

**THREE-DIMENSIONAL STRUCTURE OF A TYPE III GLUTAMINE
SYNTHETASE BY SINGLE PARTICLE RECONSTRUCTION**

Jason Macrae van Rooyen

A minithesis submitted in partial fulfilment of the requirements of the degree of Magister Scientiae (Structural Biology) at the University of the Western Cape.

Supervisors:

Associate Professor Trevor Sewell
Dr. Val Abratt

December 2004

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

KEYWORDS

Glutamine synthetase type III

Structure

Bacteroides fragilis

Post translational regulation

Single particle

3D reconstruction

Electron microscopy

Docking

Bioinformatics

Protein

ABSTRACT

THREE-DIMENSIONAL STRUCTURE OF A TYPE III GLUTAMINE SYNTHETASE BY SINGLE PARTICLE RECONSTRUCTION

J.M. van Rooyen

MSc minithesis, Faculty of Natural Sciences, University of the Western Cape and Faculty of Science, University of Cape Town.

This study represents the first structural investigation of any type III glutamine synthetase (GS). The GS, GlnA, from the medically important opportunistic human pathogen *Bacteroides fragilis* was studied with a view to better understanding its structure/functioning in relation to the extensively characterised GSIs. GSIIIs are the most recently discovered family of GSs and are the most phylogenetically distant GSs from the GSIs.

Images (160) of negatively stained rGlnA, expressed in *E.coli* YMC11 (*glnA'*), were recorded at 50K magnification using a Leo 912 operating at 120kV with energy filtering coupled to a 4 megapixel CCD camera. An angular refinement based reconstruction strategy was adopted using SPIDER. A reconstruction based on an *ab initio* starting model, derived by a common-lines based simultaneous minimization of rotationally invariant K-mean clustered class averages, converged to the same structure as a reconstruction based on a GSI starting model to a resolution of 2.1nm as determined by Fourier shell correlation).

In contrast to preliminary EM observations, which identified GlnA as a hexamer, this work has revealed a dodecameric structure, with subunits (82.8KDa) arranged in two opposing hexagonal rings with distinct handedness. This is similar to the quaternary structure of GSIs and GlnTs except that the complex is 50% longer and the two rings are not symmetrically related. They differ not only in diameter (16.5 or 15.0nm) but also the degree of separation of subunits and as such the particle possesses only C6 and not D6 symmetry. The finding that particles lie in a preferred orientation, with the larger ring in contact with the carbon support, accounted for this asymmetry, through partial staining. Hexameric views, with similar overall arrangement but larger size in comparison to GSI, were also observed. However, it was uncertain whether these were true hexamers resulting from dissociation of the dodecamers or were a consequence of partial staining.

Homology modelling was also undertaken in an attempt to predict the structure of GlnA based on GSI, with a view to interpreting the low resolution EM structure. Due to the failures of state of the art algorithms in detecting the distant homologies between GS families, manual profile-based alignment strategies, incorporating structural information, were employed. Through the first full length alignments of GS sequences from all four families, conservation of all active site residues, core active site $\alpha\beta$ barrel fold motifs, and additional previously unreported regions was demonstrated. Docking of these homology models into the 3D structure confirmed the presence of the $\alpha\beta$ barrel fold predicted by the bioinformatic analysis of the sequences alone, thus, identifying the indentations between subunits in the volume as putative active sites.

In addition to providing unequivocal proof that GlnA is a GS and confirming the presence of putative $\alpha\beta$ barrel active site folds, this work has made steps towards understanding the regulation of this enzyme. It is hypothesised that GlnA occurs as both active hexamer and an inactive dodecamer, the interconversion of which, is thought to represent a means of reversible post-translational regulation.

December 2004

DECLARATION

I declare that *Three-Dimensional Structure of a Type III Glutamine Synthetase by Single Particle Reconstruction* is my own work, that it has not been submitted for any degree or examination in any other university, and that all sources I have used or quoted have been indicated and acknowledged by complete references.

Jason Macrae van Rooyen

16 December 2004

Signed:

A handwritten signature in black ink, appearing to read 'J. Macrae van Rooyen', with a large loop at the top and a horizontal line across the middle.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Assoc. Prof. Trevor Sewell and Dr. 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.

Finally, I gratefully acknowledge the financial assistance of the National Research Foundation and the Carnegie Corporation of New York.

LIST OF ABBREVIATIONS

3D PSSM	three-dimensional position specific scoring matrix
Ap	ampicillin
ATP	adenosine triphosphate
BSA	bovine serum albumin
C(x) e.g. C6	x-fold cyclical space group symmetry
CC score	cross-correlation score
CCD	charge coupled device
CTAB	cetyl trimethyl ammonium bromide
DPR	differential phase residual
D(x) e.g. D6	x-fold dihedral space group symmetry
EDTA	ethylenediamine tetra-acetic acid
EM	Electron microscopy/microscope
FEG	field emission gun emission source
FSC	Fourier shell correlation
GS	glutamine synthetase
LaB6	lanthium hexaboride emission source
LB	Luria broth
MetSox	methionine sulfoximine
MW	molecular weight
PDB	protein databank
PMSF	phenylmethanesulfonyl fluoride
SDS-PAGE	sodium dodecyl sulphate
TEM	transmission electron microscopy/microscope
UA	uranyl acetate

TABLE OF CONTENTS

Title page	1
Keywords	2
Abstract	3
Declaration	4
Acknowledgements	5
List of abbreviations	6
Table of contents	7
CHAPTER 1	
INTRODUCTION	10
1.1 Background	10
1.2 Literature review	13
1.3 Aims	39
CHAPTER 2	
BIOINFORMATICS	40
2.1 Introduction	40
2.1.1 Phylogeny	41
2.1.2 Structures	43
2.1.3 Techniques	43
2.2 Methods	46
2.2.1 Sequences	46
2.2.2 Automated alignment strategy	46
2.2.2.1 Profile based alignment	46
2.2.2.2 Automated alignment servers	47
2.2.3 Manual alignment strategy	47
2.2.4 3D Visualization	48
2.2.5 Modelling	48
2.3 Results	49
2.3.1 Automated alignment	49
2.3.1.1 Profile alignments	49
2.3.1.2 Improved alignments	49
2.3.2 Manual alignments	51
2.3.2.1 Conserved regions/residues	51

2.3.2.2	Secondary structure conservation	67
2.3.2.3	Alternate alignments	67
2.3.3	Models	71
2.4	Discussion	71
CHAPTER 3		
PURIFICATION OF RECOMBINANT BACTEROIDES FRAGILIS GLNA		77
3.1	Introduction	77
3.1.1	Purification	77
3.1.2	Regulation	78
3.2	Methods	78
3.2.1	Origin of clones and preparation of expression host	78
3.2.2	Growth conditions	79
3.2.3	Plasmid extraction and transformation	80
3.2.4	Expression	80
3.2.5	Harvesting and lysis	80
3.2.6	PEG precipitation	81
3.2.7	Gel filtration chromatography	81
3.2.8	Ultrafiltration	82
3.2.9	Ultracentrifugation	82
3.2.9	Measurement of activity	82
3.3	Results	82
3.3.1	Transformation	83
3.3.2	Expression and purification	83
3.3.2.1	PEG precipitation	84
3.3.2.2	Gel filtration	85
3.4	Discussion	89
CHAPTER 4		
SINGLE PARTICLE RECONSTRUCTIONS		93
4.1	Introduction	93
4.1.1	General background	93
4.1.2	Single particle reconstruction	94
4.2	Methods	95
4.2.1	Sample preparation	95
4.2.2	Negative stain electron microscopy	96
4.2.3	Image processing and reconstruction	96

4.2.3.1	Data handling	97
4.2.3.2	Particle selection	98
4.2.3.3	Preprocessing	98
4.2.3.4	2D Reconstruction	99
4.2.3.5	3D Reconstruction	99
4.2.3.6	Docking	104
4.3	Results	104
4.3.1	Visualization	104
4.3.2	Sampling	105
4.3.3	2D average of pinwheel views	109
4.3.4	3D Reconstruction	111
4.3.4.1	Structural insights	117
4.3.4.2	Statistics	121
4.3.5	Docking	128
4.4	Discussion	131
4.4.1	2D average of pinwheel views	131
4.4.2	3D Reconstruction	133
4.4.3	Docking	137
 CHAPTER 5		
	SUMMARY AND CONCLUSIONS	138
5.1	Summary	138
5.2	Discussion of the important findings	140
5.2.1	Stoichiometry and inactivation	140
5.2.2	Docking	141
5.2.3	Evolutionary relationships	142
5.3	Future directions	143
5.3.1	Biochemical characterisation	144
5.3.2	Structural studies	145
5.4	Conclusions	147
 BIBLIOGRAPHY		
		149

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Nitrogen metabolism is a fundamental aspect of cellular physiology and is thus ubiquitous to all higher life. It is through these metabolic pathways that a large number of vital biomolecules are synthesised. Consequently, nitrogen metabolism is subject to extensive regulation at every level from genetic to post-translational modifications of vital enzymes to ensure that metabolic requirements for growth are met under all conditions (Merrick & Edwards 1995). In particular, the glutamine synthetase (GS) enzyme plays a central role in nitrogen metabolism, catalysing the condensation of ammonium and glutamate to form glutamine, a precursor used for the synthesis of many other critical biomolecules. 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 & Edwards 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)).

Because of their importance and complexity, GS enzymes have been the subject of study for over four decades and a wealth of knowledge has accumulated about their function, regulation, and more recently, structure (Eisenberg, Gill, et al. 2000). The insights gained from this research have not only led to a far greater understanding of the regulation of nitrogen metabolism, but also served as a paradigm for post-translational regulation (Stadtman 2001). Due to their importance, GS enzymes are found in almost all life forms from primitive to higher organisms (Pesole, Bozzetti, et al. 1991). This family of ubiquitous enzymes is, therefore, large and evolutionarily diverse. The majority of insights have been derived from studies into GSIs (prokaryotic) and GSIIIs (primarily eukaryotic) (Eisenberg, Gill, et al. 2000).

It was discovered relatively recently (1986), that the GS possessed by *Bacteroides fragilis* represented a novel type of GS, termed a GSIII, which had previously not been recognized (Southern, Parker, et al. 1987). As such, very little is known about the functioning and in particular the regulation of this enzyme. With the recent sequencing of the *B. fragilis* genome, an additional putative GS has been discovered in the organism (personal communication, Abratt; Kuwahara, Yamashita, et al. 2004). This appears to be a GSI but has not been proven to be functional.

B. fragilis is a medically important anaerobic opportunistic human pathogen. It is the primary causative agent of abdominal infections following trauma to or surgery on the digestive tract (Gibson III, Onderdonk, et al. 1998). It is known that low levels of nitrogen, as presumed to be found outside the digestive tract, can induce the secretion of proteases by *B. fragilis* (Gibson & Macfarlane 1988). This in turn allows invasion of tissues by the necrotic bacteria. GS is a potential control point for the regulation of nitrogen metabolic pathways because of its unique ability to synthesise glutamine under conditions of low nitrogen. Ultimately it is hoped that research into the role played by this critical enzyme will lead to a better understanding of the relationship between the pathogenesis and nitrogen metabolism of *B. fragilis*. Furthermore, because of its unique ability to synthesise glutamine, GS could potentially offer a target for the control of this important organism, through the design of selective inhibitors against GSIII. GS has already been targeted in plants by herbicides (Obojska, Berlicki, et al. 2004), and research is currently underway to design specific inhibitors against the extracellular secreted GS of *Mycobacterium tuberculosis* (Harth & Horwitz 1999).

Structural analyses can reveal a great deal about the function and regulation of this enzyme. Insights gained through the study of the more well known GS enzymes can assist in inferring these directly from the structure, while at the same time reveal the unique features of this enzyme. This would enhance the body of knowledge about all GSs structure-function relationships. The evolutionary distance between GSIII and, GSI, the only GS enzyme for which an atomic resolution structure has been determined, is too great to permit the structure of GSIII to be determined solely by bioinformatics and modelling tools. Thus, the intended approach in this and future

studies is to use two complementary state-of-art structural biology techniques, namely: electron microscopy together with image processing and X-ray crystallography, to achieve this aim.

1.2 LITERATURE REVIEW

PREVIOUS GS STRUCTURAL STUDIES BY EM AND IMAGE PROCESSING

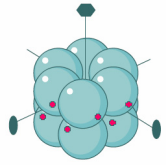
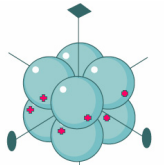
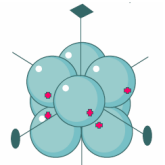
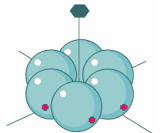
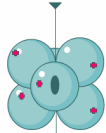
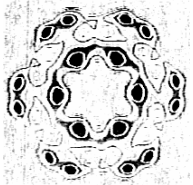
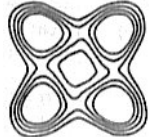

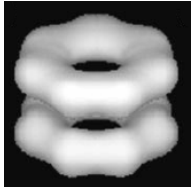
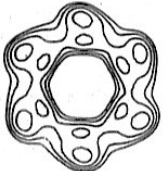
This is a historical review of the previous structural investigation of GS enzymes by EM and image processing. It aims to reveal the type of information that can be derived from EM studies, highlight any potential problems and motivate, via previous successes, why EM together with image processing has been chosen to investigate the structure of GSIII.

GS I

First direct visualization of GS structure:

The first efforts to directly determine the structure of GS were carried out in 1968 on the GS purified from *E. coli* (Valentine, Shapiro, et al. 1968). This collaboration grew out of the efforts in the lab of E.R. Stadtman to understand the regulation of the enzyme (Stadtman 2001). Prior to this study the only information about the quaternary structure of GS came from electrophoretic and ultracentrifugation studies which revealed stoichiometric and hydrodynamic volume of the GS complex. However, these results were often difficult to interpret and conflicting, for example, the resolution of density-gradient ultra-centrifugation results revealed that GS was composed of 12-14 identical subunits but a precise prediction was not possible (Woolfolk, Shapiro, et al. 1966). Electron microscopy was therefore employed to try and reveal the arrangement and stoichiometry of GS subunits. Additionally, biochemical studies had revealed that the GS enzyme could exist in two additional and distinct states, characterized by differing levels of activity, metal ion occupancy, and hydrodynamic volumes, from the native 'taut' enzyme. A catalytically inactive "relaxed" state results from treatment of native GS with EDTA at raised pH (~8) and subsequent treatment of this 'relaxed state' with 1M urea causes complete dissociation of the enzyme into its component subunits. Some catalytic activity can however be regained from the "relaxed" and even dissociated forms of the enzyme by the reintroduction of the divalent metal cations, Mn^{2+} , and lowering of the pH,

Table 1.1: Quaternary structures of glutamine synthetases investigated by EM

Family	GSI	GSII			GlnT	GSIII	?
Subunit stoichiometry	Dodecamer	Cubic octamer	Twisted cubic octamer	Tetradecamer	Dodecamer	Hexamer	Hexamer
Symmetry	D6 (622) 	D4 (422) 	D4 (422) 	D7 (722)	D6 (622)	C6 (6) 	D3 (32) 
Distribution	Prokaryotes & potentially higher plants as evidenced by gene sequence (Mathis, Gamas, et al. 2000)	Animals: rodent liver (Tate, Leu, et al. 1972), chicken neural retina (Sakar, Fischman, et al. 1972), ovine brain (Wilk, Meister, et al. 1970); Root nodule cytosol: yellow lupin (McPharland, Guevara, et al. 1976); Fungi: <i>N. crassa</i> (Palacios 1976); Green algae: <i>A. braunii</i> (Rasulov, Shakirov, et al. 1986).	Root cytosol: pumpkin; Leaf cytosol: pea; Leaf chloroplast: pea; Bacteroids: <i>R. lupine</i> . (Tsuprun, Zograf, et al. 1987).	Alfalfa and Humans (Kiang 2000).	Prokaryotes: <i>Rhizobium</i> (Espin, Moreno, et al. 1990), <i>Agrobacterium</i> .	Prokaryotes: Anaerobes - <i>Bacteroides fragilis</i> (Southern, Parker, et al. 1987), <i>Butyrivibrio fibrisolvens</i> (Goodman & Woods 1993) Cyanobacteria – <i>Synechocystis sp.</i> (Reyes & Florencio 1994).	Algae: <i>Chlorella</i> (Rasulov, Evstigneeva, et al. 1977).
Example reconstruction	 2D Reconstruction, low-dose, image processing (Kunath et al. 1984)	 (Tsuprun, Zograf, et al. 1987)	 (Tsuprun, Zograf, et al. 1987)	 (Kiang, C.H. 2001)	 (Tsuprun, Zograf, et al. 1987)	?	?

resulting in the formation of the “tightened” enzyme. The primary objective of this investigation was to ascertain whether changes in molecular architecture accompany the interconversion of these different states of the enzyme.

Direct visualization of GS molecular envelope by EM of negatively stained particles 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 D_6). This arrangement was inferred by Valentine *et al.* from the three principal views present in the micrographs: hexagonal ring shaped particles, tetramers, and double bar-like particles, with the latter two views interpreted as side-on projections of the double ring. The distinctive appearance of six-fold symmetry in the front-on views led to the conclusion that the rings must be eclipsed otherwise a homogenous ring would have been present. The tetramer was thought to be formed by the ring standing edge-on in the negative stain. In such an orientation the eight subunits, 4 at the top and 4 at bottom of the molecule would be aligned in projection and produce strong visible contrast in the negative stain compared to the 4 subunits at the sides of the hexamer which barely project into the stain and produce very little contrast as a consequence (Haschemeyer & de Harven 1974). Similarly, the double bar-like projections were thought to represent side-on views where the subunits at the top and bottom of the ring were not aligned and the negative stain between the subunit could, therefore, not superimpose to produce a distinctive divide between the subunits as in the tetramer. It has been noted that this experiment represented one of the simplest examples of image interpretation because of the abundance of these clear and recognisable projections (Haschemeyer & de Harven 1974). The argument of dihedral symmetry came from the double layered appearance seen in the side views. 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, which result in dihedral symmetry, prevent the GS molecules from stacking into long polymers, which would occur with C_6 symmetry. Consequently, the fact that addition of divalent cations caused the linear polymerization of the molecule, led to the suggestion that, new

binding sites are formed on the outer surfaces by rearrangement of the molecular symmetry in the 'relaxed' state.

Measurement of the molecular dimensions was also possible after 5-fold averaging of images in three distinct molecular views in order to reduce the noise and enhance the features of the molecules. Each subunit was shown to be ellipsoidal with dimensions $45 \times 45 \times 53$ Å (the major axis is oriented radially, perpendicular to the 6-fold axis of symmetry), which was in agreement with previous hydrodynamic data (Woolfolk, Shapiro, et al. 1966) and the overall molecular dimensions was shown to be 140×90 Å.

Unfortunately, no distinction could be made between the molecular arrangement of the 'taut' and 'relaxed' forms, despite sedimentation data to the contrary. Examination of the 'tightened' form did, however, explain their characteristically low solubility by revealing the unexpected face-to-face stacking of GS molecules that produces extended hexagonal tubes, and given enough-time, large-paracrystalline aggregates by further lateral associations of these tubes. By examining the dissociation kinetics of the relaxed enzyme, exposed to 1M urea, the important discovery was made that breaks in the molecular architecture are more likely to occur between subunits in the same ring than between rings. This finding had important implications in understanding the forces stabilizing the GS molecule by illustrating that the isologous contacts between subunits of a ring are weaker than the heterologous contacts between rings, and the bonding energies involved would later be quantified by Eisenberg *et al.* according to changes in solvation energy on oligomerization (Almassy, Janson, et al. 1986). Furthermore, of practical importance, was the demonstration that the experimental conditions used to prepare a sample for negative staining, disrupted the molecular structure. Indeed, the addition of feedback inhibitors was thought to possibly facilitate this disruption. Examination of re-associated oligomers, similarly, helped explain the transient nature of activity recovered on addition of Mn^{2+} , by demonstrating the less-defined molecular envelopes of the reconstituted oligomers and their subsequent break-down and aggregation with time.

So despite not revealing, due to the limitations of applied technique at the time (either intrinsic resolution limitation of EM in 1968 due to a lack of image processing computers or the preparative effects of negative staining), the changes associated with transitions between the ‘taut’ and ‘relaxed’ forms of the enzyme, this study contributed significantly to the understanding of the molecular architecture and properties of GS from *E. coli*.

Despite employing rudimentary image averaging to reduce the noise level and enhance molecules features, the development of the full potential of EM and associated image processing techniques to reconstruct the molecule in three dimensions only took place 7 years later.

Low resolution Reconstruction from helical cables.

Frey *et al.* described the first 3-dimensional, low resolution, reconstruction of *E. coli* GS from electron micrographs of negatively stained molecules arranged in 7-stranded cables (Frey, Eisenberg, et al. 1975). As described above, GS was known to form paracrystalline aggregates after treatment with EDTA and subsequent reintroduction of the divalent metal cations: Mn^{2+} or Co^{2+} . It was known to Valentine *et al.* that addition of Co^{2+} also caused face-to-face stacking but these lateral associations were observed to form more random aggregates than the Mn^{2+} treated enzymes. The subsequent discovery that made the 3D-reconstruction possible was the finding that addition of Co^{2+} to ‘relaxed’, or even native ‘taut’ enzyme, at low temperatures (4°C) caused the GS molecules to stack face-to-face and to wind around each other forming ordered cables rather than random aggregates like the ‘wheat sheaths’ described by Valentine *et al.* (Valentine, Shapiro, et al. 1968). These aggregates were observed as single, three and, seven (6 strands surrounding a central cable) stranded cables. The discovery by these researchers that Co^{2+} induced linear polymerization of native enzymes was significant because it inferred that new binding sites, which were suggested by Valentine *et al.* to form in the transition to the relaxed state, are already present in the native enzyme but require activation by the divalent metal cations.

This reconstruction employed the then newly described technique of 3-dimensional reconstruction from electron micrographs of helical arrays of biological molecules (De Rosier & Klug 1968). Briefly, this method involves the reconstruction of the object from the optical diffraction pattern, derived from electron micrographs, using Fourier-Bessel techniques. Because, the helical lattice is determined from the distribution of maxima in the optical diffraction pattern, the technique takes advantage of the improved signal-to-noise ratio offered by the ordered arrangement of molecules resulting in higher resolution reconstructions. Consequently, the reconstruction of GS cables was carried out from micrographs of the 6+1 cable which contained 24 views of the GS molecules. These cables were $400 \pm 30 \text{ \AA}$ in diameter with a repeat unit of 8 GS molecules, each related to its neighbour by a rotation of 7.5° .

The 3-dimensional reconstruction of GS from the cables achieved a resolution of 30-50 \AA . The reason for the range in the estimate is because the resolution was anisotropic, i.e., the increased order present along the axial direction compared to the limited order present in the finite lateral diameter of the cable, results in higher resolution along the former (30 \AA) and lower resolution along the latter (50 \AA). The results of this reconstruction confirmed the molecular architecture and packing interactions described by Valentine *et al.* in their initial study (Valentine, Shapiro, et al. 1968), such as D6 symmetry and eclipsed subunit arrangements, but also provided some novel insights. Specifically, at higher resolution, it was seen that a disk-shaped void exists between molecules packed along the cable's axial direction, demonstrating that the surface of the hexameric rings are "dished-in". The higher resolution and 3-dimensionality of the data also allowed refined estimates of the subunit dimensions. The subunits were found to be ellipsoidal when viewed radially (perpendicular to the 6-fold axis) but they appeared circular in the axial direction (parallel to the 6-fold axis). As such, the subunits were characterised as "oblate ellipsoids of revolution". The minor and major axes were measured to be, 48 \AA and 63 \AA respectively.

It was subsequently shown by the same researchers that a distribution of helical pitches exists in these cables (Frey & Eisenberg 1984). In particular the helical pitch of 6+1 cables was found to vary from 320-540nm. By implication, the authors concluded that the variation in pitch must arise either from variations in intermolecular contacts, of

the order of 4-10 Å, along the cable or from movement of the protein domains responsible for bonding relative to the remainder of the GS molecule.

EM of crystals of GS from E. coli

Further confirmation of the symmetry and arrangement of the GS oligomers came from EM analysis of crystallization solutions (mother liquor) of *E. coli* GSI (Bywater, Carlisle, et al. 1975). In agreement with previous studies, dodecameric particles with subunit diameters of 45-50 Å were observed. Furthermore, the subunits making up the particles were also found to be non-spherical, as described previously. Linear polymers of GS molecules stacked along their six-fold axes and larger aggregates of these columns in lateral association were also observed.

However, despite the agreement with previous research these were not the only particles present in solution. The numerous other configurations observed included: (i) large open square tetramers, which could also have been interpreted as cubic octamers, were seen to aggregate into squarely packed arrays. The overall dimensions of these tetramers were measured at 120 Å, in agreement with the diagonal of the tetramer projection proposed by Valentine *et al.* (ii) Narrow rods with a diameter of 60 Å, which were thought to comprise a double row of particles (subunits) spaced at intervals of 40 Å. Interestingly, these rods were found to associate lengthwise in a parallel fashion, but not in close contact throughout their length, and also transversely by lying across each other at an angle of 45°. Other configurations such as pentamers, larger rings and globules were also noted.

Haschemeyer has summarized the usefulness of periodic arrays and aggregates of biological molecules for revealing additional information about the constituent molecules structure, which would otherwise not be possible from solution studies (Haschemeyer & de Harven 1974). Firstly, in the case of periodic arrays, the large number of identical orientations visible reinforce the validity of a particular projection. Secondly, such arrays are more amenable to measurements of molecular dimensions than molecules free in solution. Indeed, from the arrays seen, it was evident that the tetramers certainly do represent real projections and precise measurements of their

dimensions were made possible. However, the variety of aggregates formed by GS under the conditions of this experiment only serve to confuse the interpretation of the solution state quaternary structure of GS. Furthermore, because breakdown products of the GS microcrystals could not be visualized, no correlation between the type of aggregation seen in solution and productive crystallization could be reached by these researchers

It appears that under the specific crystallization conditions employed, different subunit associations are more stable compared to the conditions in solution where the dodecameric structure predominates. In a subsequent study on crystalline GS from *E.coli*, the packing of GS in the crystal was evaluated based upon low resolution diffraction data, in particular, the systematic absences (Kabsch, Kabsch, et al. 1976), and the known molecular arrangement previously determined by Valentine's EM studies. This study showed that GSI was able to crystallize as a dodecamer.

Low Resolution Reconstruction of GS using computer averaging

With the introduction of computer averaging, reconstructions were no longer constrained to periodic arrays of pre-aligned molecules like crystals and helices, which provide sufficient signal through interference phenomena to allow reconstruction. By employing this technique Frank *et al.* were able to reconstruct, to a low resolution, the structure of GS molecules dispersed in negative stain (Frank, Goldfarb, et al. 1978). Prior to the introduction computer averaging, averaging involved manually excising and translating the asymmetric units of the molecular lattice into coincidence to produce a stack or array of aligned images which were then averaged by optical filtering or as a montage (Frank, Goldfarb, et al. 1978). Automatic averaging, however, utilises a computer to arrange the dispersed particles *in silico*, rather than relying on the existence of an ordered lattice. This technique employs algorithms to determine the rotational and translational parameters relating the images of the particles to each other. A final average is produced by arithmetic summation of the pixel values in the aligned images with a subsequent improvement in the signal-to-noise ratio because the random nature of the noise present in the images, from such sources as structural variation in the particles and photographic

noise, are random in nature and therefore average out. Unlike visual alignment, as used in the first direct visualization of GS (Valentine, Shapiro, et al. 1968), computer alignment automates the process allowing a far greater number of particles to be averaged, thereby recovering higher resolution information. In this specific example, the researchers averaged ten times as many particle images (50) as Valentine *et al.* and when 6-fold rotational averaging was employed the reconstruction was derived from the average of 300 subunits (50×6). The reconstruction was based on the assumption that all the GS molecules interacted with the support film in an identical manner and as such the only views down the direction of the 6-fold axis were used.

The overall shape and dimensions of the reconstructed GS particle was found to agree with the results of the original study by Valentine *et al.* and the shape of GS in the helical cables (Valentine, Shapiro, et al. 1968; Frey, Eisenberg, et al. 1975). However, several novel insights were made that were not possible to derive from the previous studies, either because the resolution was too low, as in the case of Valentine's study, or because the stain distribution was effected by the molecular packing in the cables of Frey *et al.*'s study. Firstly, the stain distribution of the final reconstruction revealed prominent ellipsoidal stain excluding regions at the periphery of the hexameric ring (61 Å radially from the ring's centre), in turn linked by a smaller stain excluding spiral arms to the centre region. The dimensions of the major and minor axes of this ellipsoid were measured to be 42.5 and 24 Å, respectively, with the former being skewed by 6° to the normal of the molecule radius. Together the arm and ellipsoid occupied a circular area of 42.5 Å in agreement with the oblate ellipsoids of revolution seen by Valentine *et al.* (Valentine, Shapiro, et al. 1968). Secondly, the reconstruction was found to possess distinct handedness, which had not been observed by any previous research. This was unexpected because images formed by the electron microscope represent projections of the mass distribution of the molecules being viewed and since the subunits in GS are arranged by dihedral symmetry their projection should possess mirror symmetry. To account for this irreconcilability the authors proposed that a non-uniformity of staining was responsible. By interacting differentially with the alternate faces of the molecule, it was suggested that negative staining could introduce this handedness. Alternatively, the authors suggested that

differential shrinkage behaviour of the particle might distort the top face differently from the bottom face leading to a disruption of the point group symmetry.

Identification of the adenylylation site by immunohistochemistry

It is also worth mentioning the work by Frink *et al.* who investigated the GSI adenylylation site by immunoelectron microscopy (Frink, Eisenberg, et al. 1978). By visualizing the negatively stained GS complexed with antibodies, with affinity against a modified adenosine moiety, 1,*N*⁶-ethanoadenosine, and the fully adenylylated and modified GS, these researchers were able to locate the site of adenylylation to within 10 Å. Although this effort was not intended to reveal any details about the gross structure of GS it did provide valuable structural background to what was known about the functional regulation of GS at the time.

Low dose imaging

The approach of reducing random noise, from for instance: variation in stain distribution, electron shot noise and particle structural variation as a result of preparation artefacts and radiation damage (Kunath, Weiss, et al. 1984), by averaging large numbers of particles, leaving only the redundant and, therefore, significant structural features, has already been discussed. An alternate approach is to minimize the source of the noise. With the introduction of low dose techniques (less than 100e/nm²), which minimize the exposure of the biological sample to the probe radiation, structural variations arising from electron beam damage can be minimized. The critical dose for biological objects, or the number of electrons passing through a defined area that result in observable structural changes, is typically from 50-300 e⁻/nm² (Kiselev, Sherman, et al. 1990). However, because the number of electrons passing through the sample is reduced, the electron shot noise is increased. This reciprocal relationship represents a trade-off between the level of exposure and the number of images that have to be averaged to provide a significant signal-to-noise ratio. The method therefore relies heavily on image processing, which involves the computational alignment of particle projections via cross-correlation algorithms, and their subsequent arithmetic superposition to produce an average.

There have been two examples of EM investigations of GSI using the low-dose technique (Kessel, Frank, et al. 1980; Kunath, Weiss, et al. 1984). These studies served to demonstrate the usefulness of the low-dose technique and associated image processing for extracting high resolution information, and in so doing, revealed further details about GSI's structure. GS has often been chosen as an example for 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.

The first study by Kessel *et al.* utilized correlation averaging to reveal the subdivision of the GSI subunits into two stain-excluding domains (Kessel, Frank, et al. 1980) in a similar fashion to that seen by Frank *et al.* previously (see above) (Frank, Goldfarb, et al. 1978). With increasing dose, the stain was seen to migrate from this groove separating the subunits leading to a reduction in attainable resolution. Images were recorded at three alternate dose levels and due to the low contrast under minimum dose conditions, the alignment parameters for translation had to be determined from the higher dose images. Based on the Rose equation (Kuo & Glaeser 1975), which reveals the number of electrons required per unit area to define a detail of specific size, and the assumption that the predominant source of noise was shot noise, the number of images calculated to generate a statistically meaningful average was 45. The reconstructions were further 6-fold symmeterized, thus, increasing number of asymmetric units by 6-fold.

In addition to standard negative staining glucose/negative stain embedding was also tested. Due to the much lower contrast in the latter, "blind" alignment had to be used to align the particles, i.e. dual exposures were taken and the higher dose with associated higher contrast were used to align the particles and the transformation parameters then applied to the low dose images. The glucose/stain revealed similar details to the uranyl acetate staining but was much more sensitive to radiation damage as judged by loss of symmetry and thus did not offer any improvements over the latter technique.

The 2D reconstructions displayed clear handedness, as described above for previous studies of GSI, which again was attributed to preferential staining of one ring of the dodecamer. Proof of the partial staining was attained by recording images at high tilt and observing that the face-on views did not lengthen as expected for a cylindrical particle tilted about an axis normal to its long axis. Furthermore, side views did not transform into the face-on views as expected for completely stained particles. Severe flattening, leading to deformation of the particle diameter by as much as 20%, was also observed for face-on views in low stained regions.

The aim of the subsequent study by Kunath *et al.* was primarily to demonstrate that time-resolved low-dose EM could be used to isolate the sources of noise discussed above and secondly, to use the resulting high-resolution data to investigate the structure of GSI's different functional states (Kunath, Weiss, et al. 1984). Whereas previous studies of GSI had involved the enzyme from *E. coli*, Kunath *et al.* visualized GS from *S. typhimurium* thus allowing them to evaluate, to within the achievable resolution, biochemical data that suggested the two enzymes were very similar (Balakrishnan, Villafranca, et al. 1977).

By using a TV camera, associated digital signal processor, and storage system, Kunath *et al.* were able to produce a series of averaged GS molecules embedded in uranyl acetate at different levels of exposure. In order to generate such averages, image processing was used to correct for drift, isolate individual molecules and align them by cross-correlation techniques before superimposing them. In addition, the use of an inline e^- counter (scintillator) made it possible to monitor the signal-to-noise ratio and consequently the contrast levels in real-time. The reconstructions with the highest contrast, least noise and resolution estimated to be better than 20 Å, were obtained after averaging 32 particles, imposing 6-fold symmetry and then filtering out high resolution noise. The shot noise present in these reconstructions was further reduced by averaging five 1 $e^-/\text{Å}$ averages of time series resulting in an effective 5 $e^-/\text{Å}$ dose average. From contour plots of these reconstructions two rings, an inner ring surrounding the central hole and a larger ring at the periphery, both of which were segregated into six elements, were apparent. The appearance of the inner ring agrees with the observations of Kessel *et al.*, however, the segregation of the outer ring is

more pronounced in this instance. Furthermore, it was observed that for the low dose, $1 \text{ e}/\text{\AA}^2$ time series, both the peripheral and inner elements appear subdivided and a bridging density (stain-excluding region) was also seen between the inner and outer rings. The arrangement of these sub domains displayed distinct right-handedness. These results are similar to the findings of Frank *et al.* in their low resolution reconstruction discussed above, except that the subdivisions visible at higher resolution were not previously noticeable. However, progression of the time series revealed that the exposure to the electron beam ($30 \text{ e}/\text{\AA}$) caused a smoothing out of high resolution details mainly attributed to stain migration, which is a well documented occurrence (Kiselev, Sherman, et al. 1990).

At the time of publication the authors noted similarities between their reconstruction and preliminary 10 \AA data from X-ray investigations of GS crystals by Eisenberg *et al.* (Almassy, Janson, et al. 1986). This agreement was later reiterated by Kiselev *et al.* (Kiselev, Sherman, et al. 1990) and Tsurprun (Tsurprun, Zograf, et al. 1987) in a comparison of the above low-dose density maps with the 3.5 \AA crystal structure of *S. typhimurium* GS determined by Almassy *et al.* (Almassy, Janson, et al. 1986). The latter investigation confirmed the two domain nature of the GS subunit and identified a low density protein region where negative stain could penetrate to divide the molecule.

In both of the low-dose experiments discussed above, the segregation of the GS subunit into two distinct stain excluding domains was thought to correspond with previous biochemical evidence from limited proteolysis experiments. This data had revealed two cleavage products of 33 and 18 KDa, with similar mass proportions to the dual domains visualized by EM (Lei, Aebi, et al. 1979). However, at the much higher resolution provided by Eisenberg *et al.*'s crystal structure (Almassy, Janson, et al. 1986), it is obvious that the GSI subunit molecule is divided into a smaller N-terminal domain (11KDa) and a larger C-terminal domain (40KDa), with the susceptible proteolysis cleavage site falling within the latter domain and not at the boundary between them. In fact the high susceptibility of this region to proteolysis has been revealed by mapping the site to an extended loop region that protrudes into the central cavity of the GSI dodecamer (Almassy, Janson, et al. 1986). Furthermore, an

analysis of the face-on view of the crystal structure superimposed on the low resolution projections of GSI (this work) reveals that the tripartite division seen by Kunath *et al.* does not follow the domain arrangement of the subunit. Instead it appears that the partially eclipsed arrangement of the two opposing rings coupled with the screw appearance of the subunits results in the two peripheral density peaks and the stronger single inner peak, when the oligomers are observed in projection. The position of the low density regions between subunits, corresponding to active sites, is also in the correct position. The ‘docked’ structure agrees most closely with the 30 e/Å ensemble average because of the mirror symmetry of these projections. As discussed above, the dihedral symmetry of GS precludes the projections of the molecule from possessing handedness. However, in the low-dose ensembles seen by Kunath *et al.* a right-handed character is readily apparent. These results tend to suggest that the high resolution features observed at lower doses are the result of non-uniform stain distribution and possibly increased shot-noise. The most reliable or precise representation of GS, therefore, is only formed after rearrangement of the stain.

From the above discussion it is apparent that the understanding of GSI’s structure has paralleled developments in image processing because of the ease of interpretability and defined orientation preference of this particle.

An additional study worth mentioning at this point was the investigation into the dissociation products of GSI by Maurizi *et al.* As discussed above, the first visualization of GS by Valentine *et al.* also revealed a distribution of lower molecular weight dissociation products, attributed to preparation artefacts. This phenomenon has important implications for the interpretation of GS subunit stoichiometry data from EM. In the previous discussion it was pointed out that GS was known to exist in two principal states, the native or active ‘taut’ state and an inactive ‘relaxed’ state produced by removing the divalent metal cations from the enzyme, specifically the n1 high affinity site (Eisenberg, Gill, et al. 2000). In addition to their activity levels, these states are also characterized by differences in hydrodynamic volumes and susceptibility to disruption by denaturants. In this study an additional ‘extremely tight’ and inactive state, brought about by the tight binding of the inhibitor L-

methionine-S-sulfoximine (MetSox) to GS, was investigated. In this state the binding of MetSox-PO₄, ADP and 2 Mn²⁺ ions per subunit constitutes a very tight complex which stabilizes the intra- and inter-subunit bonds of the GS oligomer. Compared to the other possible states of the enzyme, the inhibitor-complex is extremely resistant to disruption. For instance, it has been shown that whereas the relaxed enzyme is completely dissociated into component subunits by 1M urea as described above, the native enzyme is less susceptible but can still be completely dissociated in less than 1min by 6M guanidine-HCl at pH 7.2 and a temperature of 37°C. This is compared to the inhibitor-complex which requires 4 hours under the same conditions (Maurizi & Ginsburg 1982). Because of this ability to strengthen subunit-subunit associations, the inhibitor-complex was an ideal target to investigate the dissociation/association kinetics of GS, allowing otherwise unstable intermediates to be observed. By investigating the distribution and rate of formation of oligomers when GS from *E. coli* was treated with EDTA (divalent metal ion chelator) and DTNB (sulphydryls disruptor), Maurizi *et al*, were able to show that the kinetics of dissociation were non-first order, with dissociation occurring in discontinuous fashion through multiple intermediates states of 4,6,8,10 and 12 subunits. The distribution of oligomers was also seen to be correlated with the degree of inhibition of the enzyme prior to dissociation. For example, tetramers predominated when less than 20% of the GS subunits were inhibited before dissociation. This was explained in terms of the stabilizing effects of the inhibitors described above. Due to their stability during several rounds of purification and long term storage, these intermediates were thought to represent stable oligomers rather than simple equilibrium distributions of aggregated species. Further confirmation came from ultra-centrifugation studies which showed distinct populations of oligomers and was therefore evidence for the lack of rapid association-dissociation equilibria among the species. Most importantly with the exception of the large aggregates, these oligomers, even partially saturated with MetSox inhibitor, all showed catalytic activity. Therefore, it was concluded that the dodecamer structure is not an absolute requirement for activity. It was also demonstrated that the dissociated oligomers can reassociate to form the dodecamer by addition of mercaptoethanol (to remove the TNB groups) and Mn²⁺ ions. Using scanning-transmission electron microscopy to visualize the oligomers and estimate their mass distributions, it was shown that, as in the study by Valentine *et al*, the

oligomers originate by disruption of the heterologous subunit contacts within each ring rather than between the rings.

This study does not reproduce *in vivo* conditions because of the presence of the inhibitor. It does, nonetheless, demonstrate that a distribution of breakdown products of GS exist (which might otherwise have not been visualized without the stabilizing conditions) and that they are also active. These findings have implications for the investigation of GS quaternary structures. Depending on the dissociation-association kinetics of the oligomeric complexes, specifically the rate of ensemble exchange between different intermediate states, and the conditions under which the structure is being investigated, the enzyme will be assumed to adopt different oligomeric states with associated stoichiometries.

Structural studies have also been undertaken into the other GS families.

GSII

The determination of the structure of GSII has not been as straightforward as GSI, primarily as the result of the difficulty interpreting the projections of the particles and as a result of the apparent instability of the complex. Interestingly, it appears that the visualization of the structure of GSII from sheep brain preceded that of GSI but consensus has yet to be reached on the quaternary structure of this enzyme. The efforts to investigate the structure of GSII have been reviewed by Haschemeyer (Haschemeyer & de Harven 1974) and Kiselev (Kiselev, Sherman, et al. 1990).

To date, three models have been proposed for the quaternary structure of the oligomeric GSII family:

Eclipsed Octamer D4

The structure of negatively stained ovine GSII was first investigated by Haschemeyer in 1966 (Haschemeyer 1966). The model proposed for the structure of GSII was that of a cubic octamer with the individual protomers arranged at the vertices of a two

opposing and eclipsed squares, thus possessing D4 symmetry. This model was based on the interpretation of the projections seen in the micrographs, which included: tetramers, double-bars and complex projections resembling the letters H, N or I. In comparison to the model for GSI derived by Valentine *et al.* (Valentine, Shapiro, et al. 1968), the model for GSII required that the subunits were not spherical in order to account for the appearance of the complex projections. Furthermore, it was suggested that the isologous contacts between two tetramer rings are made by only two of these non-spherical subunits. The double-bar appearance represented the most common projection and was explained by a lack of stain penetration between these dimers.

Similar images have also been reported for GSII isolated from a diverse assortment of tissues: pea seedling (see below) (Pushkin, Antonyuk, et al. 1985), human brain (Wilk, Meister, et al. 1970), rat liver (Tate, Leu, et al. 1972), chicken liver, neural retina (Sakar, Fischman, et al. 1972), hamster liver (Tiemeier & Milman 1972), and pig brain (Wilk, Meister, et al. 1969). Independently of the first study Haschemeyer *et al.*, similar models were inferred from investigations of the latter two enzymes. This conformation of eclipsed octameric GSII is not limited to animals. The first plant GS to be visualised, that from the soya bean root nodule cytosol (McPharland, Guevara, et al. 1976), shares a similar conformation. So does the green algae *Ankistrodesmus braunii* (Rasulov, Shakirov, et al. 1986).

However, doubts about the accuracy of the cubic octamer model for GSII were raised at a later date by Haschemeyer (Haschemeyer & de Harven 1974) who noted that since the octameric model depends heavily on tetrameric views with four-fold symmetry, and these images were not only rare but frequently accompanied by other unexplainable profiles, the structure of GSII was still not completely certain. To add further difficulties to the problem of interpretation, the views of GSII were likely to suffer from the same preferential staining problems as GSI due to their large size.

Partially eclipsed Octamer D4

In comparison to the studies of GSII from higher animals and the cytosol of root nodules, the studies of GSII from higher plants, their seeds, and prokaryotes

uncovered an eclipsed octameric quaternary structure. There are large discrepancies in the molecular weights of the different GSII complexes determined by biochemical methods (Eisenberg, Gill, et al. 2000). However, Kretovich *et al.* report that the GSIIIs with eclipsed octameric structures have molecular weights less than 400KDa compared to the partially eclipsed GSIIIs which are larger (Kretovich, Evstigneeva, et al. 1984).

Investigations into pea plants by Tsuprun *et al.* revealed four different types of glutamine synthetases (Tsuprun, Samsonidze, et al. 1980). These multiple forms of GS were identified as isozymes, with different physiochemical, kinetic, and immunochemical characteristics (Pushkin, Antonyuk, et al. 1985). Of these, only three were investigated in terms of quaternary structure: chloroplasts GS (480KDa), leaf cytosol (520KDa), and seed cytosol (380KDa). Prior to the studies by Tsuprun *et al.*, the only GS structure from higher plants was from soybean root nodule cytosol. This appeared like the GS of higher animals to possess a cubic arrangement. However, based on the interpretation of 4 distinct types of views (1 face-on and three side views) of pea plant GS, a model arrangement of subunits with D₄ symmetry was proposed. In this model the subunits were arranged into two planar tetramers which were only partially eclipsed down the 4-fold axis i.e. twisted by 40° with respect to each other. Computer correlation averaging and symmeterization, according to the symmetry determined from rotationally averaged power spectra, were used to improve the signal-to-noise ratio of the images. As for GSI, deviations from the mirror symmetry expected for views normal to the 2-fold axes of a particle with dihedral symmetry were observed. This too was accounted for by an uneven stain distribution i.e. preferential staining of either the top or bottom ring. Additionally, the appearance of pea seed cytosolic GS was annular with a large central hole, which was accounted for by the particles possessing an internal cavity that could be filled with stain.

Investigations into the GS possessed by nitrogen fixing bacteroids in root nodules revealed that GSII was not exclusive to eukaryotes and the partially eclipsed octameric arrangement, which is seen in higher plants, lead to the suggestion that the bacteria might have acquired GSII by horizontal gene transfer (Tsuprun, Zograf, et al.

1987). However, the discovery of GSII in other non-nitrogen fixing bacteria such as members of the *Streptomyces* (Kumada, Takano, et al. 1990) and *Frankiacea* (Rocheffort & Benson 1990) disproved this hypothesis. For instance, *Rhizobium lupine* possesses 3 forms of GS (see below).

Thus, despite the disparate results it appears from these early studies that all GSIIIs from both eukaryotes and prokaryotes occur as an octamer 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 arrangement of subunits into an octamer has important implications according to the half-site hypothesis (Eisenberg, Gill, et al. 2000). Because the active sites of GS are formed by the latch region from the N domain of one subunit and the remaining seven β sheets from α/β barrel of the C domain of an adjacent subunit, two restraints can be predicted. Firstly the number of active sites must be a multiple of two and secondly they need to be arranged in such a way as to maximize the number of complete sites formed by opposite sides of the protomers i.e. in a ring structure. However, for an octomeric arrangement with two tetrameric rings the positions of these regions only permits two active sites per ring i.e. half the number of active sites predicted for an octamer. In support of this, some GSIIIs, such as rat liver (Tate, Leu, et al. 1972), have been shown to bind only 4 equivalents of ATP per enzyme and 5 equivalents of the inhibitor MetSOX. On the contrary, evidence also exists to suggest that GSII, as in the case of ovine brain GS, can bind 8 equivalents of ATP and inhibitor (Wilk, Meister, et al. 1969). The latter findings suggest that domain flexibility must be responsible to allow four complete active sites to form per ring.

