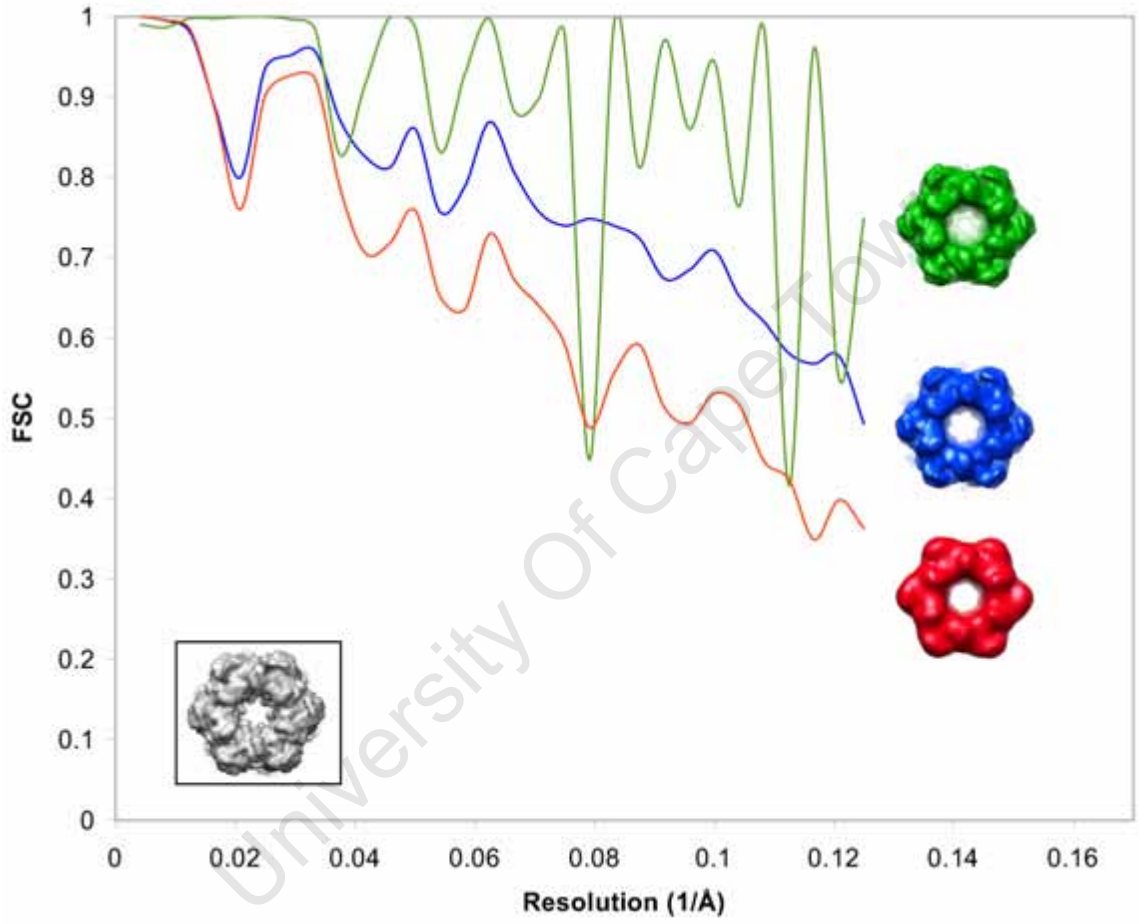


Figure 11: Simulation of the effect of limited angular coverage on resolution. The FSCs between the GSI crystal structure, [pdbid: *2gls* - Yamashita *et al.* 1989] (inset – grey coloured volume), and volumes reconstructed with difference angular samplings of the *2gls* model: all 27 projections (10°) (green), the top 4 most populous classes shown in Figure 9(a) (blue), and the single most populous class Figure 9(b) (red).

The limitations of restricted angular samplings on the resolution of the reconstruction were further investigated by simulation (Figure 11). Several volumes were created from different numbers of evenly spaced projections of the *S. typhimurium* GSI crystal structure [pdbid: *2gls* Yamashita *et al.* 1989]. The FSC between the resulting volumes and the crystal structure were

then calculated to investigate the loss of resolution brought about by back-projection with a limited angular sampling. It can be seen from Figure 11 that a reconstruction generated from only images corresponding to the most populous classes shown in Figure 10a maintained the resolution of the reconstruction up to the Nyquist limit i.e. the 2 × the highest sampling in the recorded micrographs. Back-projection from a single class, which matched the most populous class shown in Figure 10b, however, resulted in a resolution limit of ~13 Å according to the conservative 0.5 FSC criterion.

2.5
Discussion

2.5.1 GlnN quaternary structure

The cryo-EM reconstructions provided an unbiased view of the higher order arrangement of subunits in GlnN. Several lines of independent evidence proved that the GlnN dodecamer possesses D6 symmetry and not the previously suggested C6 arrangement. These results, therefore, confirm that the limited stain depth and large size of the GlnN oligomer together with its strong preferential attachment to the carbon substrate resulted in asymmetric visualization of the particles in previous reconstructions [van Rooyen 2004].

Another important advantage of the cryo-EM reconstruction was the ability to visualize internal density. The cryo-EM volume confirmed the presence of the central solvent filled cavity, which was bordered by the columns of density connecting the two hexameric rings, and also revealed previously unseen rings of density at either end of this cavity. The large solvent filled cavity at the centre of the GlnN complex explains the exaggerated molecular weight, 1.3 MDa, measured by SEC (Section 1.4.4) as these data were calibrated on the basis of molecular weight and not hydrodynamic volume.

2.5.2 Conservation of higher order structure

Despite the poor sequence similarity shared between the GlnN and GSI structures (9% global identity), the overall arrangement of subunits was similar if the handedness of the GlnN reconstruction was chosen to match the GSI crystal structure. If, however, the alternate handedness is assumed, then the GSI and GlnN structures can only be brought into coincidence through an inversion of the orientations of the hexameric rings in the GlnN dodecamer. Such an inversion would also explain the handedness of the previously observed negative stain “pinwheel” views [van Rooyen 2004] as arising from partial staining of the large complex and not as hexameric dissociation intermediates. In either event, the results of the cryo-EM reconstruction confirmed the previously noted similarities of GSI and GlnN within the individual rings [van Rooyen 2004].

Outside of the conserved active site regions forming the closed rings, the differences in the GSI and GlnN structures were more obvious as expected from their different monomer molecular weights (82 vs. 52 kDa) and the resulting larger size of the GlnN complex (24% wider & 65% longer). From such a comparison it can be seen that the GlnN structure appears more expanded with the majority of the mass surrounding a large central cavity unlike in the more compact cylindrical GSI. The domed caps, the tips of the pinwheel arms, and the inter-ring connections contribute the most to the regions of largest disagreement. In GSI the N and C-terminal region of the protein sit on opposite sites of the hexameric rings forming the dodecamer, with the latter region forming the inter-ring interface and the former, the outer solvent exposed surface. It would, therefore, appear that the large differences in the sizes of the GSI and GSIII families is due to differences in the N and C-terminal regions in addition to a large mid-sequence insertion, which presumably maps to the tips of the "pinwheel" arms, as previously suggested by bioinformatics analyses [van Rooyen 2004]. Although the intra-ring contacts forming the active site β-barrel folds must be conserved in the GSI and GlnN structures, it appears that in the latter, these interfaces are supplemented by the extensive contacts forming the large domed caps at either end of the complex. The inter-ring interfaces in GlnN also appear quite different. In GSI, this interface is formed by two isologous interactions: the H-bonding of a pair of two stranded β-sheets to form the walls of the central channel, and the hydrophobic interaction of the elongated C-terminal 'helical thong' from one subunit with a hydrophobic cavity in the opposite subunit [Almassy *et al.* 1986]. These interfaces result in a tight association between the two rings. The situation is different in GlnN where the two rings are joined by more extended processes which form the walls of the large central cavity.

A detailed understanding of the conservation and divergence of quaternary structure in the GSIII family, and the responsible sequence motifs, together with the determination of the absolute hand of the GlnN complex will addressed by the atomic resolution structure of GlnN (Chapter 3).

2.5.3 Structural insights from ligand binding

Both kinetic and structural evidence exists to suggest that the subunits of the large GS enzymes interact cooperatively during ligand binding and catalysis [Wedler *et al.* 1982; Pfluegl *et al.* 1996]. However, the structural basis for this allostery is unknown. The results of

previous spectrophotometric and sedimentation velocity studies have detected both local and larger scale changes in the GSI structure upon ligand binding [Shrake *et al.* 1980]. The recent crystals structures of GSI and GSII enzymes, with and without bound inhibitors, have confirmed the former local rearrangements (see Section 3.2). These structures reveal that the binding of ligands and the subsequent catalysis is achieved through relatively minor (< 10 Å) motions of flexible loops surrounding the active site [Krajewski *et al.* 2008] and these flexible loops are not expected to be visible in the cryo-EM reconstructions of GlnN. The equivalence of the apo-GlnN and MetSox-inhibited GlnN reconstructions presented here suggests that either no large scale domain rearrangements take place upon auto-inhibition by MetSox and ATP or the rearrangements are too small to be perceived at the currently achievable resolution. An additional point to consider is that the enforcement of symmetry during the reconstruction process is a powerful constraint which facilitates the convergence of the reconstruction but does not allow non-symmetric structural perturbations to be visualized.

2.5.4 Limitations to the achievable resolution

Despite several improvements to the reconstruction algorithm and sample, including optimization of ice thickness and particle distribution (data not shown), the resolutions of the reconstructions presented in this work were limited to ~ 16 Å. Several possible factors are suspected as having played a role in this:

Defocus coverage

The poor distribution of defocus values in the MetSox-GlnN reconstruction might account for the failure of the improvements to the sample and image processing algorithms to significantly alter the resolution of the reconstruction. In the first apo-GlnN reconstruction, the defocus coverage extended evenly between -1 and -3.5 µm. However, the defocus coverage of the later MetSox-GlnN reconstruction was skewed towards larger defocus values with an average at -3.5 µm. Such a bias has an effect on the frequency distribution present in the recorded and consequently reconstructed volumes. Higher defocii result in more low resolution contrast but also cause attenuation of higher frequencies due to the defocus dependent envelope function. These higher frequencies can in theory be recovered by applying negative *B*-factor correction but the unavoidable amplification of noise is predicted

to prevent this in practice. The level of attenuation beyond which recovery is impossible is uncertain.

Theoretical calculations of the CTF can reveal the extent of these intrinsic limitations on the resolution of the reconstruction brought about by the envelope function. For a defocus of -3.5 µm the simulated contrast reaches 0 at ~ 6 Å and 10% of the maximum value at 10 Å. The CTF passes 50% contrast at 15 Å, which correlates well with the final resolution attained for both reconstructions but this level of signal should still be well above the noise component. It also does not explain why there was no difference seen between the two reconstructions. The first reconstruction, which had a better defocus spread, centered on -2.4 µm, is predicted to be limited to only 4 Å and the contrast reaches 10% at 6 Å according to the simulation. None of the reconstructions, therefore, appears to have lost significant signal due to the high defocus values and this cannot be the sole contributor towards reducing the ultimate resolution of the reconstruction.

In the reconstruction approach implemented in this study, only the phases of the individual particle CTFs were corrected. The final signal distribution was, therefore, composed of the composite sum of all the phase-flipped particle CTFs without normalized weighting. It is, therefore, possible that under-sampling of defocii could have resulted in poorer signal in certain regions of the CTF which would have affected the appearance of the final reconstruction but not the ultimate attainable resolution.

Limited angular variance

GlnN was previously shown to adopt a preferred orientation with the 6-fold axis tilted at 45° to bring the large flat surface of the conical dome at the top of the molecule in contact with the carbon support film [van Rooyen 2004]. The results of the current cryo-EM study revealed that the GlnN particles also adopted a preferred orientation in hydrated layer prior to vitrification. The possibility that the skewed class distributions resulted from miss alignment of particles was ruled out by the results of the angular analysis of the 20° tilted data. Alignment of these data to the initial cryo-EM volume proved that the particle adopted a preferred orientation by revealing an altered distribution of views which surrounded the most populous classes after tilting.

The preferential interaction of GlnN with the air-water interface was different for each of the reconstructions but there was a common bias towards a theta angle of 30° from horizontal. In the first reconstruction, an additional pair of overpopulated classes, which can be regarded as equivalent seeing as they span the border of the projection asymmetric unit, corresponding to a theta angle of 40°, were also observed. This matches the orientation of the particles previously seen to adopt a preferred orientation on carbon [van Rooyen 2004]. The more limited sampling present in the MetSox-GlnN and the tilted data reconstructions most likely arose from the thinner ice layer used in these experiments. The thinnest ice thicknesses capable of supporting data collection have been estimated at 150-300 Å [Stagg *et al.* 2006]. It is, therefore, reasonable to assume that such geometric constraints might force the large (185 Å) particles into a more horizontal orientation. The 30° tilt from horizontal and the paucity of side views, however, suggest that some interaction is taking place between GlnN and the air-water interface. The hydrophobic nature of the GlnN complex (discussed in Section 1.5.4) further supports this notion.

The simulations of limited angular coverage revealed the expected contribution of this phenomenon to the reduction of the final reconstruction resolution. As in the case of molecules with high symmetry, very few projections were necessary for complete visualization of the GlnN complex. Only when the angular bias was sufficient to be equivalent to back projection from a single view was the resolution limited and then only to 13 Å. Considering that even in the reconstruction showing the most bias, only 5% of the images were contained in the most populous class, it does not seem likely that this is a major limitation in the resolution of the reconstruction.

Particle heterogeneity

A homogenous population of particles is an absolute requirement for successful structural determination by single particle methods. It is, therefore, possible that structural heterogeneity resulting from the intrinsic flexibility of the large GlnN complex is to blame for the limited resolution of the reconstructions. The alternate explanation that structural heterogeneity resulted from denatured or damaged particles seems unlikely considering the use of CC-score thresholds to exclude badly aligning images. The cryo-EM reconstructions presented here took place prior to the adoption of the final purification protocol and the optimizations to the

preparation stability, and it, therefore, remains to be seen what effect these improvements will have on the attainable resolution.

2.5.5 Conclusions

Although the cryo-EM reconstructions presented here appeared to be limited to low-resolution, the resulting structural information not only provided the first insights into the conservation and divergence of quaternary structure in the GSIII family, but also proved essential in solving the crystal structure of the GlnN protein (Chapter 3). In addition, the reconstruction of inhibited GlnN revealed that no large scale domain motions accompany ligand binding or auto-inhibition by MetSox and ATP.

Chapter 3

X-ray Crystallographic Studies of GlnN

3.1
Summary

To date atomic resolution structural information only exists for members of the GSI and GSII families and these results have provided a structural context to half a century of biochemical and biophysical studies into the functioning and regulation of these important enzymes. This chapter describes the X-ray crystallographic studies of the GlnN protein from *B. fragilis*.

The characterization and optimization of diffraction data from the two MetSox-inhibited GlnN crystal forms presented in Chapter 1 revealed that the full-length and the digested proteins crystallized in the P1 and $C222_1$ space groups respectively, with the latter exhibiting stronger diffraction. The initial low resolution phases (7 Å) used in the solution of the $C222_1$ structure were derived by single wavelength anomalous dispersion phasing of Ta_6Br_{12} derivative data and were of sufficient quality to identify the dodecameric GlnN complex in map sections. Density modification in the form of iterative NCS-averaging, with the envelope and NCS parameters derived from the cryo-EM volume (Chapter 2), was then used to extend the phases to the limit of the data (3.0 Å). Model building following simulated annealing torsion angle dynamics refinement produced a model of 88% of the GlnN structure including the ADP and MetSox-P ligands. Several of the missing regions were found to be the most mobile regions in other GS structures, thus, confirming their conservation in light of their suggested catalytic roles. The protease susceptibility site (Chapter 1) was one of the missing regions and was mapped to the very tips of the GlnN pinwheel arms, which were also shown by temperature factor analyses to be the most mobile. This was further supported by the lack of density for this region in the P1 map solved by molecular replacement. Crystallographic packing analyses highlighted the significance of this region in lattice formation and explained how proteolysis leads to increased inter-molecular contacts in the $C222_1$ form.

The crystal structure of GlnN revealed a surprising inversion in the orientation of the hexameric rings making up the dodecameric complex as compared to the GSI quaternary structure. Examination of the interfaces responsible for oligomerization in the GlnN dodecamer showed that the termini and several smaller mid-sequence insertions were primarily responsible for this divergence in higher order assembly. The GlnN complex was found to be stabilized primarily through intra-ring interactions from the C-terminal insertion

and the less extensive inter-ring interface was formed by the N-terminal region.

Despite the differences in quaternary structure, structural alignments of the GlnN monomer with the GSI and GSII enzymes revealed a conservation of active site residues and modes of ligand binding, thus, suggesting a conservation of reaction mechanism among all GS enzymes. The first structure-based multiple sequence alignments of representative members of the GS superfamily also revealed that the greatest sequence variance in the active site residues occurs at the nucleotide binding site and this tolerance was explained in part by a reliance on non-specific main chain interactions with the cofactor. Several residue substitutions were identified that appear to be unique to the GSIII family and these were suggested as targets for the design of selective inhibitors against the family.

The first crystal structure of a GSIII enzyme has raised interesting questions surrounding the mechanism and driving forces responsible for the evolution of higher order assembly and its role in the functioning of the ancient and essential GS enzymes. The structural insights into the determinants of ligand binding specificity in the GSIII family have also suggested several avenues for the design of targeted inhibitors against the important human pathogen *B. fragilis.*

3.2
Introduction

"X-ray Crystallographic Studies of Glutamine Synthetase Enzymes"

3.2.1 Insights into the quaternary structure of GSI enzymes

The first direct visualizations of GS enzymes by electron microscopy [Valentine *et al.* 1968] preceded the attainment of a crystallographic structure by eighteen years [Almassy *et al.* 1986]. Although crystals were available from about the same date [Bywater *et al.* 1969], the breakthrough came with the isolation of *S. typhimurium* GSI from an adenylylation deficient host [Janson *et al.* 1984], thus, avoiding the conformational heterogeneity that is thought to result from post-translational regulation of the enzyme. The resulting crystal form had been noted earlier for the *E. coli* protein grown under different conditions [Heidner *et al.* 1978] and crystals of the *S. typhimurium* protein were subsequently shown to diffract to 2.4 Å [Liaw *et al.* 1993a]. The first ever structure of a GS was solved to 3.5 Å by multiple isomorphous replacement and subsequent studies were able to extend the resolution (2.8 Å) with X-ray crystallographic refinement [Yamashita *et al.* 1989; Liaw *et al.* 1994].

The results of these early crystallographic studies confirmed the quaternary structure assignments from EM [Valentine *et al.* 1968] by revealing that the GSI enzymes are large dodecameric complexes composed of two opposing hexagonal rings arranged with 622 (or D6) symmetry (Figure 1a). The higher resolution of the crystal structures showed that the GSI polypeptide is divided into two domains, the boundary of which, did not agree with the earlier results suggested by negative stain EM [Kessel *et al.* 1980] or proteolysis [Lei *et al.* 1979]. The smaller N-terminal domain (residues 1-102) has an ubiquitin-like β-grasp fold and unlike the C-domain, it is predominantly composed of β-sheets. The C-terminal domain (102-468), which forms the majority of the structure, is built around a 6-stranded anti-parallel β-barrel fold (Figure 1b). The majority of helices in this domain pack against the rear of the solvent accessible β-sheet formed by the motif, thus, forming a crescent shaped scaffold and giving rise to the characteristic GS 'pinwheel' arms.

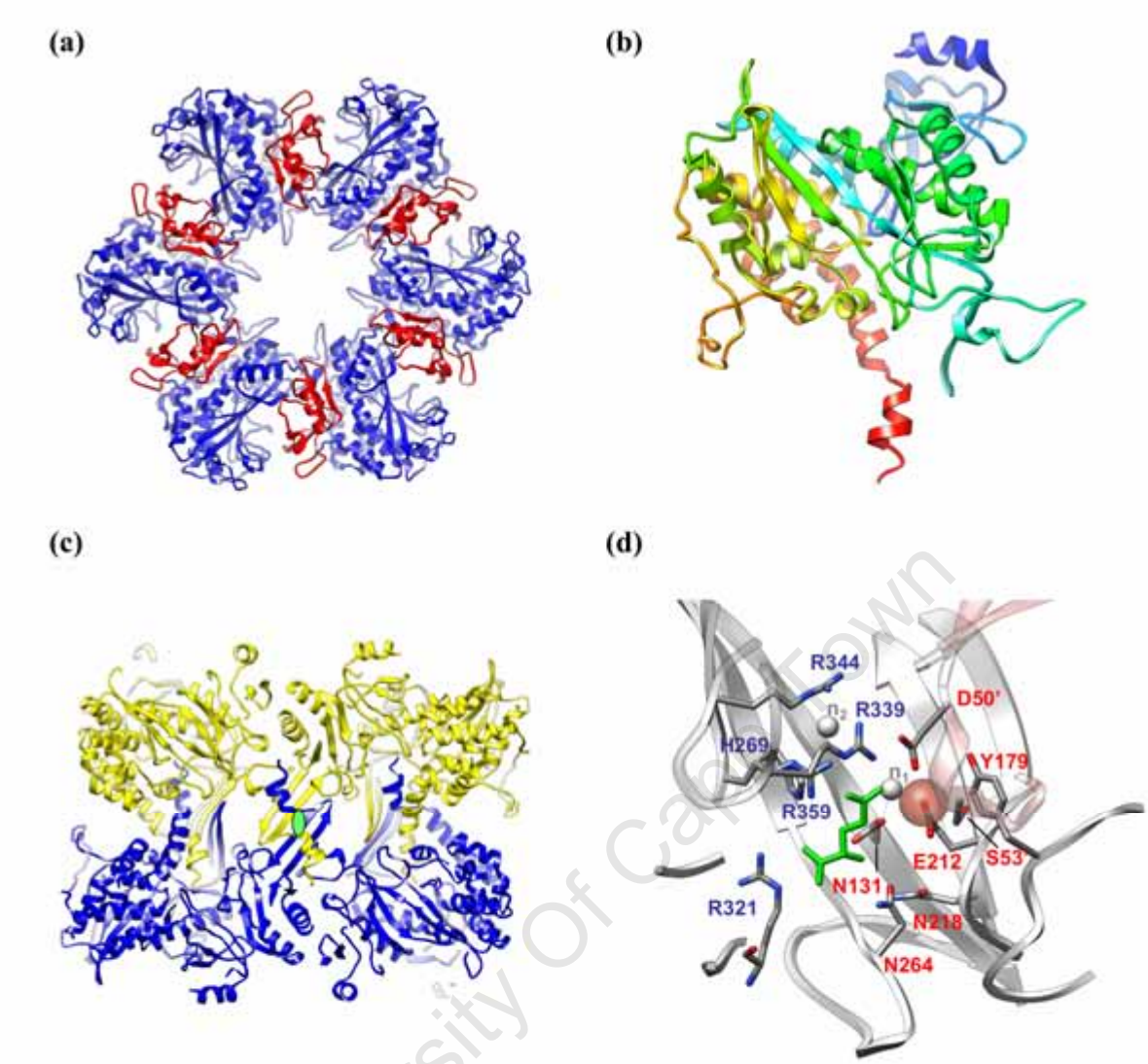


Figure 1: Structural insights from GSI crystallography. **(a)** Domain arrangements in the *S. typhimurium* GSI oligomer [pdbid: *2gls* – Yamashita *et al.* 1989]. The N and C-terminal domains within one hexameric ring are shown in red and blue respectively and the protein backbone is shown as a ribbon trace. **(b)** Crystal structure of the *S. typhimurium* monomer [pdbid: *1fpy* - Gill *et al.* 2001] coloured according to residue position. The view is into the active site with the dodecameric ring interface towards the bottom of the image. **(c)** Isologous contacts responsible for the associations between opposing rings (yellow and blue). The view is down the 2-fold axis (green ellipse). Parts of the structure were removed for clarity. **(d)** Close-up of the GS1 active site (*1fpy*) showing the positions of residues forming the glutamate (green) and NH_4^+ (brown sphere) binding sites. The view is the same as in (b) and the N-terminal domain contribution of the adjacent subunit in the ring is shown in transparent red colouring and the residues are labeled with an apostrophe. Residue labels are coloured according to the charge characteristics of their functional groups (red - negative & blue – positive).

The polar GS active site is formed by the heterologous associations of the N (2 β-strands) and C-terminal domains (6 β-strands) from adjacent subunits resulting in the characteristic GS ring structures (Figure 1a). The associations of the partially eclipsed opposing rings of the GSI dodecamer, on the other hand, are isologous in nature (Figure 1c). The inter-ring stabilizations are dominated by (a) the pairing of β-strands (136-154) with their symmetry related partners to form the 4-stranded β-sheet lining the walls of the GSI central cavity and (b) the hydrophobic C-terminal 'helical thongs' which interface extensively with the symmetry related subunit [Almassy *et al.* 1986].

3.2.2 Structural basis for catalysis

The results of several subsequent studies, which looked at the structures of GSI enzymes with metal ions, substrates, inhibitors, and cofactors bound in the actives site, have led to a detailed understanding of the structural mechanism of GS catalysis and explained the conservation of the residues involved [Reviewed by Eisenberg *et al.* 2000]. Some of the most highly conserved residues in the GS active site are responsible for chelating the divalent cations, Mn^{2+} or Mg^{2+}, which contribute to structural stability whilst playing an essential role in substrate binding and catalysis (Figure 1d) [Liaw *et al.* 1994]. The higher affinity, n_1, and lower affinity, n_2, metal binding sites divide the β-barrel into two halves and the nucleotide cofactor and glutamate substrate enter from opposite sides to interact with the metals. The charge state on the n_2 site, adjacent to where ATP binds, is only partially satisfied and this is thought to assist in the phosphoryl transfer of the γ-PO4 to glutamate during the first step of catalysis (see below).

Within the lower half of the β-barrel, the electrostatic distribution is polarized due to the different charge characteristics of the substrate and ammonium binding sites (Figure 1d) [Gill *et al.* 2001]. The substrate, glutamate, binds near the n_1 metal binding site on one side and is stabilized by both a positively charged pocket (R321, R339, R344, R359, and H269 - numbered according to the *S. typhimurium 2gls* structure), which interacts with the α and γ carboxylate groups [Liaw *et al.* 1995], and by a negatively charged pocket (N264, E131, and E218), which stabilizes the amino group [Gill *et al.* 2001]. On the opposite side of the lower β-barrel is the polar solvent exposed ammonium binding site formed by residues Y179, S53', D50', and E212 (with apostrophes indicating a contribution from an adjacent subunit within a ring) [Gill *et al.* 2001]. The high affinity of the latter negatively charged tetrahedral

arrangement for NH_4^+ over H_2O prevents the futile hydroylysis of reaction intermediates back to glutamate (see below) and the identity of the attacking species bound to this site determines the reaction catalyzed by GS [Eisenberg *et al.* 2000]. If hydroxylamine replaces NH_4^+ then the reaction will lead to the production of γ-glutamylhydroxamate Alternatively, if glutamine is supplied instead of glutamate, the γ-glutamyl group is transferred to hydroxylamine also yielding glutamylhydroxamate, via the "reverse reaction".

The catalysis of glutamine synthesis in the GS enzymes takes place via a simple two-step mechanism (Figure 2) [Meister 1989; Liaw *et al.* 1994]. In the first step, the γ-PO_4 of ATP is transferred to the carboxylic side chain of the substrate glutamate to form an activated gamma-glutamyl phosphate intermediate. In the second step, nucleophillic attack by ammonia on the carbonyl carbon leads to a tetrahedral intermediate which quickly breaks down to form glutamine with the release of inorganic phosphate. The later release of reaction products is the rate limiting step during normal catalysis [Abell *et al.* 1991].

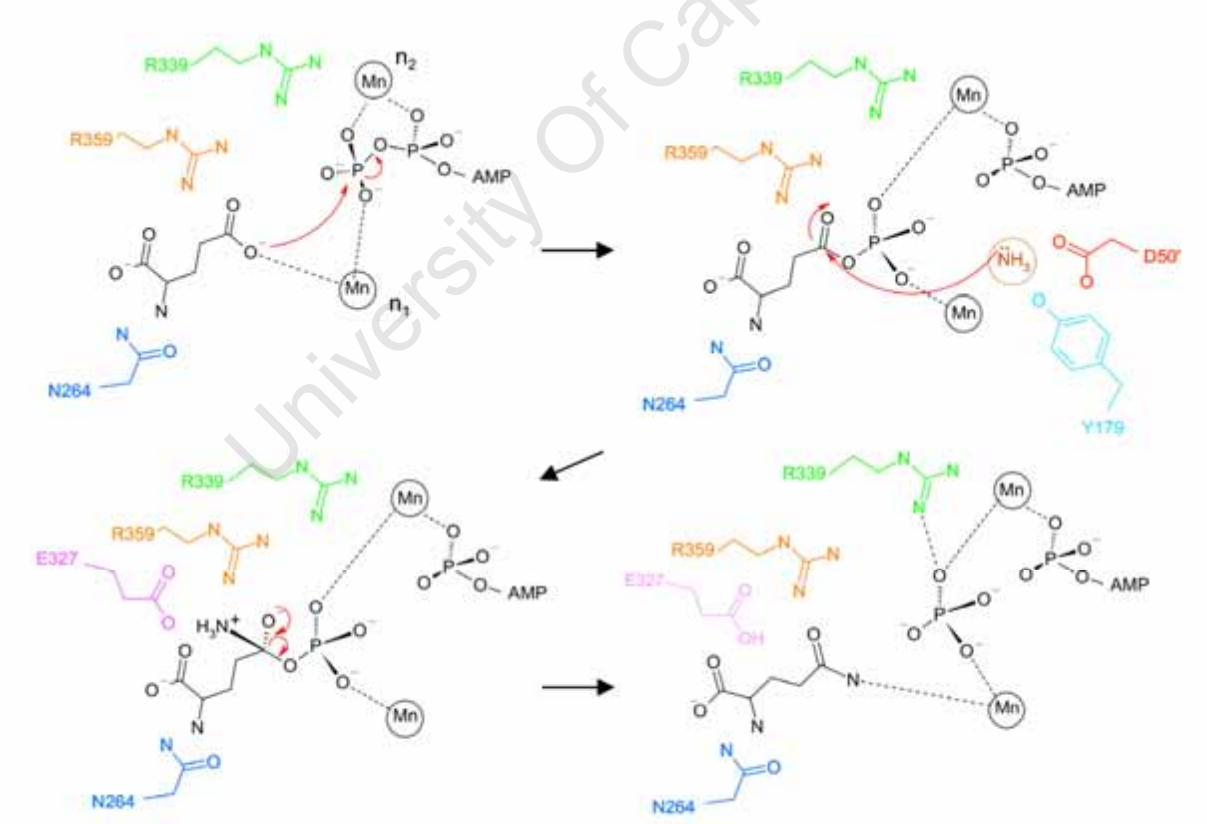


Figure 2: Summary of the GS reaction mechanism (adapted from [Liaw *et al.* 1994]) and drawn in ISIS^{TM}/Draw. Red arrows indicate the movement of electrons. Residues are coloured for ease of identification in Figure 3.

In addition to the highly conserved residues which mediate this catalysis, the importance of several mobile loop regions has been discovered through comparisons of GSI structures trapped in different stages of catalysis [Yamashita *et al.* 1989; Liaw *et al.* 1994 & 1995; Gill *et al.* 2001 & 2002] and through multi-copy X-ray crystallographic refinement [Gill *et al.* 2002]. The five important loops involved in catalysis (numbered according to the *S. typhimurium 2gls* structure) are: (a) the Asp50 loop or the 'latch' (residues 50–64) contributed by the N-terminal domain of an adjacent subunit; (b) the Tyr179 loop (153–188); (c) the Asn264 loop (255–267); (d) the Glu327 loop or the 'flap' (324–329); and (e) the Tyr397 loop or 'adenylylation' loop (388–411). One of the most important loops is the 'flap', which is involved in several aspects of catalysis, such as: closing over the active site to contain the reaction intermediates, phosphoryl transfer, and formation of the ammonium binding site [Gill *et al.* 2002]. The remaining loops such as the 'latch' act in concert to stabilize the 'flap' thereby regulating catalysis.

The loop motions facilitating the GS enzyme reaction are shown in Figures 3a-d and can be summarized as follows:

(1) First the cofactor ATP enters and binds at the top of the β-barrel near the n_2 metal ion (Figure 3a & b). It can be seen from these structures that the highly mobile 'flap' and 'latch' loops are only visible after the binding of ATP, suggesting a significant conformational change and stabilization by this ligand. This has been shown to occur via the cooperative stabilization of the β-sheet 209 which allows the N-terminal domain of the adjacent subunit to approach more closely, whereupon, D50' interacts with R339 and D64' with R344 to stabilize the latch next to the active site [Gill *et al.* 2002].

(2) The loop movements induced by ATP binding also cause R359 to move towards the glutamate binding site resulting in an increased affinity for the substrate. Glutamate binds above the E327 'flap' and the N264 loop (the analogous position of the glutamate substrate can be inferred from the location of the inhibitors in Figure 3c & 3d). Upon binding glutamate, N264 interacts with S53' and E327 of the 'flap' interacts with S52' to further stabilize the 'latch'. These loop movements also explain the stabilization of intra-ring contacts upon ligand binding.

