Soluble Expression of *Plasmodium falciparum* Glutamine Synthetase and Three-Dimensional Structure by Single Particle Reconstruction

by

Satishkumar Ishverlal Patel

A dissertation submitted in fulfillment of the requirements for the
Degree of Master of Science in the
Department of Integrative Biomedical Sciences

Faculty of Health Sciences
UNIVERSITY OF CAPE TOWN

Supervised by: **Professor Trevor Sewell**

October 2015

Declaration

I declare that this dissertation is based on my own work, both in concept and execution, and that, apart from the normal guidance from my supervisor, I have received no assistance, except where acknowledgements indicate otherwise. I also declare that all sources that I have used have been indicated and acknowledged by means of complete references. I can also state that this work has not been submitted for any degree or examination in any other university.

I, Satishkumar Ishverlal Patel, hereby grant the University of Cape Town free licence to reproduce this thesis in whole or in part, for the purpose of research.

Signed by candidate

Signature Removed

Signature: ……

Dated: …27 OCTOBER 2015

Acknowledgements

I would like to formally express my gratitude to the people that have enabled this journey to be undertaken in the fullness of time.

Firstly, to Professor Vernon Coyne, as the Head of the Department of Molecular and Cell Biology, for innovating the notion of a probationary entrance into the field of Biological Sciences for me and to his academic staff for the patience and encouragement to stay the course.

My sincere thanks to research officer, Mrs. Diane James, for identifying my plight in the early days of the probationary period, and boldly suggesting the need to bring in “reinforcements” to accelerate my learning curve.

I would like to thank my supervisor, Professor Trevor Sewell, for his efforts to accommodate my entrance into the structural biology group. To the past and present members of our laboratory family, without whose active support this project would not have been successful: Dr. Brandon Weber, for providing hope and optimism throughout each stage and always being there for support; Dr. Jason van Rooyen for grounding me in the ways of science and introducing me to Linux and his wealth of knowledge on glutamine synthetase; Dr. Serah Kimani and Dr. Jennifer Thompson for providing their helpful insights on the laboratory benches; Mohamed Jaffer for help on the electron microscope; Dr. Jeremy Woodward and Dr. Ndoriah Thuku for sharing their practical experience on single particle reconstruction and to Mr. Sean Karriem for ensuring security of my laboratory data in the midst of the chaos during the relocation.

I am grateful for the high-speed crystal trials undertaken by Dr. Amit Sharma from the ICGEB, Delhi and Dr. Nathan Van Wyk, from the department of Chemical Engineering for the use of their prized PCR machine for the melt curves.

My thanks to the colour that was added at break-times with interesting insights, which helped me get a sense of biology in the light of evolution. For this a special thanks to Dr Jeremy Woodword, Dr. Brandon Weber and Dr. Ndoriah Thuku for the sharing of their views so candidly.

I would also like to acknowledge my wife, Pratibha, for never giving up during this time and reminding me to finish what I started so as to set an example to our son Mitesh and daughter Aarti.

Abstract

Malaria infection caused by the apicomplexa pathogen *Plasmodium falciparum* has a high rate of resistance to existing anti-malarial drugs. The World Health Organisation recommended interventions are unlikely to eliminate the growth of resistance and it would therefore be prudent to continue the search for new drug targets for the continued combatting of malaria.

Plasmodium falciparum is parasitic on the host for its metabolites and therefore inhibiting the transportation of glutamine from the host, has long been considered a potential strategy for combating the spread of infection. The recently sequenced *Plasmodium falciparum* genome has however shown that pathways for independent survival are also conserved. Therefore, combating the spread of *Plasmodium falciparum* in the human host, in addition to inhibiting the transportation of glutamine, will also require the inhibition of the *de novo* expression of essential amino acids within the *Plasmodium falciparum* cell. This could be achieved by inhibiting the glutamine synthetase gene, which is an essential step in the tri-carboxylic acid cycle.

The adenine and thymine rich *Plasmodium falciparum* glutamine synthetase gene was codon optimized for expression in *Escherichia coli*. Soluble over-expression was achieved by cold-shock expression with vector pCold I (Takara) in *Escherichia coli* (DE3). A hexa-histidine tag enabled a single step Ni-affinity chromatography purification that produced a homogenous, catalytically active enzyme.

Electron microscopy showed the glutamine synthetase of *Plasmodium falciparum* monomer forms two 6-fold symmetric rings that stack back-to-back with dihedral symmetry. An atomic model was built based on the homology with type I glutamine synthetases of known structure. This model indicated conservation of the structure with type I glutamine synthetase and locates a sequence insertion of 39 amino acids in the outer surface of the dodecamer.

The study successfully expressed soluble *Plasmodium falciparum* glutamine synthetase. A negative stained 3-D reconstruction confirmed that the protocol used for the purification of *Plasmodium falciparum* glutamine synthetase, produced a homogenous dodecamer enzyme, hence paving the way for further structural studies.

Abbreviations

A	Å	angstrom
	AMP	adenosine monophosphate
	APS	ammonium per sulphate
	AT	adenine & thymine
	ATP	adenosine triphosphate
	ADP	adenosine diphosphate
B	bp	base pairs
C	CTF	Contrast Transfer Function
D	DEAE	diethylaminoethyl
	DMSO	dimethyl sulphoxide
	DNA	Deoxyribonucleic acid
E	EDTA	Ethylenediaminetetraacetic acid
F	FFT	Fast Fourier Transform
	FSC	Fourier Shell Correlation
G	GS	glutamine synthetase
	GUI	graphical user interface
H	hr	hour
I	IPTG	isopropyl-beta-D -thiogalactopyranoside
K	kDa	kilo Dalton
L	L	litre
	LA	Luria Agar
	LB	Luria Broth
M	M	Molar
	mg	milligram
	min	minute
	ml	millilitre
	mm	millimeter
	mM	millimolar
N	n	nano
	NADP	Nicotinamide adenine dinucleotide phosphate
	NADPH	Nicotinamide adenine dinucleotide phosphate-oxidase
	NMR	Nuclear Magnetic Resonance
P	PCR	polymerase chain reaction
	PDB	Protein Data Base
	*Pf*GS	*Plasmodium falciparum* glutamine synthetase
	P_i	inorganic phospate
R	RMSD	root-mean-square deviation
S	SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
T	TE	Tris / EDTA buffer
	Tris	Tris-hydroxymethyl aminomethane
U	UV	ultraviolet
µ	µ	micro
	µg	microgram
	µl	microliter
	µmol	micromole
	µM	micromolar

Table of Contents

List of Figures

List of Tables

Chapter 1 – Introduction

Malarial infection is caused by *Plasmodium*. Five species infect humans, namely *falciparum, vivax, ovale, knowlesi and malariae,* of which *Plasmodium falciparum* is the severest form and is responsible for almost all deaths. Many of the current drugs for the treatment of malaria are already ineffective in many regions of the world and *Plasmodium falciparum* induced malaria, can only be treated by Artemisinin-based combination therapies (ACTs). Furthermore, the Thai-Cambodian border region has recently confirmed resistance to Artemisinin. The World Health Organisation (WHO) has implemented interventions in terms of treatment strategies, to reduce the spread of resistance before it is globalized (WHO, 2011). This is due to the high rate of resistance to previous antimalarial drugs. As a result, Dr Margaret Chan, Director-General of the WHO has stated "The threat must be taken seriously" (WHO, 2011, p.5). Hence, an extensive global accord has been reached to invest and to implement the WHO recommended interventions. However, it is likely the growth of resistance would be slowed but not entirely eliminated. It is therefore be prudent to use this window to identify potential new drug targets for the continued combatting of malaria.

1.1 Dual Strategy for Combatting Spread of Malaria at Blood Stage

Plasmodium falciparum is parasitic on the host for its metabolites, however, it is also able to proliferate with only isoleucine provided in the growth media (Liu *et al*., 2006), which is indicative of having conserved the pathways for independent survival (Payne & Loomis, 2006). Completion of the *Plasmodium falciparum* genome shows the genes for the tricarboxylic acid cycle are conserved. Biochemical studies have indicated the metabolite, acetyl coenzyme A (acetyl-CoA), that normally activates the tricarboxylic acid cycle in the mitochondria, is absent in *Plasmodium falciparum.* Instead, the tricarboxylic acid cycle is activated by 2-ketoglutarate, generated from the intake of glutamine (Olszewski *et al*., 2010). Therefore, combatting of the spread of *Plasmodium falciparum* in the human host will require more than simply inhibiting of the transport pathway for metabolites to enter the *Plasmodium* cell. In addition, the inhibition of the *de novo* expression of essential amino acids within the *Plasmodium falciparum* cell is also required.

1.2 The Role of Glutamine Synthetase in Nitrogen Metabolism

1.2.1 Nitrogen Assimilation

In virtually all cellular activity, glutamate and glutamine serve as nitrogen-donors for biosynthetic reactions (Merrick & Edwards, 1995). Bacteria are able to scavenge numerous nitrogen compounds, but ammonia is considered to support the fastest growth rates (Reitzer, 2003).

1.2.2 Ammonia assimilation in Bacteria

Ammonia is assimilated into glutamine and glutamate by two major pathways - glutamine synthetase/glutamate synthetase (GS/GOGAT) pathway and the glutamate dehydrogenase (GDH) pathway.

The GS/GOGAT pathway, being the more important and ubiquitous pathway in bacteria, results in the production of glutamate from ammonia and 2-ketoglutarate in bacteria, as depicted Figure 1.1 and 1.2 by each of the enzymatic reactions in the respective biosynthetic direction.

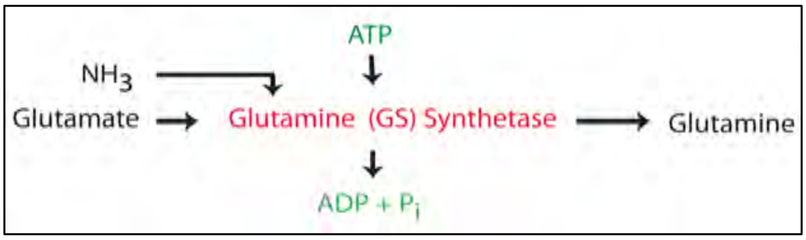


Figure 1.1 Bio-synthetic glutamine synthetase reaction of the GS/GOGAT pathway. Glutamate and ammonia with ATP yield glutamine, ADP and inorganic phosphate.

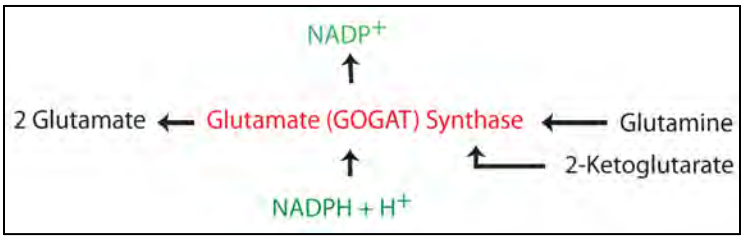


Figure 1.2 Bio-synthetic glutamate synthase reaction of the GS/GOGAT pathway. Glutamine and 2 ketoglutarate with NADPH and H^+ yield 2 Glutamate and NADP.

These pathways are ubiquitous in bacteria and the contribution of each of the pathways in assimilating ammonia and hence, nitrogen transportation, may differ for each organism (Merrick & Edwards, 1995). In both *Clostridium pasteurianum* and *Streptomyces clavuligerus* ammonia assimilation occurs exclusively via the GS/GOGAT pathway (Dainty, 1972; Brana *et al.*, 1986). On the other hand, ammonia assimilation occurs exclusively via the glutamine dehydrogenase pathway for *Streptococcus bovis* (Chen & Russell, 1989). The biosynthetic reaction for this pathway is indicated in Figure 1.3.

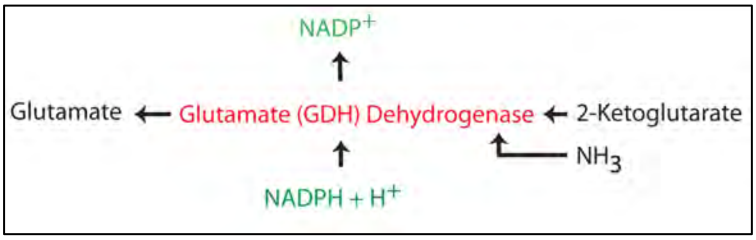


Figure 1.3 Bio-synthetic GDH reaction. Ammonia and 2 ketoglutarate with NADPH and H^+ yield Glutamate and NADP.

Nitrogen requirements for the composition of nitrogen containing compounds essential to life are derived almost exclusively from glutamate and glutamine (Reitzer & Magasnik, 1987), where:

- Glutamate provides the alpha-amino group for synthesis of amino acids, half the nitrogen for pyrimidine, purine and imidazole rings.
- Glutamine provides nitrogen for the synthesis of amino acids, p-aminobenzonate and NAD and the nitrogen for completing the formation of purines, pyrimidines, histidines and tryptophan (Wohlheuter *et al*., 1973)

Extensive studies conducted on *Escherichia coli* indicate that, at high concentrations of ammonia in the environment, sufficient quantum is able to diffuse across the cytoplasmic membrane. This is required for proliferation, however, when diffusion becomes limiting, an ammonium transporter (AmtB) is synthetized (Burkovski, 2003). In the presence of high concentrations of ammonia, GS is suppressed and glutamate dehydrogenase is the primary pathway to assimilate ammonia. In the presence of low concentrations of ammonia (<1mM), the GS/GOGAT pathways dominate (Merrick & Edwards, 1995; Burkovski, 2003). Therefore, in cells grown in nitrogen-limited media, glutamine synthetase is the dominant ammonia-assimilating enzyme.

Targeting the inhibition of glutamine synthetase with glutamate analogues in plants, showed a rapid accumulation of ammonia and caused plant death. However, the inhibitors were found to be equally effective on weeds and crops (Logusch *et al*., 1991). This inability of the inhibitors used to distinguish between weeds and crops has resulted in limited use of glutamine synthetase inhibitors as a direct herbicide. Nevertheless, ongoing structural studies by Occhipinti *et al*., (2009) to identify selective glutamine synthetase inhibition remains an active research area. Similarly, the glutamate analogue, L-methionine-S-sulfoximine (MetSOx) was found to rapidly inactivate purified *Mycobacterium tuberculosis* glutamine synthetase, and was shown to block the growth of the pathogen (Harth & Horwitz, 1999). Screening with ATP analogues for possible inhibition by irreversibly binding to the active site in glutamine synthetase is currently being pursued as a possible drug target for the *Mycobacterium tuberculosis* pathogen (Gising *et al*., 2012).

1.2.3 Glutamine Synthetase

As a result of the importance to nitrogen metabolism in bacteria, glutamine synthetase is the only known biosynthetic reaction to the formation of glutamine and has therefore been extensively studied and characterized (Brenchley *et al*., 1975; Fuchs & Keitster, 1980; Backman *et al*., 1981; Bhatnagar *et al*., 1986; Bohannon & Sonenshein, 1989; Fisher & Wray, 1989; Graziano Pesole *et al*., 1995). Glutamine synthetase is generally classified in four main classes, GS I, glnT, GS II, and GS III. To simplify discussion, it will be assumed that the glnT class is a subset of the GS I classification.

The length of the monomer enables one to distinguish the family to which the glutamine synthetase belongs. The length of the three main classes mentioned above average 360, 450, and 730 amino acids respectively. All glutamine synthetases are composed of two closed ring structures with the active sites formed between the protomers (Berman *et al*., 2000).

The physiologically relevant reaction arising from GS is the formation of glutamine as indicated in Figure 1.1 and 1.2. ATP first binds, followed by glutamate to form an intermediate γ-glutamyl phosphate and ADP. The ammonium ion then binds and attacks the intermediate γ-glutamyl phosphate to form glutamine and an inorganic phosphate (Liaw *et al*., 1995). ADP and the inorganic phosphate do not dissociate until the ammonium ion binds and the glutamine is released (Eisenberg *et al*., 2000).

The regulatory mechanism for GS enzyme activity is not uniform and can vary between tissue types for the same organism (Merrick & Edwards, 1995). GS expression level is highly regulated by nitrogen starvation, which has been studied extensively in *Escherichia coli*. A further cascade regulatory system through the adenylylation and de-adenylylation of a tyrosine associated with each active site has also been extensively studied in *Escherichia coli*, but its allosteric impact of the active site remains largely unknown.

1.2.4 Glutamine Synthetase Regulation

Review on glutamine synthetase regulation by Woods and Reid, (1993) indicates there is no single regulatory mechanism at the transcriptional and post-transcriptional level.

Post-translational modifications also regulate glutamine synthetase activity. In most cases, glutamate synthetase enzyme is regulated in response to nitrogen but it appears that there are different regulatory mechanisms that ensure optimization of the use of nitrogen substrates.

At least three post-transcriptional regulations of glutamine synthetase are well documented:

- Adenylylation and subsequent de-adenylylation of a tyrosine residue,
- Presence of divalent cation, Mg^{++} or Mn^{++}, at the active sites,
- Attachment of multiple end products of glutamine metabolism.

1.2.4.1 Adenylylation inhibition of Glutamine Synthetase

Glutamine synthetase obtained from nitrogen-starved cells is less sensitive to feedback inhibition than glutamine synthetase that is grown in a nitrogen-rich medium. This differential inhibition is the direct result of the covalent addition of an AMP moiety to a conserved tyrosine residue of each subunit of the glutamine synthetase enzyme (Woods & Reid, 1993). Once adenylylated, that specific subunit is inactive such that the enzyme can exist in a range of activity states within the cell (Stadtman, 2001). Adenylylation occurs in response to high nitrogen conditions, and progressive de-adenylylation occurs as nitrogen becomes more limiting, thereby activating the enzyme.

The regulation of glutamine synthetase activity through adenylylation and de-adenylylation is accomplished by the interaction of three proteins:

- uridylyltransferase / uridylyl-removing enzyme,
- the signal transduction protein P_{II}
- and adenylytransferase

Figure 1.4 depicts the complexity of this bicyclic cascade regulation of glutamine synthetase activity.

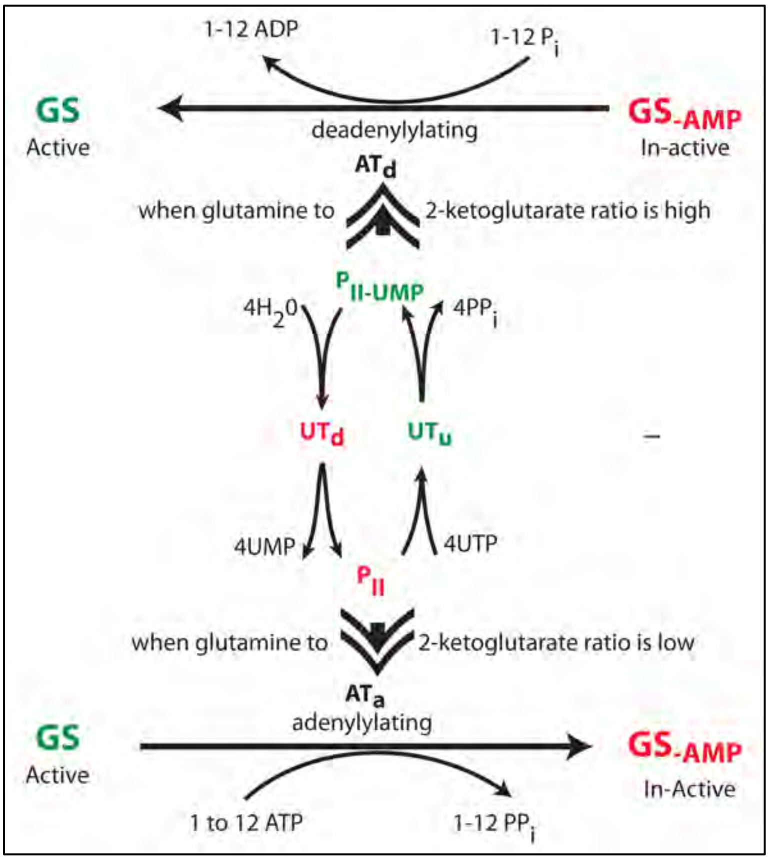


Figure 1.4 Bi-cyclic cascade regulation of glutamine synthetase activity with 12 subunits. Interrelationship between the inter-conversion of the regulatory protein between uridylylated (P_{II}-UMP) and unuridylylated (P_{II}) forms, and the cyclic interconversion of GS between adenylylated (GS-AMP) and unadenylylated forms. AT_a and AT_d denote the adenylylation and de-adenylylation sites of adenylyltransferase, respectively; UT_d and UT_u denote the deuridylylation and uridylylation sites of uridylyltransferase, respectively. (*Adapted from Merrick & Edwards, 1995*)

The enzymes catalysing the adenylylation and de-adenylylation reactions are also subject to fine metabolic control in themselves. A high intracellular concentration of glutamine activates the uridylyl-removing enzyme, which causes the de-uridylylation of the regulatory protein P_{II}. This unmodified P_{II} interacts with adenylyltransferase, which in turn catalyzes the adenylylation of glutamine synthetase, thus resulting in an inactive form of the glutamine synthetase subunit (Merrick & Edwards, 1995).

Conversely, a high intracellular concentration of 2-ketoglutarate relative to glutamine activates uridylyltransferase, transfers a UMP group to each subunit of the regulatory protein P_{II} to form P_{II}-UMP. The interaction between adenylyltransferase and P_{II}-UMP, then catalyses the removal of AMP from glutamine synthetase resulting in an active form of glutamine synthetase (Merrick & Edwards, 1995; Radchenko *et al*., 2013).

Not all glutamine synthetase enzymes, from the different organisms, have a conserved adenylylation loop and those organisms are therefore reliant on other regulatory mechanisms.