Annular Biplanar Tetradecamer D7

In stark contrast to the model proposed and held over the previous thirty years, investigation of GS from human brain tissue using single-particle techniques has recently suggested that the structure of GSII is a bilayered tetradecamer with D7 symmetry (Kiang 2001). Furthermore, it was found that the GSII oligomer existed in

a heterogenous quaternary structure, with three association/dissociation intermediates present: the tetradecamer, a planer heptamer, and completely dissociated monomers.

This study was the first application of single-particle 3D reconstruction techniques to the structure of any GS. A more detailed description of single particle techniques and the associated image-processing will be given in Chapter 4. For now it suffices to say that the technique involves reconstructing a 3D structure from a large number of images of particles embedded in vitreous water. CryoEM, unlike negative staining, requires the use of low-dose exposures ($<10 \text{ e}\text{\AA}^{-2}$ in this case) to prevent damage to the specimen, which are far more sensitive to beam damage without the protective effects of the stain. This together with the absence of electron scattering stain means that the contrast inherent in these images is much less than that of negatively stained particles, necessitating the use of image processing to extract statistically meaningful information from the noisy micrographs. As previously discussed for 2D alignments, image processing achieves this by improving the signal-to-noise ratio of representative views of particles using correlation algorithms to align particle views such that they may be averaged. However, a major distinction between 3D reconstruction and simple computer averaging is that in order to generate a 3D volume from the projections some means of determining the orientations of these views in 3D space is required. The previous 2D reconstruction attempts discussed above relied on the particular orientation preference of GS, i.e. the predominance of face-on GS views, so that the alignment involved only a 3 orientation parameter search, i.e. two translational and the azimuthal angle. Single-particle reconstructions are more complicated than simple 2D averaging requiring the orientation of each representative view to be explicitly determined (3 Euler angle parameters and 2 translation parameters) such that they can be ‘back-projected’ to form a volume. Determining the angular relationship between projections is the crux of solving the 3D structure. If prior information exists about the structure of the molecules or the views are easy to interpret then images can be matched to a template or model and the reconstruction performed according to this relationship (see Chapter 4). However, if no information exists about the structure, the relationship between the views can be determined empirically by first grouping similar projections and then using search

algorithms, based on the common-lines theorem, to determine their 3D relationship as in the common-lines based simultaneous minimization technique (see Chapter 4).

Unfortunately, Kiang does not describe the reconstruction and preparation methods used in detail, which makes it difficult to critically evaluate the contribution of this study. For instance, the number of particles used in the reconstruction is vital for assessing the resolution of the data, which in this case is claimed to extend to 25 Å. Since no mention of classifying techniques was made, it appears that the structure of the molecule was inferred by interpretation of the views present in the micrograph and based upon this model a reconstruction was performed from frontal and lateral views. Electron micrographs displayed the characteristic views seen previously for GSI and GSII, namely rounded projections corresponding to face-on views, and two types of lateral views: double-bars and tetramers. The trapezoidal projections, characteristic of a twisted octameric structure were not observed. However, a novel single-bar projection and isolated monomers were apparently obvious and interpreted as dissociation intermediates. The proposed double heptamer ring structure had a diameter of 14 nm, a height of 10 nm, and a central channel of 4 nm. These proportions together with the subunit separation of 4.5 nm are very similar to those of GSI and GlnT providing circumstantial evidence (within the uncertainty of measurements) for a large ring shaped molecule.

The quaternary structure of GS from human brain has previously been reported to be octameric with the subunits arranged in an eclipsed cubic configuration, in agreement with the structure of other mammalian brain GSs (Haschemeyer & de Harven 1974). The tetradecameric model proposed in the study by Kiang does not agree with these previous observations or with the molecular weight estimates made by direct measurement (chromatography and ultracentrifugation), which also indicate an octameric structure (Boksha, Schonfeld, et al. 2002). The observation of circular projections in this study together with their large diameter is the main argument against an eclipsed cubic configuration, which in all examples to date has a diameter less than 11 nm.

The author suggests that these differences can be reconciled by considering that previous experimental efforts to determine the structure have relied on ensemble averaged techniques. Within the timescale of negative staining or direct measurements like ultracentrifugation, Kiang argues that the rapid dissociation kinetics of GSII would lead to a distribution of oligomeric intermediates, but the heterogeneity of the population would not be resolved because they are too short lived. Single particle techniques applied to cryofixed sample, however, he argues, can reveal the heterogeneity of the population by trapping the intermediate dissociation products. However, previous dissociation studies argue against this explanation. The rapid dissociation kinetics of GSII are well characterized and even suggested as a means of regulating the enzymes activity (Denman & Wedler 1984), yet, even employing so called 'ensemble techniques', subpopulations of tetramers and octamers have been resolved under non-denaturing conditions (Palacios 1976). It, therefore, stands to reason that if these intermediates are stable enough to exist under such conditions then why are they not accounted for in this experiment? Rather, Kiang described the equilibrium as a distribution between novel intermediates: double ringed tetradecamers, single ring heptamers and monomers. The dissociation constants applied to this equilibrium originate from studies of octamer to tetramer dissociation (Denman & Wedler 1984) and it does not stand to reason that the measurements can simply be extrapolated to account for the new model. Furthermore without the use of classification analysis to group related projections it would be difficult to determine the orientation of particles. Interpretation is further complicated by the tetrameric appearance of lateral projections of GSII, which should be difficult to distinguish from dissociated tetrameric oligomers. These concerns do not, however, account for the main characteristic supporting the tetradecamer model, namely the occurrence of circular face-on views with apparent seven-fold symmetry.

Preliminary investigation into the quaternary structure of alfalfa GS using EM and crystallography appear to confirm the tetradecameric arrangement of subunit in GSII (Pfluegl et al. 2002). Due to difficulties in phasing the crystal data, only low resolution structural information is available. Both micrographs of the GSII and the results of self-rotation function calculated from the crystal diffraction data revealed a 7-fold symmetry axis.

GS glnT

The GS from *Rhizobium lupine*, with a molecular mass of 600KDa, was revealed by negative stain EM and correlation averaging to be large dodecameric complex (13-14nm diameter & 9nm height) composed of two hexagonal rings arranged with D6 symmetry, in a very similar fashion to GSI (Tsuprun, Zograf, et al. 1987). This similarity also extended to the subdomain structure, which was bilobed as in GSI, and to the left-handedness of the reconstructions. Furthermore, the GlnT appeared to form stacks as described for GSI. Thus, the similarities in quaternary structure reflect the similarity in sequence (~20%) between these GSs (see Chapter 2).

Confusingly, because some *Rhizobiaceae* spp. species were discovered to possess an additional GS to their GSIs and GSIIIs, when this new GS was discovered it was named GSIII. However, because of their similarity to GSIs (primarily their similar sequence lengths compared to the greater variety between the other families, see Chapter 2) this group was not recognised as a completely separate group by all researchers. As such, when the novel GS from *Bacteroides fragilis* was discovered, it too was termed GSIII and the term GlnT adopted for the third type of GS discovered in the *Rhizobium* spp.

GSIII

EM investigations into the structure of GS III have only been undertaken on a single occasion as part of studies into physicochemical properties of the enzyme from *B. fragilis* (Southern, Parker, et al. 1987). These investigations were preliminary and not intended as a rigorous determination of the oligomeric arrangement. They rather served to verify predictions about the hexameric subunit stoichiometry based on hydrodynamic data from gel filtration studies and SDS-PAGE, which gave molecular weight estimates for the oligomer and subunits of 490 and 75KDa respectively (Southern 1986). Subsequent sequencing of the *B. fragilis glnA* gene revealed an open reading frame of 729 amino acids with a combined predicted MW of 82.8KDa (Hill, Parker, et al. 1989). According to Southern *et al.* GSIII molecules appeared as a hexameric rings (although distorted) with a diameter of 12 nm but no tetragonal side

views were observed. Due to the poor quality of the staining and the lack of wide field views it is difficult to confirm whether the projections can be interpreted in a statistically meaningful manner.

Uncategorized GS

Interestingly, a GS isolated from the eukaryotic algae, *Chlorella*, was shown, by SDS-electrophoresis to comprise 6, 53KDa subunits (Rasulov, Evstigneeva, et al. 1977). EM revealed the subunits to be arranged in a hexamer with 32 point group symmetry (D3). This contrasts with later findings for another algae, *Ankistrodesmus braunii* already discussed above, where it was shown that both the GSs from the cytosol and chloroplast were octameric with D4 symmetry (Rasulov, Shakirov, et al. 1986). Unfortunately, because no sequence information exists for this protein or its gene sequence and the imprecision associated with electrophoretic size estimates, it is uncertain to which family the *Chlorella* GS belongs.

Summary

It is clear that EM and associated image processing techniques have played a crucial role in elucidating the structure of the GSs. Indeed, in the case of GSI, the ease of interpretation and defined orientation of GSI molecules ensured that these insights followed and in some cases motivated developments in image processing techniques. Although the insights derived from X-ray crystallography studies have exceeded those from EM due to the higher resolution of the latter technique, such atomic resolution structures only exist for two very similar (~ 47% sequence identity, see Chapter 2) GSs: the GSI from *S. typhimurium* and the GSI from *M. tuberculosis*. Furthermore, the structural insights from EM have occurred in a timely manner, with the first direct observation of GS structure by EM predating the first crystal structure by 18 years (see Figure 1.1).

Although, EM studies suffer less uncertainty than quaternary structure determinations by biochemical techniques such as chromatography, centrifugation, and electrophoresis (Gouaux, Braha, et al. 1994; Haschemeyer & De Harven 1974),

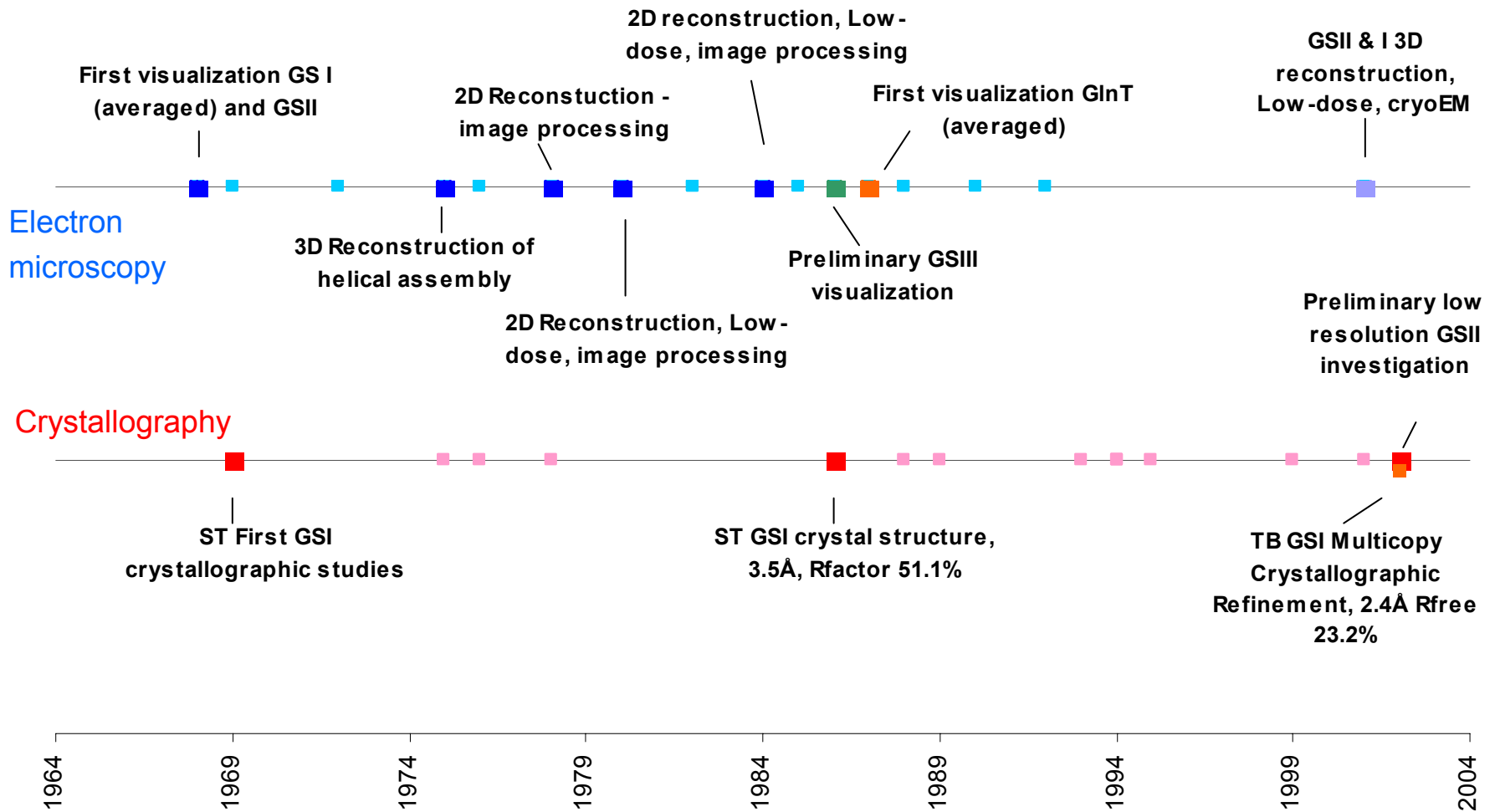


Figure 1.1. Timeline comparing GS structural studies by EM and X-ray crystallography.

conflicting results have been reported (Stahl & Jaenicke 1972; Tsuprun, Zograf, et al. 1987; Kiang 2000). In particular, the apparent heterogeneity of GS populations as a result of association-dissociation during preparation is a hindrance to interpretation of GS structure that needs to be avoided.

In a summary of the literature the following characteristic features have been noted for all glutamine synthetases except GSIII (Kretovich, Evstigneeva, et al. 1984). (1) GSs consist of an even number of subunits. (2) The number of monomers can equal 6,8, and 12. (3) GSs are arranged as double layered structures with dihedral point group symmetry.

1.3 AIMS

It is hoped that the insights gained through the structural investigation in GSIII will provide a detailed understanding of the structure, function and regulation of this important enzyme and ultimately make the design of an antibacterial drug against *B. fragilis* GSIII a possibility. Furthermore, determination of this structure would be only the second type of glutamine synthetase to ever be solved and thus have important implications for the broader understanding of the structure-function relationship and evolution of this important group of enzymes and enzymes in general.

The work presented here represents the first implementation of this investigation strategy, utilizing both bioinformatics and negative stain EM together with single particle reconstruction techniques to investigate the structure of the GSIII, GlnA, from *B. fragilis* with a view to better understanding the function and regulation of this important enzyme.

The main research questions this work seeks to answer are:

1. What is the oligomeric stoichiometry in solution and is the population homogeneous?
2. What is the 3D arrangement of subunits and substructure (domains) i.e. quaternary structure?
3. How does the structure relate to other GSs and do these structural relationships reflect what is seen at the sequence level?
4. Does enough similarity exist, at the structural level, between GSI and GSIII in order to interpret the density from negative stain EM? If so, could any inferences be made about the regulation/functioning of GSIII?

CHAPTER 2

BIOINFORMATICS

This chapter describes the parallel use of both automated and manual homology modelling strategies, relying on the same underlying bioinformatics principles, to predict the structure of GSIII with the aim of interpreting the EM structures to be investigated.

2.1 INTRODUCTION

The resolution afforded by single particle EM techniques, especially negative stain preparations illuminated with a standard thermionic source, is not high enough to allow the electron density derived during a reconstruction to be interpreted directly. It is, therefore, necessary to have prior knowledge of the subunit structure, in the form of an atomic model, permitting the electron density of the imaged complex to be interpreted by docking these building blocks into the map.

High resolution structures can be determined using biophysical techniques such as X-ray crystallography, NMR or electron crystallography, but these require lengthy experimentation. Alternatively, bioinformatic approaches offer the possibility of predicting the structure of a protein from its sequence alone. Several approaches are used at present: *ab initio*, threading (fold recognition), and comparative modelling.

The most successful of these to date is the latter, comparative or homology modelling, which is based on the fact that since structure is more conserved than sequence, proteins displaying a high level of sequence homology are likely to share similar structures. Homology modelling takes place in four distinct stages: (a) template structure selection (b) target-template alignment (c) model building (d) model evaluation. The success of homology modelling depends strongly on the accuracy of the sequence alignments, which allows subtle homologies to be detected, and also requires at least one atomic resolution structure. The higher the homology or similarity between sequences, the higher the confidence that they possess similar

structures or that conserved motifs represent conserved structural features. There is, therefore, a critical threshold of sequence similarity (<30%), called the twilight zone, below which it becomes exceedingly difficult for homology modelling to produce a meaningful structural prediction. This is not to say that comparative modelling cannot work below this threshold because, in spite of global dissimilarity, it is still possible to model the architecture (but not topology) of homologous sub-regions of a protein with confidence. The difficulty is in detecting this homology in the conservation of amino acid residues.

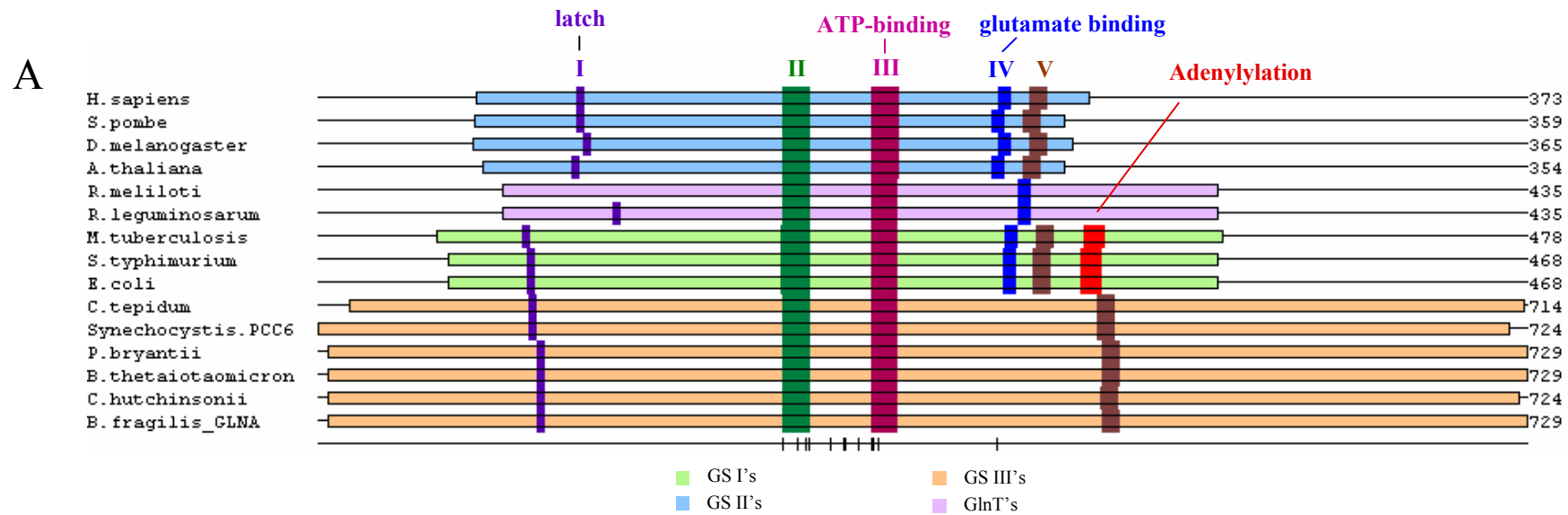
The predicted success of homology modelling can, therefore, be gauged by the degree of similarity between the target and template sequences and the existence of atomic resolution structures.

2.1.1 Phylogeny

The GS superfamily represents a diverse group of enzymes (Figure 2.1), as evidenced by their varied peptide lengths. Despite their overall diversity, five regions of high sequence similarity have been reported initially in GSIs (Almassy, Janson, et al. 1986; Janson, Kayne, et al. 1986) and subsequently in GSIIIs (Janssen, Jones, et al. 1988) and GSIIIs (Wen, Peng, et al. 2003; Reyes, et al. 1994b; Goodman & Woods 1993) (Figure 2.1). These regions are, consequently, the most promising targets for successful structure prediction by comparative modelling.

The previous studies, however, disagree on the designation of these regions. This can be mainly attributed to the diversity of sequences investigated. In particular, for the GSIII sequence from *B. fragilis* the repeated D-X-S-S- motif has led to alternate positions of this motif being reported (see below) (Hill, Parker, et al. 1989; Wen, Peng, et al. 2003). Furthermore, these two studies differ in the detection of regions IV and V. Hill *et al* report the absence of region IV but presence of the V, whereas Wen *et al* report the opposite.

With the determination of the atomic structure of the GSI from *S. typhimurium* (Almassy, Janson, et al. 1986), these five conserved regions were shown to be key



B

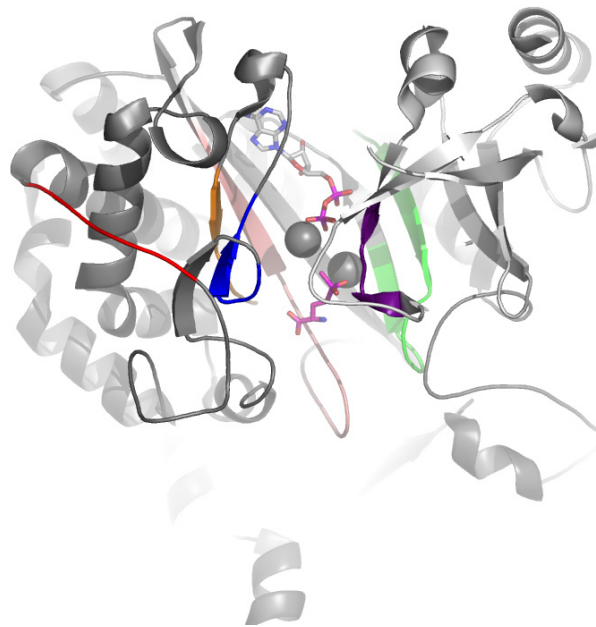


Figure 2.1. **A** - Ungapped alignment of representative sequences from all 4 GS families. Positions of the conserved regions/motifs are shown in colour, highlighting the conserved distance between the central motifs and the diversity in sequence length. Refer to Table 3 or Fig 2.2 for motif sequences. **B** – Cartoon representation of the structure of GSI from *S. typhimurium* (Alamassy *et al.* 1986). The five conserved regions are coded by the same colours used in **A**.

structural features of the active site α/β barrel (Figure 2.1B). Region I is a core component of the N-terminal latch, a region of β sheet contributed by the N domain of an adjacent subunit, which completes the active site. Regions II-V correspond to remaining β sheets from the C domain which comprise the staves of the barrel. Therefore, all five regions correspond to structural motifs enclosing the active site. Functions have been suggested for two of these regions, based upon sequence similarity to known ligand binding motifs from other proteins. Region III is thought to be the ATP-binding site because the sequence resembles the ATP-binding motif K-X(4)-G-X(2)-G-X-G-K-T found in several proteins ((Tischer, Das Sarma, et al. 1986) cited in (Hill, Parker, et al. 1989)). Region IV is thought to be the glutamate binding site because it resembles the glutamine-binding site D-R-G-A-S-I-V of two glutamine dehydrogenases ((Tischer, Das Sarma, et al. 1986) cited in (Hill, Parker, et al. 1989)).

These conserved regions are near the middle of the protein sequences indicating that the majority of the divergence between these sequences lies in the termini. This is especially true in the case of GSIIIs, which are on average 362, 290, and 254 amino acids longer than the GSIIIs, GlnTs, and GSIs respectively. The distance between the two central most conserved motifs (42 a.a. residues), however, remains the same between all the sequences (Figure 2.1).

2.1.2 Structures

Although there have been significant EM investigations into the structures of representatives of all the GS families (Chapter 1), to date atomic resolution structures, derived by X-ray crystallography, only exist for two GSIs: *S. typhimurium* (Almassy, Janson, et al. 1986) (reviewed by (Eisenberg, Gill, et al. 2000)) and *M. tuberculosis* (Gill, Pfluegl, et al. 2002).

2.1.3 Techniques

Modern protein structure prediction strategies utilize a combination of approaches and integrate information from a number of sources. This increases the accuracy of

alignments and, therefore, the chance of detecting more subtle relationships and extending the limits of successful prediction. Such strategies include:

Ab initio secondary structure prediction. Secondary structures have conserved patterns of solvent accessibility because of the way they pack against the hydrophobic core of proteins (Jones 1999b). Solvation parameters can, in turn, be predicted for amino acid residues based on their level of conservation in multiple sequence alignments. The most conserved regions of the proteins will correspond to this conserved core and the regions required for catalysis, whereas, the most variable regions are likely to be accessible to solvent (Jones 1999b). Thus, secondary structure can be predicted via the use of neural networks which have been trained to recognise such patterns of solvation parameters, as implemented in the PSIPRED server (McGuffin, Bryson, et al. 2000; Sonnhammer, Eddy, et al. 1997). This information is then used as an additional constraint in the alignment process. For instance, higher gap penalties are calculated for insertions within regions of secondary structure.

Threading or fold prediction. This technique aims to evaluate the quality of an initial alignment of a target protein sequence against a template with a known structure. This is done by calculating a scoring function based on the local environment (mean force pair wise potentials and solvation energy) of each target residue according to its position in the template as determined by the alignment. In the guise of fold recognition as performed by GenTHREADER (McGuffin, Bryson, et al. 2000), candidate templates are first identified by sequence alignments and these are then evaluated by threading. The template exhibiting the highest threading score, i.e. the lowest value of the energy function, is thus a likely prediction for the fold of the target. Threading can, therefore, predict the fold of a protein sequence but in this implementation there is no optimization of the initial alignment based on the threading energy of the correctly identified fold.

The use of **3D profiles**, on the other hand, achieves this aim of using structural information to improve alignment quality. 1D profile based alignments are able to increase the likelihood of detecting relationships between distant sequences by utilising position specific scoring tables (PSSM) to improve the sensitivity of the

alignment algorithms. Similarly, 3D profiles achieve enhanced sensitivity by incorporating information about structural homologies between proteins. Such profiles are generated by investigating the conservation of structurally important features of residues in their local environments. These include: bonding patterns, solvation parameters, and secondary structures, as determined from structural alignments of proteins contained in databases like HOMSTRAD (de Bakker, Bateman, et al. 2001) and FSSP (Holm & Sander 1994). This method can, thus, be thought of as a step up from sequence-sequence homology modelling, which works because structure is more conserved than sequence. Homology at the primary structure level, therefore, reflects homology at higher structure levels. 3D profile based alignments, on the other hand, involve sequence-structure alignments, which utilize information about conserved structural features to reveal subtle similarities in sequence.

The two automated alignment algorithms utilized in this work incorporate such features to differing degrees. mGenTHREADER's primary use is as a fold recognition tool (Jones 1999a; McGuffin & Jones 2003). However, the alignment strategy it employs to suggest initial templates incorporates a wide variety of information enabling it to detect distant homologies between sequences. Specifically it uses: (1) Bi-directional profile-based alignments, i.e. not only is the target sequence aligned to a profile of templates but individual template sequences are also aligned against a profile of generated from the target. (2) 3D PSSM's calculated from the FSSP (Holm & Sander 1994) database. (3) Secondary structure prediction from the PSIPRED server, which add additional constraints to the alignment. The second algorithm employed, FUGUE, is not as inclusive, but does utilize 3D profiles and structure-dependent gap-penalties to improve alignment accuracy (Shi, Blundell, et al. 2001). These 3D profiles are derived from multiple structure alignments in the HOMSTRAD database (de Bakker, Bateman, et al. 2001).

2.2 METHODS

2.2.1 Sequences

Table 2.1: GS Sequences used in this study

Family	Organism	Database identifier (entries marked with “gi” are Genbank identifiers. All others are Swissprot)	References
GSI	<i>Salmonella typhimurium</i>	P06201	(Janson, Kayne, et al. 1986)
	<i>Escherichia coli</i>	P06711	(Colombo & Villafranca 1986)
	<i>Mycobacterium tuberculosis</i>	Q10377	(Harth & Horwitz 1997)
GSII	<i>Homo sapiens</i>	P15104	(Gibbs, Campbell, et al. 1987)
	<i>Drosophila melanogaster</i>	P20478	(Caizzi, Bozzetti, et al. 1990)
	<i>Arabidopsis thaliana</i>	Q9LVI8	(Sato, Nakamura, et al. 2000)
	<i>Schizosaccharomyces pombe</i>	Q09179	(Wood, Gwilliam, et al. 2002)
GSIII	<i>Chlorobium tepidum TLS</i>	Gi:21674232	(Eisen, Nelson, et al. 2002)
	<i>Synechocystis</i> sp. (strain PCC 6803)	Q59982	(Reyes & Florencio 1994b)
	<i>Prevotella bryantii</i>	Q8RNI4	(Wen, Peng, et al. 2003)
	<i>Bacteroides thetaiotaomicron</i> BT0543 VPI-5482	Gi:29345953	(Xu, Bjursell, et al. 2003)
	<i>Cytophaga hutchinsonii</i>	Gi: 48856327	whole genome shotgun sequence NZ_AABD03000001
	<i>Bacteroides fragilis</i>	P15623	(Hill, Parker, et al. 1989)
GInT	<i>Rhizobium meliloti</i>	O87393	(Capela, Barloy-Hubler, et al. 2001)
	<i>Rhizobium leguminosarum</i>	P31592	(Chiurazzi, Meza, et al. 1992)

2.2.2 Automated alignment strategy

2.2.2.1 Profile based alignment

Profile based alignments of a number of sequences from each major GS family were performed using ClustalW (Thompson, Higgins, et al. 1994) accessed through the

ClustalX GUI (Thompson, Gibson, et al. 1997). The following parameters were used: Gonnet series weighting, gap opening penalties of 10, gap extension penalties of 0.2, divergent sequences delays of 30%, and residue specific penalties were enforced.

2.2.2.2 Automated alignment servers

The mGenTHREADER threading algorithm (Jones 1999a; McGuffin & Jones 2003) was accessed via the PSIPRED server (McGuffin, Bryson, et al. 2000). The sequence-structure alignment mode of the FUGUE, 3D-profile based sequence-structure alignment server (Shi, Blundell, et al. 2001), was used. The template 1f52 was chosen and a profile of GSIII sequence previously aligned with ClustalW (see below) was submitted.

2.2.3 Manual alignment strategy

The first step of the alignment procedure was the profile-based alignment of a number of sequences from each major GS family using ClustalW (Thompson, Higgins, et al. 1994) (see above). These profiles were then manually aligned to GSI profile using Genedoc (Nicholas & Nicholas 1997). The GSII family was aligned first, followed by the GlnTs and finally the GSIII profile. During this process structural information from previous studies (section 2.1.1) was incorporated in the form of the following constraints: (1) Conservation of regions I-V i.e. these served as anchor regions for the alignment. (2) Conservation of active site residues reported in the annotated alignments of GSIs and GSIIs by Eisenberg *et al* (Eisenberg, Gill, et al. 2000). (3) Conservation of an N-terminal motif reported in the Pfam database (Sonnhammer, Eddy, et al. 1997). A Blosum 62 scoring matrix was used for conservation shading and score reports. Regions with low sequence similarity were independently aligned using the dynamic programming algorithm (Feng & Doolittle 1987; Thompson, Higgins, et al. 1994) in DNAMAN (DNAMAN 1994). The parameters specified were: a gap opening penalty of 10, a gap extension penalty of 0.1, a Blosum scoring matrix, a % delay divergent sequences of 40, and no residue specific penalties were used.

The alignments were evaluated by inspecting the local environment of conserved residues in the GSI structure to determine whether their conservation was plausible. Information about the local environment of conserved residues automatically generated from structural alignments in the HOMSTRAD (de Bakker, Bateman, et al. 2001) database was also used in a similar manner. This information was in the form of sequence-sequence alignments between the two GSI structures of *S. typhimurium* and *M. tuberculosis* annotated with the JOY format (Mizuguchi, Deane, et al. 1998). The conservation of secondary structure was also evaluated by inspecting the patterns of hydrophobic amino acids.

2.2.4 3D Visualization

The molecular modelling and visualization packages SPDBV (Guex & Peitsch 1997), UCSF Chimera (Pettersen, Goddard, et al. 2004), and PyMOL (Delano 2004) were used for 3D visualization of the homology models and inspection of conserved residues.

2.2.5 Modelling

The homology models were based on the alternate alignments of *B. fragilis* GSIII and *S. typhimurium* GSI produced by the manual and automated methods. The atomic resolution crystal structure of GSI from *S. typhimurium* (PDB entry 2gls (Yamashita, Almasy, et al. 1989)) served as the template structure. The modelling procedure implemented was rudimentary and did not aim to model insertions in the GSIII sequence relative to the GSI template structure. However, regions corresponding to insertions in the template GSI sequence relative to the target GSIII sequence were deleted from the model in the UCSF Chimera package (Pettersen, Goddard, et al. 2004).

The automated modelling algorithms of the SWISS-MODEL (Guex & Peitsch 1997 ; Schwede, Kopp, et al. 2003) and 3D-PSSM (Schwede, Kopp, et al. 2003) servers were also evaluated. The fold recognition and the sequence-structure alignment

modes of SWISS-MODEL were used, while 3D-PSSM was used in the fold recognition and automatic model generation modes.

2.3 RESULTS

2.3.1 Automated alignment

2.3.1.1 Profile alignments

Initially, sequences from the GSI family were aligned independently. It was then attempted to align the GSIII sequences against this profile using PSSM's (data not shown). However, this was unsuccessful as typified by the misalignment of the region I. Alignment of a GSIII profile against the GSI profile was also unsuccessful.

2.3.1.2 Improved alignments

mGenTHREADER (McGuffin, Bryson, et al. 2000; McGuffin & Jones 2003; Jones 1999a) was therefore used in an attempt to automatically provide the most accurate alignments possible. It was decided to use mGenTHREADER as a first attempt because of the inclusiveness of the structural information incorporated into the alignments (see section 2.1.3).

The resultant alignment of *B. fragilis* GSIII against *S. typhimurium* GSI produced by mGenTHREADER is shown in Figure 2.2A. It can be seen that only regions II, III, and V were correctly aligned by the program (assuming true homology between these regions).

FUGUE (Shi, Blundell, et al. 2001) was also utilized in an attempt to automatically improve the quality of the alignments. The results are shown in Figure 2.2B. It can be seen that only regions II and III, the most highly conserved regions, were correctly aligned.

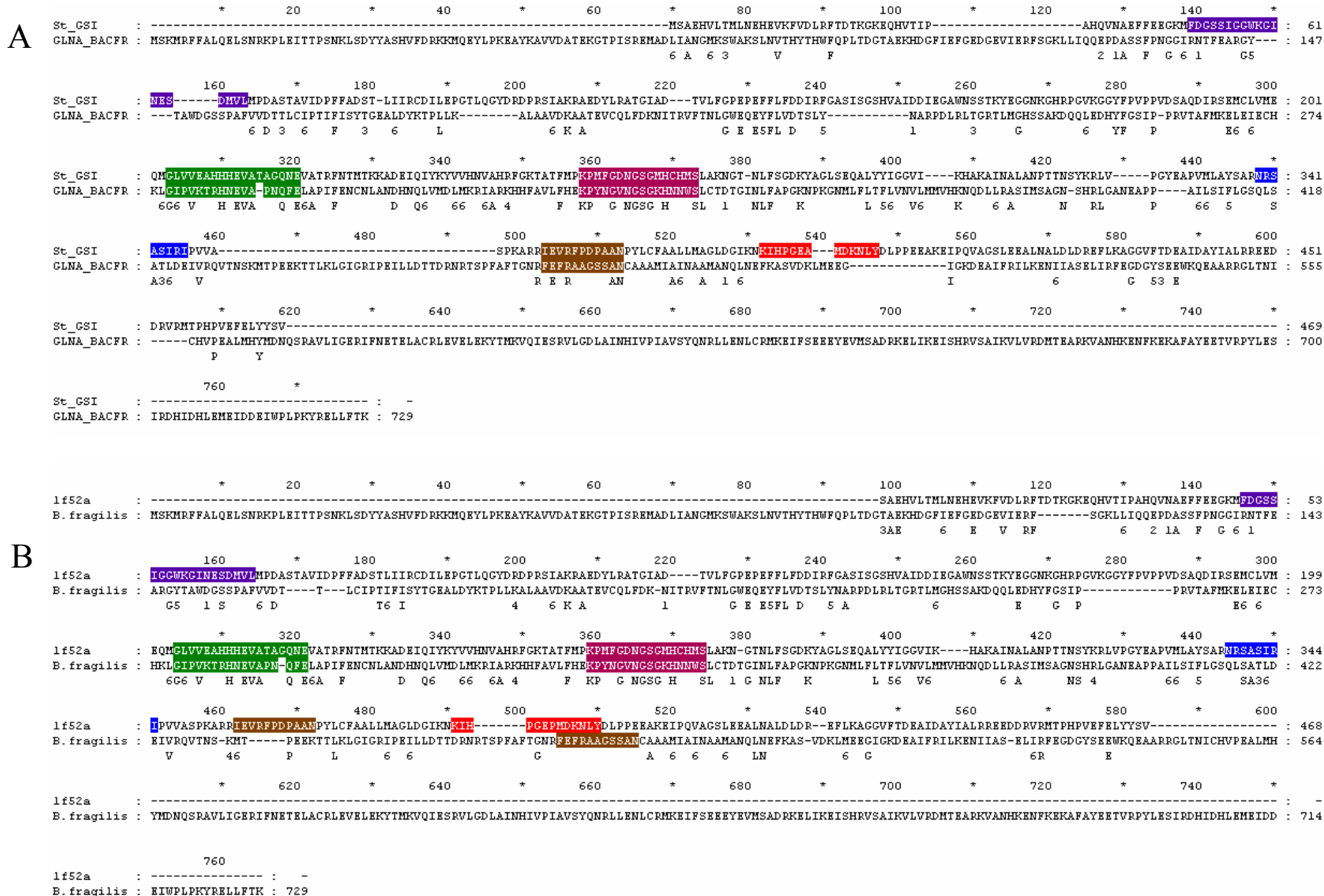


Figure 2.2. - Alignment between GSI and GSIII produced by mGenTHREADER (A) and FUGUE (B). The five conserved regions corresponding to the β sheets of active site $\alpha\beta$ barrel are shown in colour. They are designated as follows: Region (I) Latch region [PYF]-D-[GA]-S-S- ; Region (II) G-X(8)-E-[VD]-X(3,4)-Q-X-[EF]- ; Region (III) ATP-binding site K-P-[LIVMFYA]-X(3,5)-[NPAT]-G-[GSTAN]-G-X-H-X(3)-S ; Region (IV) Glutamate binding site [ND]-R-X(3)-[IV]-R-[IV]- ; Region (V) [ILF]-E-[FDV]-R-X(6)-[NDPS]-.

2.3.2 *Manual alignments*

The results of the automated alignment disagreed with the findings of previous researchers (section 2.1.1), who first identified the regions of sequence identity between GSIII and GSI (Guex & Peitsch 1997; Schwede, Kopp, et al. 2003). It was, therefore, decided to attempt to improve the alignments by incorporating these previous findings in the alignment scheme.

Details of the alignment strategy, programs, and algorithms used were given in section 2.2. Initial attempts to align individual GS family profiles to each other using profile based alignments failed, as judged by the misalignment of region I, necessitating manual alignment. Figures 2.3 & 2.4 show the results of the alignment strategy integrating the information described in section 2.2.3. During the alignment exercise, inspection of the Pfam database (Sonnhammer, Eddy, et al. 1997) entry for the N-terminal region of GSs revealed the strong conservation of a (FYW)-X(4)-G motif. This motif was discovered in all the GS sequences and the alignment was adjusted accordingly. Having produced the alignments, these were then evaluated according to more general structural principles and information about the conservation of residues in their local environments determined from aligned structures in the HOMSTRAD database (de Bakker, Bateman, et al. 2001).

2.3.2.1 Conserved regions/residues

The GS homologous superfamily can be divided into four sequence families with sequence identities $\geq 35\%$ (according to the CATH classification scheme (CATH--a hierarchic classification of protein domain structures 1997)) based on the alignments described above.

It can be seen from Table 2.2 that the global sequence identity between the GSIs, the only family for which structures exist, and all the other families and is very low and well within the twilight zone.

A

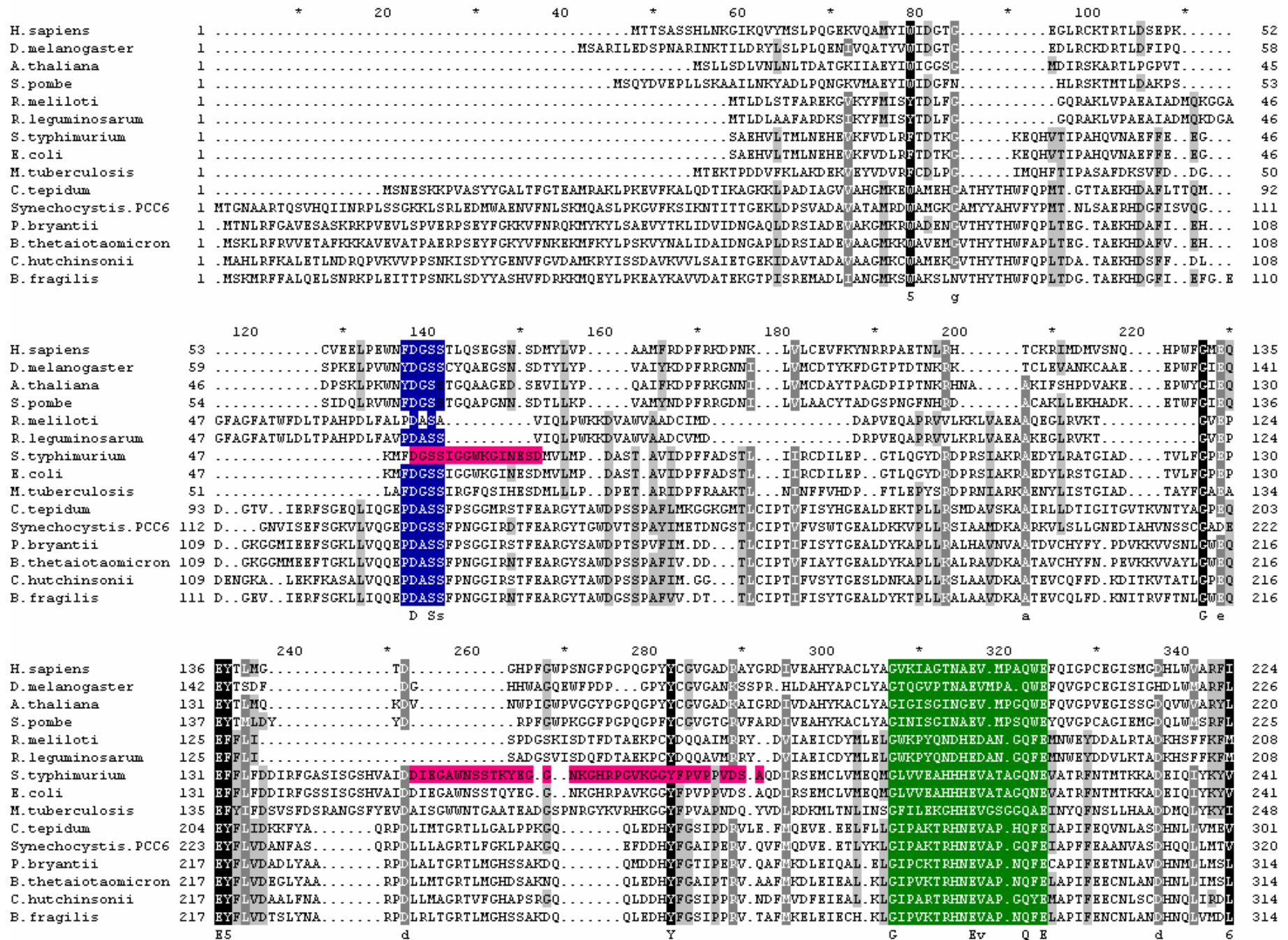


Figure 2.3.

		700	*	720	*	740	*	760	*	780	
<i>H. sapiens</i>	-	-
<i>D. melanogaster</i>	-	-
<i>A. thaliana</i>	-	-
<i>S. pombe</i>	-	-
<i>R. meliloti</i>	-	-
<i>R. leguminosarum</i>	-	-
<i>S. typhimurium</i>	-	-
<i>E. coli</i>	-	-
<i>M. tuberculosis</i>	-	-
<i>C. tepidum</i>	622	ESIGLSDAALQSQAQLLKTLAEDLSKLDLDTAILEETIEEMEEQSELDKADFC SARLLPCMNAIREVADKIEVQVDRSRWQLPTYSEMLFEH									714
<i>Synechocystis. PCC6</i>	638	. . SGLGIDFEKESAKKIADLTNQHWGRVARLSEAMAKHDFAN. . . TEERLQYCAQTLRPLMDEVRTFADALEGEIADSFWPLPTYQEMLFIK									724
<i>P. bryantii</i>	637	FGADEAKRLSERNISIIKDAERTALIEKGVVEELVEARKKANKIENEREKAIAYHD TVETKFDAIRYQIDKLELEVSDDELWTLPKYRELLFIR									729
<i>B. thetaiotaomicron</i>	637	FPADKAARLSAKNLELIEEADRTAFIKEHVDAMI EARKVANKIESEREKAIAYHDTIVPALEEIRYTHIDKLELIVDNQHWTLPKYRELLFVR									729
<i>C. hutchinsonii</i>	633	. KDIGLEKEAKPVIEVIKEISKHINAIQTDVAMDARRKANIEEHMPERAKAYCDKVKPYFETIRYAVDKLELIVDDEMWPLPKYRELLFLR									724
<i>B. fragilis</i>	637	FSEEEYEVMSADRKELIKEISHRVSAIKVLVRDMTEARKVANHKENFKKAFAYEETVVRPYLESIRDHIDHLEMEIDDEIWPLPKYRELLFTK									729

Figure 2.3. A - Gapped alignment of representative sequences from all 4 GS groups highlighting the five regions (marked in colour) conserved between GSs. They are designated as follows: Region (I) Latch region [PYF]-D-[GA]-S-S- ; Region (II) G-X(8)-E-[VD]-X(3,4)-Q-X-[EF]- ; Region (III) ATP-binding site K-P-[LIVMFYA]-X(3,5)-[NPAT]-G-[GSTAN]-G-X-H-X(3)-S ; Region (IV) Glutamate binding site [ND]-R-X(3)-[IV]-R-[IV]- ; Region (V) [ILF]-E-[FDV]-R-X(6)-[NDPS]-. The adenylation motif is shown in red. Shading in greyscale represents the conservation of residues (scored according to Blosum62 score table). Marked in pink on the *S. typhimurium* sequence are the positions of 5 important loops, in order: the Asp50 loop or ‘the latch’ (residues 50-64); Tyr179 loop (153-188); Asn264 loop (255-267); Glu327 loop or ‘the flap’ (324-329); and Tyr397 loop or ‘adenylylation loop’ (388-411).