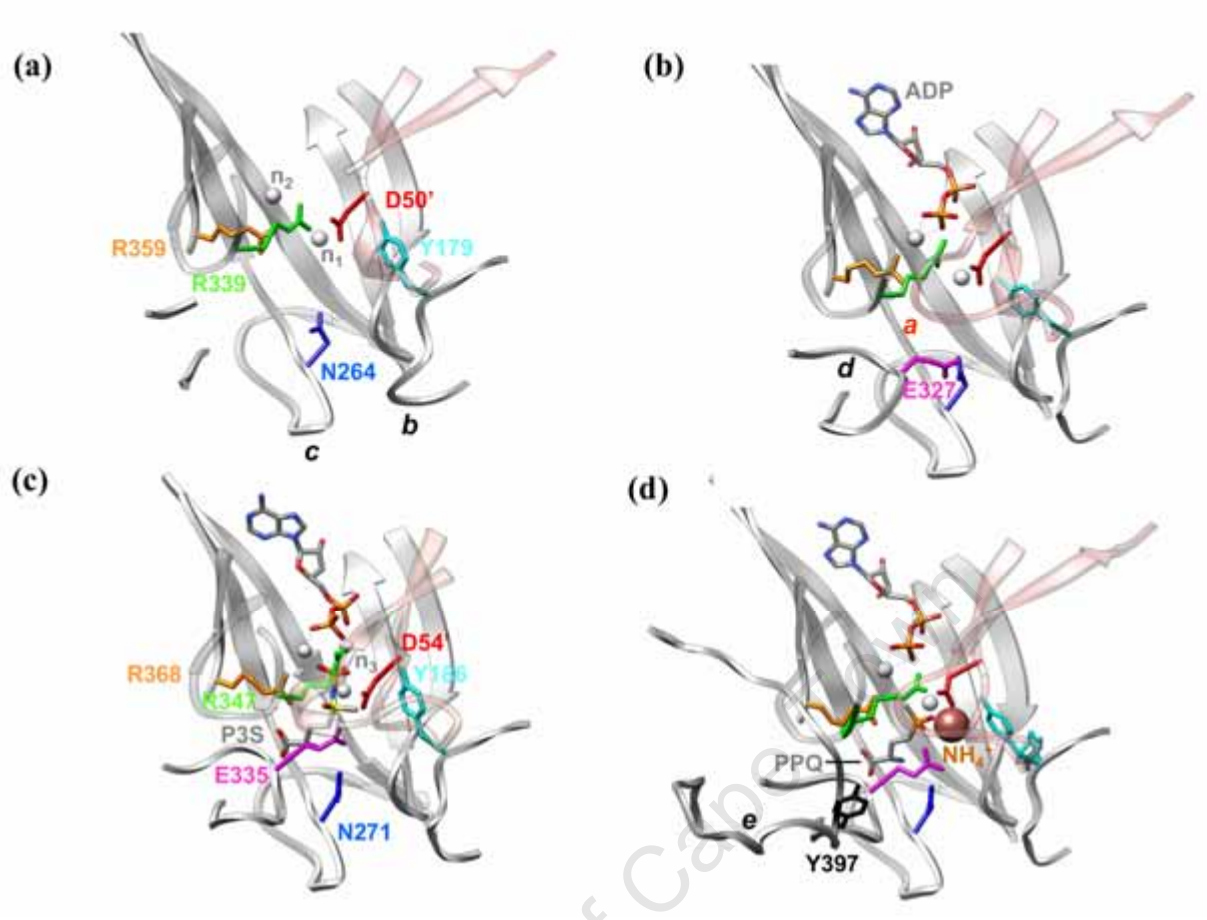


Figure 3: Overview of important side chain and loop movements accompanying the binding of ADP, phophinothricin (PPQ), and MetSox (P3S). Frames **a-d** were rendered from structures *1lgr*, *1f52*, *2bvc*, and *1fpy* respectively and the position of the NH_4^+ binding site was taken from the Tl^+ ion position in *1f1h* [references are given in Table 1 in Section 3.3.3]. The colouring of side chains matches that of Figure 2. The structures shown in this figure do not represent frames of a sequential reaction trajectory but rather distinct ligand binding states which serve to illustrate the resulting changes in the positions of residues and loops. Figures (c) and (d) represent the two end points of inhibitor binding and the labeling in (c) follows that of *M. tuberculosis*.

(3) In its new position, R339 sits near the inorganic phosphate binding site (the position of which can be inferred from the phosphorylated sulfonamide group of MetSox in Figure 3c) and participates in the phosphoryl transfer of the ATP γ-PO_4 to the glutamate γ-carboxylate group. This transfer is further facilitated by the partial charge satisfaction of the n_2 site, as described above. The altered positions of D50' and Y179 following substrate binding and phosphorylated-intermediate formation contribute to a

polar and tetrahedral arrangement of the chelating groups, which is specific for ammonium. D50' is, therefore, thought to deprotonate the ammonium ion to form ammonia [Liaw *et al.* 1994] and the subsequent nucleophilic attack by the lone pair electrons on the carbonyl carbon of the gamma-glutamyl phosphate intermediate then results in the formation of a tetrahedral intermediate.

(4) The positive charge of the resulting γ-amino group of the intermediate further stabilizes the 'flap' by forming a salt-linkage with the negatively charged E327, therefore, closing the active site to shield the reaction intermediates from hydrolysis or escape (compare Figure 3b with 3c or 3d).

(5) Following collapse of the tetrahedral intermediate and the formation of inorganic phosphate, E327 is thought is to accept a proton from the γ-amino group to yield glutamine. The stabilizing link between the 'flap' and the substrate is no longer maintained and the reaction products are free to diffuse from the active site. The reaction is then ready to repeat itself with the entry of new substrate molecules.

Such an understanding of the structural basis of the GS reaction mechanism has depended on the solution of GSI structures containing bound inhibitors. The structure of phosphinothricin (PPT) bound GSI (Figure 3d) [Gill *et al.* 2001] first led to a postulated inhibitory mechanism in which the stabilization of the E327 'flap' and N264 loop traps the active site in the closed conformation and prevents the inhibitor from leaving or substrate from binding. The 'flap' is locked in this configuration due to the persistent interaction of the positively charged and protonated nitrogen of the phosphinyl groups with the E327 residue. Even if proton transfer takes place, the phosphate group maintains electrostatic and H-bond interactions with E327 and the reaction stalls because nucleophilic attack by ammonium is no longer possible due to the disruption of the negative binding pocket by the Oε2 group of the inhibitor.

With the recent solution of a higher resolution (2.1 Å) *M. tuberculosis* GSI structure in complex with the tight-binding MetSox inhibitor (Figure 3c) [Krajewski *et al.* 2005], however, it has been shown that the methyl group of the tetrahedral transition state analogue is responsible for the shift in D50' and Y179 which lead to disruption of the ammonium binding site, and not the previously inferred Nε group or Oε2 in the case of PPT [Gill *et al.* 2001]. This structure also revealed a number of novel features which have important

implications for the GS reaction mechanism: (1) The nucleotide base of the ADP cofactor was modeled in a completely different orientation (180° about the glycosidic bond) to the majority of previously observed structures (described above). As a result, several new cofactor-protein interactions were identified and K215 (*M. tb* numbering) was noted as possibly participating in the phosphoryl transfer step. A similar orientation was previously seen on only one occasion for the orientation of the nucleotide but not the terminal phosphate of AMP in the structure *1lgr* [Liaw *et al.* 1994]. (2) All previous GSI structures, none of which were adenylylated, were crystallized with manganese in the active site. This structure, however, was crystallized with very high concentrations (200 mM) of the physiologically preferred magnesium ion with the surprising result that a third metal binding site was identified in the active site. From its position adjacent to the terminal phosphates of ADP, this site is thought to facilitate substrate binding and stabilization of the intermediates. (3) The higher resolution of the structure also revealed a previously unseen H-bond between the structural equivalents in *S. typhimurium* of E327 and D50' (see mechanism above). This linkage has implications for the stabilization of intra-ring contacts upon glutamate binding and for the formation of the ammonium binding site.

3.2.3 Structural changes accompanying post-translational regulation

Because of their central importance, GS enzymes are regulated at numerous levels by many different mechanisms and the properties of these enzymes, such as catalytic potential, reaction kinetics, pH optima, sensitivity to feedback inhibitors, and conformational states, are modified by a number of effectors and conditions in a complex fashion [reviewed by Eisenberg *et al.* 2000]. In addition to explaining the reaction mechanism of glutamine synthesis by the GS enzymes, structural insights have also helped explain the effects of these varied levels of post-translational regulation on GS activity. (1) One of the first major discoveries was the finding that the heterotropic 'cumulative feedback' inhibitors of GS enzymes did not bind at distinct effector sites as previously suggested [Woolfolk *et al.* 1967a] but rather all bound to the same glutamate binding site described above [Liaw *et al.* 1994 & 1993b]. (2) The first GSI structure also identified the protease susceptible region [Lei *et al.* 1979] as a surface loop (156-173) inside the central channel and it has subsequently been shown that this is one of the most mobile regions alongside the five regulatory loops described above [Gill *et al.* 2002]. The abolishment of activity following digestion of this loop is now understood in the context of the 'Y179' loop's contribution to the ammonium

binding site [Almassy *et al.* 1986]. (3) Another flexible loop, the 'adenylylation' loop is involved in the negative regulation of GS activity after the covalent attachment of AMP to the Y397 [Stadtman 1990] which changes the rate limiting step of the reaction to intermediate formation rather than substrate release [Abell *et al.* 1991]. This loop sits next to the 'flap' and is thought to stabilize E327 in the 'taut' form of the enzyme (Figure 3d). No crystal structure of an adenylylated GS has ever been solved but the mechanism of inhibition has been postulated in analogy to the 'relaxed' form of the enzyme [Gill *et al.* 2002]. In this model, the covalently attached AMP group of Y397 interacts with the hydrophobic portion of the 'latch', preventing its closure and disabling the N264 and Y179 loops. Without the interactions mediated by these loops, the enzyme is thought to adopt the configuration found in the 'relaxed' state, thus, disrupting the substrate, ammonium, and n_1 metal binding sites in addition to preventing phosphoryl transfer. A comparison of the crystal structures of 'taut' and 'relaxed' *M. tb* GSI [Gill *et al.* 2002; Krajewski *et al.* 2005] has also explained the previously noted differences in hydrodynamic volume [Ginsburg 1972] that could not be discerned with SPEM techniques [Valentine *et al.* 1968]. Because the E209 β-sheet sits near the surface of the GSI molecule, it is suggested that the changes (although slight) in this region, which result from the missing n_1 metal ion in the 'relaxed' state, are responsible for this difference and the inability to form higher order tubular aggregates [Valentine *et al.* 1968]. (4) Finally, one feature of GS activity and regulation that structural studies have failed to adequately explain is the cooperative interaction of GS subunits. Evidence for allosteric interactions have come from kinetic studies [Wedler *et al.* 1982; Rhee *et al.* 1981; Shrake *et al.* 1980] and the crystal structure of a GSI enzyme in the presence of low ADP concentrations, which revealed that binding of one cofactor increases the binding affinity of the opposing subunit [Pfluegl *et al.* 1996]. The structural mechanism of this suggested cooperative behavior, however, is still uncertain and due to the large size of the GS enzymes, the crystallographic refinement procedures required to examine such behaviors are poorly behaved without the enforced symmetry relations between subunits.

3.2.4 Structures of other GS enzymes

As described earlier (Section 2.2.2), there has been much controversy surrounding the quaternary structures of the GSII enzymes and this has been exacerbated by the presence of multiple genes within each species encoding enzymes with different sizes, stabilities, and catalytic properties. With the recent achievement of the first GSII crystal structure yet another

quaternary structure arrangement has emerged. The crystal structure of the cytosolic GSII from *Zea mays*, confusingly called GS1a, was solved using multiple isomorphous replacement with anomalous scattering, and revealed a decameric complex composed of two staggered pentameric rings arranged with 52 symmetry (or D5) (Figure 4a) [Unno *et al.* 2006]. The overall fold and the domain boundaries (N:1-103; C:104-356) are remarkably conserved in comparison to the GSI enzymes (Figure 4a & 4b) and consequently the intra-ring quaternary structure associations that form the active site β-barrels are maintained except for the tighter diameter of the pentameric rings. The inter-ring interface, on the other hand, is substantially less extensive than in GSI, with an order of magnitude decrease in the contacting surface area (Figure 4c). This difference arises because the shorter GSII enzymes are lacking the sequence regions which form the extended C-terminal 'helical-thong' and the mid-sequence 'β-loops', both of which, interact with opposing subunits to stabilize the inter-ring interface. Instead, a short section (145-152), which does not appear to have any secondary structure, is responsible for inter-ring associations in GSII. The GSII structure also has an extended N-terminal region (1-18) which forms a short helical segment called the 'meander' that packs inside the central solvent filled channel of the decamer and, thus, appears to contribute towards intra-ring associations. Interestingly, through the creation of chimeric constructs of two different plant isozymes with different stabilities, Unno *et al.* [2006] were able to attribute these physiologically relevant differences to a single hydrophobic residue involved in stabilizing intra-ring associations.

The results of several subsequent structural investigations have confirmed the conservation of this quaternary structure arrangement in yeast [He *et al.* 2009], dicotyledonous plants [Seabra *et al.* 2009], and mammals [Krajewski *et al.* 2008], dispelling concerns that the higher order arrangement was species specific. The structure of a human GSII from the latter study has also explained the differences in the stabilities of several mutants with altered inter-ring interfaces, and together with the similar experiments in the *Z. mays* GSII (listed above) these results confirm the weak nature of this higher order interaction.

In addition to the overall similarity in the folds of the *Z. mays* GSII and the earlier GSI structures, this study revealed a remarkable conservation of the active site residues involved in metal ion and substrate binding [Unno *et al.* 2006]. The structures of GS1a containing bound PPT and MetSox revealed that this similarity extends to the modes of interaction with the substrates, suggesting a common reaction mechanism with the GSI enzymes (see above).

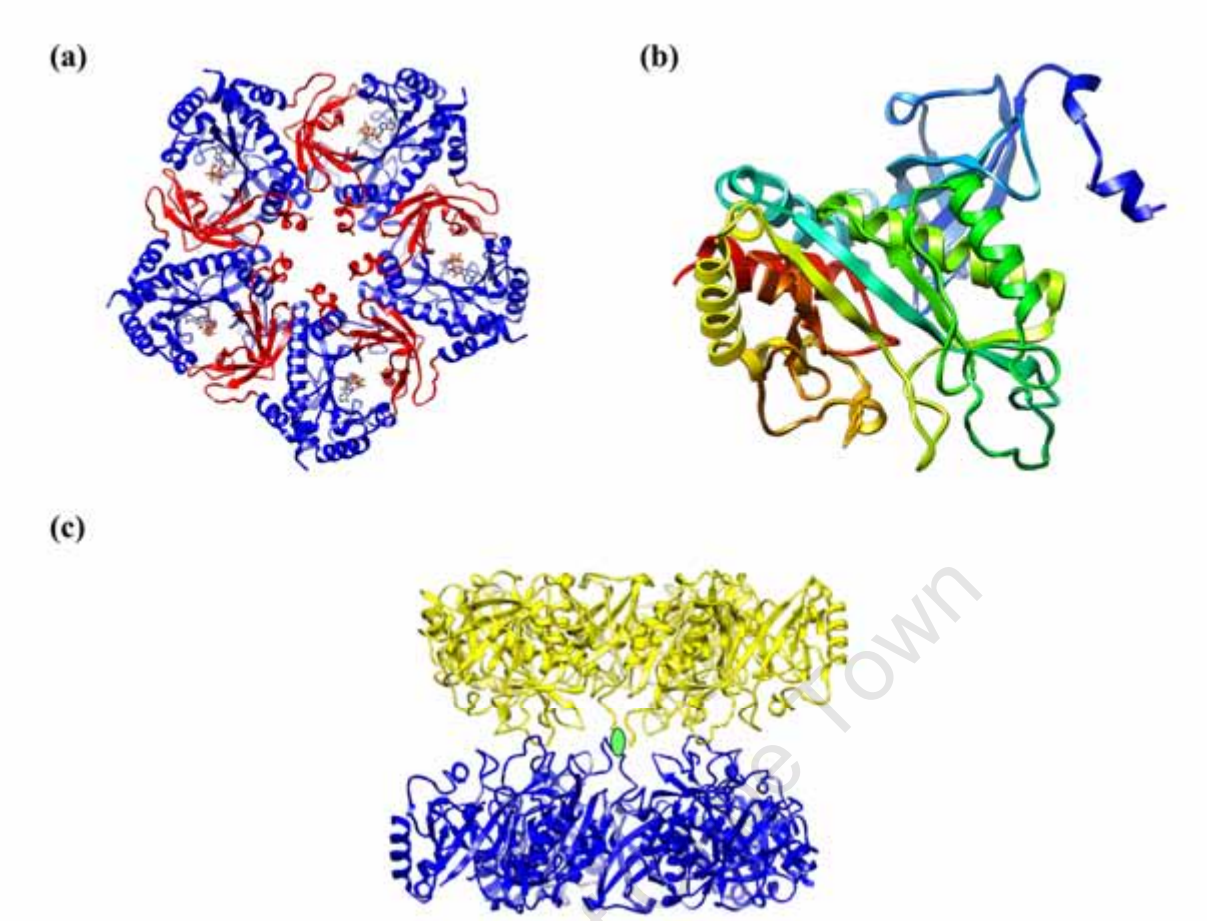


Figure 4: Structural insights from GSII crystallography. **(a)** Domain arrangements in the *Z. mays* GSII pentameric rings [pdbid: *2d3a* - Unno *et al.* 2006]. The rendering and colouring is the same as in Figure 1(a). **(b)** Crystal structure of the *Z. mays* monomer coloured according to residue position. The view is into the active site with the decameric ring interface towards the bottom of the image. **(c)** Isologous contacts responsible for the associations between opposing rings (yellow and blue).

The nucleotide binding site, however, was found to be significantly different, both in terms of residue identity and their mode of interaction with the cofactors, and this discrepancy was suggested as a possible avenue for the design of selective GS inhibitors to avoid cross-targeting the human GSII enzymes.

Despite their divergence from the GSI family, the structures of GSII enzymes from plants, mammals, and yeast have additionally provided some structural insights that have a bearing on all GS structures:

(1) All of the crystal structures of GSII enzymes in the 'taut' state contain the additional third metal ion binding site first seen in the *M. tb* GSI [Krajewski *et al.* 2005], despite the use of lower concentrations of divalent cations (~10 mM vs 200 mM). Interestingly, the human GSII structure, crystallized with ADP and phosphate in the active site [Krajewski *et al.* 2008], possesses a fourth Mn^{2+} ion near the ammonium binding site and the ability of this site to bind monovalent cations has been noted before [Liaw *et al.* 1995]. The presence of free phosphate at the center of the metal ion binding sites in the human GSII structure, however, is at odds with the previously suggested mechanism of direct phosphoryl transfer proposed by Unno *et al.* [2006].

(2) In all the structures containing bound MetSox and PPT inhibitors, the methyl group of the sulphonimide and phosphinyl groups were found to point towards the ammonium binding site, thus, confirming the orientation first suggested by Gass *et al.* [1970] and visualized by Krajewski *et al.* [2005].

(3) The orientation of the nucleotide bases in the GSII structures match those of the cofactors in the earlier structure of AMP bound GSI from *S. typhimurium* [Liaw *et al.* 1994] and the later 'taut' GSI from *M. tb* structure [Krajewski *et al.* 2005]. These results are in disagreement with the earlier *S. typhimurium* structures wherein the base was rotated by 180° about the glycosidic bond (see above).

(4) A comparison of the different GSII enzymes solved to date shows that the differences in their structures are localized to several distinct regions. The largest variance is seen at the N (before the first β-strand) and C-termini (after the last α-helix), where differences in sequence length produce structurally variable unconstrained sections of random coil. Truncation of the yeast N-terminal 'meander' region was in fact essential for crystallization but did not effect the oligomeric state of the complex, thus, suggesting that these N-terminal intra-ring associations are not essential [He *et al.* 2009]. In addition to these seemingly non-essential extensions, the important catalytic loops arranged around the entrance to the lower half of the active site are the most mobile and disordered regions in the GSII structures, as found for the GSI structures (see above). Changes in the 'flap' and 'latch' have also been seen upon ligand binding as originally reported for GSI [Krajewski *et al.* 2008]. Furthermore, the shifted position of β-sheet 209 which characterized the transition from 'relaxed' to 'taut' enzyme upon cofactor and metal

binding [Gill *et al.* 2002] has also been observed in all the GSII structures, thus, suggesting that it is a common feature of GS structures.

(5) Comparisons of the most recent GSII structures with the GSI family have confirmed the varied nature of the nucleotide binding site in comparison to the high degree of conservation of the glutamate, ammonium, and especially the metal binding sites first seen by Unno *et al.* [2006]. The crystal structure of the human GSII enzyme has also allowed a detailed comparison of these sites, and the results have not only led to a reiteration of the importance of the ATP binding site in designing inhibitors against GS enzymes, but also highlighted the difficultly in achieving selectivity between the more closely related enzymes of mammals and plants [Krajewski *et al.* 2008]. The former strategy represents a divergence from the most widely available GS inhibitors, like MetSox and PPT, which target the glutamate binding site [reviewed by Eisenberg 2000], but due to cross-reactivity they are toxic to humans and other mammals [Lamar 1968]. These efforts have already led to investigation of a purine analogue with a 60-fold higher selectivity for inhibiting the *M. tb* GSI than the human GSII [Nilsson *et al.* 2009].

3.2.5 Summary

The application of X-ray crystallography has had a major impact on our understanding of GS function and regulation by giving a structural context to several decades of biochemical and biophysical data. The resulting atomic resolution information has provided insights into the unique quaternary structure of these enzymes and, through ligand binding studies, elucidated the role played by subtle structural changes in the multi-step reaction mechanism. The recently determined GSII structures have also overturned a number of misinterpretations from lower-resolution EM studies and contradictory biochemical information about the quaternary structures of these enzymes. Comparisons with the well characterized GSI family have revealed the relationship between these families and permitted the design and optimization of selective inhibitors against GS enzymes of medicinal and commercial interest.

3.3
Methods and Materials

3.3.1 Data collection

Cryo-preservation

Crystals were looped from their droplets under oil in the microbatch plates (Hampton Research) using LithoLoops^{TM} (standard SPINE pin lengths) fitted in Hampton HT Crystal Caps or Molecular Dimensions Magnetic Cryo-Caps. The crystals were either placed into drops of cryoprotectant under oil for several seconds or dragged through drops of Paratone-N oil (Hampton Research) before being re-looped and flash-cooled in the cryo-stream. The final crystallization conditions described in the previous chapter served as the basis of all cryo-protectant solutions.

Heavy atom derivatization

Stocks of heavy-atom (HA) solutions were made up in cryo-preservation solution containing 20% PEG 400. Ta_6Br_{12} (Jenna Biosciences) was prepared at a stock concentration of 10 mM and 0.2 µl was added to the 2 µl crystallization droplet under oil. Crystals were then rapidly transferred into the droplets and incubated for varying time periods before retrieval and cryo-protection as described above.

Data collection

Final data collection from a $C222_1$ (digested) crystal (0.1 mm) preserved under optimized cryo-conditions (Paratone-N oil) was performed at BM14 at the ESRF, in Grenoble, France. BM14 was configured in the standard setup for single crystal protein diffraction with the MAATEL MDII microdiffractometer, SC3 robotic sample changer, Oxford Cryo-stream, and MAR CCD detector (225 × 225 mm), except that a custom designed low-resolution beamstop manufactured by the ESRF instrument group was fitted. Data (3 Å) were collected at 100 K with a wavelength of 0.95373 Å over 189 frames with 1° oscillations, 3 spindle passes per

image, and a crystal-to-detector distance of 336.3375 mm. The synchrotron was operating in 4 × 10 mA filling mode and exposure times of 90 seconds were used.

During characterization of the $C222_1$ crystal form's diffraction quality, the HC1 humidity control device, in place of the cryo-stream, was used to vary the relative humidity (RH) during data collection [Sanchez-Weatherby *et al.* 2009]. Crystals were fished and rapidly mounted in the humidified air stream set to the same RH as found in the crystallization droplet. This stable RH value was estimated by mounting an empty loop of either crystallization solution or cryo-protectant and noting the value which maintained a stable droplet diameter as the RH was varied. The RH was then changed and left to equilibrate before collecting a series of diffraction images to evaluate the changes in diffraction quality.

A complete data set was collected from crystals of full-length GlnN by Dr. Hassan Belrhali with assistance from Joachim Fleury. Data (2.9 Å) were collected at 100 K from a crystal, preserved in Paratone-N and flash-cooled in liquid nitrogen, using 0.95373 Å wavelength X-rays produced by the synchrotron operating in 7/8 + 1 filling mode (200 mA intensity). A complete data set was collected from 360 frames of 1° oscillations (4 passes) exposed for 120 seconds with a crystal-to-detector distance of 324.8 mm.

Collection of data from HA derivative crystals was carried out at ID14-2 (ESRF) equipped in standard configuration with a mini-diffractometer, SC3 robotic sample changer, Oxford Cryostream, and Quantum 4 CCD detector (Area Detector Systems Corporation). Data from a crystal of the digested GlnN form (0.1 mm), derivatized in Ta_6Br_{12} for 30 minutes and cryo-preserved in Paratone-N oil, were collected at 100 K over 360 frames with 1° oscillations (3 passes per oscillation) with 0.933 Å wavelength X-rays and a crystal-to-detector distance of 282.5401 mm. The synchrotron was operating in 4 × 10 mA filling mode and exposures were 90 seconds in duration.

3.3.2 Structure determination and refinement

Data processing and reduction

Integration of the $C222_1$ native and derivative data sets was carried out using iMOSFLM [Leslie 2006] and space group assignments were also evaluated in POINTLESS [Evans

2005]. Scaling, reduction, and merging of diffraction data was carried out using SCALA [Evans 2005] within the CCP4 suite [Collaborative Computational Project 1994]. Diffraction images for publication were converted to JPEG format in MOSFLM. For the P1 dataset, the initial space group assignment and evaluation of the data quality was carried out in the same manner, however, the final integration, scaling and merging was performed using XDS [Kabsch 2010].

Initial data analysis

Cell content analysis was carried out in MATTHEWS_COEF (CCP4) and self-rotation functions were calculated in MOLREP [Vagin *et al.* 1997].

SAD phasing

The Autosol script [Zwart *et al.* 2008] within the python wrapper PHENIX [Adams *et al.* 2010] together with the programs HYSS [Grosse-Kunstleve *et al.* 2003], RESOLVE [Terwilliger 2000], and PHASER [McCoy *et al.* 2007] were used to solve the initial structure of the single anomalous dispersion (SAD) data set from the Ta_6Br_{12} derivatized $C222_1$ crystal form. Direct methods were used by HYSS to find the positions of HA sites using data from 50-7 Å with *f'* = - 2.73878 and *f''* = 10.2589. PHENIX.Xtriage [Zwart *et al.* 2005] was used to investigate the anomalous signal statistics.

Phase extension by density modification with NCS averaging

Phase extension by iterative NCS averaging was carried out with a set of scripts, kindly provided by Dr. Jorge Navaza, which called on the Uppsala Software Factory package RAVE [Kleywegt *et al.* 2006] and the CCP4 suites of programs to carry out the multiple steps involved.

An all-atom model of the GlnN dodecamer, generated from the cryo-EM volume with the best matching handedness (Chapter 2) using SITUS [Wriggers *et al.* 1999] was manually placed into the initial SAD-phased electron density map and a mask was calculated in MAMA (USF RAVE). The initial NCS parameters were obtained from this mask to improve the signal-to-noise ratio during the auto-correlation search by the SUPERMAP program [A. Volbeda,

unpublished work] using AMORE [Trapani *et al.* 2008]. The map was then averaged with AVE using these parameters and the mask updated with COMA and MAPMAN (all from USF RAVE).

Several hundred rounds single crystal averaging with solvent flattening were then carried out using AVEREST which called on various CCP4 and USF programs. Following this density modification, NCS operators were improved using IMP (USF RAVE). An improved mask was then calculated from the output averaged map and the density modification cycle was repeated.

The resulting 7 Å phases from the averaged map were combined with the structure factors from the higher resolution native $C222_1$ dataset to generate an electron density map by Fourier synthesis in CCP4. Phase extension was carried out from 7 – 3.5 Å, in 470 steps each with 10 cycles of averaging, including an additional 5% of reflections at each resolution step, and re-optimizing NCS operators every 1 Å and before the final resolution was reached. A fraction (5%) of the reflections resulting from this phase extension process was set aside as a cross-validation data set.

Model building

Initial model building was carried out automatically in Buccaneer [Cowtan 2006] using the averaged electron density map output at the end of the phase extension procedure. A maximum likelihood target function, incorporating phase information in the form of bimodal probability distributions, was utilized over the three cycles of model building. Manual editing, chain tracing, and fragment extension were then carried out in MIFit [McRee *et al.* 2008]. The identification and connection of discontinuous density was aided by comparisons with GSI crystal structures (see Table 1 below) taking into account the multiple sequence alignments (MSA) produced by van Rooyen [2004] and secondary structure predictions from PSIPRED [McGuffin *et al.* 2000]. Only one of the six monomers present in the asymmetric unit was modelled.

The final GnN monomer structure contained 9 cysteines. However, none of these were modeled as interacting via a disulphide bridge because the closest distance between potential disulphide bonding partner Cα atoms was ~9 Å.

Refinement

Prior to refinement, the geometry of the initial model was checked both in CNS [Brunger 2007] and in Coot [Emsley *et al.* 2004], and any significant distortions were corrected in the latter program using real space refinement. During real space refinement: planar peptide restrains were included, the weight of the electron density fit term relative to the geometrical term was set at 20, and where applicable, secondary structure torsion angle restraints were imposed.

Crystallographic refinement was carried out in CNS by NCS-constrained simulated annealing with torsional angle dynamics using data between 500 – 3.5 Å. Bulk solvent and overall anisotropic B-factor corrections were applied, and phase information, in the form of Hendrickson-Lattman coefficients, were included with amplitudes in the maximum likelihood refinement target. Simulated annealing was started at a temperature of 5000 K and carried out with a slow cooling rate of 50 K per cycle. Energy minimization, taking into account the X-ray term, was carried out before (200 rounds) and after simulated annealing (100 rounds), and minimization, to effect geometric regularization (100 rounds), was also carried out before torsional molecular dynamics. The weights relating the X-ray diffraction terms to the stereochemical restraints terms in the refinement target were determined automatically in CNS. Grouped temperature factor refinement (30 cycles) was then performed on the refined coordinates treating the main chain atoms and side chain atoms as separate groups. The cross-validation data set (5%), assigned in CCP4, was excluded from the refinement.

Missing density including the ligands ADP and MetSox, together with a Mg^{2+} ion, visible in cross-validated, σA weighted, phase-combined 2*Fo-Fc* maps was manually built in Coot. The initial position and configuration of the active site ligands and ions prior to refinement were initially modeled exactly on their counterparts in the GSII structure, *2d3a* [See Table 1]. The parameterization and topology files for refinement of these ligands were generated by PRODRG [Schuettelkopf *et al.* 2004]. This process of manual model building and real-space refinement, followed by torsion angle simulated annealing refinement was iterated several times until no further improvement in the match between the model and the electron density could be achieved, as judged by the R_{free} statistics and the appearance of difference maps.

Refinement optimizations

Several improvements where subsequently made to the original refinement scheme. Where refinement was repeated, the best model from the previous refinement approach served as the starting point, with crystallographic refinement alternating with real-space inspection, manipulation, and refinement in Coot.

(1) The NCS operators, initially determined during the phase extension process, were re-optimized by carrying out rigid body refinement in CNS. A hexameric model was generated from the refined monomer model using XPAND (USF X-UTIL) [Kleywegt *et al.* 2006] and each subunit was treated as a rigid body during 2 cycles (20 minimization steps) of maximum likelihood based refinement incorporating experimental phase information in CNS. These updated operators were used for all subsequent NCS constrained refinement steps.

(2) Additional data, to the limit of 3.0 Å were included in the refinement procedure. This was accomplished by reindexing the original data to higher resolution and scaling and merging as described above. A new cross-validation dataset, comprising 10% of the total number of reflections in this higher resolution dataset, was then selected in thin shells from 60 – 3.0 Å using DATMAN [Kleywegt *et al.* 2006]. The best monomer model of GlnN was then perturbed using 100 rounds of simulated annealing torsion-angle molecular dynamics at a constant temperature of 500 K, without any contribution of the X-ray term to the refinement target [Fabiola *et al.* 2006]. This was done to avoid any bias resulting from overlapping test and work data in the new and old datasets used for refinement. Simulated annealing refinement was then repeated as described above with the X-ray term "on", and followed by grouped B-factor refinement.

(3) Phase extension from 7 – 3.0 Å was repeated in CNS using density modification with NCS averaging to produce phase probability distributions in the form of Hendrickson-Lattman coefficients rather than the bimodal distributions used initially. The first values for these coefficients were derived from the phase extended centroid phases and figures-of-merit by application of a B-factor (30) and scale factor (0.75). As before, the starting phases in the form of HL coefficients were taken from the best

RESOLVE map (7 Å), and the structure factors were taken from the native digested $C222_1$ data set (3.0 Å). Phase extension was carried out in 200 cycles of density modification by density trunctation, solvent flipping, and NCS averaging. The NCS averaging mask was calculated in MAMA (USF RAVE) from a refined monomer and the solvent mask was calculated from the model and shrunk automatically during phase extension. These improved phase probability estimations were included in all subsequent refinements and map calculations.