1.2.4.2 Cation Mg^{++} / Mn^{++} attachment

Abell *et al.,* (1995) note that "the enzyme requires two divalent metal ions in each subunit for catalysis, which are distinguished from one another by their dissociation constants. The more tightly bound metal ion is required to keep the enzyme in its catalytically active conformation while the less tightly bound metal ion is thought to facilitate nucleotide binding". In the review paper on structure-function relationships of glutamine synthetase Eisenberg *et al.*, (2000) also note the importance of divalent metal ions (either Mg^{++} or Mn^{++}) has been recognized for both activity and structural stability. Activity can vary depending on which divalent metal is bound (Mg^{++} or Mn^{++}) to the enzyme, however, it is generally accepted that Mg^{++} is physiologically more prevalent.

1.2.4.3 Feedback regulation

Glutamine synthetase activity is subject to cumulative feedback inhibition by the products of glutamine metabolism. The adenylylated (active) form of glutamine synthetase is inhibited by alanine, glycine, histidine, tryptophan, CTP, AMP, carbamoyl phosphate and glucosamine-6phosphate (Reitzer, 2003). Each inhibitor was found to decrease glutamine synthetase activity partially so that the residual activity in the presence of several inhibitors equalled the product of the individual residual activities. This was interpreted as the result of each inhibitor acting/binding at a different site on the enzyme. Acting together, therefore, the feedback products were found to abolish activity.

However, alanine, serine and glycine have been seen in crystal structures to bind to the glutamate substrate (S. Liaw, *et al.*, 1993a; S. Liaw, *et al.*, 1993b; S. Liaw *et al.*, 1994; S. Liaw & Eisenberg, 1994). Similarly, GDP, ADP and AMP have been seen to bind to the ATP site by Eisenberg *et al.* (2000), suggesting that there may be a simpler mechanism for feedback control than that of cumulative inhibition from separate sites.

1.2.4.4 Transcriptional regulation

Woods and Reid (1993) elucidate the regulation of glutamine synthetase in a range of organisms from enterobacteriaceae *(Escherichia coli, S. typhimurium* and *K.aerogenes*) *Bacillus subtilis, Clostridium acetobutylicum, Bacteriodes fragilis, Butyrivibro fibrisolvens, Actinomycetes and Rhizobiaceae.* In general, the expression of glutamine synthetase is regulated by nitrogen. However, it appears that bacteria that have more than one glutamine synthetase encoded in the genome, the GSI coded gene is expressed constitutively but with post-translational regulation by adenylylation and the GSII coded gene is regulated by nitrogen.

In the case of *Escherichia coli*, the glutamine synthetase gene has three promoters, two upstream and one downstream of the glutamine synthetase DNA sequence. In addition, there are two nitrogen regulatory genes downstream of the glutamine synthetase DNA sequence, which also form part of the overall mechanism for regulating glutamine synthetase levels.

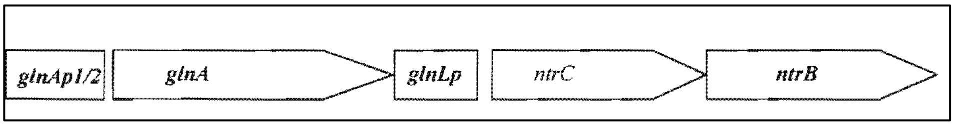


Figure 1.5 *Escherichia coli* glnA (glutamine synthetase) DNA map showing upstream and downstream regulatory sequences for glnA (glutamine synthetase). The operon has three promoters, glnAp1, glnAp2 and glnLp, and further downstream two nitrogen regulatory genes.

The three promoters shown in relation to glnA (glutamine synthetase DNA sequence) in Figure 1.5 enable the cell to maintain a low level glutamine synthetase (product of glnA) and NR$_I$ (product of ntrC) during growth with an excess of nitrogen and to increase the level of glutamine synthetase and NR$_I$ rapidly in response to a low level of nitrogen (Reitzer & Magasnik, 1987). In cells growing under carbon limitation, but with an excess of nitrogen, glnA is transcribed from glnApl, which is partially repressed by NR$_I$, resulting in a low level of glutamine synthetase. The cells also maintain a low level of NR$_I$, through transcription initiated at glnLp (Reitzer & Magasnik, 1987).

In the case of *Escherichia coli* the following key genes and their associated products that are involved in transcriptional regulation are listed in Table 1.1.

Gene	Product
glnA	glutamine synthetase
glnB	P$_{II}$
glnD	uridylyltransferase / uridylyl removing enzyme
glnE	Adenylyltransferase
ntrB	nitrogen regulator I (NR$_{I)}$
ntrC	nitrogen regulator II (NR$_{II)}$
rpoN (glnF, ntrA)	σ^{60}

Table 1.1 *Escherichia coli* regulatory genes and product involved for glutamine synthetase regulation (reproduced from Reitzer & Magasnik, 1987)

The regulation of the transcription of glnA in response to the availability of nitrogen is achieved through the action of uridylyltransferase/uridylyl-removing enzyme (P_{II}) nitrogenregulator I (NR_{I},), and nitrogen regulator II (NR_{II}). This requires core RNA polymerase and σ^{60}. When the intracellular concentration of 2-ketoglutarate is high, uridylyltransferase converts P_{II} to P_{II}-UMP. In the absence of P_{II}, NR_{I} catalyses the conversion of NR_{I} to NR_{I}-phosphate, and this phosphorylated protein then activates the initiation of transcription at the σ^{60}dependent promoter glnAp2. When the intracellular concentration of glutamine is high, uridylyl-removing enzyme converts P_{II}-UMP to P_{II}. This subsequently causes NR_{II} to remove the phosphate group from NR_{I}-phosphate, which in turn, halts the initiation of transcription from glnAp2.

In cells growing under carbon limitation, but with an excess of nitrogen, glnA is transcribed from glnApl, which is partially repressed by NR_{I}, resulting in a low level of glutamine synthetase (Reitzer & Magasnik, 1987). The cells also maintain a low level of NR_{I}, through transcription initiated at glnLp. Cells maintained in a nitrogen-deprived medium result in the activation of uridylyltransferase by the increase in intracellular 2-ketoglutarate. When cells are maintained in a nitrogen deprived medium, there is an increase in intracellular ketoglutarate which results in the activation of uridylyltransferase. The uridylyltransferase then converts P_{II} to P_{II}-UMP and allows NR_{II} to convert NR_{I}, to its active form, NR_{I}-phosphate. NR_{I}-phosphate can, in turn, fully activate the initiation of transcription at glnAp2 by σ^{60}-RNA polymerase. The level of intracellular NR_{I} is increased through the transcription of glnG initiated at glnAp2. The increase in the level of NR_{I} results in complete repression at glnApl and glnLp. A shift of the cells from nitrogen starvation to nitrogen excess causes the activation of uridylyl-removing enzyme through the increased intracellular glutamine and the removal of UMP from P_{II}-UMP by uridylyl-removing enzyme. The resulting P_{II} causes NR_{II} to remove the phosphate from NR_{I},-phosphate, thereby stopping the initiation of transcription from glnAp2. Maintaining growth in this medium results in decreased levels of glutamine synthetase and NR_{I} dilution, which eventually leads to sufficient lifting of the repression at glnApl and glnLp, to allow both glutamine synthetase and NR_{I} to be maintained at low levels (Reitzer & Magasnik, 1987).

Whilst the complexities of transcriptional feedback regulatory mechanisms have been extensively studied in enterobacteriaceae, especially *Escherichia coli,* studies in some of the other organisms have indicated interesting differences. In the case of *Bacteriodes* *fragilis,* GS activity is maintained only by transcription of glutamine synthetase which is regulated by nitrogen (Abratt *et al*., 1993). In the case of *Clostridium acetobutylicum,* there is an antisence coded glutamine synthetase sequence downstream of the glutamine sequence. Fierro-monti *et al.* (1992) found that there was reduced GS activity when expression of a antisense RNA was increased. The antisense RNA expression was repressed in nitrogen limiting media but induced in nitrogen-rich media. Hence the level of antisense RNA expression, determines the translation of glutamine synthetase thereby regulating glutamine synthetase activity

1.2.5 Glutamine Synthetase Structure and Phylogeny

Structure determination through electron microscopy and crystallography for Glutamate synthetase was started as early as 1969. Sub-atomic depositions in PDB indicate structures consisting of 10 or 12 monomers arranged in a double ring structure with dihedral group symmetry.

1.2.5.1 Typical Glutamine Synthetase structures

Since the recent deposition of the crystal structure of type III glutamine synthetase from *Bacteroides fragilis* (PDB ID: 3O6X; Rooyen *et al*., 2011) structures off all three glutamine synthetase families derived by X-Ray crystallography are now represented. The monomer protein sequences for each of the depositions below can be found in Appendix D.

GS Type	Organism	Monomer amino acid	Sequence used in alignment	PDB Deposition Variants
II	*Zea mays*	356	2D3A (2005)	2D3B (2005); 2D3C (2005)
	Saccharomyces cerevisiae	370	3FKY (2008)	-
	Canis familiaris	381	2UU7 (2007)	-
	Homo sapiens	384	2OJW (2007)	2QC8 (2007)
I	*Salmonella typhimurium*	468	1F52 (2000)	2GLS (1989); 2LGS, 1LGR (1994); 1F51, 1FPY (2000)
	Synechocystis sp	473	3NGO (2010)	-
	Mycobacterium tuberculosis	486	2BVC (2005)	1HTO, 1HTQ (2001); 2WGS, 2WHI (2009); 3ZXR, 3ZXV (2011)
III	*Bacteroides fragilis*	729	3O6X (2010)	-

Table 1.2 PBD glutamine synthetase depositions. The different colours reflect glutamine synthetase type. The typical difference between type II and type 1 is 100 amino acids and a much larger difference between type 1 and type III of around 250 amino acids.

Size of the monomer easily distinguishes glutamine synthetase type as can be seen in Table 1.2. The typical macromolecule structure in each of the GS type and their relative size differences are indicated in the Figure 1.6.

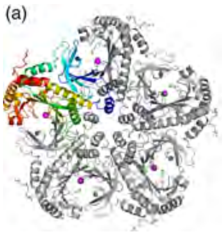
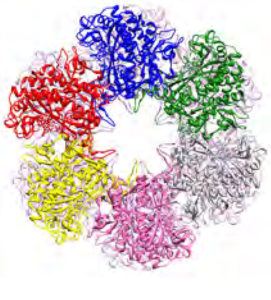
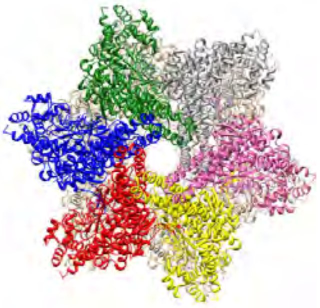
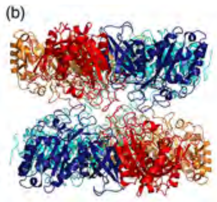
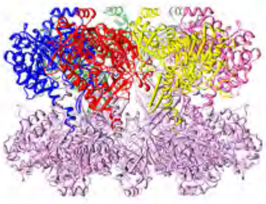
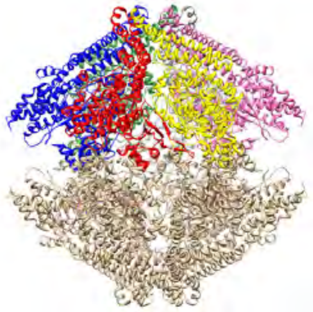
GS II	**GS I**	**GS III**
Homo Sapiens	*Salmonella typhimurium*	*Bacteroides fragilis*
ɸ115 x 90 Å	ɸ 140 x 110 Å	ɸ 180 x 170 Å
		
		

Figure 1.6 Glutamate synthetase structures of type II,I and III. The double ring structure is characteristic of glutamine synthetase. Figure a and b reproduced from W. Krajewski *et al*., 2008, GS II from PDB 1F52 and GS III from PDB 3O6X visualized in Chimera.

Rooyen *et al*., 2011 makes an interesting observation that in the case of GS III the orientation of the hexameric rings is inverted in comparison to other GS families.

1.2.5.2 Glutamine synthetase Phylogeny

GS I has been primarily associated with Prokaryotes, and GS II with Eukaryotes but more recently multiple isoforms can sometimes be found in the same organism. GS III on the other hand, is much longer than both GS I and GS II. Pesole *et al*. (1995) showed that the divergence between the GS I and GS II forms occurred about 2500 million years ago. The common ancestor of the GS II genes found in plants and animals precedes the divergence between plants and animals.

When considering all three GS families simultaneously, a high level of sequence similarity in five regions are conserved (Janssen *et al*., 1988), which would be helpful towards a structure prediction using existing models as templates. The length between regions II and III is also similar across all glutamine synthetase types. The adenylylation loop sequence is only conserved in GS I family of enzymes.

The sequences of the conserved regions are indicated in the Table 1.3 using *Salmonella typhimurium* as the template sequence. These conserved regions are indicated using Clustral W (Larkin *et al.*, 2007) pairwise alignment tool (but with the gaps removed) in Figure 1.7 to also show the conservation of length between region 2 and region 3 in all GS types.

R1	*50*-FDGSSIGGWKGINESDMVLMPDASTAVIDPFF	**R3**	*254*-TFMPKPMFGDNGSGMHCHMSL	**R5**	*357*-IEVRFPDPAANPYLCFA
R2	*211*-HEVATAGQNEVAT	**R4**	*337*-RNRSASIRIP	**A**[1]	*398*-IHPGEAMDKNLYDLPPEE

Table 1.3 Conserved DNA sequences in glutamine synthetase family using *Salmonella typhimurium* template for sequence reference.

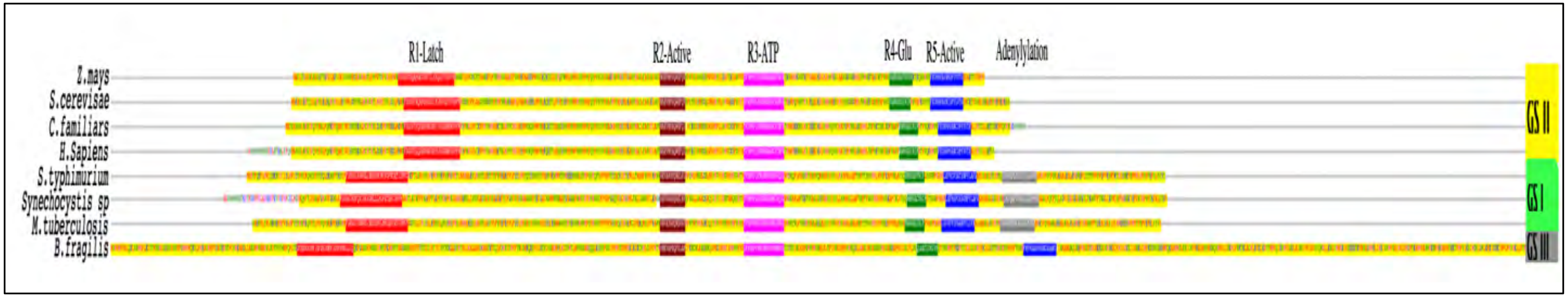


Figure 1.7 Glutamine synthetase multiple alignments of PDB depositions, centered on regions II and III. The Five conserved regions and adenylylation loop are indicated and shown according to the colour key: Red=Region1 – Latch; Brown=Region2 – Enzyme Active Site; Magenta=Region3 ATP binding site; Green=Region4 Glutamate Binding Site; Blue=Region5 Enzyme Active Site and Grey= Adenylylation Loop. The gaps have been removed in the alignment to indicate the length of the region between R2 and R3 is conserved in all GS types whilst GS I and GS II share a similar length between R2 through to R5.

---

[1] Adenylylation Loop

The multiple alignment scores between the GS's are indicated in Table 1.4.

GS TYPE	**Deposition & Chain Length**		**Organism**	**II**				**I**		
				2D3A	**3FKY**	**2UU7**	**2OJW**	**1F52**	**3NG0**	**2VBC**
II	**2D3A**	**356**	***Zea mays***							
	3FKY	**370**	***Saccharomyces cerevisiae***	57						
	2UU7	**381**	***Canis familiaris***	54	53					
	2OJW	**384**	***Homo sapiens***	54	54	96				
I	**1F52**	**468**	***Salmonella typhimurium***	21	23	23	24			
	3NG0	**473**	***Synechocystis sp***	22	24	21	22	57		
	2VBC	**486**	***Mycobacterium tuberculosis***	22	21	19	20	56	52	
III	**3O6X**	**729**	***Bacteroides fragilis***	16	15	14	14	19	17	16

Table 1.4 Sequence Identity between glutamine synthetases of different organisms. The sequence similarity between GS types is low (between 14 -24%), whereas in the same GS type the similarity increases to greater than 50%. GS I=Green;GS II=Yellow and GS III=Grey.

The sequence similarity within each type (highlighted in yellow and green in Table 1.4) is greater than 50%.

1.2.5.3 Glutamine Structure function

The five regions of high sequence similarity as shown in Table 1.3 have the following functional importance:

- Regions II and V – enzyme active site
- Region I – Latch region
- Region III – ATP binding site
- Region IV – Glutamate binding site
- Adenylylation loop – in GS I

Eisenberg *et al.* (2000) have extensively reviewed the structure of glutamine synthetase, describing the active site as a bi-funnel in which the glutamate and ATP bind at opposite ends. The two divalent cation binding sites (referred to as n1 and n2 cation) are found at the joint of the bi-funnel. Three glutamate side chains bind the n1 cation and help in stabilizing the protein. The n2 cation is involved in phosphoryl transfer held in place by two glutamate ligands and a histadine side chain. Pesole and Lanave (1991) showed that the cation ligands are highly conserved in glutamine synthetase.

Both the N-terminus and C-terminus of each monomer are helical. The two rings of the GS macromolecule are held together mainly by hydrophobic and hydrogen bonding interactions (Almassy *et al.*, 1986). In general, the N-terminus helix sits above the ring exposed to solvent, whilst the C-terminal helix is inserted into a hydrophobic hole in the subunit of the opposite ring.

The adenylylation loop containing a tyrosine residue is covalently modified by the addition of AMP. This loop sits just outside the bottom entrance of the bi-funnel.

The functional study of *Mycobacterium tuberculosis* glutamine synthetase by Eisenberg *et al*. (2000) postulates the workings of the latch region as follows:

1. Residues 323-330 form a loop that guards the glutamate entrance to the active site.
2. The loop closes the active site, shielding the γ-glutamyl phosphate intermediate from aberrant hydrolysis.
3. Glu-327 carboxylate forms a part of the ammonium site.
4. When the loop closes the glutamate entrance to the active site, Asp-50 deprotanates the ammonium ion, forming ammonia, which in turn attacks the γ-glutamyl phosphate intermediate thereby forming a tetrahedral intermediate, yielding glutamine.
5. The loop in the closed position is thought to increase the inter-subunit stabilization through interaction with Ser-52P or Ser-53P of the adjacent subunit.

Krajewski (2008) has crystallized *Mycobacterium tuberculosis* glutamine synthetase with different analogues in the binding sites. Some of the inhibitors identified from high throughout screening and subsequently optimized on *Mycobacterium tuberculosis* glutamine synthetase (glnA1) were carried out on whole cell *Mycobacterium tuberculosis* (Couturier *et al.*, 2015). None of the inhibitors was found to be active on the whole cell M. tuberculosis. Couturier *et al.* (2015) concluded that *Mycobacterium tuberculosis* glutamine synthetase (glnA1) is not a suitable anti-tubercular drug target. Further structural studies on *Mycobacterium tuberculosis* glutamine synthetase remains an ongoing research area.

1.3 Aims of this Study

Enabling the design of new drugs that will inhibit Plasmodium falciparum glutamine synthetase will require its structure. At present the structural knowledge of *Pf*GS is limited to its sequence. In this study we hope to gain further structural and other information that will contribute ultimately to the atomic resolution study that will be necessary for the interpretation of drug binding data.

The specific aims of this project were to:

a) Clone and produce soluble and active *Plasmodium falciparum* glutamine synthetase.
b) Purify the *Plasmodium falciparum* glutamine synthetase to homogeneity
c) To determine the 3-D structure of *Pf*GS by negative stain single particle electron microscopy.
d) Use homology modelling to generate a model for *Pf*GS to interpret the EM density.

The known *Plasmodium falciparum* genome sequence facilitated the synthesis of the gene encoding glutamine synthetase protein that was the starting point for this study. The work presented here demonstrates the expression and purification of a soluble, intact, active enzyme. This enzyme was further characterized with respect to its thermostability as a prelude to crystallization trials. Structural data in the form of negative stain electron micrographs were obtained and these enabled a three-dimensional reconstruction that demonstrated that the enzyme was a dodecamer.

Chapter 2 – Expression and purification of recombinant *Plasmodium falciparum* Glutamine Synthetase

This chapter describes the initial insoluble expression and the soluble expression of recombinant *Plasmodium falciparum* glutamine synthetase from a synthetically made DNA sequence. Soluble expression was followed by purification of *Plasmodium falciparum* glutamine synthetase protein.

2.1 Introduction

Glutamine synthetase sequence, which was found to be conserved (Berman *et al.*, 2000), in the *Plasmodium falciparum* genome (PLASMODB gene id: PF3D7_0922600) was used for synthetising the DNA sequence.

There were multiple unsuccessful attempts made to express soluble *Pf*GS using the expression plasmid pETM-11_D216 prepared by GeneArt[2]. Subsequent analysis identified a further key error in the deposited sequence of this enzyme in PLASMODB (Bahl *et al.*, 2003). Following the alignment of the sequence of GS I from *Salmonella typhimurium,* it was noticed that the α-helix central to the inter-ring interface, was not conserved. The significance of this difference became clear once the important structural role of this element was understood. On re-examination of the sequence, the presence of an intron was missed which resulted in the reading of a stop codon at residue 511, which would have also have caused a 17 amino acid shorter protein compared to the native length. At the same time this discrepancy was identified, the PLASMODB website also reflected the corrected *Pf*GS sequence.