Table 2.2: Comparison of the identity of the amino acid sequences of GS enzymes from various organisms representing the four GS families.

	GSII				GlnT		GSI			GSIII					
	<i>H. sapiens</i>	<i>D. melanogaster</i>	<i>A. thaliana</i>	<i>S. pombe</i>	<i>R. meliloti</i>	<i>R. leguminosarum</i>	<i>S. typhimurium</i>	<i>E. coli</i>	<i>M. tuberculosis</i>	<i>C. tepidum</i>	<i>Synechocystis</i>	<i>P. bryantii</i>	<i>B. thetaiotaomicron</i>	<i>C. hutchinsonii</i>	<i>B. fragilis</i>
<i>H. sapiens</i>	373	54%	50%	51%	12%	12%	16%	16%	15%	8%	7%	7%	7%	8%	7%
<i>D. melanogaster</i>		365	50%	48%	10%	10%	14%	14%	13%	7%	8%	8%	8%	7%	7%
<i>A. thaliana</i>			354	52%	12%	12%	15%	15%	15%	8%	8%	8%	8%	9%	8%
<i>S. pombe</i>				359	13%	13%	14%	14%	13%	8%	8%	7%	7%	8%	8%
<i>R. meliloti</i>					435	85%	20%	20%	19%	9%	9%	8%	8%	9%	8%
<i>R. leguminosarum</i>						435	20%	20%	19%	9%	8%	8%	8%	9%	8%
<i>S. typhimurium</i>							468	97%	47%	10%	11%	10%	10%	10%	9%
<i>E. coli</i>								468	48%	10%	10%	10%	10%	10%	9%
<i>M. tuberculosis</i>									478	10%	12%	10%	10%	10%	9%
<i>C. tepidum</i>										714	45%	44%	46%	44%	44%
<i>Synechocystis</i>											724	42%	40%	43%	41%
<i>P. bryantii</i>												729	73%	55%	57%
<i>B. thetaiotaomicron</i>													729	57%	60%
<i>C. hutchinsonii</i>														724	56%
<i>B. fragilis</i>															729

Results are given as the percentage identity of the a.a. residues in the alignments shown in Figure 2.3 (as calculated by Genedoc). Origins of the sequence information and references are described in section 2.2.1. The values marked in bold reflect the % identity between all other GS families and *S. typhimurium* GSI, the template structure upon which the homology models were based.

A

<i>H. sapiens</i>	1	MTTSASSHLNKGIKQVYMSLPQGEKRWQAMYYIWDIGTE	EGLRCKTRTLDSEPK	52
<i>D. melanogaster</i>	1	MSARILEDSPNARINKTILDRLYSLPLQENIVQATYVWIDGTG	EDLRCKDRTLDLFIQ	58
<i>A. thaliana</i>	1	MSLSDLVNMLNLTDATCKLIIAEYIWIIGGSG	MDLRKARTLPGPVT	45
<i>S. pombe</i>	1	MSQYDVEPLLKAAALNHYADLPQNGKVMAYIYIWDIGFN	HLRSKTMTLDAKPS	53
<i>R. meliloti</i>	1	MTLDLSTFAREKVKYFMISYTDLFG	GQRAKLVPAEAIADMQKGA	46
<i>R. leguminosarum</i>	1	MTLDLAAAFADKSIKFFMISYTDLFG	GQRAKLVPAEAIADMQKGA	46
<i>S. typhimurium</i>	1	SAEHVLTMLNEHEVKFVDLRFDTK	KEQHVITPAHQVNAEFFE	..EG	46
<i>E. coli</i>	1	SAEHVLTMLNEHEVKFVDLRFDTK	KEQHVITPAHQVNAEFFE	..EG	46
<i>M. tuberculosis</i>	1	MTEKTPDDVFKLAFDEKVEYVDFRCDLPG	IMQHFTIPASAFDKSVD	..DG	50
<i>C. tepidum</i>	1	MSNESKRPVASYGALTFCETAMRAKLPKEVFKALQDTIKACKKLPADIGVVAHCKREKAMEHGATHYTHWFQPM	GTTAEKHD AFLTQM	92
<i>Synechocystis. PCC6</i>	1	MTGNAARTQSVHQIINRPLSSGKLSRLEDMMWAEVFNLSKMQASLPRGVFKSIKNTITTEKRLDPSVADAVATAMRDWAMCKCANYAHVFPMT	NLSAERHDCGISVQG	111	
<i>P. bryantii</i>	1	MTNLRFGAVESASRRKRPVEVLSVPERPSEYFGKRVFNQKMYKLSAEVYTKLIDVIDNGAQLDRSIADVAKMKRWADEMGVTHYTHWFQPLTEG	TAEKHD AFI	..EH	108	
<i>B. thetaiotaomicron</i>	1	MSKLRFRVVEATAPKKKAVEVATPAERPSEYFGKRVFNKKEKPKYLPKRVYNALIDAIDNCAPLDRSIADVAAGMKKWADEMGVTHYTHWFAPLTEG	TAEKHD A FV	..EH	108	
<i>C. hutchinsonii</i>	1	MAHLRFKALETLDNRQPVKVVPPSNKISDYEGENVFQVDAWKRYISSDAVKKVLSAETGKIDAVTADAVAAGMKCWAMEKGVTHYTHWFQPLTDA	TAEKHD SFF	..DL	108	
<i>B. fragilis</i>	1	MSKMRFFALQELSNRKLPEITTPSNKLSDYASHVDFRKKHQEYLPKREAYKAVVDATKCTPISREMLIANCHKMSWAKSLNVTHYTHWFQPLTDG	TAEKHD CFI	..EFG	110	

5 g

<i>H. sapiens</i>	53	CVEELPEWNFDCSS	TLQSEGSN	SDMYLVP	AAMFRDPFRKDPNK	LVLCEVFKYNRRPAETNLEH	TCKRIMDMVSMQ	HPWFQMEQ	135																							
<i>D. melanogaster</i>	59	SPKELPVWNYDCSS	CYQAECSN	SDTYLYP	VAIYKDPFRRCGNI	LVMCDTYKFDGTPDTNKRK	TCLEVANKCAA	EPWFGIEQ	141																							
<i>A. thaliana</i>	46	DPSKLPKWNYYDCSS	TCQAAGED	SEVILYP	QAIFKDPFRRCGNI	LVMCDAYTPAGDPIPTNKRHNA	AKIFSHPDVAKE	EPWYFGIEQ	130																							
<i>S. pombe</i>	54	SIDQLRVWNFDCSS	TCQAQGN	SDTLLKP	VAMYNDPFRRCGNI	LVLAACTADGSPMGFNHRD	ACAKLLEKHADK	ETWFGIEQ	136																							
<i>R. meliloti</i>	47	CFACFATWFDLTPAHPDLFALPDASA	VIQLPWKRDVAWVAADCIMD	DAPVEQAPRVVLRKLVAAEAQEGLRVKT	GVEP	124																												
<i>R. leguminosarum</i>	47	CFACFATWLDLTPAHPDLFAVPDASA	VIQLPWKRDVAWVAADCIMD	DRPVEQAPRVVLRKLVAAEAQEGLRVKT	GVEP	124																												
<i>S. typhimurium</i>	47	KMFDGSS	ICGWKGINESDMVLM	DAST	AVIDPFFADSTL	IIRCDILEP	CTLQCYDRDPRSI	AKRAEDYL	RATGIAD	TVLFGPEP	130																					
<i>E. coli</i>	47	KMFDGSS	ICGWKGINESDMVLM	DAST	AVIDPFFADSTL	IIRCDILEP	CTLQCYDRDPRSI	AKRAEDYL	RSTGIAD	TVLFGPEP	130																					
<i>M. tuberculosis</i>	51	LAFDCSS	IRCFQSIHESDMLLLP	DPET	ARIDPFFAAKTL	NINFFVHDP	FTLEPYSRDP	PNRIAR	KAENYL	LISTGIAD	TAYFGAEA	134																				
<i>C. tepidum</i>	93	D..GTV	..IERFSGEQLIQEPPDASS	FPNGGIRSTFEARCYTAMD	PSSPAFLMKCGKCM	TLCIPTVFISYHG	BALDEKTP	LLRSM	DAVSKAA	IRLLD	TIGIT	GVTK	WNTY	AGPEQ	203																						
<i>Synechocystis. PCC6</i>	112	D....GN	VISSEFSKRVLVQCEPPDASS	FPNGGIRDTFEARCYT	GDVTS	PAYIM	TDNGS	TLCIPTVFIS	WTS	GEALD	DKVPL	LLRS	IAAMD	KAARKV	LSL	LGND	IAHWN	SSCGADE	222																		
<i>P. bryantii</i>	109	D..GR	CGMIIEFSGKLLVQCEPPDASS	FPNGGIRSTFEARCYSA	MDPTSP	VFIM	DD...	TLCIPTIF	VISY	TEALD	YKAP	LLRAL	HAVN	VAA	TDV	CHYFF	PDVK	KVVSN	LGWEQ	216																	
<i>B. thetaiotaomicron</i>	109	D..GR	CGMIEEFTKLLVQCEPPDASS	FPNGGIRNTFEARCYSA	MDPSS	PAFIV	DD...	TLCIPTIF	VISY	TEALD	YKAP	LLRAL	HAVN	VAA	TDV	CHYFF	PDVK	KVVSN	LGWEQ	216																	
<i>C. hutchinsonii</i>	109	DEN	KA..L	EKFAS	ALVQ	CEPPDASS	FPNGGIRSTFEARCYT	AMD	PSS	PAFIM	GC...	TLCIPTIF	VISY	TE	SLDN	KAP	LLK	SLAA	VDR	KA	TEVC	QFFD	..KD	ITR	V	T	A	T	L	G	P	E	Q	216			
<i>B. fragilis</i>	111	D..GE	V..IER	FSGKLLIQEPPDASS	FPNGGIRNTFEARCYT	AMD	GSS	PAFV	DT...	TLCIPTIF	VISY	TE	ALD	YK	T	PL	K	ALAA	VDR	KA	TEVC	Q	L	F	D	..KN	IT	R	V	T	N	L	G	W	E	Q	216

D Ss

a

C e

<i>H. sapiens</i>	136	EYTL	LMC	TD	GHPFCWPSNGFP	GPQGPY	CCVGAD	RAYC	RD	IVE	AH	YRAC	LYA	G	V	K	I	A	G	T	N	A	E	V	..MP	A	QW	E	F	Q	V	C	P	C	E	C	I	S	M	G	D	H	L	W	V	A	R	F	I	224																																																												
<i>D. melanogaster</i>	142	EYTS	D	F	DC	HHWACQ	EWFPDP	CPYY	CCVG	GANK	SSPR	..H	L	D	A	H	Y	A	P	C	L	Y	A	G	T	Q	G	V	P	T	N	A	E	V	..M	P	A	QW	E	F	Q	V	C	P	C	E	C	I	S	I	C	H	D	L	W	M	A	R	F	L	226																																																
<i>A. thaliana</i>	131	EYTL	M	D	KDV	NWPI	GW	PV	GGY	PG	Q	P	Y	CC	V	G	A	D	K	A	I	C	R	D	I	V	A	H	Y	K	A	C	L	Y	A	G	I	S	I	G	I	N	G	E	V	..M	P	G	QW	E	F	Q	V	C	P	V	E	C	I	S	S	C	Q	V	W	V	A	R	L	220																																								
<i>S. pombe</i>	137	EYTM	L	D	YD	RPFC	W	P	K	G	F	P	G	Q	P	F	Y	CC	V	G	T	C	R	V	F	A	D	I	V	E	A	H	Y	K	A	C	L	Y	A	G	I	S	I	G	I	N	A	E	V	..M	P	S	QW	E	F	Q	V	C	P	C	A	G	I	E	M	G	D	L	W	M	S	R	F	L	225																																			
<i>R. meliloti</i>	125	EFFL	I	SPD	G	S	K	I	S	D	T	F	D	T	A	E	K	P	C	D	Q	Q	A	I	M	R	Y	..D	V	I	A	E	I	C	D	Y	M	L	E	L	G	W	K	P	Y	Q	M	D	H	E	D	A	N	..G	F	E	M	N	W	E	Y	D	D	V	L	K	T	A	D	K	H	S	F	F	K	F	M	208																																
<i>R. leguminosarum</i>	125	EFFL	I	SAD	G	S	V	I	S	D	Q	F	D	T	A	E	K	P	C	D	Q	Q	A	V	M	R	Y	..D	V	I	A	E	I	C	D	Y	M	L	E	L	G	W	K	P	Y	Q	M	D	H	E	D	A	N	..G	F	E	M	N	W	E	Y	D	D	V	L	K	T	A	D	K	H	S	F	F	K	F	M	208																																
<i>S. typhimurium</i>	131	EFFL	F	D	G	N	K	G	H	R	P	A	V	K	G	G	Y	F	V	P	P	V	D	S	..A	Q	D	I	R	S	E	M	C	L	V	M	E	Q	M	G	L	V	V	E	A	H	H	H	E	V	A	T	A	G	Q	N	E	V	A	T	R	F	N	T	M	T	K	K	A	E	I	E	I	Q	I	Y	K	Y	V	241																													
<i>E. coli</i>	131	EFFL	F	D	G	N	K	G	H	R	P	A	V	K	G	G	Y	F	V	P	P	V	D	S	..A	Q	D	I	R	S	E	M	C	L	V	M	E	Q	M	G	L	V	V	E	A	H	H	H	E	V	A	T	A	G	Q	N	E	V	A	T	R	F	N	T	M	T	K	K	A	E	I	E	I	Q	I	Y	K	Y	V	241																													
<i>M. tuberculosis</i>	135	EFYI	F	D	S	V	S	F	D	S	R	A	N	G	S	F	Y	E	V	D	A	I	S	G	W	N	T	G	A	A	T	E	A	D	S	P	N	R	G	Y	K	V	R	H	K	G	Y	F	P	V	A	P	N	D	Q	Y	V	D	L	R	D	K	M	L	T	N	L	I	N	S	G	F	I	L	E	K	H	H	E	V	C	S	G	Q	A	E	I	N	Y	Q	F	N	S	L	L	H	A	A	D	D	M	Q	L	Y	K	Y	248
<i>C. tepidum</i>	204	EYFL	I	D	Q	R	P	D	L	I	M	T	G	R	T	L	L	G	A	L	P	P	K	Q	Q	L	E	D	H	Y	F	G	S	I	P	D	R	V	L	E	..F	M	Q	E	V	E	..E	E	L	F	L	L	C	I	P	A	K	T	R	H	N	E	V	A	P	..H	Q	F	E	L	A	P	I	F	E	C	N	L	A	S	D	H	N	L	L	V	M	E	V	301																				
<i>Synechocystis. PCC6</i>	223	EYFL	V	D	Q	R	P	D	L	L	A	G	R	T	L	F	G	K	L	P	A	K	Q	E	F	D	D	H	Y	F	G	A	I	P	E	R	V	..Q	V	F	M	Q	D	V	E	..E	T	Y	K	L	C	I	P	A	K	T	R	H	N	E	V	A	P	..H	Q	F	E	L	A	P	I	F	E	C	N	L	A	S	D	H	N	L	L	V	M	E	V	320																						
<i>P. bryantii</i>	217	EYFL	V	D	R	P	D	L	A	L	T	G	R	T	L	M	G	H	S	A	R	D	Q	Q	M	D	D	H	Y	F	G	T	I	P	E	R	V	..Q	A	F	M	K	D	L	E	I	A	L	..E	L	G	I	P	C	K	T	R	H	N	E	V	A	P	..H	Q	F	E	L	A	P	I	F	E	C	N	L	A	S	D	H	N	L	L	V	M	E	V	314																						
<i>B. thetaiotaomicron</i>	217	EYFL	V	D	R	P	D	L	L	T	G	R	T	L	M	G	H	S	A	K	N	Q	Q	L	E	D	H	Y	F	G	A	I	P	E	R	V	..A	A	F	M	K	D	L	E	I	A	L	..E	L	G	I	P	V	K	T	R	H	N	E	V	A	P	..H	Q	F	E	L	A	P	I	F	E	C	N	L	A	S	D	H	N	L	L	V	M	E	V	314																							
<i>C. hutchinsonii</i>	217	EYFL	V	D	R	P	D	L	L	A	G	R	T	V	F	G	H	A	S	R	G	Q	Q	L	D	D	H	Y	F	G	S	I	P	E	R	V	..N	D	F	M	V	D	F	E	I	A	L	..E	L	G	I	P	A	R	T	R	H	N	E	V	A	P	..H	Q	F	E	M	A	P	T	F	E	C	N	L	S	D	H	N	L	L	V	M	E	V	314																								
<i>B. fragilis</i>	217	EYFL	V	D	R	P	D	L	R	L	T	G	R	T	L	M	G	H	S	A	R	D	Q	Q	L	E	D	H	Y	F	G	S	I	P	E	R	V	..T	A	F	M	K	E	L	E	I	E	C	H	..K	L	G	I	P	V	K	T	R	H	N	E	V	A	P	..H	Q	F	E	L	A	P	I	F	E	C	N	L	A	S	D	H	N	L	L	V	M	E	V	314																					

E5

d

Y

C

E_v

Q

E

d

6

Figure 2.4. conserved in GS I's conserved in GS II's conserved in GS III's conserved active site conserved in all species conserved in GlnT's

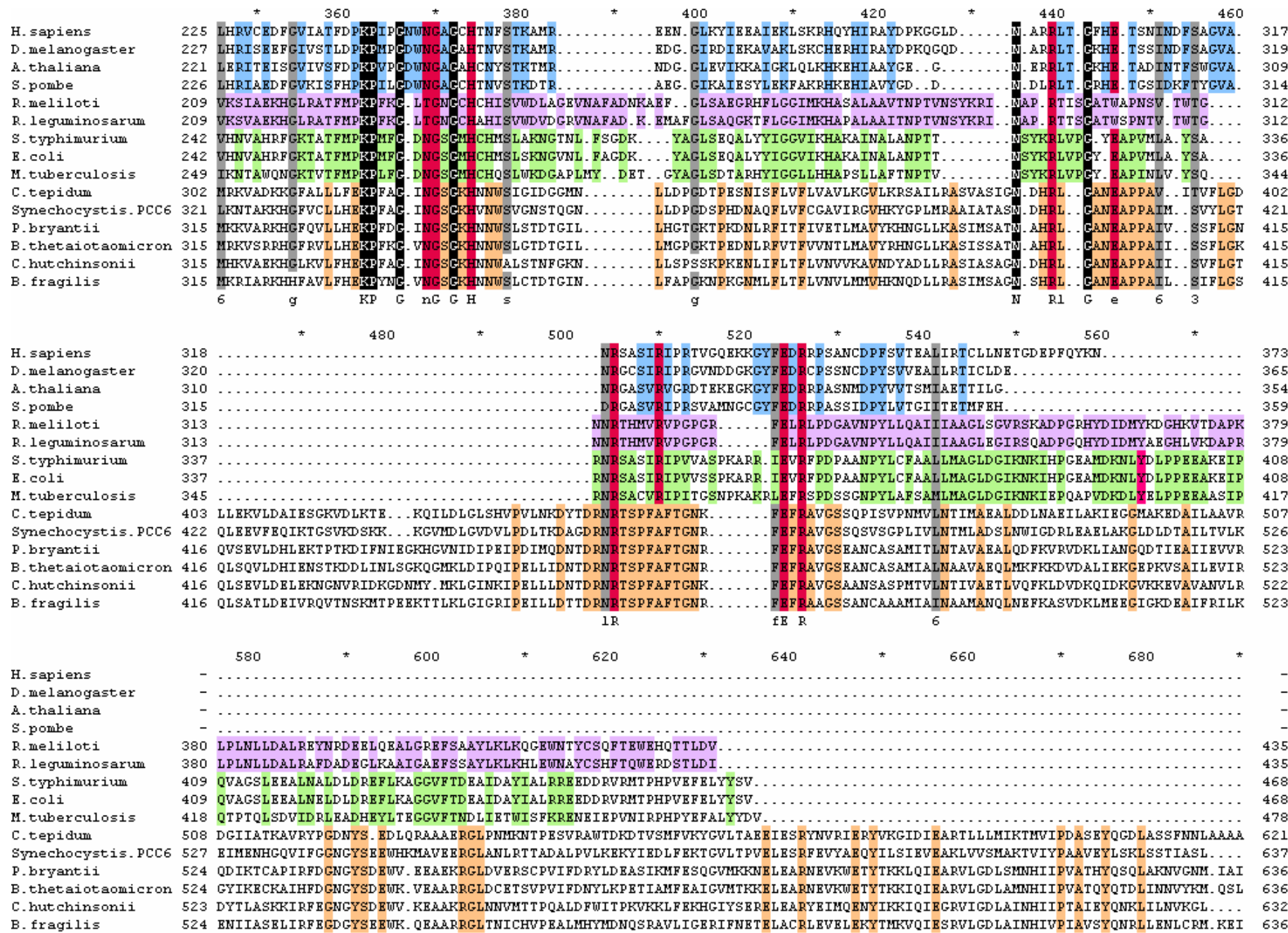


Figure 2.4.

		700	*	720	*	740	*	760	*	780	
<i>H. sapiens</i>	-	-
<i>D. melanogaster</i>	-	-
<i>A. thaliana</i>	-	-
<i>S. pombe</i>	-	-
<i>R. meliloti</i>	-	-
<i>R. leguminosarum</i>	-	-
<i>S. typhimurium</i>	-	-
<i>E. coli</i>	-	-
<i>M. tuberculosis</i>	-	-
<i>C. tepidum</i>	622	ESIGLSDAALQSQAGLLKTLAEDLSKLIIDLTAILEETIEMEEQSELDKADFCSARLLPCMNAIREVADKIEVQVDRSRWQLPTYSEMLFEH									714
<i>Synechocystis. PCC6</i>	638	..SGLGIDFEKESAKKIADLTNQMVGRVAKLSEAMAKHDFAN...TEEKLOYCAQTLRPLMDEVRTFADALEGRIADSFWPLPTYQEMLFIK									724
<i>P. bryantii</i>	637	FGADEAKKLSEERNISIIKDIAERTALIEKGVVEELVEAPKANKIENEREKAIAYHDTVETKFDAIRYQIDKLELEVSDLELWTLPKYRELLFIR									729
<i>B. thetaiotaomicron</i>	637	FPADKAARLSAKNLELIBEIIADRTAFIKEHVDAMIEAPKVKANKIESEREKAIAYHDTIVPALEEIIRYHIDKLELIVDNQMWTLPKYRELLFVR									729
<i>C. hutchinsonii</i>	633	.KDIGLEKBAKPVIEVIKEISKHINAIQTDVDMTDARKKANIIEHMPERAKAYCDKVKPYFETIRYAVDKLELIVDDEMWPPLPKYRELLFLR									724
<i>B. fragilis</i>	637	FSEEEYEVMSADRKELIKEISHRVSAIKVLVPRDMTEARKVANHKENFKERAFAYEETVVRPYLESIRDHIDHLEMEIDDEIWFPLPKYRELLFTK									729

B

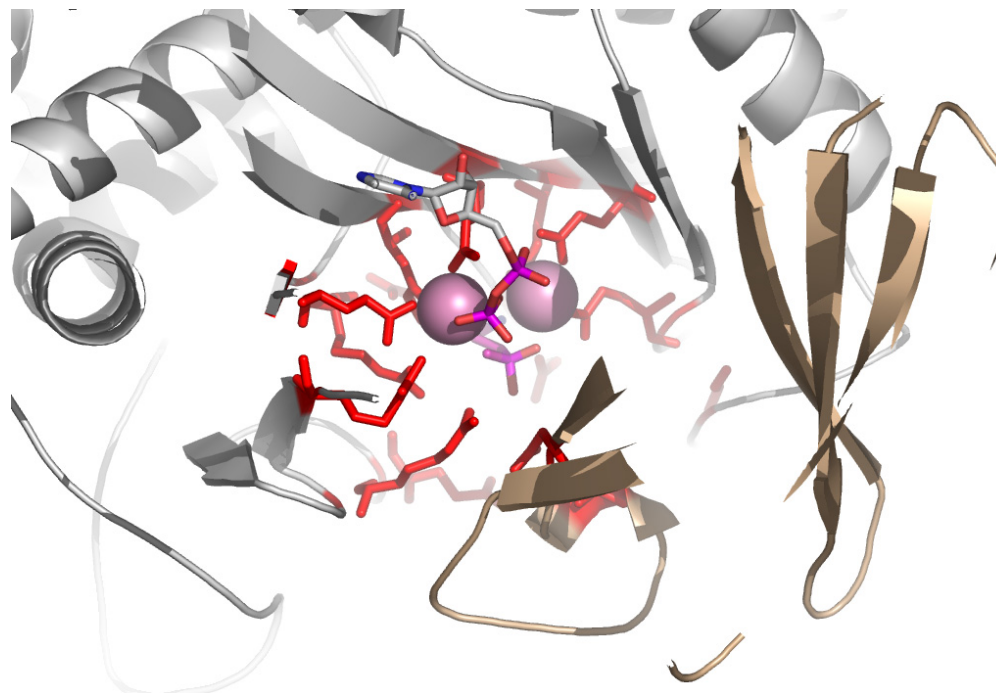


Figure 2.4. A - Multiple sequence alignment of representative sequences from all 4 GS groups. Conserved active site residues, deduced from alignments of GSIs and GSIIIs and atomic structures of GSIs, are shown in red. Despite their poor global sequence homology, the GS sequences possess many regions of high similarity and all appear to possess the critical active site residues implying a similar catalytic mechanism. These inferences and the key residues for involving GSIs and GSIIIs were derived from the alignments by Eisenberg *et al.* (Eisenberg, Gill, et al. 2000) but here all 4 groups are aligned not just GSIs and GSIIIs and the alignments differ substantially as a result. **B** – Structure of GSI from *S. typhimurium* (Yamashita, Almasy, et al. 1989) showing the conserved active site residues, highlighted in A, in red.

However, as described in section 2.1.1, all GSs share five regions of high local similarity. Table 2.3 lists these regions together with their percentage sequence identity to highlight the level of similarity in these regions.

Table 2.3: Regions of high sequence similarity

	Proposed function	Consensus	% identity
I	β sheet of active site $\alpha\beta$ barrel - Latch region	[PYF]-D-[GA]-S-S-	80%
II	β sheet of active site $\alpha\beta$ barrel	G-X(8)-E-[VD]-X(3,4)-Q-X-[EF]-	59%
III	β sheet of active site $\alpha\beta$ barrel - ATP-binding site	K-P-[LIVMFYA]-X(3,5)-[NPAT]-G-[GSTAN]-G-X-H-X(3)-S	67%
IV	β sheet of active site $\alpha\beta$ barrel - Glutamate binding site	[ND]-R-X(3)-[IV]-R-[IV]-	58%
V	β sheet of active site $\alpha\beta$ barrel	[ILF]-E-[FDV]-R-X(6)-[NDPS]-	61%

See Figure 2.1A and 2.3A for the positioning of these regions in the sequence alignments.

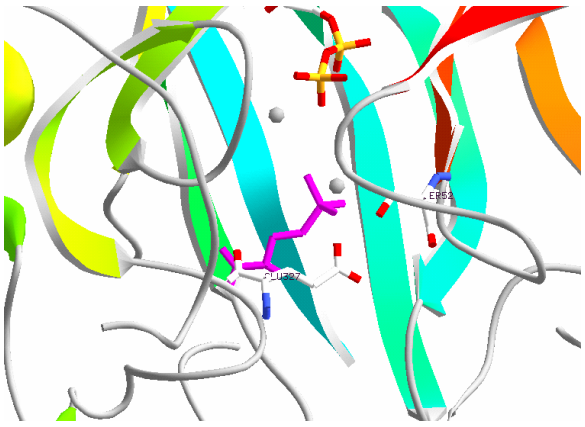
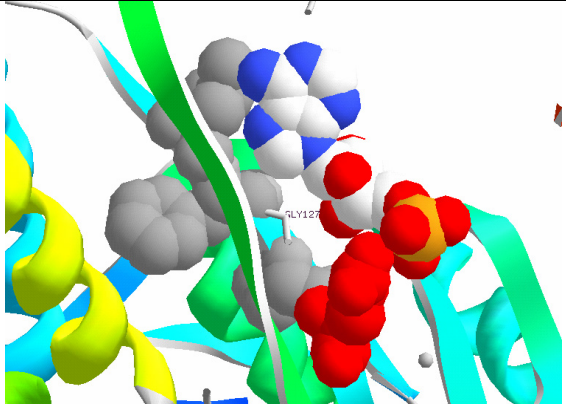
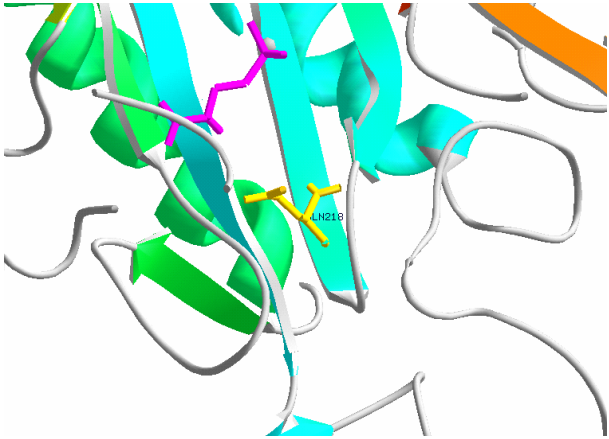
Furthermore, it can be seen from Figures 2.3 A and B that, in addition to the five regions discussed above, further regions of high sequence identity were also discovered. In particular, the two regions corresponding to *S. typhimurium* residues 127-134 and 317-335 were particularly well conserved, with sequence identities of 74% and 54% respectively.

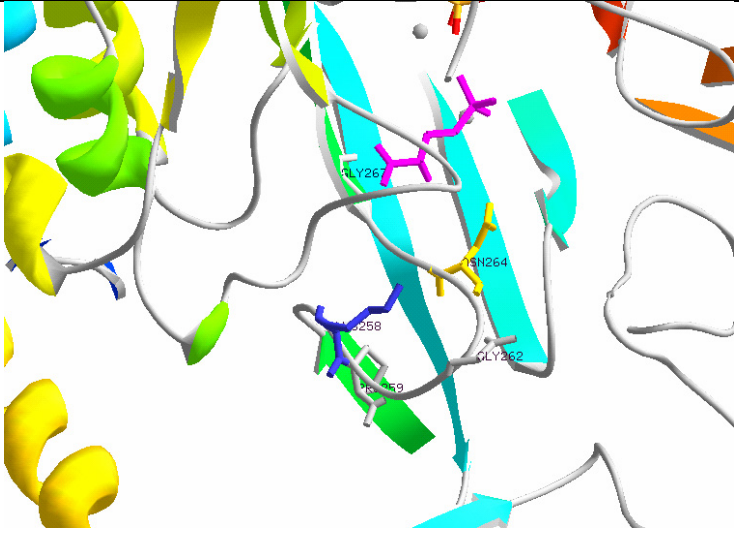
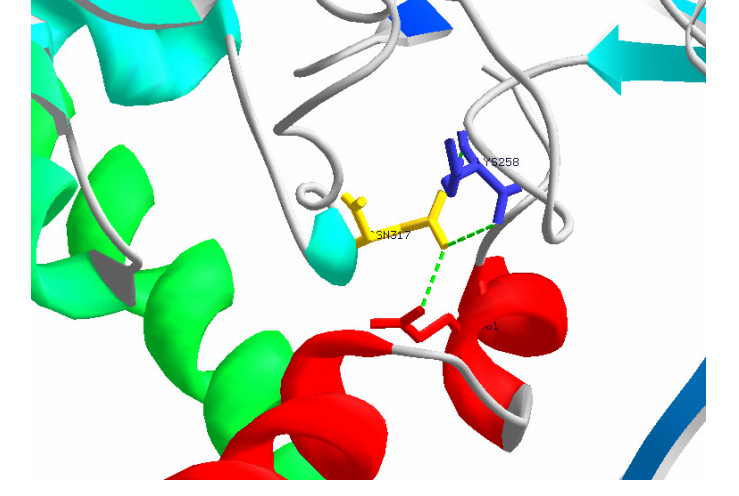
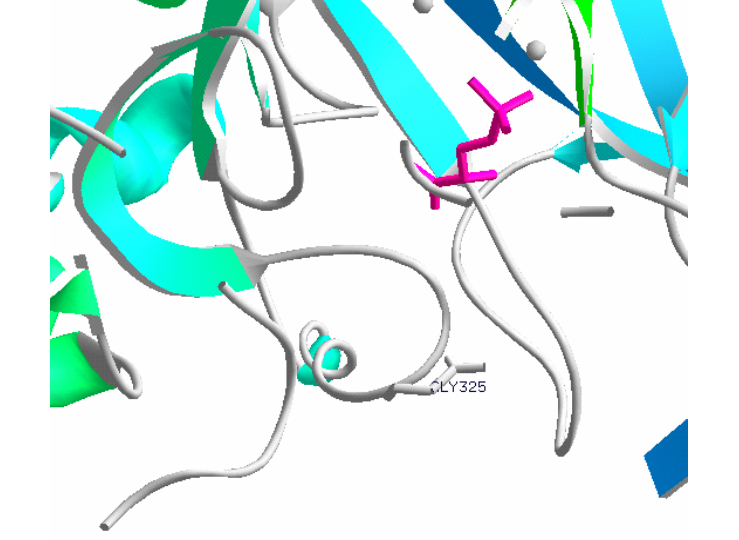
During the initial alignment it was possible to conserve, between all the GS sequences investigated, all but one of the active site residues reported by Eisenberg *et al* (Eisenberg, Gill, et al. 2000) (Figure 2.4). Arg344 in region IV, the position of which was conserved in GS families I, II, and GlnT was replaced by a phenylalanine in the GIIIs. An alternate alignment (discussed below) based only on the GSIs and GSIIIs was however, able to conserve both IV and V.

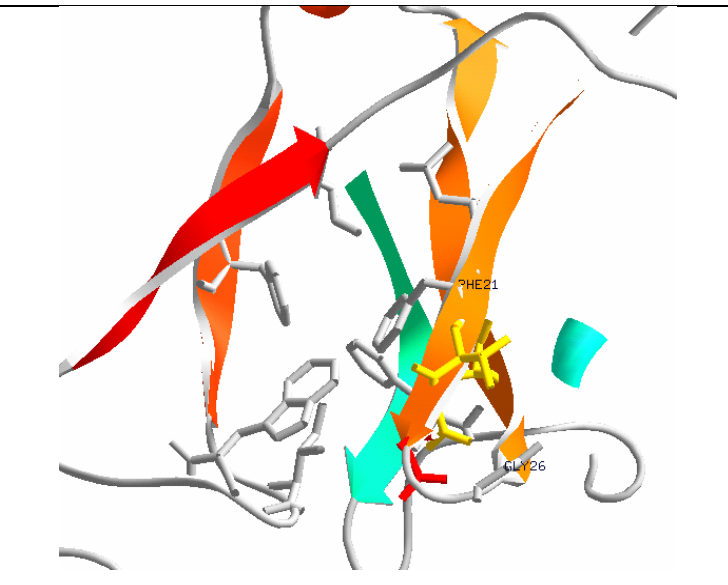
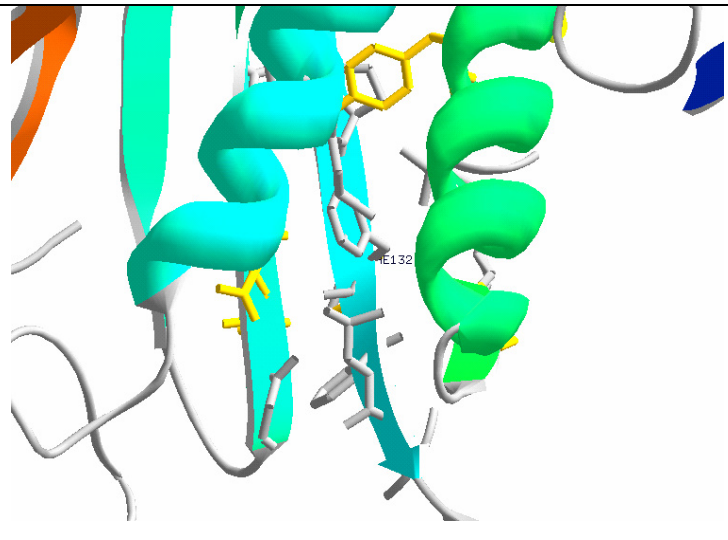
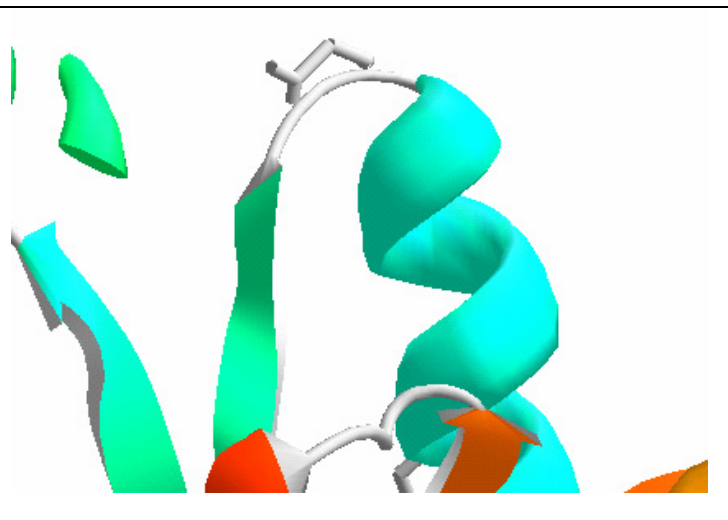
In addition to these previously reported conserved residues, the alignments revealed many other residues conserved between the GS families. These are marked in grey in Figure 2.4A. Possible reasons for their conservation are discussed in light of their local environments in Table 2.4.

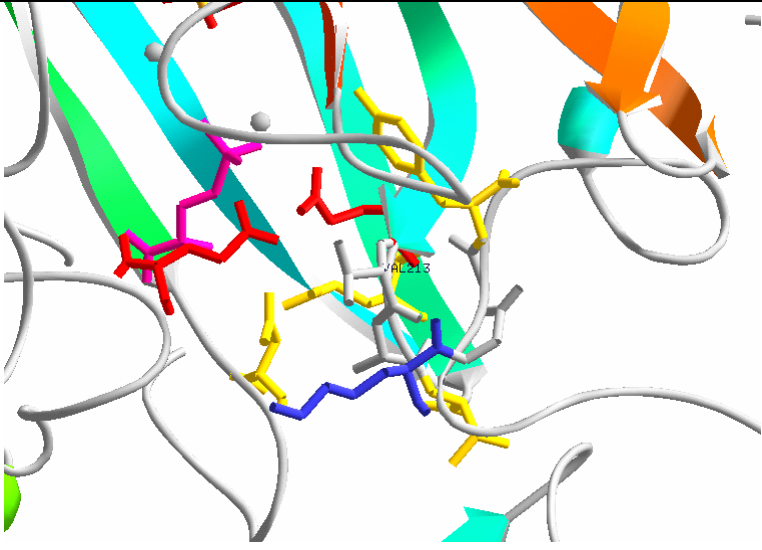
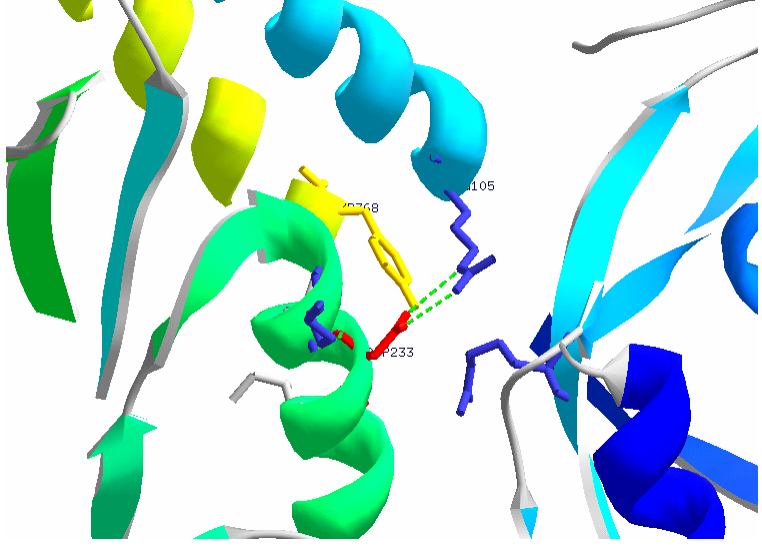
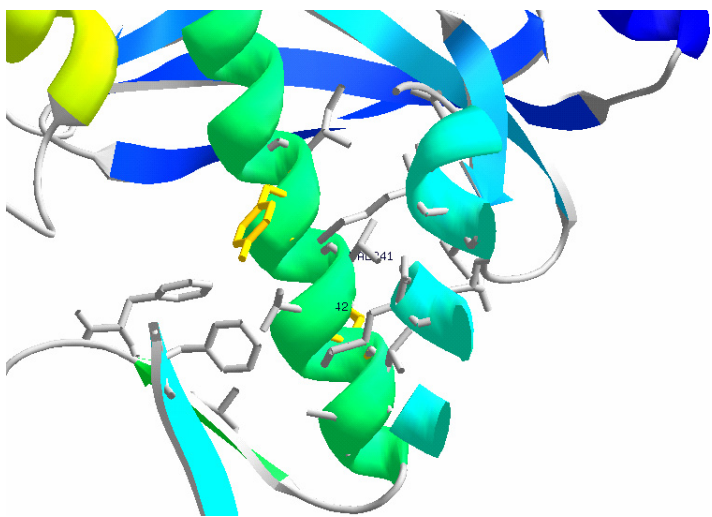
Table 2.4: Local environment of conserved residues not directly involved in the active site.

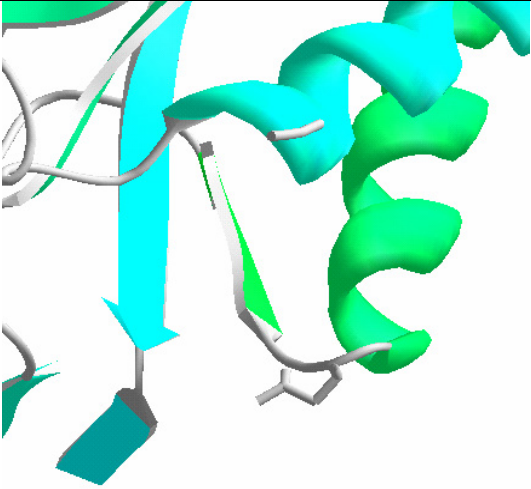
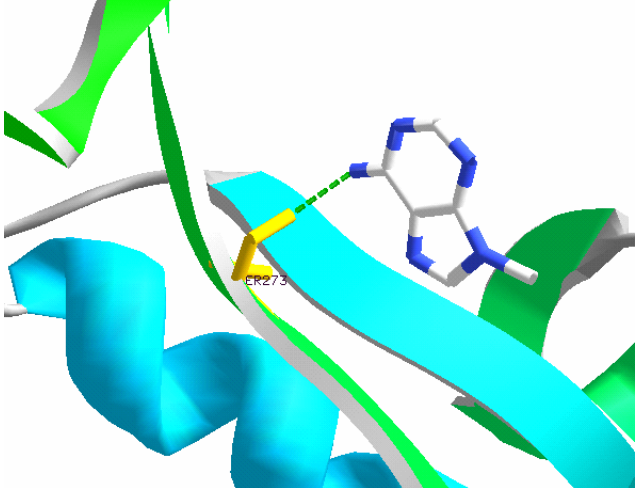
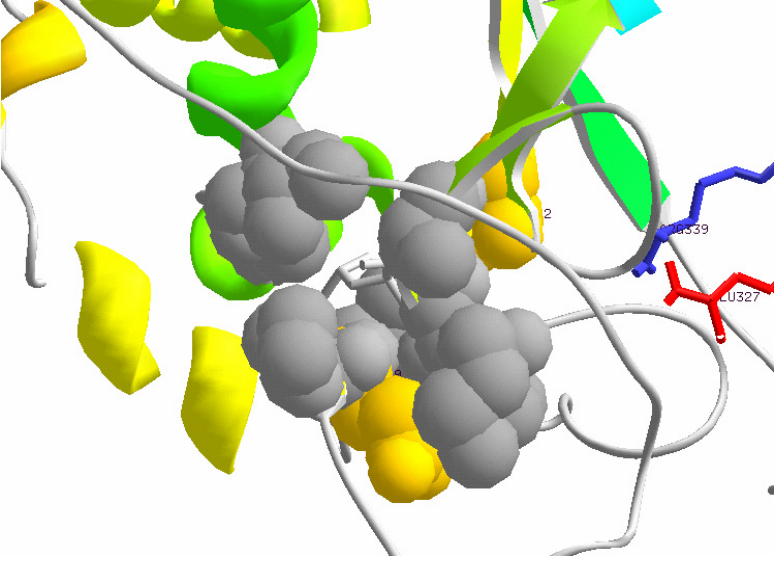
For details of the conservation of active site residues see (Eisenberg, Gill, et al. 2000). The structure shown is of GSI from *S. typhimurium* (Gill & Eisenberg 2001).