(4) The final positions of the ions and ligands in the active-site were reconsidered after examination of cross-validated, σA weighted, phased combined 2*Fo-Fc* electron density maps which took into account the updated NCS parameters and higher resolution data. An additional Mg^{2+} ion was modeled in the active site, and to accommodate this change, the positions of the MetSox and ADP ligands were restored to their positions in the *2d3a* GSII structure [see Table 1]. Grouped B-factor refinement was carried out first, treating each ligand or ion as an individual residue, followed by 200 rounds of conjugate gradient coordinate minimization refinement in CNS.

HA binding analysis

Anomalous Fourier difference maps were calculated in CCP4 from the $C222_1$ Ta_6Br_{12} derivative data with calculated phases derived from the 3.5 Å averaged map.

Validation

Geometric distortions and cis-peptides were all identified and corrected in Coot. Coordinate error estimations were taken from the model statistics analysis carried out in CNS. Temperature factor analysis and real space map vs. model correlation analyses were also carried out in CNS. RAMPAGE (CCP4) [Lovell *et al.* 2003] was used to calculate Ramachandran plots from the final refined GlnN monomer model and MOLMAN2 (USF X-UTIL), accessed via the STAN server, was used to calculate a Cα–Ramachandran [Kleywegt 1997].

Molecular replacement

The crystal structure of the full-length P1 crystal form was solved by molecular replacement. A dodecameric search model was generated from the $C222_1$ crystal structure, the refined NCS operators, and the space group symmetry operators, using the USF programs XPAND and MOLEMAN2 (USF X-UTIL). Molecular replacement was carried out in MOLREP (CCP4) [Vagin *et al.* 1997] and the most promising solution was subjected to rigid body refinement in CNS, treating each of the 12 protomers as a rigid body. This maximum likelihood refinement was carried out over 10 cycles with 20 rounds of minimization, each with overall anisotropic B-factor and bulk-solvent corrections.

To avoid potential bias from the molecular replacement solution and to rapidly evaluate the quality of the P1 dataset, phase extension by density modification with NCS averaging, was carried out in CNS as described above for the $C222_1$ crystal structure. A low-resolution subset of phases (up to 15 Å) taken from the rigid-body-refined model served as the starting point for phase extension. Phase extension from 15 - 3 Å was carried out in CNS, with bulk solvent correction, over 500 rounds of density truncation, solvent flipping, and NCS averaging.

3.3.3 Structure analysis

Structure-derived multiple sequence alignments

Multiple sequence alignments (MSAs) were derived from structural superpositions of the GlnN $C222_1$ crystal structure with structures from the GS type I (pdbid: *2gls* chain A) and II (2d3a:A) families [see Table 1 below]. N and C-terminal domains from each structure were aligned separately using the MATCHMAKER feature in UCSF CHIMERA [Pettersen *et al.* 2004; Meng *et al.* 2006]. The Smith-Waterman alignment algorithm with the Blosum 62 substitution scoring matrix and a secondary structure scoring term, weighted at 30% relative to the residue similarity term, was used together with iterative pruning of atom pairs greater than 6 Å to produce the structural alignment. The default secondary structure scoring parameters and gap extension parameters were utilized. A MSA was then derived from the aligned structures using the MATCHALIGN [Meng *et al.* 2006] feature with an inter-residue cut-off value of 6 Å. The resulting N and C-terminal domain alignments were then manually merged.

Representative sequence profiles for each GS family were then manually added to the MSA to create a larger structure-based GS superfamily profile. Details of the sequences used to compile this MSA are given in van Rooyen [2004]. These profiles were previously aligned in CLUTALW [Larkin *et al.* 2007] and the intra-profile sequence alignment was constrained during the overall alignment.

The fit between this initial structure-based profile and the three aligned structures was then optimized by manual inspection and realignment of secondary structure elements (SSEs) in 3D space to account for the significant divergence between the families. Briefly, the sequences corresponding to SSEs were each in turn used as a reference to superimpose the three structures using MATCHALIGN. Regions that clearly displayed similar topologies and SSEs, but were not well matched in the 3D space, were then locally realigned using the same MATCHMAKER procedure by only considering the sequence belonging to the SSEs. Equally, regions that showed clear similarity in 3D space but poor alignment in sequence space were realigned using CLUTALW. Where it was clear that no relationship existed between sequences within loop regions the alignments were manually edited to reflect this. The MSAs were prepared for presentation in the GENEDOC package [Nicholas *et al.* 1997]. Shading according to sequence conservation, manual annotation of important residues, and block diagram summaries were all performed in this program.

Structural alignments

The alignment of GlnN to all of the current publicly available GS crystal structures was achieved in MULTIPROT [Shatsky *et al.* 2004]. In total 23 structures were aligned with GlnN: 9 GSI structures (*2gls, f1h, 2bvc, 1f52, 1fpy, 1hto, 1lgr, 2lgs, 1htq*) plus an additional 8 multi-copy models from the last structure, and 6 GSII structures (*2d3a, 2d3b, 2d3c, 2ojw, 2qc8, 2uu7*) were included in the alignment [see Table 1]. The A chains of all the GSI and GSII structures were all split into N and C-terminal domains which were then automatically aligned in MULTIPROT. The N and C-terminal domains of the GlnN monomer were then separately aligned to these structural superpositions using the MSA described earlier. The conservation of residues within this MSA was also mapped to the structure in CHIMERA.

For the superposition of the entire oligomeric assemblies of GlnN and GSI, one GSI dodecamer was first aligned to the top ring of GlnN, considering only the A chain from each

molecule, using the MATCHALIGN feature in CHIMERA. This procedure was then repeated for the alternate ring of the GlnN complex, thus bringing the second GSI dodecamer into alignment.

Table 1: GS structures investigated in this study.

Organism	Structure	pdbid	Reference	Resolution (Å)
S. typhimurium	StyGSI/ADP/2Mn/Tl	*1f1h*	[Gill *et al.* 2001]	2.67
	StyGSI/ADP/2Mn	*1f52*	"	2.49
	StyGSI/phosphinothricin-P /ADP/2Mn	*1fpy*	"	2.89
	StyGSI/AMP/2Mn	*1lgr*	[Liaw *et al.* 1994]	2.79
	StyGSI/2Mn	*2gls*	[Yamashita *et al.* 1989]	3.50
	StyGSI/Glu/2Mn	*2lgs*	[Liaw *et al.* 1993b]	2.80
M. tuberculosis	Mtb/citrate/AMP/1Mn	*1hto*	[Gill *et al.* 2002]	2.40
	Mtb/citrate/AMP/1Mn	*1htq*	"	2.40
	Mtb/Metsox-P/ADP/3Mg	*2bvc*	[Krajewski *et al.* 2005]	2.10
Z. mays	ZmGSI/Metsox-P/ADP/3Mn	*2d3a*	[Unno *et al.* 2006]	2.63
	ZmGSI/AMPPNP/3Mn	*2d3b*	"	3.50
	ZmGSI/phosphinothricin-P /ADP/3Mn	*2d3c*	"	3.81
H. sapiens	HsGSII/ADP/P/4Mn	*2ojw*	[Krajewski *et al.* 2008]	2.05
	HsGSII/Metsox-P/ADP/3Mn	*2qc8*	"	2.68
C. familiaris	CfGSII/1Mg	*2uu7*	"	3.00

Ligand binding residue analysis

The residues responsible for ligand and ion binding in the GlnN structure were compared with those found in the GSI (*2bvc*) and GSII (*2d3a*) crystal structures [Table 1] which contained MetSox and ADP. All residues within 3.5 Å of the ligands and ions present in the active sites of the structures were deemed to interact with the ligands. Additional residues contributing to the binding of ligands were discovered with LIGPLOT [Wallace *et al.* 1995] (running through the PDBSUM web server [Laskowski *et al.* 2001]). Any residue which was found to contribute to the active site was inspected for similar counterparts in the remaining two structures by analysing the aligned structures and their structure-based MSA. If residues were detected in a similar position but not necessarily with similar identities, these residues were then included in the list of interacting resides for later comparison. The percentage

conservation of all the interacting residues was then mapped to the solvent accessible surface in CHIMERA. The corresponding table of residues substitutions for each structure was shaded according to the percentage conservation and displayed with GENEDOC.

Crystal packing analysis

Symmetry related neighbors, within 10 Å of atoms in the asymmetric unit (ASU), were generated from the final refined structures using PYMOL. For the $C222_1$ crystal structure, the ASU was first generated from the refined model coordinates and the NCS operators using USF XPAND (USF X-UTIL). Crystal contacts were then identified in UCSF CHIMERA using the CLASH/CONTACT plug-in with a VDW overlap limit of ≥ -3.5 Å.

Real space rigid body docking

Rigid-body docking of the GlnN crystal structure into the high-pass filtered 16 Å cryo-EM map (Chapter 2) was carried out automatically using the fast Fourier transform (FFT) accelerated, correlation based, algorithm in SITUS [Wriggers *et al.* 1999]. A dodecameric search model was used in the 6D search.

Quaternary interface analysis

Interacting interfaces were identified and selected in CHIMERA according to a distance cut-off of 3.5 Å. The nature of the interacting residues, their contributing surface areas, and positions in helical segments (as revealed by helical wheel plots) were all determined with PDBSUM [Laskowski *et al.* 2001] and marked on the MSA with GENEDOC. The percentage conservation statistics for the MSA were calculated in Jalview [Waterhouse *et al.* 2009] and plotted in SigmaPlot. Interaction distance plots were calculated by the CSU program running on the "Tools for Structure Prediction and Analysis based on Complementarity with Environment" (SPACE) web server [Sobolev *et al.* 2005]. Additional calculations to ascertain the changes in surface area upon oligomerization and crystallization were carried out in AREAIMOL (CCP4) [Lee *et al.* 1971] with a solvent probe radius of 1.4 Å.

Molecular inspection and visualization

All renderings of molecular coordinates and electron density maps were carried out in UCSF CHIMERA. Secondary structure was assigned using the DSSP algorithm in the latter program and PROMOTIF [Hutchinson *et al.* 1996] accessed via the PDBSUM web server. Topology diagrams were drawn by hand in Microsoft® Visio® and the aspects of the layout for the N-terminal domain were inspired by Burroughs *et al.* 2007. Electrostatic surface potentials were calculated using the DELPHI plugin [Rocchia *et al.* 2001] in UCSF CHIMERA.

3.4
Results

Part A: "Structure Solution"

3.4.1 Diffraction studies

Room temperature experiments showed that the two crystal forms diffracted strongly to 7 Å but only weakly to 3.5 Å with the full-length GlnN protein and the digested protein crystallizing in P1 and $C222_1$ space groups respectively. At this temperature, the crystals exhibited radiation sensitivity, only surviving 100 seconds in the beam, and the diffraction quality between crystals was variable but routinely stronger for the $C222_1$ form. The latter crystals also displayed a reversible transition during *in situ* dehydration experiments, whereas, the P1 crystals lost all diffraction upon dehydration below 90% relative humidity (data not shown).

Cryo-protection in standard buffers containing 20% PEG 400 and 25% glycerol resulted in a reduction of the diffraction limit to 7 Å for both crystal forms following flash-cooling in the cryo-stream (data not shown). It was subsequently shown that these crystals were losing order during the freezing process, as they still diffracted to 3.5 Å at RT after soaking in 30% glycerol for 15 minutes. A systematic assessment of cryo-conditions revealed that Paratone-N oil was able to preserve the higher resolution diffraction from $C222_1$ crystals. A complete and highly redundant (7.8) data set was collected to 3.0 Å (Table 2). The data were integrated, scaled, and merged to yield 94 656 unique reflections with an overall R_{merge} of 0.301. Initial phase extension and refinement, however, only included data up to a resolution of 3.5 Å (see below). At this resolution the mean I/σI after merging was 4.6 in the highest resolution shell and data were subsequently included to an extended limit of 3 Å (I/σI 1.6). This was also supported by the point at which the correlation between random data half-sets was equal to 0.5 (data from ScaleA not shown). The overall R_{merge}, however, was predictably increased from 0.2 to 0.3 when the higher resolution cut-off was adopted.

Table 2: Data collection statistics.

	Digested GlnN		Full-length GlnN
	Native	Ta_6Br_{12}	Native
Space group	$C222_1$	$C222_1$	P1
Cell dimensions			
a,b,c (Å)	198.25, 203.96, 234.59	199.20, 204.94, 235.03	141.79,148.30, 159.32
α,β,γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	115.76, 93.65, 115.99
Cell contents analysis			
Unit cell volume ($Å^3$)	9 485 668	9 594 876	2 579 829
Vm ($Å^3$/Da)	2.41	2.44	2.62
Molecules in the ASU	6	6	12
Solvent content (%)	48.99	49.57	53.11
Resolution range (Å)	62.9 – 3.0 (3.2 – 3.0)	142.9 - 4.0 (4.2 - 4.0)	130.0 – 3.0 (3.1 - 3.0)
No. unique reflections	94 656 (13 703)	40 521 (5 864)	195 958 (14 507)
No. observed reflections	737 299 (104 258)	313 622 (45 823)	777 667 (57 569)
Completeness	100 (100)	100 (100)	97.9 (97.5)
Redundancy	7.8 (7.6)	7.7 (7.8)	3.97 (3.97)
I / σI	7.9 (1.6)	13.4 (8.5)	8.92 (1.27)
R_{merge}	0.301 (1.350)	0.140 (0.221)	0.157 (1.160)

Values in parentheses are for the highest resolution shell.

Heavy atom (HA) derivatization was carried out *in situ* and a 5 minute soak in a Ta_6Br_{12} solution was sufficient to detect incorporation both in terms of obviously green crystals and anomalous signal. A full data set was collected from crystals soaked for 30 minutes but the resolution was limited to 4 Å (Figure 5a). The unit cell dimensions of the derivative crystal were very similar but not identical to the native digested $C222_1$ crystal (1% average increase). The complete and highly redundant (7.7) data were integrated, scaled, and merged to yield 40 521 unique reflections with an overall R_{merge} of 0.140. The anomalous data were complete and measured on average 4 times for each Bivoet pair giving a detectable anomalous signal value of 0.0468.

Data could not initially be collected from crystals of the P1 form because of the weak diffraction after cryo-protection but subsequent analysis employing a higher source intensity (200 mA vs 40 mA) and longer exposure times (120 vs 90 seconds) was successful in collecting a full 3 Å dataset (Figure 5b and Table 2). However, the quality of the diffraction at this higher dose was inferior to the data obtained from the $C222_1$ crystals, with tails and

irregular spot profiles occurring towards the upper resolution limit. XDS was, therefore, used to process this data in order to utilize the advantages of 3D profile integration. After scaling and merging 195 958 unique reflections in XDS, the complete (97.9%) and redundant (3.97 fold) dataset yielded an overall R_{merge} of 0.157. The mean I/σ(I) was only 1.27 in the outer resolution shell which extended from 3.08 – 3 .00 Å.

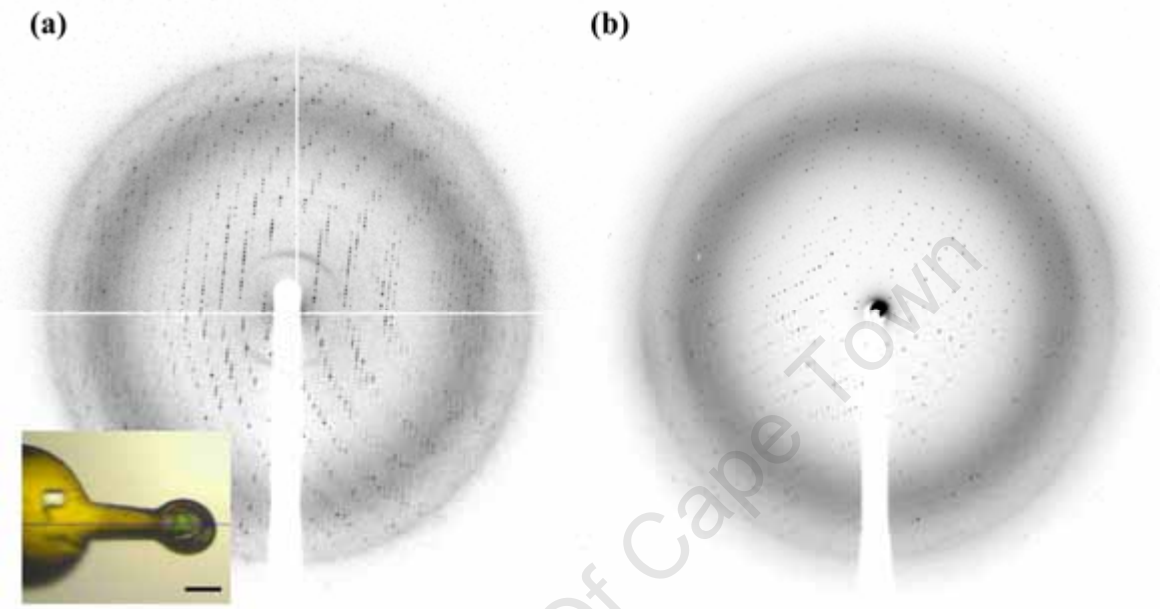


Figure 5: Diffraction images from the **(a)** Ta_6Br_{12} derivative $C222_1$ and **(b)** P1 GlnN crystals. Details of the collection conditions are given in the text. The contrast levels in the diffraction images were adjusted for presentation purposes.

3.4.2 Structure solution

An overview of the structure solution process is given in Figure 6.

Initial Data Analysis

Solvent content analyses predicted 6 monomers of GlnN to be present in the ASU of the $C222_1$ crystal and 12 in the P1 crystal form (Table 2). This agreed with the D6 symmetry revealed by cryo-EM (Chapter 2) and placed a constraint on the translational position of the complex in the $C222_1$ crystal. The double rings of GlnN had to lie on a crystallographic 2-fold axis in order to generate the full biological unit (dodecamer) from the hexameric asymmetric unit contents.

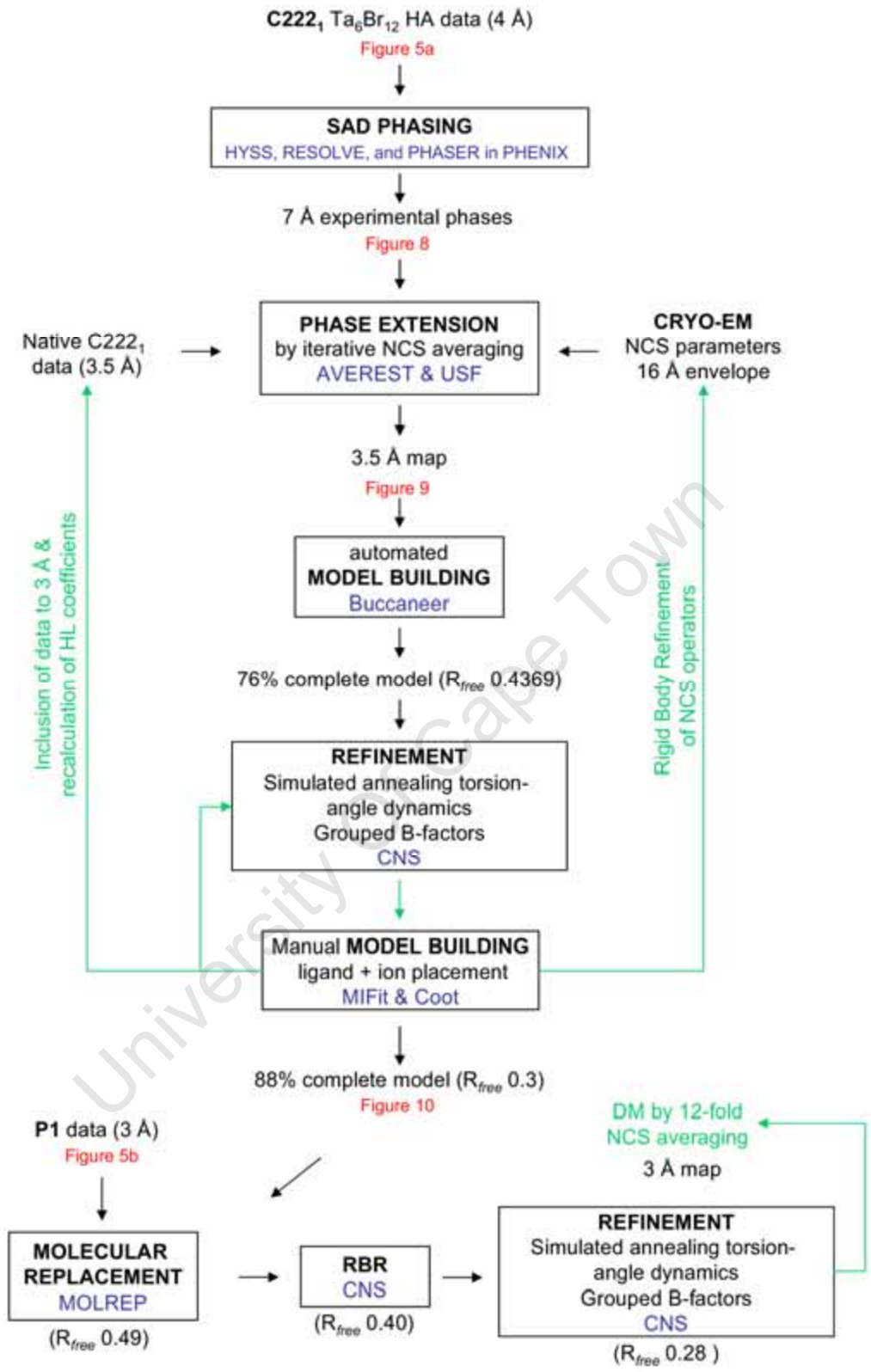


Figure 6: Summary of the GlnN crystal structure solution. Processing steps are outlined by boxes and the software used is highlighted in blue. Subsequent optimizations to the solutions are shown in green and relevant figures are labelled in red. Details and reference for these methods are given in Section 3.3.2.

The 180° section of the SRF calculated from the $C222_1$ native dataset displayed 6 pairs of evenly spaced unique peaks (mirrored about Y). There were also two smaller peaks situated along X at 15° either side of the origin in the 60° section. In the light of the expected 6-fold NCS in the ASU and the knowledge of the D6 symmetry of the GlnN dodecamer from the cryo-EM study (Chapter 2), the SRF was interpreted as arising from the GlnN dodecamer packing in such way that the 6-fold axis was lying in the XZ plane inclined at -15° from Z. This interpretation was supported by the SRF of the previously solved $C222_1$ GSI structure which was similar [pdbid:*2bvc* - Krajewski *et al.* 2005]. The SRF from the P1 crystal data is shown as a comparison (Figure 7).

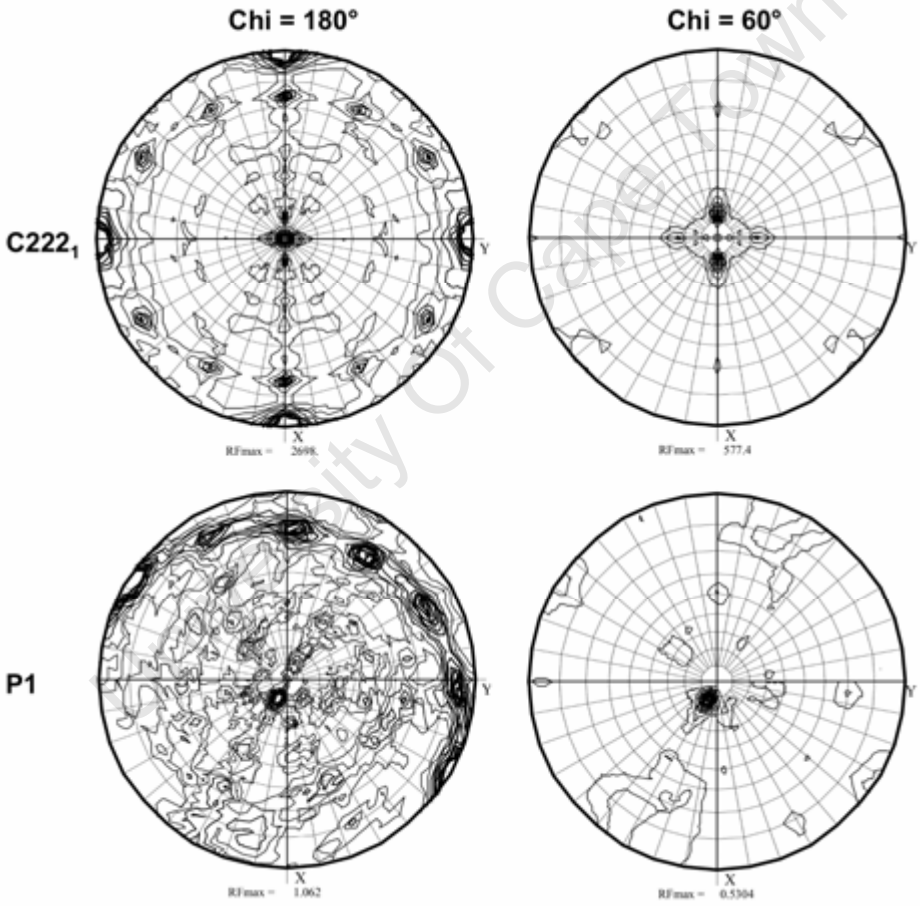


Figure 7: Self-rotation functions (theta, phi, chi) calculated from native data of the digested protein crystal ($C222_1$) and the full-length (P1) crystal forms. These polar plots were calculated by MOLREP [Vagin *et al.* 1997] with theta plotted as the radius and phi as the azimuthal angle.

$C222_1$ Phasing

Only the data from the most promising derivative, Ta_6Br_{12}, were used in the SAD phasing of the $C222_1$ crystal structure (Table 3). It was felt that the phasing power of the HA clusters would be sufficient at low-resolution (7 Å) despite the large number of predicted residues in the ASU. Initially, only 5 Ta sites were searched for and the best resulting solution had a cross-correlation (CC) score of 0.244. PHASER was subsequently able to find and refine an additional 6 sites using Fourier difference maps calculated from the phases output after density modification (DM) in RESOLVE (Table 4).

Table 3: Structure solution statistics for $C222_1$ Ta_6Br_{12} derivative crystal.

SAD phasing statistics	
Anomalous data	
Completeness (%)	100
Multiplicity	4
Signal*	0.0958
No. heavy atom sites	11
Phasing resolution (Å)	51.24 - 7
Total Figure of merit for map-probability phasing	0.76
HA experimental phase information (fraction of phase information)	0.425 (0.3)
DM map-based phase information (fraction of phase information)	0.68 (0.7)
CC for experimental map (DM map)	0.56 (0.80)
R_{factor} for experimental map (DM map)	0.46 (0.26)
NCS overlap	1.01

Phase extension statistics	
Resolution upper limit: starting - final (lower limit) (Å)	7 - 3.5 (100)
Reflections added per step (% of current reflections)	5
Resolution extension steps	47
Cycles of averaging per step	10
NCS	6
NCS operator optimization resolution steps (Å)	7,6,5,4,3.5
Final reliability index from SFALL (starting)	0.199 (0.266)
Final FOM from SigmaA (starting)	0.91317 (0.82184)
Final sigma from FFT (starting)	15.81 (18.280)

* anomalous measurability as defined in AutoSol: The observed anomalous measurability in AutoSol was defined as the fraction of Bijvoet related intensity differences for which the following inequality holds:

|delta_I|/sigma_delta_I > 3.0 min[I(+)/sigma_I(+), I(-)/sigma_I(-)] > 3.0

Table 4: Ta heavy atom sites.

Site no.	x	y	z	occupancy	B-iso
1	0.4085	0.3875	0.1356	1.4402	1.7525
2	0.5105	0.0891	0.1058	0.9962	1.7161
3	0.1319	0.3926	0.8041	1.2738	1.7381
4	0.183	0.0408	0.8194	0.8729	1.717
5	0.0466	0.5000	0.5000	0.9402	1.7088
6	0.2389	0.0196	0.0615	1.1237	1.7197
7	0.2093	0.1322	0.548	0.8457	1.7209
8	0.1474	0.0975	0.5463	0.8523	1.7169
9	0.104	0.0189	0.7931	0.7455	1.7035
10	0.0071	0.114	0.5441	0.4246	1.7025
11	0.0448	0.099	0.5675	0.3734	1.6969

The 6-fold NCS predicted from the solvent content analysis, SRF, and the prior knowledge of the cryo-EM structure was confirmed by RESOLVE which was able to determine the NCS operators from the HA substructure solution. Anomalous difference maps calculated from Ta_6Br_{12} derivative $C222_1$ structure factors and refined phases (see below) revealed both specific symmetrical and non-specific HA binding sites (Figure 8).

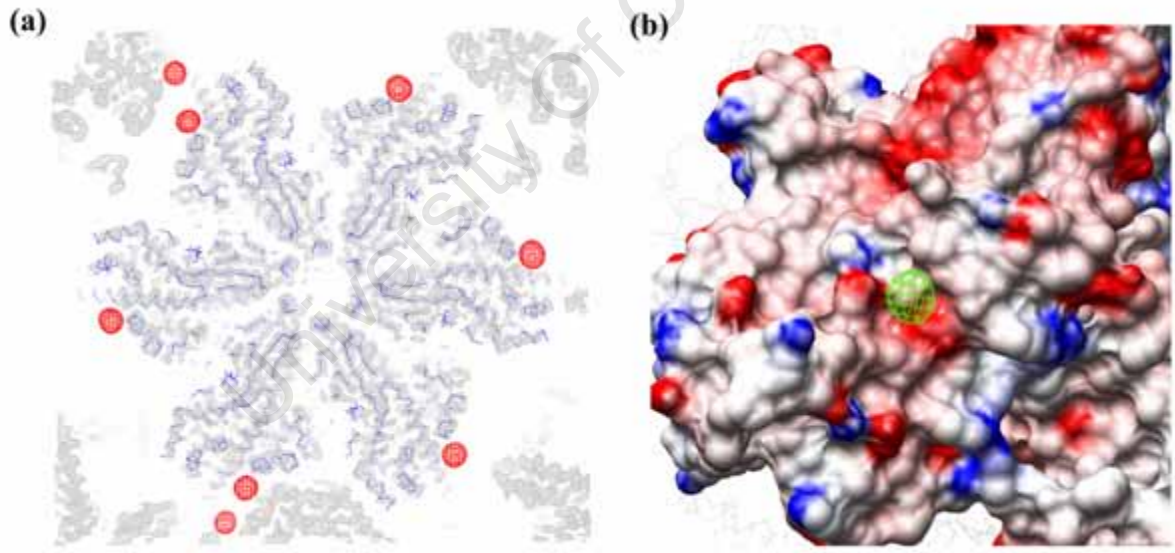


Figure 8: Location of the heavy atom binding sites. **(a)** Anomalous difference map calculated from the Ta_6Br_{12} derivative $C222_1$ structure factors and 3.5 Å NCS extended phases is shown in red. The simple Fourier synthesis averaged map calculated from native data and the same phases is shown in grey wire mesh and the Cα trace of the final refined atomic model is rendered in blue. **(b)** The electrostatic surface corresponding to the specific heavy atom binding site at the tips of the subunit is shown together with the anomalous density corresponding to the Ta_6Br_{12} cluster (green wire frame).

The best solution produced by RESOLVE exhibited a strong agreement between the map calculated with DM-derived phases and the map calculated from HA-derived phases, as demonstrated by the high scores for CC (0.56), R_{factor} (0.46), and NCS overlap (1.01). The FOM score (0.425), however, only had an associated Bayes-CC score of half the value of the other statistics (27 vs 60) and both the total map-probability phasing FOM (0.76) and R_{factor} (0.26) were much improved after DM in RESOLVE. Furthermore, the HA solution only contributed 30% of the overall RESOLVE solution's phase information with the other 70% coming from the DM derived map.

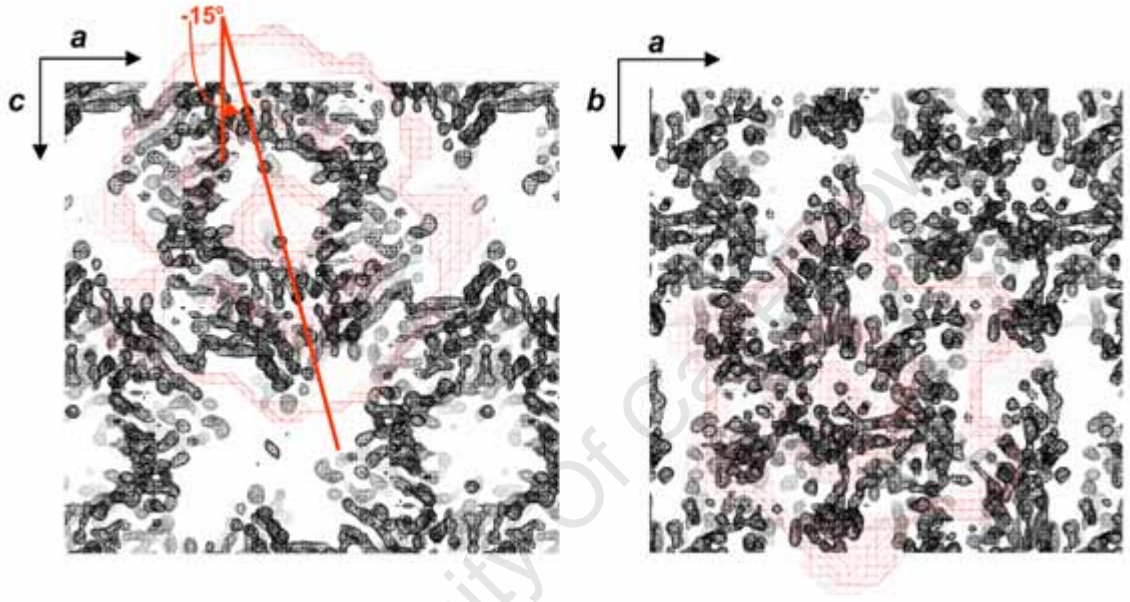


Figure 9: Identification of the GlnN dodecamer in the initial SAD-phased electron density maps. Simple FOM-weighted averaged electron density maps calculated with *Fc* and *PhiCalc* from the best HA substructure solution are shown in black. An outline of the all-atom model used to calculate the initial NCS parameters is shown in red. The orientation of the 6-fold axis of the dodecamer is highlighted for comparison with the predictions made by the SRF.