2.2 Materials and Methods

Growth media, stock solutions and buffer compositions are all detailed in Appendix A.

2.2.1 *Pf*GS DNA Sequence

2.2.1.1 Codon optimization of DNA sequence

Different organisms show marked differences in codon preferences for their gene coding. The *Plasmodium falciparum* genome is a very AT rich sequence (~80%) and is known to be a challenge for recombinant soluble over expression in *Escherichia coli*. The host codon

---

[2] A biological venture for commercial gain in the USA.

preferences and tRNA abundance relative to the requirements for the heterologous gene may vary vastly with the host, which could result in either slowing down or absence of over expression. (Bonekamp & Jensen, 1988; Robinson *et al*., 1984).

Expression in *Escherichia coli* for the recombinant protein was preferred due to the availability and low cost of the expression system. The *Plasmodium falciparum* glutamate synthetase sequence was codon optimized by GeneArt for expression trials taking into account codon preferences (Ikemura, 1981), where possible, for *Escherichia coli* as the host strain. The native and codon optimized sequences, together with the translated protein sequence are included in Appendix B.

2.2.1.2 Plasmids used for synthesized gene from Gene Art

GeneArt (USA) synthetized the codon optimized *Pf*GS gene and delivered it in two plasmid constructs, a cloning plasmid pMA with 4018 base pairs (Figure 2.1) and an expression plasmid pETM-11_D216 with 7005 base pairs (Figure 2.2).

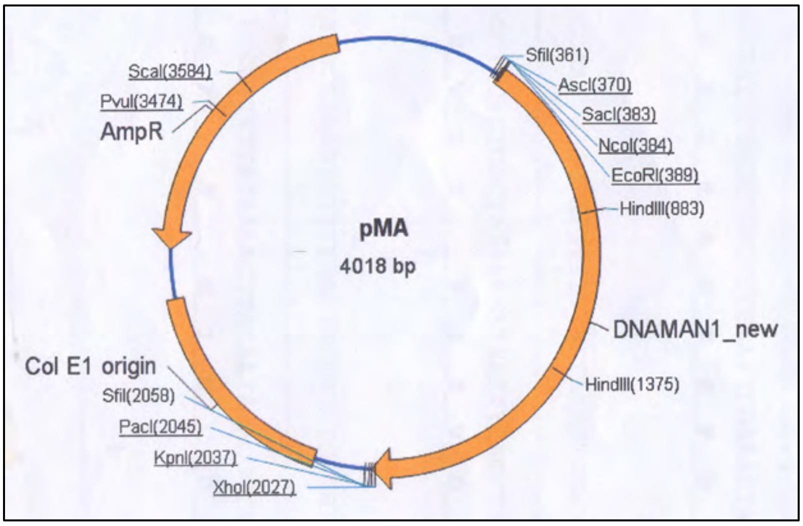


Figure 2.1 Map of pMA Cloning plasmid used for insertion of synthesised gene. pMA is ampicillin resistant. (*Reproduced from GeneART documentation*)

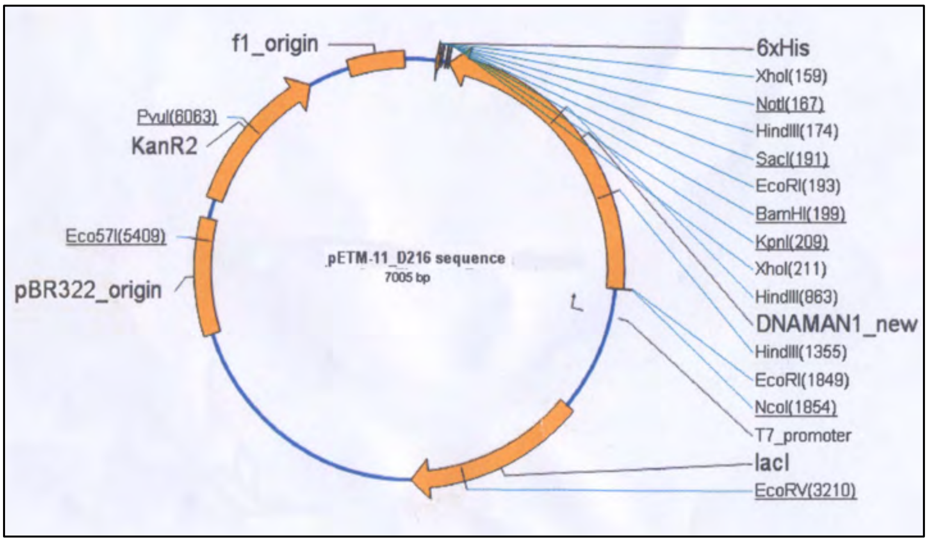


Figure 2.2 Map of pETMM1_D216 expression plasmid used for the insertion of synthesised gene. pETM-11_D216 is kanamycin resistant. (*Reproduced from GeneART documentation*)

Five micrograms of each plasmid construct was received and re-suspended in 50 µl TE buffer (2.2.9 Appendix A). The pMA (GeneArt) plasmid is ampicillin resistant whilst the pETM11-D216 (GeneArt) is kanamycin resistant.

2.2.1.3 pCold I *Plasmid*

Reducing the incubation temperature sufficiently induces a group of “cold shock” proteins in *Escherichia coli* to be expressed. The pCold I (Takara) plasmid (Figure 2.3) utilizes one of the cold shock genes as a promoter for inducing over-expression. Takara markets four kinds of pCold plasmids, whose arrangements vary in the existence of translation enhancing element (TEE), hexa-histidine tag sequence and Factor Xa cleavage site.

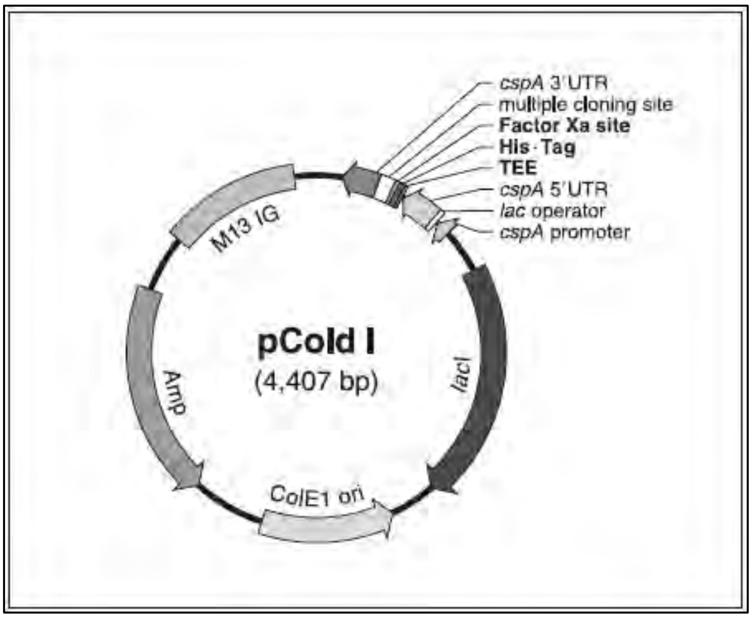


Figure 2.3 pCold I plasmid map used for the cold shock over-expression. pCold I is ampicillin resistant. (*reproduced from GeneART documentation*)

The pCold I vector shown in Figure 2.3 was selected for expression trials, with the view of using the N-terminal hexa-histidine tag towards a single step purification process. The pCold I vector is optimized for expression at 15°C.

2.2.2 Transformation of Competent Cells

Competent cells that had been stored at -20°C were thawed on ice for 10 minutes. Plasmid DNA was (aim for 1-50 ng) added to 100 µl of competent cells. Incubated on ice for 20 minutes. Heat shock cells at 42°C for 5 minutes and quenched of ice for 2 minutes before adding 900 µl of culture media (LB). Incubated at 37°C or 60 minutes. Spread 100 µl cell onto LA plate with antibiotic selection. Incubated petri dish at 37°C overnight then stored plate at 4°C.

Competent cells of *Escherichia coli* strains BL21(DE3), BL21(DE3) pLysS, BL21(DE3) C41, and BL21(DE3) Rosetta Gami were used in the transformations. Three plasmids were used in the transformations, namely pMA (GeneArt), pETM11_D216 (GeneArt) and pCold I (Takara). Plasmids pMA and pCold I are ampicillin resistant and pETM-11_D216 is kanamycin resistant.

2.2.3 Expression Protocols

2.2.3.1 Standard expression protocol

A single colony was inoculated in 10 ml LB growth media with antibiotic selection and incubated at 37°C for 16 hrs to produce a starter culture. Expression trials were conducted in either 250 ml or 1000 ml growth media preparations with antibiotic selection inoculated with starter culture 1:500. Cell growth was monitored by measurement of absorbance at 600 nm. Expression was induced with 0.2 mM IPTG when absorbance at 600 nM reached 0.5. Incubation was continued for a further 16 hours. Cells were collected by centrifuging media at 4500 rpm and discarding the supernatant. At this stage, the cells were stored at -18°C.

Cells were re-suspended in a precooled (4°C) phosphate buffer at 20 mM pH7.5. Cell lysis was achieved by 15 seconds intermittent sonication for 4 minutes on ice using a microtip operating at a power output of 65W (Misonix Sonicator). The whole cell extract was centrifuged at 4°C for 20 minutes at 20000 x g and the cell free extract was collected.

2.2.3.2 Cold Shock Expression

The cold shock expression is similar to the standard expression protocol except for:

a) Induction takes place after growth culture was cooled for 30 minutes at 15°C.
b) Induction with IPTG at a final concentration of 0.4 mM in growth media.
c) Post induction incubation was for 24 hours at 15°C.

2.2.4 Standard Protein Protocols

2.2.4.1 SDS-PAGE

Visualisation of protein size was carried out according to Laemmli using a Mini-PROTEAN 3 cell from Bio-Rad with 10% separating gels. Pre-stained molecular markers (Fermentas Life Sciences) were included in the SDS-PAGE electrophoretic separations.

Gel separating solution was pipetted into an assembled glass slide and overlaid with 2 ml isopropanol. Polymerization was allowed to take place for 30 minutes. Removal of the overlay and pipette prepared stacking gel solution on top of the polymerized separating gel was followed immediately by inserting the gel comb, ensuring a neat fit and without trapping any bubbles. We allowed polymerization of stacking gel for 30 minutes and removed the comb carefully.

We assembled the glass slide into the running tank according to the manufacturers instruction and filled the tank with running buffer. Steps followed were:

Prepared 20 µl of sample volumes in fresh eppendorfs (16 µl sample/dH_2O with 4 µl 5 x sample buffer) and denatured protein for 5 minutes in boiling water before loading. Samples were loaded in respective lanes and order of loading was noted.

Ran the gel at room temperature at a constant 30 mA until detection of the blue front was within 5 mm of the slide edge. Disassembled the gel slide plates and carefully freed the gel and put into staining solution, ensuring gel was submerged for one hour with gentle agitation. The submerged gel was placed into destain solution for 12 hours with gentle mixing and then we stored the destained gel in 7% acetic acid solution until visualized in light box and findings recorded.

2.2.4.2 Bradford Protein quantification

Freshly prepared Protein Dye Reagent (Bio-Rad) was diluted 1:5 in dH2O before use. The Bradford microtitre plate procedure was used. 10 µl of protein sample in triplicate was loaded in the microtitre 96 well plate followed by 200 µl of diluted Protein Dye Reagent. Absorbance was read at 595 nm in a Multiskan spectrophotometer (Titertek, Helsinki) after 5 minutes.

A standard curve was prepared by replacing the sample with a known dilution series between 0.5 and 0 mg/ml that was made from 1 mg/ml of BSA stock. The 0 mg/ml (origin) was represented by a sample buffer. Plotting the absorbance against the BSA dilution series allowed the concentration of the sample to be computed.

2.2.5 Standard DNA Protocols

2.2.5.1 Agarose Gel electrophoresis

The DNA fragments resulting from the PCR amplification were visualized by using the standard 1% Agarose Gel electrophoresis technique in TAE (Sigma-Aldrich) buffer and using EtBr as the intercalating dye.

Samples for loading were prepared in a fresh eppendorf, by adding 2 µl DNA gel loading dye (Fementas), 2 µl of amplified DNA and 8 µl water. Together with the sample, a calibrated DNA ladder (Fementas) was also loaded in a separate lane.

The electrophoresis was carried out at 120V until visible dye line was approximately 75-80% of the way, down the gel. DNA bands were visualized under UV light and interpret bands using the calibrated DNA ladder as the reference.

2.2.5.2 Restriction Digests

All restriction enzymes and the respective buffers were obtained from NEB (New England BioLab). In a clean eppendorf we prepared a 20 µl double digest reaction mixture, by the addition of 1 µg DNA, 1 unit enzyme A, 1 unit enzyme B, 2 µl of the recommended buffer made up with dH_2O. The reaction mixture was all settled in the bottom of eppendorf by centrifuging briefly. The reaction was allowed to incubate for 1 hour at 37°C then heat inactivated at 65°C for 20 minutes. Cooling was carried out on ice for 10 minutes before storing at 4°C.

2.2.5.3 Plasmid Dephosphorylation

Dephosporylation of the digested plasmid was done using the calf intestine alkaline phosphatase. The manufacturers digest protocol was followed, whereby 3 µl de-phosphorylation buffer, 3 µl calf intestine alkaline phosphatase (Sigma-Aldrich) and 4 µl dH_2O was added to the completed plasmid digest, incubated for 30 minutes at 37°C and heat inactivated at 65°C for 20 minutes. The dephosphorylated plasmid was stored at 4°C.

2.2.5.4 Ligation Protocol

The ligation protocol assumed the plasmid was dephosphorylated. We added in a clean Eppendorf 25 ng DNA insert fragment, 100 ng dephosphorylated plasmid DNA, 2 µl ligation buffer (Fermentas), 1 unit of T4 DNA ligase (Fementas) and made up reaction volume to 20 µl using dH_2O. Reaction was allowed to incubate for 16 hours at 16°C and then heat inactivated at 65°C for 10 minutes. Cooling was carried out on ice for 10 minutes before storing at 4°C.

2.2.5.5 PCR Amplification

Amplification was done using Phusion® Hot Start high fidelity DNA polymerase kit (New England BioLabs). Reaction mixtures of 50 µl were prepared as recommended for the use of the polymerase kit.

Step	**Temperature**	**Duration**
Initial Denaturation	98 °C	30 seconds
25 cycles - Denaturation - Annealing - Extension	98 °C Primer T_m + 3 °C 72 °C	10 seconds 30 seconds 15-30 seconds / kb of DNA
Final Extension	72 °C	10 minutes
Hold	4 °C	

Table 2.1 PCR thermo cycling conditions recommended by the manufacturer. The lower of the two primers melt temperature was used and extension time rounded to the nearest 30 seconds using the lower end of the extension per kb.

The manufacturer set out recommended conditions for the thermo cycling is indicated in Table 2.1, which was used for the amplification.

2.2.6 DNA Primer Designs

*2.2.6.1 For Amplification of Pf*GS *from vector pMA (GeneArt)*

Insertion of *Pf*GS was chosen to be between Nde I and Kpn I in the multiple cloning site of the pCold I vector to ensure correct orientation and maximizing utilisation of existing restriction sites at the multiple cloning site as shown in Figure 2.4.

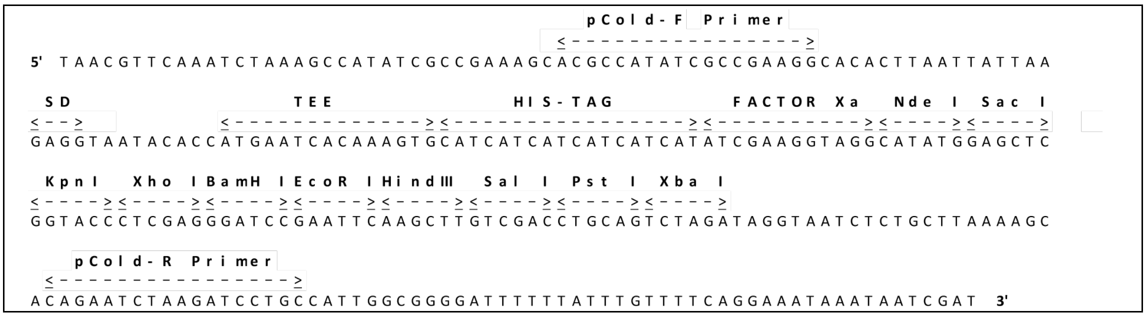


Figure 2.4 pCold I Multiple Cloning Site shown from 5' to 3'. Recommended forward and reverse amplification primers are shown.

Ensuring correct orientation during insertion and DNA reading frame alignment for *Pf*GS sequence required the introduction of a Nde I restriction site during PCR amplification.

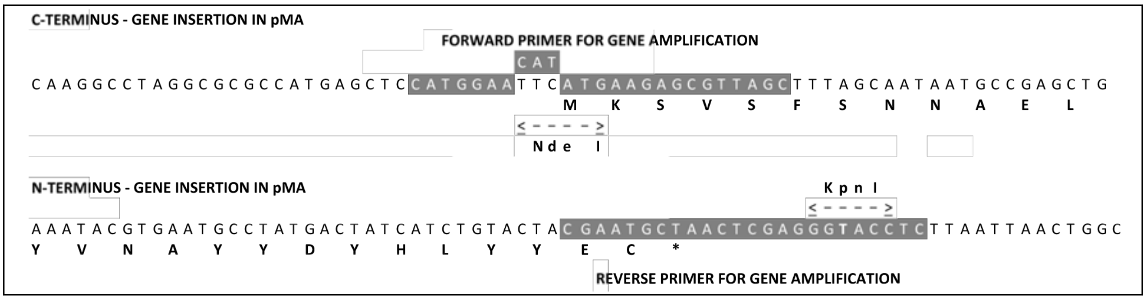


Figure 2.5 Primer design for amplification of *Pf*GS in pMA. DNA sequence in the vicinity of the N-terminus and C-terminus of *Pf*GS insertion region is indicated. To enable a Nde I restriction site to form a 3 codon, CAT insertion is required instead of TTC for the forward primer.

As there was no Nde I restriction site available in the pMA plasmid, (Figure 2.1), a site-directed mutagenesis of three codons (CAT) was introduced, as can be seen in Figure 2.5, to ensure a Ndel I site results during gene amplification. The Kpn I site already existed in both the pMA construct at the C-terminal end and at the pCold I multiple cloning site.

Primer	Sequence	Length (bp)	T_m (°C)
Forward	5' CATGGAACATATGAAGAGCGTTAGC	25	56.0
Reverse	5' GAGGTACCCTCGAGTTAGCATTCG	24	59.1

Table 2.2 Primers to amplify *Pf*GS from pMA vector (GeneArt).

The forward and reverse primers as indicated in Table 2.2 were synthesized by Integrated DNA Technologies and diluted to 100 µM stock.

*2.2.6.2 For Amplification of Pf*GS *from vector pCold I (Takara)*

Primers as recommended by Takara as shown in Figure 2.4 for the verification of the insert DNA sequence in pCold I were used.

Primer	Sequence	Length (bp)
Forward	5' ACGCCATATCGCCGAAAGG	19

Reverse	5’ GGCAGGGATCTTAGATTCTG	20

Table 2.3 Primers to amplify P*f*GS in pCold I vector (Takara).

The forward and reverse primers as indicated in Table 2.3 were synthesized by Integrated DNA Technologies and diluted to 100 µM stock.

2.2.7 DNA Purification

2.2.7.1 DNA from PCR amplification

In preparation for efficient restriction enzyme digest, the amplified DNA was purified using a proprietary NucleoSpin® Extract II kit. Essentially the DNA is bound to a silica membrane and contaminants removed by a few washing steps. The pure DNA is then eluted with a slightly alkaline buffer (Tris/HCl, pH 8.5). Dilution of the binding buffer determines the size of the DNA fragments that will bind. To exclude binding of less than 1000 bp of DNA, the binding buffer provided in the kit was diluted 1:5 with water.

2.2.7.2 DNA purification after Restriction Digest

The digested DNA was loaded on 1% Agarose gel and after the electrophoresis run, the required DNA bands were excised and placed into clean Eppendorf tubes. The excised agarose gel with DNA fragment was dissolved and DNA fragment extracted in accordance with the proprietary NucleoSpin® Extract II kit recommendations.

2.2.8 Protein Purification

Purification was achieved by immobilized metal affinity chromatography and competitive binding of imidazole eluted *Pf*GS.

2.2.8.1 Cell Lysis

The standard protocol for lysis by sonication was used.

Cells were re-suspended with 3 ml pre-cooled buffer[3] at 4°C per gram of cell weight. Power output was set at 65 W and sonication for 4 minutes with 15 seconds intermittent cycles (Misonix Sonicator) using a microtip. Cell free extract was collected following centrifuge at 4°C for 20 minutes at 20 000 x g.

It was important that the subsequent purification after cell lysis is done in quick succession and as far as possible below room temperature.

---

[3] Re-suspension buffer used was 20 mM Na_2HPO_4 buffered to pH7.5 with NaH_2PO_4 + 50 mM Imidazole + 0.1 mM $MnCl_2$.

2.2.8.2 Affinity Chromatography

A 5 ml His-trap Fast Flow™ (GE Healthcare) Ni-affinity column was loaded with filtered cell free extract after filtering with a 0.45 µm filter at 1 ml/min in a 4°C chamber. Two column washes (10 ml) were effected with pre-cooled wash buffer (50 mM imidazole). The loaded and washed affinity column was placed in a modular Gilson system, with Gilson 321 pump, Waters detector and Gilson fraction collector. Protein was eluted using a linear gradient of elution buffer (500 mM imidazole at a flow rate of 2.5 ml/min over 25 minutes. Fraction collector was set to collect in intervals of 1 minute.

Consolidation of the fractions was based on collecting the fractions around the peak absorbance tube and presence of GS activity using the ϒ-glutamyl transferase (GGT) assay.