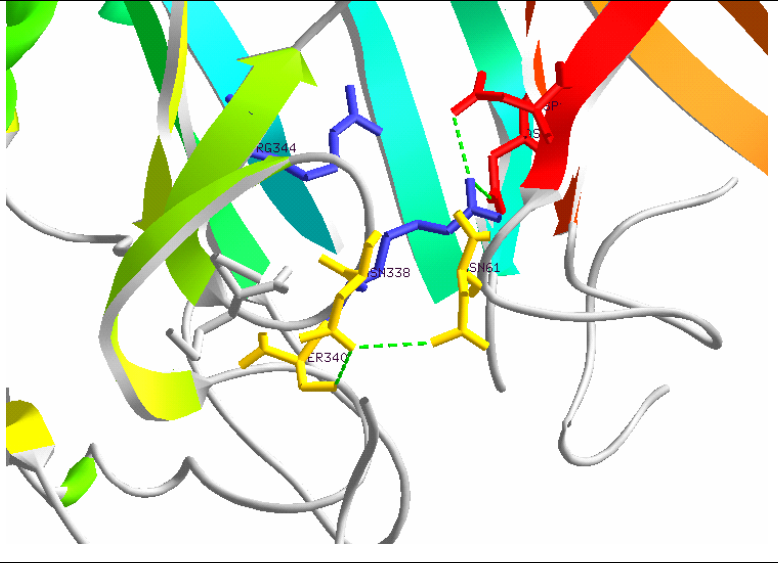
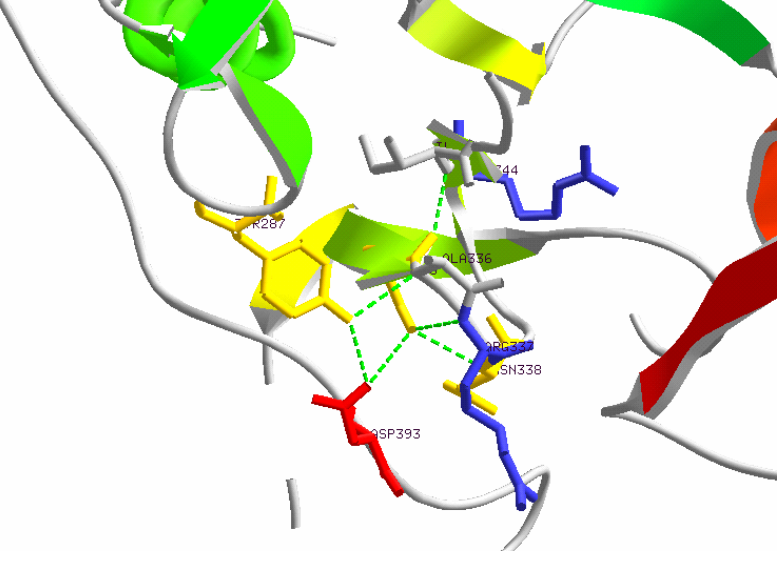
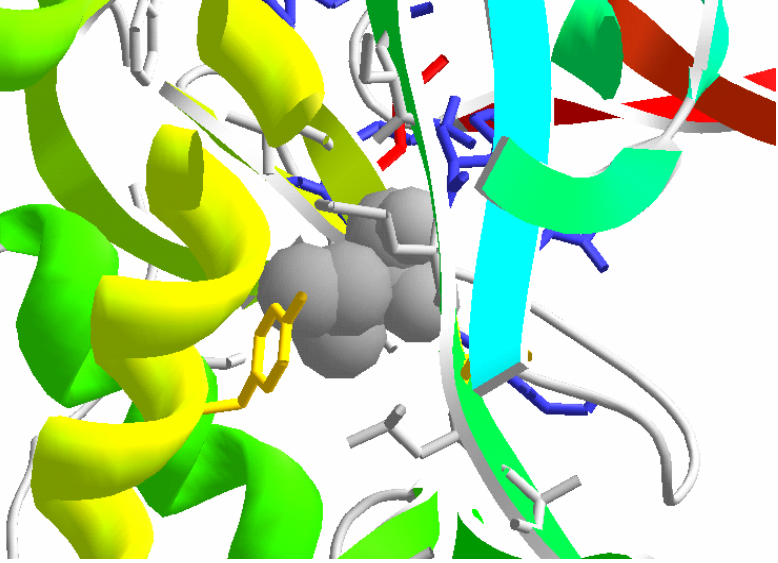
	Sequence #	Consensus	Possible Roles	Local structural environment
Completely conserved	52	S	Forms one of the catches on the N terminal latch (H-bonds to Glu327' on the flap, closing the active site) (Gill, Pfluegl, et al. 2002)	
	127	G	ATP binding site. Avoids steric clash.	
	218	Q	May contribute to the second ammonium binding site (Gill & Eisenberg 2001)	

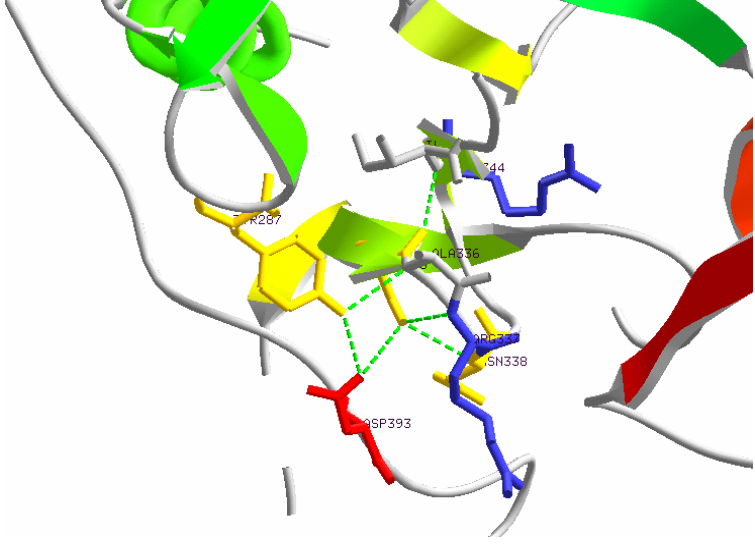
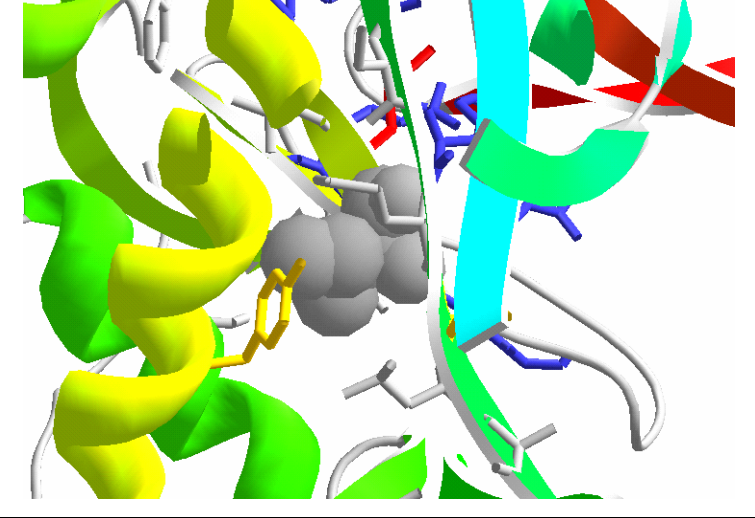
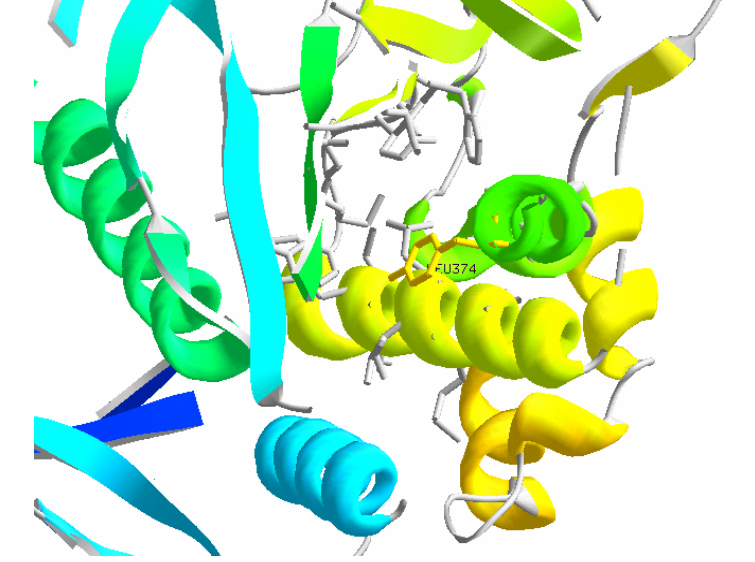
	Sequence #	Consensus	Possible Roles	Local structural environment
Completely conserved	258-259	K-P	Form part of the Asn264 loop. Gly262 is at the centre of this reverse turn	
	262	G		
	267	G		
	317	N	Just before the Glu327 flap. H-bonds to Asn264 loop main chain Lys258 and Glu461 of the subunit from the other ring.	
325	G	Centre of the Glu327 loop turn. Appears to be type II reverse turn.		

	Sequence #	Consensus	Possible Roles	Local structural environment
Highly conserved	21	Yfw (hydrophobic)	Centre of hydrophobic core of N domain	
	26	g	Gly is the i+3 residue of this type I reverse B-turn between B-sheets 1 & 2. Positive phi (HOMSTRAD)	
	132	Yfw (hydrophobic)	Centre of hydrophobic core between B-sheets 1,3,4, and A-helices 3,4 of the C domain i.e. at the back of the B-barrel.	
	203	g	Centre of the reverse turn between A2 and B2.	

	Sequence #	Consensus	Possible Roles	Local structural environment
Highly conserved	213	v	Close to the ammonium binding pocket	
	233	d	H-bonds to Arg105 at the start of A1 (C domain) and Tyr368 of A8 but is also close to Arg88 of the N domain.	
	241	lvim	Participate in same hydrophobic core as discussed for 132 above. Val241 participates in stacking of A3 and A4.	
242	lvim			

	Sequence #	Consensus	Possible Roles	Local structural environment
Highly conserved	250	g	Gly is the i+3 residue of this type I B-reverse turn	
	273	s	H-bonds to the N6 of the adenosine moiety of ADP	
	332	ivl	Positioned in the middle of the loop between Arg344 and Glu327 on the back of the B-sheet. It faces away from the active site towards a hydrophobic cavity	

	Sequence #	Consensus	Possible Roles	Local structural environment
Highly conserved	338	n	Next to Arg339 and Arg344 and similarly interacts with latch residues. H-bonds to Asn61?	
	335	st	H-bonds to Asp393 on the adenylation loop and to the two main chain N's of the following 2 residues	
	356	ilf	Participates in the hydrophobic stacking of A5 against B-sheets 4,5, and 6.	

	Sequence #	Consensus	Possible Roles	Local structural environment
Highly conserved	335	st	H-bonds to Asp393 on the adenylation loop and to the two main chain N's of the following 2 residues	
	356	ilf	Participates in the hydrophobic stacking of A5 against B-sheets 4,5, and 6.	
	374	ilm	Participates in the hydrophobic stacking of A8 against B-4,5, 6 and against A5.	

2.3.2.2 Secondary structure conservation

The active site of GSI is formed by the staves (β sheets) of an α/β barrel (Figure 2.1B), the outside of which is surrounded by α helices which pack against the β sheets hydrophobically. The interior surface of the barrel, however, contains numerous polar residues involved in binding effector molecules and ligands to mediate enzymolysis. In order to achieve a polar interior and hydrophobic exterior, the sequence of amino acid residues comprising the β sheets needs to alternate with the periodicity of the β sheets, which is equals two. This means that every 2nd amino acid in a β sheet faces the opposite direction. Similarly for α helices to achieve a hydrophobic stripe running the length of the helix, every 3-4 residue in a helix must be hydrophobic. Such patterns are evident in the β sheets and α helices of GSI. The conservation of these patterns, as determined by inspecting the patterns of amino acid physiochemical properties, adds confidence to the alignments. In particular, three regions (numbers corresponding to the *S. typhimurium* sequence), which otherwise did not display high sequence similarity, were found to posses the correct patterns of hydrophobic residues for the secondary structures they were aligned against: 180-218 (α helix), 220-241 (β sheet), and 288-316 (α helix). Figure 2.5 demonstrates this relationship for α helix 3 of the C domain.

The structurally annotated (JOY (Mizuguchi, Deane, et al. 1998)) alignment of *S. typhimurium* and *M. tuberculosis* GSIs in the HOMSTRAD database (de Bakker, Bateman, et al. 2001) confirmed the conservation of these patterns. From such a representation it was possible to confirm that these patterns of hydrophobic residues represent conserved local structural features, thus indicating the conservation of the secondary structures they comprise. Furthermore, all three of the regions described above were predicted by the PSIPRED server (McGuffin, Bryson, et al. 2000).

2.3.2.3 Alternate alignments

The task of aligning the terminal regions, with their poor sequence similarity, was made more difficult by the fact that the GSIII sequences contained repeats or alternate choices for a number of the conserved motifs used to guide the alignments (Figure

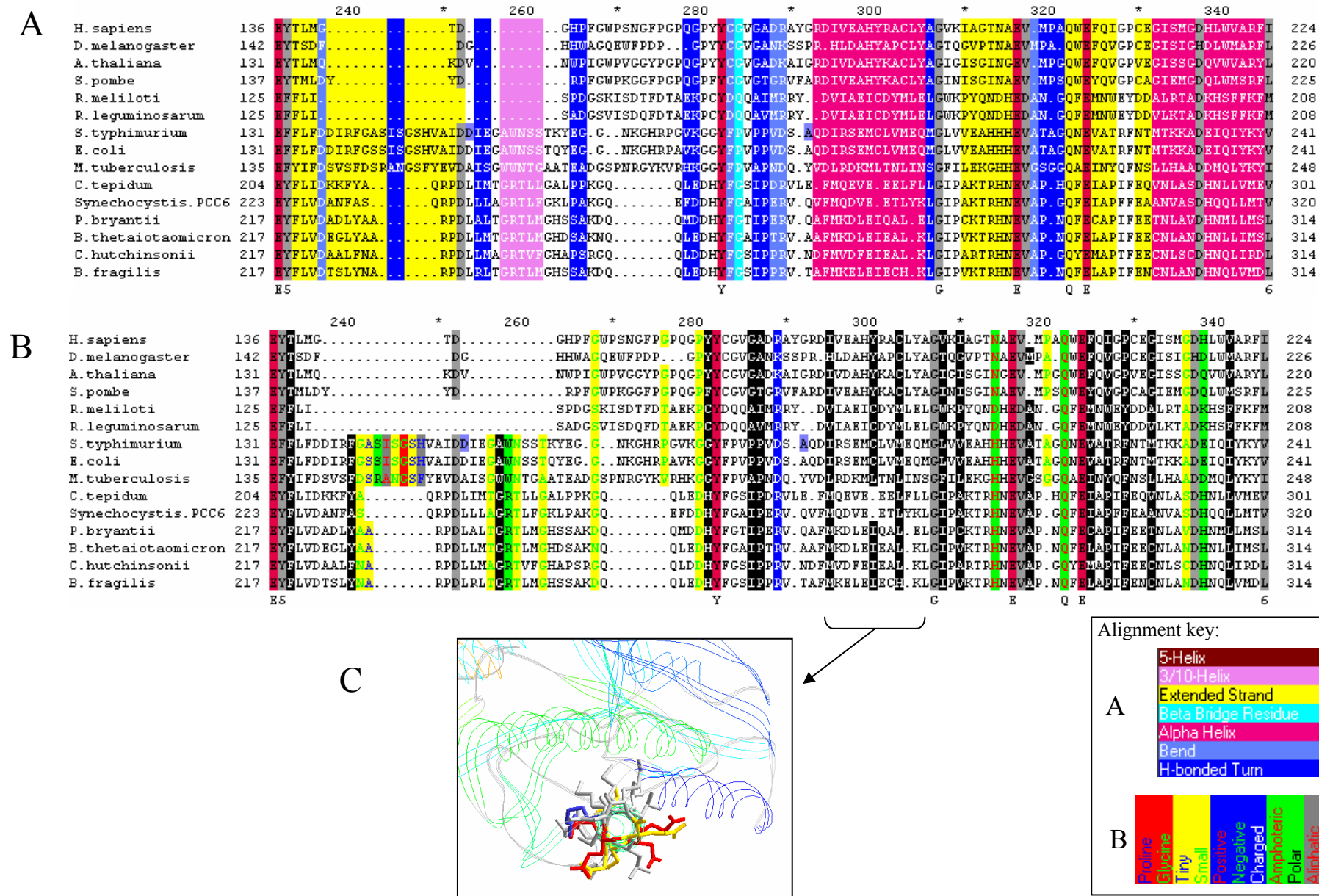


Figure 2.5. **A** – Multiple sequence alignment (same as Figure 2.3A) coloured according to structural features (see key). **B** – same as A but coloured according to conservation of physiochemical properties. **C** – View down helix 3 to emphasise the amphipathic nature of the helix packing (hydrophobic residues are shown in grey, polar residues in yellow, positive in blue and negative in red). The hydrophobic residues marked in black in B are repeated with the correct period to appear on the same side of the helix allowing the helix to stack against the hydrophobic surfaces of the β sheets formed from similar patterns of hydrophobic residues. The other side of the helix is covered in polar residues enabling the helix to interact with the solvent.

2.3). The first repeat is in region I. Eighteen residues downstream of the [PYF]-D-[GA]-S-S motif (128-132) is the sequence W-D-X-[ST]-[ST] (151-154). In *B. fragilis* this repeat sequence is D-G-S-S, which is identical to the region I consensus for GSIs and GSIIIs. The first instance of the motif was chosen to be aligned with the GSI and II region I for the following reasons. Firstly, on closer inspection, GSIs were also shown to possess an alternate D-X-[SA]-T sequence. Secondly, the hydrophobic residues Leu19, Phe21, Trp57, Phe49, Ile54, and Leu67 found in β sheets 4 and 6 all participate in the hydrophobic core of the N domain and thus should be conserved. Only Trp57 cannot be accounted for by matching the first motif to the GSI consensus motif. Leu19, Leu67, and Trp57 cannot, however, be conserved by matching the second occurrence of the motif. Finally, Cys89 and Ile91 are present in the correct β sheet position in the original but not the alternate alignment.

The second area of confusion involved regions IV and V, where repeats of the Asn338-Arg339 and Glu357-X-Arg359 active site residues were found. This alternate alignment of the C-terminus is shown in Figure 2.6. The second E-X-R (position 603-605) sequence is 131 amino acid residues downstream of the first (position 472-474) and the second N-R (position 469-470) sequence is 10 amino acid residues downstream of the first (position 459-460).

The alternate alignment arising from this repetition could not be disregarded based upon inspection of the sequences alone and was considered for the following reasons: (1) The alternate alignment conserves the hydrophobic nature of C terminus, which is involved in inter-ring bonding in GSI (Almassy, Janson, et al. 1986). (2) All active site residues can be conserved. Specifically, Arg344 is now conserved at a similar distance from the Asn338-Arg339 as in *S. typhimurium*. (3) A 2% higher score between the GSIII and GSIs is achieved. This is up from 9% with the first alignment. (4) The patterns of hydrophobic amino acids in C terminal domain helix 10 (starting at 414) are conserved. This is discussed further below.

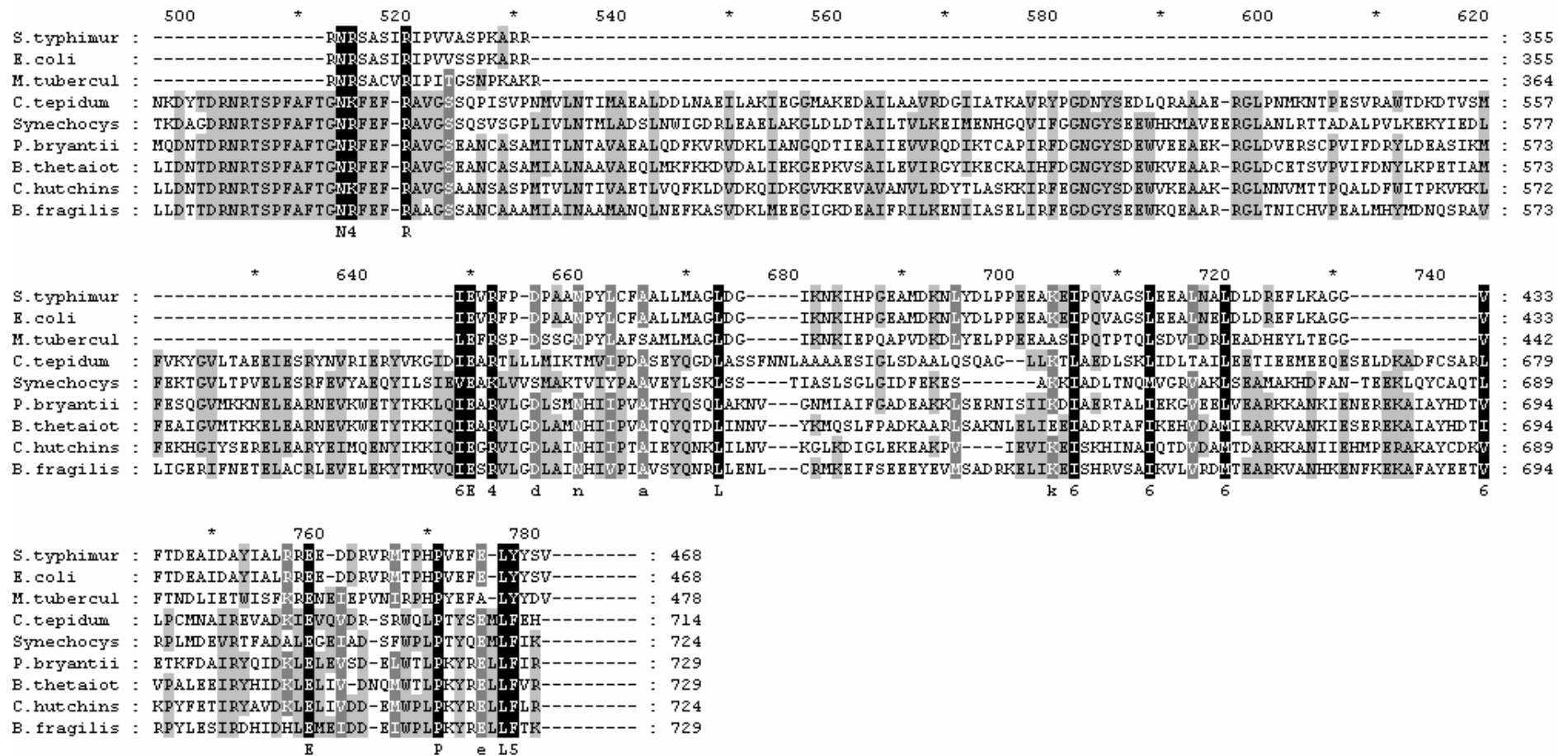


Figure 2.6. Alternate alignment of the C-terminal end of GSIIIs. Shading in greyscale represents the conservation of residues (scored according to Blosum62 score table).

2.3.3 Models

Four models were produced based upon the alternate alignments of *B. fragilis* GSIII and the structure of *S. typhimurium* GSI (Yamashita, Almassy, et al. 1989) with two from the manual alignments and two from the automated alignment programs mGenTHREADER (Jones 1999a; Schwede, Kopp, et al. 2003) and FUGUE (Shi, Blundell, et al. 2001) (Figure 2.7).

There are several primary differences and similarities between the alignments and, therefore, models. Although all the alignments had N terminal insertions, the two manually produced alignments differed only in the C terminus and both had N terminal insertions of 57 residues. The automated alignments have larger N terminal insertions of 70 and 99 for mGenTHREADER and FUGUE respectively. Both automated alignments also have large C terminal insertions of 181 and 162 residues for FUGUE and mGenTHREADER respectively. The first manual alignment has a 145 residue C terminal extension whereas the second only has a 122 residue insert upstream.

No attempt was made to model the insertions in GSIII relative to GSI because they were too large to be contained in loop databases, as demonstrated by error reports from failed attempts to use the SWISS-MODEL homology-modelling server (Schwede, Kopp, et al. 2003). In addition, they displayed no homology to known folds as determined by threading (data not shown). When the *B. fragilis* GSIII sequence was submitted to the automated SWISS-MODEL server for fold prediction and model building, a prediction was only possible for residues 198-307, which correspond to the most conserved regions of alignment 1. Additionally, the server could not produce a model based on alignment 1 and a choice of GSI template structure (PDB entry 1fpy (Gill & Eisenberg 2001)).

2.4 DISCUSSION

The results of the automated alignments did not agree with those of previous researchers (Section 2.1.1). This suggests that either these algorithms are not sensitive

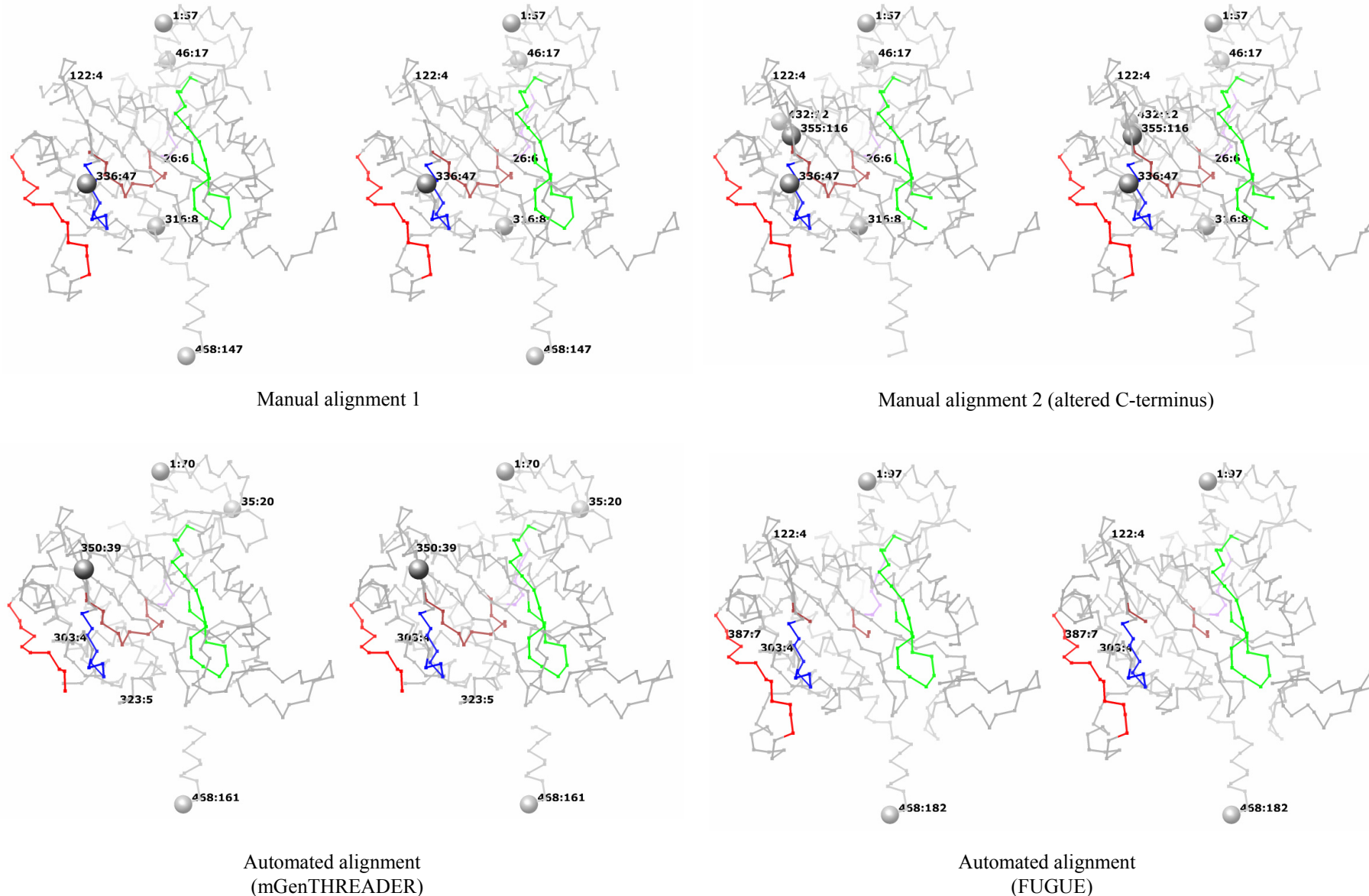


Figure 2.7. Stereo α -carbon traces of models of GSIII based on the four alternate alignments of GSIs and GSIIIs (described in the text) and derived from the structure of *S. typhimurium* GSI. The five conserved regions, corresponding to β sheets, are coded by the same colours used in Table 2.3. Regions of the peptide back-bone corresponding to deletions in the alignment between GSIII and GSI are not shown and insertions are marked by text labels (the number before the colon is the residue position and the number after is insertion size in residues). Spheres mark the position of large insertions for easy reference. The view is into the active site of the enzyme with the 6-fold axis vertical.

enough to detect the subtle sequence homology between these proteins or that not all the regions previously reported to display sequence homology, represent conserved features in the protein. The former seems more likely scenario for the following reasons. Firstly, the global similarity between the GS families is very low. Secondly, empirical evidence in the form of biochemical studies (Eisenberg, Gill, et al. 2000) supports the notion that all GS enzymes share a common catalytic mechanism. Thus, the key active site residues and structural folds should be conserved (see below). In particular, GSIII has been shown to require the same substrates and cofactors, to show similar kinetic profiles, and to be inhibited by similar compounds as GSI and GSII (Southern 1986; Garcia-Dominguez, Reyes, et al. 1997; Eisenberg, Gill, et al. 2000).

So, why did the automated procedures fail? They appear to have failed at the point of profile alignment using 1D PSSMs, as evidenced by the unproductive attempts to align profiles of the GS families together. This is likely the result of too great a distance between the families, for example, 9% sequence identity between GSIIIs and GSIs. The sensitivity of the profile based alignment methods at detecting subtle homologies improves with the number of sequences. The automated procedures employed, including those not shown (3D-PSSM (Kelley, MacCallum, et al. 2000) and SWISS-MODEL (Schwede, Kopp, et al. 2003) utilise algorithms which searched public databases and incorporated many sequences in the profiles. Despite this they were unable to correctly align the regions described above. Structural knowledge about conserved active site residues and binding motifs together with underlying structural principles can be used to detect these subtle homologies at the sequence level. The use of 3D profiles attempts to achieve this in an automated manner. However, just like 1D profiles, this method relies on a large number of templates to generate sensitive 3D-PSSMs. The only structures that exist for GSs, *S. typhimurium* and *M. tuberculosis*, are both GSIs with 47% sequence identity between their sequences. This suggests that the 3D-PSSM derived from their structural alignment is not likely to be sensitive towards homology between families.

Manual alignment, on the other hand, was able to demonstrate the presence of numerous conserved regions throughout the GS sequences by incorporating previously reported homologies and structural insights. In some regions, where

sequence identities were not high, it was still possible to demonstrate the conservation of patterns of amino acids associated with the interactions of secondary structures, thereby inferring their conservation. The nature and importance of these patterns was determined by inspecting the local environment of these residues in the GSI structure. Such structural homology is not represented by sequence identity figures and adds confidence to the alignment and, therefore, the homology models. The regions of high local similarity were found predominantly toward the centre of the sequences in the alignments, indicating that the diversity between the GS families lies predominantly in their terminal regions. This is especially true for the GSIII proteins who have a much longer sequence, resulting in them sharing the lowest sequence similarity to the other families and making them the hardest to align. It was for this reason that two alternate alignments were produced.

According to the manual alignments, the most conserved regions (shown in red in Figure 2.8) correspond to (a) the β sheets comprising the main structural motif of the C domain, the active site α/β barrel, (b) the short region of the N domain (the latch), which completes the active site on an adjacent subunit, and (c) the loops entering or leaving the barrel, which in parallel α/β barrels are critically involved in the active site reactivity (Brandon & Tooze 1999). The intervening sequences between the β sheets of the barrel form loops and helices, with the latter packing against the back of the barrel to form a compact globular structure. It seems likely, therefore, that all GSs contain a core conserved α/β barrel fold with the major differences between families occurring in the terminal and loop regions. The conservation of these structural features together with the finding that all the critical active site residues reported by Eisenberg are conserved between the GS families (Eisenberg, Gill, et al. 2000), suggests that all GS enzymes utilize a common fold to carry out similar catalytic mechanisms. The low global sequence similarity of the GSIIIs to the GSIs, suggests that comparative modelling will not be able to predict the entire structure of GSIII from sequence-sequence alignments alone with any degree of confidence. The regions of high local similarity, however, are more promising targets for successful prediction. The highest confidence in the predicted structure, therefore, resides in the central regions forming the α/β barrel. This is further supported by the model produced by the SWISS-MODEL server (Schwede, Kopp, et al. 2003) (data not shown)

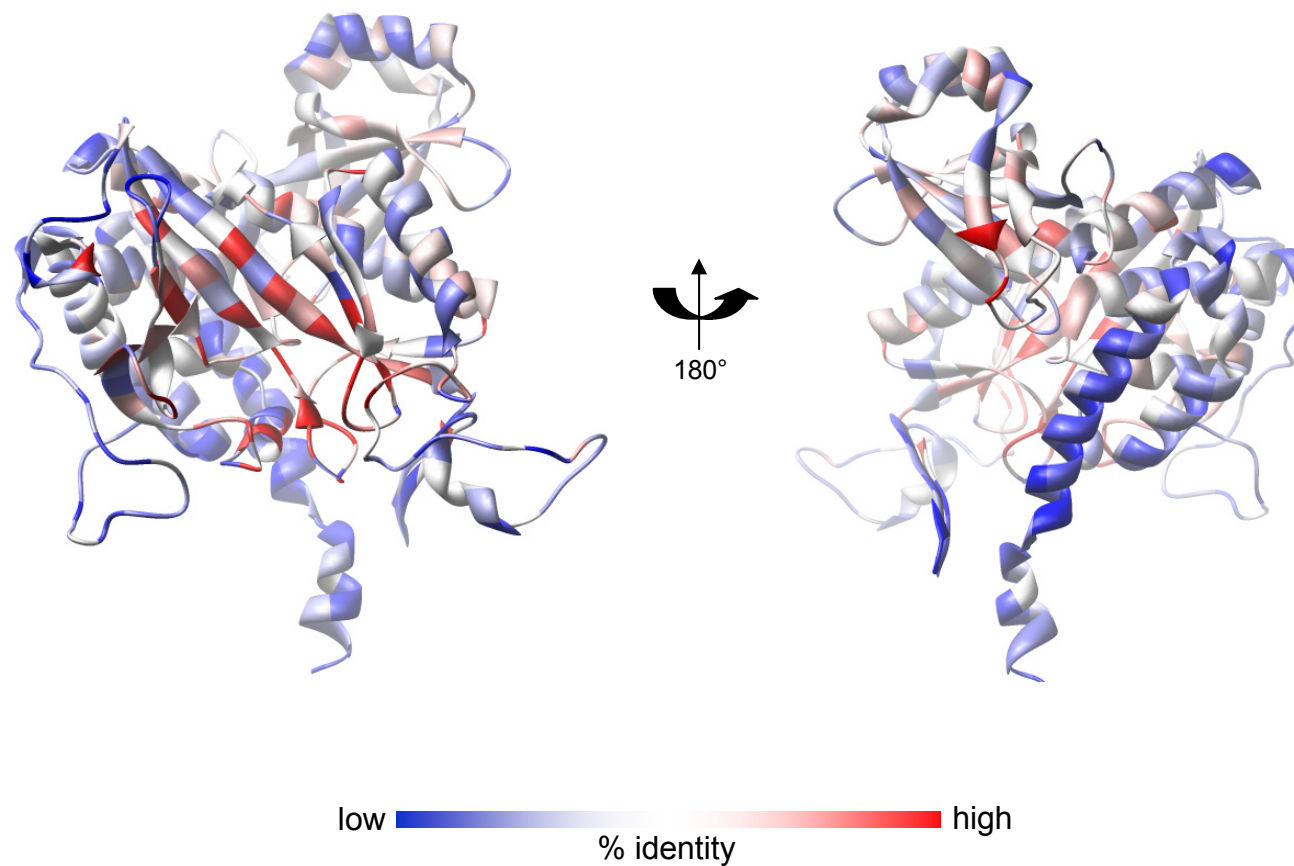


Figure 2.8. Confidence of the predicted homology model. Residue conservation between all GSs. A cartoon representation of the structure of GSI from *S. typhimurium* colour coded according to residue conservation (% sequence identity) on a linear scale from blue (13%) to white (30%) to red (100%) is shown. The view on the left is into the active site with the 6-fold axis vertical.

which only corresponded to this central region. However, it is only the architecture of this fold that is likely to be conserved and not the topology. It is, therefore, hoped that, if the resolution permits, it will be possible to determine the presence and position of the GSIII active site by inspection of the EM structure for such a predicted α/β barrel. The large open channel through the centre of the barrel and the fact that the active site of GSI is formed between subunits suggest that this will be possible even at resolutions below those required to interpret secondary structure directly. Outside of this central region, the confidence in the predicted structure wanes as a result of uncertainties in the correct alignment. It was for this reason that several models were produced according to the alternate alignments and based upon the GSI structure with no loop insertions modelled or residues mutated. The two models derived from manual alignments of GSIII and GSI possess alternate insertions in these loop regions that both seem reasonable.

This chapter has revealed how bioinformatic methods can propose predictions for the structure of *B. fragilis* GSIII. In the absence of close global homology to a known structure, however, these methods cannot substitute for empirical structural data. The docking of these homology models into the 3D volume, to be determined in Chapter 4, will not only aid in interpreting the structure but at the same time evaluate the accuracy of these predictions. The following chapter describes the purification of GSIII in preparation for the empirical structure determination by negative stain EM and single particle reconstruction techniques described in Chapter 4.

CHAPTER 3

PURIFICATION OF RECOMBINANT *BACTEROIDES FRAGILIS* GLNA

This chapter describes the expression and purification of recombinant *B. fragilis* GlnA from the glutamine synthetase deficient *E. coli* host strain, YMC-11. The aim of this purification was to achieve a homogeneous particle preparation, which is an absolute requirement for structure determination by single particle techniques (Chapter 4).

3.1 INTRODUCTION

3.1.1 Purification

A number of organisms have been shown to possess functional GSIIIs, including: the anaerobes *Bacteroides fragilis* (Southern, Parker, et al. 1987), *Butyrivibrio fibrisolvens* (Goodman & Woods 1993) and *Prevotella bryantii* B₁₄ (Wen, Peng, et al. 2003); the cyanobacteria *Synechocystis* sp. PCC 6803 (Reyes & Florencio 1994a) and *Pseudanabaena* sp. PCC 6903 (Crespo, Garcia-Dominguez, et al. 1998). Additionally, sequence information from the genomes of a several organisms has revealed genes possibly coding for GSIII enzymes, for example, the radiation resistant bacterium *Deinococcus radiodurans* (White, Eisen, et al. 1999) and the anaerobe *Bacteroides thetaiotaomicron* (Xu, Bjursell, et al. 2003). However, GSIIIs have only been purified from *B. fragilis* (Southern, Parker, et al. 1987) and *Synechocystis* sp. PCC 6803 (Garcia-Dominguez, Reyes, et al. 1997). These two proteins share 41% sequence identity (Chapter 2) and are, therefore expected, to exhibit similar properties. In the first isolation of GSIII from *B. fragilis*, Southern *et al* (Southern, Parker, et al. 1987) showed the enzyme to be a hexamer as deduced from the results of calibrated size exclusion chromatography and non-denaturing electrophoresis. The later isolation from *Synechocystis* sp. PCC 6803, Garcia-Dominguez *et al.* arrived at the same conclusion using similar techniques. However, the latter purification was methodologically quite different, employing ion-exchange and affinity chromatography, followed by ultrafiltration to purify the enzyme.

3.1.2 Regulation

Because of their central importance, GSs are regulated at numerous levels by many different mechanisms (Eisenberg, Gill, et al. 2000). Since only two GSIII have been purified to date and they have been poorly characterized in comparison to the GSIs and GSII, little is known about the regulation of the GSIIIs. What is known is as follows: GSIII expression is repressed by nitrogen (Southern 1986). The control elements responsible for this regulation are thought to reside near the GSIII structural gene, as repression is seen for the cloned *glnA*. Additionally, it appears that the activity of GSIII is also rapidly regulated in response to nitrogen. When nitrogen starved YMC-11 *E. coli* cells containing the cloned GSIII are exposed to ammonium sulphate (0.25 mM) just prior to protein extraction, the GS activity is twofold less than in untreated cells. Similarly, for *Synechocystis* sp. PCC 6803 it has been shown that transcription of the *glnN* gene (GSIII) increases under condition of nitrogen deprivation and the GSIII enzyme is rapidly and reversibly inactivated on addition of ammonium to nitrogen starved cells (Garcia-Dominguez, Reyes, et al. 1997). It, therefore, appears that in addition to control at the level of transcription, these GSIIIs are also subject to post-translational regulation. How this is achieved and whether these two enzymes share a common mode of inactivation is unknown. The possibility of adenylation, as seen in the GSIs, has been ruled out by Southern *et al.* because snake venom phosphodiesterase treatment does not lead to reactivation of the enzyme (Southern 1986). The other forms of modification, such as feedback inhibition, metal ion dependency, pH dependency, and oxidation etc. are all possible. A novel mechanism could potentially also be at play. It is interesting to note that the first incidence of GS being inhibited by protein-protein interaction was reported for GSI in *Synechocystis* sp. PCC 6803 (Garcia-Dominguez, Reyes, et al. 1999).

3.2 METHODS

3.2.1 Origin of bacterial clones and plasmids

Plasmid pJS139 contains the structural gene coding for *B. fragilis* BF-1 glutamine synthetase in a 8.7kb insert. Southern originally isolated this plasmid from a genetic

library prepared from *B. fragilis* Bf-1 strain by screening for clones that could complement the glutamine synthetase (*glnA*) deficiency in an *E. coli* (YMC11) deletion strain.

Table 3.1: Bacterial strains and plasmids used in this study

Strain	Relevant genotype	Reference
<u><i>Escherichia coli</i></u>		
YMC-10	<i>glnA</i> ⁺ , <i>ntxB</i> ⁺ , <i>ntxC</i> ⁺ , Ap ^S	(Backman, Chen, et al. 1981)
YMC-11	<i>glnA</i> ⁻ , <i>ntxB</i> ⁻ , <i>ntxC</i> ⁻ , Ap ^S	(Backman, Chen, et al. 1981)
<u>Plasmids</u>		
pJS139	Ap ^r , <i>glnA</i> ⁺	(Southern 1986)
pMET104	Ap ^r , <i>glnA</i> ⁻	(Wehnert, Abratt, et al. 1990)

Plasmid pJS139 was originally harboured in the *E. coli* wild type strain YMC-10 and, therefore, was transferred to *E. coli* YMC-11, a *glnA* deficient strain, to ensure that only recombinant GlnA would be purified. *E. coli* YMC-11 was also transformed with pMET104 which is derived from the same pEcoR251 vector but contains an arbitrary DNA insert to serve as a negative control (Wehnert, Abratt, et al. 1990).

3.2.2 Growth conditions

E. coli was routinely grown in Luria Bertani broth (LB) (1% tryptone, 0.5% yeast Extract, 0.5% NaCl, in distilled water) (Davis, Botstein, et al. 1980). All percentage concentrations described here are expressed as weight per volume. Where solid media was required such as for colony selection and strain storage, 1.5% (w/v) agar was used. Cold Spring Harbour (CSH) minimal medium (Davis, Botstein, et al. 1980) (0.45% KH₂PO₄, 1.05% K₂HPO₄, 0.047% Sodium Citrate, 0.26% (NH₄)₂SO₄, 0.01% MgSO₄.7H₂O, 1.5% agar, 1ug/ml vitamin B1, 0.2% glucose), was used during transformant selection and GS induction. Ampicillin (Ap) was used at a final concentration of 100 µg/ml where necessary.

3.2.3 Plasmid extraction and transformation

Plasmid DNA was isolated from the YMC-10 (pJS139) strain using the Qiagen mini-prep plasmid isolation kit according to the manufacturer's instructions. Competent *E. coli* YMC-11 cells were prepared by the standard procedure of cold glycerol permeabilization and then transformed with pJS139 by electroporation (Bio-Rad). Transformants were selected on LB agar (LA) medium (Davis, Botstein, et al. 1980) containing Ap. The ability of *E. coli* YMC11 (pJS139) to utilize $(\text{NH}_4)_2\text{SO}_4$ as a sole nitrogen source was demonstrated by growing transformants on CSH minimal medium and Ap, confirming the ability of pJS139 to complement the glutamine synthetase deficient *E. coli* YMC-11 strain and verifying the presence of the recombinant structural gene, *glnA*.

3.2.4 Expression

For protein purification, *E. coli* YMC11 (pJS139) colonies were picked from freshly plated stock cultures (LA + Ap), inoculated into precultures (5 ml LB + Ap) and grown with aeration for ~5.5 hours at 37°C. Larger (750 ml) cultures (LB + Ap) were then inoculated with the entire 5ml of preculture and incubated with aeration for approximately 13 hours at 37°C. The YMC-11 (pJS139) overnight cultures were diluted into equal volumes (750ml) of nitrogen-free CSH minimal medium and then incubated with aeration for 3 hours at 37°C. CTAB, a cationic detergent and stabilizer of GS adenylation state, was added to the diluted culture to a final concentration of $0.1\text{mg}\cdot\text{ml}^{-1}$ and incubation continued for 10 minutes.

3.2.5 Harvesting and lysis

The CTAB-treated *E. coli* YMC11 (pJS139) cells were collected by centrifugation for 10 minutes at 7500 rpm at temperature of 4°C, and resuspended in 1/100th the original culture volume (i.e. 7.5ml) in extraction buffer (50 mM Imidazole-HCL, 20 mM MgCl_2 , 0.1 mg/ml PMSF, and 20 mM 2-mercaptoethanol, pH 7.15). Lysozyme was then added to a final concentration of 10 mg/ml and the cells aerated at 37°C for 30 minutes before being chilled on ice for an additional 30 minutes. Cells were

disrupted by sonication for 8 minutes (2 minutes cycles with 30 seconds cooling periods between). An ethanol/ice slurry was used to cool the sample during sonication. The supernatant was then clarified by centrifugation at 15000 rpm for 30 minutes and kept on ice. This fraction is referred to later in the text as a cell free extract.

3.2.6 PEG precipitation

NaCl was added to the whole cell extract to a final concentration of 0.1 M. PEG (6000 MW) from a filter-sterilized 20% stock solution was then added dropwise, with stirring, to a concentration of 11% (w/v). The sample was maintained on ice throughout the precipitation (~ 4hrs) and then stored at 4°C overnight. The precipitate was then collected by centrifugation and redissolved in 4 M NaCl at half the original precipitation volume.

3.2.7 Gel filtration chromatography

The redissolved precipitate (4.5 ml) was loaded onto a pre-packed 60×1.6cm Sephacryl 300 HR column (AP Biotech), equilibrated and eluted with column buffer (10 mM Imidazole-HCL, 1% (w/v) KCl, pH 7.15) (C buffer) at 0.2 ml/min. This column has a fractionation range of 1×10^4 - 1.5×10^6 . Fractions (3ml) were collected, and selected fractions were assayed for GS activity (see below) and protein concentration, and analysed by SDS-PAGE (7.5% w/v acrylamide concentration and coomassie blue staining) (Laemmli 1970). The protein concentration of the eluent was measured at 280 nm and recorded by an inline spectrophotometer coupled to a PC. The Bradford (Bradford 1976) assay (with BSA as a standard) was also used to determine the protein concentration of selected fractions.

The three fractions (3ml) corresponding to the centroid of the activity peak were pooled, concentrated (see 3.2.8), and subjected to an additional round of gel filtration chromatography. The second gel filtration was performed as above except that all of the concentrated protein solution (1.5 ml) was loaded onto the pre-equilibrated column and 1.5 ml fractions were collected.

3.2.8 Ultrafiltration

The 3 most active fractions of the first gel filtration step were pooled and concentrated to 1.5ml by ultrafiltration using a 50ml Amicon bioseparations stirred cell with a Millipore PM10 10KDa exclusion membrane.

3.2.9 Ultracentrifugation

In order to avoid the detrimental effects of high salt concentrations on the quality of negative staining and to concentrate the sample (~1mg/ml), the most active fraction (37) from the second gel filtration run was concentrated and the buffer exchanged for a low-salt buffer (Imidazole-HCl, 10 mM; MnCl₂, 10 mM; pH 7.15) using a Millipore YM10 centricon centrifugal filtration device with a MW exclusion of 100KDa. The constituents for this buffer were based on consensus constituents of other purification protocols and the conditions used by Southern et al. (Southern, Parker, et al. 1987).

3.2.9 Measurement of activity

GS activity was assayed by the γ -glutamyl transferase (GGT) assay as described by Bender *et al.* (Bender, Janssen, et al. 1977). The modifications determined by Southern to provide the optimum conditions for the assay were adopted (Southern 1986). Specific activity was expressed as μ M of glutamyl hydroxamate formed per min, per mg protein. According to Southern 1 μ mol of glutamyl hydroxamate gave 0.525 absorbance units at wavelength 540nm.