The GlnN complex was first identified in principal sections of the resulting maps using the prior knowledge of the direction of the 6-fold symmetry axis, which from the SRF described above was expected to lie at -15° from **c** in the **ac** plane, and the known structure of the dodecamer from cryo-EM (Figure 9). Sufficient detail was revealed in these low-resolution (7 Å) discontinuous maps to assign the absolute handedness of the cryo-EM volume. Taking this into account, the NCS parameters derived from the positioned cryo-EM volume were exploited during several hundred rounds of solvent flattening and density modification by 6-

fold NCS averaging to extend the phases to the full resolution of the native data (Table 3 & Figure 10). This approach was very successful and the final averaged map calculated with extended phases displayed obvious secondary structure features and density for amino acid side chains (Figure 10d). The striking improvement in the resolution of the map during density modification (DM) was reflected in the improved statistics comparing the initial phase probability estimates to their final values (Table 3).

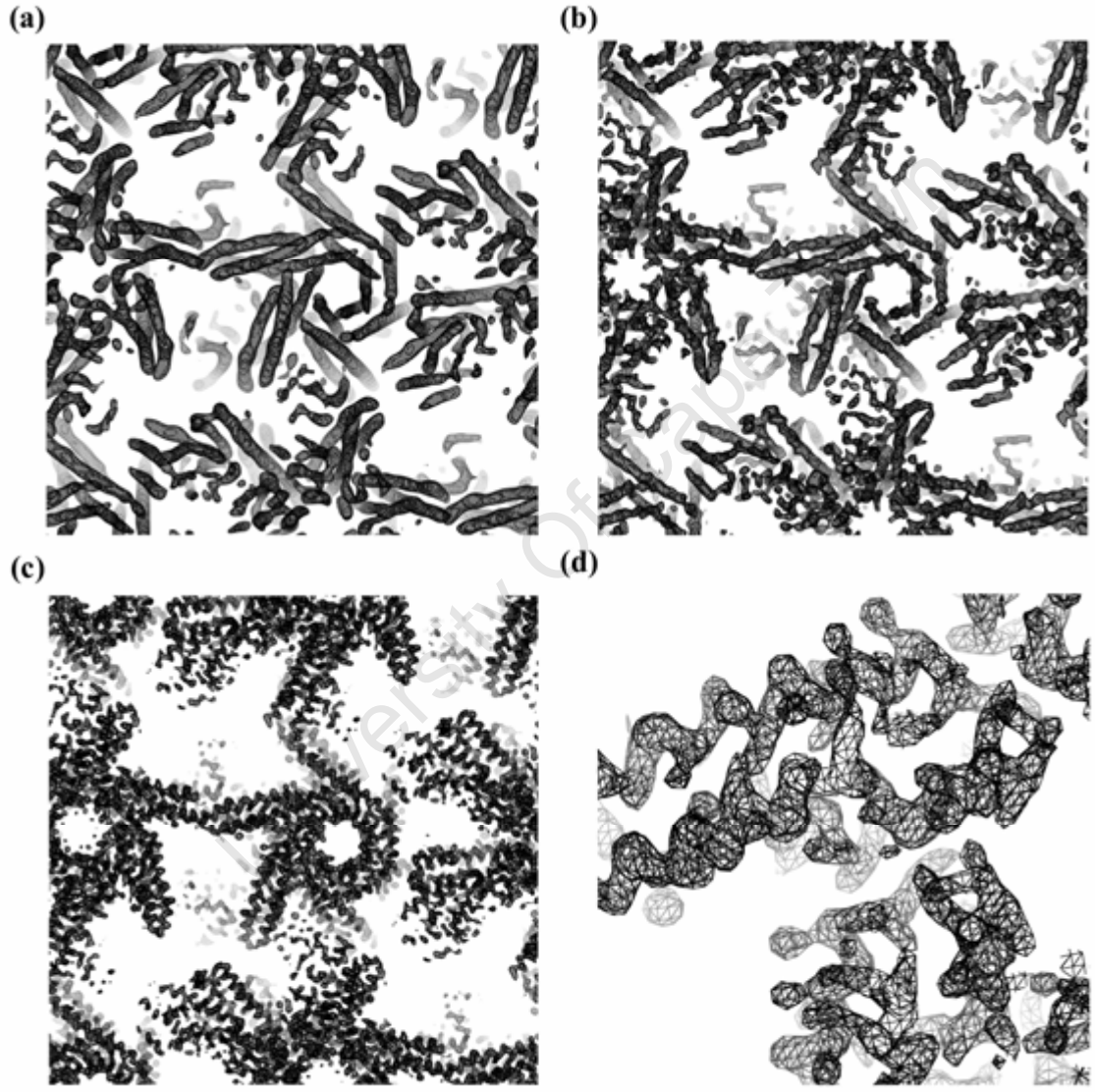


Figure 10: Improvement in the resolution of the map during phase extension by density modification with iterative NCS averaging. Averaged Fourier synthesis maps calculated from **(a)** 7 Å, **(b)** 5 Å, and **(c)** 3.5 Å phases. **(d)** Close-up view of the 3.5 Å averaged map.

$C222_1$ Model building and refinement

The final averaged map calculated from the native structure factors and 3.5 Å phases was of sufficient quality to support model building and refinement. Despite the low-resolution, automated model building succeeded in building 3 336 out of 4 374 (76%) mostly α-helical residues in the ASU. During the following manual model completion several regions of disordered density were identified and the completed structure contained 85% of the predicted GlnN molecule.

The initial simulated annealing torsion angle dynamics refinement of this model followed by grouped temperature factor refinement decreased the R_{free} from 0.4369 to 0.3607. Subsequent

Table 5: Refinement and validation statistics for $C222_1$ structure.

Refinement	
Resolution range	130 – 3.0
Total no. reflections in refinement	94 614
No. reflections in test set	9 465
No. refined parameters:	
Torsion angle dynamics	2 502
Gradient decent coordinate minimization	15 468
Grouped B-factors	1 287
No. atoms	
Protein	5 111
Ligand/ion	45
Water	0
Minimum data/parameter ratio	5.08
Final R_{work}	0.2868
Final R_{free}	0.2993
Model Validation	
Luzzati coordinate error for resolution 5 – 3 (Å)	0.52
rms deviations from target values for:	
Bond distances	0.009140
Bond angles	1.48177
Average isotropic B-factors	
Overall	56.84
Main chain	49.04
Side chain	60.88
Ligand/ion	98.91
SigmaB	25.5

rebuilding of the model and inspection of difference maps revealed density within the active site, which corresponded to the positions of the MetSox, ADP, and divalent cations in the GSI and GSII structure (after alignment). In addition to the missing active site density, inspection of the Fourier difference maps revealed a region of very strong electron density (visible at 16 σ) near the top of the solvent accessible cleft between the GlnN monomers which was modeled as a zinc ion. After several rounds of alternating manual model building, real-space refinement, and simulated annealing refinement, which included higher resolution data (3.0 Å) and optimized NCS operators (see Section 3.3.2 for details), no further change in the final R_{free} of 0.30 was measured (Table 5). A total of 85 149 working reflections were utilized in the refinement of the model to 3.0 Å. With only 2 502 parameters being refined during torsion angle dynamics and 1 287 during grouped B-factor refinement, the data-to-parameter ratio (22:1) was very favourable during the simulated annealing refinement stages. The final electron density map calculated from the refined model was only slightly improved upon in comparison to the map calculated from the phases refined to 3.0 Å by density modification (Figure 11), thus, confirming the quality of the initial phase estimates.

In total the final refined model comprised 640 residues, totalling 88% of the GlnN structure. The majority of the missing and problematic regions occurred in the surface exposed loop regions joining secondary structure elements (Table 6), including the largest region (33 residues) at the tips of the pinwheel arms, which also corresponded to the site of native proteolytic digestion by the *P. fluorescens* protease (Chapter 2). Details of the missing regions, their predicted structure and functions will be given below.

Validation

Considering the significant number of missing regions, the results of standard crystallographic validation procedures were expected to be less than ideal (Table 5). The error in the model coordinates estimated from the Luzzati plot in CNS was 0.52 Å with RMS deviations from ideality in bond lengths of 0.009 and in angles of 1.48. The majority of the geometric errors negatively impacting on these statistics were located on either side of the regions of missing density discussed above. The overall average G-factor output by PROCHECK was 0.14, indicating no unusual stereochemistry.

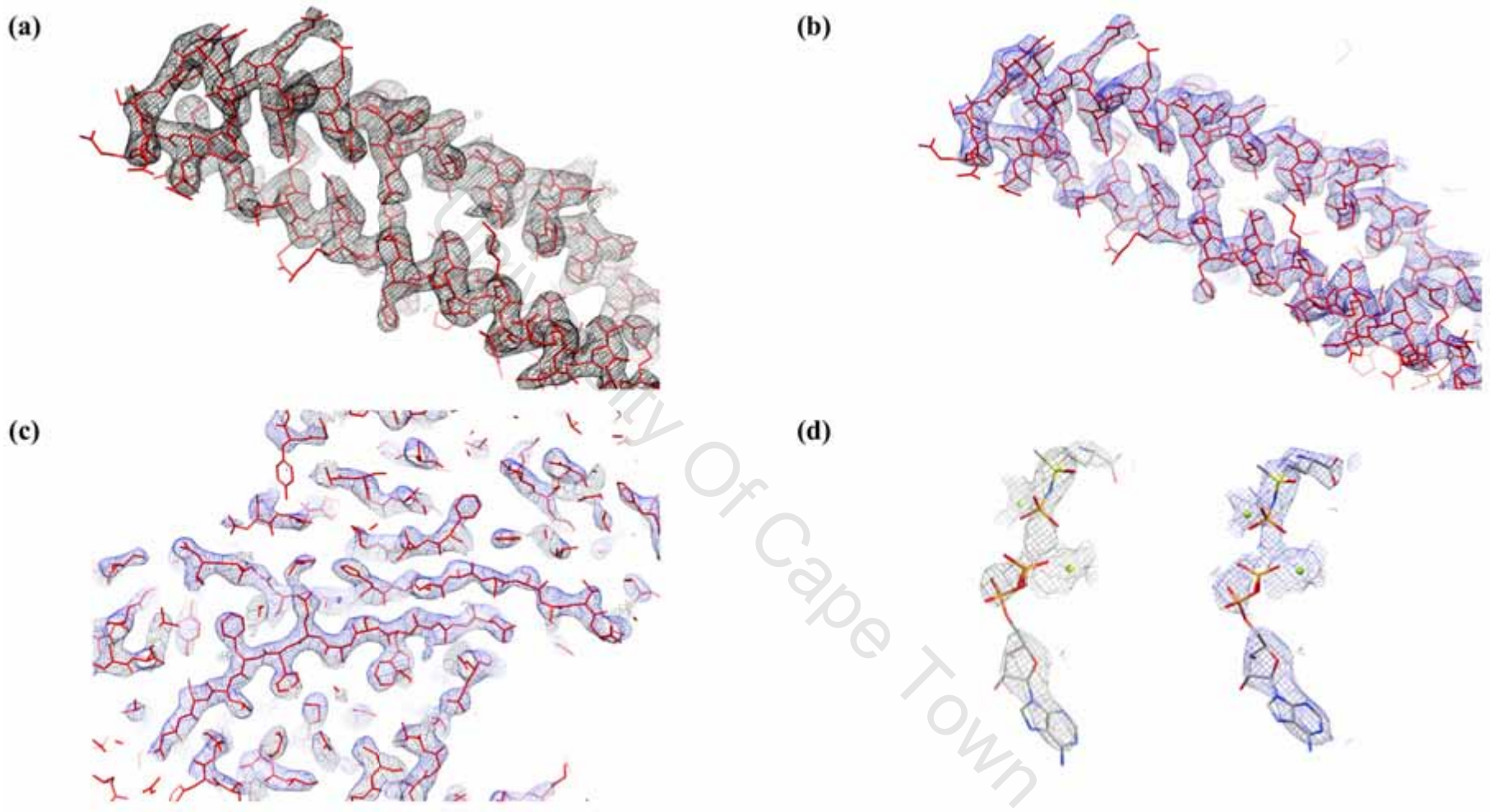
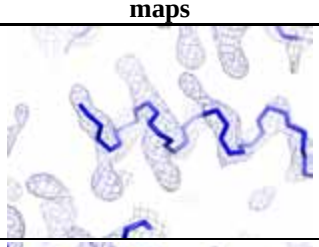
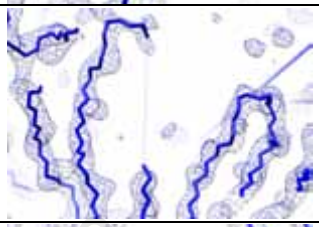
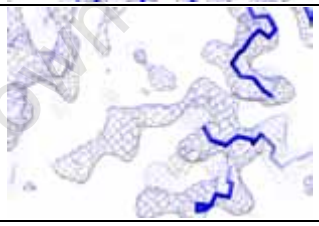
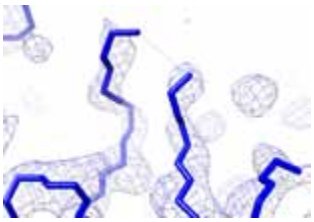
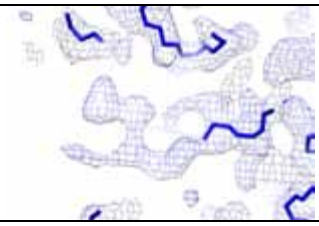
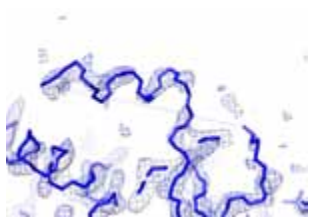
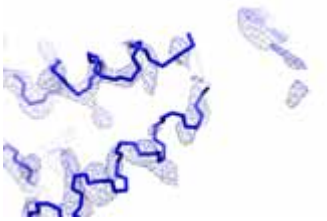
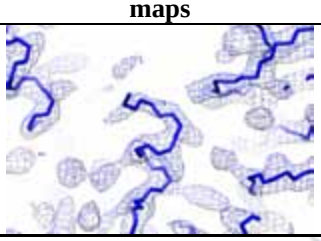
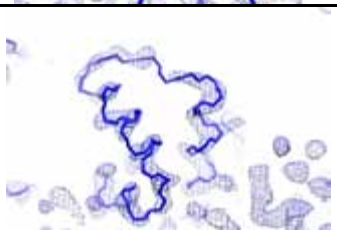
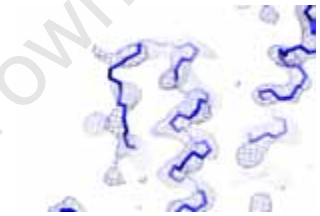
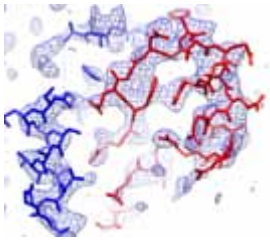


Figure 11: Comparison of map quality before and after model building and refinement. **(a)** Fourier synthesis map calculated from experimental SAD phases refined by density modification employing iterative NCS averaging. The final refined atomic model is shown in red for comparison. **(b)** & **(c)** 2*Fo-Fc* cross-validated, sigma-A weighted, phased combined maps calculated from refined model coordinates. **(d)** Comparison of active-site ligand density. The colouring and interpretation of the maps is the same as for (**a)-(c)**.

Table 6: Problematic regions of the GlnN atomic model*

	No. residues effected	Residue range	Description	cross-validated, σA weighted, phased combined 2*Fo-Fc* maps
Missing regions	2	1-2	N-terminal residues.	
	5	96-100	Partial loop region and β-strand of the active-site β-barrel section of the N-domain.	
	24	124-147	Large section of the N-domain region forming the opposite wall of the active-site β-barrel.	
	4	171-174	Turn region just prior to the end of the N-domain. Some density is present at low contour levels but real space modelling could not trace the chain path. Residues 170 and 175 were modelled as alanines because no side chain density was visible.	
	14	244-257	Extended coil region which exits the centre of the molecule at the solvent channel between pinwheel arms and rejoins a helix near the active site.	
	33	425-457	Region containing the site of native proteolytic cleavage. The partial structure of a helix is visible. Residues 458-461 were modelled as alanines due to the presence of weak side chain density for this region.	
	6	509-514	Loop region linking two surface exposed helices at the tip of the pinwheel arm. At very low contour levels some density is visible but real space modelling was unsuccessful.	

	No. residues effected	Residue range	Description	cross-validated, σA weighted, phased combined 2*Fo-Fc* maps
Missing regions cont.	1	729	The C- terminal residue.	
Poorly defined regions	7	59-65	Loop region connecting two helices at the ring interface.	
	6	200-205	The loop region joining two surface exposed helices at the tip of the pinwheel arm.	
	13	353-365	Extended coil linking an active site β-strand with a core α-helix.	
	31	517-547	Surface exposed loop linking two α-helices (red trace). This region was the last to be modelled and proved to be the most difficult. Although, the main chain fit is reasonable, many of the side chains in this region and not well defined.	

*Individual residues missing side chain density are not discussed in this table. They are dealt with in the real space CC analysis results of the validation section.

It can also be seen in Table 5 that the overall B-factor for the GlnN model was quite high (56.84 Å^2) as too was the variance in the B-factors (σB = 25.5 Å^2). A comparison of the real-space model-map correlations and the average B-factor per residue confirmed that the regions flanking missing density were the largest contributors to these poor statistics in addition to region 517-547 (Figure 12). The latter region, which displayed poor side chain density, comprised a long surface exposed loop joining two helices adjacent to the missing protease susceptibility site. The final model building for this region took into account the homology of

GlnN with the GSI structure *2gls* [Yamashita *et al.* 1989], the predicted secondary structure from PSIPRED [McGuffin *et al.* 2000], and sequence-distance constraints. The weak density for the ligands and ions in the active site were also reflected by the high B-factors (average 98.91 $Å^2$) and real space R-factors.

The Ramachandran plot for the refined GlnN model is shown in Figure 13a. The percentage of residues in the most favoured regions was 83.3% compared to the expected 98%, and 12.2% of all residues were found in allowed regions compared to the expected 2%. A number of outliers (4.5%) were also detected according to this analysis. As a further verification of the validity of the refined model, an analysis of the CA backbone angles in the form of a Cα Ramachandran plot was also calculated (Figure 13b) [Kleywegt 1997]. The percentage of residues in the core allowed regions was 78.33% and the percentage of disallowed residues was 2.39%, both of which were less than one standard deviation from the mean values for these statistics. Of the 14 residues belonging to the latter class, five were in or near one of the regions of poor or missing density.

P1 structure solution

The P1 crystal structure was solved by molecular replacement using the final refined model from the $C222_1$ crystal structure. With the prior knowledge of 12 expected monomers in the asymmetric unit from the solvent content and SRF analysis, the molecular replacement search was carried out with a dodecameric search model and produced a solution with an R_{free} of 0.4947. This solution was refined by rigid-body refinement of all 12 monomers in the ASU to give an R_{free} of 0.3792. Simulated annealing torsion-angle dynamics refinement was then carried out with the imposition of 12-fold NCS constraints. This resulted in a large reduction in the R_{free} value to 0.3111 and grouped temperature factor refinement pushed this even lower to 0.2756. Unexpectedly, no significant improvement in the density of the previously missing regions in the $C222_1$ structure could be seen after the preliminary refinement of the model (data not shown). Of particular significance, was the lack of expected density for the corresponding loop which was digested in the $C222_1$ crystals. Model building was, therefore, attempted to extend the protein chains in these poor regions. However, as with the $C222_1$ structure, alternating cycles of model building and crystallographic refinement did not improve the missing density and neither did density modification by 12-fold NCS averaging.

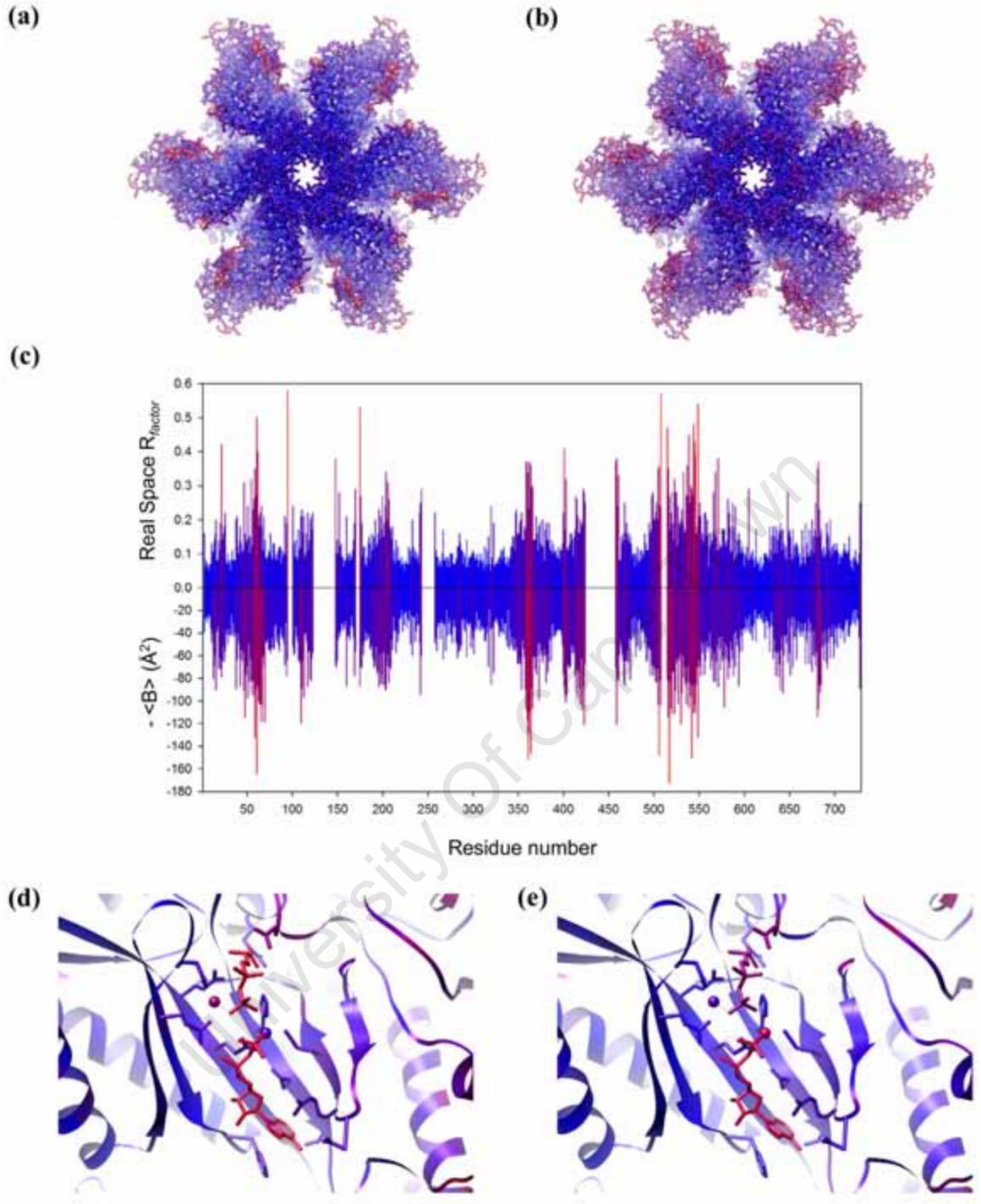


Figure 12: Temperature factor and real space model-map correlation analysis of the refined GlnN model. **(a)** All atom representation of the GlnN ASU coloured according to the real space R_{factor} with a linear gradient between 0 (blue) and 0.57 (red). **(b)** All atom representation of the GlnN ASU coloured according to the average temperature factor with a linear gradient between 0 (blue) and 180 (red) **(c)** Comparison between the model-map real space correlation values and the average temperature factors for each residue of the GlnN model, coloured as in (a) & (b). **(d)** & **(e)** Close-up views of the active site ligands and residues coloured according to the same scheme in (a) and (b) respectively.

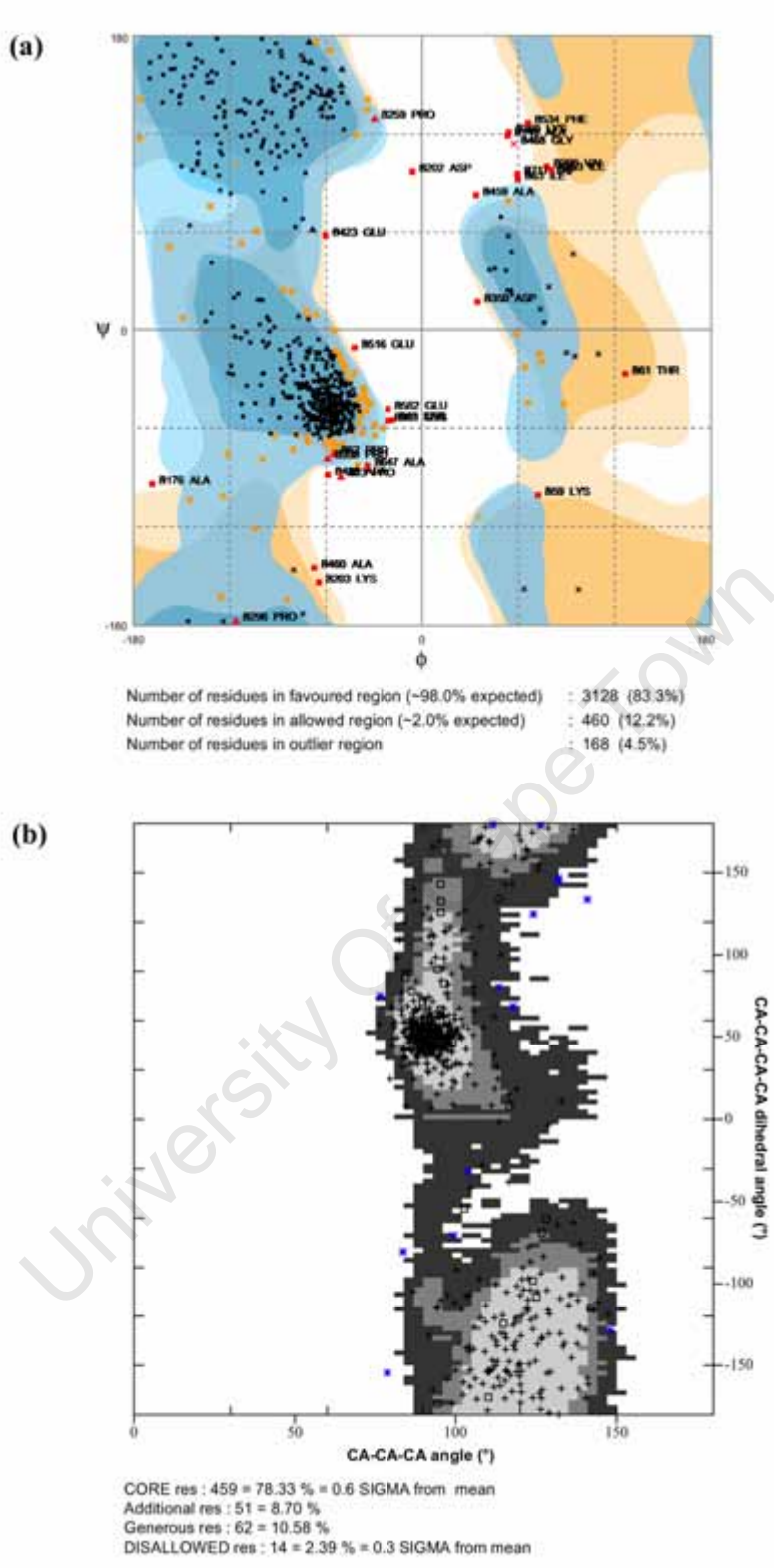


Figure 13: **(a)** Ramachandran plot calculated in RAMPAGE [Lovell *et al.* 2003]. **(b)** Cα-Ramachandran calculated in MOLMAN2 (USF) and adapted from the output of the STAN server **[**Kleywegt 1997]. Non-glycine residues in disallowed regions are marked by blue asterices and glycines are marked by open squares.

3.5
Results

Part B: "Structural Analysis"

3.5.1 GlnN higher order associations

Quaternary structure

The results of the crystal structure solutions confirmed that the double-ringed dodecameric complex, determined in low-resolution by cryo-EM, was the biological unit that crystallized in both the P1 and $C222_1$ space groups, and in so doing, verified the prior assignment of D6 symmetry and the predicted locations of the active site motifs (Chapter 2). The higher resolution of the GlnN crystal structure, however, also allowed the determination of the absolute handedness of the complex, and surprisingly, it was revealed that the hexameric rings forming the GlnN dodecamer were inverted in relation to their orientations in the GSI and GSII structures. Unlike in GSI, where the C-terminal region is primarily responsible for inter-ring association, in GlnN, it is the N-terminal region (Figure 14). It, therefore, appears that despite an overall similar arrangement of subunits within each ring, there are fundamental differences in the residues responsible for the higher order association of the rings in GlnN compared to the other GS families (these novel interfaces are discussed in detail below).

The crystal structure of the GlnN dodecamer displayed a reasonable agreement with the low-resolution cryo-EM reconstruction (Figure 15) but several discrepancies were noted as evidenced by the failure of docking attempts employing single monomer of the GlnN (data not shown). The highest agreement occurred in domed caps at either end of the complex, the pinwheel arms, and the connecting density between the hexameric rings. The areas of strongest disagreement lay between the rings lining either end of the central cavity and the outer surfaces forming the domed caps. These regions appeared hollow instead of comprising solid density as expected from the crystal structure.

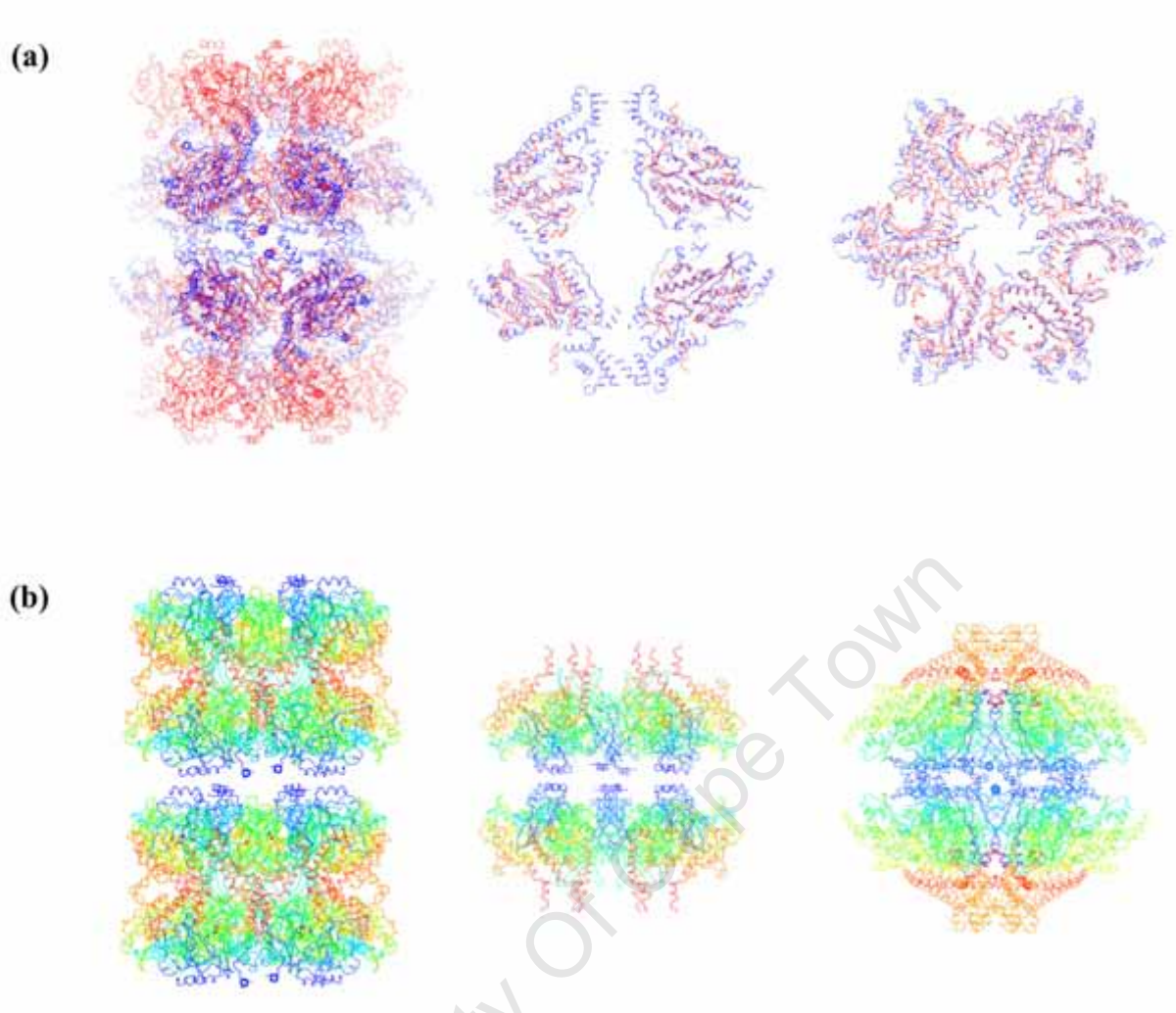


Figure 14: Comparison of the GlnN and GSI [*S. typhimurium* pdbid: *2gls* - Yamashita *et al.* 1989] quaternary structures. **(a)** Structural alignment of GSI (red) and GlnN (blue) rendered as Cα backbone traces. Two GSI dodecamers are present because of the inverted orientation of the rings in the GlnN dodecamer. The outer GSI rings are removed for clarity in the two figures on the right. **(b)** Inversion of the inter-ring interface in GlnN. Structures are shown in Cα backbone representation and coloured in a rainbow scheme according to sequence position. From left to right are the "back-to-back" stacking of two GSI dodecamers, the two inner rings of GSI, and the GlnN dodecamer.