The protein was exposed to room temperature only during the fractionation step. This exposure was further minimised by ensuring the buffers are pre-chilled and kept on ice during the entire fractionation and each fraction collection tube on completion was transferred immediately to ice.

GS activity was assayed by the γ-glutamyl transferase (GGT) assay as described by Bender *et al.*, (1977). Assay volumes were reduced to allow measurements in the Titertek Multiscan plate reader. In a clean eppendorf 66.7 µl of reaction mix, 8.33 µl of Glutamate (0.2mM) and 8.33 µl of sample were added. A blank reaction eppendorf is also included (i.e. sample replaced with dH_2O). The solution was incubated at 37°C and the stop reaction solution added after 5 minutes. The stopped reaction was then centrifuged at 5000 rpm for 5 minutes to remove any precipitants. The supernatant was loaded in the microtiter plate and absorbance at a wavelength of 540 nm recorded. The reaction product is glutamyl hydroxamate. One unit of GS activity is defined as production of 1 µmol of glutamyl hydroxamate per minute. 1 µmol of glutamyl hydroxamate is represented by 2.916 of absorbance units at wavelength of 540 nm.

2.2.9 Buffer Exchange and Protein Concentration

The Amicon® stirred cells device with a 100 kDa filter was used. The device used has a capacity of 50 ml.

The starting solution was diluted 1:1 with the final buffer and concentrated in the Amicon® stirred cell device to ~5 ml. The Amicom device was refilled until all diluted solution was loaded. The cycle of concentrating to 5 ml and reloading with the final buffer was repeated for two cycles. The final preparation was concentrated to around 2 mg/ml and flash frozen in Cryo Tubes under liquid nitrogen for storage.

For concentrations greater than 2 mg/ml the addition of 10% glycerol and 0.4 mM beta-octyl glucoside was made to the buffer.

2.2.10 Differential Scanning Fluorimetry

Differential scanning fluorimetry (DSF), using a SYPRO orange probe (Niesen *et al.*, 2007), was used to rank the thermal stability of *Pf*GS. DSF essentially monitors the unfolding of the protein in the presence of a fluoro probe, using a real time PCR instrument.

The ROTO-GENE 6000 PCR machine, having the necessary excitation and emission wavelengths but without refrigeration attachment, was available. The experiment started at 25°C. Each reaction volume was 40 µl, with 0.003 mg of *Pf*GS with recommended dilution of SYPRO orange and buffer conditions being tested.

2.2.11 Protein Sequencing

The synthesized primers, together with the isolated plasmid DNA, were sent to Macrogen (South Korea) laboratories for analysis by using the standard Sanger sequencing.

2.2.12 Mass Spectrometry

Sequencing of purified protein was done using MALDI-TOF and by BLAST using the trypsin digest peptide mass fingerprint. The services were outsourced to the Molecular and Cell Biology department at the University of Cape Town.

2.3 Results

2.3.1 Visualisation of Soluble Expression with pETM-11_D216 in *Escherichia coli*

Transformations into the four strains[4] of *Escherichia coli* were successful, as indicated by growth of anti-biotic (kanamycin) resistant colonies on LA-plates.

---

[4] BL21(DE3), BL21(DE3) pLysS, BL21(DE3) Rosetta-gami and BL21(DE3) Overexpress C41

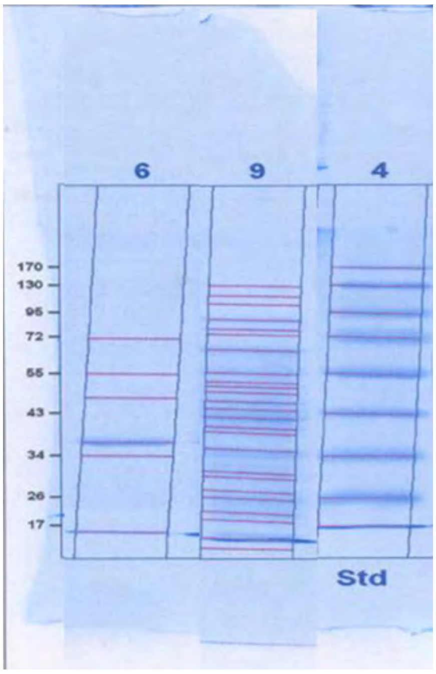


Figure 2.6 SDS-PAGE visualisation using Gel Doc XR™ imager of soluble expression in *Escherichia coli* . The lane headers 6 is strain BL21 (DE3), 9 is strain BL21 (DE3) pLysS and 4 is the ladder. No visual bands could be seen at the 63 Kda band, however the Gel Doc at low sensitivity emphasizes a potential band at this region (red line) in the case of BL21 (DE3) pLysS.

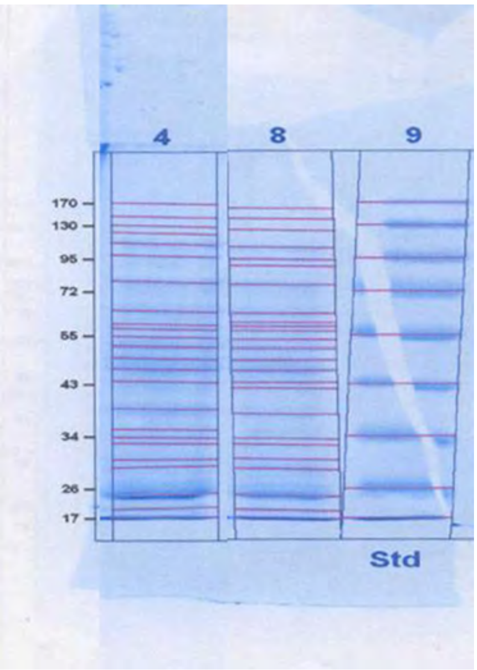


Figure 2.7 SDS-PAGE visualisation using Gel Doc XR™ imager of soluble expression in *Escherichia coli* . The lane headers 4 is strain BL21 (DE3) RossetaGami, 8 is strain BL21 (DE3) Overexpress C41 and 9 is the ladder. No visual bands could be seen at the 63 Kda band, however the Gel Doc at low sensitivity emphasizes a potential band at this region (red line) for both strains.

Over-expression in the 63 kDa (P*f*GS) band was expected. No band was visible in the expected range when viewed under the light box. The gel was then also analyzed using the

Gel Doc XR™ (Bio-Rad) imager and Quantity One Version 4.5.2 software, with the sensitivity set to low to allow low levels of detection. Three of the strains indicated a possible band around 63 kDa as indicated in Figure 2.6 and Figure 2.7 by the red lines drawn by the software in attempting to indicate an identified band.

Ni-affinity purification was then carried out on the cell free extract of *Escherichia coli* BL21 (DE3) expression which indicated no absorbance peaks beyond 30% of the elution buffer (150 mM imadazole) and no GS activity as measured by the GGT assay in any of the elution tubes. It was therefore concluded that no soluble expression of *Pf*GS was achieved.

2.3.2 Confirmation of DNA Sequence of the Insertion in pCold I

The returned sequence chromatograms from Macrogen were analyzed using software version 2.01 of Chromas (Goodstadt & Ponting, 2001) and the DNS sequence was confirmed to align with the codon optimized gene sequence, received from GeneArt for *Pf*GS.

2.3.3 Size of Amplified DNA Fragment from pCold I (*Pf*GS) construct

The expected fragment size of the amplified DNA fragment including the overhangs of both the primers was 1661 base pairs.

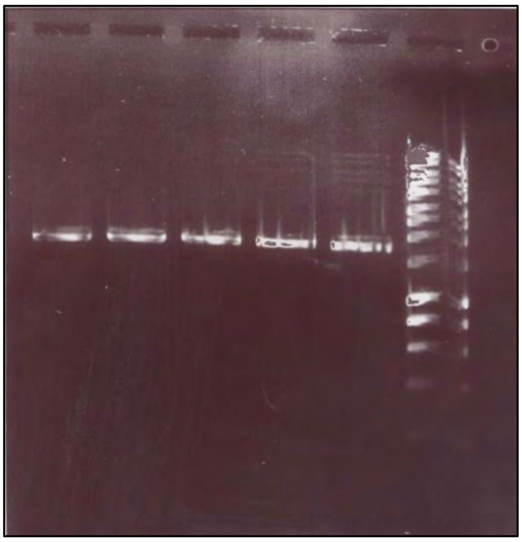


Figure 2.8 Agarose Gel of PCR amplified product in first five lanes and the DNA ladder in the last lane. The amplified product band is indicated at the band size of ±1.7 kbp.

The bands from the agarose gel electrophoresis run as shown in Figure 2.8 confirms amplified DNA sequence at the expected size of 1.7 kbp.

2.3.4 Affinity Chromatography Fractionation Profile

A summary of the fractionation elution profile indicates a peak elution after 18.5 minutes.

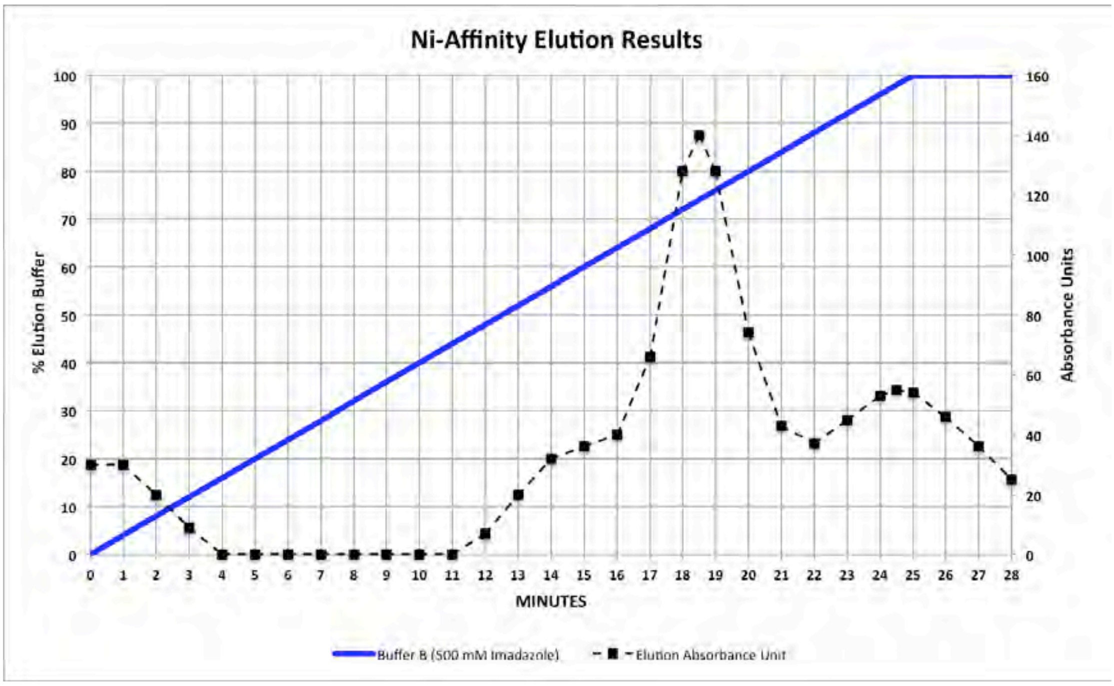


Figure 2.9 Absorbance profile of elution with linear gradient of elution buffer. Absorbance units at 280 nm indicated on right hand scale. Elution peak at 350 mM imidazole (in tube 19). A second characteristic peak towards 500 mM imidazole in tube 25.

Figure 2.9 indicates that elution towards the first first peak starts at ±45 % of buffer B (± 225 mM imidazole concentration) and peaks at ±75 % of buffer B (± 375 mM imidazole concentration).

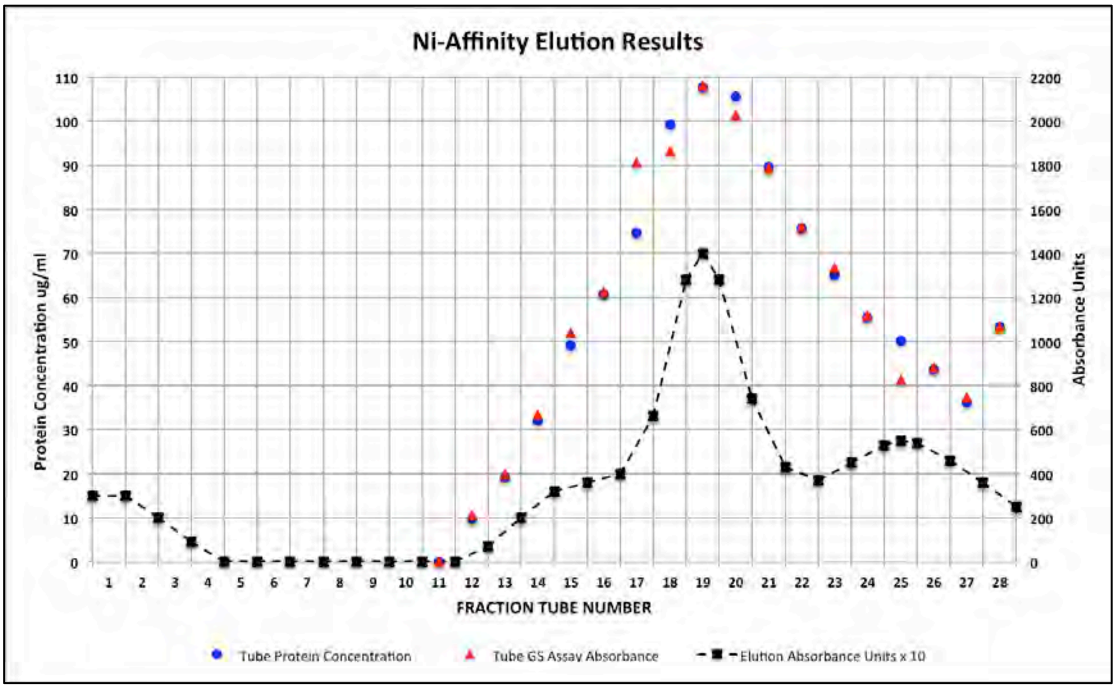


Figure 2.10 Protein concentration (µg/ml) vs GS activity (absorbance units) for the collected fractions in each of tubes 12 -28. The original elution absorbance at 280 nm (x10) is superimposed for comparative purposes. The Protein concentration and GS assay were not done in triplicate. Tubes 16 to 22 were consolidated as purified protein.

Figure 2.10 indicates that protein concentration measured in the fractionated tubes is in line with the elution profile of absorbance at 280 nm.

Tube		12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Activity	**U**	15	27	46	71	84	124	128	148	139	123	104	91	77	57	60	51	73
Protein	**µg/ml**	10	19	32	49	61	75	99	108	106	90	76	65	55	50	44	36	53
Specific Activity	**U/mg**	3810	3572	3572	3634	3459	4164	3227	3438	3291	3419	3441	3514	3462	2827	3450	3530	3429

Table 2.4 GS activity and specific activity in fraction tubes 12 -28. Both Protein and GS assay were not done in triplicate as no purification table was being attempted.

Table 2.4 indicates activity in Units (µmol of glutamyl hydroxamate per minute) The average specific activity over all the tubes (12-28) averages 3485 U/mg with an 8% (269 U/mg) standard deviation. Samples were not tested in triplicate as the measure was meant to provide a qualitative indication of activity. The total eluted protein between tubes 12 to 28 and the consolidated tubes 16-22 were estimated to be 2.6 mg and 1.5 mg respectively.

*2.3.4.1 Verification of and purity of fractionated Pf*GS

Third party mass spectrometer (Maldi-TOF) of the purified *Pf*GS, confirmed using the peptide mass fingerprint from a trypsin digest that the sequence represented “Glutamine synthetase, putative *Plasmodium falciparum* (3D7)” with a p-value = 0.00012 (<0.05 implies very significant).

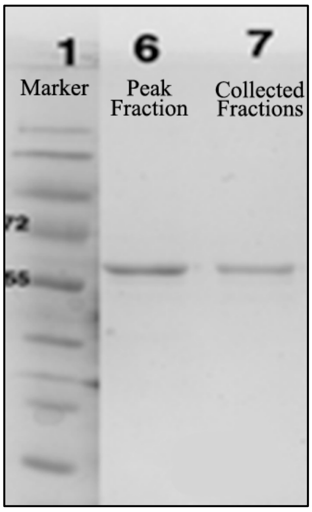
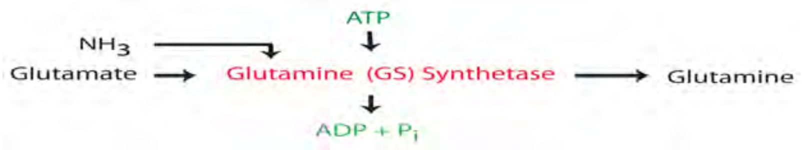


Figure 2.11 SDS –PAGE visualisation of purified *Pf*GS from the peak fraction (tube 19) and the consolidated fractions of tubes 16-22. Under lightbox conditions only a single band was visible at ±63 kDa.

SDS-PAGE gel in Figure 2.11 of both the peak fraction (tube 19) and the consolidate sample (tubes 16-22) indicates a single band, indicative of high purity fractionation.

2.3.5 Thermal Stability

The ligands involved in the reaction of GS were used to check impact on the thermal stability. The cases considered were Glutamate/ATP (forward reaction), glutamine/ADP (reverse reaction) and methionine sulfoxide (a known inhibitor). Melt curves for four cases each with either Mn^{++} or Mg^{++} as cation were undertaken in the phosphate buffer. Cation and Ligands were introduced at 0.1 mM in the final concentration.



Ligands	**Designation**	**Mn^{++}**	**Mg^{++}**	**Delta T_m**
None	Initial	With no cation < 25		
None	Cation	35	27	8
ATP+GLU	Reverse	35.9	27.8	8.1
ADP+GLN	Forward	35.9	31.4	4.5
MetSox	Inhibitor	36.2	33.4	2.8

Table 2.5 *Pf*GS melt temperatures (T_m) using DSF with different ligands. The ROTOGENE PCR machine was unable to record below 25°C. The Mn^{++} cation provides a higher thermal stability than the Mg^{++} Cation. The impact on improved stability due to the addition of ligands to either cation is marginal.

With no cation present in *Pf*GS, unfolding at 25°C was already indicated on the PCR machine. The results of the DSF experiment in Table 2.5 demonstrated that the enzyme (*Pf*GS) was stabilized by the presence of either Mn^{++} or the Mg^{++} cation. The Mn^{++} cation provided a higher thermal stability. The ligands did not increase the thermal stability significantly.

2.4 Discussion

The initial trials resulted in no soluble expression as indicated by SDS-PAGE. Using the Ni-affinity chromatography to purify the expression products from the BL21 (DE3) strain indicated no visible elution peaks between 150 mM imadizole to final elution buffer concentration of 500 mM imidazole. No GS activity was found in any of the same fractions.

Other expression trials with the pETM11_D216 at lower incubation temperature post induction to 20°C, addition of magnesium (to final concentration of 5 mM) to the culture medium, were undertaken on all four *Escherichia coli* strains. The lower temperature was introduced in an attempt to reduce metabolic strain; whilst the cation was introduced to ensure two cations per subunit of glutamine synthetase was available for structural stability. None of the expressions trials resulted in any visible soluble expression being visible on the SDS-Page gel viewed under the light box.

A different approach to using the traditional *Escherichia coli* expression protocols was necessary. Splicing of a known folding gene upstream (also known as a chaperone gene) was considered, but given the already large DNA sequence length of Glutamate Synthetase, this approach was not selected.

As the formation of the macromolecule of *Pf*GS was likely to consist of a oligomer of 8, 10 or 12 monomers, splitting the sequence into smaller domains was considered to be unfavourable towards the formation of the macromolecule and hence eliminated as an option.

The favoured approach was to reduce the metabolic expression during the growth phase of the strain, thereby maximizing resources for over-expression and folding of *Pf*GS. The pCold vector from Takara was designed to achieve these objectives.

Placement of the hexa-histidine tag was included in the N-terminus, informed by other GS sub atomic resolutions.

The standard Sanger sequencing confirmed the inserted DNA sequence as the codon optimized sequence of *Pf*GS. Mass spectrometry of the purified protein indicated with a high degree of confidence of the sequence being *Pf*GS. Both the peak and consolidated fractions indicated a single visible band on SDS-PAGE gel, indicating that a high level of purification was achieved.

*Pf*GS was found to precipitate when concentrating *Pf*GS following purification. Krajewski (2008) also found that the introduction of cations in their buffer for the *Mycobacterium tuberculosis* GS stabilised the protein. The forward and reverse ligands only marginally increased thermal stability and were therefore considered optional in the buffer.

Attempting to concentrate *Pf*GS beyond 2mg/ml resulted in significant loss of protein due to precipitation. Solubility was enhanced by the addition of 10% glycerol and beta-octyl glucoside (at a 0.4mM final concentration) and protein concentrations as high as 8 mg/ml could be achieved.

Chapter 3 – Negative Stain 3-D Reconstruction

After failing to obtain P*f*GS crystals using standard Hampton crystal trial screens, HR-110 and HR-112, it was decided to pursue single particle 3-D reconstruction using negative stain electron microscopy of the purified P*f*GS towards further characterization.

Recent advances in electron microscopy have brought cryo-electron microscopy alongside X-ray crystallography and NMR for structure determination and functional interpretation. The structure of β-galactosidase at a resolution of 3.2 Å was recently obtained by cryo-electron microscopy (Bartesaghi *et al.*, 2014) . This has provided the impetus for increased use of direct electron detectors as well as automation leading to the collection of large image data sets that can be processed to yield 3-D reconstructions.

Negative stain electron microscopy is an essential prerequisite for such a high resolution study. Three-dimensional reconstruction of the negatively stained protein can be used, not only to provide a starting model for further refinement using other techniques, but also as a basis for homology modelling using templates from known structures.