3.3 RESULTS

Preliminary purification trials were undertaken to demonstrate the effectiveness of Southern *et al.*'s methods (Southern, Parker, et al. 1987) at producing a homogenous sample for study by EM and single particle methods.

3.3.1 Transformation

The *B. fragilis* structural gene product was shown to be the only GS produced by the transformed *E. coli* expression host YMC-11 (pJS139) and to be functional. *In vivo*; the pJS139 plasmid enabled this glnA deletion strain to utilize (NH₄)₂SO₄ as the sole nitrogen source, thus, rescuing this auxotroph, whereas transformation of the same strain with a control plasmid, pMET104, (same vector but noncoding insert) did not alter the auxotrophic phenotype. *In vitro*, investigation of the specific activity of GS in whole cell extracts derived from YMC-11 (pJS139) and YMC-11 (pMET104) revealed that only the former had significant activity (20 U/mg) compared to the latter (0.0 U/mg).

3.3.2 Expression and purification

The results of the purification trials are summarized in Table 3.2, the most critical of these being a significant loss in activity occurring during the PEG fractionation step. Results from the initial trial purifications are included to illustrate the significant loss of activity.

Table 3.2: Purification table

	Volume (ml)	Concentration (mg/ml)	Total protein (mg)	Total activity (U)	Specific activity (U/mg)
Culture	2250				
Cell extract	19.5	46.3*		4563*	19.7*
Redissolved PEG ppt (4 M NaCl)	11			25.3	
Gel filtration volume loaded	4.5			10.3	
Pooled fractions from gel filtration run 1	9	1.08	9	10.43	1
Gel filtration run 2 fraction 36	1.5	0.53	0.79	0.71	0.9
Gel filtration run 2 fraction 37	1.5	0.49	0.74	0.89	1.2

* From first trial purification

3.3.2.1 *PEG precipitation*

In Southern's original protocol (Southern, Parker, et al. 1987), a selective PEG (6000) fractionation between 4% and 6% was used as the first crude purification step after cell lysis. The 4% step was reported to leave GS in solution but precipitate other soluble macromolecules. Thereafter, raising the PEG concentration to 6% would selectively precipitate GS. However in this study, significant difficulty was encountered with the PEG precipitation step using the values described by Southern *et al* (Southern, Parker, et al. 1987). Even after empirically determining the precipitation conditions, scale-up of the precipitation resulted in premature precipitation of GS in the first step (as determined by supernatant activity). It was, therefore, necessary to adopt the protocol of Streicher *et al.* (Streicher & Tyler 1980). In this protocol the GS

enzyme is first completely precipitated by 10% PEG (6000) and then redissolved in a high concentration (3-4 M) NaCl solution. The high salt ensures that GS is selectively redissolved whereas other proteins and DNA are not.

3.3.2.2 *Gel filtration*

The first gel filtration step did not produce well resolved peaks (Figure 3.1A). However, when the purity of representative fractions was assessed by SDS-PAGE (Coomasie blue staining), only one band corresponding to 82KDa was apparent in the fractions and its abundance correlated with the protein concentration measured by the inline spectrophotometer at 280nm. Furthermore, the main peak of the activity profile was quite broad but its central moment did correspond with one of the protein peaks. A preparation from the most active fraction was also negatively stained (as described in Chapter 4) and viewed under the EM to investigate its homogeneity (Figure 3.1B). Large particles were clearly apparent in the preparation including some resembling those seen by Southern (with a dark central region) (Southern, Parker, et al. 1987). However, the preparation was not homogenous, necessitating an additional gel filtration passage.

The results of the second gel filtration step were consistent with the first. The maximum activity eluted at a very similar elution volume (Figure 3.2) but there was also a broad distribution of unresolved peaks which was unexpected since only the fractions corresponding to the centre of the peak in the first gel filtration passage were loaded the second time. Representative fractions of the protein profile were again assessed for homogeneity by EM and all were shown to contain particles of a similar size and shape (Figure 3.3A) indicating homogeneity. Fraction 28, one of the first fractions to elute, did however also contain what appeared to be protein aggregates.

The final preparation appeared homogenous after desalting except for a small amount of contamination by putative GroEL (Figure 3.3B). Due to the quality of the resultant purification preparation and the time constraints, it was decided to commence structural analysis on this sample and no further preparations or biochemical investigations were undertaken.

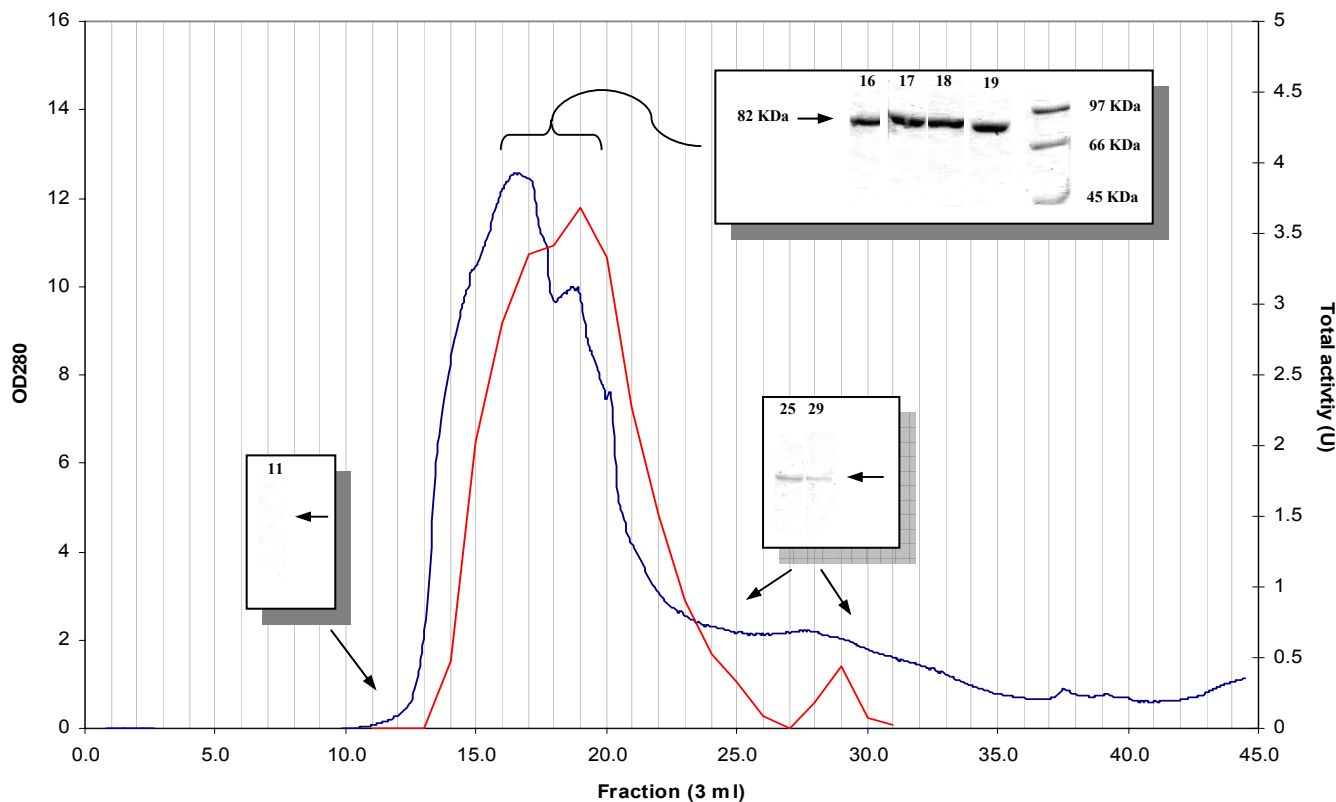
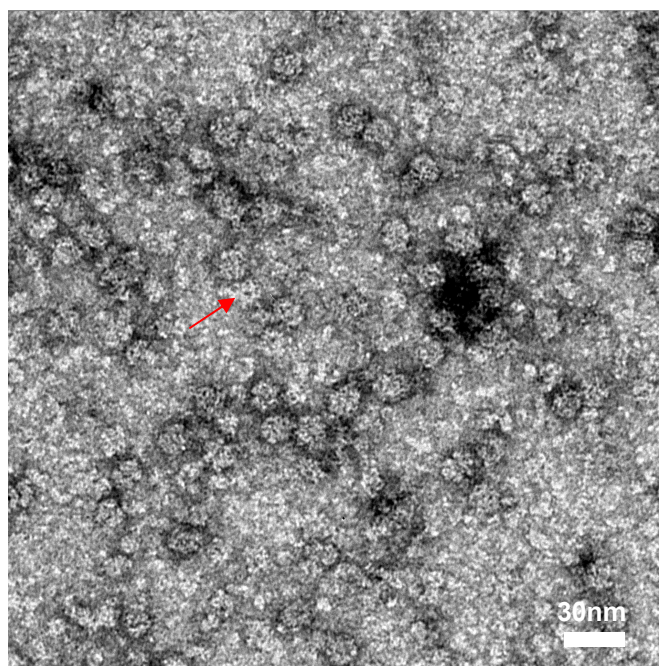
A**B**

Figure 3.1. First gel filtration passage. **A** – Gel filtration. Blue – OD280, Red – Total activity. Inset – Coomassie stained SDS-PAGE of selected fractions. The arrow indicates where the 82 KDa GlnA monomer was expected to run. **B** – Electron micrograph of negatively stained (Chapter 4) fractions 18,19 and 20, pooled from the activity peak in A (C buffer). Arrow- characteristic particle with central hole.

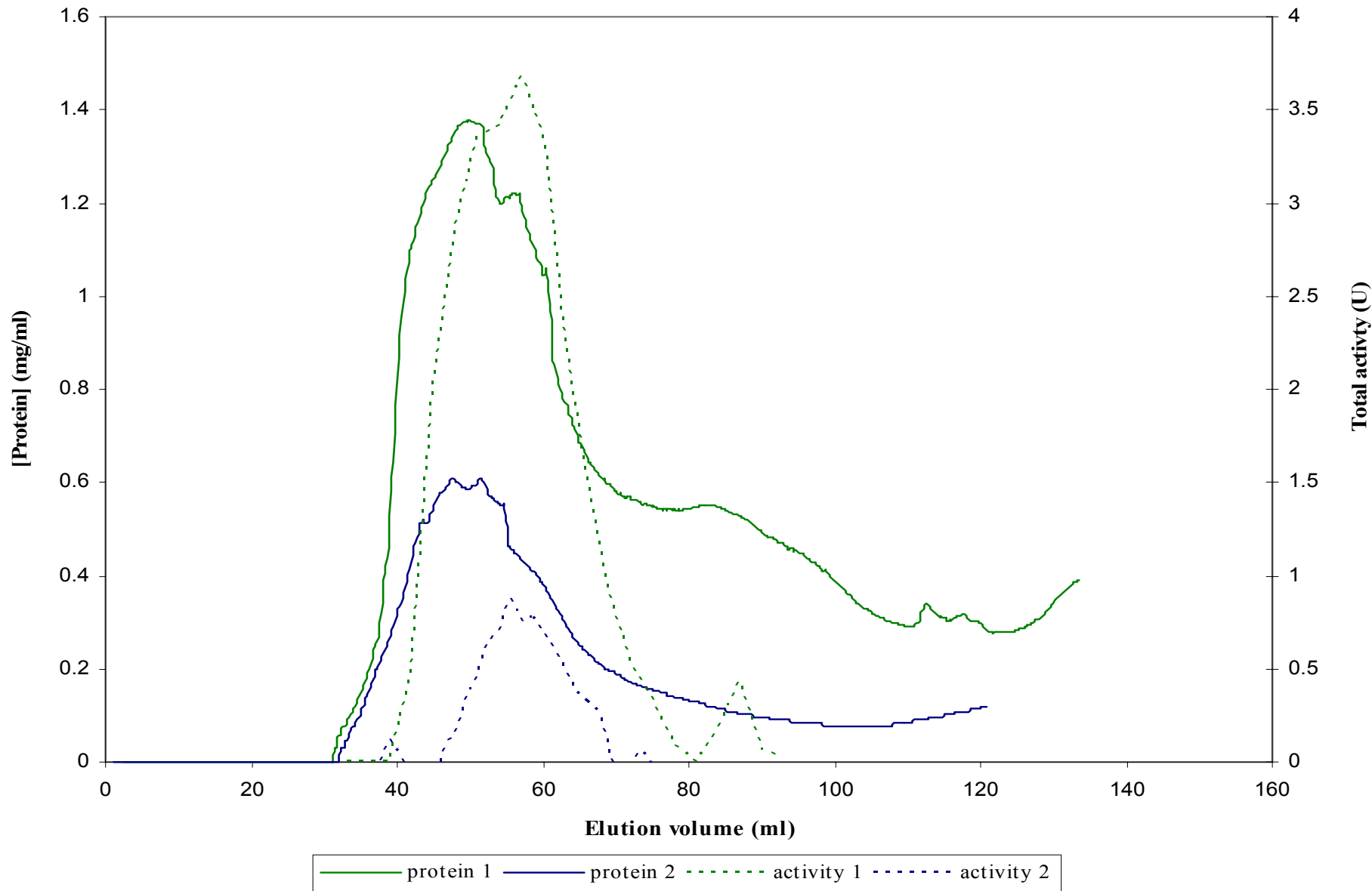


Figure 3.2. Comparison of gel filtration passages. The three most active fractions (3 ml each) of the first gel filtration passage were pooled, concentrated to 1.5 ml, loaded onto the pre-equilibrated column for the second passage. Both gel filtrations were performed under exactly the same conditions except that 1.5 ml fractions were collected the second time around.

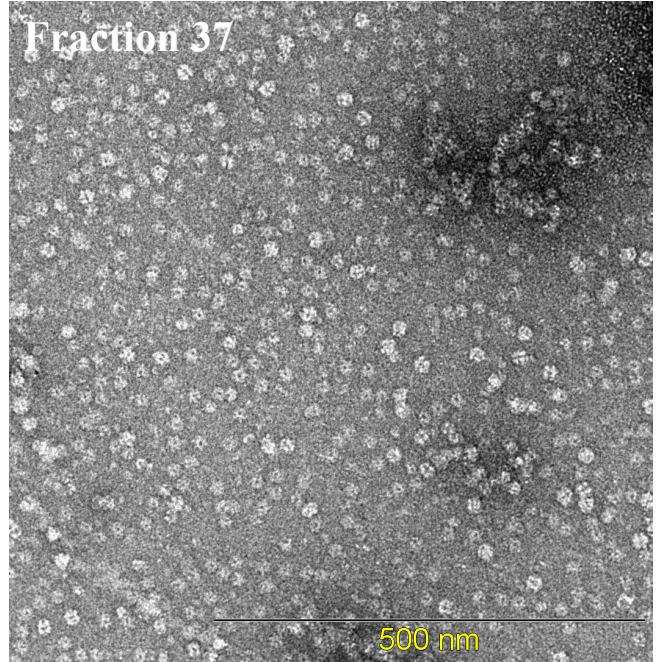
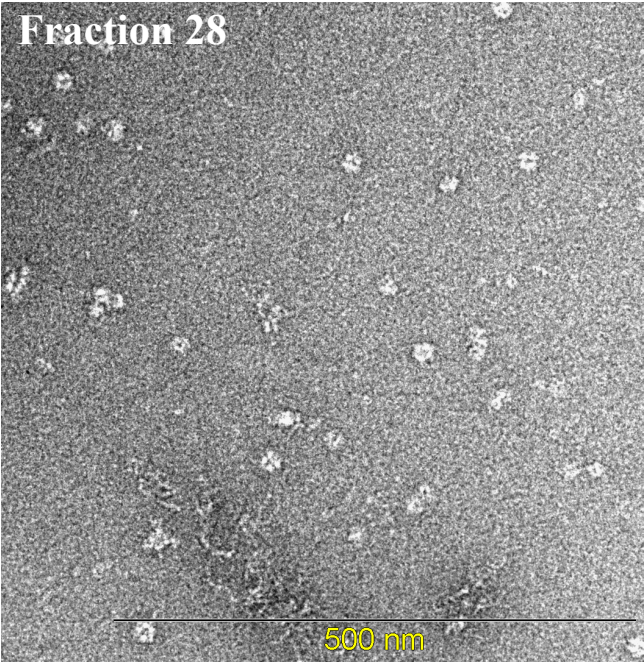
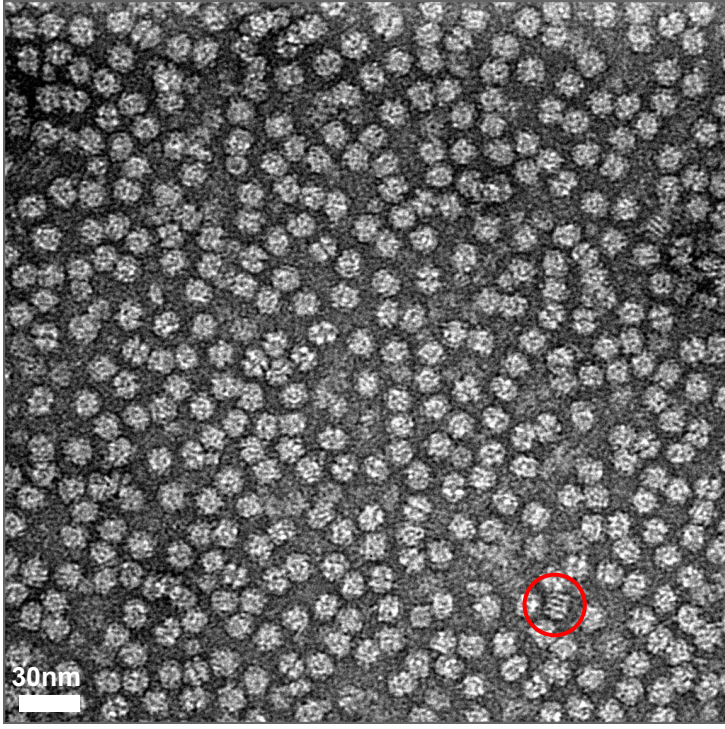
A**B**

Figure 3.3. Electron micrograph of the negatively stained samples during purification. **A** - Electron micrographs of negatively stained (see Chapter 4) examples fractions (C buffer) from the final gel filtration passage. **B** - Electron micrographs of final purified sample (in low salt 10 mM Imidazole-HCl, 10 mM MnCl₂, pH 7.15 buffer), negatively stained as described in Chapter 4. The ringed particle is putative GroEL.

3.4 DISCUSSION

The large loss of activity seen during the purification was unexpected. This occurred primarily during the PEG precipitation, with a 180-fold loss in total activity. Subsequent loss of activity did take place but to a much lesser extent. Initially the amount of total activity contained in the crude cell extract (~200U/ml) was far larger than that achieved by Southern *et al.* (2.5U/ml) (Southern, Parker, et al. 1987) and twice that accomplished for *Synechocystis* GSIII (~100U/ml) (Garcia-Dominguez, Reyes, et al. 1997). However, the final specific activity (~1 U/mg) was lower than achieved by Southern (101.4 U/mg) and for *Synechocystis* GSIII (36.8 U/mg).

Losses in activity were not exclusive to this purification and were also reported for the GSIII purification from *Synechocystis* sp. PCC 6803 (Garcia-Dominguez, Reyes, et al. 1997). Despite including stabilizing agents in their extraction buffer (see below) only 18% recovery was possible. The opposite was true for the original Southern *et al.* purification (Southern, Parker, et al. 1987), where in fact, an increase in specific activity was recorded during the purification resulting in a final recovery of 114%. This was attributed by these authors to the removal of an inhibitor during the 6% PEG (6000) precipitation. Further evidence for the presence of such an inhibitor was cited as coming from failed attempts to purify GSIII from native *B. fragilis* (Southern 1986). Whole cell extracts displayed almost non-existent levels of activity despite Western blot analysis confirming the presence of the expressed protein, and all activity was subsequently lost during the 4% PEG (6000) precipitation. Furthermore, whole cell extracts were shown to be capable of inhibiting purified recombinant GSIII isolated from *E. coli* by 91%. What is most interesting is that extracts from YMC-11 were also shown to be capable of inhibiting cloned GSIII by 80%. The inhibitor was never isolated.

The extent of knowledge about GSIII regulation was introduced in section 3.1.2, where it was revealed that both GSIIIs investigated to date are subject to rapid, and, in the case of *Synechocystis* sp. PCC 6803 GlnN, reversible post-translational inactivation in response to ammonium ions. In the light of this apparent nitrogen regulated inhibition, the question can be asked whether the inactivation seen during

the purification of GSIII from *E. coli* and *B. fragilis* is a result of some specific inhibition, by for example, covalent modification/feedback inhibition, or a result of conditions during the purification, such as pH or cofactor concentrations. The properties of GS enzymes are modified by a number of effectors and conditions in a complex fashion (Eisenberg, Gill, et al. 2000). One of the most important of these is the effect of metal ion cofactors. In GSI family, the enzymatic properties of this enzyme, activated by different cations, varies in catalytic potential, reaction kinetics, pH optima, sensitivity to feedback inhibitors, and in conformational state via altered heterologous subunit interactions (Segal & Stadtman 1972a; Segal & Stadtman 1972b). Furthermore, there appears to be complex interplay between the different cations, with for example Mg^{2+} enhancing the affinity of GSII for Mn^{2+} which has been shown to reduce the activity of the enzyme (Eisenberg, Gill, et al. 2000).

Alteration in oligomeric state is another enzymatic property that is strongly affected by the type and concentration of divalent metal ions, which together with accompanying changes in activity have been reported for both GSIs and GSIIIs (Denman & Wedler 1984; Maurizi & Ginsburg 1982). Complete removal of metal ions by chelating agents results in dissociation of the complexes into lower molecular weight species, such as octamers, tetramers, and monomers. These lower molecular weight species still exhibit activity but at a much reduced level. Enzyme concentrations and the presence of substrates can also alter the oligomeric state of these enzymes. For instance, the binding of ligands is thought to enhance intersubunit contacts just as increased concentration of divalent metal ions do, causing reassociation into higher molecular weight complexes (Maurizi & Ginsburg 1982). In particular, the addition of Mn^{2+} leads not only to reactivation but also to the formation of higher order complexes (Denman & Wedler 1984).

It is, therefore, apparent that slight changes in the concentration of effectors, or alteration in conditions during purification, can have drastic effects on the stability and activity of GS enzymes. There are, thus, several possible explanations for the loss of activity seen during this purification. (1) The S300HR column is ideal for separation of a 500KDa protein as it falls in the middle of its selectivity range. It was, therefore, unexpected that the gel filtration steps did not resolve the protein

constituents satisfactorily. Instead, a distribution of peaks all appearing to be comprised of an apparently homogenous (according to subunit MW determined by SDS-PAGE) population was obtained. Repassage of sample from the centre of the most active peak, which should have produced a single peak, revealed the same distribution of peaks indicating that the preparation redistributes to range of molecular weights or volumes during the period between gel-filtration runs. The presence of aggregates in the first peak to elute suggests that this equilibrium distribution might represent degradation products. Dissociation into smaller oligomeric states with subsequent loss of activity could also explain the results. The lack of any Mn^{2+} in the gel filtration buffer would support the latter hypothesis. However, Southern achieved a homogenous gel filtration profile using the same conditions (Southern 1986), and the purified protein corresponded to a hexamer, in agreement with the work of Garcia-Dominguez *et al.* who used 2 mM Mn^{2+} (Garcia-Dominguez, Reyes, et al. 1997). (2) Because it is known that PEG precipitation is achieved through increasing the excluded volume of proteins in solution or molecular crowding, conditions that promote protein-protein association or those that cause dissociation should facilitate or impede precipitation respectively (Ingham 1984). Thus, it is possible that the difficulties seen during PEG precipitation were a result of other factors affecting the oligomeric state of GSIII. Alternatively, PEG, via its effect on excluded volume, could have altered the degree of oligomerization and, therefore, activity of the enzyme. Similarly, the use of glycerol by Garcia-Dominguez could account for their better yield (Garcia-Dominguez, Reyes, et al. 1997). (3) Additionally, any covalent or specific inhibition of the enzyme would most likely be noticeable during the PEG precipitation because it takes place over lengthy time periods (4hrs and 24hrs). The possibility of feedback inhibition seems unlikely because the metabolic products required would have been present from the beginning of the purification. However, they might only become effective under different conditions. (4) Irreversible denaturation due to inappropriate conditions could account for the loss of activity but this seems unlikely because of the homogeneity and structural integrity of the complexes visualized by EM. Furthermore, the temperature optimum of this enzyme is 45°C (Southern, Parker, et al. 1987) excluding the possibility of rapid thermal denaturation at room temperature. (5) The high concentrations of salt used to redissolve the PEG 6% precipitate could have further effected activity, but the

majority of inactivation was seen before this. (6) The activity of preparations was only measured by the transferase activity of GSIII. It is, therefore, possible that GSIII might have exhibited higher activity if measured by the biosynthetic assay, because the catalytic potential of these reactions differ according to effectors and conditions. Furthermore, it has been shown previously for GSI that the characteristics of inhibition, as measured by inhibitor kinetics, can differ between the two principal assays used: the forward biosynthetic assay, and the reverse gamma-glutamyl transferase assay (Liaw, Pan, et al. 1993). This discrepancy was explained in terms of the different metal ion dependencies of each assay (the biosynthetic assay is inhibited by Mn^{2+} compared to Mg^{2+}) and also the relative stabilities of the different GS-cofactor complexes, GS-ADP and GS-ATP.

It is clear from the above possibilities that further investigation into the mechanism of inactivation is required to understand the low levels of GSIII activity in *B. fragilis* and during purification of the recombinant enzyme. In particular, it needs to be ascertained whether the inhibition is mediated by a specific inhibitor or by perturbation of environmental condition required for activity. Even if the latter turns out to be the case, it will still remain to be determined whether this inactivation is of physiological significance. Furthermore, Garcia-Dominguez *et al* mentioned that the inactivation seen in *Synechocystis* sp. PCC 6803 is reversible (Garcia-Dominguez, Reyes, et al. 1997). It also uncertain whether the reactivation procedures described for GSI and GSII can restore the activity of purified GSIII (Eisenberg, Gill, et al. 2000; Maurizi & Ginsburg 1982; Denman & Wedler 1984).

It is acknowledged that more rigorous investigation is required to improve the efficiency and reproducibility of purification protocol. Nonetheless, the primary objective of the purification, to obtain a homogenous sample for analysis by EM, was achieved. The following chapter describes the visualization of the purified GlnA by negative stain EM and the elucidation of its 3D structure by single particle reconstruction techniques.

CHAPTER 4

SINGLE PARTICLE RECONSTRUCTIONS

4.1 INTRODUCTION

4.1.1 General background

Previous structural studies have revealed that glutamine synthetases are, in general, double ringed structures, with an even number of subunits and dihedral point group symmetry (Kretovich, Evstigneeva, et al. 1984). Biochemical studies and preliminary EM investigations into the structure of GlnA from *Bacteroides fragilis* (Southern, Parker, et al. 1987), however, have suggested a hexameric quaternary structure for GSIII. This is in agreement with the quaternary structure (determined by size-exclusion chromatography) of *Synechocystis* GlnN, the only other GSIII isolated to date (Garcia-Dominguez, Reyes, et al. 1997). The results of biochemical techniques such as size-exclusion chromatography, centrifugation, and gel electrophoresis are often ambiguous due to experimental and theoretical difficulties associated with the techniques (Gouaux, Braha, et al. 1994; Haschemeyer & De Harven 1974). Even direct visualizations of particles by negative stain EM have yielded conflicting results (Stahl & Jaenicke 1972; Tsuprun, Zograf, et al. 1987; Kiang 2000). This uncertainty together with the finding that some GS enzymes occur in a range of oligomeric states, the inter conversion of which is particularly sensitive to environmental conditions and effectors concentrations (Chapter 3), detracts from the confidence of the previous findings for *B. fragilis* GlnA.

Negative stain EM and single particle reconstruction techniques were, thus, undertaken in order to better understand the oligomeric stoichiometry of this protein in solution and to reveal the 3D arrangement of subunits and substructure of a GSIII. In addition to the structure determination, this work aimed to reveal the relationship between GlnA and GSI via the docking of homology models (Chapter 2) into the 3D volume, thus allowing interpretation of the structure and shedding light on the evolution of this enzyme. Depending on the accuracy of the models and the resolution

achieved, such interpretations might permit inferences to be made about the function and regulation of GlnA based upon the extensive understanding of the structure/function relationships of GSI (Eisenberg, Gill, et al. 2000).

4.1.2 Single particle reconstruction

In order to reconstruct a particle in three dimensions it is necessary to determine the angular relationship between all projections or views of the particles. This can be achieved by collecting views with predetermined angular relationships as in the case of single axis tilt series/tomography or, alternatively, single particle methods can be used. In the latter technique the views are distributed over multiple randomly orientated particles.

In the simplest implementation of the single particle technique, 2D averaging, images from multiple particles are aligned together and averaged arithmetically with the primary aim of improving the signal to noise ratio of the images. Thus, single particle techniques can reduce the effects of noise arising from sources such as electron shot noise due to minimal dose imaging, which otherwise hamper interpretation of molecular details. 2D averaging presumes a defined orientation of the particle on the grid and thus only requires the determination of the two translational and single in plane rotational angle that relate the images to each other. Combining views from different particles also effectively distributes the effect of damaging radiation among all members of the particle population because each particle will incur damage in a different random location (review by (Glaeser 1999)). On averaging, the conformational variation resulting from this damage is averaged out. Negative staining also circumvents the sensitivity of biological samples to radiation by forming a cast of the molecular structure which is more resistant to radiation induced damage. The stain, thus, preserves an imprint of the hydrated structure, with greatly improved contrast due to the heavy atom composition, before desiccation and irradiation (Kiselev, Sherman, et al. 1990).

In three dimensions i.e. assuming a random distribution of particle orientations, five parameters relate projections to each other: three angular parameters (Euler or

alternate coordinate system) and the two translational parameters. These need to be determined such that the projections may be, by analogy to the 2D example, ‘averaged’ together by backprojection in three dimensions to form a model volume. Multi-reference alignment (Joyeux & Penczek 2002) or in its more complete implementation, angular refinement (Penczek, Radermacher, et al. 1992), offers a method for determining the parameters by matching projections to views of a reference volume. This method therefore requires the availability of a starting model of sufficient resolution to either determine all these parameters in the first search or in the more sensitive form, the iterative angular refinement procedure, provide a starting point for the process to bootstrap the search to higher resolution.

An alternate, means of deriving the relationship between all projections in the absence of prior information is given by the simultaneous minimization program based on common lines (Penczek, Zhu, et al. 1996). Because of the low signal-to-noise ratio of individual images, this method works best with averages of images produced by classification. This requires a very large data set so that enough particles occur in similar directions to allow averaging whilst maintaining a fine angular sampling. The initial model can then be improved by angular refinement as described above or by “alignment by classification” which only uses multi-reference alignment against projections of the previous model to improve the alignment of images which are thereafter subjected to reclassification and the process iterated.

4.2 METHODS

4.2.1 Sample preparation

Recombinant *B. fragilis* GSIII was isolated from a GS deficient *E. coli* expression host and purified as described in Chapter 3. The final preparation containing ~1mg/ml of GlnA in a low salt buffer (10 mM Imidazole-HCl, 10 mM MnCl₂ pH 7.15) was stored at 4°C.

4.2.2 Negative stain electron microscopy

10 ul of a $\frac{1}{4}$ dilution (in low salt buffer) of GSIII solution was applied to EM grids previously coated with thin carbon support film and additionally glow discharged in air for 20 seconds, before being stained with 2% uranyl acetate solution using the droplet method. Specifically, one 10 ul droplet of sample and three 10 ul droplets of stain solution were placed on ParafilmTM. The carbon coated grid was first floated on the protein solution for 25 seconds before being removed, blotted and, in quick succession with blotting in between, floated for 1 second on the next two stain droplets. The protein was left to stain for ~ 7 seconds on the final droplet, before being slowly blotted and air dried. The adsorption and final staining times required, for optimal particle concentration on the grid and stain contrast, were optimized empirically.

Electron micrographs (160) were recorded by a Proscan 2048 \times 2048 slow-scan CCD camera, with 14 μ m pixels, using a Leo 912 TEM operating at 120kV and a nominal magnification of 50000 times magnification with zero-loss energy filtering. Images were recorded by the minimum dose technique (Williams & Fisher 1970) programmed using the TVIPS software (Tietz, 1986; EM-MENU 2003), to reduce the exposure to damaging radiation. During the focussing stage of this procedure, astigmatism and defocus were corrected by inspection of the real time power spectra calculated from the images.

4.2.3 Image processing and reconstruction

Image processing and 3D reconstruction were performed on an Intel server with two 2.2 Mhz Xeon processors and 2Gb physical memory operating under the RedHat 9 Linux operating system. The majority of image processing steps and all reconstruction steps were accomplished using the program, SPIDER (Frank, Rademacher, et al. 1996). Two programs from the MRC image processing package IMAGE2000 (Crowther, Henderson, et al. 1996) were used in the initial preprocessing stages (see below).

4.2.3.1 Data handling

Micrographs which had been recorded as unformatted 2048×2048 pixel 16 bit arrays were converted to MRC format before being downsized by a factor of 2 using the IMAGE2000 program, Label (Crowther, Henderson, et al. 1996), to average adjacent pixels. Binning to reduce images size and speed up processing was possible because the final predicted sampling, 5.6 Å/pixel still corresponds to an over-sampling by a factor of 3.5 for an expected maximum resolution of 1/15 Å and a 14 μm CCD pixel. The theoretical maximum sampling required is given by the Whittaker-Shannon sampling theorem (Shannon 1949), which states that, in order to avoid aliasing effects, the sampling needs to be at least twice the highest frequency present in the signal. Over-sampling, however, is necessary to minimise the deleterious effects of accumulating interpolation errors during the reconstruction.

Although useful, this value only represents a rough estimate of the sampling because the microscope was uncalibrated at 50000× magnification. An accurate determination of the sampling is essential for determining correct filter parameters and for ensuring the accuracy of model scaling and measurements. Fortunately, contaminating putative GroEL particles were present in preparations. Contaminating putative GroEL particles were preprocessed as described below and then aligned together by reference free alignment (Penczek, Radermacher, et al. 1992). An atomic structure of GroEL (Braig, Otwinowski, et al. 1994) was then imported at different sampling values, low-pass filtered to 15 Å, and an average image calculated from 40 different projections at 9° intervals around the particle's long axis for each differently scaled particle. The maximum cross-correlation between the differently scaled atomic structure averages and the average from the micrograph yielded the final sampling of the micrographs.

Average periodograms (equivalent to the power spectrum for large number of samples) of the micrographs were also calculated using scripts obtained from the SPIDER (Frank, Radermacher, et al. 1996) web site (www.wadsworth.org/spider_doc/spider/docs) in order to screen for astigmatism, drift, and incorrect defocus ranges.

4.2.3.2 Particle selection

Particles (12587) were interactively picked from micrographs displayed in the MRC program, Ximdisp (Smith 1999), at reduced sampling (factor $\frac{1}{2}$). The files produced by this program containing the particle coordinates were then used by SPIDER to window these particles, placing them in 80×80 image boxes which were twice the particle diameter to avoid the influence of edge effects (ringing in the Fourier transforms) produced by the sharp edges of the box in the cross-correlation calculations during alignment.

4.2.3.3 Preprocessing

Filtering - Images were band-pass filtered according to standard single particle procedures to reduce the influence of noise during the alignment procedure. The high-pass filter function applied was an inverse top-hat function with the filter radius set to exclude frequencies greater than the particle diameter (+20%), i.e. $1/200 \text{ \AA}$. The low-pass function used a Gaussian high-frequency fall-off set to suppress frequencies greater than $1/15 \text{ \AA}$.

Masking - Masking was used to decrease the influence of neighbouring particles in the alignment process which resulted from the high particle density on the grid. It was not possible to achieve good particle separation by varying dilution of adsorption time without obtaining preparations that were too sparse. The radius of the circular mask (25 pixels) was larger than the longest particle dimension (21 pixels) and a Gaussian fall off masking function was used to avoid the above mentioned edge effects.

Normalization - Images were normalized after masking to zero average density and a constant standard deviation of one. Normalization of images, to achieve similar background grey levels and particle contrast, is necessary to prevent differences in these values affecting the cross-correlations based alignment procedure. Projections from a single specimen should have the same mean density but different variances, assuming a homogenous sample. It is, therefore, more correct to set the background average to zero and the background variance to one. However, it has been shown that

this difference does not have an effect on reconstruction quality (Sorzano, de la Fraga, et al. 2004). Normalization was performed after masking to ensure that the mean density was only contributed from single particles.

Images displaying dust or inconsistent grey levels were detected from visual inspection of all particles in WEB (Frank, Radermacher, et al. 1996) and deleted at this point.

4.2.3.4 2D Reconstruction

A 2D-average was generated from 150 pinwheel pre-processed views by reference-free alignment (Penczek, Radermacher, et al. 1992). The average was further 6-fold rotationally symmeterized about its centre.

4.2.3.5 3D Reconstruction

The iterative angular refinement (Penczek, Radermacher, et al. 1992) reconstruction strategy employed, was divided into three steps: reference-based alignment (Joyeux & Penczek 2002) to classify and align particle images (12587) to a starting model (see below), calculation of a new reference volume from the class averages by back-projection, and projection of the new reference volume to generate templates for the following round of alignment. This procedure was iterated until the reconstruction converged on a stable volume. Details are given below.

Starting models:

Two independent 3D reconstructions were calculated by angular refinement from alternate starting models:

(1) *Ab initio* model derived by simultaneous minimization based on common lines

Particle images (12587) were aligned together and centred by iterative reference-free alignment (Penczek, Radermacher, et al. 1992). Rotationally invariant K-means

clustering (Penczek, Zhu, et al. 1996) was then used to classify the aligned particles into 100 classes. Images were then recentred according to the centre of gravity of their class average, and reclassified. After 3 rounds, iterative multi-reference alignment was then used to refine class membership by matching and aligning images to the most similar of the 100 class templates, recalculating the averages, and repeating the alignment. ‘Trap classes’ i.e. characteristic views of contaminating particles (putative GroEL) and pinwheels were included in an attempt to exclude these particles from the reconstruction. Classes were inspected and any averages that did not adequately represent the class members were deleted. The angular relationships between the 86 remaining class averages were determined by a simultaneous minimization program based on common lines (Penczek, Zhu, et al. 1996) and a 3D model produced by backprojection (see below). Class averages were used instead of individual images because of their higher signal-to-noise ratios. The choice of 100 classes was based on balancing the number of redundant class averages with the homogeneity and signal-to-noise ratio of the averages (determined visually).

(2) Model based on crystal structure of GSI.

An atomic resolution structure of GSI (Yamashita, Almassy, et al. 1989), low-pass filtered to 1/15 Å resolution, was used as the other starting model.

Multi- refinement alignment

Multi-reference alignment was used to classify and align the pre-processed particle images by comparison (cross-correlation) with templates, derived by projection of, in the first round, these starting models and in subsequent rounds, the previous reconstructions. Initially the symmetry of the large particles was unknown and the model volumes were projected in quasi-evenly spaced directions covering 3D space. The sampling of the projection directions was chosen to adequately represent the theoretically attainable resolution. The formula used to calculate the number of projections required to obtain a given resolution was first derived for a single-axis tilt series reconstruction by Crowther *et al.* (Crowther, Amos, et al. 1970) and later

generalized for all projections directions by *van Heel et al.* (van Heel & Harauz 1986):

$$\text{number of projections} = \text{Diameter of the particle} \times 2 / \text{desired resolution}$$

Since the multi-reference algorithm employed did not use mirror check it was necessary for the projections to cover the entire surface of the angular distribution sphere instead of only a hemisphere. Thus, 166 templates at 15° increments represents an over-sampling by $166 / (165 \times 2 / 15) / 2 \approx 4$, to achieve 1/15 Å resolution. Furthermore, if the fact that symmetry related pixels do not represent independent signal (Orlova, Dube, et al. 1997) is considered, then the number of templates required is reduced even further by the symmetry factor, in this case 6. Additionally, 4 ‘trap’ classes composed of 2 images of the pinwheel view of GSIII (from the simultaneous minimization based on common lines class averages) and 2 images of GroEL (from the sampling determination) were included in the template library. In each alignment cycle, the same untransformed pre-processed images were aligned to the reference volume i.e. transformations did not accumulate during reconstruction, necessitating the use of large translational search parameters. These projections were pre-processed as described for the raw images. During the alignment procedure, a threshold cross-correlation score was imposed to exclude 20% of the particles from the reconstruction. Furthermore, those images requiring translational shifts greater than 8 pixels were also excluded. Images were aligned by translation and rotation according to the parameters determined during multi-reference alignment. Class averages were then calculated and empty classes were replaced by their template images.

Reconstruction in the absence of circularity

Two approaches were attempted to prove the reliability of the reconstruction procedure and the validity of the final model. Firstly, reference-free alignment was performed within each class after each round of multi-reference alignment. After convergence of the iterative reference-free algorithm (Penczek, Radermacher, et al. 1992) the 2D average was aligned by reference-based alignment to the original

projection template used to classify the images. These 2D class averages were then used for the reconstruction. Secondly, aligned images within each class were further classified by K-means rotationally invariant clustering after each round of multi-reference alignment. The number of groups was based on the size of each class with 200 images being sorted into 10 classes. After classification, class averages were calculated for each of the sub-classes. The sub-class average best matching the original template projection was then determined by multi-reference alignment and the 2D average transformed by the given parameters. These 2D class averages were then used for the reconstruction.

These strategies were tested on the final model derived by angular refinement (as described above). For sake of time these reconstructions also utilized several time saving modifications not found in the previously described angular refinement procedure. (1) After each round of alignment transformed particles were stored for use in the following round of alignment. This enabled the search radius to be decreased as the alignment proceeded. (2) Only projections spanning the asymmetric unit were generated. (3) A mirror-check was initiated in the multi-reference alignment algorithm so that only projection directions spanning a hemisphere i.e. $\theta = 0-90^\circ$ were required instead of twice the number.

3D Reconstruction

Backprojection of these class averages, using interpolation in Fourier space, was used to reconstruct the volumes.

C6 symmetry was enforced during reconstructions. This was determined initially from a self-rotation function about the putative symmetry axis (Z-axis) of an unsymmetrized reconstruction. The reconstruction was projected along its long axis and the projection was compared to itself by a 2D rotational search algorithm in SPIDER (Frank, Radermacher, et al. 1996). The six highest cross-correlation peaks were found to correspond to multiples of 60° confirming the C6 symmetry (Table 4.1).

In the final evaluation of the model (see below) it was decided to reduce the projection directions to the asymmetric unit only and employ a mirror-check to speed up processing.

Convergence

Convergence of the algorithm was measured by monitoring two statistics, namely, the average cross-correlation score between all images and the templates derived from the reference volume, and the number of images jumping to different classes between alignment cycles.

Resolution criteria

The resolution of the reconstructions was determined by the Fourier shell correlation (FSC) (Harauz & van Heel 1986) and Differential Phase residual (DPR) methods (Penczek, Grassucci, et al. 1994). The image data set was divided into two and independent reconstructions were calculated from each set using the angular refinement procedure outlined above with one set using the GSI-based model and the other using the *ab initio* model. After convergence, the structures were aligned and their resolution compared by the above methods. As a conservative estimate, the resolution was taken as the point where the FSC between data half-sets equalled 0.5, which is equivalent to a phase error of 35° (Rosenthal & Henderson 2003). 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.

Visualization

Reconstructed volumes were visualized using Web (Frank, Radermacher, et al. 1996), SpiderViewer (Burford, Sewell, et al. 2004), and Chimera (Pettersen, Goddard, et al. 2004). The threshold for contouring the density was decided based upon the volume enclosed by the surface as determined by integrating the voxel histogram in SpiderViewer. This was determined from the predicted molecular weight of GSIII

(82.7KDa) (DNAMAN 1994), the estimated number of subunits, and an average value for protein density of $0.73 \text{ Da}/\text{\AA}^3$.

4.2.3.6 Docking

Initially, an entire ring (6 subunits) from the crystal structure of *S. typhimurium* GSI (without any alteration to the subunit spacing) (Yamashita, Almasy, et al. 1989) was manually docked into the larger ring of the GSIII density.

In order to improve the fitting, an individual subunit of GSI was docked automatically using a fast Fourier transform (FFT) accelerated, correlation based, docking in SITUS (Wriggers, Milligan, et al. 1999) to perform the 6 dimensional search required to find the best docking of the high resolution model to the low resolution density. Rigid-body docking was performed with 15° angular sampling of the search space and subsequent optimization of the top 6 results. To speed up the docking, only a sub-region, corresponding to a radial span of 90° of the larger ring and two-thirds of the length of the molecule in the axial direction (cutting off a portion of the smaller ring) was searched.

Individual dockings were not performed for each homology model because they did not explicitly model inserted density. Instead, the automatically docked GSI subunit was used to evaluate the models by annotating the positions of insertions and deletions for each alternate alignment.

4.3 RESULTS

4.3.1 Visualization

Initial inspection of the grids under the optimised conditions and in regions of high stain depth, revealed a homogenous population of roughly spherical particles with a diameter of $\sim 160 \text{ \AA}$ (Figure 3.3B). Putative GroEL molecules were also observed in the final preparation but only as a very minor fraction. It was, however, discovered that in lightly stained regions, two distinct particle morphologies existed. In addition

to the spherical particles, pinwheel-shaped particles with a similar diameter but clear left-handedness, 6-fold symmetry, and a well stained central channel were observed (black arrows in Figure 4.1). These pinwheels made up roughly 1.2% of the total particles. In Figure 4.1, a low-pass filtered micrograph (not used in the 3D reconstruction) is shown, which highlights the differences between the larger more heavily stained particles and the lighter end-on pinwheels. Furthermore, at very low stain levels, positive staining effects, revealed what appeared to be views of a double-ringed structure for the larger particle (see white arrows in Figure 4.1). Together with the observation that these particles exhibited all the characteristic views seen for other double ringed GSs (Chapter 1) and the fact that smaller hexamers were present, this finding led to the putative interpretation of the larger particles as double-ringed dodecamers. Two separate reconstructions were calculated from the particle population: a 2D average of the pinwheel face-on views and a 3D reconstruction of the larger particle.

An example of one of the 160 micrographs used for the reconstructions is shown in Figure 4.2. The final depth of stain decided upon for data collection was not as low as shown in Figure 4.1 but was low-enough to reveal the pinwheels. Double-carbon negative-staining preparations were attempted but due to the thickness of the EM grid bars, the useable area was too small and a large degree of particle flattening occurred near the centre of each grid square (data not shown).

4.3.2 Sampling

The final sampling in the micrographs was 4.25 Å/pixel, as determined by the scaling of a 2D average of GroEL particles against an atomic resolution structure of the particle (Figure 4.3). The predicted value, based on a 14µm pixel and a magnification 50000, was 5.6 Å/pixel (taking 2-fold binning into account). This represents a magnification error of 24%.