Crystal packing

Analysis of the arrangement of the dodecamers in the two crystal forms revealed substantial differences in the molecular contacts responsible for crystalline packing (Figure 16). The P1 crystal form is made up of layers of GlnN dodecamers arranged in a hexagonal fashion in the **ab** plane. These layers then stack on top of each other in an offset manner with the large flat

surfaces of the GlnN conical domes fitting into each other like saw teeth (Figure 16 P1 **a'c'** and **b'c'** views). In the $C222_1$ crystals, the dodecamers are also arranged in parallel arrays but the packing of the adjacent molecules in the layers is more orthogonal (Figure 16 $C222_1$ **ab** view). Every alternate layer is composed of parallel arrays of molecules with their 6-fold axes tilted in opposite directions from **c** in the **ac** plane (Figure 16 $C222_1$ **ac** view).

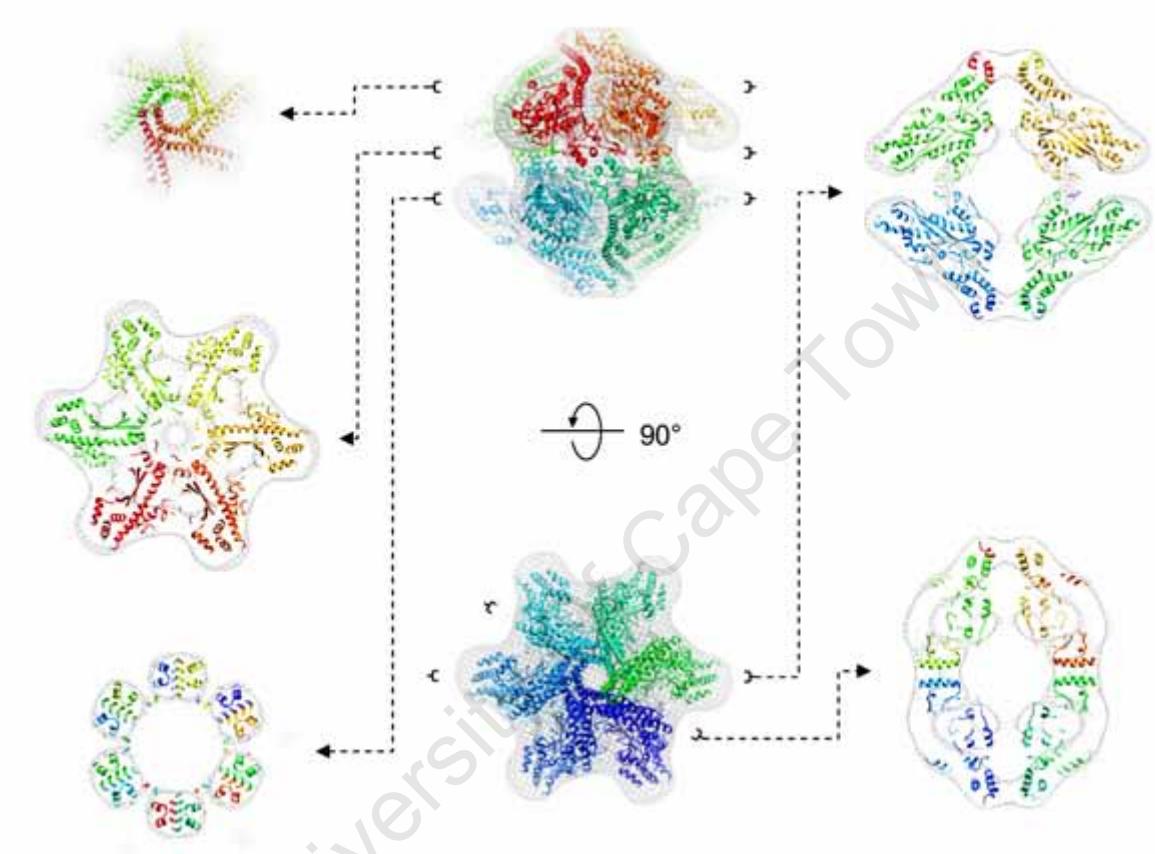


Figure 14: Agreement between the X-ray and cryo-EM structure. The reconstructed cryo-EM volume (16 Å) (Chapter 2) is shown as a mesh representation with the density contoured to enclose 1 341 nm^3. The dodecameric $C222_1$ crystal structure is shown in ribbon representation and coloured according to subunit identity. The locations of the "slices" through the structure are indicated by parentheses and dotted arrows.

It can be seen from these packing analyses that the protease susceptibility site, located at the tips of the pinwheel arms, contributes to crystalline packing through the formation of several crystal contacts with neighbouring dodecamers (Figure 16 - inset). In the P1 crystal, these regions interact primarily with the adjacent dodecamers in the same layer. However, after proteolysis, the nature of the interaction changes resulting in a greater number of contacts

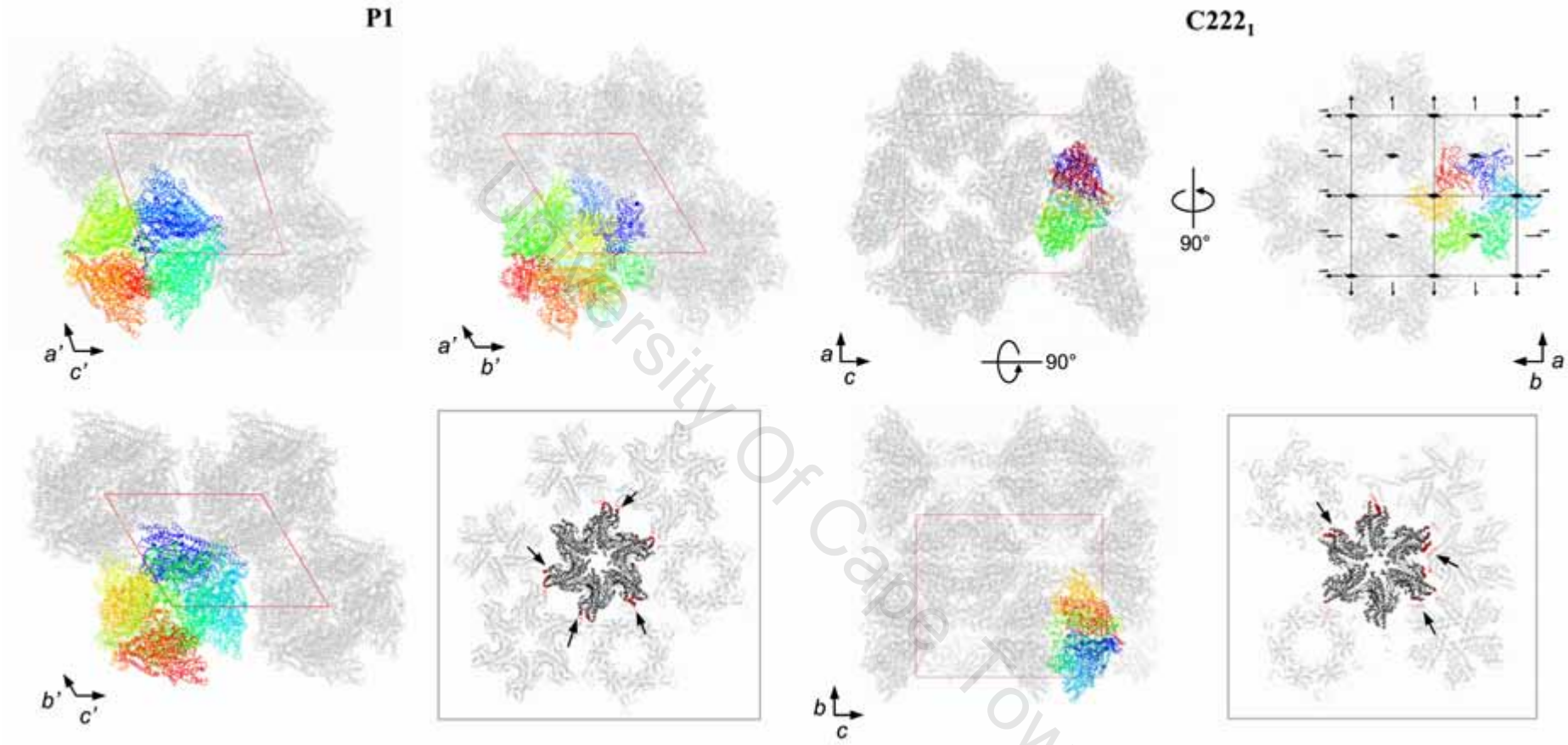


Figure 16: Molecular packing in the two crystal lattices. The unit cell is highlighted in red and the subunits in the biological ASU are highlighted in different colours. **Insets:** Analysis of the crystal contacts. The GlnN ASU is shown in thick Cα-trace representation and the symmetry related molecules within 10 Å are shown in thin Cα-trace representation. Residues from one ring of the dodecamer which approach closer than 3.5 Å to symmetry related neighbouring molecules are highlighted in red. Arrows mark the positions of contacts involving the region of protease susceptibility.

between the molecules of the ASU and adjacent dodecamers. The resulting tilting of the dodecamers in the **ac** plane increases the total surface area upon crystallization from -2 771 $Å^2$ in the P1 crystal form to -5 056 $Å^2$ in the $C222_1$ crystal form (excluding the inter-ring interactions). The predominant interface formed by the large domed surface in both crystals is composed of a central hydrophobic stripe which is surrounded on either side by charged residues (data not shown).

3.5.2 The GlnN monomer structure

Active site conservation and ligand binding

Despite the differences in oligomeric arrangement, the structure of the GlnN monomer displayed remarkable similarity to the other GS structures. This agreement is highest in the active site region (Figure 17) as first predicted by bioinformatic and electron microscopy analyses [van Rooyen 2004]. As in other GS enzymes, the active site β-barrel motif in GlnN is formed by the association of 2 β-strands from the N (residues 1-178) and 6 β-strands from the C-terminal domains (179-729) of adjacent subunits. This similarity was sufficient to bring the remaining more divergent regions into juxtaposition and to produce the first structure based MSAs for the superfamily (Figure 17a & 17b). From these results it can be seen that that despite the low global sequence identity of GlnN to the GSI (9%) and GSII (6%) families, the structures share several highly conserved regions which are either structural components of the active site or extended loops in close proximity to the active site (Figure 17c). The role of the latter loops in catalysis was discussed in Section 3.2.2 and their position is highlighted in Figure 17a.

The degree of conservation is most clear when considering the 6 residues forming the two metal binding sites in the structures, which are conserved both in sequence and structure (Figure 18a). The original 3.0 Å electron density map calculated from unbiased DM phases shows that this conservation of the active site residue configuration is reflected in the position and conformation adopted by the ligands in GlnN when compared to the other structures. All of the residues that were found to interact with the ligands and metal ions in GlnN are shown together with their final refined electron density in Figure 18b.

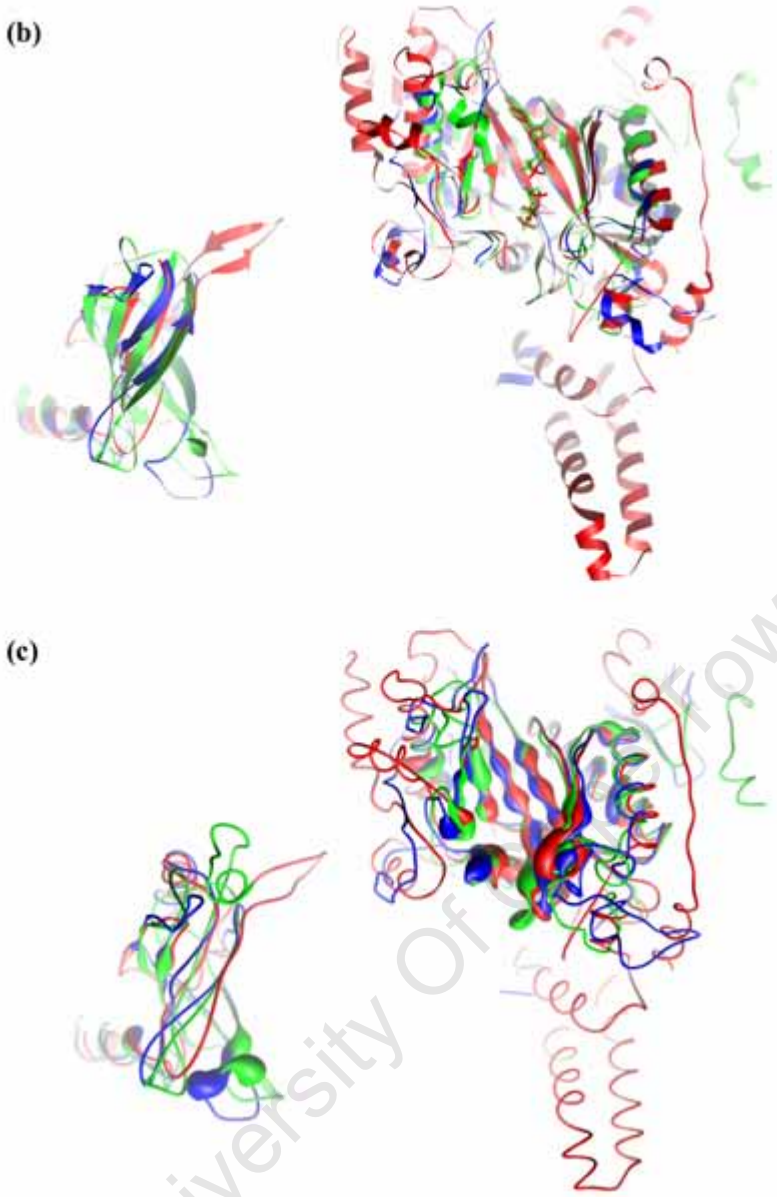


Figure 17: Conservation of sequence and structural features within the GS superfamily **(a)** Structure-derived (see (b)) multiple sequence alignments of representative sequences from the three GS families: GSIII (red), GSI (blue), and GSII (green). Shading represents the conservation of residues within the three GS families and inter-group shading represents conservation of residues across group divisions. Highly conserved residues are shown in black, strongly conserved residues are shown in grey, and the consensus sequence is printed below the multiple sequence alignment. Conserved active site residues, deduced from previously reported alignments of the GSIs and GSIIs [Eisenberg *et al.* 2000] and their atomic structures (Section 3.2) are shown in solid red. Residues outlined by black boxes and labeled a-e in the *S. typhimurium* sequence mark the position the five important loops introduced in section 3.2.2. **(b)** Structural alignments of the N and C-terminal domains from GlnN with those from GSI and GSII [pdbid: *2gls* & *2d3a* – see Table 1] (shown in ribbon representation and coloured as in (a)). **(c)** Mapping of sequence conservation onto the GS structural alignments. The percentage sequence conservation derived from the alignments in (a) is represented by the width of the cartoon backbone from a range of 8 to 100% at the maximum ribbon width.

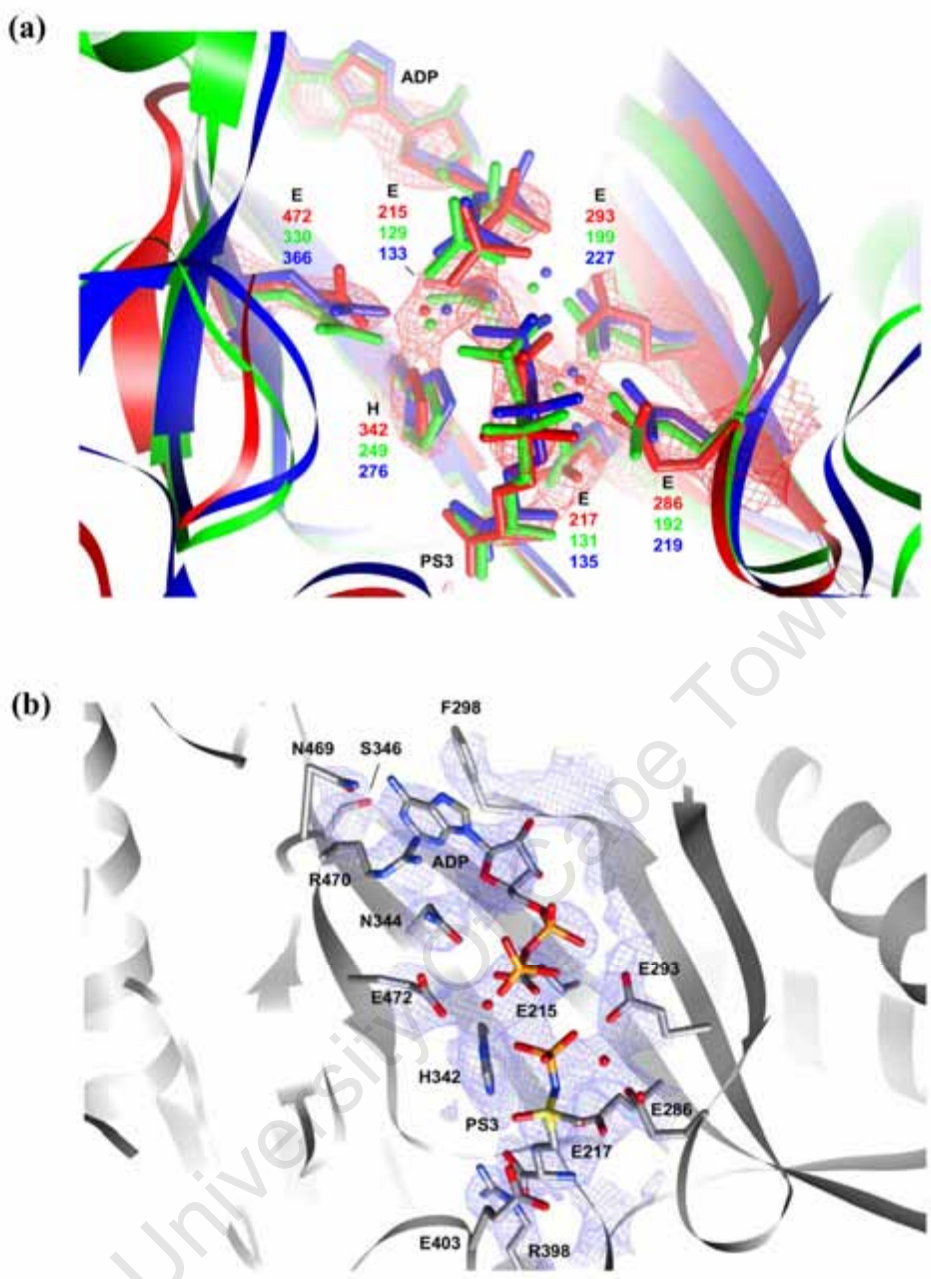


Figure 18: Conservation of ligand-binding residues in the GlnN active site. **(a)** Overlay of GlnN (red) with aligned structures of GSI (pdbid: *2bvc* - blue) and GSII (pdbid: *2d3a* - green) containing MetSox-P and ADP [see Table 1 for references]. Only the highly conserved residues binding the two ligands and metal binding sites are shown in stick representation, whereas, the remaining residues are shown as ribbon traces. Spheres mark the positions of the metal ions and residue labels are given in black with the respective numbering in each structure shown in the appropriate colour. The simple Fourier synthesis electron density map for the GlnN structure calculated from DM phases up to 3.0 Å is shown in red wire frame representation. The view is into the active site β-barrel from the direction of the N-domain, matching that in Figure 1. **(b)** Similar to (a) except that all residues within 3.5 Å of the ligands and ions in the active site are shown in stick representation. The 2*Fo-Fc* cross-validated, σA weighted, phased combined electron density calculated from refined model coordinates is shown in blue wire frame representation.

The absolute conservation of the ligand binding configuration seen around the metal binding sites is mirrored to a lesser extend in all the residues that are found to interact with the MetSox and ADP ligands (Table 7 and Figure 19). In total, 13 residues are conserved in both sequence and structure between all three structures (highlighted in red in Figure 19). The contributions to ligand binding by an additional three residues, which are conserved in sequence and structure in the GSI and GSII models, could not be determined because they occurred in one of the missing regions (the predicted structure of these regions is discussed below). The counterparts of these missing active site residues (bold entries in Table 7): D50, R339, and Y179 were all shown to be critical to GSI catalysis as described in Section 3.2.2. In addition to these residues, poor side chain density prevented the modelling of R474, which in the corresponding position in *S. typhimurium* GSI, R359, stretches into the active site and interacts with the terminal phosphate group of ADP. At very low contour levels weak density is visible for this residue but it points towards the adjacent E472 residue and not towards the ligands. The important E327 and N264 equivalents in GlnN, E403 and N337, were however, conserved in sequence and structure.

The remaining active site residues, which cluster around the ADP ligand, are far less conserved despite occupying similar positions in the structures. Half of these 16 residues are found at distances greater than 3.5 Å from the ligands, but contribute to binding through main chain atoms which form the walls of concave binding pockets (marked with a "**b**" in Figure 19b). The other half of the less conserved active site residues are the most informative as they all interact with the ligands via their side chains and, therefore, represent novel binding modes (underlined in Figure 19a and marked with black arrows in Figure 19b). One of these residues, K281, was in a region of missing density.

Conservation of domain architecture and fold topology

In addition to the obvious conservation of active site residues in the three GS structures, several regions of tertiary structure also share common architectures and topologies (coloured secondary structure elements (SSE) in Figure 20). These more subtle structural conservations, which are not discernible from the sequence conservations alone, are obvious when presented side-by-side and the GlnN protein shares 7 α-helices and 13 β-sheets with the two other structures. This does not take into account the missing regions in the GlnN structure which are predicted to share an additional 2 β-sheets with the other structures (see below). These

Table 7: Active site residues interacting with the metal ions and ligands in GS structures containing MetSox and ADP.

	GSIII (GlnN)	GSII (*2d3a*)	GSI (*2bvc*)	*2gls* equivalents†	Ligand interaction or closest atom
Conserved in structure and sequence	E286	E192	E219	E212	Divalent cation binding site n_1
	E217	E131	E135	E131	
	E293	E199	E227	E220	
	E215	E129	E133	E129	Divalent cation binding site n_2
	E472	E330	E366	E357	
	H342	H249	H276	H269	
	R398	R291	R329	R321	COOH MetSox
	G338	G245	G272	G265	NH_2 MetSox
	E403	E297	E335	**E327**	OE MetSox
	S346	S253	S280	S273	N1ADP
Conserved in GSI & GSII but not visible in GlnN	D129	D56	D54	**D50**	CE of MetSox
	R474	R332	R368	**R359**	O2A PO_4 & SO MetSox
	R460	R311	R347	**R339**	PO_4 of ADP
	Y254	Y158	Y186	**Y179**	CE of MetSox
> 3.5 Å but conserved in sequence and structure	N337	N244	N271	**N264**	NH_2 of MetSox
	G213	G127	G131	G127	O3' sugar ring ADP - hollow pocket
	V287	V193	V220	V213	CE of MetSox
> 3.5 Å & interacting via backbone contacts	I297	G203	Q231	R224	Backbone contact - O2' ADP
	N211	W125	Y129	L125	Backbone pocket - CA and CB interact with nucleotide of ADP near N1
	S339	A246	S273	S266	Backbone contact - OT of MetSox
	W345	Y252	Q279	M272	Backbone contact - C2 of ADP
	A404	T298	A336	A328	Backbone contact - OT of MetSox
	N402	H296	Y334	Y326	Backbone contact - CA of MetSox
	P296	V202	Y230	T223	Backbone contact CO - O3' sugar ring of ADP
Unique residues	T282	G188	K215	A208	K215 interacts with O2A of first ADP PO_4 G188 & T282 are > 4 Å away in GSII and GlnN
	K281	S187	E214	E207	E214 and S187 interact with the sugar ring of ADP near O2' Density is missing in GlnN
	N344	N251	H278	H271	Perpendicular interaction with O4' sugar ring ADP
	R470	Y328	R364	R355	Stacking interaction with nucleotide ring of ADP near N7
	F298	P204	F232	F225	Adenosine ring Pi stacking and mainchain N - O2' of sugar ring
	A465	R316	R352	R344	R forms part of n_2 metal site and interacts with terminal PO_4 of ADP in GSI & II
	A295	N201	N229	A222	N close to first PO_4 of ADP in GSI and II
	N469	G327	K363	R354	In GSI and GSII backbone contact - N6 adenosine ring but N469 - N6 adenosine ring in GlnN
	n/a*	n/a	K361	K352	CO – N6 adenosine ring, main chain

Residues coloured in pink and green interact with the ADP and MetSox ligands respectively.

† The structural equivalents of residues in the StyGSI/Mn structure are given as a reference so that the conservations can be understood in the context of the reaction mechanism explained in Section 3.2.2. The most critical residues involved in this mechanism are highlighted in bold.

*Entries labelled n/a mean that no structural counterpart existed in the structure for that amino acid.

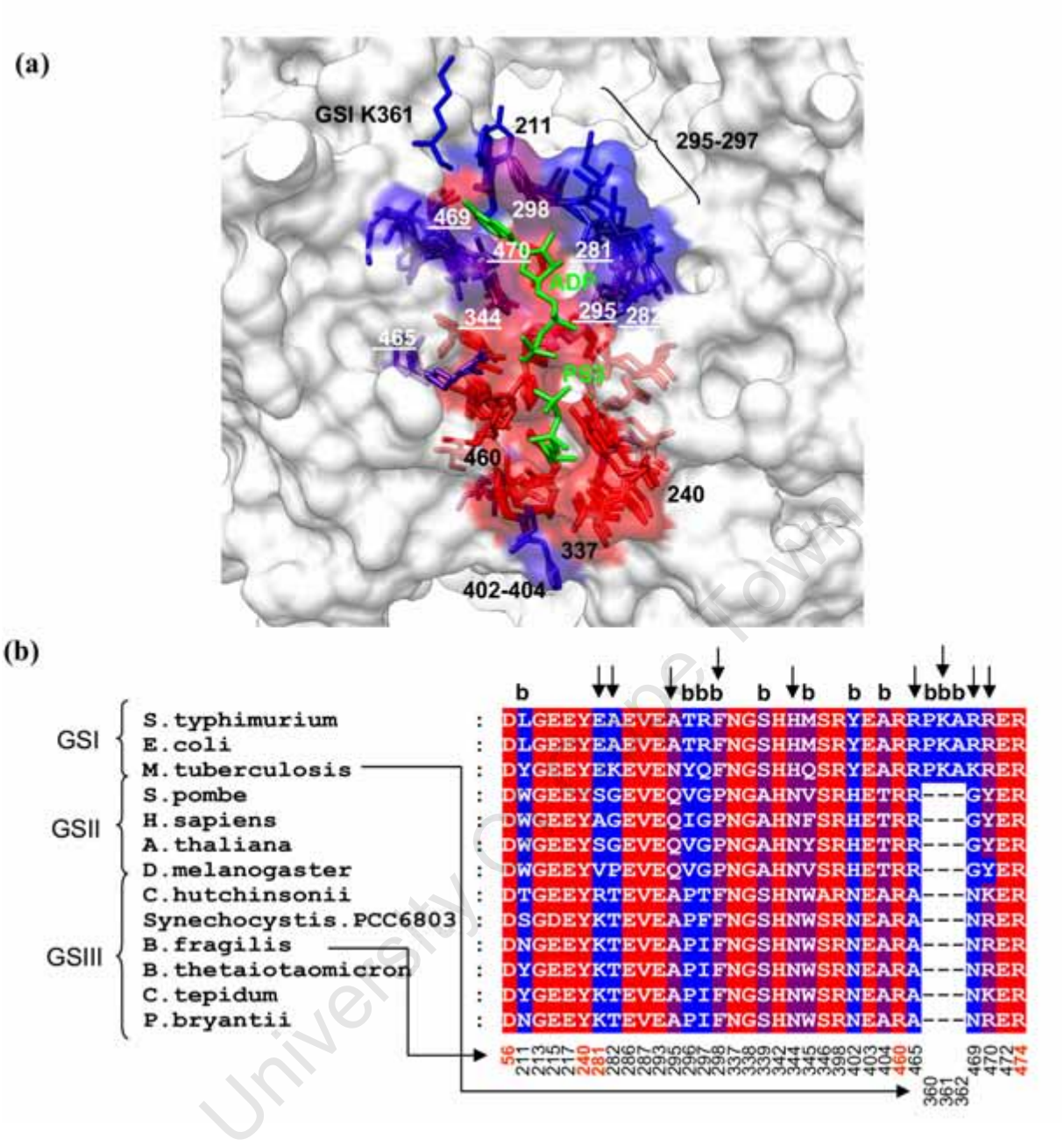


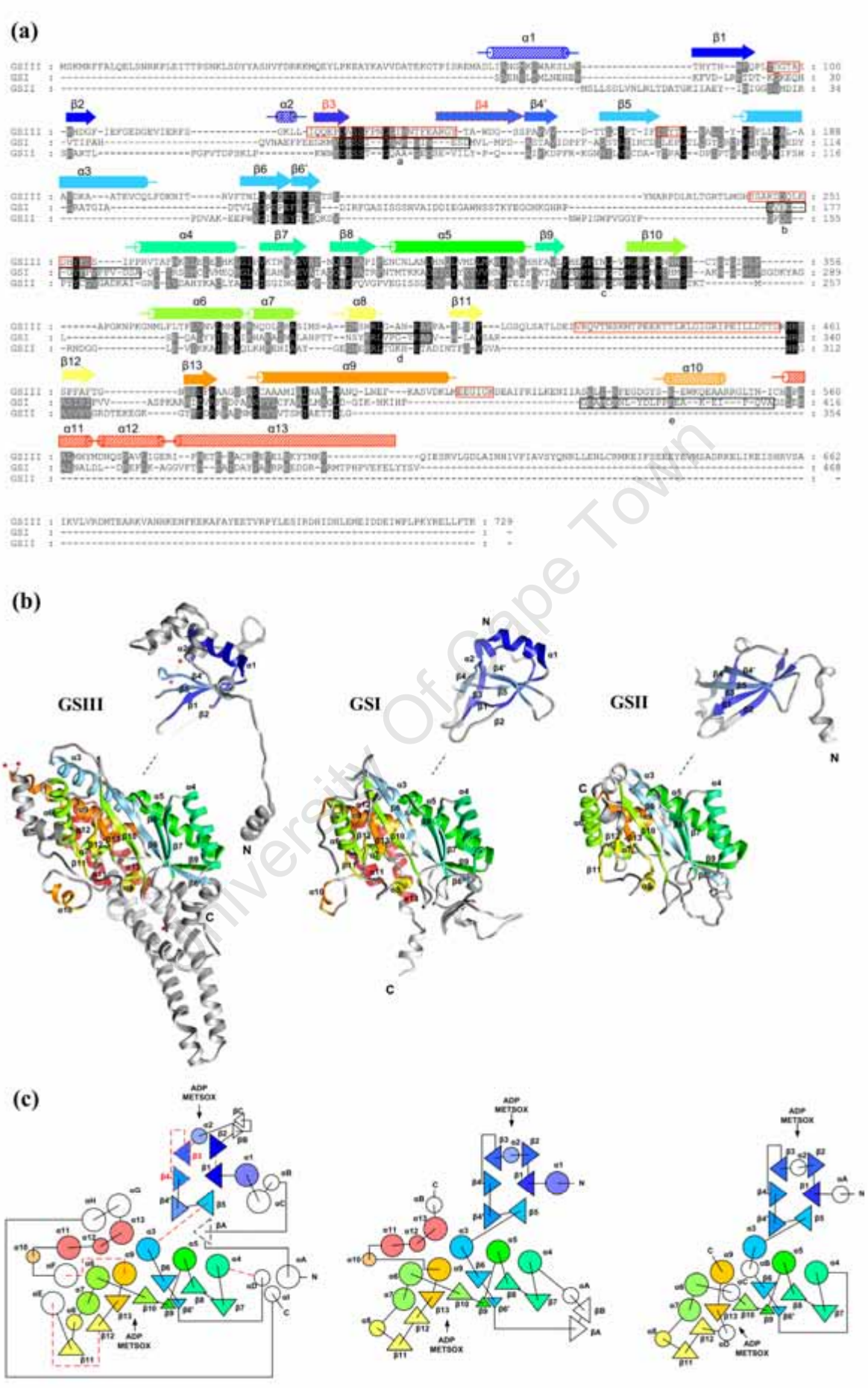
Figure 19: Conservation of residues responsible for ligand binding specificity in the GS superfamily. **(a)** All of the residues that bind MetSox-P (PS3) and ADP and their homologous partners in the 3 GS structures (GSIII – GlnN, GSI - *2bvc*, and GSII - *2d3a*) are shown in stick representation and their percentage conservation is represented by their colouring (linear scale between blue =30% and red = 100% also projected onto the semi-transparent surface). **(b)** MSA corresponding to (a). All active site residues are shown side-by-side and their numbering in the image follows that of the GlnN structure unless there is no counterpart, in which case, the numbering follows *M. tb* GSI. Residues interacting with the ligands via main chain atoms are marked with a "**b**" above the MSA. Unique residues interacting with the ligands via side chain associations are marked by **arrows** above the MSA and their respective labels are underlined in the image. Residues that were not visible or displayed weak side-chain density in the final electron density maps are highlighted in red below the MSA. The view is into the active site β-barrel from the direction of the N-domain matching Figure 1.

conserved SSEs pack together to form the active-site β-barrel and in so doing create the conserved core region in the C-domain. The compact 6-stranded N-terminal β-grasp domain is also well conserved and packs against the C-domain on the side opposite to the active site. An additional 6 α-helices are shared only between the GSI and GSIII structures with no structural counterparts in the GSII family.