3.1 Single Particle Reconstruction and Resolution

Single particle reconstruction involves the following major steps:

1. Specimen Preparation
2. Image Acquisition
3. Particle Selection
4. 2-D Image Alignment and Clustering
5. Starting Model Generation
6. Back Projection and Refinement

Steps 2 to 6 require large data handling and computational ability. Recent advances in computer and imaging devices have significantly impacted on electron microscopy reconstruction. Automated 3-D reconstruction algorithms (detailed below) are now readily available with user graphical interface options with a goal towards unbiased 3-D reconstruction.

Two of the earliest electron microscopy software packages SPIDER (Frank *et al.*, 1996) and IMAGIC (Heel *et al.*, 1996) are still widely used. EMAN (Guang Tang *et al.*, 2007) has gained usage, as the software provides some advances in overcoming model bias and also provides a helpful graphical interface. Other standalone software platforms widely used are SPARX (Hohn *et al.*, 2007), FREALIGN (Grigorieff, 2007), XMIPP (de la Rosa-Trevín *et al.*, 2013) and RELION (Scheres, 2012). A web-based development, APPION

(Lander *et al.*, 2009), allows a complete workflow from the control of microscope image through to the solving of the 3-D reconstruction. This workflow development also allows the use of functions from various software platforms mentioned above with appropriate graphical interfacing moving the bottleneck for reconstruction towards image acquisition.

The key to producing a 3-D reconstruction is to determine the angular relationships between the different images captured downstream of the high-energy electrons passing through the macromolecule. This is achieved by taking multiple images of the same particle or single image of the particle in different orientations.

Obtaining projections of the same particles as depicted in Figure 3.1(A) with pre-determined angular relationships is also known as a single axis tilt series. The pre-determined angular relationships are then used in the reconstruction protocol as depicted in Figure 3.1(B).

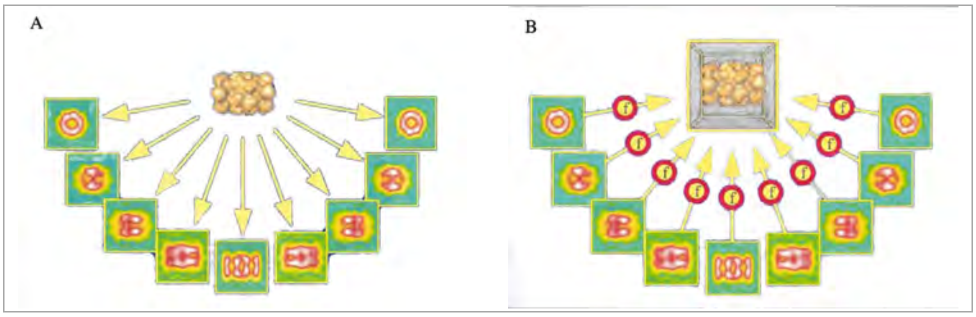


Figure 3.1 Visualizing single axis tilt protocol. (A) Single particle images are taken at pre-determined angles leading to multiple views of the same particle. (B) Reconstruction of the single particle using the pre-determined angular views in the back projection (*reproduced from Koster et al., 1997)*

Multiple exposures can lead to rapid degradation of a biological macromolecule and this method of image capture for reconstruction was excluded.

Using randomly orientated particles, also known as single particle reconstruction, is depicted below. This consists of grouping and aligning the extracted particle images from multiple micrographs into a limited number of classes and then generating a back-projected 3-D model.

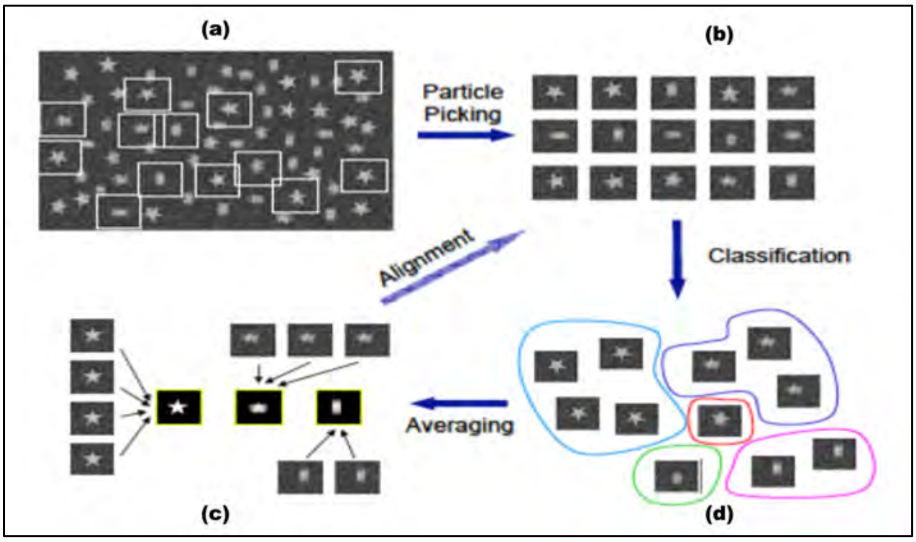


Figure 3.2 Visualizing the basics of classification and alignment of selected particles. (a) Particle selection from the micrographs. (b) Extracting of selected particles from micrograph. (c) Classifying selected particles into pre-determined number of classes. (d) Aligning and averaging particles within each class for improving signal to noise ratio. (*reproduced from lecture notes: http://people.cryst.bbk.ac.uk/~ubcg16z/cpn/uclEM2.pdf*)

The steps undertaken to classify and align particles is indicated in Figure 3.2. In all reconstruction techniques, the averaging of similar orientated particles are undertaken to reduce the overall signal to noise ratio that is inherent in individual images.

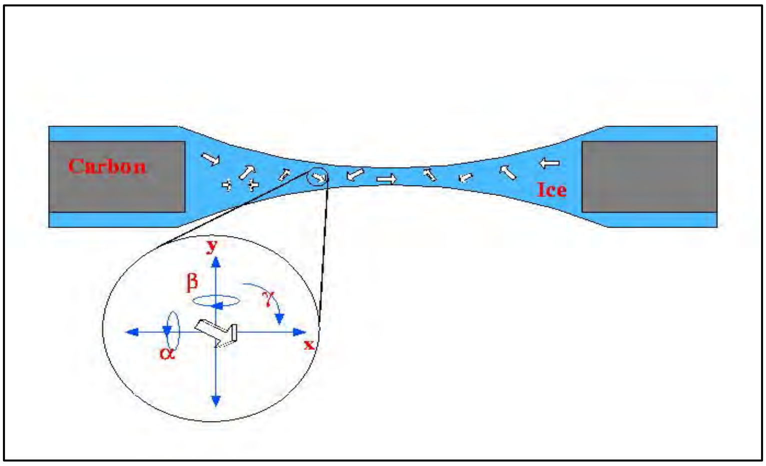


Figure 3.3 Translational and angular parameter convention in orientating the different particles against a reference point. (*reproduced from lecture notes: http://people.cryst.bbk.ac.uk/~ubcg16z/cpn/uclEM2.pdf*)

Assignment of five spatial parameters, categorized into two transitional and three angular parameters orientate each particle relative to each other. The relative orientations are initially not known, as can be seen from the random orientations on the cryo-EM grid in Figure 3.3, but nevertheless remain essential to orientate the particle. However, various strategies for an initial assignment are available. Following initial assignments, they are reassigned in an iterative process called refinement. One such technique assumes "a starting / assumed" 3-D model which is used to assign an initial orientation based on the best fit of the macromolecule image to give a projection of the starting model.

The typical workflow, as indicated in Figure 3.4, uses a starting model to project 2-D images at pre-determined angles of the 3-D model. The 2-D raw images are then aligned to the projected 2-D image (also known as reference image). The angular orientations of the reference image that best fit the raw image is assigned to that particular raw image. It is the raw images and the assigned angular orientations that are used to regenerate a new 3-D model (also known as back-projection). When more than one raw image is aligned to a particular reference image then the raw images are averaged.

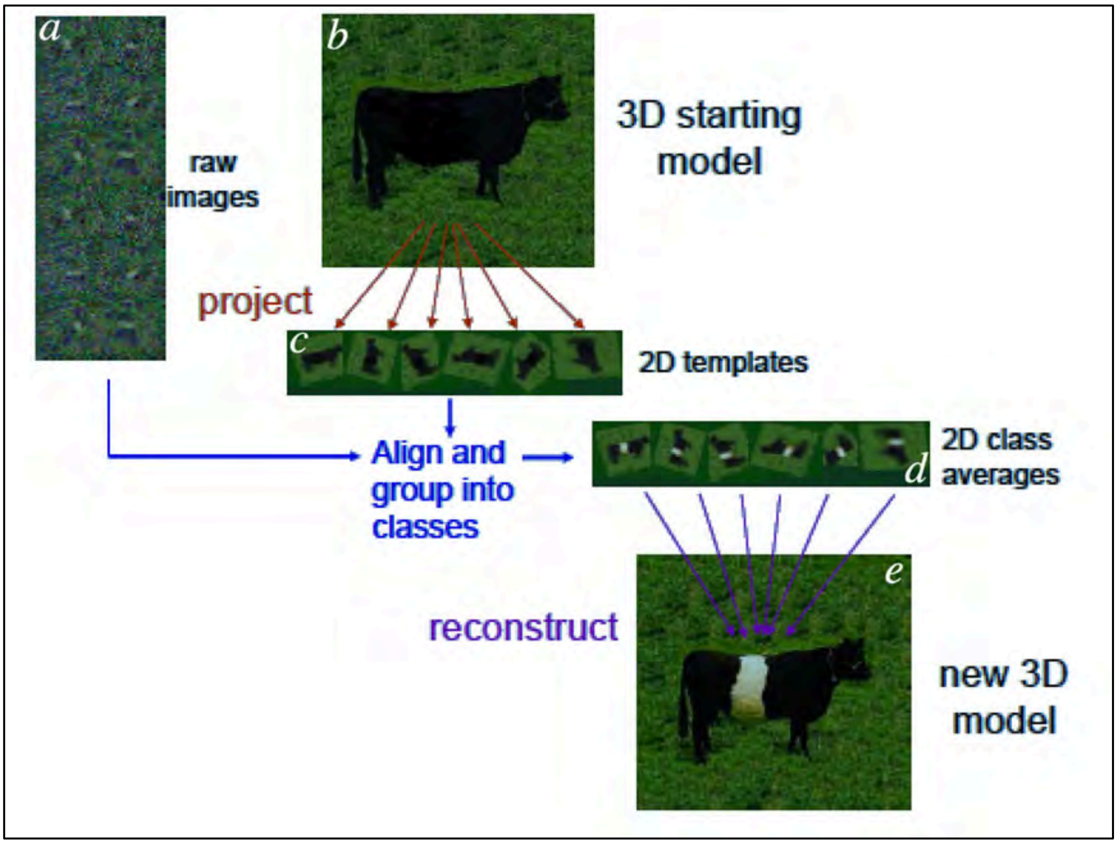


Figure 3.4 The basic steps in 3-D reconstruction using randomly orientated images of same object. (a) Raw 2-D images of object (b) known 3-D starting model (c) 2-D projections of the known starting model (d) Align 2-D raw images to the 2-D projections to obtain angular and transitional parameters. (e) back project a 3-D model using angular and transitional parameters obtained in the previous step. (*reproduced from lecture notes: http://people.cryst.bbk.ac.uk/~ubcg16z/cpn/uclEM2.pdf*)

The workflow is then iterated (refinement) until no further improvement to resolution can be obtained with the extracted 2-D particle image set. Increasing the number of classification groups (angular refinement), may further improve the alignment of the images and thus increase the resolution of the 3-D back-projected volume. However, there is a concern that the starting model could introduce a bias towards the final model. To overcome this concern, a strategy to independently group and align the macromolecule image set on the basis of their similarity in pre-determined number of classes is used. This technique is known as common lines, from which an initial model is generated from the created classes.

Determining the resolution for single particle reconstructions (Liao & Frank, 2010) is obtained by splitting the images into two sets and looking for data consistency between the two independent reconstructions. Comparisons are made over rings (shells) with increasing radii in the Fourier space of the two models. The Fourier shell correlations (FSC) are plotted against their respective ring and a FSC threshold of 0.5 is used to quote the resolution obtained.

3.2 Materials and Methods

Where no specific mention is made, the same protocol was applied to both negative stain samples. Visualizations of 3-D volumes and sub-atomic structures was done using the CHIMERA (Meng *et al*., 2006) software.

3.2.1 Sample Preparation

Recombinant *Pf*GS was isolated from *Escherichia coli* using a cold shock expression protocol and purified as described in chapter 2. The stored protein was diluted to approximately 0.1 mg/ml with buffer[5].

3.2.2 Negative Stain Grid

Electron microscopy grids that were pre-coated with thin carbon support film were glow discharged in air for 20 seconds. Four 10 µl droplets, one of 0.1 mg/ml of sample and three of 2% uranyl acetate solution were eluted on Parafilm™. The carbon-coated side of the grid was floated on the sample droplet for 25 seconds, before being lifted and blotted. Immediately the same side was floated on each of the remaining three droplets with blotting before moving on the next droplet. The grid was left floating for five seconds on each droplet. The grid was air dried at room temperature for 30 minutes, before storage in a closed Petri-dish awaiting microscope time.

---

[5] 20 mM Na_2HPO_4 buffered to pH7.4 with NaH_2PO_4+ 50 mM NaCl + 5mM $MnCl_2$

3.2.3 Negative stain EM Grid Imaging

Micrographs were recorded, using a 4096 x 4096 slow scan CCD camera with 15 µm pixel size, attached to the TEM-F20 operated at 200 kV. Dosage was set at 20 e/Å^2. Astigmatism and defocus were checked and manually adjusted after a change to a new location on the grid was made. The first image set was captured at a 50 000 x magnification and the second image set at a 100 000 x magnification setting on the TEM-F20 console. The calibrated sampling was 2.11 and 1.05 Å /pixel

3.2.4 Micrograph Rejection and Contrast Transfer Function Parameter Estimation

An initial screening was undertaken by visually inspecting the spatial frequency power spectrum of the micrograph and by rejecting those with Thon rings that indicated astigmatism or did not or a resolution of one quarter of the maximum for the magnification in question. Acceptable micrographs were then included in the reconstruction pipeline software for further processing. Estimated defocus, astigmatism and confidence of parameter fits for each micrograph was obtained using CTFFIND (Mindell & Grigorieff, 2003) software which was further used to exclude low quality micrographs.

3.2.5 Negative Stain Image Processing and 3-D Reconstruction

The APPION web application software (Lander *et al.*, 2009) was chosen for the 3-D reconstruction.

3.2.5.1 Particle Selection and Image Stack Creation

The expected particle size, as measured from a typical micrograph, was 160 Å (Figure 3.5). With a practical maximum resolution of 15 - 20 Å expected from a negative stained sample for the reconstruction, sampling at 7.5 Å at a minimum was necessary to avoid aliasing and accumulating interpolation errors during reconstruction. With sampling at 2.11 and 1.05 Å respectively, downsizing the image beyond a factor of 3 and 6 respectively was avoided.

For the first negative stain image set, an initial small set of particles were manually selected and classified into categories. These were then used as templates for auto-picking using the FindEM software (Roseman, 2004). For the second negative stain image set particles were all picked manually, due to the close proximity of the side view particles.

Particle extraction was done on phase-flipped micrographs and stacked, using the smallest recommended box size to minimise the possibility of incorrect alignment of the side views with adjacent particles.

The first image set extraction was done at 112 pixels and for the second image set at 224 pixels. The 224 box size particles were binned to 112 pixels. Images were padded to 160 pixels during the alignment and reconstruction phases to avoid edge effects when transformations are carried out on the particle images.

3.2.5.2 Particle Alignment

Alignment of the stacked images were undertaken using the EMAN (Ludtke *et al.*, 1999) protocol and unsupervised classification using the Xmipp (Pascual-Montano *et al.*, 2001) software package.

3.2.5.3 Ab Initio Model

Due to the close proximity of the side views, constructing of an ab initio model using common lines was not practical. Having confirmed a six-fold symmetry from the first negative stain data set, filtering of a glutamine synthetase from PDB:1F52 to 50 Å was used as the starting model for refinement. Concerns of model bias are discussed in the section 3.4

3.2.5.4 Model Refinement and Reconstruction

The EMAN1 (Ludtke *et al.*, 1999) projection-matching refinement protocol was used, whereby the extracted particles are classified according to the angular projections of the previously constructed model. Standard recommended parameters were used with a phase-flipped particle set. No amplitude corrections were invoked.

3.2.5.5 Post Processing of 3-D constructed model

The final model was, however, low-pass filtered to 15A.

3.3 Results

3.3.1 First Negative Stain Image Set

3.3.1.1 Visualisation

Inspection of the first negative stain images reveal roughly spherical particles (referred to

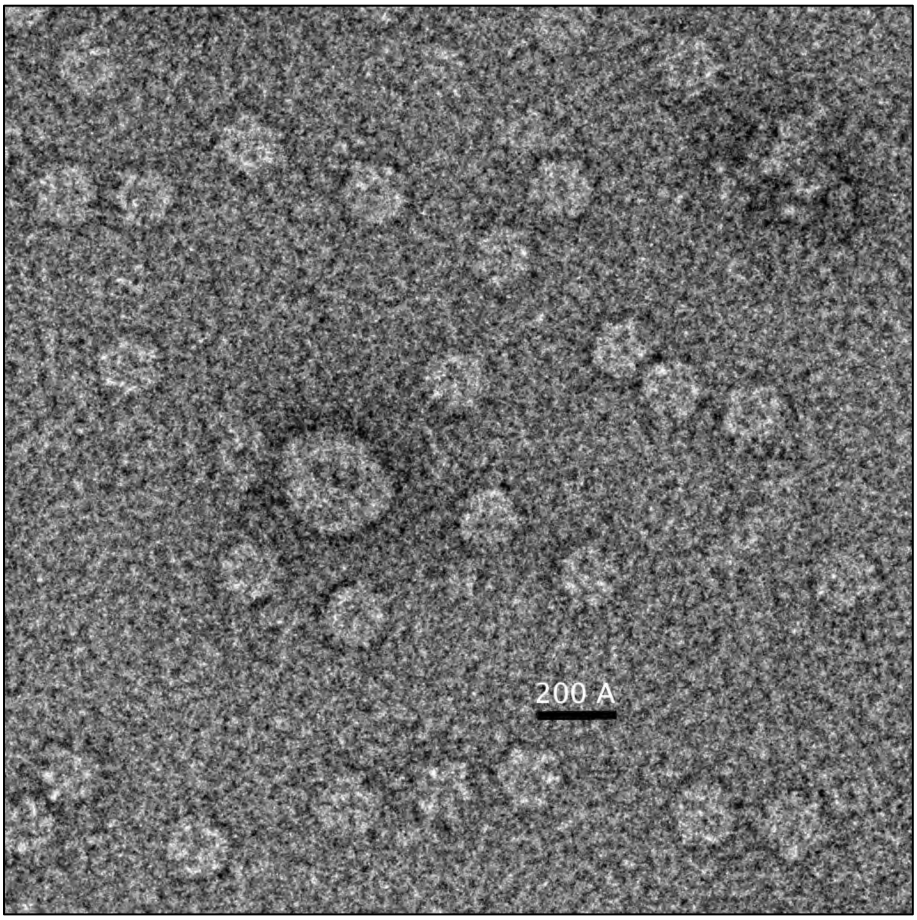


Figure 3.5 A typical micrograph from the first negative stained image set. Spherical images only were visible. Shown above is *(1024x1024 pixels extracted from the captured image of 4096x4096 pixels).*

as top views). A micrograph in Figure 3.5 typifies this. Particle orientations were however, limited to these top (spherical) views measuring approximately 160 Å.

3.3.1.2 Micrograph rejection based on FFT

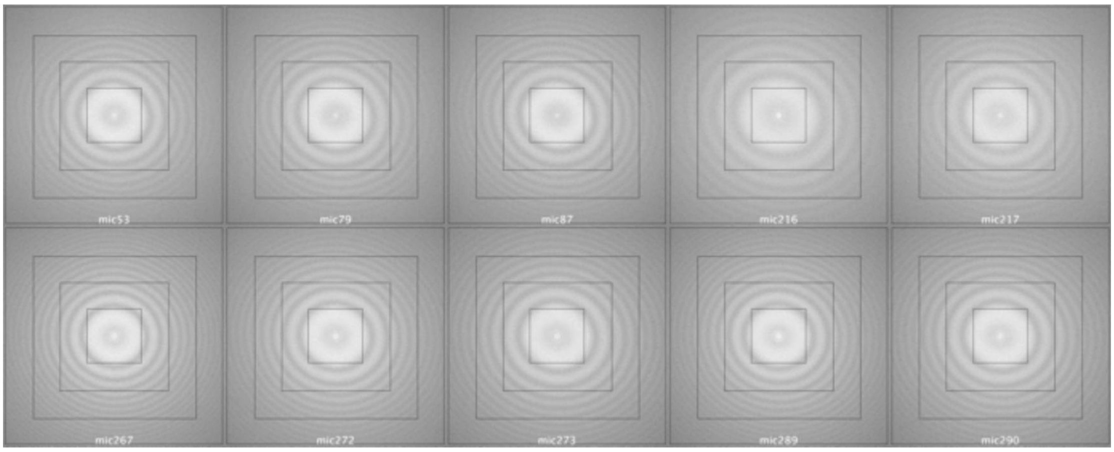


Figure 3.6 Spatial frequency power spectra of a subset of 10 accepted micrographs. These micrographs were accepted due to near spherical Thon rings. The reference boxes indicate 12.5, 25 and 37.5% of Nyquist frequency.

From the micrographs accepted for further processing, the power spectra of ten micrographs are shown in Figure 3.6. These power spectra indicate minimal astigmatism and Thon rings extending beyond the 37.5% Nyquist frequency (corresponding to a resolution of 5.6 Å).