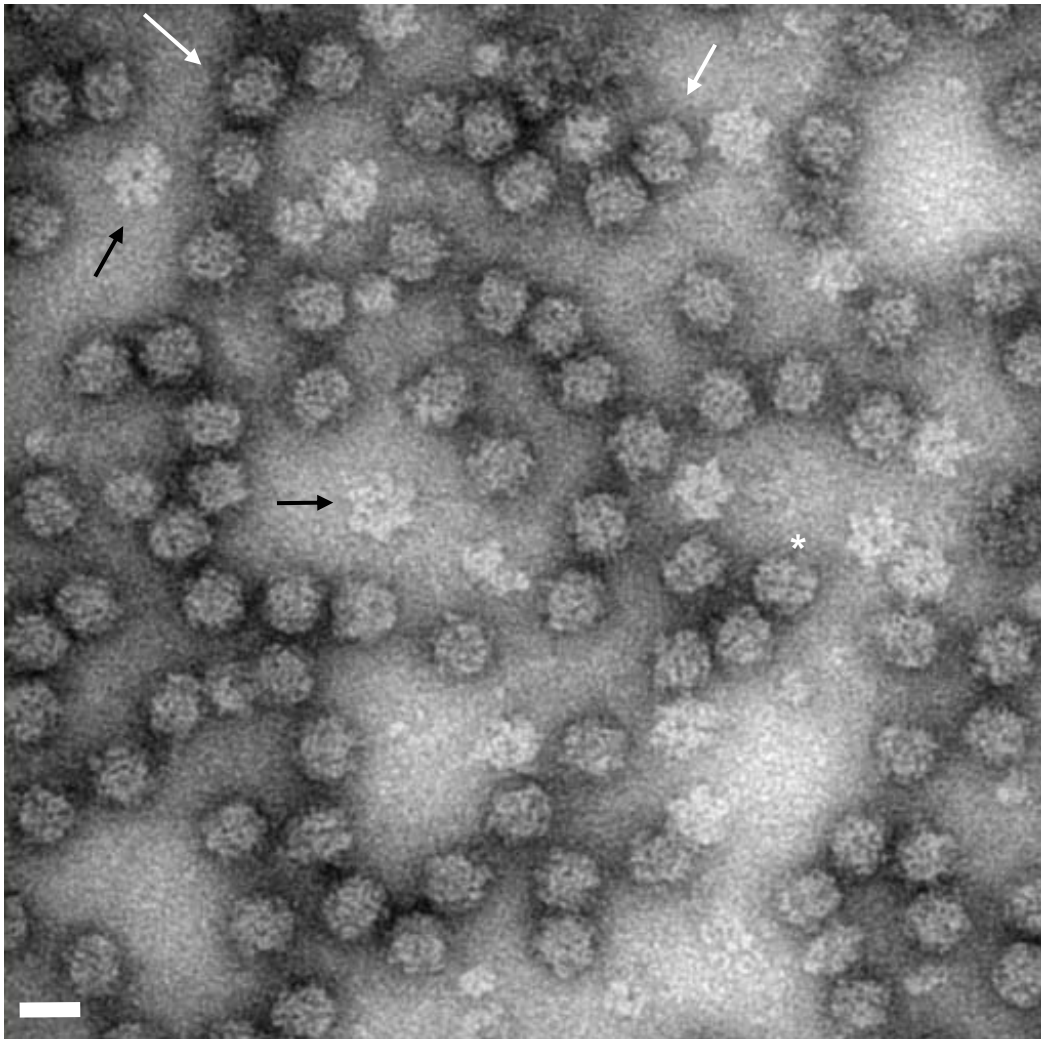


Figure 4.1 Electron micrograph (low-pass filtered to 20 Å) of negative stain (2% UA) preparation of GlnA. Particles marked with black arrows are thought to represent single hexamers (from dissociation or partial staining) in comparison to the better contrasted double-ringed dodecamers (examples marked with white arrows). Scale bar = 18 nm. The particle marked by an asterisk is referred to later in the text.

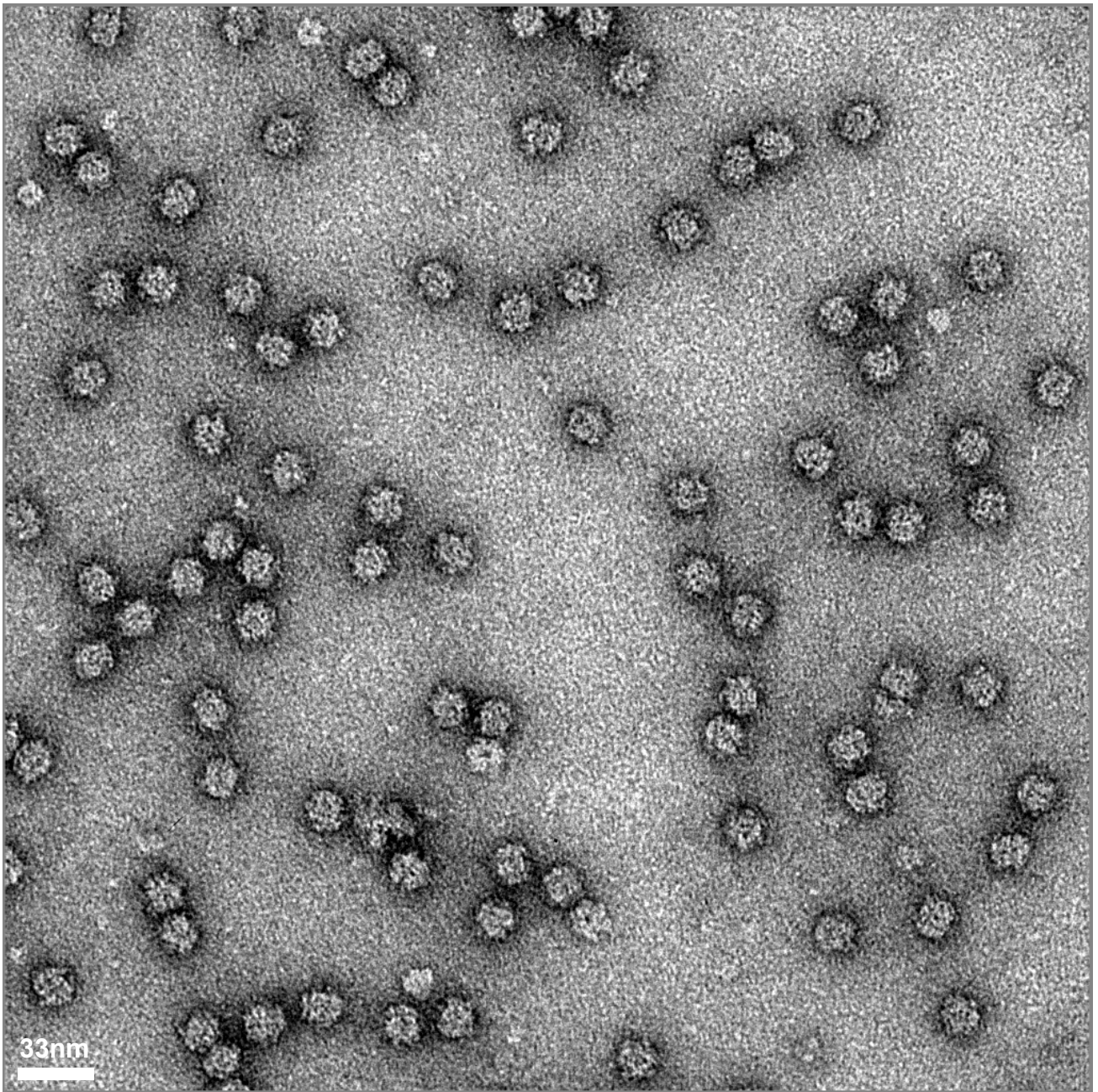


Figure 4.2. Example micrograph of GlnA single particles stained with uranyl acetate (2%) used in the reconstruction. The image has been reduced from the original sampling of 4.25 Å/pixel and the grey levels have been adjusted for presentation. The sample was prepared and imaged as described in chapter 3.

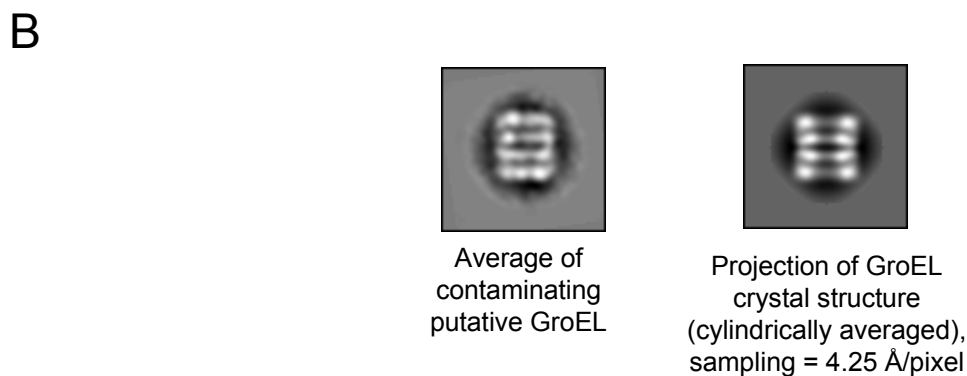
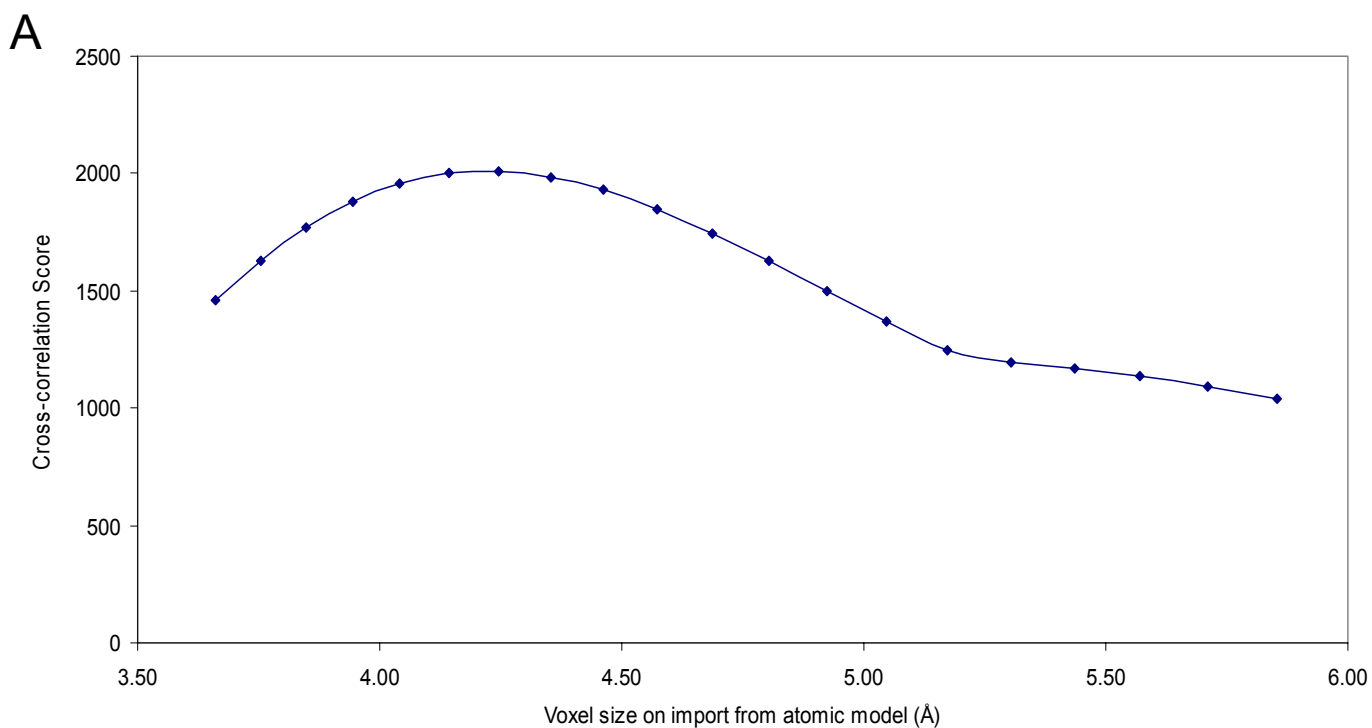


Figure 4.3. Determination of final micrograph sampling by comparison of an average of all contaminating putative GroEL side views against differentially scaled models of an atomic structure (Braig, Otwinowski, et al. 1994). **A** – Plot of the CC score between the 2D average and cylindrically averaged projections of the atomic model as a function of sampling. **B** – Comparison of the 2D average and the projection displaying the highest CC score

4.3.3 2D average of pinwheel views

The pinwheel views were all assumed to lie in the same orientation on the grid permitting a simple 2D average to be calculated by reference-free alignment (Penczek, Radermacher, et al. 1992). This is a valid assumption considering that for a spherical particle of diameter 180 Å, 1/19 Å resolution is still possible for tilts as large as 12° from the normal (Crowther, Amos, et al. 1970).

The 2D-average generated from 150 pinwheel views (Figure 4.4A) with 6-fold symmetry enforced is shown in Figure 4.4B. The stain distribution of the 2D average revealed a particle with an overall similar structure to the previously reported 2D-averages of partially stained *E. coli* GSI (Kessel, Frank, et al. 1980) (Figures 4.4C & 4.4D), which only contained information about the structure of the ring closest to the carbon (for a population of face-on views). However, the structures differed in several regards. Firstly, the diameter of GlnA pinwheel is 180 Å, 30% longer than that of the GSI 2D-average (137 Å). Secondly, both proteins possess prominent ellipsoidal stain excluding regions at the periphery of the hexameric rings (61 vs. 60 Å radially from the ring's centre for GSI and GlnA respectively), which are bilobed being divided into two subdomains. These are the “oblate ellipsoids of revolution” first seen by Valentine *et al.* investigating GSI (Valentine, Shapiro, et al. 1968). Thirdly, the dimensions of the major axes of these ellipsoid domains were measured to be 42.5 and 60 Å for GSI and GlnA respectively. The minor axes were measured to be 24 and 30 Å GSI and GlnA respectively. The former axis of these ellipsoidal domains is skewed with respect to the normal of the molecule radius by 6° in GSI and 25° in GlnA. The ellipsoidal domains are in turn linked by a smaller stain excluding spiral arms to the central ring region. Thus, the predominant difference between the structures is the larger size of the ellipsoidal domains, and the greater degree of rotation of these domains relative to the normal of the particle radius (Figures 4.4C & 4.4D).

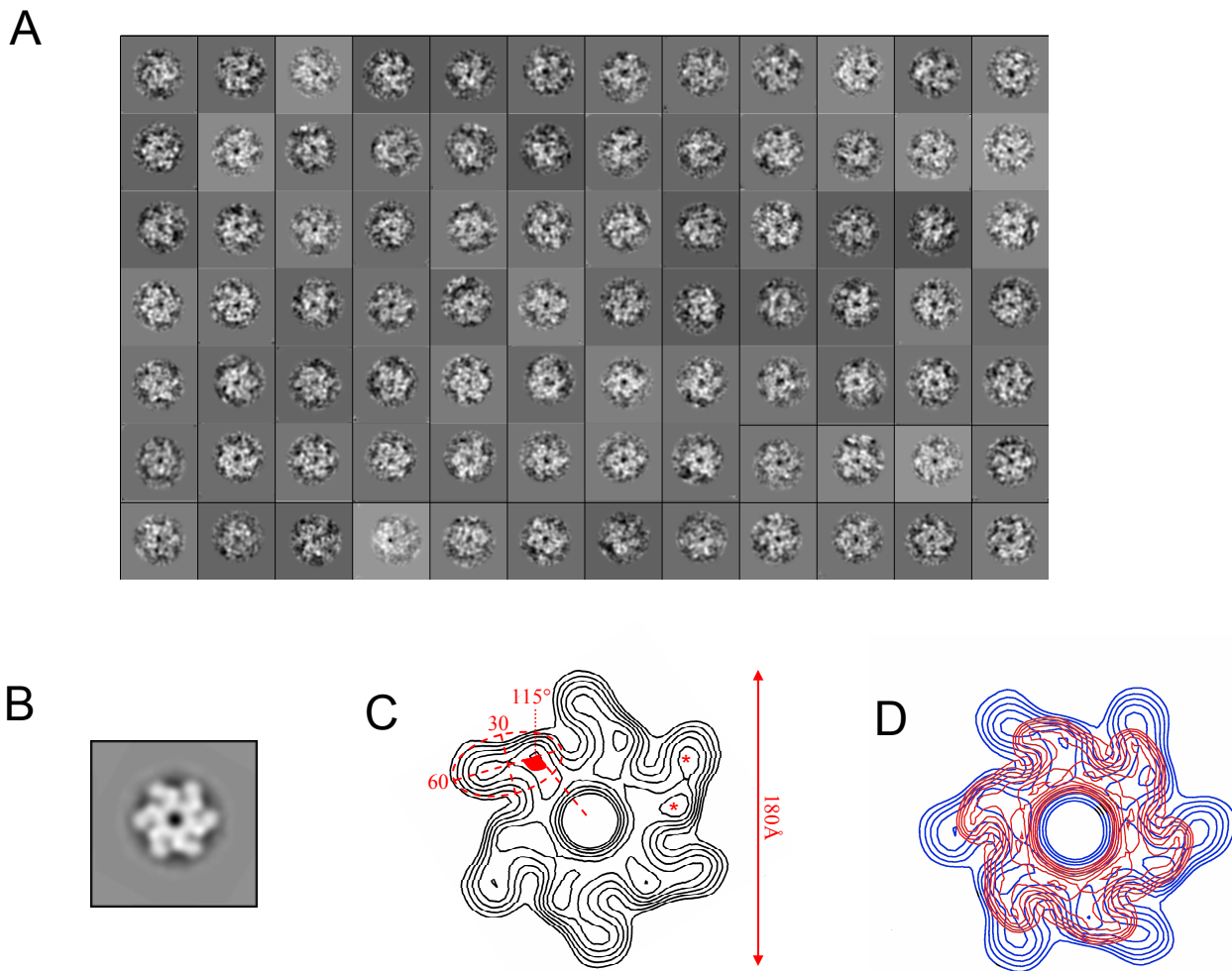


Figure 4.4. 2D average of putative hexameric pinwheel form. **A** – Pre-processed images of face-on hexamer, aligned by iterative reference-free alignment, averaged, and 6-fold rotationally symmetrized to produce the image shown in **B**. **C** - Contoured representation of **B** (-3.24 to 2.48 in steps of 0.95) with annotated measurements of the ellipsoidal domains shown in red. The positions of the two density peaks comprising the bilobed domain are marked with asterisks. **D** – overlay of a projection of one ring of the GSI crystal structure (shown in red) with the 2D-average of GlnA (shown in blue).

4.3.4 3D Reconstruction

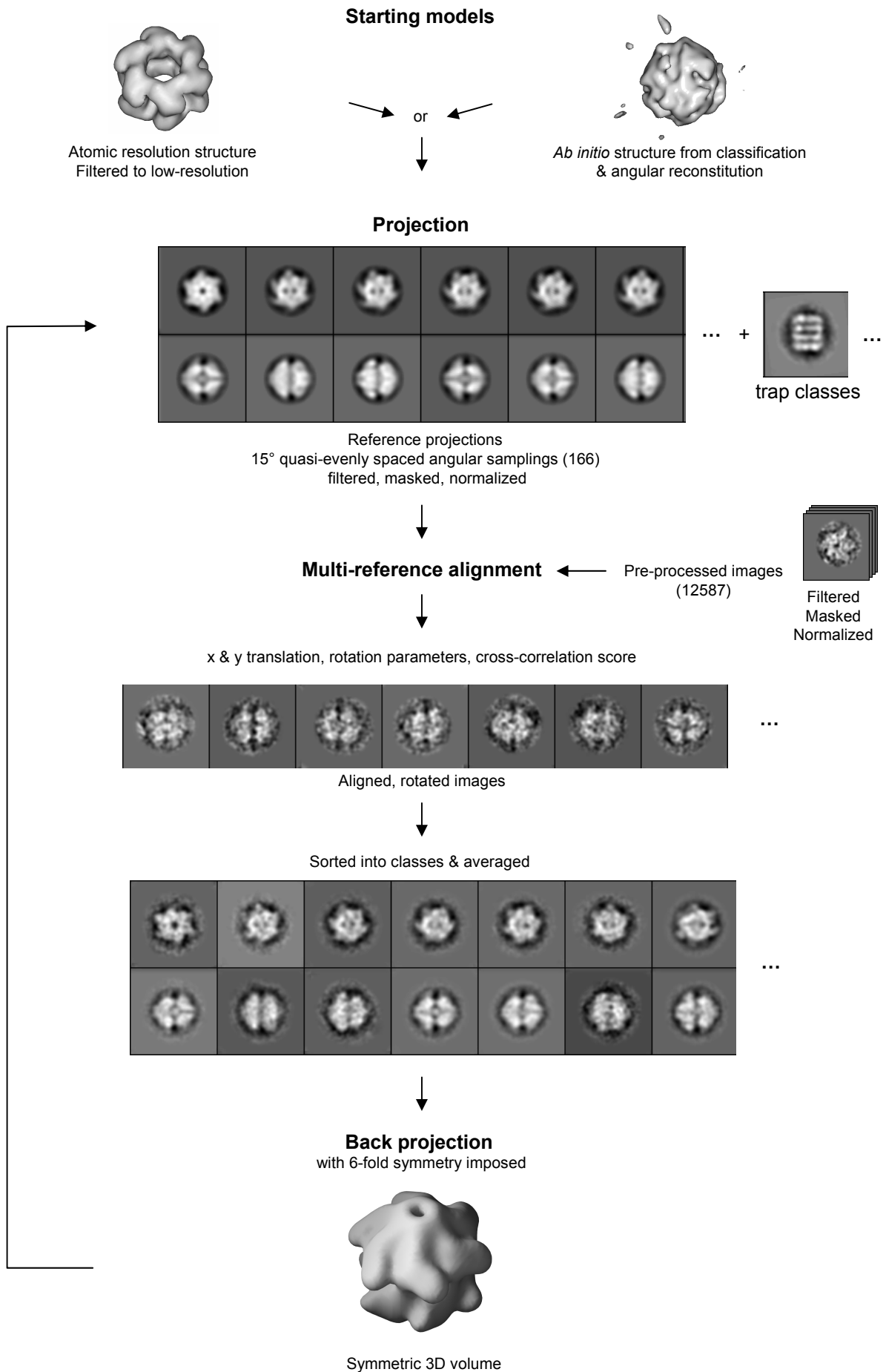
Unlike the pinwheel views, the large molecules appeared to lie in numerous different orientations on the support film, making these structures amenable to 3D reconstruction by single particle techniques. Angular refinement, based upon two independent starting models, was used to determine the angular relationships between all projections such that they could then be backprojected to form a 3D volume for GlnA (Figure 4.5). One starting model was derived from the atomic structure of GSI and the other was derived *ab initio* by classification and common-lines based simultaneous minimization. The 89 class averages, derived by the latter approach, together with the resulting model are shown in Figure 4.6. Preliminary reconstructions in the absence of symmetry displayed clear cyclical 6-fold symmetry and self-rotation functions about the long axis of the particle confirmed this (Table 4.1). This symmetry was, therefore, enforced during the reconstruction procedures. Examples of the reference projections and corresponding class averages are shown in Figure 4.7. The reason for calculating two independent reconstructions from different models (as determined by FSC) was to avoid the problem of extension of resolution due to the correlation of Fourier coefficients in adjacent shells that results from the finite boundedness of particles (Yang, Yu, et al. 2003).

Table 4.1: Detection of C6 symmetry by self-rotation function about the putative symmetry axis (Z).

Peak number*	Peak value	Rotational angle	Multiple of 60°
1	0.29633	0	0
2	0.26449	61.989	1
3	0.26449	298.01	5
4	0.26066	235.19	4
5	0.26066	124.81	2
6	0.25594	180	3

*only the six highest peaks are reported

Both reconstructions, converged to similar structures (judged by FSC, see below) as shown in Figure 4.8A. Figure 4.8B shows the convergence of the reconstruction based



...(from previous page)

Figure 4.5. Outline of the angular refinement procedure (details are given in section 4.2.3.5). Two alternate starting volumes were used: the crystal structure of GSI filtered to low resolution and a model derived by common-lines based simultaneous minimization of class averages from rotationally invariant K-mean classification. Reference templates (166) were generated by projecting the starting model volumes at evenly spaced 15° samplings in 3D space. ‘Trap classes’ i.e. characteristic views of contaminating particles (putative GroEL and the single hexamer) were included in an attempt to exclude these incorrectly picked particles from the reconstruction. Pre-processed images (12587), floated from the original micrographs, were then aligned to the reference templates. Multi-reference alignment yielded the three orientation parameters: x-translation, y-translation and in-plane rotation angle used to align all the images and a correlation score between the images and references. Based upon the latter parameter, the images were sorted into 171 different classes and averaged. These averaged classes (excluding the trap classes) were then used to reconstruct the volume by back projection in Fourier space, with cyclical six-fold symmetry imposed during the reconstruction. The resulting symmetric volume was then used as the reference volume for the next round of reconstruction. The cycle was iterated until the reconstruction algorithm converged, as determined by the average CC score.

A



B

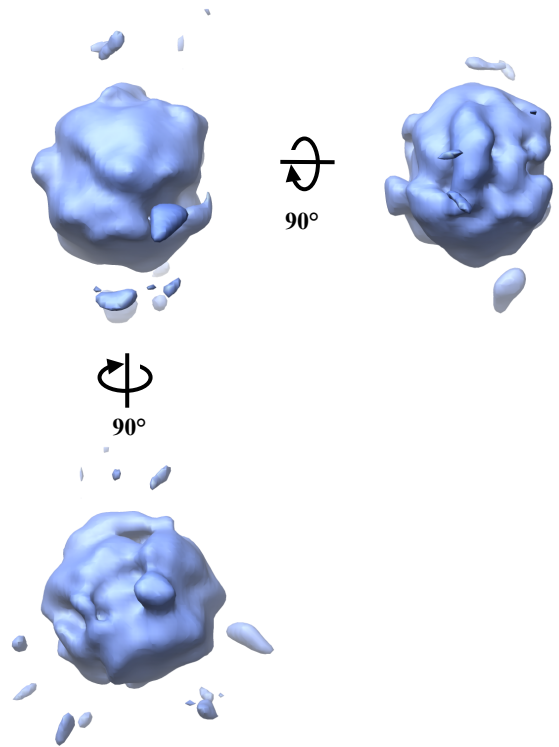
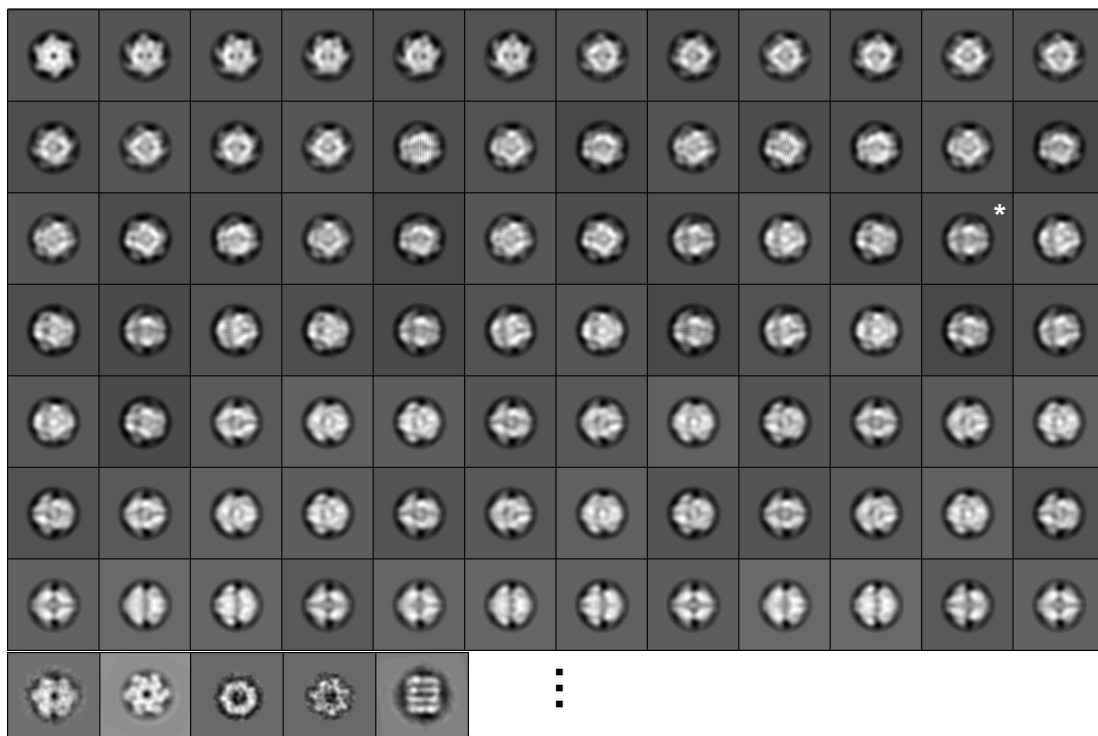


Figure 4.6. Summary of the generation of an *ab initio* starting model by angular reconstitution of class averages of rotationally invariant K-mean classified images (details are given in the text). **A** – class averages and ‘trap classes’ (final 3) subjected to angular reconstitution to generate the model shown in **B**.

A



B

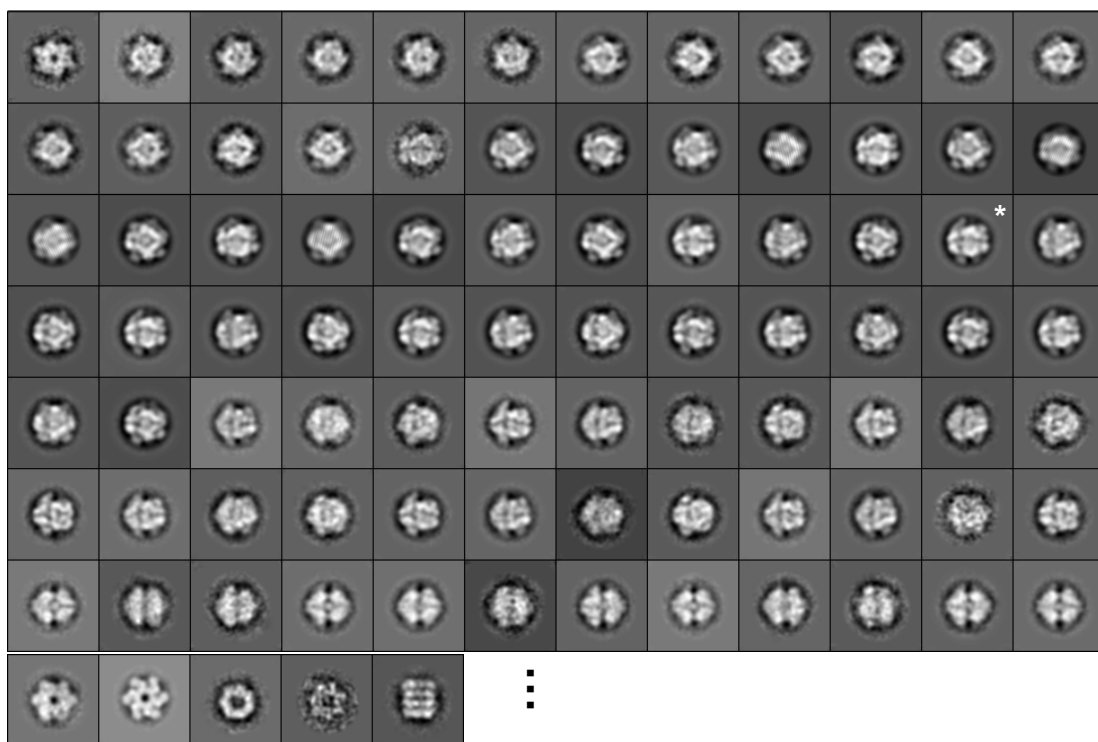


Figure 4.7. Examples of the projections and corresponding class averages used in the reconstruction based on the *ab initio* starting model (final round). **A** – Subset of the 166 template projections at quasi-evenly spaced 15° increments. ‘Trap classes’ are shown in the last row. **B** – Subset of the corresponding 166 class averages. The success of the reconstruction can be directly gauged from a comparison of **B** with **A**. The template and class average marked by asterisks are referred to later in the text.

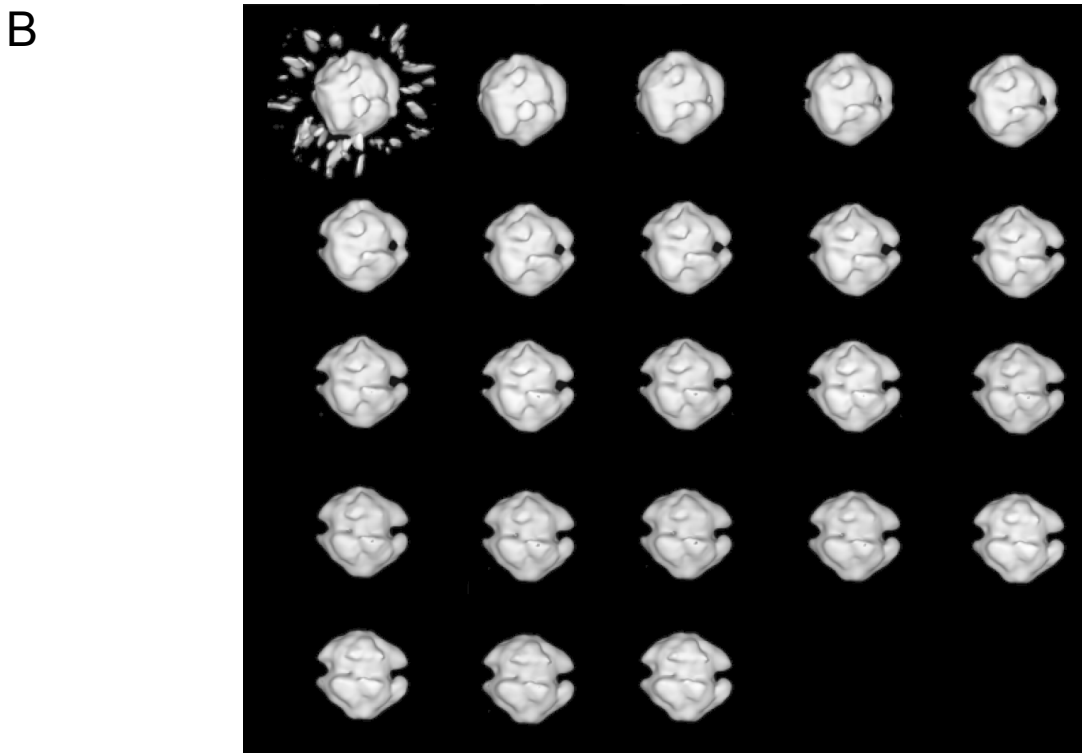
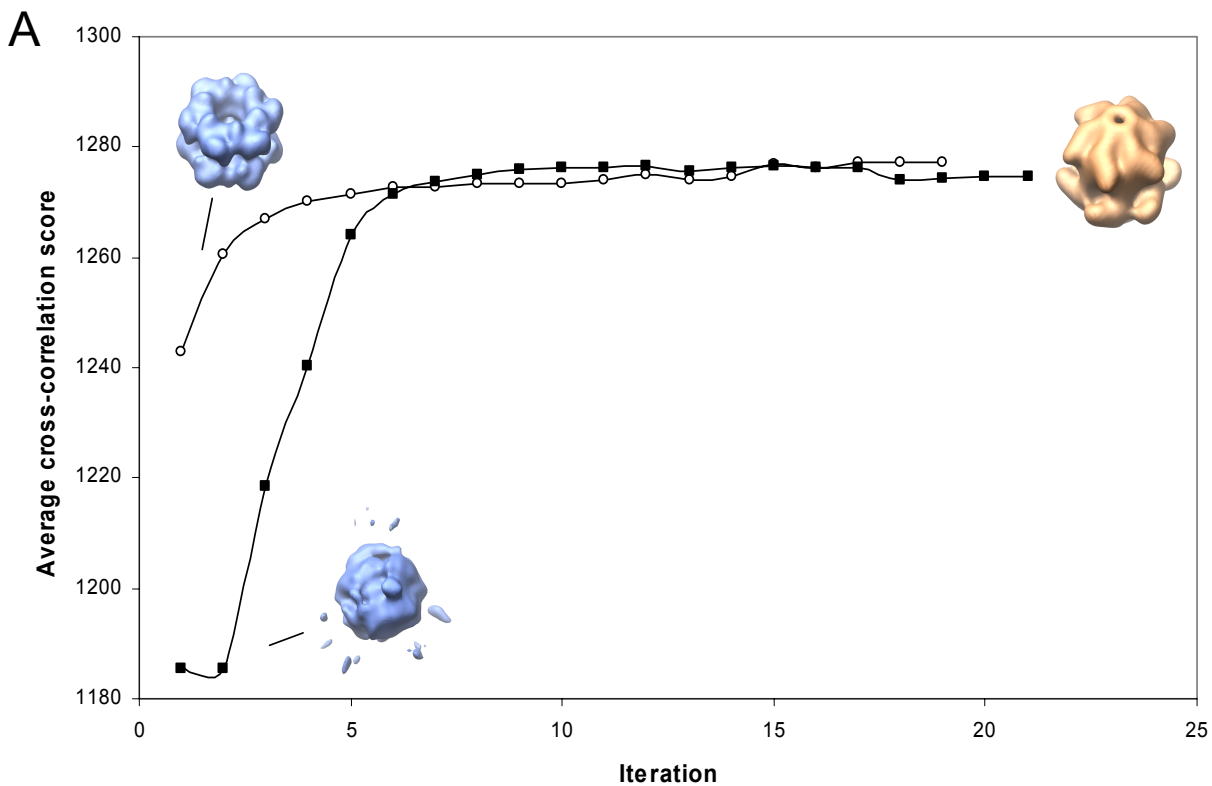


Figure 4.8. Reproducibility of the reconstruction. **A** – Convergence of the two independent reconstructions, calculated from data half-sets and each starting from a different reference volume, as determined by the average CC score. *Ab initio* model (solid squares), GSI based model (open circles). **B** – Surface representations of the volume reconstructed from the *ab initio* model at incremental stages of the reconstruction. No symmetry was enforced during for this reconstruction, unlike in **A**.

on the *ab initio* model in the absence of symmetry. Alternate views of the molecular envelope of GlnA as determined by the reconstructions are shown in Figure 4.9 and discussed below. Although the *ab initio* reconstitution did converge to the same structure as the GSI based reconstruction (Figure 4.8A), the angular distribution of the class averages did not agree with the angular assignments determined by comparing the same class averages with projections of the final model using multi-reference alignment (Figure 4.10).

4.3.4.1 Structural insights

In contrast to preliminary EM observations, (Southern 1986) and the results of the 2D averaging discussed above, which both identified GlnA as a hexamer, the 3D reconstruction revealed a large dodecameric complex with average dimensions 16.5×16.0 nm. The subunits (82.8KDa) of the complex are arranged in two opposing hexagonal rings with distinct handedness, each resembling the 2D pinwheel average discussed above (Figure 4.9). The two rings are assumed to be face-to-face forming a closed space group, otherwise the oligomer would not be limited to only two rings. This is similar to the quaternary structure arrangements of the GSIs and GlnTs except that the GSIII complex is 60% longer (16.5 vs. 10.3nm) in the axial direction and on average 12% wider than GSI (16.0 vs. 14.3 nm) (see Figure 4.11). The two rings are, however, only pseudo-symmetrically related, differing not only in diameter (16.5 or 15.5nm vs. 14.3nm) but also the degree of separation of subunits (Figure 4.9). As such, the particle possesses only cyclic 6-fold symmetry in comparison to the dihedral symmetry of GSI (Figure 4.11) and consequently projections of the complex in the axial direction possess handedness. As discussed earlier, projections of a molecule with dihedral symmetry, in a direction normal to any of the 2-fold axes, should possess mirror symmetry. The complex also possesses a large central cavity, which tapers in diameter towards the ends of the complex in the axial directions. GSI on the other hand has a channel of constant diameter extending throughout the axial direction (Figure 4.11).

At the contour level chosen, so as to enclose the volume corresponding to 12 subunits, subunit contacts are not well resolved. At higher thresholds however, these become

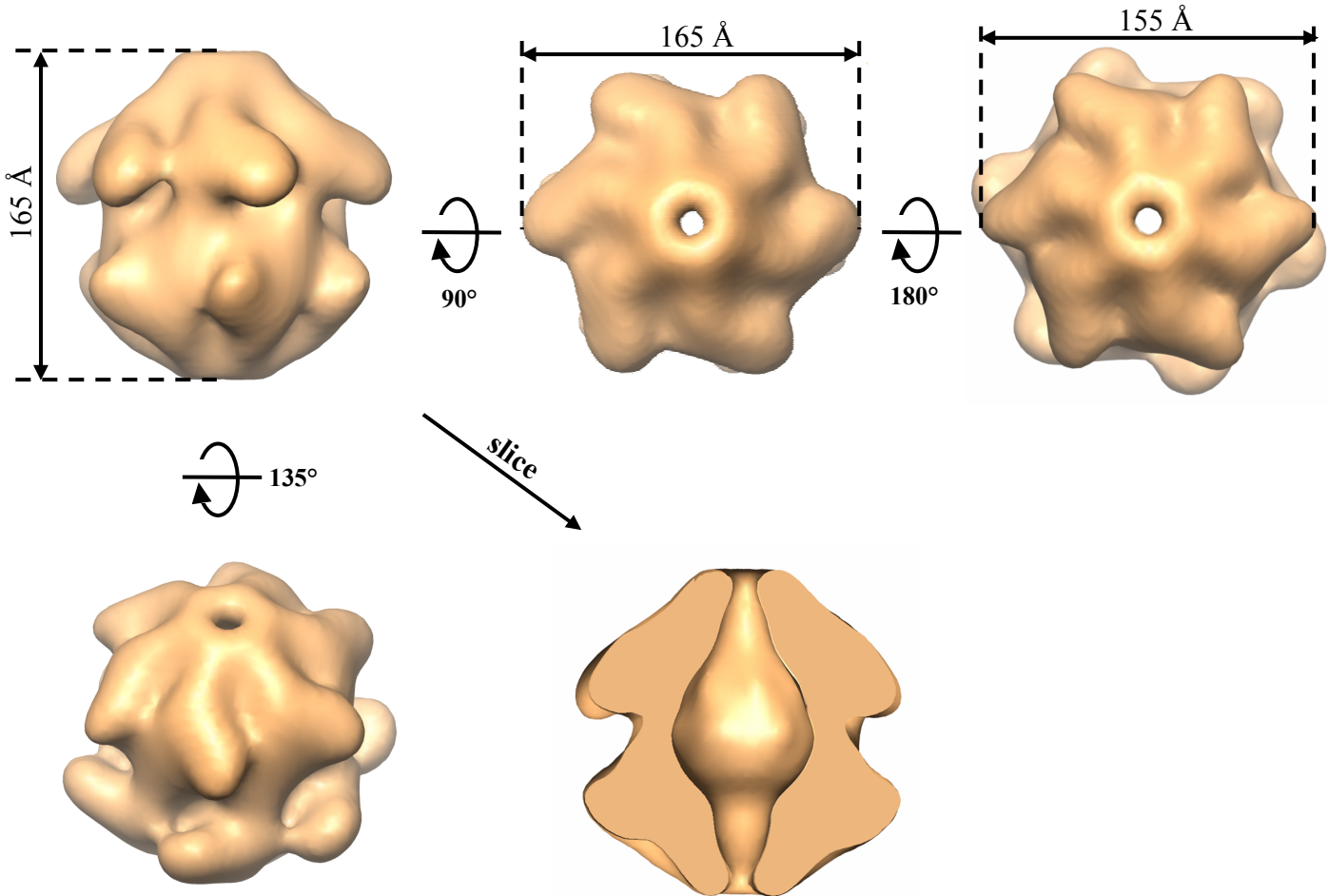


Figure 4.9. Alternate views of the molecular envelope of GlnA. The reconstructed volume is shown as a surface representation with the density contoured to enclose 1360 nm^3 . GlnA is a dodecameric complex, with subunits (82.7 KDa) arranged in two opposing hexagonal rings showing distinct handedness. The two rings are not symmetrically related, differing not only in diameter but also the degree of separation of subunits and as such the particle possesses only cyclic symmetry. There is a large central cavity which tapers in diameter towards the poles of the molecule and a central sphere of density resulting from averaging and not present in models made from fewer images has been excluded.

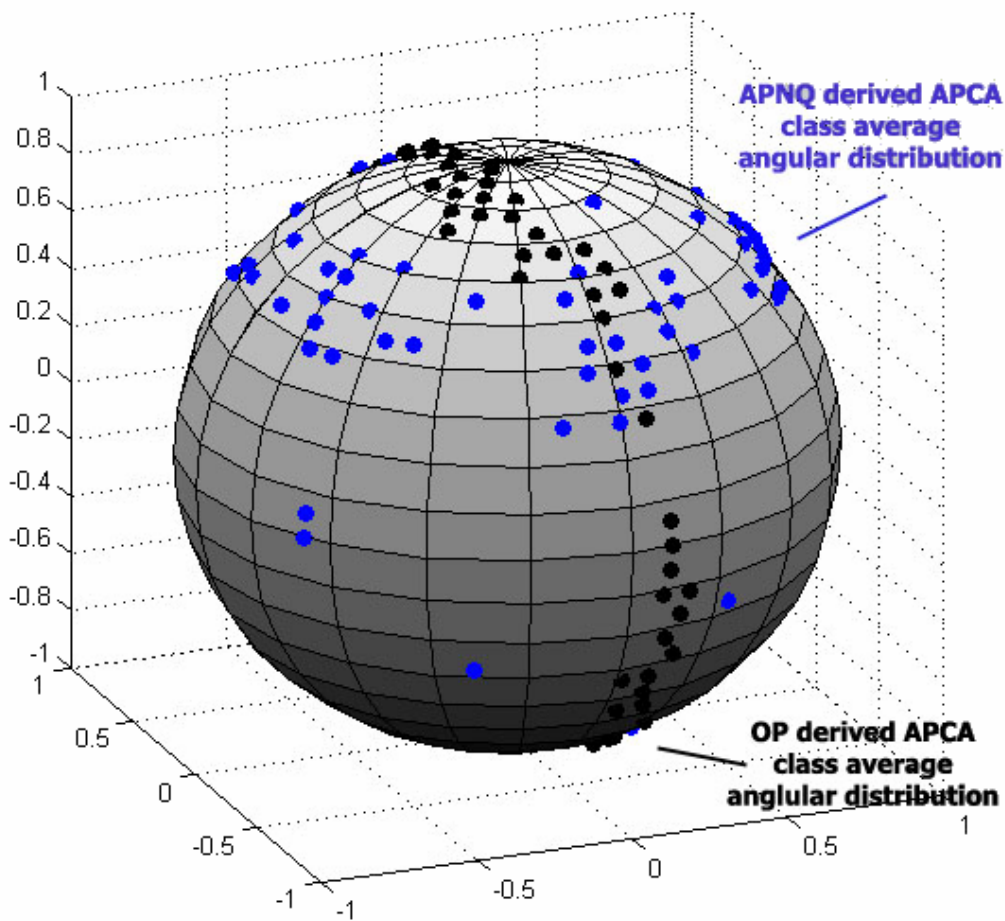


Figure 4.10. Comparison of the angular distributions of the class averages in **A** as determined by angular reconstitution (black) and by multi-reference alignment (blue) against the final model. Points on the surface of the sphere represent the viewing direction, with longitudinal lines representing θ and latitudinal lines, Φ . It can be seen that according to the multi-reference alignment, the angular refinement failed to correctly determine the angular relationship between the class averages.