Protein flexibility

In order to investigate the flexibility of the missing regions, with a view to predicting their structures, alignments were produced from all known GS atomic models (Figure 21). This superposition shows that of the regions shared by GlnN and the other GS families, the regions of missing density in GlnN were also the most mobile in the other structure solved to date. The most variable loop in the N-domain was the **β3-β4** loop ("latch" in GSI) followed by the **β1-β2** loop. In the C-domain, the **α9-α11** ("adenylylation loop" in GSI) was the most flexible region followed by the adjacent **α8-β11** loop ("E327 flap" loop). The **β6'-α4** loop showed some flexibility near its midpoint but far less in region just prior to the **α4** helix, corresponding to the **"**Y179 loop" in GSI. The **β12-β13** and **β11-β12** regions also displayed considerable flexibility. Further evidence for mobility in these regions comes from the fact that they are missing in several other GS structures. Residues 324-329 (**α8-β11**), 393–407 (**α9-α11**), and 57-64 (**β3-β4**) were not modelled in the GSI structure *2lgs*. Residues 301-307 (**α8-β11**) were not present in the GSII structure *2ojw* and residues 292-301 (**α8-β11)** and 70-73 (**β3-β4)** were missing in the newly determined GSII structure *3fky* [references are given in Table 1]. In the GlnN structure, the flexible "latch", **β1-β2** loop, "adenylylation", and "Y179" loops are all missing but the "flap" and "N264" loop are both visible.

Therefore, although some of missing regions in GlnN are highly mobile in other structures, it is possible to suggest topologies for those regions that demonstrate sequence homology. This is true for the three missing regions in the N-terminal domain (Figure 20d & 21) which likely correspond to the regions: **β1 - β2 (**96-100**), β5 - α3 (**171-174), and **β3-turn-β4 (**124-147). Similarly, the last section of the missing region 244-257 in the C-terminal domain is thought to correspond to the Y179 loop in GSI.



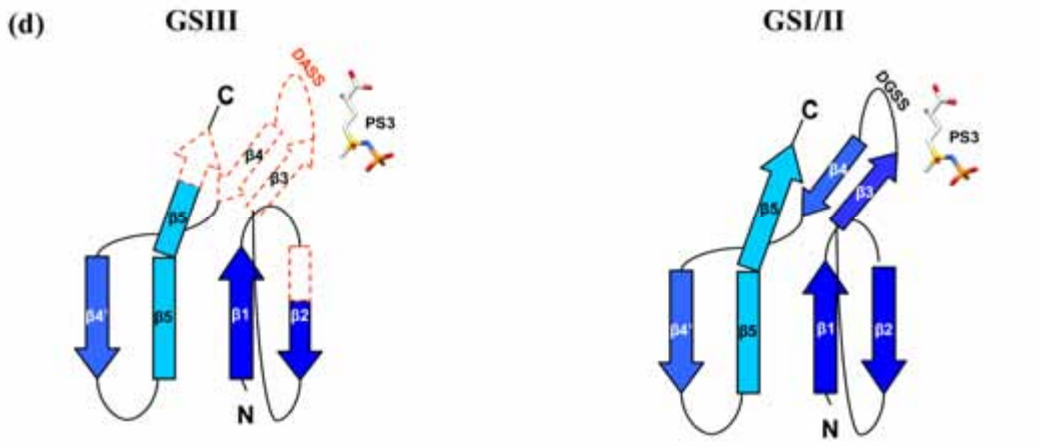


Figure 20: Conservation of core secondary structure elements (SSEs) and fold topologies between members of the GS superfamily. **(a)** MSA of members of three GS families: GSI - *S. typhimurium*, GSII – *A. thaliana* and GSIII – *B. fragilis*. This MSA was derived from the structural alignment of GlnN against the two GS structures: GSI - *2gls* and GSII- *2d3a* [see Table 1]. Sequence conservation, scored according to the Blosum62 substitution scoring matrix, is represented by the grey-scale shading with darker shading representing higher conservation. SSEs common to all three GS structures are annotated above the MSA in solid arrows (β-strands) or cylinders (α-helices) and labelled numerically in sequential order. Striped annotations represent SSEs that are only conserved in the GSI and GSIII families. The colour of both of these annotations follows the order of appearance of the SSEs in the structures. Missing regions in the GlnN sequence are outlined in red and missing SSEs are outlined in red dashed lines and labelled in red text. Residues contributing to the five important loops, labelled a-e, mentioned in Section 3.2.2 are highlighted by black boxes **(b)** Comparison of the three GS structures shown in ribbon representation and coloured and labelled as in (a). Conserved SSEs are colored according to (a). Red spheres, connected by dashed lines, mark the positions of the terminal residues of missing density. The view is into the active site β-barrel. The N-terminal domain has been separated from the C-terminal domain for the sake of clarity (black dashed line) but the orientation has not been altered. **(c)** Topology diagrams for the three GS structures. α-Helices are represented by circles and β-strands by triangles. Conserved SSEs are labelled and coloured as in (a) and SSEs unique to each structure are labelled sequentially in alphabetic order. The proximity between schematic SSEs implies association. The directionality of the connections is representative of the actual connectivity and square connections represent extended loops. The annotation of missing regions follows that of (a). **(d)** β-sheet connectivity diagrams of the N-domains. Only the main β-strands forming the β-grasp motif are shown. The annotation of missing regions follows that of (b) and the position of the MetSox (PS3) ligand is also indicated.

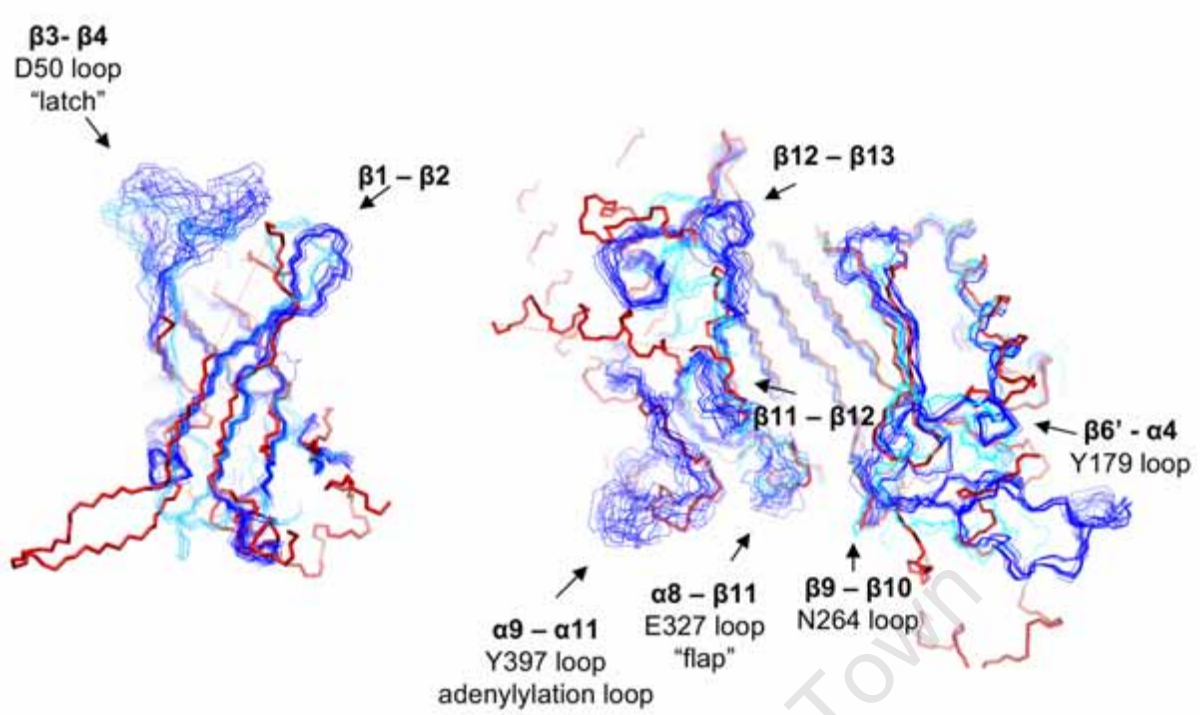


Figure 21: Structural variability in the GS superfamily. The structural alignment of GlnN (thick red trace) against 17 GSI (thin dark blue trace) and 6 GSII (thin cyan trace) structures is shown in Cα-backbone representation. The most flexible loops are labelled according to the scheme in Figure 17a and Section 3.2.2. The structures are listed in Table 1 in Section 3.3.3.

3.5.3 Conservation and divergence of quaternary structure

Sequence motifs responsible for divergence

Outside of the catalytic fold lie the non-conserved gap regions containing the sequence insertions and deletions responsible for the unique oligomeric structures of the different GS families (greyscale SSEs in Figure 20). The largest differences between the structures occur in the following regions:

N & C-termini. GlnN has long N and C-terminal insertions compared to the other GS structures. As with the GSII N-terminal 'meander', the N-terminus of GlnN starts with a short segment of α-helix which packs against SSEs near the core of the C-domain and is linked to

the compact N-domain by a region of extended coil. This N-terminal extension comprises 4 short α-helices and also contributes an additional β-strand, **βA**, to the N-domain. Unlike the GSII structure which is shorter than the GSI and, therefore, does not share the last 5 α-helices, GlnN is longer and in addition to sharing a section of **α13,** it also contains additional helices which contribute to the domed structures at either end of the complex.

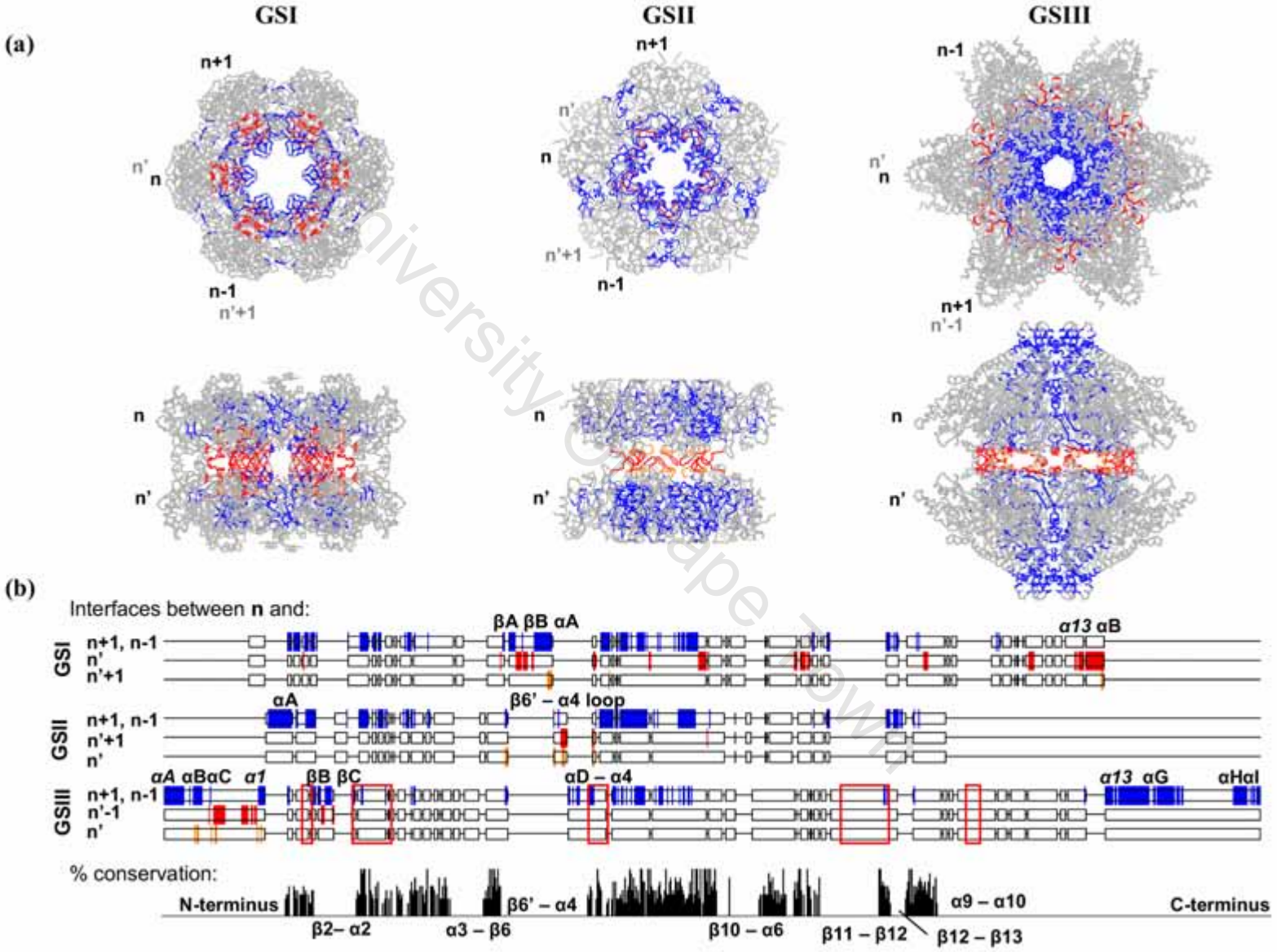
α9 - α10. This extended loop is only found in the GSI and GlnN structures and is near the tips of the GS 'pinwheel' arms. It connects the last of the conserved core helices, **α9**, to the C-terminal helices. In GSI, this loop contains a short stretch of α-helix, **α10**, just prior to the site of the adenylylation motif. In GlnN, this loop only starts after an additional helix but the chain direction and the position of **α10** are conserved. The density for the short loop connecting **α9** and this additional helix, **αF**, is missing in the GlnN structure (residues 509-514 in Table 6) and predicted to be a random coil. Interestingly, there is also a tyrosine residue (Y539) present in the loop region preceding the **α10** helix in GlnN.

β6' - α4. This region, below the active site in GSI and GSII, is involved in inter-ring associations. In the former family, this region contains a 2-stranded β-motif whereas in the latter, this region does not contain any SSEs. In GlnN, the backbone segment exiting **β6'** is also extended and contains two short segments of α-helix but the remaining section of the loop cannot be seen in the density (residues 244-257 in Table 6).

β11 - β12. This region links the final two adjacent parallel staves of the active site β-barrel at the periphery of the pinwheel arms and forms the top of the nucleotide binding region. Not only is this the site of the largest insertion in the GlnN structure but it is also the site of the largest missing density (residues 425-457 in Table 6) and the location of the proteolysis susceptibility site. Unlike in GSI and GSII, where a small loop connects these SSEs, in GlnN, the start of an additional helix can be found after the **β11** strand.

Intra-ring oligomeric contacts

Because the domain arrangements and active site folds are conserved in all GS families, a sub-set of the interactions that are responsible for the characteristic closed rings structures are highly conserved and map predictably to the interacting elements of the N and C-terminal domains (blue contacts in Figure 22). The nature of the interacting surfaces in GSII is still



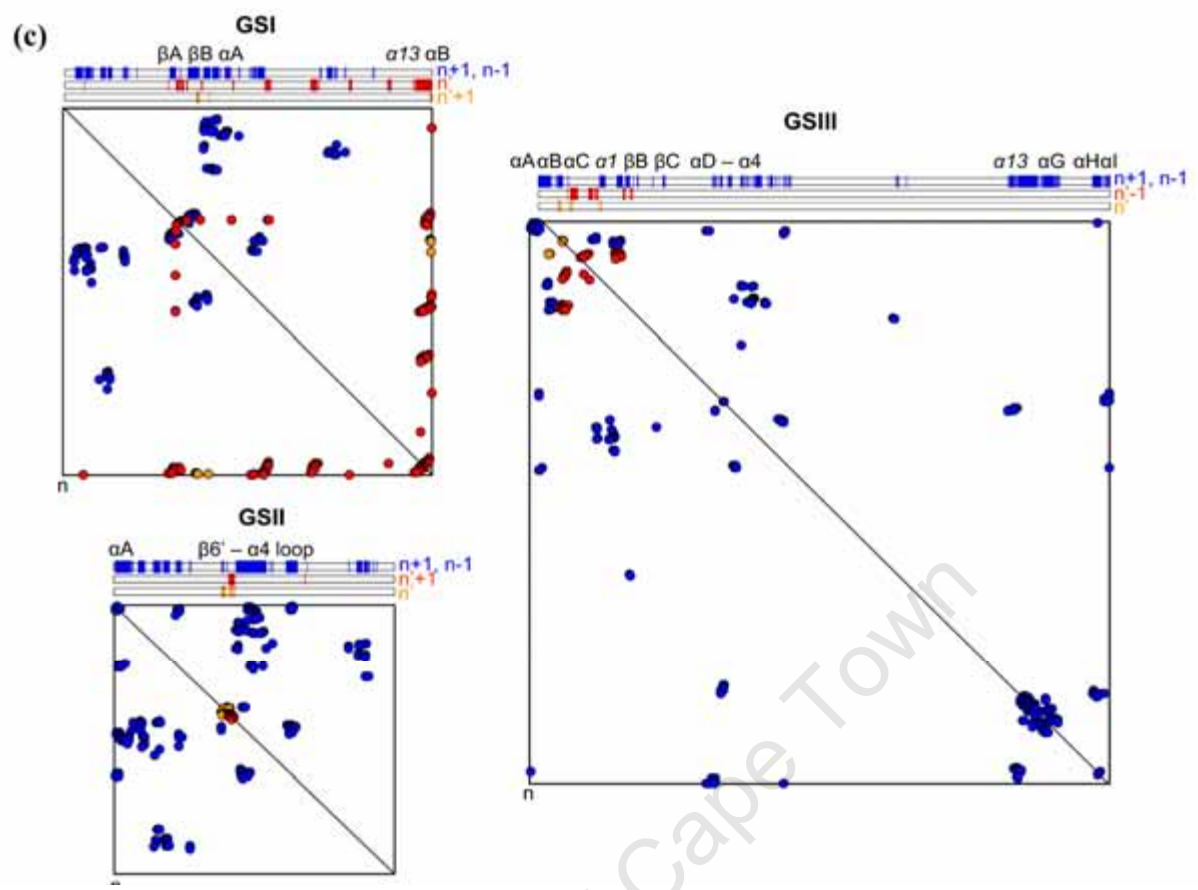


Figure 22: Quaternary structure interfaces in the GS superfamily. **(a)** Comparison of the interfaces responsible for higher order association in GlnN, GSI (*2gls*) and GSII (*2d3a*) [references in Table 1]. The three structures are rendered in Cα backbone representation and residues within 3.5 Å of their contacting partners are coloured as follows: intra-ring contacts are coloured blue and reflect the interacting residues of the **n**th subunit with the subunits, **n+1** and **n-1**, on either side of it (the numbering within the ring is positive in the direction of the N-terminal domain). Inter-ring contacts are highlighted in red and reflect the interaction of the **n**th subunit with the closest subunits in the symmetry related ring (marked with an apostrophe). The subunits in either ring of the GlnN and GSI structures are eclipsed and the diametrically opposite subunit to **n** is the **n'** subunit. In GSII, the rings are only partially eclipsed and the **n'** subunit is taken to be the closest subunit in the opposite ring. The less substantial inter-ring contacts between the **n**th subunit and the subunit on either side of the **n**' are coloured in orange. The labelling of the latter subunit interaction is not the same for all 3 structures because the orientation of the rings in GlnN is reversed in comparison to the other structures. **(b)** These same interacting residues are highlighted in the schematic multiple sequence alignments (MSA) shown below the structures. Unique interfaces contributing to inter and intra-ring quaternary associations are labelled as in Figure 20. Labels in italics represent conserved SSEs which contribute to interfaces in more than one GS family. A graph of the percentage conservation of the residues is plotted below the MSA and the positions of the most important gap regions are marked. **(c)** Residue interaction plots. All contacts (within 3.5 Å) between the **n**th subunit of the oligomer (vertical axis) and its contacting partner (horizontal axis) in the three oligomers are shown. Each of the three sets of data: blue, red, and orange, represents the interactions between different pairs of subunits, which are labelled and coloured as in (a). The SSEs contributing to the unique interfaces in each structure are annotated above the ungapped sequence schematic as in (b).

similar to those in the other structures despite the pentameric arrangement of the rings which results from a difference in the angle between the N and C-terminal domains.

Several additional unique regions within the non-conserved gap regions contribute to intra-ring stabilization. In GSI the **βA-βB** sheet region (residues 135-154 in *S. typhimurium*) within the major mid-sequence gap, which forms the interior wall of the double ring, associates with a short helix, **αA**, in an adjacent subunit to stabilize the ring. The stretch of helix in the neighbouring subunit is also located in the same gap region. In GSII, the N-terminal helix packs inside the cavity at the centre of the rings where it associates with residues from the C-domain and the equivalent residues in symmetry related N-terminal helices from neighbouring subunits. GlnN has several novel insertions which stabilize the intra-ring associations between subunits:

N-terminal insertions: The N-terminal helix **αA** (1-17), as in GSII, packs against the upper wall of the large inner cavity but from within the complex rather than from the outside (Figure 23a). This SSE interacts primarily with symmetry equivalent N-terminal helices and it also packs against residues in the C-domain of the same subunit, specifically, the short helix **αD** (222-227) and the C-terminal helix **αI** (723-729). These associations are dominated by hydrophobic surfaces on either side of the helix, which arise from the helical patterning of residues along its length. The single turn helix segment, $\boldsymbol{\alpha}^{\mathbf{29}}$, between **αA** and **αB** (22-32) interacts with the **βB-βC** (108-116) double stranded insertion and the loop between **α1** and **β1** (79-84) in the neighbouring subunit (Figure 23b). The former **βB-βC** insertion also strengthens the common side-by-side β-sheet stacking between **β2** of the N-domain and **β7** from the neighbouring subunit's C-domain.

Mid-sequence insertion: The region between **αD** and **α4** contains a short segment of helix, $\boldsymbol{\alpha}^{\mathbf{235}}$ (232-235), and the $\mathbf{loop}^{\mathbf{240}}$ (240-244), prior to disappearing into a region of missing density (Figure 23a & c). The $\boldsymbol{\alpha}^{\mathbf{235}}$ helix interacts with both the C-terminal **αI** helix and the $\mathbf{loop}^{\mathbf{240}}$ from an adjacent subunit. The $\mathbf{loop}^{\mathbf{240}}$ region interacts with **αD** and **α13** (582-633) of the adjacent subunit (Figure 23c).

C-terminal insertions: The large C-terminal extension in GlnN, comprising a three helix bundle, extends above the active site core region to form the dome shaped "cap" at either end of the complex (Figure 23d). Portions of the very long helices **α13** and **αG** (617-667) interact

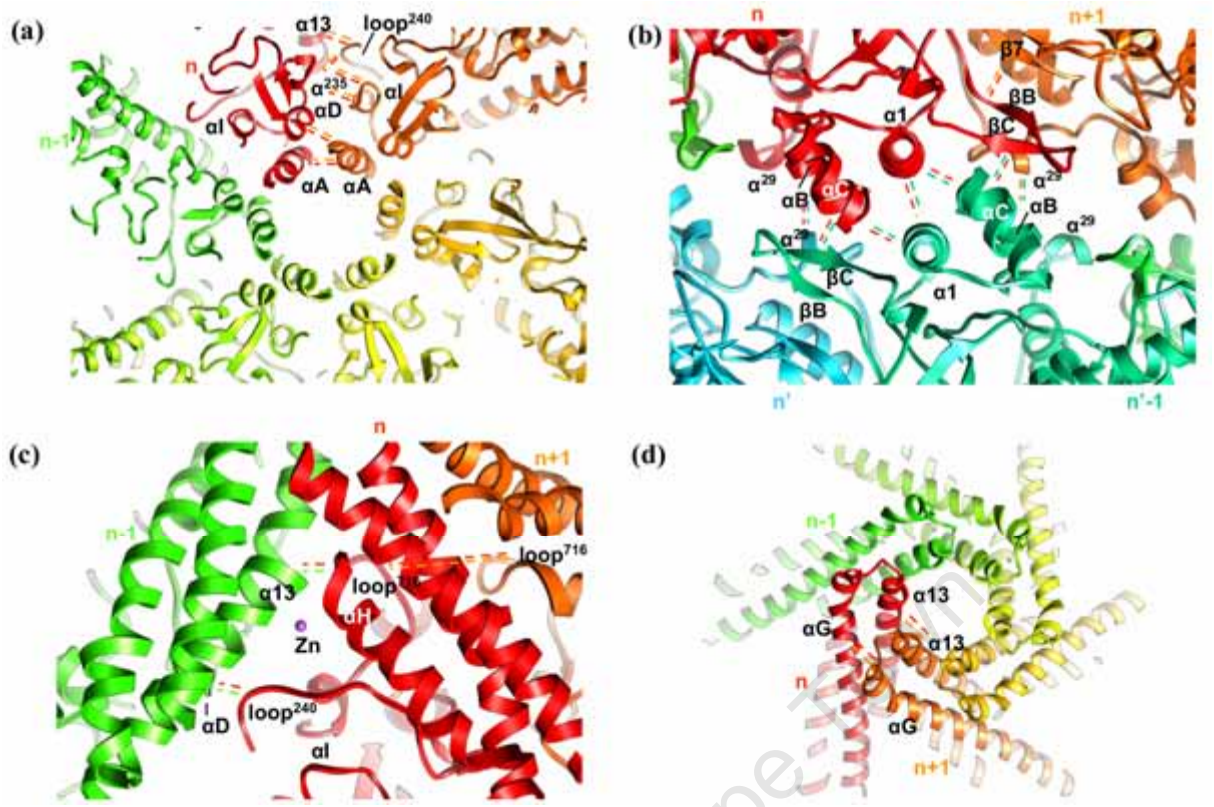


Figure 23: Subunit interactions responsible for the novel inter- and intra-ring interfaces in GlnN. The interacting subunits are labelled relative to the **nth** subunit, as in Figure 20, but their colouring is arbitrary. **(a)** View down the 6-fold axis from the inside of the complex. **(b)** View down the 2-fold axis at the interface between subunits from opposite rings of the complex. **(c)** Close-up view of the metal ion binding site formed between by the C-terminal helical extensions from adjacent subunits within one ring. **(d)** View down the 6-fold axis looking towards the centre of the complex. Interacting SSEs are labelled as in previous diagrams but 2 new elements and 2 loops are introduced: **α29**, **α235**, **loop240**, and **loop716** (superscript numbering representing the residue number at the midpoint of the SSE or loop). Interactions between SSEs are marked with dashed lines which are coloured according to the subunits involved.

with their symmetric equivalents from adjacent subunits to form a large contact area. The upper region of this interface is dominated by a hydrophobic patch and the lower region near the top of the cleft between subunits is formed by polar interactions (Figure 23c). Further down the cleft, **loop^{716}**, between the last two helices, **αH** and **αI,** in the C-terminus, together with residues from **α13** (612-615) forms a metal binding site between subunits and also interacts with their symmetric equivalents to stabilize the ring structure. The **loop^{716}** region contains residues D714, K721, and R723 and the **α13** of the neighboring subunits contributes

residues D609, N613, and H614 to the metal binding site. All of these residues are within 5 Å of the modeled zinc ion but only N613 and K721 are within 3.5 Å. Finally, the C-terminal helix, **αI**, also interacts with this metal binding loop and **α13** to further contribute to intra-ring stability.

Inter-ring oligomeric contacts

The non-conserved gap regions are also the locations of the differences in the inter-ring associations introduced above (orange and red contacts in Figure 22). Not only do these different interfaces assemble double-ringed structures with inverted ring interfaces but they also change the identity of the interacting subunits. In GSI the **nth** subunit interacts mainly with the opposite subunit, **n'**, (interface area of 2 648 Å^2) and to a lesser extent with the diagonally positioned **n'+1** subunit (interface area of only 193 Å^2) (Figure 22a and Table 8). The former interface is generated by the mid-sequence insertion, **βA-βB**, which forms a 4 stranded β-sheet with its symmetry related partner and the C-terminal helical insertions which extend across the inter-ring interface to pack hydrophobically into the core of the symmetry related subunit (Figure 22c). In GSII the **nth** subunit interacts weakly with both the **n'** (interface area of 147 Å^2) and the **n'+1** (interface area of 185 Å^2**)** subunits because the opposing rings in the GSII structure are only partially eclipsed. The mid-sequence insertion in GSII is also responsible for the inter-ring contacts and it is the **β6' – α4** loop region that interacts with its symmetrical equivalents to bring about association (Figure 22c). In GlnN, however, the **nth** subunit interacts primarily with the **n'-1** subunit in the other ring (interface area of 880 Å^2) and to a lesser degree with the eclipsing **n'** subunit (interface area of 168 Å^2). In addition, unlike in the other structures, the unique interfaces forming the inter-ring associations in GlnN are all found in the N-terminal region (Figure 22b & c). These unique insertions include:

(1) The conserved helix just prior to the first β-strand in the N-domain, **α1** (65-77) (Figure 23b). This helix displays a stripe of polar charged residues which interact with the equivalent residues in the symmetrically related **n'-1** subunit and with the short loop between the **αB** and **αC** helices (39-50).

(2) The double stranded **βB-βC** (108-116) insertion, which was also shown to stabilize intra-ring associations, interacts with the **αB - αC** helices in the **n'-1** subunit (Figure

23b). Similarly, the short **α**[29] helix, which also participated in intra-ring contacts, interacts with the **αB - αC** helices in the **n'** subunit. Both interactions are dominated by polar and charged amino-acids.

Interface contributions to oligomerization

In comparing the associations in the three structures it can be seen that the different stabilizing interfaces, contributed by the non-conserved regions, result in different predicted stabilities for the 3 complexes (Table 8). In GSI the surface area for the total intra-ring interface contributed by each monomer is 1 552 Å^2. For GSII this value is 1 890 Å^2 and the value for the GlnN structure is the highest at 2 292 Å^2. However, as a percentage, GSII buries the largest proportion of its surface area, 5% (25÷5) per monomer, when forming the pentameric ring versus roughly half of this value for the other two families (14.4÷6).

Table 8: Surface Area (SA) changes upon oligomerization in the GS superfamily.

	Monomer	Ring			Double-ring			
	SA (1E3 Å^2)	SA (1E3 Å^2)	% buried	Interface SA/subunit (1E3 Å^2)	SA (1E3 Å^2)	% buried	Δ% buried	Interface SA/subunit (1E3 Å^2)
GSI *2gls*	21.3	109.1	14.4	1.5	184.9	27.6	13.2	2.8
GSII *2d3a*	15.8	59.4	25.0	2.0	115.3	27.0	2.0	0.4
GSIII GlnN	28.2	141.5	16.2	2.3	272.7	19.2	3.0	0.9

In comparison, the inter-ring interfaces are the largest (2.8%), by an order of magnitude, in GSI family, where the C-terminal helices form extensive isologous interactions with the symmetry related ring. GSII has the weakest interaction which unlike in GlnN, is not mediated by SSEs but rather random coil as reflected by the lowest statistics (2%). Although GlnN is the largest complex, it buries the least amount of surface area during assembly in relation to overall complex size (19.2%).

3.6
Discussion

3.6.1 Structure solution

The knowledge of the GlnN structure from cryo-EM (Chapter 2) and the results of the self-rotation function analyses were essential in identifying the position of the complex in the discontinuous SAD phased maps. The quality of these initial maps suggests that the starting phase estimates were poor. The 6-fold NCS constraints, however, proved to be very powerful, and subsequent density modification, carried out with information from the cryo-EM volume, greatly improved the phases and extended their resolution to the limit of the data. Even subsequent X-ray crystallographic refinement failed to significantly improve upon these estimates.

The higher resolution of the crystal structure has allowed a direct evaluation of the quality of the cryo-EM reconstruction presented in Chapter 2. The two structures display a reasonable agreement in the arrangement of subunits and overall shape, as evidenced by the usefulness of the cryo-EM structure in solving the X-ray data. The helical bundles forming the domed caps and the SSEs forming the column-like inter-ring interfaces, in addition to the previously noted β-barrel folds forming the active sites, are all represented in the lower-resolution cryo-EM volume. However, significant differences can be found in the region between the outer surface of the structure and the walls of the inner cavity. The location of the latter density, which forms two rings at either end of the inner cavity, appears to correlate with the inner surface of the subunits but there is no intervening density between this and the outer surface. The results of the X-ray temperature factor analyses showed that these regions were not very mobile and it is, therefore, surprising that density is not visible. A possible explanation is the bias towards higher defocus data (Chapter 2) which could have resulted in a limited representation of frequencies and the presence of contrast variations in the final cryo-EM reconstruction.