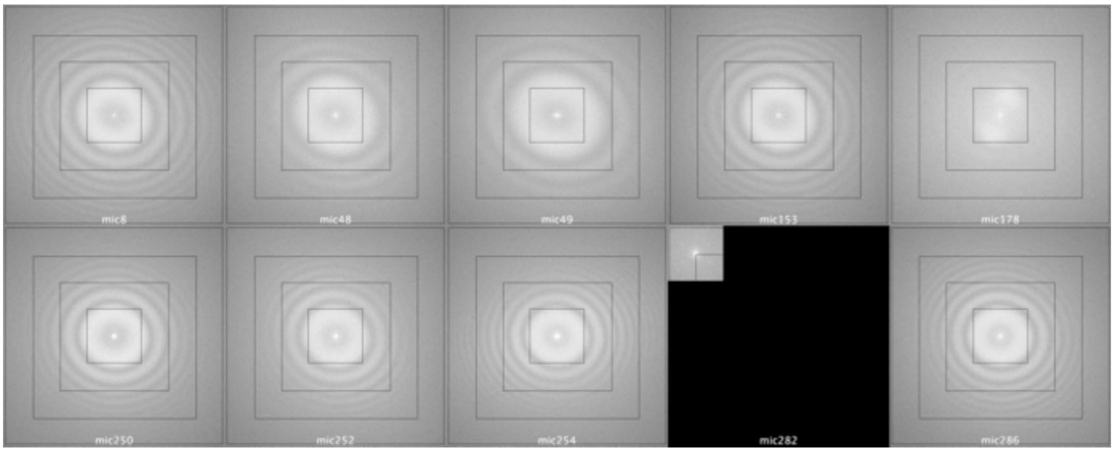


Figure 3.7 Spatial frequency power spectra of a subset of 10 rejected micrographs. These micrographs were rejected due to non-circular or limited extension of the thon rings. Mic282 was obviously corrupted. The reference boxes indicate 12.5, 25 and 37.5% of Nyquist frequency.

The power spectra of ten micrographs that were rejected for further processing are shown in Figure 3.7. Micrograph 282 was corrupted during image recording. Micrograph 178 indicates drift whilst micrographs 48, 49 and 259 indicate limited spread of Thon rings. The remaining micrographs indicate unacceptable astigmatism. A total of 16 micrographs were rejected leaving 119 for further processing.

3.3.1.3 Micrograph rejection based on CTF parameters

		Min	Max	Average	S/Deviation
Average Defocus	µm	0.49	3.10	0.84	0.22
Astigmatism	^{o} (deg)	15.1	38.3	21.7	3.1
Confidence	%	83.2	99.4	95.1	3.4

Table 3.1 CTF parameter fit of the 119 accepted micrographs using the CTFFIND algorithm. Confidence levels average 95% with a standard deviation of 3.4%.

Table 3.1 summarises the fitting of CTF parameters which indicate the range of astigmatism between 15° to 38° and a confidence of parameter fitting in the range 1/30 to 1/10 Å, of between 83 and 99%. Confidence limits greater than 80% are considered acceptable. Based on these results, no further micrographs were rejected.

3.3.1.4 Auto-Picking of particles using Templates

After selecting 73 particles manually, they were divided into four classes and used as templates for auto-picking.

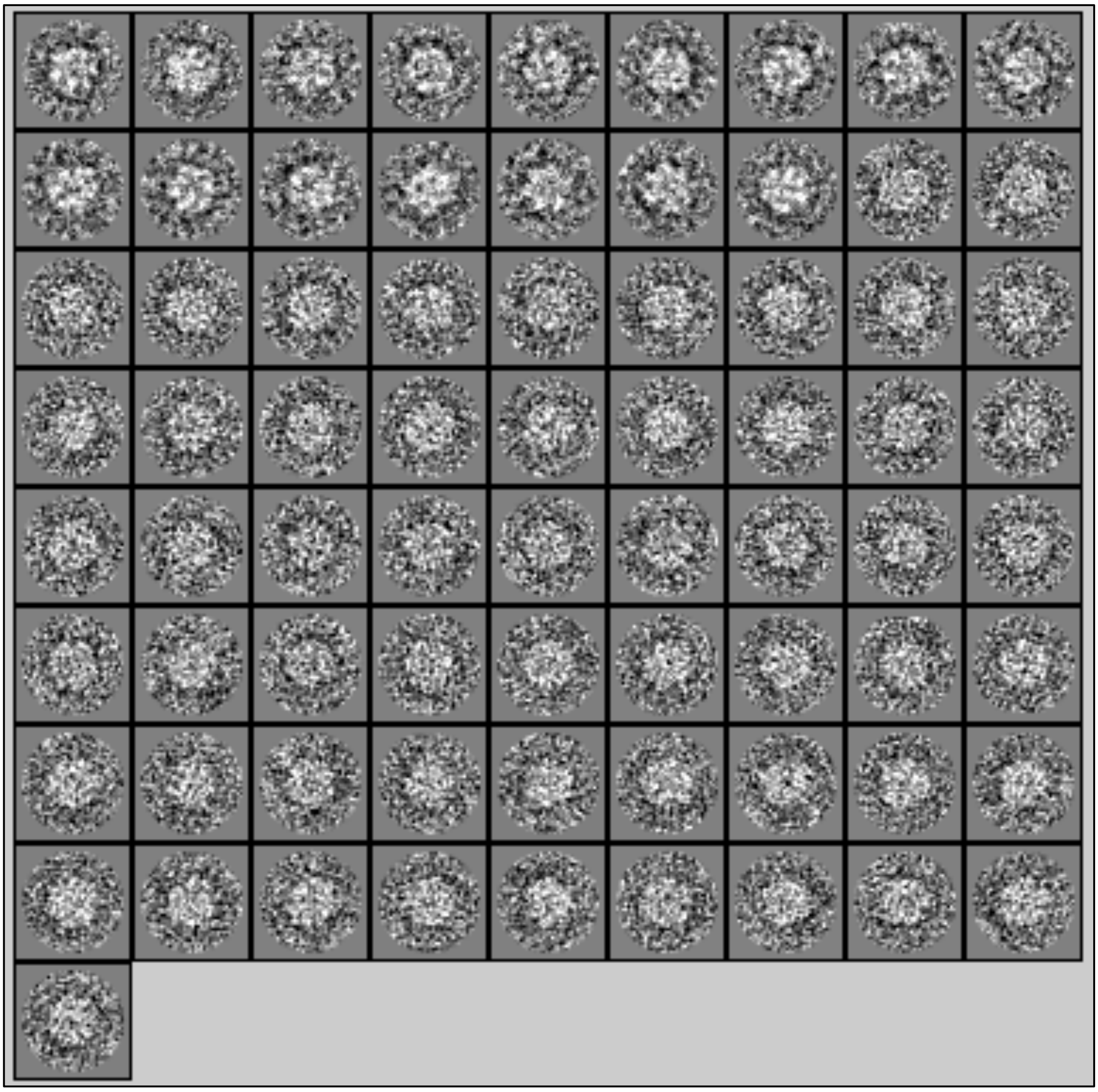


Figure 3.8 The initial stack of the manually picked particles used for creating four templates.

The 73 particles extracted, as shown in Figure 3.8, were classed into four groups, which in turn were used to auto pick particles from all the micrographs. Images of the four templates can be seen in section 1 of Appendix C.

Figure 3.9 is a typical micrograph indicating the selection of particles by each of the four templates, by way of the four different colours.

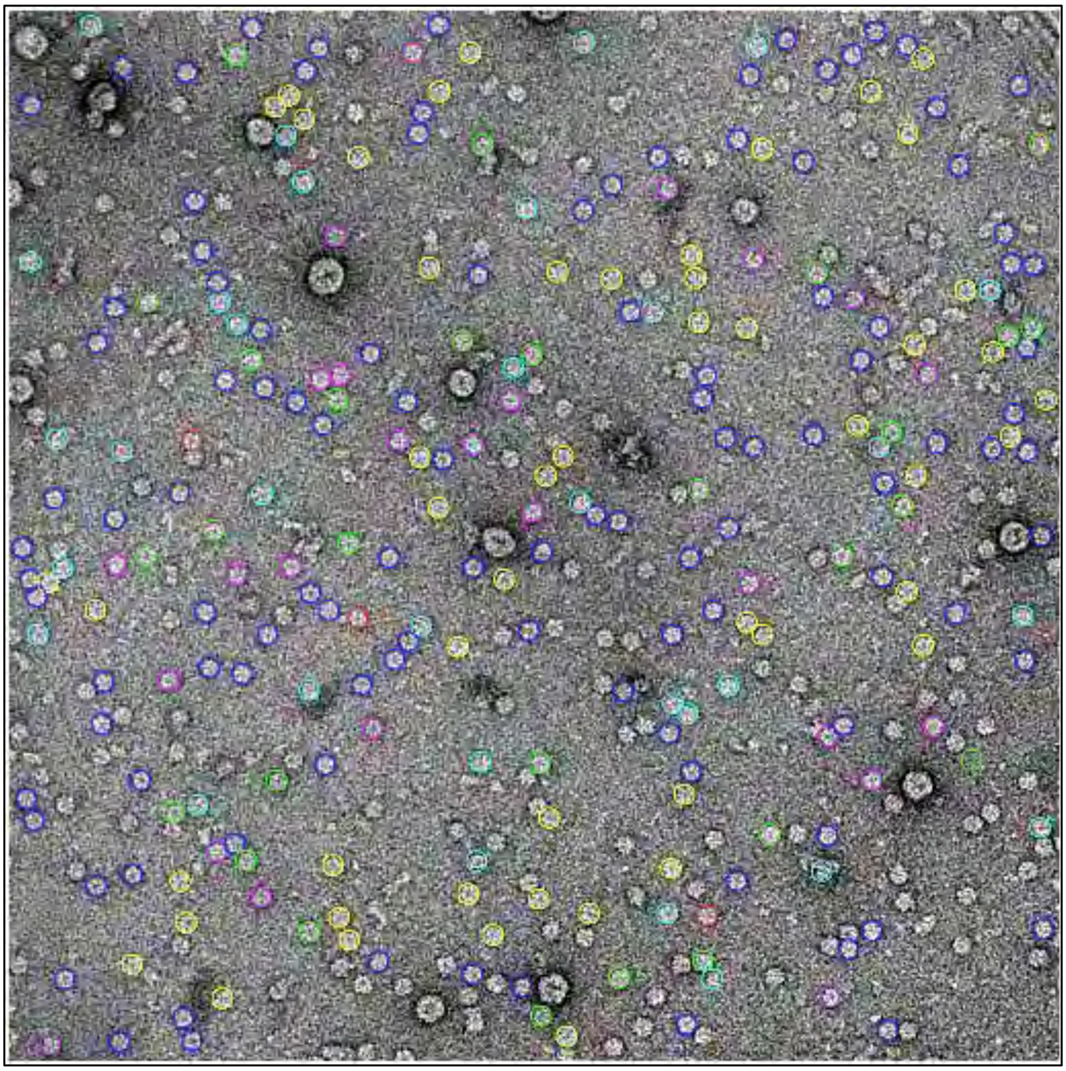


Figure 3.9 Typical micrograph after auto-picking of particles using the FindEM software embedded in Appion. The four colours (yellow, green, turquoise and blue) represent the template used as the basis for selecting the particular particle.

A total of 29665 particles were selected using the default parameters and restricting the number of particles to 250 per image (the Appion GUI for template particle picking is shown in section 1.0 Appendix C). The selection of some touching particles could have been avoided by rerunning the auto picking with changed settings in the advanced options section.

3.3.1.5 Particle Stack creation

Phase flipping and CTF correction was applied at the micrograph level before particle extraction was made. Following the stack creation (the Appion GUI for particle stack creation is shown in section 2.0 Appendix C), further filtering based on mean and standard deviation of pixel values, was undertaken to reject the outlier images, resulting in a final stack of 28001 particles.

3.3.1.6 Reference Free Alignment

The 28001 particles were aligned into twenty classes. The average of each class is shown in Figure 3.10 (the Appion GUI for creating reference free classes is shown in section 3.0 Appendix C). By inspection, it is clear that there is a lack of side view particle orientations.

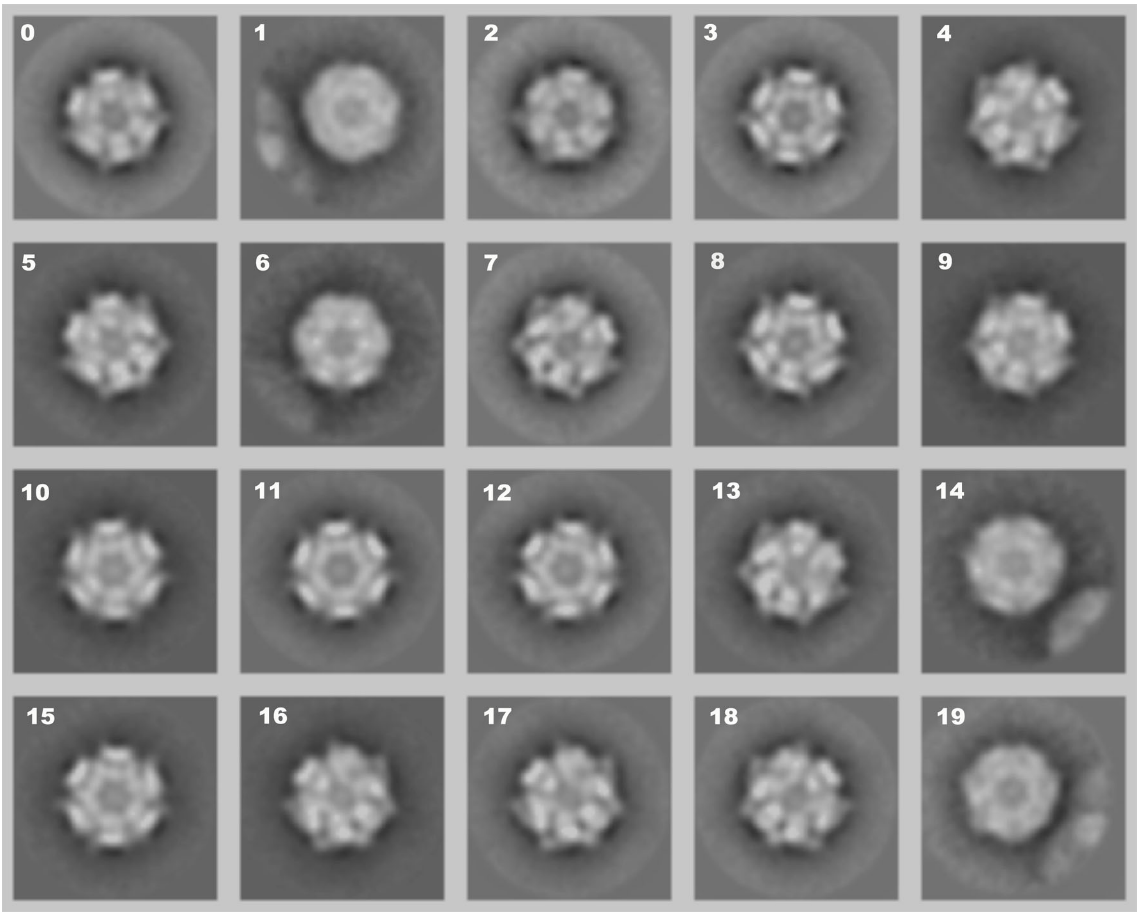
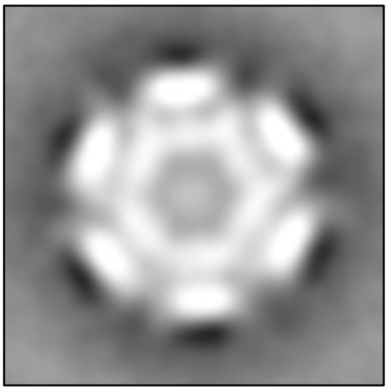


Figure 3.10 Alignment of the 28001 particles into 20 reference free classes. Only near spherical reference free classes were created (top views). A full spectrum of orientations are required for 3-D reconstruction.

However, a very important conclusion was made from undertaking a rotational cross correlation of 11th class, which is enlarged in Figure 3.11. Visibly a six fold symmetry looks apparent. The result of the rotational cross-correlation on this class is indicated in Figure 3.12.



3.11 Reference free class image enlarged of class number 11 from Figure 3.10.

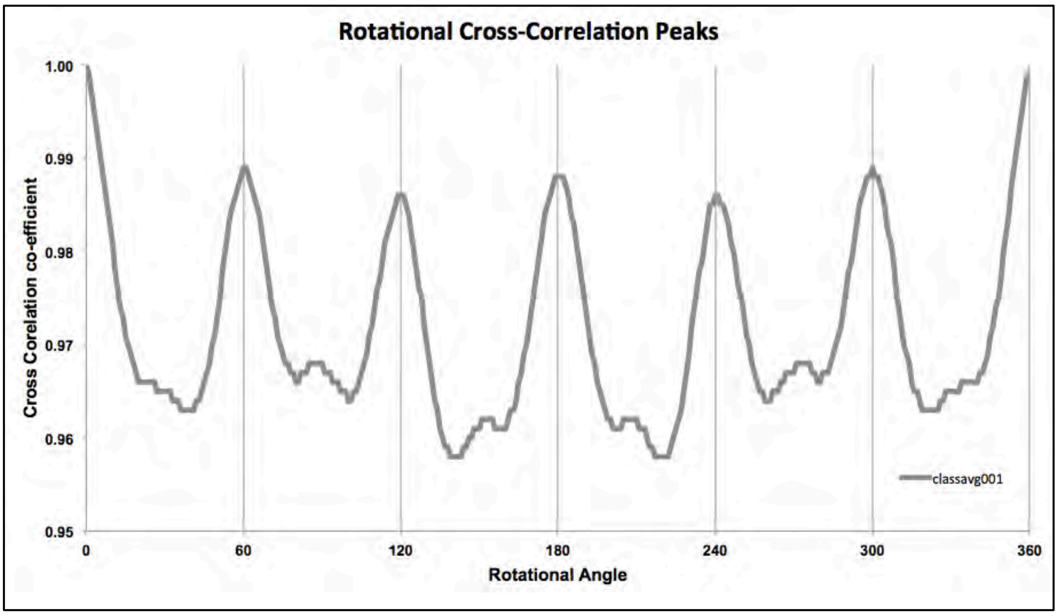


Figure 3.12 Rotational cross correlation of the 11th class. The plot indicate peaks of the cross-correlation coefficient every sixty degrees.

Cross-correlation values peak at sixty degree intervals as can be seen in Figure 3.12, is strongly indicative of a six-fold symmetry.

3.3.1.7 Initial model and Reconstruction

As side views were not represented in the data set, an initial model and reconstruction was not attempted with this data set.

3.3.2 Buffer trials for EM Grid Preparation

To enable reconstruction, side views of *Pf*GS on the EM grid was essential. Different salt concentrations and glycerol concentrations were tried for negative stain grid preparations towards obtaining all orientations of the particle.

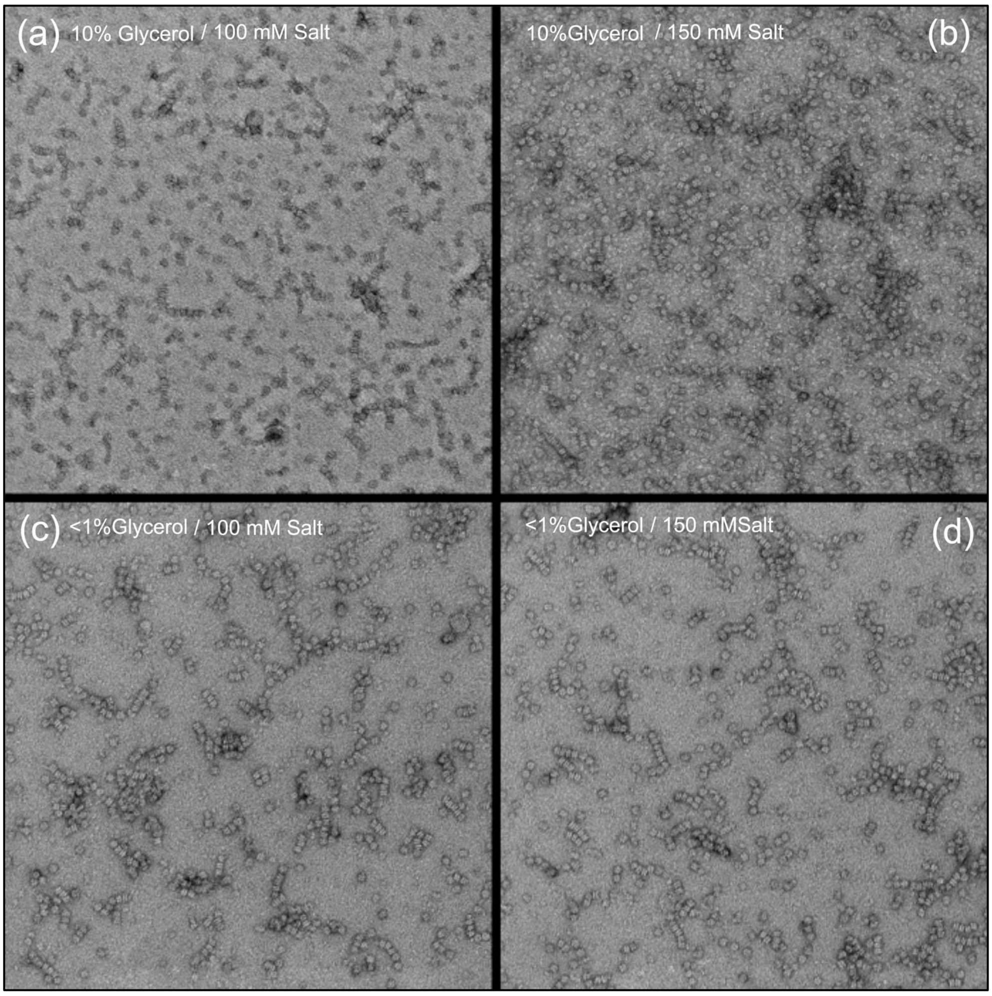


Figure 3.13 Visualisation of EM grids prepared at 100 and 150 mM NaCl and glycerol at less than 1% and 10%.