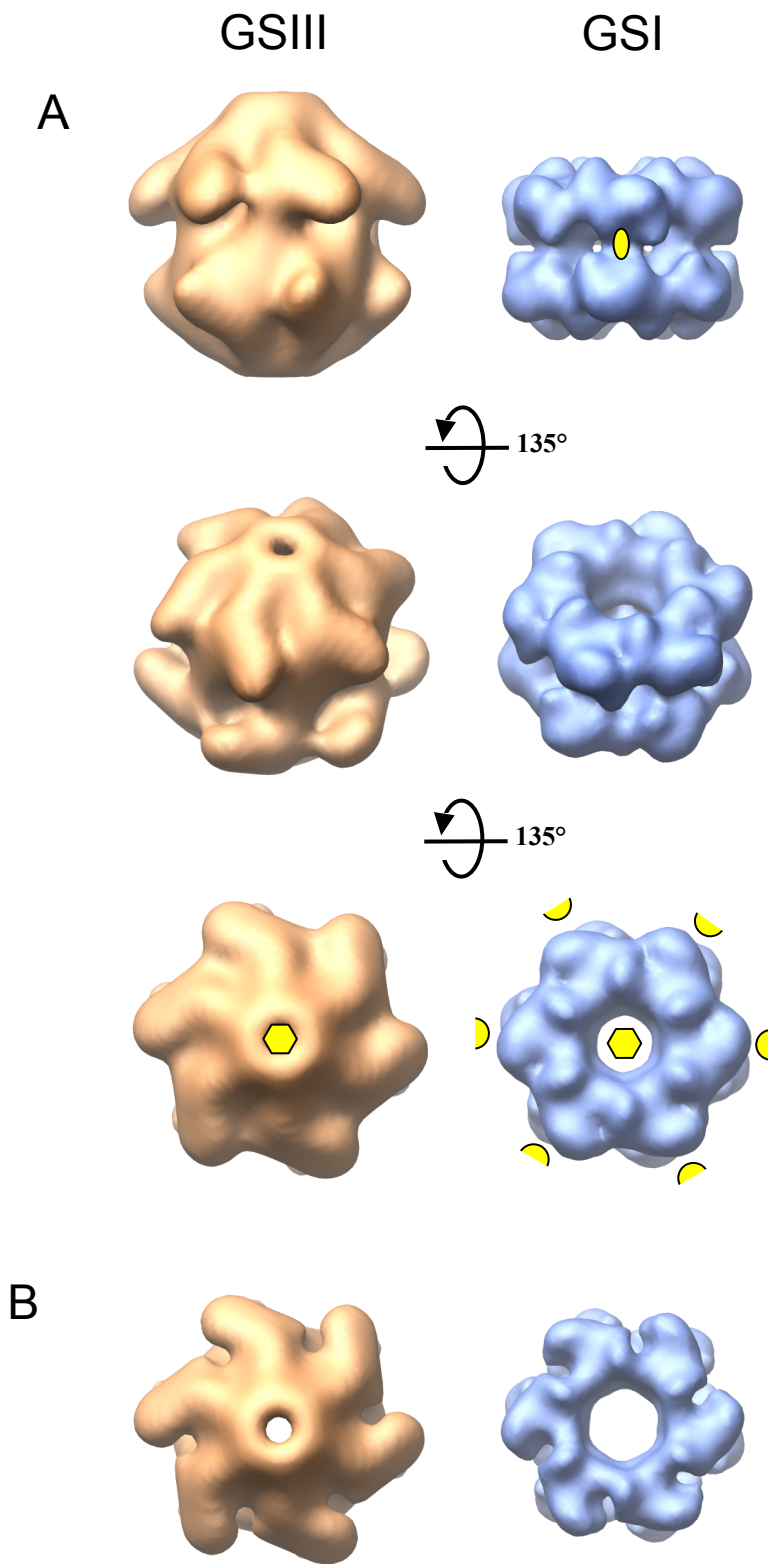


Figure 4.11. Comparison between the structure of GSIII and GSI (from the crystal structure Yamashita, Almasy, *et al.* 1989). **A** - Both structures were low-pass filtered to 21Å and are contoured to enclose the volume calculated for 12 subunits of: GSIII 82.7 KDa & GSI 51.6 KDa. Symmetry elements are shown in yellow: ovals = 2-fold axes and hexagons = 6-fold axes. It can be seen that despite both being dodecamers, GSIII has cyclical 6-fold symmetry whereas GSI has dihedral 6-fold symmetry. GSI is also 60% longer and on average 12% wider than GSIII (143 Å in diameter 103 Å along the 6-fold axis). The central cavity of GSIII is larger than that of GSI but tapers to a smaller diameter at the top and bottom of the molecule. **B** - Same as **A** but contoured at higher density threshold to emphasise handedness of each ring. GSI is left-handed as determined by crystallography but the handedness of GSIII is not known. 120

more defined (Figure 4.11). Like the pinwheel averages, each ring has semi-circular indentations between subunits. GSI has a similar configuration and these indentations correspond to the $\alpha\beta$ -barrel active sites formed by the heterologous associations between adjacent subunit. The inter-ring contacts (presumably isologous) are more extended than in GSI with six stalk-like processes extending below the arms of each subunit. In the larger ring, a further region of continuous density can be seen branching off from these stalks and joining with the lower region of the density surrounding the invaginations between subunits (data not shown).

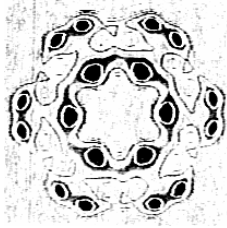
Figure 4.12 compares the two structures of the GSIII, GlnA, determined in this work with the structures of the other GSs reported previously. It is obvious from this comparison that GlnA represents the largest GS visualized to date with a molecular mass close to a mega-Dalton. As mentioned above, the dodecameric structure of the complex is similar to GSI and GlnT as reflected by the similarity of their structures determined by negative stain EM. It must also be noted that differences do exist between the 2D average of the pinwheel and the 3D reconstruction. The pinwheel is larger (18.0nm), even compared to the bigger ring of the 3D reconstruction (16.5nm) and the central channel appears to be of constant diameter as judged by the density of contour lines.

4.3.4.2 Statistics

Class distribution

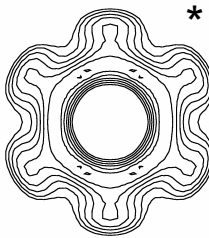
The classes were not equally populated (Figure 4.13A). Classes 2-50 possessed far greater numbers of member images than all the others. When the angular assignments of the projections were investigated (at higher angular sampling of 5° increments – Figure 4.13B) it was revealed that this skewed distribution resulted from a tendency of the particle to adopt a preferred orientation. In particular, inspection of the histogram of the number of images as a function of theta angle (as defined in the Euler coordinate system of SPIDER (Frank, Radermacher, et al. 1996) and corresponding to the tilt of the particle about an axis normal to the projection direction) revealed the most abundant orientation to be theta 45° and to a lesser extent

GSI



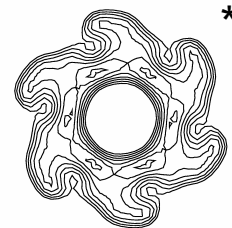
S. typhimurium

(Kunath, Weiss, et al. 1984)



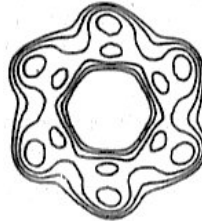
S. typhimurium

(Yamashita, Almassy, et al. 1989)



S. typhimurium
- one ring

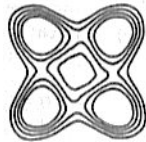
GInT



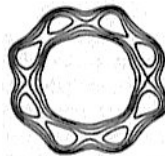
Rhizobium lupini

(Tsuprun, Zograf, et al. 1987)

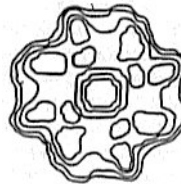
GSII



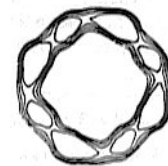
Root nodule cytosol



Pea leaf cytosol



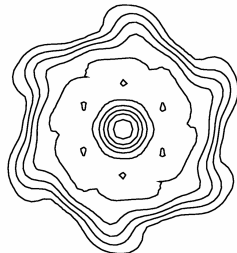
Pea seed cytosol



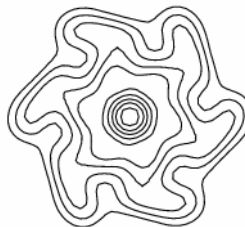
Bacteroid

(Tsuprun, Zograf, et al. 1987)

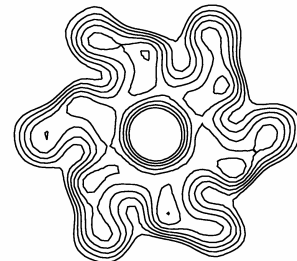
GSIII
(this work)



B. fragilis GlnA



B. fragilis GlnA
- larger ring



Pinwheel average

10nm

Figure 4.12. Comparison of face-on views of GlnA from this work with images from previous electron microscopy studies of GSs. Images are to scale and those marked with asterisks are projections of a low-pass filtered model generated from atomic resolution crystal structures of GSI. All others are from negative stain work.

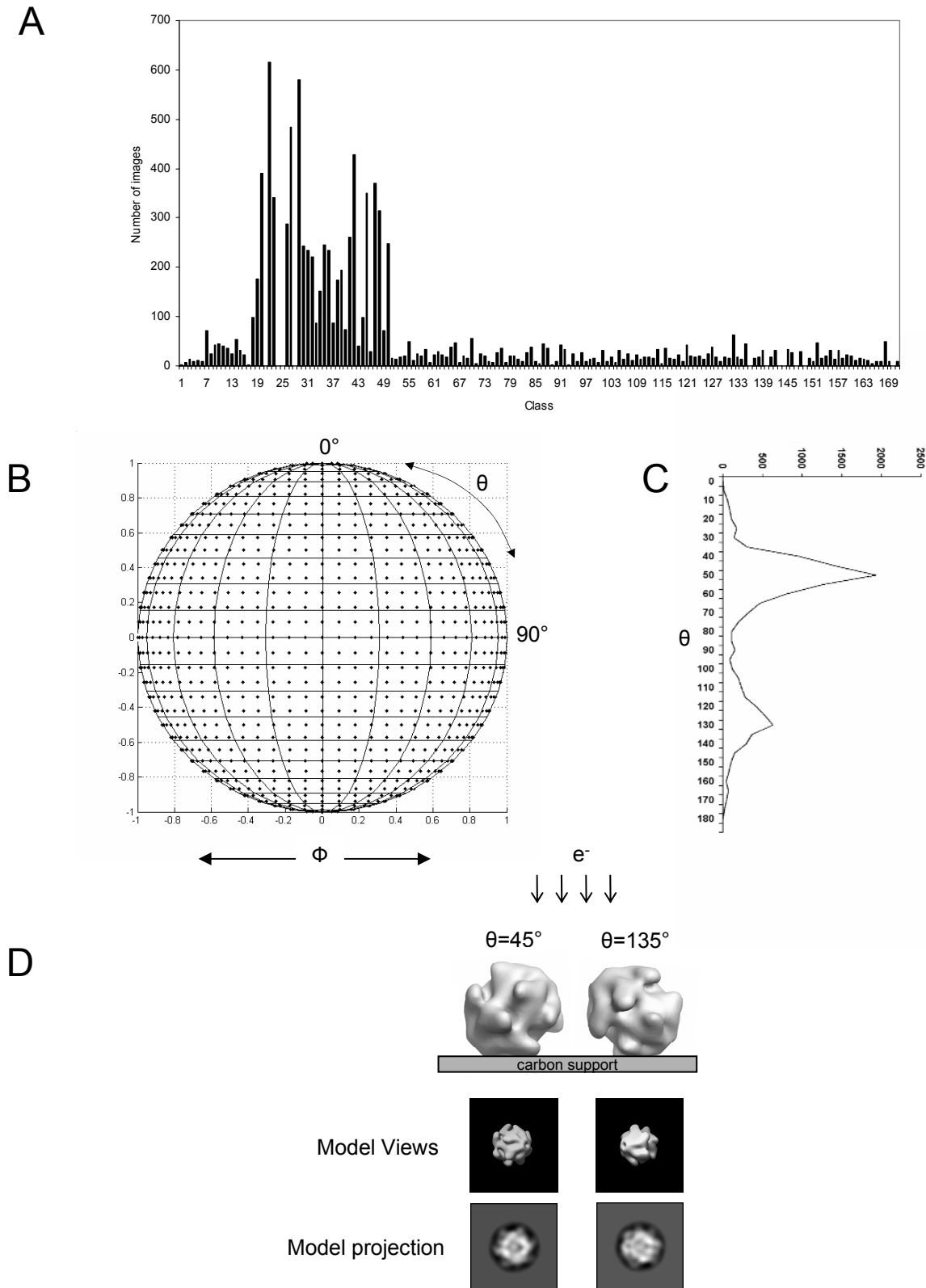


Figure 4.13. Preferential orientation of single particles. **A** – Histogram of class membership for final reconstruction. **B** – Spherical plot of projection angular distribution showing the even sampling of 3D space at 5° increments. **C** – The corresponding histogram of the number of images as a function of theta angle reveals a the most abundant orientation to be theta 45° and to a lesser extent theta 135° . This explains the abundance of views belonging to classes 20-50 in **A**. **D** – Explanation for the preferred orientation. Transformation of the reconstructed volume according to these preferred orientations demonstrates that the particles lie with their largest flat surface in contact with the support film. Views of the models in these preferred orientations are shown below, together with their corresponding class averages.

theta 135° (Figure 4.13C). These orientations correspond to the particles lying with their largest flat surface in contact with the support film (Figure 4.13D).

Unidirectional platinum shadowing was attempted to determine the absolute handedness of the particles but the results were inconclusive.

Resolution and quality of the reconstruction

Reproducibility and resolution - Reconstructions calculated from completely different starting models (as measured by FSC, data not shown) and independent datasets, converged to the same structure (Figure 4.8A). These reconstructions were equivalent to 21 Å resolution as measured by FSC with the conservative 0.5 FSC measure (Figure 4.14) and 20 Å according to the 0.33 measure for full datasets (Rosenthal & Henderson 2003).

Self-consistency - The validity of the reconstruction can be directly gauged from the similarity of the characteristic views, generated by multi-reference alignment and averaging, to reprojections of the model. A summary of the reconstruction process is shown in Figure 4.15, demonstrating how the model can account for the characteristic views observed in the micrographs. Even views that at first appear to belong to different particles can be accounted for. For instance, the particle marked with an asterisk in Figure 4.1 appears rectangular but is in fact a tilted view of the GlnA particle as can be seen in Figure 4.7A & 4.7B (asterisk).

To avoid the possibility of circularity, reference-free and classification based strategies were implemented during multi-reference alignment (section 4.2.3.5). Figure 4.16A shows the convergence of the final model to a similar structure after reference-free alignment is initiated between multi-reference alignment and back-projection, in the original angular refinement procedure. It can be seen from Figure 4.16B that, with the exception of the underrepresented face-on view (most likely due to the trap classes), the reference-free class averages are all very similar to the reference-aligned class averages. Further classification of the classes determined by

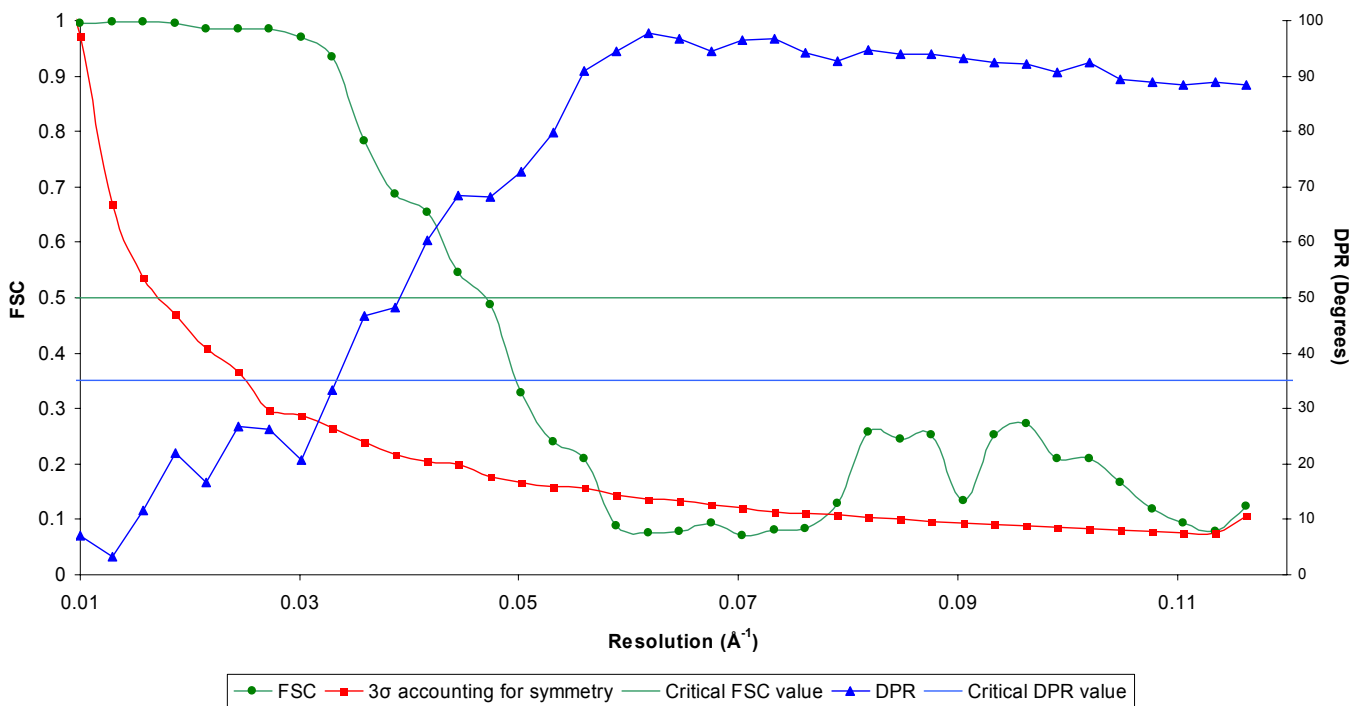


Figure 4.14. Resolution of the reconstruction. Fourier shell correlation and differential phase residual obtained from comparison of the final reconstructions derived from different starting models using separate halves of the image dataset. The FSC falls below 0.5 at a resolution of 21 \AA . The DPR shows the structures to be in agreement to only 30 \AA according to the conservative 35° criterion.

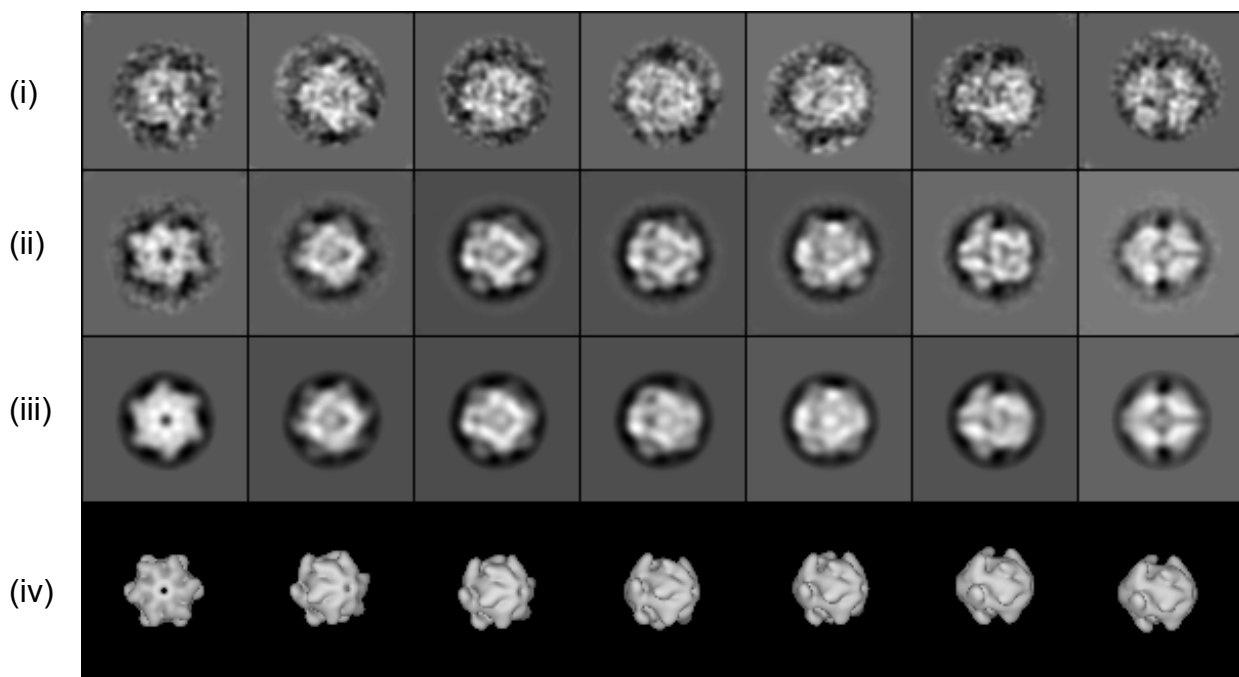


Figure 4.15. Summary of the reconstruction process. (i) Examples of pre-processed (filtered, masked, and normalized) images belonging to the classes averages shown in (ii). (iii) Projections of the 3D reconstruction in the directions corresponding to the class average views. (iv) Surface representations of the 3D reconstruction, contoured to enclose 1360 nm^3 , and orientated in the same direction as (iii).

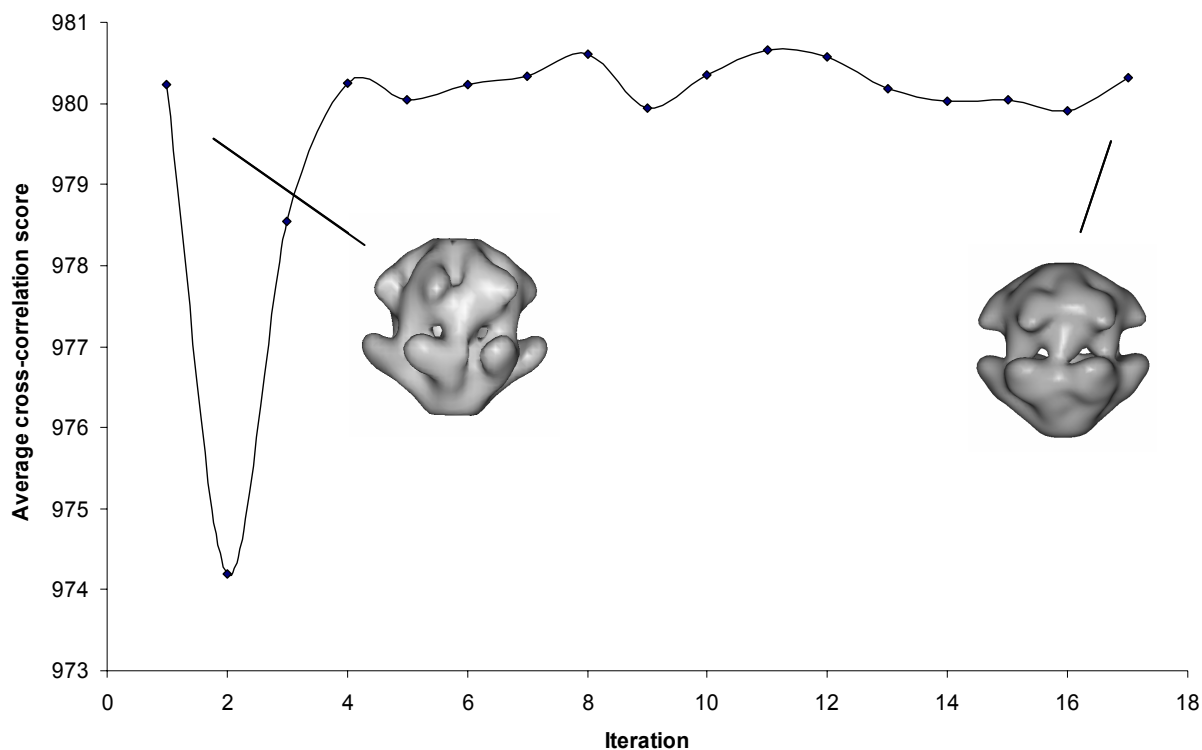
A**B**

Figure 4.16. Reconstruction to avoid circularity. **A** – The final, converged, model derived previously by angular refinement (shown on the left) was subjected to further rounds of refinement as before but with reference-free alignment performed within each class after multi-reference alignment to break the circularity between the class averages and the templates. **B** – Class averages as produced by multi-reference alignment alone (i) and by reference-free alignment within each class (ii). With the exception of the first class, the reference-free class averages are all very similar to the template-aligned class averages and therefore there is little change in the 3D reconstruction as shown in **A**.

multi-reference alignment into subclasses also yielded a similarly stable result (data not shown).

To address concerns raised about orthogonal misalignment of the side views displaying 4-fold symmetry (see last column of Figure 4.15) (personal communication, Orlova, E.V., 2004), the reference volume was masked by a cylinder with a length shorter by about a third than that of the reference and a possessing a central cylindrical channel. In spite of the masking, the reconstruction converged to a similar structure as that shown in Figure 4.9 but sharply truncated towards the poles of the volume's long axis.

4.3.5 Docking

In order to try interpret the reconstructed volume and to evaluate the homology models proposed in Chapter 2, these were docked into the density of the molecular envelope derived by negative stain EM described above. The smaller ring was not considered because its reduced size is thought be an artefact of the reconstruction (discussed below). As shown by Figure 4.17, the docking of one ring from the crystal structure of GSI into the molecular envelope of GSIII again suggests the similarity of the overall arrangement of these two structures but at the same time revealed their large differences in volume. The automated docking of a single GSI subunit into the density achieved a similar result but improved the agreement between the model and the map (inset, Figure 4.17). In particular, the coincidence between the GSI α/β barrel fold, which forms the active site together with the N-terminal 'latch' from an adjacent subunit, and the indentations between the arms of the pinwheel was clear.

The evaluation of the homology models discussed in Chapter 2, according to this docking of the GSI subunit structure, revealed large volumes of empty density near the N and C termini of the models in agreement with the positions of the major insertions in all but one (the manually aligned model 2) of the alignments of the GlnA sequence relative to GSI (blue and red spheres in Figure 4.18). The C terminal extensions do not, however, have all of the density in this region available for modelling the added residues because this space will also be occupied by the C

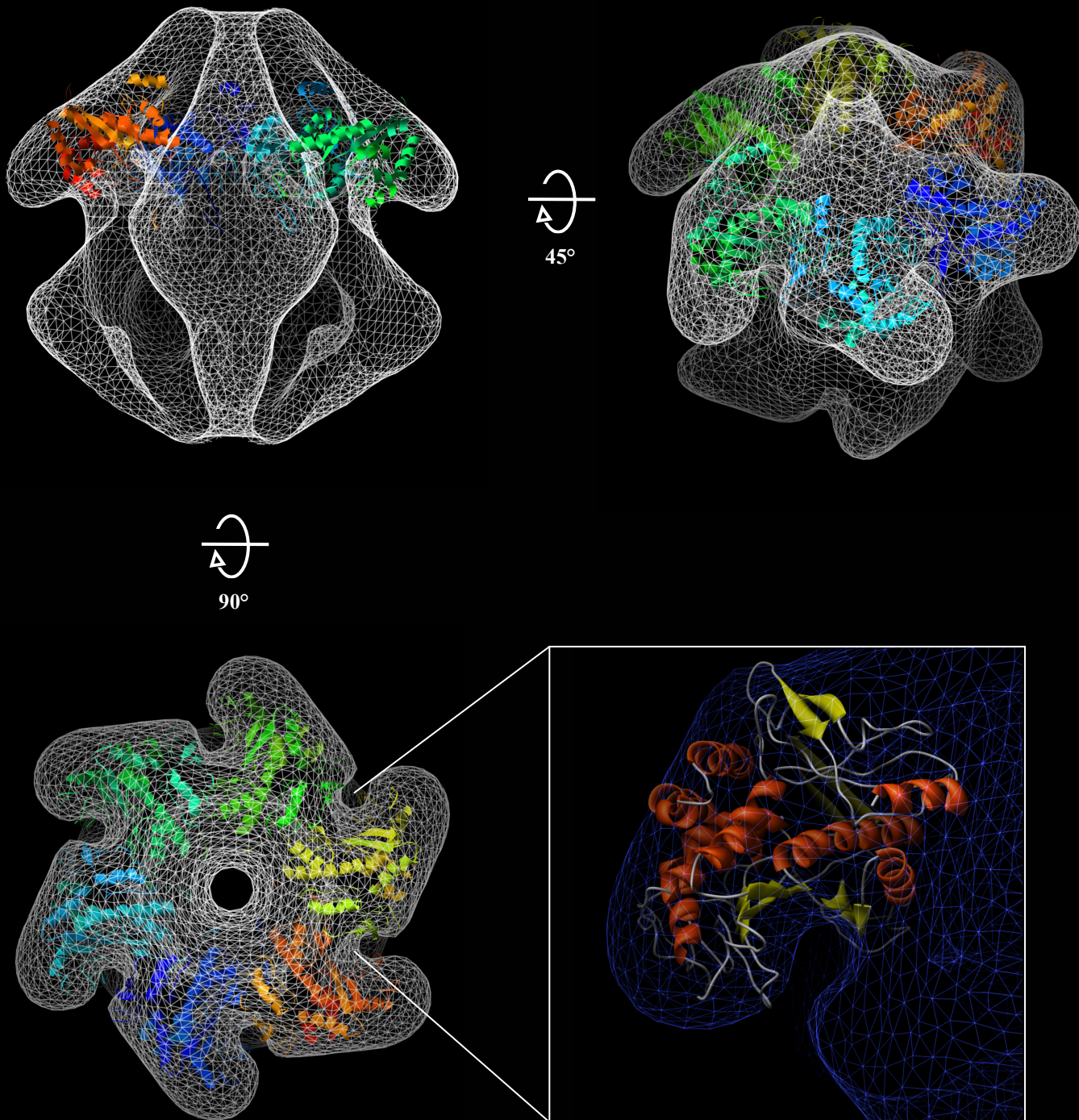


Figure 4.17. Docking of GSI crystal structure into the molecular envelope of GSIII. The three white wire frame volumes show the docking of an entire ring from the crystal structure of GSI into the molecular envelope of GSIII. Inset – close-up of the docking of one subunit demonstrating the coincidence of the GSI α/β barrel fold, which forms the active site together with the N-terminal ‘latch’ from an adjacent subunit, and the indentation between the spokes of the pinwheel. The contour level was chosen to highlight the similarity of the map to the model.

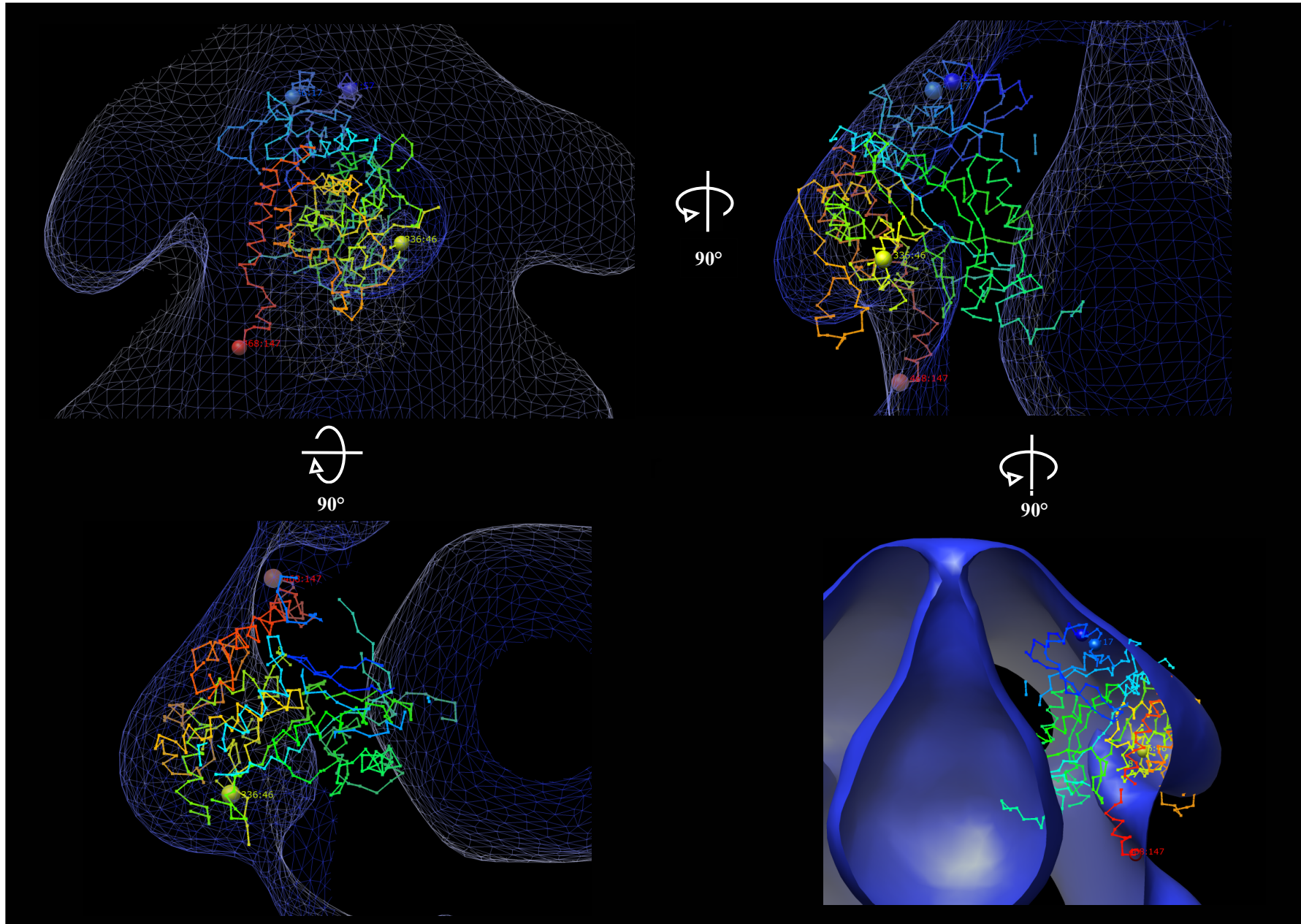


Figure 4.18. Example docking of one of the four alternate homology models, interactive alignment 1, into the molecular envelope of GlnA. All dockings were based on the automated docking of an entire GSI subunit by the program Situs. Regions of the peptide back-bone corresponding to deletions in the alignment between GSIII and GSI are not shown and insertions are marked by text labels (the number before the colon is the residue position and the number after is insertion size in residues). Spheres mark the position of large insertions for easy reference. The volume is contoured to enclose 1360 nm³ (12 subunits of 82.7 KDa MW each).

terminus of the subunit from the opposite ring. The evaluation also revealed several incompatibilities between the density and the models. A view from the inside of the particle reveals the poor fitting of the β -sheets that form the walls of the channel in GSI, to the interior surface of the particle. Most critically, with the exception of the alignment produced by FUGUE (Shi, Blundell, et al. 2001) all the other alignments have large insertions near residues 236-250, which corresponds to a position near the tip of the arms. Based on the current docking, the density does not seem to be able to account for the required insertions of 39-56 residues (yellow spheres in Figure 4.18). Furthermore, in the case of the second manual alignment, an additional 115 amino-acid residue insertion is found at a similar position and needs to be accounted for.

4.4 DISCUSSION

4.4.1 2D average

The pinwheel views were immediately recognizable as GS-like with well defined arms displaying striking 6-fold symmetry, left-handedness, and a clearly defined central hole. This led to the interpretation of these particle as hexamers with C₆ (cyclical 6-fold) symmetry. The possibility of dihedral symmetry was excluded because projections, in the axial direction, of particles with dihedral symmetry (at least one 2-fold axis of symmetry perpendicular to the direction of projection) possess mirror symmetry (Frank, Goldfarb, et al. 1978; Frank, Goldfarb, et al. 1979). Pseudo 6-fold symmetry resulting from staggered stacking of two rings with 3-fold symmetry was also excluded because, on tilting (30°), the pinwheel shortened in diameter (data not shown), suggesting that the particle was a flat disk rather than a cylinder viewed end-on which would be expected to elongate in diameter when tilted (Frank, Goldfarb, et al. 1978). If GlnA does exist in a hexameric form, then this work supports stoichiometric data from biochemical studies, including calibrated gel filtration and native SDS-PAGE analysis, which previously identified GlnA as a hexamer (Southern 1986; Southern, Parker, et al. 1987). Additionally, the only other GSIII purified to date, GlnN from *Synechocystis* sp PCC 6803 was also isolated as a hexamer as determined by gel-filtration (Garcia-Dominguez, Reyes, et al. 1997).

The exclusive appearance of left-handed pinwheels could only be achieved by hexamers with a specific orientation preference. This might result from charge or hydrophobically mediated interactions on one side of the dissociated hexamer interacting with the glow-discharged carbon support film. Determining whether this is the same surface involved in heterologous inter-ring contacts will require the absolute handedness of both particles to be determined. Furthermore, based on the small number of hexamers present in the preparations (1.2%) it appears that under the preparation conditions employed the association-dissociation kinetics lies in favour of the dodecamer.

The handedness and number of subunits could alternatively be explained by partial staining of dodecameric GlnA particles lying normal to the grid, leading to only immersion of one ring in the stain as suggested for 2D-averages of *E. coli* GSI (Frank, Goldfarb, et al. 1978). This would explain the exclusivity of left-handed pinwheels because, no matter which ring of a dihedral complex is partially embedded in stain, the two-fold axes ensure a similar handedness when viewed in projection normal to the grid. However, the distinct appearance of the pinwheels in a field of otherwise homogeneously stained particles suggests that they are actual hexamers.

Regardless of whether the particle is a hexamer or dodecamer, the 2D average contains information about the structure of only a single ring. A comparison with the EM structure of partially stained GSI (thought to correspond to one ring) (Frank, Goldfarb, et al. 1978) revealed the overall similarity between the two structures despite only 9% sequence identity. As expected for a particle of almost double the MW, the subunits were, however, larger and are more extended in arrangement, accounting for the extra mass.

The possibility of stain induced deformation needs to be considered when interpreting negative stain reconstructions. Stain deformation is primarily a problem for particles embedded in low stain depths because the stain is not thick enough to support the particles during drying and those with internal cavities are most susceptible (Frank 1996). Deformations as large as 40% in the direction normal to the grid have been reported for air dried negatively stained preparations (Frank 1996). As such, there is

an uncertainty in the measurements of the hexamer's dimensions, which is not easily estimated. Lightly stained GSI particles from *E. coli*, viewed face-on, have been reported to be as much as 20% larger in diameter than their more heavily stained equivalents (Kessel, Frank, et al. 1980). However, the close agreement (4% error) between diameter of a single GSI ring determined by negative stain EM (13.7nm) (Frank, Goldfarb, et al. 1978) and that determined by crystallography (14.3nm) (Almassy, Janson, et al. 1986), suggests that under the correct conditions little deformation takes place. Furthermore, the resemblance of GlnA to GSI in overall morphology infers that the possible deformation suffered by the particle is likely to be similar in extent to that seen previously for the latter. Scaling errors, due to deformation of negatively stained GroEL reference standard seem unlikely due to the agreement between the average and templates (Figure 4.3). The average matches the template most closely according to the height of the particle and the spacing between the 'stripes' formed by the subunits. These features are less likely to be effected by stain deformation than the particle width which could vary with stain depth and particle flattening.

4.4.2 3D Reconstruction

The validity of a reconstruction can be directly assessed by comparing the characteristic views generated by multi-reference alignment and averaging with reprojections of the model. However, because images are aligned against reference templates, their class averages will always resemble the templates (Yang, Yu, et al. 2003). This problem has been termed circularity (Schatz & van Heel 1990). Circularity is particularly a problem for images with low signal-to-noise ratios and when large numbers of images are used (Grigorieff 2000). Reference-free alignment of images within each class or alternatively further classification within classes can be used to break this circularity (Yang, Yu, et al. 2003). The convergence of independent data sets to the same structure from different starting models confirmed the reproducibility of the reconstruction and indicated the presence of strong signal in the images, as expected for negative stain EM. This was particularly evident in the reconstruction based upon the *ab initio* model, which despite the complete incorrect assignment of the class average angular relationships (6-fold axis was out by 90°),

still converged to the same model as the GSI-based reconstruction. It, therefore, seems unlikely that circularity was a major problem during the reconstruction because, in the presence of real signal and an iterative reconstruction approach, as employed here, artefactual similarity between the images and references, developed in the first alignment cycle, would diverge. In support of this, both attempts to break circularity resulted in a similar structure, albeit with lower resolution. Furthermore, the model could account for all views seen in the micrographs, even those that at first appeared to not be internally consistent with the other measured particle dimensions. This consistency together with the failed attempts to enforce a GSI-like structure by masking annulled fears that misalignment of side views was causing the domed appearance of the particle.

The putative identification of the larger particles as dodecamers based on the appearance of characteristic double-ring views and the visualization of the hexamers was confirmed by the 3D reconstruction. GlnA was revealed to be a large dodecameric complex formed by two opposing hexagonal rings resembling the pinwheels but each differing in diameter. This work, therefore, not only confirms the discovery of hexameric views of GlnA reported previously (Garcia-Dominguez, Reyes, et al. 1997; Southern 1986) but also demonstrates that GlnA is capable of higher order oligomerization to form dodecamers. This dodecameric arrangement of GSIII is similar to the quaternary structure of GSI and GlnT, except that these particles possess dihedral symmetry, despite sharing only 9% sequence identity with GSI compared to the 20% shared between the GSI and GlnT structures. Furthermore, the amino-acid sequence of GlnA is 56% longer than that of GSI, and the expected size difference is clear from a comparison of the two structures. This structure represents, to my knowledge, the largest GS visualized to date in terms of both MW and, consequently, volume. Like the pinwheel average, the rings of the structure are larger than those of GSI but in three dimensions it can be seen that the complex is also much longer in the axial direction accounting for a large proportion of the mass difference.

The apparent asymmetry of the GlnA the particle is not consistent with the findings for other double-ringed GS enzymes, but similar asymmetry has been found for other

multi-ringed structures with otherwise dihedral point group symmetry, such as in the chaperonins GroEL (Bochkareva & Girshovich 1994) and CCT (Llorca, Smyth, et al. 1999). In this case, however, it is most likely to be an artefact of the preparation and reconstruction process. The depth of staining in the micrographs was low as indicated by the sharp (strongly scattering) stain meniscus surrounding each particle. This enhances particle contrast with respect to the background leading to the false impression of better staining but can in fact result in a significant loss of finer detail as a result of partial staining for particles with diameters greater than 150 Å (Frank 1996). A low level of staining was ultimately chosen to allow the pinwheels to be visualized despite attempts at uniform staining by the double carbon method. This low stain depth (relative to the large particle diameter) together with the preferential orientation of the particle suggests that the asymmetry results from partial staining. The reason for this belief is based primarily on the following lines of evidence.

- 1) The preferential orientation of the particle is understandable in terms of its surface topology. The preferred orientations of GlnA correspond to the most energetically favourable orientations for the particle, maximizing contact with the support film. In this preferred orientation, half the arms of the ring furthest away from the carbon film would be partially embedded in stain because of the 45° angle (Figure 4.13D). Therefore, one ring of the reconstructed volume is back-projected from incomplete projections which reduce the overall diameter of the ring when the volume is rotationally symmetrized.
- 2) Projections related by $\theta = x$ and $\theta = 180 - x$, termed flip/flop projections (Frank 1996), are related by mirror symmetry (hence the use of mirror checks in multi-reference alignment requiring fewer templates). This relationship only holds for projections of entire molecules or equivalently complete projections because flip/flop particles truncated from the same direction will be missing different regions of density leading to a violation of the mirror symmetry. It was clear from the theta angle frequency distribution (Figure 4.13B) that such a violation was occurring, because the number of images was skewed towards $\theta = 45^\circ$.
- 3) The possibility of particle deformation due to contact with the support film or stain induced deformations as a result of air drying are not likely because flip/flop

related particles would be equally effected. Equivalently, even a truly asymmetric particle would possess projections identically related by mirror symmetry.

- 4) The difference in diameter between rings is much smaller at low contour levels than at higher levels suggesting that some complete staining did occur.
- 5) The most compelling evidence is lack of asymmetry in the side views (see Figure 4.7B). Potentially, given enough side views, a single-axis (tilt) reconstruction could be calculated about the 6-fold axis to reveal this.

Dihedral symmetry was not imposed during the reconstruction because it would have led to a reduction in the diameter of the large rings due to averaging effects.

A critical point to consider is the difference (8%) in diameter between the 2D average of the hexamer (18.0nm) and the 3D reconstruction of the dodecamer (16.5nm) (Figure 4.12). Because of the uncertainty (~20%) associated with the measurements of the pinwheels's diameter, due to stain deformation from low staining, a meaningful comparison is not possible. If the hexamer's dimensions are assumed to be correct then it is likely that the diameter of the 3D reconstruction is underrepresented. Four possibilities exist to explain this difference. (1) The hexamer could adopt a more extended conformation compared to the dodecamer. The molecular contacts required to form a double-ringed structure would transmit conformational changes to the outer regions of the individual rings leading to a smaller diameter. (2) Stain induced deformation could be responsible. (3) Partial staining (as explained for the smaller ring of the dodecamer) could equally result in a smaller volume. (4) The possibility of flexibility in the outer regions of the pinwheel arms cannot be excluded. This would lead to missing density in the 3D reconstruction process because the technique relies on correlation-based averaging to improve the strength of signal relative to noise. Scaling cannot be a factor because images of both particles originated from the same micrographs. Considering the resemblance of the hexamer to partially stained GSI, point 1 seems unlikely. The complete coverage of 3D space by template projection directions makes systematic stain deformation of the volume unlikely. The most likely explanation for the truncated arms is, therefore, partial staining due to the large diameter of the complex and also possible backbone flexibility in the outer arms

which is stabilized by contact with the carbon support in hexamers allowing the entire protein to be visualized.

This is a serious issue because docking will not be able to account correctly for all the extra density in GlnA relative to GSI. Contouring the density to enclose the volume corresponding to the molecular mass of GlnA will lead to too much density being displayed in the remainder of the molecule as a result of the missing density at the tips of the pinwheel arms not being considered.

4.4.3 Docking

It was obvious even from the 2D hexamer average that significant similarity existed between the structure of GlnA and GSI, the only GS for which a crystal structure is available. In particular the “oblate ellipsoids of revolution” which corresponded to GSI subunits and the indentations between subunits which corresponded to the active site $\alpha\beta$ barrels were shared between both structures, as confirmed by docking an entire ring of GSI into the 3D density map thus allowing a low resolution interpretation of the density.