3.6.2 Effects of proteolysis

The position of the missing density at the tips of the “pinwheel” arms (425-457) matches the predictions from the peptide fragment masses (Chapter 1) and the point at which the density

disappears corresponds with a predicted stretch of random coil. The 33 missing residues in this region appear to be disordered because no additional density can be seen in the P1 crystal structure of the full-length protein and temperature factor analysis revealed that the tips of the pinwheel arms are the most mobile regions of the GlnN complex. Comparisons of the cryo-EM volume and the crystal structure also appear to reveal small regions of empty density at the tips of the "pinwheel" arms but it is difficult to interpret such comparisons due to the limited resolution of the reconstructions and the presence of other spurious features (described above). The suggestion that this surface region is highly mobile would account for the previously observed changes in the mobility of the digested complex upon proteolysis (Chapter 1) and agrees with the consensus understanding of proteolytic susceptibility [Hubbard 1998]. The implication of this disorder, however, is that it is impossible to pin-point the exact site of digestion.

The results of the packing analyses showed that, due to their peripheral location, the tips of the pinwheels arms are involved in the higher order association of GlnN dodecamers in the crystal lattice of the P1 crystals. Their modification by proteolysis results in an altered molecular packing wherein the contact areas between dodecamers are maximized, thus, leading to the formation of the $C222_1$ crystal form. The presence of density for the ligands ADP and MetSox-P in density modified (iterative 12-fold averaged) maps of the P1 crystal data rules out the possibility that differences in ATP concentrations are to blame for the change in crystal packing [Liaw *et al.* 1993a]. The changes to the crystalline order, therefore, explain the improved mechanical stability and faster growth characteristics of the orthorhombic crystals. The stronger diffraction from these crystals is also understandable because the number of ASUs contributing to the Bragg scattering is doubled for the same number of dodecamers.

One of the primary contacts responsible for crystal packing in both the P1 and $C222_1$ crystal forms is contributed by the large flat surfaces of the domes at either end of the GlnN complex. These surfaces are formed by the helical bundles which make up the spokes of the "pinwheel"arms and they present a stripe of hydrophobic residues which is surrounded on either side by polar residues (data not shown). It is, therefore, possible that these surface features are also responsible for the preferential attachment of GlnN to the carbon support film or air water interface observed during electron microscopy studies (Chapter 2).

3.6.3 Active site structure and ligand binding

Electron density was clearly visible in the active site region of maps calculated from the unbiased density modified SAD phases and when the GlnN model was aligned with homologous structures it was clear that this density matched the position of the ADP and MetSox-P ligands in previously solved GSI and GSII structures [Krajewski *et al.* 2005; Unno *et al.* 2006]. The orientation of the nucleotide ring in GlnN was found to match the earlier StyGSI/AMP/2Mn structure [Liaw *et al.* 1994] and only two Mg^{2+} binding sites were identified, which corresponded to the commonly observed n_1 and n_2 sites. The arrangement of residues surrounding the n1 metal binding site and the structure of the **β7** strand (near residue 209 in *S. typhimurium*) confirm that all GS families share a common configuration in this region, corresponding to a "taut" conformation. However, as indicated by the high B-factors, there appears to be considerable flexibility in some regions of these ligands, such as the terminal phosphate of ADP, or their overall occupancy.

In addition to the weak density for some regions of the ligands, four out of the six essential residues involved in catalysis (Section 3.2.2 & Table 7) were not visualized. The residues D129 and Y254, belong to the structural equivalents of the missing "latch" and the Y179 loop and only very weak density exists for the side chains of residues R460 (R339 – equivalent in *S. typhimurium*) and R474 (R359). Density was also missing for the side chain of residue K281. In contrast, the important E327 and N264 residues, which belong to the "flap" and "N264" loops guarding the glutamate entrance to the active site, were clearly conserved in the corresponding GlnN residues: E403 and N337. Both the E327 'flap' and N264 loops, are known to be destabilized in holo-GS structures and they only become ordered following ATP and substrate binding [see Section 3.2.2]. The presence of these two loops in the GlnN structure, therefore, confirms their stabilization by the MetSox-P inhibitor, which traps the active site in the closed configuration. The absence of density for the remaining catalytic residues and loops, however, is surprising considering their known interactions with the MetSox inhibitor. Because the inhibition step was carried out at 4º C and excess inhibitor was substantially diluted in the preparation of material for crystallization (Chapter 1), it is possible that partial occupancy is to blame for the disorder seen in the remaining loops and catalytic residues.

3.6.4 The potential for the design of selective GSIII inhibitors

In total 17 of the 33 residues found to interact with the metal ions, ADP cofactor, and MetSox inhibitor were conserved in sequence. Although the configurations of the five missing side chains described above cannot be evaluated, it appears that the remaining residues are highly conserved both in sequence and structure. Together with the inferred conservation of the flexible "latch" and Y179 catalytic loops, these findings confirm the early bioinformatic analyses [Hill *et al.* 1989] and inhibition studies [Southern *et al.* 1987], which predicted that GlnN shares the catalytic mechanism of the well characterized GSI and GSII families.

The remaining 16 residues in the active site display a higher degree of sequence and conformational variability between the GS structures and these mutations, therefore, reflect the different modes of ligand recognition and binding present in the divergent GS families. For the most part, these substitutions cluster around the ADP binding site as previously noted from an analysis of the GSI and GSII structures [Unno *et al.* 2006; Krajewski *et al.* 2008]. This tolerance in side chain orientation and chemistry can in part be explained by the large fraction of these residues which were shown to interact with the ADP cofactor via their main chain atoms (7/18). Equally the residues N334, R470, and F298 are not likely to confer selectivity due to the similarity of their interaction with the ligands within each family. For instance, F298 in GSIII, P204 in GSII, and F232 in GSI, all interact with the nucleotide base of ADP via Pi stacking, with the side chains all adopting a similar parallel orientation with respect to the nucleotide base.

The remaining six unique active site residues, which all interact with the ADP cofactor via their different side chains, therefore, represent the most promising locations for the design of selective inhibitors against GlnN. Of these, the residues which potentially could provide the most selectivity between the GSIII and human GSII families are as follows:

(1) A465 - in the GSI and GSII families this is an arginine involved in binding the terminal phosphate of the ADP molecule and it also forms part of the n2 metal binding site. It is possible that the disorder seen at the terminal phosphate of ADP in GlnN is due to the missing stabilizing effect of this positively charged residue in addition to disordered R474 which also interacts with ADP in GSI and GSII.

(2) K281 – in the four GSII sequences investigated, this position is not highly conserved and is either occupied by a serine, valine, or alanine. In the GSI sequences, the position is always occupied by a glutamate. The negatively charged groups of the residues in both families interact with the sugar ring of the ADP ligand. However, the density for this side chain is missing in GlnN and its role in ligand binding is, therefore, uncertain.

(3) A295 - in GSII, the residue at position 201 is always an asparagine which interacts with the first phosphate of the ADP ligand, whereas, in all GSIII enzymes the position is occupied by an alanine.

(4) N469 - in GlnN this residue points towards the N6 atom of the adenosine ring of the ADP ligand, whereas, in GSI and GSII structures the interaction with the ligand is via backbone contacts.

In addition, there is no structural counterpart in GSII or GlnN to the backbone interaction between K361 and the nucleotide ring of the ADP molecule in GSI because of the variability in the topology of the **β11-β12** region in these structures. Equally, K215 in *M. tb* GSI interacts with the first phosphate of the ADP cofactor but in GlnN, *Z.mays* GSII, and *S. typhimurium* GSI the equivalent residue is a threonine, glycine, or alanine (respectively) which are all greater than 4 Å away from the ligand. Therefore, although these two positions also appear to contribute towards varied protein-cofactor interactions between the families, they are not likely to be exploitable due to the similarities between the GSII and GSIII families.

3.6.5 Conservation and divergence of quaternary structure interfaces

Both of the crystal structures revealed a dodecameric arrangement of subunits with D6 symmetry, thus, confirming the predictions by cryo-EM that this is the quaternary structure adopted by the GlnN in solution. At the low resolution provided by the former technique, the handedness of the GlnN complex was undetermined, but, a strong agreement with the GSI structure was noted for quaternary structure of the hexameric rings and the location of the active sites (Chapter 2). Surprisingly, however, the higher resolution of the GlnN crystal structures revealed an inversion in the orientation of the hexameric rings, which associate via the opposite face to that seen in the GSI and GSII structures. This discovery also explains the

"pinwheel" views discovered during negative stain EM as arising from partial staining of the large GlnN complex (Section 2.2.3) [van Rooyen 2004].

The near-atomic resolution of the GlnN crystal structure allowed a detailed inspection of the sequence features responsible for the conservation of intra-ring, and the divergence of, inter-ring higher order assembly interfaces in the GSIII family.

Intra-ring interfaces – Predictably, core components of the intra-ring interfaces are conserved in all GS structures because functional active sites β-barrel folds are only formed between adjacent subunit within the closed rings structures. However, each of the three families appears to be stabilized by additional intra-ring contacts outside of the core conserved regions. In comparison to the structure of GSI, the GlnN and GSII oligomers appear to have stronger intra-ring associations. GlnN in particular, has long helical insertions at the C-terminus which interact with symmetry related regions in the domed caps at either end of the complex. This is in addition to the N-terminal helical "meander" insertion which also interacts with symmetry related helices and the metal ion binding site between subunits.

Inter-ring interfaces – Unlike the intra-ring contacts, none of these interfaces are conserved between the GS families. GSI has evolved the most extensive inter-ring associations that, in addition to interacting with symmetry related partners in opposite rings via isologous contacts, also interact with a number of residues outside of the non-conserved regions. The calculated dominance of these interactions has been confirmed by electron microscopy investigations of GSI dissociation intermediates [Valentine *et al.* 1968]. One of these stabilizing interfaces, the double-stranded sheet that forms the central channel in GSI, is missing in the GS enzyme from the GlnT family and from several Gram-positive bacteria. Both GSII and GlnN, on the other hand, have much less extensive inter-ring interfaces that only associate via isologous interactions with residues from symmetry related non-conserved regions. In addition, in GlnN, this interface is formed by the N-terminal region and not the C-terminal regions present in GSI/II. The N-terminal **α1** helix, which is a major component of the inter-ring interface of GlnN, has also been implicated in higher order associations in GSI in response to divalent cations and during salting out [Dabrowski *et al.* 1994; Dabrowski *et al.* 1996]. The burial of this SSE in GlnN, therefore, confirms the predictions made in Chapter 1 that GlnN cannot be precipitated by divalent cations such as Zn^{2+} in the same manner as GSI

and also suggests that the GSI and GSIII structures can form higher order associations across the same interface.

In summary, the primary determinants of the divergence in quaternary structures between the three GS families were found to cluster in non-conserved regions at the termini of the structures and in several smaller mid-sequence regions.

3.6.6 Evolution of GS higher order structure

It has been suggested that the assembly/disassembly intermediates of homo-oligomeric proteins resemble intermediates along the pathway of their evolution from monomeric proteins to higher order assemblies [Levy *et al.* 2008]. The GS enzymes appear to refute this hypothesis because, as described earlier, the catalytically functional unit is formed within ring structures, whereas, the dissociation intermediates of GSI are not single rings but rather dimers formed by the pairing of subunits from opposite sides of the rings [Valentine *et al.* 1968]. An alternate evolutionary origin for the GS enzymes was suggested by Llorca *et al.* [2006] who hypothesised that the oligomeric GS enzymes arose via the fusion of two monomeric proteins, corresponding to the regulatory N and catalytic C-terminal domains, and this configuration constrained the structures to all form closed rings, thus maximizing number of functional active sites. The subsequent evolution of the double-ringed GS structures can then be imagined to have occurred under the pressure of the commonly accepted driving forces for the evolution of higher-order protein assemblies [reviewed by Goodsell *et al.* 2000]:

Increased stability – Large proteins, which are most efficiently formed by oligomers, are thought to achieve higher stability through the cooperative action of multiple weak stabilizing interactions, thereby, avoiding the entropic costs of more rigid monomers [Monod 1968; Lumry *et al.* 1970]. In line with this theory, the large number of flexible catalytic loops surrounding the active sites together with the large size and extensive quaternary structure interfaces in the GS oligomers suggest that these enzymes have achieved stability without sacrificing the necessary catalytic flexibility. Studies have also shown that the different stabilities of various plant GSII isozymes can be attributed to mutations in a single residue (I161) involved in intra-ring contracts between adjacent subunits [Unno *et al.* 2006], thus, providing clear evidence of the weak but cooperative nature of the oligomerization interfaces.

It also appears that symmetry has been exploited to efficiently explore the sequence space necessary to introduce new quaternary structure interfaces during evolution, as evidenced by the number of interfaces associating with symmetry related or other non-conserved regions.

Decreased surface area – By burying a large percentage of their surface area upon oligomerization larger proteins are thought to limit deleterious interactions with other proteins such as non-specific aggregations and susceptibility to proteolysis. It is clear from the non-progress of the proteolyitc degradation of GlnN by the *P. fluorescens* protease (Chapter 1) that the accessible area for proteolysis was limited. It has also been suggested that the reduced surface areas of oligomers enhances the rate of catalysis by limiting the unproductive interactions of substrates with non catalytic areas of the enzyme [Sharp *et al.* 1987]. The results of the comparative surface area analyses from the three families, however, are at odds with these theories. GlnN, the largest protein, appears to bury the least surface area upon oligomeriation; only 19% in comparison to the ~30% buried by the others. However, if it is considered that the central cavity of the GlnN dodecamer is not freely accessible to larger molecules, then the fraction of surface area buried becomes much more favourable (35%) and the highest out of all the GS enzymes. The very large GSIII enzymes, thus, appear to have evolved in order to bury the largest possible surface area per active site.

Cooperativity – Through allosteric mechanisms, the direct interaction of subunits in oligomeric proteins can achieve enhanced regulation of catalysis. Several lines of evidence suggest that the GS enzymes display catalytic cooperativity in addition to their modulation by allosteric effectors such as adenylylation [Wedler *et al.* 1982].

The structures of the three different GS families, therefore, all derive the benefits of improved catalytic potential and the enhanced stability that result from the formation of large oligomeric assemblies. However, the GlnN structure raises an interesting question as to how two double-ringed structures could have evolved with inverted inter-ring interfaces, given the evolutionary pressure to maintain these benefits. Two possible hypotheses can be suggested to explain this. (A) It is possible that alternate interfaces evolved in parallel on opposite sides of the single-ringed ancestor, following gene duplication. However, considering the conservation of the N and C-terminal helical regions of GSI, **α1** and **α13**, in GlnN and the large size difference between the two enzymes, this seems unlikely. (B) It seems more plausible that the inversion of the ring interfaces took place after the initial formation of a GSIII-like double-

ringed ancestor through the disruption of existing interfaces and the subsequent evolution of new associations. The disruption is most likely to have occurred through truncation as suggested by the terminal location of the sequence motifs responsible for inter-ring association in GlnN and the shorter length of the **α1** and **α13** SSEs in GSI. For instance, truncation of the N and C-terminal regions of a GSIII-like ancestor could have resulted in a single-ringed structure which then evolved to form a GSI-like dodecamer through alterations in the C-terminal "helical thong" and mid-sequence double-stranded regions. The fact that some gram-positive GSI enzymes do not possess the latter insertion suggests that the inter-ring associations formed by the C-terminal region would alone be sufficient for oligomerization. This would presume that such an intermediate was tolerated by the host organism and that sufficient time existed for the formation of the new interface. Alternatively, it is also possible that truncation of the C-terminal region in a GSIII-like ancestor led directly to the formation of a GSI-like inter-ring interface. In this case, if the N-terminal interface was also disrupted, GSI-like dodecamers would result, or if it was still maintained, tubular structures possessing both interfaces could have formed. The subsequent disruption of the N-terminal interface in this intermediate would then result in a GSI-like dodecamer. Regardless of the intermediate steps, the truncation hypothesis would also explain the ability of GSI to associate across an N-terminal interface in the tubular aggregates that form in the presence of divalent cations [Valentine *et al.* 1968].

A better understanding of the evolutionary relationships between these interfaces will require a detailed investigation of their conservation in higher-order GSI aggregates together with truncation experiments to test their inter-convertibility, and phyologentic studies to estimate the time scales involved in the suggested transitions.

3.6.7 Conclusion

The surprising divergence in quaternary structure revealed by the first crystal structure of a GSIII protein has raised interesting questions surrounding the mechanism and driving forces responsible for the evolution of higher order assembly and its role in the functioning of these ancient and essential enzymes. Although the ligand binding motifs and reaction mechanism appears to be conserved across all three GS families, several residues that appear to be determinants of cofactor specificity were identified in GlnN, thus, opening the way for the design of selective inhibitors against the GSIII enzymes and their medically relevant hosts.

Conclusions

The primary aim of this work was to solve the structure of the GlnN protein from *B. fragilis* in sufficient detail to enable inferences to be made about the evolution and functioning of the divergent GSIII family in the light of the known structure-function relationships of the well characterized GSI and GSII families. It was also hoped that such a detailed understanding of the structural relationship between these families would identity the residues responsible for conferring substrate specificity, thereby, allowing an evaluation of the potential for the design of selective inhibitors against the GSIII enzymes. The complimentary structural techniques of X-ray crystallography and cryo-EM were chosen to achieve these goals, and improvements to the isolation protocols and sample conditions were sought to produce material suitable for such analyses. In addition to providing a model for the solution of the X-ray structure, the cryo-EM investigations also sought to elucidate any quaternary structure changes accompanying cooperative catalytic behavior in the GS enzymes.

The achievements presented in this work can be summarized as follows:

Novel purification procedure

It has been demonstrated that it is possible to create a rapid and scalable isolation procedure for the production of pure GlnN, of suitable quality and quantity for crystallization studies, by exploiting the partially reversible precipitation that occurs in the presence of Zn^{2+} ions. The previously encountered aggregation phenomenon has been attributed to the aggregation of denatured GlnN and improvements to the stability and solubility of the GlnN preparations have been accomplished.

Low-resolution reconstruction

Reconstructions of frozen-hydrated GlnN have yielded low-resolution structural information (16 Å) of the complex, revealing a D6 symmetric arrangement of subunits in solution. Comparisons with the previously solved GSI dodecamer have reiterated the similarity of the subunit arrangements in the hexameric rings from both complexes, but at the same time, revealed significant differences between the two in the “domed caps” at either end of the structures, the tips of the “pinwheel” arms, and the inter-ring interfaces. The reconstructions of the GlnN enzymes auto-inhibited with MetSox and ATP have revealed that no

conformational change is perceivable upon binding of the ligands or irreversible inhibition at the currently achievable resolutions of the reconstructions.

Crystallization and structure solution

Conditions for the production of diffraction quality crystals (P1 crystal form containing bound ATP and MetSox-P) have been discovered following the improvements to the purity, solubility, and stability of the GlnN material. A protease has been discovered that is capable of nicking a surface exposed loop region of GlnN to yield crystals of a different space group ($C222_1$) with improved mechanical and diffraction properties. Conditions for the optimal preservation of diffraction at cryogenic temperatures have been discovered for both crystal forms and derivatization of the $C222_1$ crystals with Ta_6Br_{12} has been accomplished. The crystal structure of the digested GlnN complex has been solved to near atomic resolution (3.0 Å) by using an iterative density modification scheme (taking into account the structure of the cryo-EM reconstruction) to extend the resolution of the initial low-resolution SAD phases (7 Å). From these data, it has been possible to build an atomic model of 88% of the GlnN structure but several surface exposed loop regions are not visible.

Structural consequences of proteolysis

The site of protease susceptibility has been mapped to a region of missing density at the tips of the GlnN “pinwheel” arms in both structures and these regions are the most mobile areas according to temperature factor analyses. Crystallographic packing analyses have highlighted the significance of these protease sites in crystal lattice formation and explained how their modification leads to the alternate $C222_1$ crystal form through increased crystalline contacts.

Conservation of catalytic fold, active site residues, and ligand interactions

Comparisons with structures from the GSI and GSII families have confirmed the predicted conservation of the domain arrangements and catalytic folds in GlnN. The positions of the ADP and MetSox-P ligands in the active site of the $C222_1$ and P1 structures, together with the conservation of the sequence identity and configuration of the majority of active site residues, and the missing density for several flexible catalytic loops, suggest a common reaction mechanism for the three GS families. Inspections of the first structure-based MSAs of

representative members of the superfamily indicate that the greatest sequence variance in the active site residues occur at the nucleotide binding site and this tolerance can be partially explained by a reliance on protein main chain interactions with the cofactor. Six active site residue positions have been found to demonstrate unique interactions with the ligands and of these, A295 and A465, are suggested as the best sites to exploit in the design of selective inhibitors against the GSIII family.

Divergence of quaternary structure interfaces

The higher resolution of the crystal structure has revealed that the GlnN dodecamer is formed by the association of hexameric rings, which associated across the opposite interface to that in the GSI and GSII structures. Comparisons have highlighted that in addition to the conserved active-site interactions, the different GS structures are stabilized by unique higher order interfaces and the sequence motifs responsible have been identified. This divergence of the quaternary structure in the three GS families has raised interesting questions surrounding the mechanism and driving forces responsible for the evolution of quaternary structure and its role in the functioning of these ancient and essential enzymes. The large size of the GSIII enzymes, together with the conservation of the GSI terminal helices, which are involved in higher-order associations, suggests that the GSI and GSII enzymes could have arisen from the truncation of a GSIII-like ancestor.

Future Work

Rational drug design

It has been demonstrated that GlnN is essential for the *in vitro* growth of *B. fragilis* under conditions of low-nitrogen [Tumba 2007]. In order to validate GlnN as a drug target, further investigations of its role in the nitrogen metabolism of *B. fragilis* are necessary. In particular, the viability of a *B. fragilis* strain lacking a functioning GlnN gene needs to be evaluated under conditions more closely resembling those encountered during infection. This could either be achieved through inhibitory studies, knock-out mutagenesis, or oligonucleotide interference technologies.

If these results prove that GlnN is essential to the viability or pathogenesis of *B. fragilis,* it then remains to be determined whether existing inhibitors can be modified or novel ligands designed to provide selectivity towards GlnN and not the human GSII enzyme. In particular, it would appear that the best route to achieving this selectivity would be to design ADP analogous with 'bulkier' moieties in place of the phosphate groups to take advantage of the more accommodating A295 and A465 sites on either side of the cofactor binding pocket in the GSIII enzymes.

The altered ring interfaces of the GSIII and GSII families also offer the interesting possibility for selective drug design due to the suggestion that the nucleotide cofactor and substrate enter the active site from opposite sides [Liaw *et al.* 1994]. Such inhibitors would need to be polyvalent to discriminate between the inverted geometries of the active sites in the two families. It can be imagined that a ring of six linked substrate analogs would have access to the GSIII active sites but not the GSI because of the inter-ring interface.

The recent discovery of a GSIII in the important human pathogen *Trichomonas vaginalis* has further motivated efforts to design selective inhibitors against these unique enzymes. The role of this GS in the nitrogen metabolism of *T. vaginalis* needs to be investigated with a view to validating its nomination as a drug target.

Specialization of the divergent GS families

The occurrence of multiple different GS families within the same species, despite the presumably large energetic cost of producing multiple forms of the very large enzymes, begs

the question as to what the specialized roles are for each family. To answer this question, a more detailed understanding of the regulation and kinetic activity of the divergent GSIII family is required.

Very little is known about the post-translational regulation of GlnN and there has been a persistent failure to recover active GlnN from the native host, *B. fragilis* [Southern *et al.* 1987]. The purification protocol presented in this work provides an opportunity to rapidly recover protein from *B. fragilis* for further biochemical characterization.

A more detailed understanding of the reaction mechanism will require higher resolution structures of GlnN and co-crystals of different inhibitors/non-hydrolysable cofactors with high occupancies are necessary to reveal the contribution of the missing density to ligand binding. In addition to improving the occupancy of bound ligands, crystals of homologous proteins should also be tested for higher resolution diffraction.

Selective inhibitors of GS enzymes would greatly advance our understanding of the different catalytic roles played by these enzymes. The ability to "turn off" each of the different GS families, both in whole-cell physiological investigations and in kinetic analyses of mixtures of the different GS enzymes, would allow their individual and synergistic contributions to nitrogen metabolism to be evaluated. Therefore, even if commercially viable and non-toxic drugs cannot be produced against the anaerobic pathogens, the design of selective inhibitors against the divergent GSIII enzymes is a priority.

Structural basis for cooperativity

The factors suggested to be responsible for limiting the resolution of the cryo-EM reconstructions need to be assessed in the light of the new protocols for the isolation of stable and homogenous preparations of GlnN. Reconstructions of protein recovered from dissolved crystals offers the best opportunity to improve the resolution of the cryo-EM reconstructions. Comparisons with the previously determined structures will reveal the contribution of structural heterogeneity to the limitations in the reconstruction resolution. In addition, if higher resolution (< 8 Å) can be obtained, SPEM together with 3D variance analysis could be used investigate the role of cooperative domain motions in the GlnN reaction mechanism.

Evolution of GS quaternary structure

Due to the conservation of the domain boundaries, catalytic folds, and topologies in the GSI and GSIII families, it should be possible to investigate the inter-convertibility of the different inter-ring interfaces and, therefore, their evolutionary origins, through simple truncation experiments. This will reveal whether the truncation of the terminal regions in GSIII can result in the formation or abolishment of higher-order associations. The resulting oligomeric species would most easily be verified with electron microscopy.

Equally, the role played by quaternary structure in the function of the GS enzymes can be further investigated by domain swapping experiments. For instance, if the N-terminal domain of the *E. coli* GSI is swapped for the GlnN domain then higher order fibrous structures should arise. Alternatively, if the GlnN N-terminal domain is swapped for the *E. coli* one then single rings should result. The kinetic potential of the resulting species could then be evaluated to determine the contributions of oligomerization to GS activity and to provide evidence for the possible existence of such intermediates in the evolutionary history of the GS enzymes.

In addition to these studies, higher resolution reconstructions of cation-induced GSI cables are needed to compare the dodecamer interfaces present in these structures with the GlnN inter-ring interface, thereby confirming whether these higher order associations share an evolutionary history.

Identification and analysis of GS enzymes using bioinformatics

The structure-based MSAs presented in this work have been used to create sensitive position specific scoring profiles for the identification of divergent GS enzymes from sequence homology alone and the structural information provided by the GlnN protein will facilitate the structural predictions of unknown GSIII enzymes. Phylogenetic analyses of all the identified GS proteins are needed to provide a better understanding of their distribution and to provide estimates of the time scales involved their divergence. The evolution of quaternary structure and its role in the functioning of these ancient and ubiquitous enzymes can only be understood in this context.

References

Abell, L.M. & Villafranca, J.J. (1991). Effect of metal ions and adenylylation state on the internal thermodynamics of phosphoryl transfer in the *Escherichia coli* glutamine synthetase reaction. *Biochemistry,* **30**, 1413-1418.

Abrahams, G. L. (1996). Nitrogen metabolism of *Bacteroides fragilis bf1*, *B.Sc (Honours) dissertation*, UCT.

Abramoff, M.D., Magelhaes, P.J. & Ram, S.J. (2004). Image Processing with ImageJ. *Biophotonics International,* **11**(7), 36-42.

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C. & Zwart, P.H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr*. **D66**, 213-221.

Almassy, R.J., Janson, C.A., Hamlin, R., Xuong, N.H. & Eisenberg, D. (1986). Novel subunit-subunit interactions in the structure of glutamine synthetase. *Nature,* **323(**6086), 304-309.

Amaya, K.R., Kocherginskaya, S.A., Mackie, R.I. & Cann, I.K.O. (2005). Biochemical and mutational analysis of glutamine synthetase type III from the rumen anaerobe *Ruminococcus albus* 8. *J. Bacteriol.* **187**(21), 7481-7491.

AVEREST, Kleywegt, J., Duquerroy, S. & Trapani, S., (2008), computer program.

Backman, K., Chen, Y.M. & Magasanik, B. (1981). Physical and genetic characterization of tthe glnA-glnG region of the *Escherichia coli* chromosome. *Proc. Nat. Acad. Sci. U.S.A.* **78**, 3743-3747.

Bartolucci, C., Lamba, D., Grazulis, S., Manakova, E. & Heumann, H. (2005). Crystal structure of wild-type chaperonin GroEL. *J. Mol. Biol.* **354**, 940-951.

Bender, R.A., Janssen, K.A., Resnick, A.D., Blumenberg, M., Foor, F. & Magasanik, B. (1977) Biochemical parameters of glutamine synthetase from *Klebsiella aerogenes*, *J. Bacteriol.* **129**(2), 1001-1009.

Bennett, M.J. & Cullimore, J.V. (1989). Glutamine synthetase isoenzymes of *Phaseolus vulgaris* L.: subunit composition in developing root nodules and plumules. *Planta*, **179**(4), 433-440.

Betti, M., Márquez, A.J., Yanes, C. & Maestre, A. (2002). ATP binding to purified homopolymeric plant glutamine synthetase studied by isothermal titration calorimetry. *Thermochim. Acta*, **394**, 63-71.

Blum, H., Beier, H. & Gross, H.J. (1987). Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis*, **8**, 93-99.

Boksha, I.S., Schonfeld, H.J., Langen, H., Muller, F., Tereshkina, E.B. & Burbaeva, G.S. (2002). Glutamine synthetase isolated from human brain: octameric structure and homology of partial primary structure with human liver glutamine synthetase. *Biochemistry (Mosc.)*, **67**(9), 1012-1020.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248-254.

Brunger, A.T. (2007). Version 1.2 of the Crystallography and NMR System. *Nat. Protoc.* **2**, 2728-2733.

Burroughs, A.M., Balaji, S., Iyer, L.M. & Aravind, L. (2007). Small but versatile: the extraordinary functional and structural diversity of the beta-grasp fold. *Biol. Direct.* **2**, 18.

Burton, Z.F. & Eisenberg, D. (1980). A procedure for rapid isolation of both groE protein and glutamine synthetase from *E. coli*. *Arch. Biochem. Biophys.* **205**(2), 478-488.

Bywater, R.P., Carlisle, C.H. & Jackson, R.B. (1969). An x-ray crystallographic study of *Escherichia coli* glutamine synthetase. *J. Mol. Biol.* **45**(2), 429-31.

Carvalho H., Sunkel C., Salema R., & Cullimore J.V. (1997). Heteromeric assembly of the cytosolic glutamine synthetase polypeptides of Medicago truncatula: complementation of a glnA Escherichia coli mutant with a plant domain-swapped enzyme. *Plant Mol. Biol.* **35**(5), 623-32.

Chayen, N.E., Shaw-Stewart, P.D., Maeder, D.L. & Blow, D.M. (1990). An automated system for micro-batch protein crystallization and screening. *J. Appl. Crystallogr.* **23**, 297-302.

Chen, J.Z. & Grigorieff, N. (2007). SIGNATURE: A single-particle selection system for molecular electron microscopy. *J. Struct. Biol.* **157**(1), 168-73.

Chirokov, A. Adobe® Photoshop® FFT/iFFT plugin, (2005), computer program.

Collaborative Computational Project, Number 4, (1994). The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr.* **D50**, 760-763.

Cowtan, K. (2006). The Buccaneer software for automated model building. *Acta Crystallogr.* **D62**, 1002-1011.

Crowther, R.A., Henderson, R. & Smith, J.M. (1996). MRC Image Processing Programs. *J. Struct. Biol.* **116**, 9-16.

Dabrowski, M.J., Dietze, E.C. & Atkins, W.M. (1996). Engineering the aggregation properties of dodecameric glutamine synthetase: a single amino acid substitution controls 'salting out'. *Protein Eng*. **9**(3), 291.

Dabrowski, M.J., Yanchunas, J., Villafranca, B.C., Dietze, E.C., Schurke, P. & Atkins, W.M. (1994). Supramolecular self-assembly of glutamine synthetase: mutagenesis of a novel intermolecular metal binding site required for dodecamer stacking. *Biochemistry,* **33**, 14957.

D'Arcy, A., Elmore, C., Stihle M. & Johnston J.E. (1996). A novel approach to crystallising proteins under oil. *J. Cryst. Growth*, **168**(1), 175-180.

Davis, R.W., Botstein, D. & Roth, J.R. (1980). Advanced Bacterial Genetics Cold Spring Harbour Laboratory, New York.

Denman, R.B. & Wedler, F.C. (1984). Association-dissociation of mammalian brain glutamine synthetase: effects of metal ions and other ligands. *Arch. Biochem. Biophys.* **232**(2), 427-440.

DeRosier, D.J. & Klug, A. (1968). Reconstruction of three dimensional structures from electron Micrographs. *Nature*, **217**, 130-134.

Eisenberg, D., Gill, H.S., Pfluegl, G.M. & Rotstein, S.H. (2000). Structure-function relationships of glutamine synthetases. *Biochim. Biophys. Acta,* **1477**(1-2), 122-145.

Elliott, W.H. (1951). Studies on the enzymatic synthesis of glutamine. *Biochem. J,* **49**, 106-112.

Elliott, W.H. (1953). Isolation of glutamine synthetase and glutamotransferase from green peas. *J. Biol. Chem.* **201**(2) 661-672.

Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr*. **D60**, 2126-2132.

Evans, P.R. (2005). Scaling and assessment of data quality. *Acta Crystallogr*. **D62**, 72-82.

Fabiola, F., Korostelev, A. & Chapman, M.S. (2006). Bias in cross-validated free R factors: mitigation of the effects of non-crystallographic symmetry. *Acta Crystallogr*. **D62**, 227-238.