Figure 3.1.3 indicates that increasing buffer salt concentration to 100 mM (from an initial 50 mM) in the buffer, side views orientations became dominant but there was an absence of top views. The particles also appear to aggregate into different length 'fibre' like arrangements. Glycerol concentrations seem to have very little impact on particle orientation on the EM grid.

Subsequent grid preparations were buffer exchanged to 100 mM salt concentrations and to less than one percent glycerol. Reducing particle concentration to 0.05 mg/ml and with a higher salt concentration of 100 mM was chosen to achieve both top and side orientations being represented.

3.3.3 Second Negative Stain Image Set

3.3.3.1 Visualization

Inspection of micrograph images recorded from the second negative stain grid revealed both side and top views being represented. This can be seen in a typical micrograph shown in Figure 3.14. This was achieved after altering buffer conditions in the phosphate buffer to 100 mM NaCl and less than 1% glycerol and protein concentration to 50 µg/ml. Measure of the spherical particles is around 160 Å in diameter.

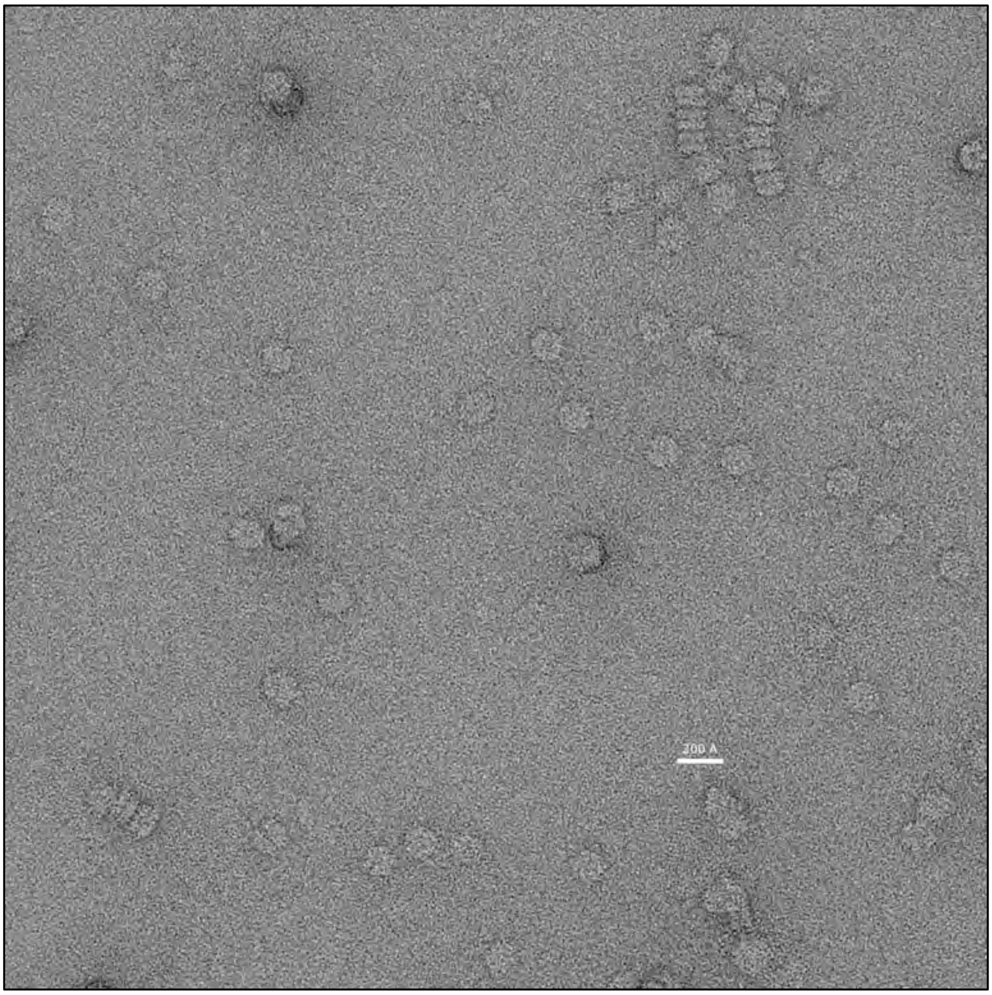


Figure 3.14 Typical micrograph from second negative stain image set. EM grid prepared at 100 mM NaCl and 50 µg/ml protein concentration. Both spherical and side views are visible (*captured image on of 4096x4096 pixels*)

3.3.3.2 Micrograph acceptance based on FFT

Of the 319 micrographs that were recorded, a subset of 167, based on visually inspecting the power spectra of the micrographs, were accepted for further processing.

3.3.3.3 Micrograph acceptance based on CTF parameters

		Min	Max	Average	S/Deviation
Average Defocus	µm	0.44	0.88	0.56	0.09
Astigmatism	°(deg)	11.1	27.4	19.07	5.5
Confidence 1/30 to 1/10 Å	%	97.2	99.9	99.3	0.5

Table 3.2 CTF parameter fit of the 167 accepted micrographs using the CTFFIND algorithm. Confidence levels average 99% with a standard deviation of 0.5%.

Table 3.2 summarises the fitting of CTF parameters which indicate the range of astigmatism between 11° to 28° and a confidence of parameter fitting in the range 1/30 to 1/10 Å, of between 97 and 99%. Based on these results, no further micrographs were rejected.

3.3.3.4 Manual Particle Picking

Particles were manually selected and carefully centralized within the selection box.

Figure 3.15 Manually picked particle on a typical micrograph. Selection of particles is unable to avoid overlapping between particles of side orientations.

Selection of particles on a typical micrograph can be seen in the Figure 3.15. A total of 672 particles were manually selected in this manner from the 167 micrographs that were accepted for further processing.

3.3.3.5 Particle Stack creation

Phase flipping and CTF correction was applied at the micrograph level before particle extraction was made. Following the stack creation (the Appion GUI for particle stack creation is shown in section 4.0 Appendix C), further filtering, based on mean and standard deviation of pixel values, was undertaken to reject the outlier images resulting in a final stack of 663 particles.

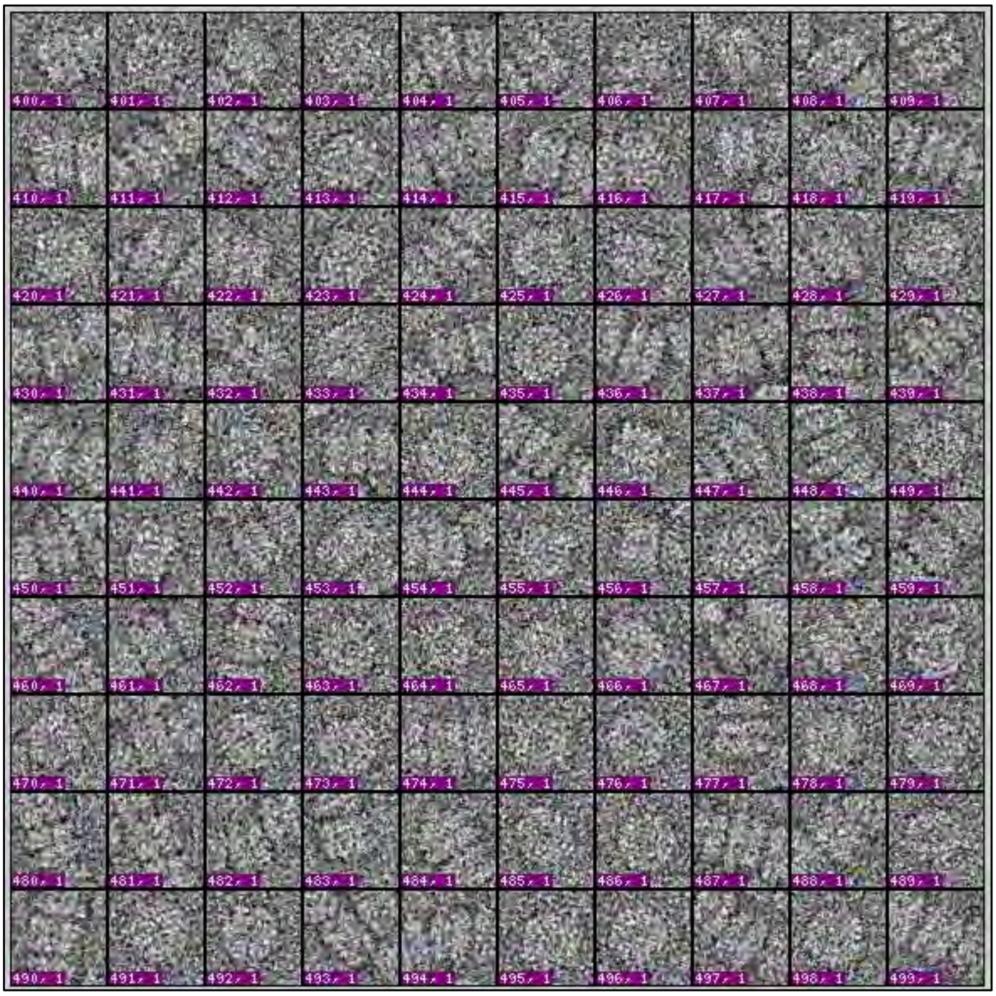


Figure 3.16 A subset of particles extracted from the stack after manual picking. Note the adjacent particles in the images where side views are represented.

A subset of 100 particles shown in Figure 3.16, indicates both top and side views. The side views are crowded with adjacent particles that will require tight masking during the initial reconstruction and refinement stages.

3.3.3.6 Reference Free Alignment

Of the thirty-six reference free class averages requested, (the Appion GUI for reference free alignment is shown in section 5.0 Appendix C) twenty seven were generated, as shown in the montage of the classes in Figure 3.17. Both top and side view orientations are

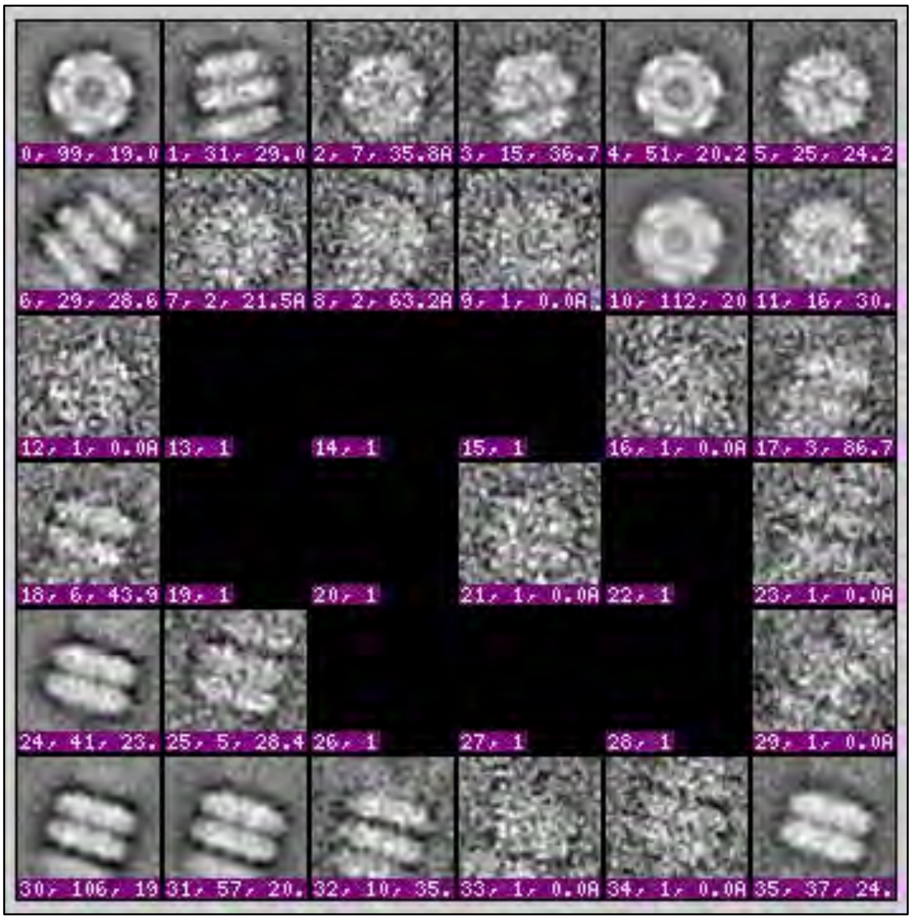


Figure 3.17 Reference free classes for second negative stain data set. Of the 36 classes requested, only 27 classes were outputted. Both circular views and side views are represented in the classes.

represented in the 27 reference free classes computed. No further particle rejections were made at this stage due to the limited number of particles available for processing.

3.3.3.7 *Initial model*

It was not practical to generate a reference free model using common lines given the

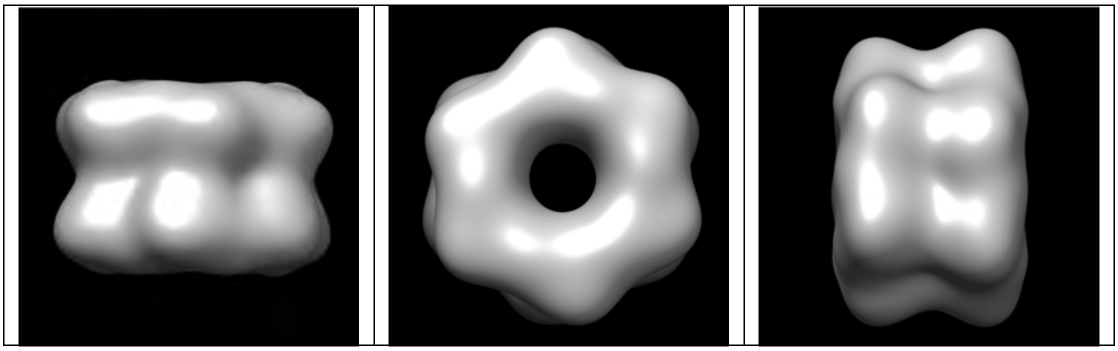


Figure 3.18 Starting model PDB:1F52 filtered to 50 Å.The central image is referred to as the top view. The left and right hand images are side views.

adjacent particles in the side view orientations. It was decided to use, as a starting model, a homology PDB[6] structure, filtered to 50Å to enable alignment of the side views. Three views of the filtered starting model filtered is shown in Figure 3.18. Concerns on model bias are addressed in the discussion section.

3.3.3.8 Model Refinement

The refinement was set for 20 iterations, using the EMAN (Guang Tang *et al*., 2007) projection matching refinement protocol. A D6 symmetry was imposed, together with angular sampling, and was set at 7.5° for the first 10 iterations and 5° for the next 10 iterations (the Appion GUI for single particle reconstruction is shown in section 6.0 Appendix C).

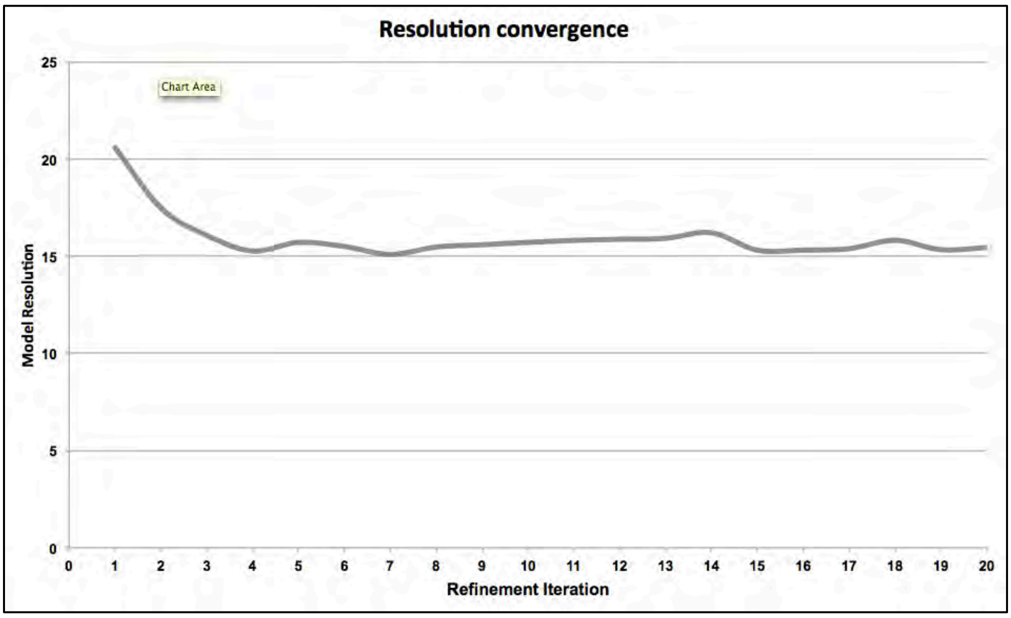


Figure 3.19 Model resolution converges vs refinement cycle.

The results of the model resolution following each iteration, shown in Figure 3.19, indicate convergence in the first five iterations. The final model resolution after 20 iterations was 16 Å. The number of particles used in the back projection during each iteration averaged 576.

The final model filtered to 15 Å, is shown in Figures 3.20 and alternate views in Figure 3.21. On inspection, there is a central density occurring in the final reconstruction, which is attributed to the tightness of adjacent particles that was not avoided in the side views and was used for the reconstruction.

---

[6] Protein Data Base structure PDB:1F52 – Glutamine synthetase of *Salmonella typhimurium*

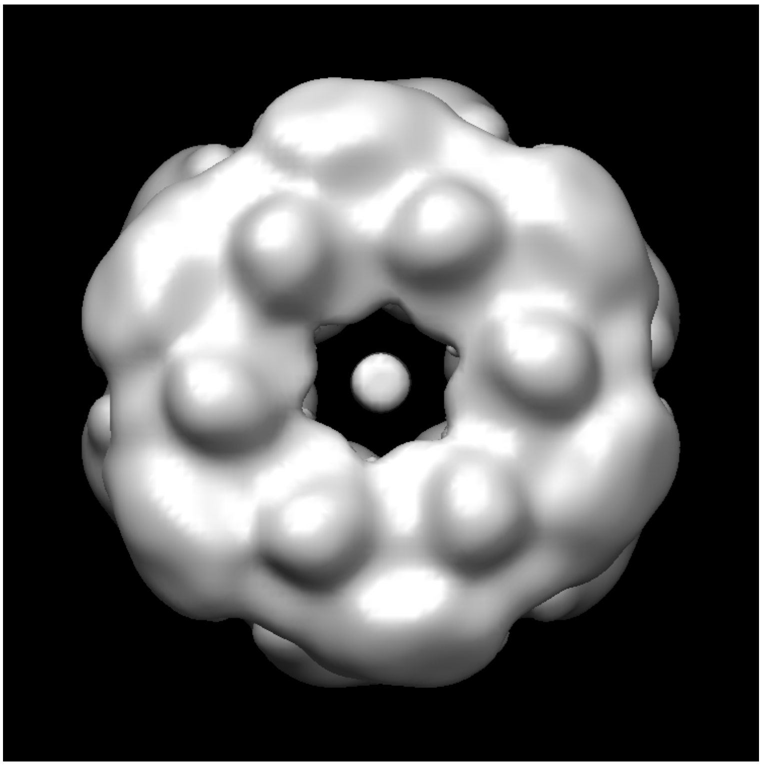


Figure 3.20 Top view of *Pf*GS 3-D reconstruction filtered to 15 Å. The central density is attributed to tightness of adjacent particles that was not avoided in the side views and was used for the reconstruction.

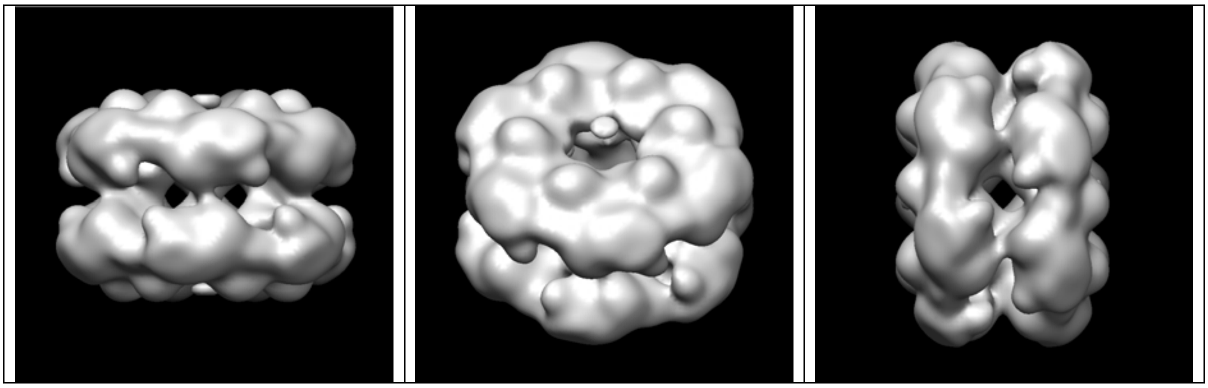


Figure 3.21 Different views of the final reconstructed model. The center image shows a 45^o rotation on the y-axis. The left and right hand images are side views.

3.4 Discussion

3.4.1 Model Bias

In order to put aside any suspicion of a model bias being introduced by using a PDB starting model, alignment of the first negative strain particle stack was made to projections of the final constructed model, using the second negative stain particle stack. Projections at three degree intervals resulted in a total of 179 projections of the model after applying a D6 symmetry.