The resolution of the density map was not, however, high enough to allow rigorous evaluation of these dockings and, consequently, confident interpretation of any but the low resolution features. The models did reveal, however, where the extra density of GlnA was positioned relative to the much smaller GSI as predicted by the alignments. For the most part these could be accounted for by the predicted flexibility of the polypeptide chain and the large areas of unoccupied density at this low resolution and contour level. However, unlike the N and C terminal insertions, the large mid-sequence insertions in several of the models could not be accounted for. It is uncertain whether this is a failure of the model, which might suffer from reduced density (as discussed above) or the primary differences in sequence between GlnA and GSI is localized to the protein's termini.

CHAPTER 5

SUMMARY AND CONCLUSIONS

5.1 SUMMARY

Bioinformatics approaches were undertaken to investigate the relationship between GSIII and the other GS families and to attempt to predict the structure of the enzyme with a view to interpreting the low resolution EM structure.

Algorithms utilizing state of the art alignment techniques, capable of detecting subtle homologies between sequences, appeared unable to correctly align the GSIII sequences with those of the GSIs and GSIIIs. This was most likely a result of the low global homology between the sequences.

In spite of this, manual profile-based alignment strategies, incorporating structural information, were able to reveal regions of significant homology between GSs from all four families, indicating the presence of a common core fold of an α/β barrel and the conservation of active site residues suggesting that they possess similar catalytic mechanisms.

Homology modelling was employed to produce four models for the structure of GSIII based on alternate alignments against GSI, with two being generated by automatic modelling algorithms and two by manual alignment. However, based on the disagreement with previous findings, the former alignments were doubtful. The level of sequence identity between the GSIII and GSI enzymes was not globally high enough to yield a confident prediction for the entire GSIII molecule. The highest confidence in the predicted structure resided in the highly conserved central regions forming the α/β barrel fold.

During the **purification** of recombinant GSIII from *E. coli*, a significant and sudden loss of activity was encountered. The reason for the loss in activity was not resolved. In particular, it was uncertain whether this loss was due to specific inhibition of the

enzyme or the effect of the conditions employed during the purification and in the case of the latter whether this was physiologically relevant.

A distribution of molecular weight species, of an otherwise pure preparation (as determined by SDS-PAGE), was discovered during the gel filtration stage of the purification. However, the activity profile only coincided with a single lower molecular weight (or volume) species in this distribution. A second gel filtration passage of fractions corresponding to the centroid of the previous run's activity peak produced similar results with lower molecular weight species redistributing to higher molecular weights and activity again associating with the same lower MW species. This suggested that an equilibrium distribution of molecular weights or hydrodynamic volumes existed in the preparations, but only one of the lower MW species was active.

GSIII was shown via **negative stain EM and 3D reconstruction** to associate into a higher order dodecameric form, in addition to the hexameric form determined previously (Southern, Parker, et al. 1987). Hexameric views, with similar overall arrangement but larger size in comparison to GSI, were also observed. However, it was uncertain whether these were true hexamers resulting from dissociation of the dodecamers or were a consequence of partial staining.

The structure of the dodecamer was solved using single particle technique to a resolution of 2.1 nm, revealing a dodecameric structure with subunits (82.8 KDa) arranged in two opposing hexagonal rings with distinct handedness. This was similar to the quaternary structures of the GSI and GlnT enzymes except that the complex is 50% longer and the two rings are pseudo-symmetrically related. The finding that particles lie in a preferred orientation, with the larger ring in contact with the carbon support, accounted for this asymmetry, through partial staining.

Indentations between subunits in the model were likely candidates for active sites as confirmed by the docking of an atomic structure of GSI. Although the N and C-terminal insertions of the proposed homology models could be accounted for, the 3D structure did not possess enough volume to account for insertions at the periphery of

the pinwheel arms. Possible truncation of the molecular envelope's density due to partial staining or inherent protein flexibility was thought to be responsible.

5.2 DISCUSSION OF THE IMPORTANT FINDINGS

5.2.1 Stoichiometry and inactivation

The sudden loss of activity together with the equilibrium distribution of molecular weights, the discovery of hexameric views, and the previous findings showing GSIII to be an active hexamer suggests that the GSIII exists in two distinct forms: an active hexamer and an inactive dodecamer. It has already been discussed in Chapter 3 that many GSs are capable of adopting a range of oligomeric states, the interconversion of which is particularly sensitive to environmental conditions, effector concentrations, and subunit concentrations. In fact GSII is known to be regulated by association-dissociation (Denman & Wedler 1984). And it has also been revealed that the particular N-source contained in the cultivation medium can effect the ratio of octamer to tetramer and hence the activity of GSII (Kretovich, Evstigneeva, et al. 1984).

The discovery of a heterogenous population of particles composed of both apparent hexamers and dodecamers, in a biochemically (SDS-PAGE) pure sample helps to explain the size distribution seen in the column chromatography by suggesting that GSIII is also subject to association-dissociation. The kinetics of this association-dissociation appears to favour the inactive dodecamer under the conditions of the purification. The association appears to occur at a rate commensurate with redistribution towards the dodecamer in the time period between elution from the gel filtration column and visualization, preventing the isolation of large numbers of hexamers. Furthermore, the discovery of aggregates in the apparently higher molecular weight fractions also suggests that this dissociation can progress to a point where random aggregation rather than specific oligomeric associations dominate. However, such amorphous aggregation cannot be responsible for the size distribution because it is unable to account for the distinct peaks during gel filtration. The presence of dodecamers also explains why the size-exclusion chromatography was not

able to resolve the individual peaks satisfactorily, because the MW of the purified protein is twice that predicted, placing the size-exclusion selectivity of the column outside of its optimum range.

In the introductory chapter it was discussed that both GSIIIs discovered to date undergo rapid, and in the case of GlnN, reversible inactivation in response to the addition of ammonium. Possible reasons for the inactivation were given in Chapter 3. The association of two active hexameric rings into an inactive dodecamer would account for this and in particular the reversibility of the inactivation. Whether this is a result of specific modification of the hexamer producing new isologous bonding sets or the effect of environmental conditions such as the decreased excluded volumes during PEG precipitation (the main point of activity loss) will need to be determined.

If the alternate hypothesis, that the pinwheel views represent partially stained dodecamers, is considered, then the equilibrium distribution of molecular weights and loss of activity remain unexplained. In light of the previous discovery of hexameric GSIIIs (Southern, Parker, et al. 1987; Garcia-Dominguez, Reyes, et al. 1997), the conditions encountered during the purification must have caused the higher order association of GSIII subunits into the dodecamer. The primary difference between the purification reported in this study and the previous work by Southern *et al.* (Southern, Parker, et al. 1987) was the purification step involving high salt concentrations to redissolve the PEG precipitate. This might have altered the quaternary structure of GSIII but not before activity was already lost in the PEG fractionation. Alternatively, the results reported by the previous researchers, which were based upon electrophoresis and gel filtration experiments, could have been misinterpreted, possibly due to dissociation of the GSIII complex.

5.2.2 *Docking*

The inability of the reconstructed volume to account for the insertions located in the periphery of the pinwheel arms of the two most confident models, which were produced manually after the failure of automated algorithms, was surprising. Two possibilities must be considered. (i) The volume is correct and one of the less likely

models, the FUGUE alignment, is, therefore, the most accurate. This would imply that GSIII only has large N and C-terminal insertions relative to GSI. (ii) The arms of the larger ring, like those of the smaller, could be artefactually too small (Section 4.4.2). If the 2D averages reflect the correct diameter of the particles then the extra density would help account for large insertions near the tips of the arms making the two manually derived models more plausible. The second manual alignment, with an alternate C-terminal alignment, had an extra 115 amino-acid residue insertion in a similar position to the insertion shared by both manual alignments, which would additionally have to be accounted for. It, therefore, seems that the first alignment agrees most strongly with volume even if it is an under-representation.

5.2.3 *Evolutionary relationships*

The low resolution of the maps did not allow a detailed interpretation of the density but they did reveal similarities between GSIII and GSI despite them sharing only 9% sequence identity. In particular, the overall similar arrangements of subunits and the characteristic indentations between them led to the putative identification of the positions of the active sites in GSIII.

The previous reporting of a hexameric GSIII (Southern, Parker, et al. 1987; Garcia-Dominguez, Reyes, et al. 1997), and the subsequent confirmation in this work, do not agree with the characteristic features of GS quaternary structure (Kretovich, Evstigneeva, et al. 1984). To review: (i) all GS enzymes consist of an even number of monomers. (ii) The number of monomers can be equal to 6, 8, and 12. (iii) all GSs are double-ringed structures with monomers possessing dihedral point group symmetry. The C₆ symmetry of the hexamer, therefore, disagrees with the last point. A hexameric GS has been isolated before but it possessed D₃ symmetry (Rasulov, Evstigneeva, et al. 1977). The discovery of a dodecameric two-ringed state of GSIII is consistent with previous findings.

An interesting point to consider is which conserved regions of the structures and, therefore, sequence are responsible for the characteristic quaternary structure features seen for all GS enzymes visualized to date? The requirement for the subunits to

arrange into rings can be understood by the fact that the active site is formed between adjacent subunits (Eisenberg, Gill, et al. 2000). In the case of the octameric form of GSII this requirement has led to the ‘half-site’ hypothesis where only two active sites are present per ring. This limit results from the angular restraints required to form a ring (Eisenberg, Gill, et al. 2000). GSIII is clearly arranged in a cyclic fashion and features similar in shape and position to GSI are obvious between its subunits, thus implying the existence of similar heterologous contacts.

The isologous contacts (related by 2-fold axis) between rings must also be conserved. In GSI the two main structural interactions responsible for heterologous interactions are 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 α -helical thong from one subunit with a hydrophobic cavity in the opposite subunit (Almassy, Janson, et al. 1986). The latter is thought to be the energetically more important interaction responsible for the association (Almassy, Janson, et al. 1986). However, the sequence alignments discussed in Chapter 2 reveal that this helical thong is only present in GSI. It, therefore, remains to be determined whether the alignments are incorrect, or whether, the former associations and perhaps other unknown interactions are more likely responsible for heterologous associations. Considering, the well defined central channels of all GSs studied to date, it seems likely that the β -sheets associations are important. However, the presence of a β sheet was not predicted by the alignments reported in Chapter 2. In his discussion of the symmetry considerations of protomers forming ‘closed crystals’, Hanson, raises the important point that information about the symmetry of an oligomer can only provide limited insights into the number and types of binding sets because additional binding is possible that does not compromise the symmetry between protomers of the oligomers (Hanson 1966). This makes it difficult, in this case, to identify the conserved feature responsible for oligomerization from the quaternary structures of GSs alone.

5.3 FUTURE DIRECTIONS

Still very little is known about the GSIII family in contrast to the well characterised GSIs and GSII, despite their discovery seventeen years ago. The work presented in

this thesis represents the first implementation of a strategy to elucidate the structure/functioning of the GSIII enzymes and has made significant progress in this direction. However, the lack of knowledge surrounding the GSIII family still needs to be addressed if the functioning of this enzyme is to be fully understood. Further structural studies and biochemical characterisation of this enzyme are required.

5.3.1 Biochemical Characterisation:

Extensive enzymatic characterization of GSIII is required to better understand its functioning and regulation. This would include studies into the cooperativity, stability, association-dissociation kinetics, and the regulation by effectors and environmental conditions such as pH and molecular crowding. The short time-scale post-translational modification responsible for the loss of activity needs to be identified and so does its physiological relevance. Specifically, biochemical evidence is needed in support of, or against, the hypothesis of inactivation due to higher order association.

As a foundation for this and future structural studies, several improvements to the protein purification can be made. A gel filtration matrix with a bigger molecular weight exclusion limit is required to achieve better separation of the different molecular weight species of GSIII. An additional step such as affinity chromatography on a Cibracon blue column (ADP analogue) or Reactive Red agarose (NADP analogue) is also required to avoid contamination by GroEL. Activity assays measuring biosynthetic activity need to be implemented to avoid uncertainty in the relevance of activity measurements. Stability studies are required and specifically the dissociation kinetics of GSIII needs to be investigated by techniques such as, but not limited to: gel filtration, ultracentrifugation, cross-linking studies, small-angle scattering and native PAGE. Most importantly, an attempt needs to be made to isolate hexamers as described by previous researchers. The protocol of Garcia-Dominguez et al. (Garcia-Dominguez, Reyes, et al. 1997) would be the starting point for this attempt. It would also be of interest to purify GSIII in the presence of the inhibitor MetSox, which stabilises inter-subunit contacts (Maurizi & Ginsburg 1982) to reveal which MW species can be isolated.

In Chapter 1, the ultimate goal of targeting this important enzyme as a means of controlling the pathogenicity of *B. fragilis*, was introduced. If GSIII is to be a candidate point of control against *B. fragilis*, then the functioning of this enzyme needs to be understood in its biological context. Firstly, without a better understanding of the human GSII enzyme it will be difficult to rationally target the pathogen's enzyme in such a way as to not affect the human GSII enzyme. Secondly, the genome sequence of *B. fragilis* (personal communication Abratt, V.R. 2004) suggests the presence of another GSI-like enzyme, and this needs to be investigated. If another GS does exist as in the case of *Synechocystis* sp. PCC 6803 (Reyes, Muro-Pastor, et al. 1997; Reyes & Florencio 1994a), the interplay of these seemingly redundant isozymes needs to be elucidated to reveal whether inhibition of GSIII could be compensated for by the other GS.

5.3.2 Structural studies

Future structure determinations by EM will need to circumvent the partial staining problem. This would not only reveal the full particle morphology but also discriminate between true dissociation of the dodecamer into hexamers and artefacts from partial staining. This could be achieved by utilizing alternate staining/embedding techniques such as: the double-carbon method, negative stain-sugar embedding on holey-carbon, and cryoEM. The latter is the route of choice to avoid any perturbations of the molecular structure as a result of the embedding technique and to reveal internal density within the structure.

An understanding of GSIII dissociation kinetics would certainly benefit from an in depth EM investigation of the nature carried out by Wringley *et al.* into the interconversion of glucose 6-phosphate dehydrogenase (G6PD) tetramers, dimers, and monomers (Wringley & Heather 1972). Detailed ultracentrifugation studies revealed that G6PD exists as a distribution of oligomeric intermediates, the equilibrium of which depends on environmental conditions such as pH and ionic strength. By investigating a time series of dissociation and association intermediates, Wringley *et al.* were able to relate the activity of preparations with the oligomeric species present,

allowing them to elucidate the kinetics and detail the subunit conformational changes associated with the interconversions.

This study has highlighted the difficulties of interpreting a low resolution structure according to the docking of homology models with low global homology to a known structure. Within the immediate future, bioinformatic modelling tools are not likely to reach a point where they can successfully predict the structure of proteins with little similarity to regions of already mapped fold-space. Therefore, higher resolution structural studies are required if insights are to be gained from the structure of GSIII and attempts made to specifically disrupt this enzyme. The ultimate goal of rational inhibitor design against GSIII will require a detailed understanding of the active site configuration that could only be achieved by X-ray crystallography. Later studies could include co-crystallization with different ligands, cofactors and inhibitors. An atomic resolution structure of a GS from another family, apart from, GSI would greatly facilitate the understanding of the evolution and conservation of the structure/function relationships amongst the GS enzymes. This would also improve the ability of automated methods to detect such distant relationships by allowing the construction of 3D PSSMs reflecting the structural homology between families.

EM and image processing still has a major role to play in elucidating the structure of GSIII. Higher resolution EM studies in the presence of different effectors, ligands and environmental conditions are required for a better understanding of the relationship between structure and function in GSIII. In addition, greater knowledge of the inactivation and dissociation kinetics of this enzyme is needed. Even if an atomic resolution crystal structure can be solved, the EM data would not be superfluous because EM can reveal the dynamic nature of molecules by trapping intermediate states that otherwise might not crystallize, and is also a faster route to this information. Flexible docking of the higher resolution atomic structure could reveal the dynamic motions of GSIII domains (structural transitions) as a result of the binding of different ligands, different conditions, and different oligomeric states, thus revealing important clues about the functioning and allosteric regulation of this enzyme. Additionally, a high resolution EM density map could contribute to the initial phasing of the X-ray crystallographic data (Stark 2002) via molecular replacement.

Before such studies are attempted, several improvements in the bioinformatic and EM methodologies applied in this study, in addition to those listed above, could be achieved. Firstly, the accuracy of the homology models might be increased by improving the quality of the alignments using larger profile based alignments. Incorporating feedback from the threading algorithms i.e. using threading energies and 3D-PSSM to generate the best alignments for the correctly identified folds, would also improve the alignment accuracy. Secondly, higher resolution data is required to allow more accurate interpretation of the molecular density. Data acquisition on machines capable of higher resolution e.g. EMs fitted with a FEG or Lab6 emission sources, which can perceive higher resolution details, would provide the greatest step towards achieving this. Thirdly, due to the low resolution of the density maps, the accuracy and efficiency automated docking could be improved by imposing constraints during the docking procedure, thus reducing the number of degrees of freedom and therefore increasing the data-to-parameter ratio. For example, enforcing symmetry during docking i.e. simultaneously find the maximum correlation between all subunits of a symmetric model at the same time. This would be equivalent to the docking of a single subunit in only the asymmetric unit but translational and rotational parameters of the particle would be constrained by the volumes of the symmetry related protomers. Additional requirements derived from knowledge (biochemically or bioinformatically derived) of the sites of interaction between subunits, such as the requirement the formation of the active site α/β barrel by adjacent subunits, could be incorporated.

5.4 CONCLUSIONS

This work has revealed the first ever structure of a GSIII, the most phylogenetically distant GS from GSI. These efforts have also highlighted the relationships between all GS families and through the first full length alignments of GS sequences from all four families, demonstrated the conservation of all active site residues, core active site $\alpha\beta$ barrel fold motifs, and additional previously unreported regions. In addition to providing unequivocal proof of the structural similarity of GSIII and GSI, and confirming the presence of putative $\alpha\beta$ barrel active site folds, this work has made

steps towards understanding the regulation of this enzyme and has put forward a hypothesis regarding this regulation. It is proposed that GSIII occurs as both active hexamer and an inactive dodecamer, the interconversion of which is thought to represent a means of reversible post-translational regulation.

This work has made progress towards understanding the structure/function of the poorly characterised GSIII family and has paved the way for future structural studies into this novel enzyme.

REFERENCES

- DNAMAN. (Version 4.13), 1994, Lynnon BioSoft, Computer Program.
- 1997, "CATH-a hierarchic classification of protein domain structures", vol. 5, no. 8, pp. 1093-1108.
- EM-MENU, 2003, Tietz Video and Image Processing Systems, Germany. Computer Program.
- Abrahams, G. L. 1996, *Nitrogen metabolism of Bacteroides fragilis bfl*, B.Sc (Honours) dissertation, UCT.
- Almassy, R. J., Janson, C. A., Hamlin, R., Xuong, N. H., & Eisenberg, D. 1986a, "Novel subunit-subunit interactions in the structure of glutamine synthetase", *Nature*, vol. 323, no. 6086, pp. 304-309.
- Backman, K., Chen, Y.-M., & Magasanik, B. 1981, "Physical and genetic characterization of the glnA-glnG region of the Escherichia coli chromosome.", *Proceedings of the National Academy of Sciences USA*, vol. 78, pp. 3743-3747.
- Balakrishnan, M. S., Villafranca, J. J., & Brenchley, J. E. 1977, "Glutamine synthetase from Salmonella typhimurium: manganese(II), substrate, and inhibitor interaction with the unadenylylated enzyme", *Arch.Biochem Biophys.*, vol. 181, no. 2, pp. 603-615.
- 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.*, vol. 129, no. 2, pp. 1001-1009.
- Bochkareva, E. S. & Girshovich, A. S. 1994, "ATP induces non-identity of two rings in chaperonin GroEL", *J Biol.Chem.*, vol. 269, no. 39, pp. 23869-23871.
- 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.)*, vol. 67, no. 9, pp. 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*, vol. 72, pp. 248-254.

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L., & Sigler, P. B. 1994, "The crystal structure of the bacterial chaperonin GroEL at 2.8 Å", *Nature*, vol. 371, no. 6498, pp. 578-586.

Brandon, C. & Tooze, J. 1999, *Introduction to Protein Structure*, second edn, Garland Publishing, Inc., New York.

Burford, D., Sewell, T., Lyness, C., & Marte, O.-C. SpiderViewer. [1.0], 2004, Cape Town, Collaborative Visual Computing Lab, Department of Computer Science, UCT, Computer Program.

Bywater, R. P., Carlisle, C. H., Jackson, R. B., Mackay, A. L., & Timmins, P. A. 1975, "Crystals of glutamine synthetase from *Escherichia coli*" *Journal of Molecular Biology*, vol. 91, no. 3, pp. 293-300.

Caizzi, R., Bozzetti, M. P., Caggese, C., & Ritossa, F. 1990, "Homologous nuclear genes encode cytoplasmic and mitochondrial glutamine synthetase in *Drosophila melanogaster*", *J Mol Biol.*, vol. 212, no. 1, pp. 17-26.

Capela, D., Barloy-Hubler, F., Gouzy, J., Bothe, G., Ampe, F., Batut, J., Boistard, P., Becker, A., Boutry, M., Cadieu, E., Dreano, S., Gloux, S., Godrie, T., Goffeau, A., Kahn, D., Kiss, E., Lelaure, V., Masuy, D., Pohl, T., Portetelle, D., Puhler, A., Purnelle, B., Ramsperger, U., Renard, C., Thebault, P., Vandenbol, M., Weidner, S., & Galibert, F. 2001, "Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021", *Proc.Natl.Acad.Sci.U.S.A*, vol. 98, no. 17, pp. 9877-9882.

- Chiurazzi, M., Meza, R., Lara, M., Lahm, A., Defez, R., Iaccarino, M., & Espin, G. 1992, "The *Rhizobium leguminosarum* biovar phaseoli *glnT* gene, encoding glutamine synthetase III", *Gene*, vol. 119, no. 1, pp. 1-8.
- Colombo, G. & Villafranca, J. J. 1986, "Amino acid sequence of *Escherichia coli* glutamine synthetase deduced from the DNA nucleotide sequence", *J Biol.Chem.*, vol. 261, no. 23, pp. 10587-10591.
- Crespo, J. L., Garcia-Dominguez, M., & Florencio, F. J. 1998, "Nitrogen control of the *glnN* gene that codes for GS type III, the only glutamine synthetase in the cyanobacterium *Pseudanabaena* sp. PCC 6903", *Mol.Microbiol.*, vol. 30, no. 5, pp. 1101-1112.
- Crowther, R. A., Amos, L. A., Finch, J. T., De Rosier, D. J., & Klug, A. 1970, "Three dimensional reconstructions of spherical viruses by fourier synthesis from electron micrographs", *Nature*, vol. 226, no. 5244, pp. 421-425.
- Crowther, R. A., Henderson, R., & Smith, J. M. 1996, "MRC Image Processing Programs", *J Struct.Biol.*, vol. 116, pp. 9-16.
- Davis, R. W., Botstein, D., & Roth, J. R. 1980, *Advanced Bacterial Genetics* Cold Spring Harbour Laboratory, New York.
- de Bakker, P. I., Bateman, A., Burke, D. F., Miguel, R. N., Mizuguchi, K., Shi, J., Shirai, H., & Blundell, T. L. 2001, "HOMSTRAD: adding sequence information to structure-based alignments of homologous protein families", *Bioinformatics.*, vol. 17, no. 8, pp. 748-749.
- Delano, W. L. PyMOL, 2004, San Carlos, CA,USA, DeLano Scientific LLC, Computer Program.
- Denman, R. B. & Wedler, F. C. 1984, "Association-dissociation of mammalian brain glutamine synthetase: effects of metal ions and other ligands", *Arch.Biochem Biophys.*, vol. 232, no. 2, pp. 427-440.

Eisen, J. A., Nelson, K. E., Paulsen, I. T., Heidelberg, J. F., Wu, M., Dodson, R. J., Deboy, R., Gwinn, M. L., Nelson, W. C., Haft, D. H., Hickey, E. K., Peterson, J. D., Durkin, A. S., Kolonay, J. L., Yang, F., Holt, I., Umayam, L. A., Mason, T., Brenner, M., Shea, T. P., Parksey, D., Nierman, W. C., Feldblyum, T. V., Hansen, C. L., Craven, M. B., Radune, D., Vamathevan, J., Khouri, H., White, O., Gruber, T. M., Ketchum, K. A., Venter, J. C., Tettelin, H., Bryant, D. A., & Fraser, C. M. 2002, "The complete genome sequence of *Chlorobium tepidum* TLS, a photosynthetic, anaerobic, green-sulfur bacterium", *Proc.Natl.Acad.Sci.U.S.A*, vol. 99, no. 14, pp. 9509-9514.

Eisenberg, D., Gill, H. S., Pflugel, G. M., & Rotstein, S. H. 2000, "Structure-function relationships of glutamine synthetases", *Biochim.Biophys.Acta*, vol. 1477, no. 1-2, pp. 122-145.

Feng, D. F. & Doolittle, R. F. 1987, "Progressive sequence alignment as a prerequisite to correct phylogenetic trees", *J Mol Evol.*, vol. 25, no. 4, pp. 351-360.

Frank, J. 1996, "Three-Dimensional Electron Microscopy of Macromolecular Assemblies", Academic Press, inc., London.

Frank, J., Goldfarb, W., Eisenberg, D., & Baker, T. S. 1978, "Reconstruction of glutamine synthetase using computer averaging", *Ultramicroscopy*, vol. 3, no. 3, pp. 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.*, vol. 116, no. 1, pp. 190-199.

Frey, T. G. & Eisenberg, D. 1984, "Formation and structure of 1-, 3- and 6 + 1-stranded helical cables of glutamine synthetase.", *International Journal of Biological Macromolecules*, vol. 6, no. 1, pp. 2-12.

Frey, T. G., Eisenberg, D., & Eiserling, F. A. 1975, "Glutamine synthetase forms three- and seven-stranded helical cables", *Proc.Natl.Acad.Sci.U.S.A*, vol. 72, no. 9, pp. 3402-3406.

Frink, R. J., Eisenberg, D., & Glitz, D. G. 1978, "Localization of the site of adenylylation of glutamine synthetase by electron microscopy of an enzyme-antibody complex", *Proc.Natl.Acad.Sci.U.S.A.*, vol. 75, no. 12, pp. 5778-5782.

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.*, vol. 244, no. 1, pp. 258-264.

Garcia-Dominguez, M., Reyes, J. C., & Florencio, F. J. 1999, "Glutamine synthetase inactivation by protein-protein interaction", *Proc.Natl.Acad.Sci.U.S.A.*, vol. 96, no. 13, pp. 7161-7166.

Gibbs, C. S., Campbell, K. E., & Wilson, R. H. 1987, "Sequence of a human glutamine synthetase cDNA", *Nucleic Acids Res.*, vol. 15, no. 15, p. 6293.

Gibson III, F. C., Onderdonk, A. B., Kasper, D. L., & Tzianabos, A. O. 1998, "Cellular Mechanism of Intraabdominal Abscess Formation by *Bacteroides fragilis*", *Journal of Immunology*, vol. 160, pp. 5000-5006.

Gibson, S. A. & Macfarlane, G. T. 1988, "Characterization of proteases formed by *Bacteroides fragilis*", *J Gen.Microbiol.*, vol. 134 (Pt 8), pp. 2231-2240.

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*, vol. 40, no. 7, pp. 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*, vol. 41, no. 31, pp. 9863-9872.

Glaeser, R. M. 1999, "Review: electron crystallography: present excitement, a nod to the past, anticipating the future", *J Struct.Biol*, vol. 128, no. 1, pp. 3-14.

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.*, vol. 139, pp. 1487-1493.

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.Natl.Acad.Sci.U.S.A*, vol. 91, no. 26, pp. 12828-12831.

Grigorieff, N. 2000, "Resolution measurement in structures derived from single particles", *Acta Crystallogr.D.Biol Crystallogr.*, vol. 56 (Pt 10), pp. 1270-1277.

Guex, N. & Peitsch, M. C. 1997, "SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling", *Electrophoresis*, vol. 18, no. 15, pp. 2714-2723.

Hanson, K. R. 1966, "Symmetry of protein oligomers formed by isologous association", *J Mol Biol*, vol. 22, no. 2, pp. 405-409.

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

Harth, G. & Horwitz, M. A. 1997, "Expression and efficient export of enzymatically active *Mycobacterium tuberculosis* glutamine synthetase in *Mycobacterium smegmatis* and evidence that the information for export is contained within the protein", *J Biol.Chem.*, vol. 272, no. 36, pp. 22728-22735.

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.*, vol. 189, no. 9, pp. 1425-1436.

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

- Haschemeyer, R. H. & De Harven, E. 1974, "Electron microscopy of enzymes", *Annu.Rev.Biochem*, vol. 43, no. 0, pp. 279-301.
- 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.*, vol. 135 (Pt 12), pp. 3271-3279.
- Holm, L. & Sander, C. 1994, "The FSSP database of structurally aligned protein fold families", *Nucleic Acids Res.*, vol. 22, no. 17, pp. 3600-3609.
- Ingham, K. C. 1984, "Protein Precipitation with Polyethylene Glycol", *Methods in Enzymology*, vol. 104, pp. 351-356.
- 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*, vol. 46, no. 2-3, pp. 297-300.
- Janssen, P. J., Jones, W. A., Jones, D. T., & Woods, D. R. 1988, "Molecular analysis and regulation of the *glnA* gene of the Gram-positive anaerobe *Clostridium acetobutylicum*", *J.Bacteriol.*, vol. 170, pp. 400-408.
- Jones, D. T. 1999a, "GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences", *J Mol Biol.*, vol. 287, no. 4, pp. 797-815.
- Jones, D. T. 1999b, "Protein secondary structure prediction based on position-specific scoring matrices", *J Mol Biol*, vol. 292, no. 2, pp. 195-202.
- Joyeux, L. & Penczek, P. A. 2002, "Efficiency of 2D alignment methods", *Ultramicroscopy*, vol. 92, no. 2, pp. 33-46.
- Kabsch, W., Kabsch, H., & Eisenberg, D. 1976, "Packing in a new crystalline form of glutamine synthetase from *Escherichia coli*." *Journal of Molecular Biology*, vol. 100, no. 3, pp 283-291.

- Kelley, L. A., MacCallum, R. M., & Sternberg, M. J. 2000, "Enhanced genome annotation using structural profiles in the program 3D-PSSM", *J Mol Biol.*, vol. 299, no. 2, pp. 499-520.
- Kessel, M., Frank, J., & Goldfarb, W. 1980, "Averages of Glutamine Synthetase Molecules as Obtained With Various Stain and Electron Dose Conditions", *Journal of Supramolecular Structure*, vol. 14, pp. 405-422.
- Kiang, C.-H. 2000, "Single-particle study of protein assembly", *Physical Review E*, vol. 64, pp. 041911-1 – 041911-2.
- Kiselev, N. A., Sherman, M. B., & Tsuprun, V. L. 1990, "Negative staining of proteins", *Electron Microsc.Rev.*, vol. 3, no. 1, pp. 43-72.
- Kretovich, W. L., Evstigneeva, Z. G., Pushkin, A. V., & Tsuprun, V. L. 1984b, "Evolution of the quaternary structure of glutamine synthetase", *Quaderni de La Ricerca Scientifica*, vol. 113, no. Macromol. Funct. Cell, pp. 109-115.
- Kumada, Y., Takano, E., Nagaoka, K., & Thompson, C. J. 1990, "Streptomyces hygrosopicus has two glutamine synthetase genes", *J Bacteriol.*, vol. 172, no. 9, pp. 5343-5351.
- Kunath, W., Weiss, K., Sack-Kongehl, H., Kessel, M., & Zeitler, E. 1984, "Time-resolved low-dose microscopy of glutamine synthetase molecules", *Ultramicroscopy*, vol. 13, no. 3, pp. 241-252.
- Kuo, I. A. & Glaeser, R. M. 1975, "Development of methodology for low exposure, high resolution electron microscopy of biological specimens", *Ultramicroscopy*, vol. 1, no. 1, pp. 53-66.
- Kuwahara, T., Yamashita, A., Hirakawa, H., Nakayama, H., Toh, H., Okada, N., Kuhara, S., Hattori, M., Hayashi, T., & Ohnishi, Y. 2004, "Genomic analysis of *Bacteroides fragilis* reveals extensive DNA inversions regulating cell surface adaptation", *Proc.Natl.Acad.Sci.U.S.A*, vol. 101, no. 41, pp. 14919-14924.

- Laemmli, U. K. 1970, "Cleavage of structural proteins during the assembly of the head of bacteriophage T4", *Nature*, vol. 227, no. 259, pp. 680-685.
- Lei, M., Aebi, U., Heidner, E. G., & Eisenberg, D. 1979, "Limited proteolysis of glutamine synthetase is inhibited by glutamate and by feedback inhibitors", *J.Biol.Chem.*, vol. 254, no. 8, pp. 3129-3134.
- Liaw, S. H., Pan, C., & Eisenberg, D. 1993, "Feedback inhibition of fully unadenylylated glutamine synthetase from *Salmonella typhimurium* by glycine, alanine, and serine", *Proc.Natl.Acad.Sci.U.S.A.*, vol. 90, no. 11, pp. 4996-5000.
- Llorca, O., Smyth, M. G., Carrascosa, J. L., Willison, K. R., Radermacher, M., Steinbacher, S., & Valpuesta, J. M. 1999, "3D reconstruction of the ATP-bound form of CCT reveals the asymmetric folding conformation of a type II chaperonin", *Nat.Struct.Biol.*, vol. 6, no. 7, pp. 639-642.
- Maurizi, M. & Ginsburg, A. 1982, "Active Site Ligand Stabilization of Quaternary Structures of Glutamine Synthetase from *Escherichia coli*", *J.Biol.Chem.*, vol. 257, no. 12, pp. 7246-7251.
- McGuffin, L. J., Bryson, K., & Jones, D. T. 2000, "The PSIPRED protein structure prediction server", *Bioinformatics.*, vol. 16, no. 4, pp. 404-405.
- McGuffin, L. J. & Jones, D. T. 2003, "Improvement of the GenTHREADER method for genomic fold recognition", *Bioinformatics.*, vol. 19, no. 7, pp. 874-881.
- 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*, vol. 153, no. 3, pp. 597-606.
- Merrick, M. J. & Edwards, R. A. 1995, "Nitrogen control in bacteria", *Microbiol.Rev.*, vol. 59, no. 4, pp. 604-622.
- Mizuguchi, K., Deane, C. M., Blundell, T. L., Johnson, M. S., & Overington, J. P. 1998, "JOY: protein sequence-structure representation and analysis", *Bioinformatics.*, vol. 14, no. 7, pp. 617-623.

- Nicholas, K. B. & Nicholas, H. B. GeneDoc [2.6.002], 1997, Computer Program.
- Obojska, A., Berlicki, L., Kafarski, P., Lejczak, B., Chicca, M., & Forlani, G. 2004, "Herbicide pyridyl derivatives of aminomethylene-bisphosphonic acid inhibit plant glutamine synthetase", *J Agric. Food Chem.*, vol. 52, no. 11, pp. 3337-3344.
- Orlova, E. V., Dube, P., Harris, J. R., Beckman, E., Zemlin, F., Markl, J., & van Heel, M. 1997, "Structure of keyhole limpet hemocyanin type 1 (KLH1) at 15 Å resolution by electron cryomicroscopy and angular reconstitution", *J Mol Biol.*, vol. 271, no. 3, pp. 417-437.
- Palacios, R. 1976, "Neurospora crassa glutamine synthetase. Purification by affinity chromatography and characterization of subunit structure", *J Biol Chem.*, vol. 251, no. 15, pp. 4787-4791.
- Penczek, P., Grassucci, R. A., & Frank, J. 1994, "The ribosome at improved resolution: new techniques for merging and orientation refinements in 3D cryo-electron microscopy of biological particles.", *Ultramicroscopy*, vol. 53, pp. 251-270.
- Penczek, P., Radermacher, M., & Frank, J. 1992, "Three-dimensional reconstruction of single particles embedded in ice", *Ultramicroscopy*, vol. 40, no. 1, pp. 33-53.
- Penczek, P. A., Zhu, J., & Frank, J. 1996, "A common-lines based method for determining orientations for $N > 3$ particle projections simultaneously", *Ultramicroscopy*, vol. 63, no. 3-4, pp. 205-218.
- Pesole, G., Bozzetti, M. P., Lanave, C., Preparata, G., & Saccone, C. 1991, "Glutamine synthetase gene evolution: a good molecular clock", *Proc.Natl.Acad.Sci.U.S.A*, vol. 88, no. 2, pp. 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.*, vol. 25, no. 13, pp. 1605-1612.

Pfluegl, G.M., Kiang, C-H, 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 Crystallographica* A58 (Supplement), 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", *Biochimica et Biophysica Acta*, vol. 828, no. 3, pp. 336-350.

Rasulov, A. S., Evstigneeva, Z. G., Kretovich, V. L., Stelmashuk, V. J., Samsonidze, T. G., & Kiselev, N. A. 1977, "Purification, properties and quaternary structure of *Chlorella* glutamine synthetase", *Biokhimiya (Moscow)*, vol. 42, p. 350.

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 (Moscow)*, vol. 51, no. 3, pp. 413-419.

Reyes, J. C. & Florencio, F. J. 1994a, "A mutant lacking the glutamine synthetase gene (*glnA*) is impaired in the regulation of the nitrate assimilation system in the cyanobacterium *Synechocystis* sp. strain PCC 6803", *J.Bacteriol.*, vol. 176, no. 24, pp. 7516-7523.

Reyes, J. C. & Florencio, F. J. 1994b, "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.*, vol. 176, no. 5, pp. 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.*, vol. 179, no. 8, pp. 2678-2689.

Rochefort, D. A. & Benson, D. R. 1990, "Molecular cloning, sequencing, and expression of the glutamine synthetase II (*glnII*) gene from the actinomycete root nodule symbiont *Frankia* sp. strain Cp11", *J Bacteriol.*, vol. 172, no. 9, pp. 5335-5342.

- Rosenthal, P. B. & Henderson, R. 2003, "Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy", *J Mol Biol.*, vol. 333, no. 4, pp. 721-745.
- Sarkar, P. K., Fischman, D. A., Goldwasser, E., & Moscona, A. A. 1972, "Isolation and characterization of glutamine synthetase from chicken neural retina", *J Biol Chem.*, vol. 247, no. 23, pp. 7743-7749.
- Sato, S., Nakamura, Y., Kaneko, T., Katoh, T., Asamizu, E., & Tabata, S. 2000, "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence features of the regions of 4,504,864 bp covered by sixty P1 and TAC clones", *DNA Res.*, vol. 7, no. 2, pp. 131-135.
- Schatz, M. & van Heel, M. 1990, "Invariant classification of molecular views in electron micrographs", *Ultramicroscopy*, vol. 32, no. 3, pp. 255-264.
- Schwede, T., Kopp, J., Guex, N., & Peitsch, M. C. 2003, "SWISS-MODEL: An automated protein homology-modeling server", *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3381-3385.
- Segal, A. & Stadtman, E. R. 1972a, "Effects of cobaltous ion on various catalytic parameters and on heterologous subunit interactions of Escherichia coli glutamine synthetase", *Arch.Biochem Biophys.*, vol. 152, no. 1, pp. 356-366.
- Segal, A. & Stadtman, E. R. 1972b, "Variation of the conformational states of Escherichia coli glutamine synthetase by interaction with different divalent cations", *Arch.Biochem Biophys.*, vol. 152, no. 1, pp. 367-377.
- Shannon, C. E. 1949, "Communication in the presence of noise", *Proc.Instit.Radio Engin.*, vol. 37, pp. 10-22.
- Shi, J., Blundell, T. L., & Mizuguchi, K. 2001, "FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties", *J Mol Biol.*, vol. 310, no. 1, pp. 243-257.

Smith, J. M. 1999, "Ximdisp--A visualization tool to aid structure determination from electron microscope images", *J Struct.Biol.*, vol. 125, no. 2-3, pp. 223-228.

Sonnhammer, E. L., Eddy, S. R., & Durbin, R. 1997a, "Pfam: a comprehensive database of protein domain families based on seed alignments", *Proteins*, vol. 28, no. 3, pp. 405-420.

Sorzano, C. O., de la Fraga, L. G., Clackdoyle, R., & Carazo, J. M. 2004, "Normalizing projection images: a study of image normalizing procedures for single particle three-dimensional electron microscopy", *Ultramicroscopy*, vol. 101, no. 2-4, pp. 129-138.

Southern J.A. 1986, *Molecular Genetic Studies of Bacteroides Fragilis*, Ph.D Thesis, University of Cape Town.

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.*, vol. 133, pp. 2437-2446.

Stadtman, E. R. 2001, "The story of glutamine synthetase regulation", *J.Biol.Chem.*, vol. 276, no. 48, pp. 44357-44364.

Stahl, J. & Jaenicke, L. 1972, "Structure of glutamine synthetase from pig brain", *European Journal of Biochemistry*, vol 29, no 3, pp 401-407. 1972.

Stark, H. 2002, "Three-dimensional electron cryomicroscopy of ribosomes", *Curr.Protein Pept.Sci.*, vol. 3, no. 1, pp. 79-91.

Streicher, S. L. & Tyler, B. 1980, "Purification of glutamine synthetase from a variety of bacteria", *J Bacteriol.*, vol. 142, no. 1, pp. 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.*, vol. 247, no. 17, pp. 5312-5321.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., & Higgins, D. G. 1997, "The CLUSTAL_X windows interface: flexible strategies for multiple sequence

alignment aided by quality analysis tools", *Nucleic Acids Res.*, vol. 25, no. 24, pp. 4876-4882.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. 1994, "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice", *Nucleic Acids Res.*, vol. 22, no. 22, pp. 4673-4680.

Tiemeier, D. C. & Milman, G. 1972, "Chinese hamster liver glutamine synthetase. Purification, physical and biochemical properties", *J Biol Chem.*, vol. 247, no. 8, pp. 2272-2277.

Tietz, H.R. 1986, "On-line image processing in real time for transmission electron microscopy", *Journal de Microscopie et de Spectroscopie Electroniques*, vol 11.

Tischer, E., Das Sarma, S., & Goodman, H. M. 1986, "Nucleotide sequence of an alfalfa glutamine synthetase gene", *Molecular and General Genetics*, vol. 203, pp. 221-229.

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*, vol. 626, no. 1, pp. 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", *Biochimica et Biophysica Acta*, vol. 913, no. 3, pp. 368-376.

Tyler, B. 1978, "Regulation of the assimilation of nitrogen compounds", *Annu.Rev.Biochem*, vol. 47, pp. 1127-1162.

Valentine, R. C., Shapiro, B., & Stadtman, E. R. 1968a, "Regulation of Glutamine Synthetase. XII. Electron Microscopy of the Enzyme from *Escherichia coli*.", *Biochemistry*, vol. 7, no. 6, pp. 2143-2152.

van Heel, M. & Harauz, G. 1986, "Resolution criteria for three dimensional reconstruction", *Optik*, vol. 73, no. 3, pp. 119-122.

van Oostrum, J., Smith, P. R., Mohraz, M., & Burnett, R. M. 1987, "The structure of the adenovirus capsid. III. Hexon packing determined from electron micrographs of capsid fragments", *J Mol Biol*, vol. 198, no. 1, pp. 73-89.

Wehnert, G. U., Abratt, V. R., Goodman, H. J., & Woods, D. R. 1990, "Cloning of *Bacteroides fragilis* plasmid genes affecting metronidazole resistance and ultraviolet survival in *Escherichia coli*", *Plasmid*, vol. 23, no. 2, pp. 155-158.

Wen, Z. T., Peng, L., & Morrison, M. 2003, "The glutamine synthetase of *Prevotella bryantii* B(1)4 is a family III enzyme (GlnN) and glutamine supports growth of mutants lacking glutamate dehydrogenase activity", *FEMS Microbiol. Lett.*, vol. 229, no. 1, pp. 15-21.

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., Fraser, C. M., & . 1999, "Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1", *Science*, vol. 286, no. 5444, pp. 1571-1577.

Wilk, S., Meister, A., & Haschemeyer, R. H. 1969, "Studies on the subunit structure of ovine brain glutamine synthetase", *Biochemistry*, vol. 8, no. 8, pp. 3168-3174.

Williams, R. C. & Fisher, H. W. 1970, "Electron Microscopy of TMV under conditions of minimal beam exposure", *Journal of Molecular Biology*, vol. 52, no. 121-123.

Wood, V., Gwilliam, R., Rajandream, M. A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., Basham, D., Bowman, S., Brooks, K., Brown, D., Brown, S., Chillingworth, T., Churcher, C., Collins, M., Connor, R., Cronin, A., Davis, P., Feltwell, T., Fraser, A., Gentles, S., Goble, A., Hamlin, N., Harris, D., Hidalgo, J., Hodgson, G., Holroyd, S., Hornsby, T., Howarth, S., Huckle, E. J., Hunt, S., Jagels, K., James, K., Jones, L., Jones, M., Leather, S., McDonald, S.,

McLean, J., Mooney, P., Moule, S., Mungall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'Neil, S., Pearson, D., Quail, M. A., Rabinowitsch, E., Rutherford, K., Rutter, S., Saunders, D., Seeger, K., Sharp, S., Skelton, J., Simmonds, M., Squares, R., Squares, S., Stevens, K., Taylor, K., Taylor, R. G., Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward, J., Volckaert, G., Aert, R., Robben, J., Grymonprez, B., Weltjens, I., Vanstreels, E., Rieger, M., Schafer, M., Muller-Auer, S., Gabel, C., Fuchs, M., Dusterhoft, A., Fritzc, C., Holzer, E., Moestl, D., Hilbert, H., Borzym, K., Langer, I., Beck, A., Lehrach, H., Reinhardt, R., Pohl, T. M., Eger, P., Zimmermann, W., Wedler, H., Wambutt, R., Purnelle, B., Goffeau, A., Cadieu, E., Dreano, S., Gloux, S., Lelaure, V., Mottier, S., Galibert, F., Aves, S. J., Xiang, Z., Hunt, C., Moore, K., Hurst, S. M., Lucas, M., Rochet, M., Gaillardin, C., Tallada, V. A., Garzon, A., Thode, G., Daga, R. R., Cruzado, L., Jimenez, J., Sanchez, M., del Rey, F., Benito, J., Dominguez, A., Revuelta, J. L., Moreno, S., Armstrong, J., Forsburg, S. L., Cerutti, L., Lowe, T., McCombie, W. R., Paulsen, I., Potashkin, J., Shpakovski, G. V., Ussery, D., Barrell, B. G., Nurse, P., & Cerrutti, L. 2002, "The genome sequence of *Schizosaccharomyces pombe*", *Nature*, vol. 415, no. 6874, pp. 871-880.

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.*, vol. 116, no. 1, pp. 177-192.

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.*, vol. 125, no. 2-3, pp. 185-195.

Wrigley, N. G., Heather, J. V., Bonsignore, A., & De Flora, A. 1972, "Human erythrocyte glucose 6-phosphate dehydrogenase: electron microscope studies on structure and interconversion of tetramers, dimers and monomers", *J Mol Biol*, vol. 68, no. 3, pp. 483-499.

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*, vol. 299, no. 5615, pp. 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.*, vol. 264, no. 30, pp. 17681-17690.

Yang, S., Yu, X., Galkin, V. E., & Egelman, E. H. 2003, "Issues of resolution and polymorphism in single-particle reconstruction", *J Struct.Biol.*, vol. 144, no. 1-2, pp. 162-171.