Frank, J., Goldfarb, W., Eisenberg, D. & Baker, T. S. (1978). Reconstruction of glutamine synthetase using computer averaging. *Ultramicroscopy*, **3**(3), 283-290.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M. & Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**(1), 190-199.

Frey, T.G., Eisenberg, D. & Eiserling, F.A. (1975). Glutamine synthetase forms three- and seven-stranded helical cables. *Proc. Nat. Acad. Sci. U.S.A.* **72**(9), 3402-3406.

Garcia-Dominguez, M., Reyes, J.C. & Florencio, F.J. (1997). Purification and characterization of a new type of glutamine synthetase from cyanobacteria. *Eur. J. Biochem.* **244**(1), 258-264.

Gass, J.D. & Meister, A. (1970). Computer analysis of the active site of glutamine synthetase. *Biochemistry*, **9**(6), 1380-90.

Gibson III, F. C., Onderdonk, A. B., Kasper, D. L., & Tzianabos, A. O. (1998). Cellular mechanism of intraabdominal abscess formation by *Bacteroides fragilis*. *J. Immunol.* **160**, 5000-5006.

Gill, H.S., Pfluegl, G.M.U. & Eisenberg, D. (1999). Preliminary crystallographic studies on glutamine synthetase from *Mycobacterium tuberculosis*. *Acta Crystallogr*. ***D*55**, 865-868.

Gill, H.S., Pfluegl, G.M.U. & Eisenberg, D. (2002). Multicopy crystallographic refinement of a relaxed glutamine synthetase from *Mycobacterium tuberculosis* highlights flexible loops in the enzymatic mechanism and its regulation. *Biochemistry*, **41**(31), 9863-9872.

Gill, H.S. & Eisenberg, D. (2001). The crystal structure of phosphinothricin in the active site of glutamine synthetase illuminates the mechanism of enzymatic inhibition. *Biochemistry*, **40**, 1903-1912.

Gill, H.S., Pfluegl, G.M. & Eisenberg, D. (2002). Multicopy crystallographic refinement of a relaxed glutamine synthetase from *Mycobacterium tuberculosis* highlights flexible loops in the enzymatic mechanism and its regulation. *Biochemistry*, **41**, 9863-9872.

Ginsburg, A. (1972). Glutamine synthetase of *Escherichia coli*: some physical and chemical properties. *Adv. Protein Chem*. **26**, 1-79.

Gipson, B., Zeng,X., Zhang, Z.Y. & Stahlberg, H. (2007). 2dx - User-friendly image processing for 2D crystals. *J. Struct. Biol*. **157**, 64-72.

Goodman, H. J. K. & Woods, D. R. (1993). Cloning and nucleotide sequence of the *Buyrivibrio fibrisolvens* gene encoding a type III glutamine synthetase. *J. Gen. Microbiol.* **139**, 1487-1493.

Goodsell, D.S. & Olson, A.J. (2000). Structural Symmetry and Protein Function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105-153.

Gouaux, J.E., Braha, O., Hobaugh, M.R., Song, L., Cheley, S., Shustak, C. & Bayley, H. (1994). Subunit stoichiometry of staphylococcal alpha-hemolysin in crystals and on membranes: a heptameric transmembrane pore. *Proc. Nat. Acad. Sci. U.S.A.* **91**(26), 12828-12831.

Grigorieff, N. (2000). Resolution measurement in structures derived from single particles. *Acta Crystallogr*. **D56**(10), 1270-1277.

Grosse-Kunstleve, R.W. & Adams, P.D. (2003). Substructure search procedures for macromolecular structures. *Acta Crystallogr*. **D59**, 1966-1973.

Harauz, G. & van Heel, M. (1986). Exact filters for general geometry three-dimensional reconstruction. *Optik*, **73**, 146-156.

Harth, G., Clemens, D.L. & Horwitz, M.A. (1994). Glutamine synthetase of *Mycobacterium tuberculosis*: extracellular release and characterization of its enzymatic activity. *Proc. Nat. Acad. Sci. U.S.A.* **91**(20), 9342-9346.

Harth, G. & Horwitz, M. A. (1999). An inhibitor of exported *Mycobacterium tuberculosis* glutamine synthetase selectively blocks the growth of pathogenic mycobacteria in axenic culture and in human monocytes: extracellular proteins as potential novel drug targets. *J. Exp. Med.* **189**(9), 1425-1436.

Haschemeyer, R.H. & De Harven, E. (1974). Electron microscopy of enzymes. *Annu. Rev. Biochem*, **43**, 279-301.

Haschemeyer, R.H. (1966). *Division Biological Chemistry, American Chemistry Society conference proceedings*, New York 12-16 September 1966. New York, ACS, 46.

Haschemeyer, R.H., Wall, J.S., Hainfeld, J. & Maurizi, M.R. (1982). Scanning transmission electron microscopy of submolecular oligomers of stabilized glutamine synthetase from Eschericia coli. *J. Biol. Chem.* **257**(12), 7252-7253.

He, Y.X., Gui, L., Liu, Y.Z., Du, Y., Zhou, Y., Li, P. & Zhou, C.Z. (2009). Crystal structure of *Saccharomyces cerevisiae* glutamine synthetase Gln1 suggests a nanotube-like supramolecular assembly. *Proteins*, **76**(1), 249-54.

Heidner, E.G., Frey, T.G., Held, U., Weissman, L.J., Fenna, R.E., Lei, M., Harel, M., Kabsch, H., Sweet, R.M. & Eisenberg, D. (1978). New crystal forms of glutamine synthetase and implications for the molecular structure. *J. Mol. Biol*. **122**(2), 163-73.

Henrick, K., Newman, R., Tagari, M. & Chagoyen, M. (2003). EMDep: a web-based system for the deposition and validation of high-resolution electron microscopy macromolecular structural information. *J. Struct. Biol*. **144**(1-2), 228-37.

Hill, R.T., Parker, J.R., Goodman, H.J., Jones, D.T. & Woods, D.R. (1989). Molecular analysis of a novel glutamine synthetase of the anaerobe *Bacteroides fragilis*. *J. Gen. Microbiol.* **135**(12), 3271-3279.

Hofmeister, F. (1888). Zur lehre der wirkung der salze. Zweite mittheilung. *Arch. Exp. Pathol. Pharmakol*. **24**, 247-260.

Hubbard, S.J. (1998). The structural aspects of limited proteolysis of native proteins. *Biochim. Biophys. Acta*, **1382**, 191-206.

Hutchinson, E.G. & Thornton, J.M. (1996). PROMOTIF-A program to identify and analyze structural motifs in proteins. *Protein Sci.* **5**(2), 212-20.

Ingham, K.C. (1984). Protein precipiation with polyethylene glycol. *Methods Enzymol.* **104**, 351-356.

ISIS^{TM}/Draw (Version 2.4) (2001). MDL Information Systems, Inc., computer program.

Jaenicke, L. & Berson, W. (1977). Glutamine synthetase from pig brain: binding of adenosine triphosphate. *Hoppe-Seyler's Z. Physiol. Chem.* **358**(7), 883-889.

Janson, C.A., Kayne, P.S., Almassy, R.J., Grunstein, M. & Eisenberg, D. (1986). Sequence of glutamine synthetase from *Salmonella typhimurium* and implications for the protein structure. *Gene*, **46**(2-3), 297-300.

Joyeux, L. & Penczek, P.A. (2002). Efficiency of 2D alignment methods. *Ultramicroscopy*, **92**(2), 33-46.

Kabsch, W. (2010). XDS. *Acta Crystallogr.* **D66**(2), 125-132.

Kessel, M., Frank, J. & Goldfarb, W. (1980). Averages of glutamine synthetase molecules as obtained with various stain and electron dose conditions. *J. Supramol. Struct.* **14**, 405-422.

Kiang, C.H. (2001). Single-particle study of protein assembly. *Phys. Rev. E*, **64**, 041911-1 – 041911-3.

Kinoshita, S., Isu, S., Kaneko, G., Yamada, H., Hara, T., Itoh, Y. & Watabe, S. (2009). The occurrence of eukaryotic type III glutamine synthetase in the marine diatom *Chaetoceros compressum*. *Mar. Genomics*, **2**(2), 103-111.

Kleywegt, G.J., Zou, J.Y., Kjeldgaard, M. & Jones T. A. (2006). Around O. In *International Tables for Crystallography, Vol. F. Crystallography of Biological Macromolecules* (Rossmann, M.G. & Arnold, E., eds), pp. 353-356, 366-367. Kluwer Academic Publishers, The Netherlands, Dordrecht.

Kleywegt, G.J. (1997). Validation of protein models from C[alpha] coordinates alone. *J. Mol. Biol.* **273**, 371-376.

Krajewski, W.W., Collins, R., Holmberg-Schiavone, L., Jones, T.A., Karlberg, T. & Mowbray, S.L. (2008). Crystal structures of mammalian glutamine synthetases illustrate substrate-induced conformational changes and provide opportunities for drug and herbicide design. *J. Mol. Biol.* **375**(1), 217-28.

Krajewski, W.W., Jones, T.A. & Mowbray, S.L. (2005). Structure of *Mycobacterium tuberculosis* glutamine synthetase in complex with a transition-state mimic provides functional insights. *Proc. Nat. Acad. Sci. U.S.A.* **102**(30), 10499-10504.

Krajewski, W.W., Collins, R., Holmberg-Schiavone, L., Jones, T.A., Karlberg, T. & Mowbray, S.L. (2008). Crystal structures of mammalian glutamine synthetases illustrate substrate-induced conformational changes and provide opportunities for drug and herbicide design. *J. Mol. Biol.* **375**, 217-228.

Krebs, H.A. (1935). Metabolism of amino-acids: The synthesis of glutamine from glutamic acid and ammonia, and the enzymic hydrolysis of glutamine in animal tissues. *Biochem. J.* **29**, 1951-1969.

Kumada, Y., Benson, D.R., Hillemann, D., Hosted, T.J., Rochefort, D.A., Thompson, C.J. Wohlleben, W., & Tateno, Y. (1993). Evolution of the glutamine synthetase gene, one of the oldest existing and functioning genes. *Proc. Nat. Acad. Sci. U.S.A.* **90**, 3009-3013.

Kumada, Y., Takano, E., Nagaoka, K., & Thompson, C. J. (1990). *Streptomyces hygroscopicus* has two glutamine synthetase genes. *J Bacteriol.* **172**(9), 5343-5351.

Kunath, W., Weiss, K., Sack-Kongehl, H., Kessel, M. & Zeitler, E. (1984). Time-resolved low-dose microscopy of glutamine synthetase molecules. *Ultramicroscopy*, **13**(3), 241-252.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**(259), 680-685.

Lamar, C. Jr. (1968). The duration of the inhibtion of glutamine synthetase by methionine sulfoximine. *Biochem. Pharmacol*. **17**(4), 636-40.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. & Higgins, D.G. (2007). ClustalW and ClustalX version 2. *Bioinformatics*, **23**(21), 2947-2948.

Laskowski, R.A. (2001). PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res*. **29**(1), 221-2.

Lee, B. & Richards, F.M. (1971). The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol*. **55**(3), 379-400.

Lei, M., Aebi, U., Goldsmith Heidner, E. & Eisenber, D. (1979). Limited proteolysis of glutamine synthetase is inhibited by glutamate and by feedback inhibitors. *J. Biol. Chem*. **254**(8), 3129-3134.

Leslie, A.G. (2006). The integration of macromolecular diffraction data. *Acta Crystallogr*. **D62**(1), 48-57.

Levy, E.D., Erba, E.B., Robinson, C.V. & Teichmann, S.A. (2008). Assembly reflects evolution of protein complexes. *Nature* **453**, 1262-1265.

Li,H., Pfluegl, G.M.U., Kiang, C., Gingery, M. & Eisenberg, D. (2002). Difficult phasing in solving the structure of eukaryotic glutamine synthetase II. XIX IUCR congress, Geneva, Switzerland 6-16 August 2002. *Acta Crystallogr*. **A58**(Supplement), C77.

Li, Y., Yang, G., Huang, X., Ye, B., Liu, M., Lin, Z., Li, C. & Cao, Z.A. (2009). Recombinant glutamine synthetase (GS) from *C. glutamicum* existed as both hexamers and dodecamers and C-terminal His-tag enhanced inclusion bodies formation in *E. coli*. *Appl. Biochem. Biotechnol*. **159**, 614-622.

Liao C.H. & McCallus D.E. (1998). Biochemical and genetic characterization of an extracellular protease from *Pseudomonas fluorescens* CY091. *Appl. Environ. Microbiol.* **64**(3), 914-921.

Liaw, S.H., Kuo, I. & Eisenberg, D. (1995). Discovery of the ammonium substrate site on glutamine synthetase, a third cation binding site. *Protein Sci*. **4**(11), 2358–2365.

Liaw, S.H., Jun, G. & Eisenberg, D. (1993a). Extending the diffraction limit of protein crystals: The example of glutamine synthetase from *Salmonella typhimurium* in the presence of its cofactor ATP. *Protein Sci*. **2**, 470-471.

Liaw, S.H., Pan, C., Eisenberg, D. (1993b). Feedback inhibition of fully unadenylylated glutamine synthetase from *Salmonella typhimurium* by glycine, alanine, and serine. *Proc. Nat. Acad. Sci. U.S.A.* **90**, 4996-5000.

Liaw, S.H., Jun, G. & Eisenberg, D. (1994). Interactions of nucleotides with fully unadenylated glutamine synthetase from *Salmonella typhimurium. Biochemistry*. **33** 11184-11188.

Llorca, O., Betti, M., González, J.M., Valencia, A., Márquez, A.J. & Valpuesta, J.M. (2006). The three-dimensional structure of an eukaryotic glutamine synthetase: Functional implications of its oligomeric structure. *J. Struct. Biol.* **156**(3), 469-479.

Lovell, S.C., Davis, I.W., Arendall, W.B. III., de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S. & Richardson, D.C. (2003). Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins*, **50**(3), 437-50.

Lumry, R. & Rajender, S. (1970). Enthalpy-entropy compensation phenomena in water solutions of proteins and small molecules: a ubiquitous property of water. *Biopolymers*, **9**, 1125-1227.

Manco, G., Rossi, M., Defez, R., Lamberti, A., Percuoco, G. & Iaccarino, M. (1992). Dissociation by NH4Cl treatment of the enzymatic activities of glutamina synthetase II from *Rhizobium leguminosarum* biovar viceae. *J. Gen. Microbiol*. **138**, 1453-1460.

Mathis, R., Gamas, P., Meyer, Y. & Cullimore, J. V. (2000). The presence of GSI-like genes in higher plants: support for the paralogous evolution of GSI and GSII. *J. Mol. Evol.* **50**, 116–22.

MATLAB (7.3.0.267), (2006), The MathWorks, computer program.

Maurizi, M. & Ginsburg, A. (1982a). Active site ligand stabilization of quaternary structures of glutamine synthetase from *Escherichia coli. J. Biol. Chem*. **257**(12), 7246-7251.

Maurizi, M. & Ginsburg, A. (1982b). Reactivation of Glutamine Synthetase from *Eschericia coli* after auto-inactivation with L-Methionine-S-sulphoximine, ATP, and Mn2+. *J. Biol. Chem*. **257**(8), 4271-4278.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. & Read, R.J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr*. **40**(4), 658-674.

McGuffin, L.J., Bryson, K. & Jones, D.T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, **16**, 404-405.

McKeller, R.C. & Cholette, H. (1987). Effect of temperature shifts on extracellular proteinase-specific mRNA Pools in *Pseudomonas fluorescens* B52. *Appl. Environ. Microbiol*. **53**(8), 1973-1976.

McParland, R.H., Guevara, J.G., Becker, R.R. & Evans, H.J. (1976). The purification and properties of the glutamine synthetase from the cytosol of Soya-bean root nodules. *Biochem. J*, **153**(3), 597-606.

McRee, D.E. & Israel, M. (2008). XtalView, protein structure solution and protein graphics, a short history. *J. Struct. Biol.* **163**(3), 208-213.

Meister, A. (1989). In *Glutamine Metabolism, Enzymology and Regulation*. (Mora, J. & Palacios, R. eds), pp. 1-40, Academic Press, New York.

Meng, E.C., Pettersen, E.F., Couch, G.S., Huang, C.C. & Ferrin, T.E. (2006). Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics*. **7**, 339.

Merrick, M. J. & Edwards, R. A. (1995). Nitrogen control in bacteria. *Microbiol.Rev.*, **59**(4), 604-622.

Miller, R.E., Shelton, E. & Stadtman, E.R. (1974). Zinc-induced paracrystalline aggregation of glutamine synthetase. *Arch. Biochem. Biophys*. **163**, 155-174.

Monod J. (1968). On symmetry and function in biological systems. Nobel Symp. Symmetry Funct. Biol. Syst. Macromol. Lev., 11th, Stockholm, 15-27. Wiley, New York.

Newman, J. (2006). A review of techniques for maximizing diffraction from a protein crystal *in stilla*. *Acta Crystallogr*. **D62**, 27-31.

Nicholas, K.B., Nicholas H.B. Jr. & Deerfield, D.W. II. (1997). GeneDoc: Analysis and Visualization of Genetic Variation. *EMBNEW.NEWS* **4,** 14.

Nilsson, M.T., Krajewski, W.W., Yellagunda, S., Prabhumurthy, S., Chamarahally, G.N. Siddamadappa, C., Srinivasa, B.R., Yahiaoui, S., Larhed, M., Karlén, A., Jones, T.A. & Mowbray, S.L. (2009). Structural basis for the inhibition of *Mycobacterium tuberculosis* glutamine synthetase by novel ATP-competitive inhibitors. *J. Mol. Biol*. **393**(2), 504-513.

Ornstein, L. (1964). Disc electrophoresis. I. Background and theory. *Ann. N.Y. Acad. Sci*. **121**, 321-349.

Obojska, A., Berlicki, L., Kafarski, P., Lejczak, B., Chicca, M., & Forlani, G. (2004). Herbicidal pyridyl derivatives of aminomethylene-bisphosphonic acid inhibit plant glutamine synthetase. *J. Agric. Food Chem.* **52**(11), 3337-3344.

Paint Shop Pro. (Version 8.1), (2003), Jasc Software, Inc., computer program.

Palacios, R. (1976). *Neurospora crassa* glutamine synthetase. Purification by affinity chromatography and characterization of subunit structure. *J. Biol. Chem.* **251**(15), 4787-4791.

Penczek, P.A., Zhu, J. & Frank, J. (1996). A common-lines based method for determining orientations for N > 3 particle projections simultaneously. *Ultramicroscopy*, **63**(3-4), 205-218.

Penczek, P., Radermacher, M. & Frank, J. (1992). Three-dimensional reconstruction of single particles embedded in ice. *Ultramicroscopy*, **40**(1), 33-53.

Pesole, G., Bozzetti, M. P., Lanave, C., Preparata, G., & Saccone, C. (1991). Glutamine synthetase gene evolution: a good molecular clock. *P. Natl. Acad. Sci. U.S.A.* **88**(2), 522-526.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. & Ferrin, T.E. (2004). UCSF Chimera-A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**(13), 1605-1612.

Pfluegl, G.M.U., Gill, D. & Eisenberg, D. (1996). How to avoid overfitting in refinement of structures with approximate non-crystallographic symmetry. High resolution structure refinement of glutamine synthetase reveals evidence for homotropic cooperative binding. *IUCR*, **24**, C238.

Pfluegl, G.M.U., Kiang, C-H. Gingery, M. & Eisenberg, D. (2002). Difficult phasing in solving the structure of eukaryotic glutamine synthetase II. *IUCR*, A**58**, C77.

Pushkin, A.V., Antonyuk, L.P., Solov'eva, N.A., Shubin, V.V., Evstigneeva, Z.G., Kretovich, V.L., Cherednikova, T.V., Tsuprun, V.L., Zograf, O.N. & Kiselev, N.A. (1985). Glutamine synthetases of pea leaf and seed cytosol. Structure and properties. *Biochim. Biophys. Acta*, **828**(3), 336-350.

PyMOL Molecular Graphics System, DeLano Scientific LLC, San Carlos, CA, U.S.A., computer program.

Quillin, M. L. & Matthews, B. W. (2000). Accurate calculation of the density of proteins. Acta Crystallogr. D56, 791-794.

Rasulov, A.S., Shakirov, Z.S., Evstigneeva, Z.G., Tsuprun, V.L. & Kretovich, V.L. (1986). Quaternary structure of glutamine synthetase of *Ankistrodesmus braunii*. *Biokhimiya (Mosc.)*, **51**(3), 413-419.

Reyes, J.C. & Florencio, F.J. (1994). A new type of glutamine synthetase in cyanobacteria: the protein encoded by the glnN gene supports nitrogen assimilation in *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* **176**(5), 1260-1267.

Reyes, J.C., Muro-Pastor, M.I. & Florencio, F.J. (1997). Transcription of glutamine synthetase genes (glnA and glnN) from the cyanobacterium *Synechocystis* sp. strain PCC 6803 is differently regulated in response to nitrogen availability. *J. Bacteriol.* **179**(8), 2678-2689.

Rhee, S.G., Chock, P.B. & Stadtman, E.R. (1985). Glutamine Synthetase from *Eschericia coli*. *Methods Enzymol.* **113**, 213-241.

Rhee, S., Chock, P., Wedler, F.C. & Sugiyama, Y. (1981). Subunit interaction in unadenylylated glutamine synthetase from *Eschericia coli*. *J. Biol. Chem.* **256**, 644-648.

Robertson, D.L. & Alberte, R.S. (1996). Isolation and characterization of the glutamine synthetase from the marine diatom *Skeletonema costatum*. *Plant Physiol.* **111**, 1169–1175.

Rocchia, W., Alexov, E. & Honig. B. (2001). Extending the applicability of the nonlinear Poisson-Boltzmann Equation: multiple dielectric constants and multivalent ions. *J. Phys. Chem.* **B105**(28), 6507-6514.

Rosenfeld, J., Capdevielle, J., Guillemot, J.C. & Ferrara, P. (1992). In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Anal. Biochem.* **203**(1), 173-179.

Rowe, W.B., Ronzio, R.A., Wellner, V.P. & Meister, A. (1970). Glutamine synthetase (Sheep Brain). *Methods Enzymol.* **17**(1), 900-910.

Sakakibara, H., Shimizu, H., Hase, T., Yamazaki,Y., Takao, T., Shimonishi,Y. & Sugiyama, T. (1996). Molecular identification and characterization of cytosolic isoforms of glutamine synthetase in maize roots. *J. Biol. Chem*. **271**(47), 29561-29568.

Sanchez-Weatherby, J., Bowler, M.W., Huet, J., Gobbo, A., Felisaz, F., Lavault, B., Moya, R., Kadlec, J., Ravelli, R.B. & Cipriani, F. (2009). Improving diffraction by humidity control: a novel device compatible with X-ray beamlines. *Acta Crystallogr*. **D65**, 1237-46.

Sarkar, P.K., Fischman, D.A., Goldwasser, E. & Moscana, A.A. (1972). Isolation and characterization of glutamine synthetase from chicken neural retina. *J. Biol. Chem*. **247**, (23), 7743-7749.

Saxton, W.O. & Baumeister, W. (1986). Principles of organization in S layers. *J. Mol. Biol.* **187**, 251–253.

Schein, C.H. (1990). Solubility as a function of protein structure and solvent components. *Bio/technology* **8**, 308-317.

Schuettelkopf, W. & van Aalten, D.M.F. (2004). PRODRG - A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr*. **D60**, 1355-1363.

Seabra, A.R., Carvalho, H. & Pereira, P.J.B. (2009). Crystallization and prelimary crytstallographic characterization of glutamine synthetase from *Medicago truncatula*. *Acta Crystallog.* **F65**, 1309-1312.

Shapiro, B.M. & Stadtman, E.R. (1970). Glutamine synthetase (Eschericia coli). *Methods Enzymol.* **17**, 910-922.

Sharp, K., Fine, R. & Honig, B. (1987). Computer simulations of the diffusion of a substrate to an active site of an enzyme. *Science*, **236**, 1460-1464.

Shatsky, M., Nussinov. R. & Wolfson, H.J. (2004). A method for simultaneous alignment of multiple protein structures. *Proteins*, **56**(1), 143-156.

Shrake, A., Whitley, E.J. & Ginsburg, A. (1980). Conformational differences between unadenylylated and adenylylated glutamine synthetase from *Eschericia coli* on binding L-methionine sulfoximine. *J. Biol. Chem*. **255**, 581-589.

SigmaPlot (windows version 8.0), (2002), SPSS Inc., computer program.

Singh, J., Joshi, M.C. & Bhatnagar, R. (2004). Cloning and expression of mycobacterial glutamine synthetase gene in Escherichia coli. *Biochem. Biophys. Res. Commun*. **317**, 634-638.

Smith, J. (1999). XIMDISP—A visualization tool to aid structure determination from electron microscope images. *J. Struct. Biol.* 125(2-3), 223-228.

Sobolev, V., Eyal, E., Gerzon, S., Potapov, V., Babor, M., Prilusky, J. & Edelman, M. (2005). SPACE: a suite of tools for protein structure prediction and analysis based on complementarity and environment. *Nucleic Acids Res*. **33**(web server issue), W39-W43.

Soper, D. (2004). Trichomoniasis: under control or undercontrolled? *Am. J. Obstet. Gynecol*. **190**(1), 281-90.

Southern, J.A., Parker, J.R. & Woods, D.R. (1986). Expression and purification of glutamine synthetase cloned from *Bacteroides fragilis*. *J. Gen. Microbiol*. **132**, 2827-2835.

Southern, J.A., Parker, J.R. & Woods, D.R. (1987). Novel structure, properties, and inactivation of glutamine synthetase cloned from *Bacteroides fragilis*. *J. Gen. Microbiol*. **133**, 2437-2446.

Speck, J.F. (1949). The enzymatic synthesis of glutamine, a reaction utilizing adenosine triphosphate. *J. Biol. Chem*. **179**, 1405-1426

Stadtman, E.R. (1990). Discovery of glutamine synthetase cascade. *Methods Enzymol*. **182**, 793-809.

Stadtman, E.R., Hohman, R.J., Davis, J.N., Wittenberger, M.,Chock, P.B. & Rhee, S.G. (1980). Subunit interaction of adenylylated glutamine synthetase. *Mol. Biol. Biochem. Biophys*. **32**, 144-156.

Stagg, S.M., Lander G.C., Pulokas, J., Fellmann, D., Cheng, A., Quispe, J.D., Mallick, S.P., Avila, R.M., Carragher, B. & Potter, C.S. (2006). Automated cryo-EM data acquisition and analysis of 284742 particles of GroEL. *J. Struct. Biol*. **155**, 470-481.

Streicher, S.L. & Tyler, B. (1980). Purification of glutamine synthetase from a variety of bacteria. *J. Bacteriol*. **142**(1), 69-78.

Tate, S.S., Leu, F.Y. & Meister, A. (1972). Rat liver glutamine synthetase. Preparation, properties, and mechanism of inhibition by carbamyl phosphate. *J. Biol. Chem*. **247**(17), 5312-5321.

Terwilliger, T.C. (2000). Maximum likelihood density modification. *Acta Crystallogr*. **D56**, 965-972.

Thain, A., Gaston, K., Jenkins, O. & Clarke, A.R. (1996). A method for the separation of GST fusion proteins from co-purifying GroEL. *Trends in Genetics*, **12**(6), 209-210.

Tiemeier, D.C. & Milman, G. (1972). Chinese hamster liver glutamine synthetase. Purification, physical and biochemical properties. *J. Biol. Chem.* **247**(8), 2272-2277.

TOPCAT - Tool for OPerations on Catalogues And Tables, (Version 3.5-1), (2009), Taylor, M., computer program.

Trapani, S. & Navaza, J. (2008). AMoRe: classical and modern. *Acta Crystallogr*. **64**(1), 11-6.

Tsuprun, V.L., Samsonidze, T.G., Radukina, N.A., Pushkin, A.V., Evstigneeva, Z.G. & Kretovich, W.L. (1980). Electron microscopy of glutamine synthetase from pea leaf chloroplasts. *Biochim. Biophys. Acta*, **626**(1), 1-4.

Tsuprun, V.L., Zograf, O.N., Orlova, E.V., Kiselev, N.A., Pushkin, A.V., Shiffelova, G.E., Solov'eva, N.A., Evstigneeva, Z.G. & Kretovich, W.L. (1987). Electron microscopy of multiple forms of glutamine synthetase from bacteroids and the cytosol of yellow lupine root nodules. *Biochim. Biophys. Acta*, **913**(3), 368-376.

Tsuprun, V.L., Boekema, E.J., Pushkin, A.V. & Tagunova, I.V. (1992). Electron microscopy and image analysis of the GroEL-like protein and its complexes with glutamine synthetase from pea leaves. *Biochim. Biophys. Acta*, **1099**, 67-73.

Tumba, N. (2007). Glutamine synthetase in *Bacteroides fragilis*. *MSc. Thesis*. University of Cape Town.

Tyler, B. (1978). Regulation of the assimilation of nitrogen compounds, *Annu. Rev. Biochem*. **47**, 1127-1162.

Unno, H., Uchida, T., Sugawara, H., Kurisu, G., Sugiyama, T., Yamaya, T., Sakakibara, H., Hase, T. & Kusunoki, M. (2006). Atomic structure of plant glutamine synthetase: a key enzyme for plant productivity. *J. Biol. Chem*. **281**(39), 29287-96.

Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Cryst*. **30**, 1022-1025.

Valentine, R.C., Shapiro, B. & Stadtman, E.R. (1968). Regulation of glutamine synthetase. XII. Electron microscopy of the enzyme from *Escherichia coli*. *Biochemistry*, **7**(6), 2143-2152.

van Heel, M.(1987). Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy*, **21**(2), 111-23.

van Heel, M. & Harauz, G. (1986). Resolution criteria for 3-dimensional reconstruction. *Optik* **73**, 119-122.

van Heel, M., Harauz, G., Orlova, E.V., Schmidtb, R. & Schatz, M. (1996). A new generation of the IMAGIC image processing system. *J. Struct. Biol*. **116**, 17-24.

van Rooyen, J.M. (2004). Three-dimensional structure of a Type III glutamine synthetase by single particle reconstruction. *MSc. Thesis*, University of Western Cape and the University of Cape Town.

Wallace, A.C., Laskowski, R.A. & Thornton, J.M. (1995). LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127-134.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A, Clamp, M. & Barton, G. J. (2009). Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**(9), 1189-1191.

Wedler, F.C., Sugiyama, Y. & Fisher, K.E. (1982). Catalytic cooperativity and subunit interactions in *Escherichia coli* glutamine synthetase: binding and kinetics with methionine sulfoximine and related inhibitors. *Biochemistry*, **21**(9), 2168-77.

White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W., Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S., Smith, H.O., Venter, J.C. & Fraser C.M. (1999). Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. *Science*, **286(**5444), 1571-1577.

Wilk, S., Meister, A. & Haschemeyer, R.H. (1969). Studies on the subunit structure of ovine brain glutamine synthetase. *Biochemistry*, **8**(8), 3168-3174.

Woolfolk, C.A., Shapiro, B. & Stadtman, E.R. (1966). Regulation of glutamine synthetase. I. Purification and properties of glutamine synthetase from *Escherichia coli*. *Arch. Biochem. Biophys.* **116**(1), 177-192.

Woolfolk, C.A. & Stadtman, E.R. (1967a). Regulation of glutamine synthetase - III. Cumulative feedback inhibition of glutamine synthetase from *Escherichia coli*. *Arch. Biochem. Biophys.* **118**, 736-755.

Woolfolk, C.A. & Stadtman, E.R. (1967b). Regulation of glutamine synthetase - IV. Reversible dissociation and inactivation of glutamine synthetase from *Escherichia coli* by the concerted action of EDTA and urea. *Arch. Biochem. Biophys.* **122**(1), 174-89.

Wriggers, W., Milligan, R.A. & McCammon, J.A. (1999). Situs: a package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.* **125**, 185-195.

Wyatt, K., White, H.E., Wang, L., Bateman, O.A., Slingsby, C., Orlova, E.V. & Wistow, G. (2006). Lengsin is a survivor of an ancient family of class I glutamine synthetases re-engineered by evolution for a role in the vertebrate lens. *Structure*, **14**(12), 1823-1834.

Xu, J., Bjursell, M.K., Himrod, J., Deng, S., Carmichael, L.K., Chiang, H.C., Hooper, L.V. & Gordon, J.I. (2003). A genomic view of the human - *Bacteroides thetaiotaomicron* symbiosis. *Science*, **299**(5615), 2074-2076.

Yamashita, M.M., Almassy, R.J., Janson, C.A., Cascio, D. & Eisenberg, D. (1989). Refined atomic model of glutamine synthetase at 3.5 Å resolution. *J. Biol. Chem*. **264**(30), 17681-17690.

Yanchunas, J., Dabrowski, M.J., Schurke, P. & Atkins, W.M. (1994). Supramolecular self-assembly of *Eschericia coli* glutamine synthetase: characterization of dodecamer stacking and high order association. *Biochemistry*, **33**(50), 14949-14956.

Zwart, P.H., Afonine, P.V., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., McKee, E., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., Storoni, L.C., Terwilliger, T.C. & Adams, P.D. (2008). Automated structure solution with the PHENIX suite. *Methods Mol. Biol*. **426**, 419-35.

Zwart, P.H., Grosse-kunstleve, R.W. & Adams, P.D. (2005). Xtriage and Fest: automatic assesment of X-ray data and substructure factor estimation. *CCP4 Newsletter*, **43**.