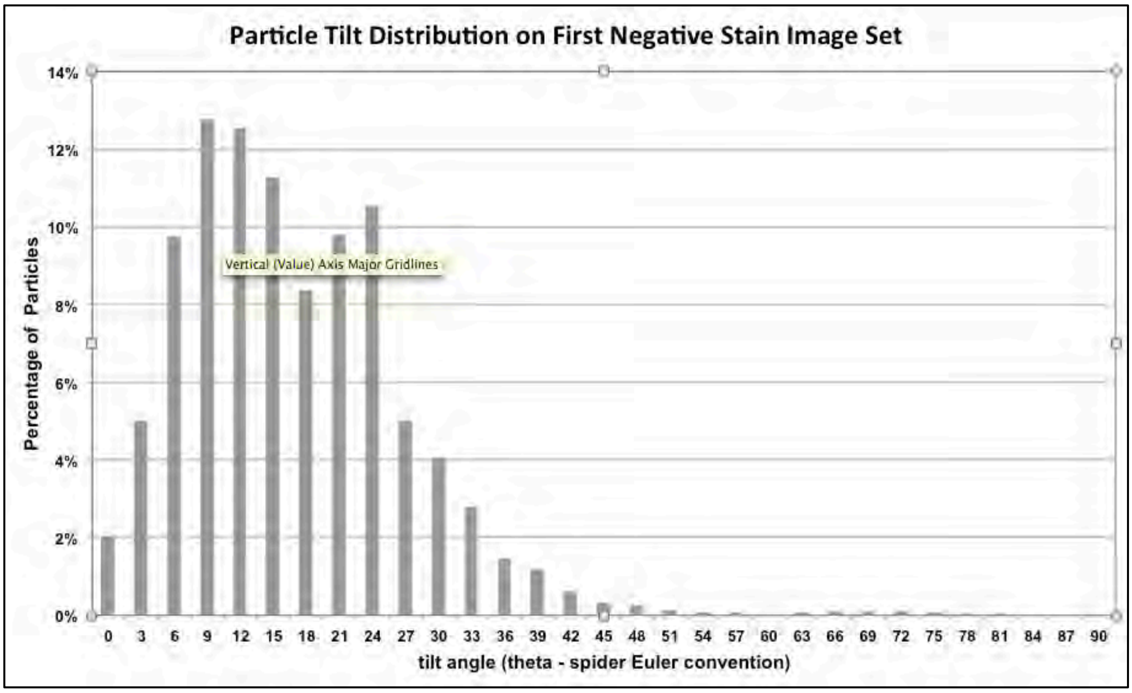


Figure 3.22 Particle orientation distribution of the first negative stain data set.

Summarizing the number of particles graphically in Figure 3.22 in three degree intervals of theta[7], confirms the bias towards top views.

Visual comparisons are made of the class averages of the first negative stain data set to the model projection in Figure 3.23 below.

---

[7] based on the Euler angle convention used in the Spider functions.

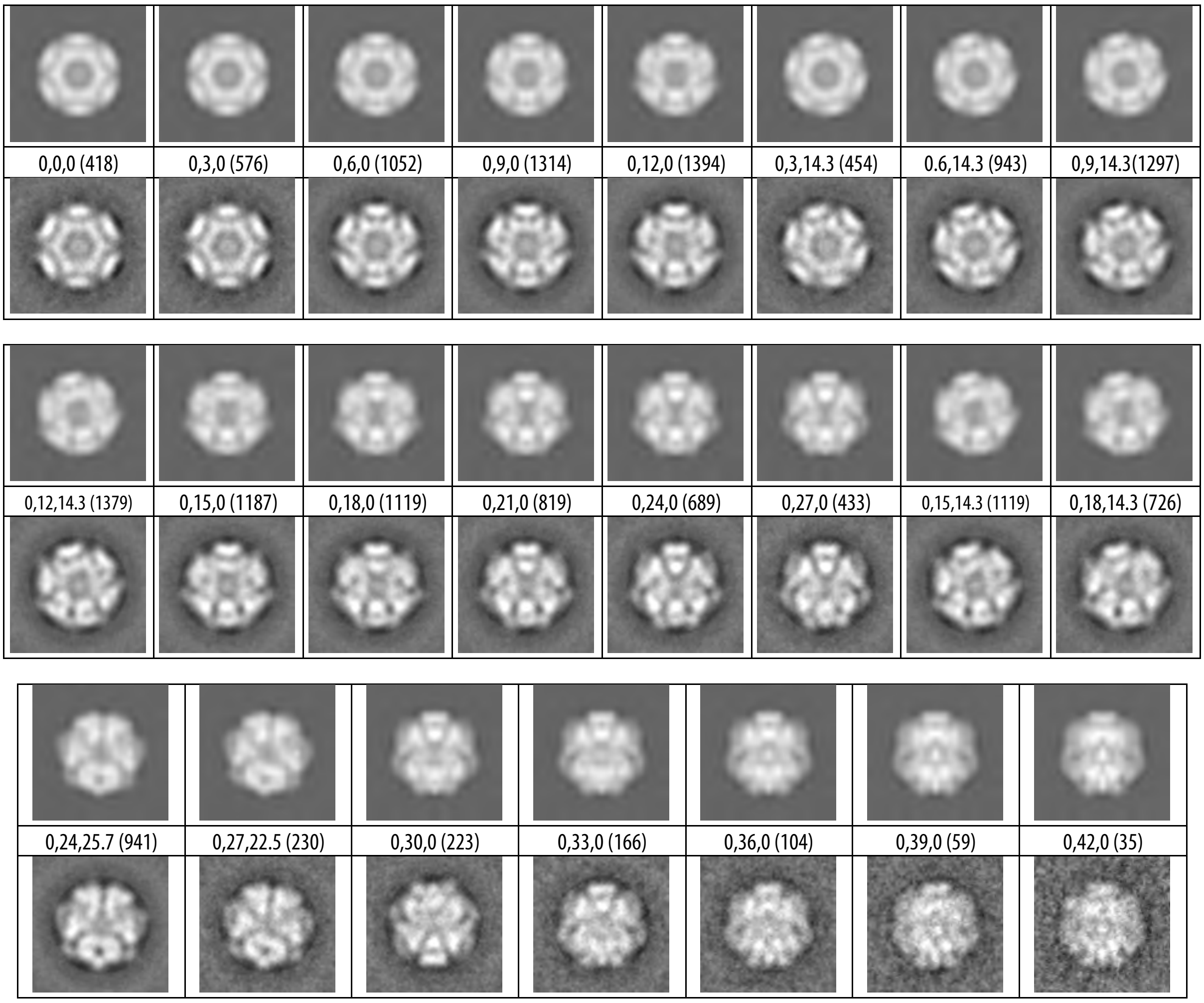


Figure 3.23 Model reference projections against class averages. Between the pair of images of each class and its model reference projection is shown the Euler angles used in Spider functions and the number of particles in the class.

Visually there is a close resemblance in each of the classes to the reference model projection, which is indicative that model bias was not introduced to the structure resolution due to the use of a PDB starting model, which was used to aid the initial alignment of particles.

3.4.2 Model Resolution using Fourier Shell Correlation

The marginal increase of FSC values as can be seen in Figure 3.24 towards the Nyquist frequency, is indicative of some reconstruction artifacts. In the context of this reconstruction, tight masking was essential, and this phenomenon can be expected at the higher frequencies. Given the relatively marginal increase observed, the resultant resolution obtained may marginally be over-estimated.

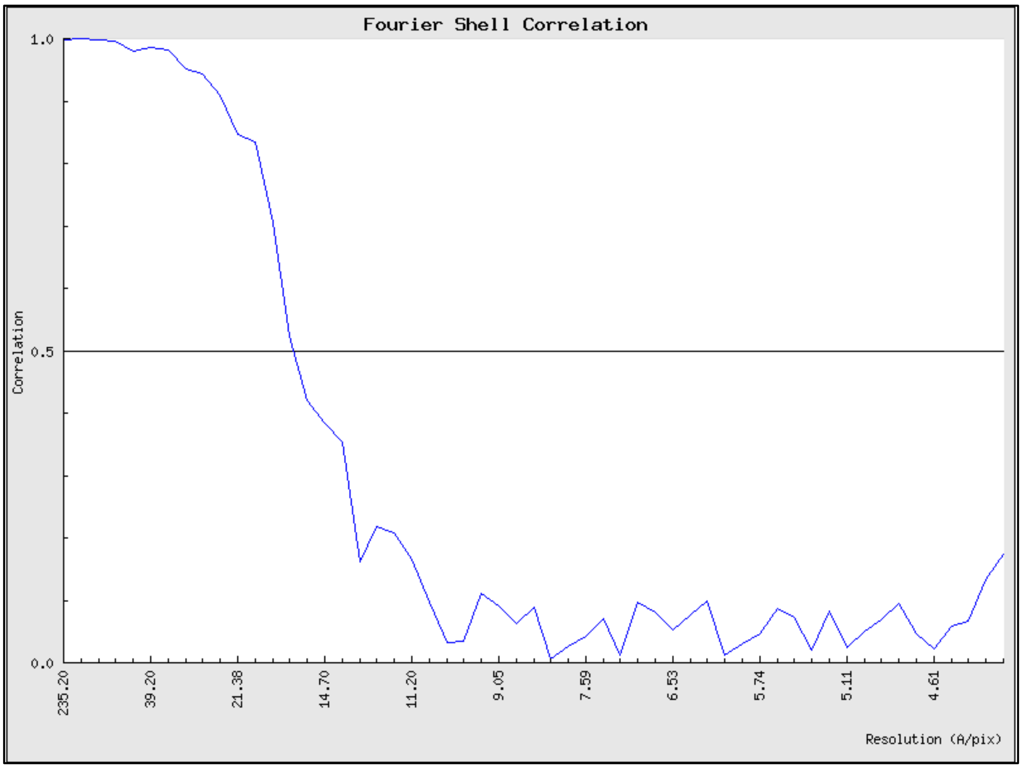


Figure 3.24 FSC plot of the final 3-D construction.

Using the practice of FSC at 0.5, the overall resolution of the reconstruction is 16 Å,

3.4.3 Negative Stain inherent Limitations

In the preparation of the EM grid, there is deposition of the stain over the specimen (protein). The resulting dehydration may result in flattening of the specimen (protein). This may result in the diameter of the molecule being overestimated. Hence using the density of the negative stain 3-D reconstruction volume for fitting any homology model will have a similar impact.

3.4.4 Homogeneity of *Pf*GS produced by the Purification Protocol

The resulting reconstruction to a dodecamer is indicative of the single step purification protocol being robust enough to be used for further structural studies.

Chapter 4 – Homology Modelling and Docking

This chapter describes the basis for the prediction of *Pf*GS structure and the docking of this predicted structure in the 3-D reconstructed volume in the previous section. Multiple alignments of *Pf*GS with other GS depositions in PDB are also indicated.

4.1 Introduction

The purpose of Homology Modelling is to be able to extrapolate an atomic resolution structure from an associated structure for a known protein sequence that has been previously solved. With the increasing availability of experimental structural data for diverse protein families, the predictive (extrapolative) models are constantly being enhanced. In general, the higher the sequences similarity of the gene of interest to the known atomic structure, the higher the confidence of the possession of similar structures. As a rule of thumb, more than a 30-percentage point sequence similarity between the gene of interest and a known atomic structure, for sequences of more than 250 amino acids, may produce meaningful structure prediction.

Additionally, success also depends on the extent of gaps and insertions between the protein of interest and the template sequence alignments. However, dealing with gaps/insertions greater than eight in a particular domain can be very challenging.

4.2 Methods

4.2.1 Automatic Model Prediction

The PSIPRED (Buchan *et al.*, 2013) server was used for structure prediction using the BioSerf 2 automated homology modelling protocol and visualization of the model in CHIMERA (Meng *et al.*, 2006).

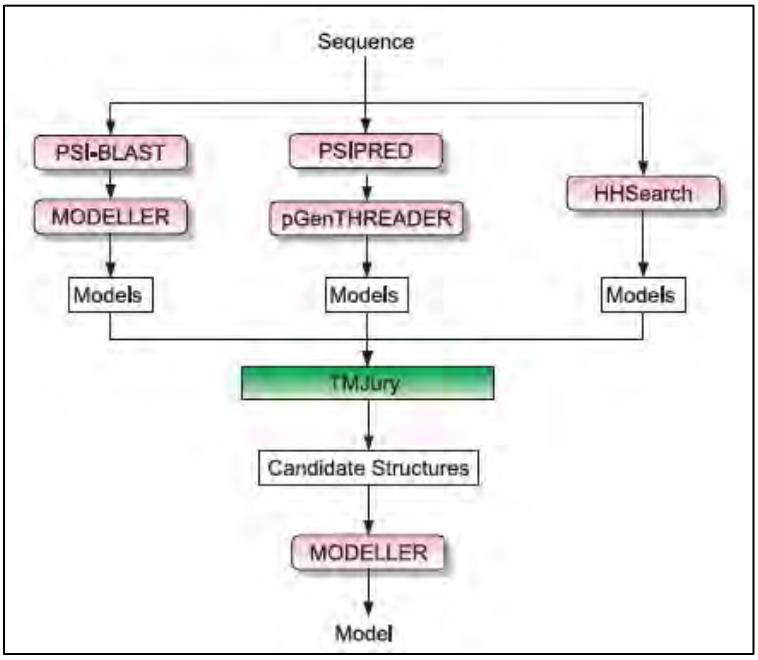


Figure 4.1 Flowchart of the BioSerf2 automated homology modelling protocol.Incoming query sequences are independently matched to PDB chains using PSIBLAST and pGenTHREADER and HH. The three sets of models produced are then compared by TMJury process, which produces up to 10 candidate homologous structures. These structures are their alignments to the query sequences and are used as input for MODELLER to produce a single final model. (*Reproduced from Buchan et al. (2013)*)

The BioSerf2 automated protocol as explained in Figure 4.1, also outputs the pGenThreader (Lobley *et al.*, 2009) and PSIPRED (Jones, 1999) scoring which is used to recognise the closest structural homologs and secondary folding probabilities during the computational protocol.

The predicted model was also independently validated using programs ProSA (Sippi, 1993 & Wiederstein, *et al.*, 2007) and Verify-3D (Bowie, *et al.*, 1991 & Luethy, *et al.*, 1992). ProSA compares the predicted model to current PDB models deposited in PDB and an energy plot as a function of sequence position. Problematic parts of the model can then be identified based on residues that show up as positive energy levels. Verify-3D on the other hand attaches to each residue three attributes obtained from the query 3D protein structure. These are:

i) the total area of the side chain area covered by other protein atoms
ii) the fraction of the side chain area that is covered by polar atoms
iii) the local secondaty structure.

A composite score, referred to as 3D-1D score for each position of the query sequence is calculated. Using the 3D-1D scores in the Veryfy-3D algorithm in a moving-window scan against the database of known sequences is able to identify incorrectly modeled segments when the compatibility score is below 0.2.

4.2.2 Multiple Sequence Alignments with *Pf*GS

Multiple sequence alignments were done on the EMBL-EMI server using Clustal W (Larkin *et al.*, 2007).

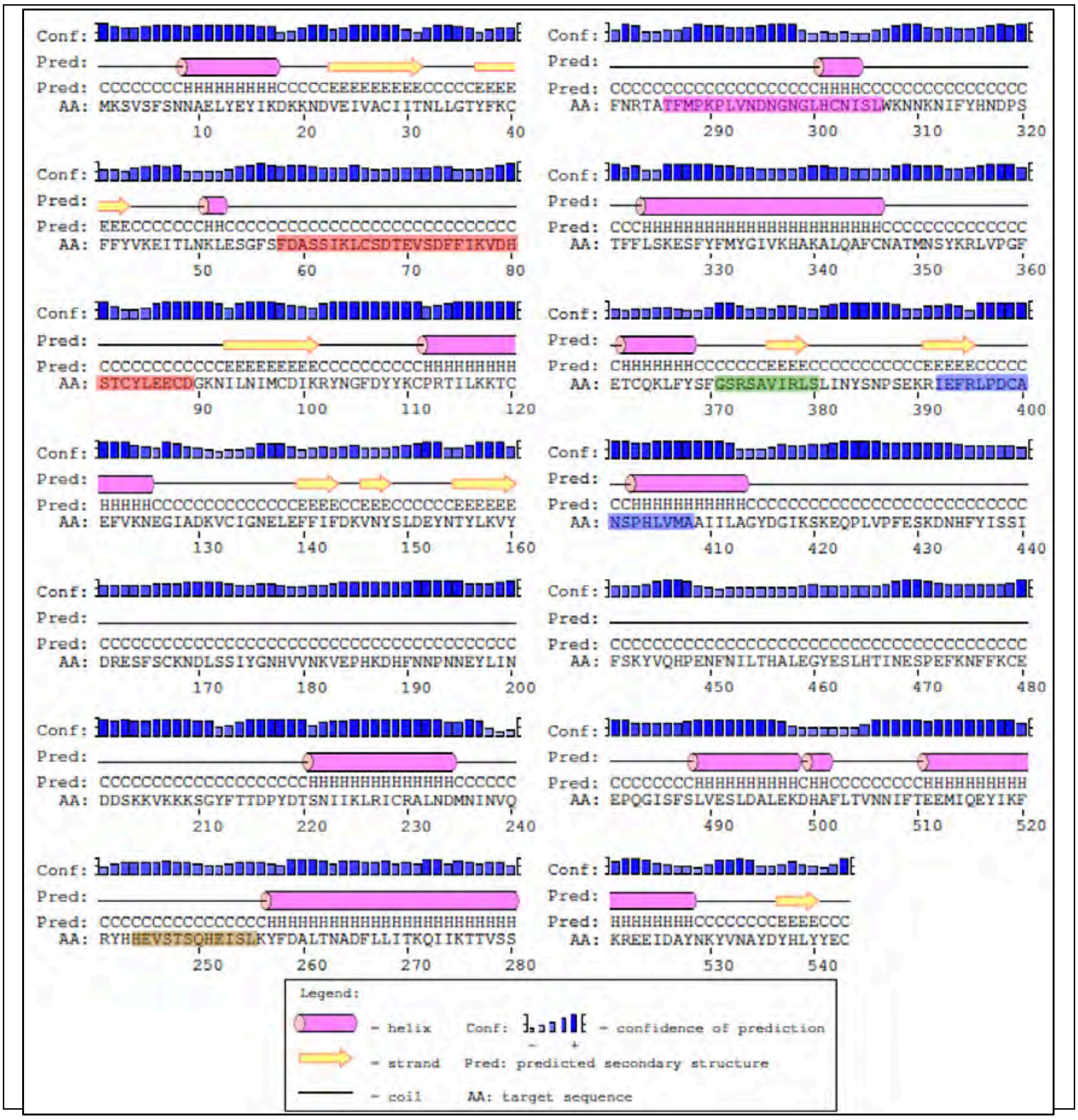
4.2.3 Docking the Predicted *Pf*GS Model into the 3-D Resolved EM Map of *Pf*GS

Rigid docking of the model was undertaken using two of the Situs (Wriggers, 2010) functions,

- “colores” to dock the monomer to the reconstructed 3-D map
- “pdbsymm” to construct the macro-molecule to a D6 symmetry.

4.3 Results

4.3.1 Secondary Fold Predictions and Confidence Scores



```
Conf:
Pred:
Pred: CCCCCCCCHHHHHHHHHCCCCCEEEEEEEEEECCCCCEEEE
  AA: MKSVSFSNNAELYEYIKDKKNDVEIVACIITNLLGTYFKC
              10         20         30         40

Conf:
Pred:
Pred: EEECCCCCCCHHCCCCCCCCCCCCCCCCCCCCCCCCCCCC
  AA: FFYVKEITLNKLESGFSFDASSIKLCSDTEVSDFFIKVDH
              50         60         70         80

Conf:
Pred:
Pred: CCCCCCCCCCCCEEEEEEEEEECCCCCCCCCHHHHHHHHH
  AA: STCYLEECDGKNILNIMCDIKRYNGFDYYKCPRTILKKTC
              90        100        110        120

Conf:
Pred:
Pred: HHHHHCCCCCCCCCCCCCCCEEEECCEEECCCCCCEEEEEE
  AA: EFVKNEGIADKVCIGNELEFFIFDKVNYSLDEYNTYLKVY
             130        140        150        160

Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
  AA: DRESFSCKNDLSSIYGNHVVNKVEPHKDHFNNPNNEYLIN
             170        180        190        200

Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHCCCCCC
  AA: DDSKKVKKKSGYFTTDPYDTSNIIKLRICRALNDMNINVQ
             210        220        230        240

Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHH
  AA: RYHHEVSTSQHEISLKYFDALTNADFLLITKQIIKTTVSS
             250        260        270        280

Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCHHHHCCCCCCCCCCCCCCCC
  AA: FNRTATFMPKPLVNDNGNGLHCNISLWKNNKNIFYHNDPS
             290        300        310        320

Conf:
Pred:
Pred: CCCHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCC
  AA: TFFLSKESFYFMYGIVKHAKALQAFCNATMNSYKRLVPGF
             330        340        350        360

Conf:
Pred:
Pred: CHHHHHHHCCCCCCCEEEECCCCCCCCCCCEEEEECCCCC
  AA: ETCQKLFYSFGSRSAVIRLSLINYSNPSEKRIEFRLPDCA
             370        380        390        400

Conf:
Pred:
Pred: CCHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCC
  AA: NSPHLVMAAIILAGYDGIKSKEQPLVPFESKDNHFYISSI
             410        420        430        440

Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
  AA: FSKYVQHPENFNILTHALEGYESLHTINESPEFKNFFKCE
             450        460        470        480

Conf:
Pred:
Pred: CCCCCCCCHHHHHHHHHHHCHHCCCCCCCCCHHHHHHHHHH
  AA: EPQGISFSLVESLDALEKDHAFLTVNNIFTEEMIQEYIKF
             490        500        510        520

Conf:
Pred:
Pred: HHHHHHHHCCCCCCCCEEEEECCC
  AA: KREEIDAYNKYVNAYDYHLYYEC
             530        540
```

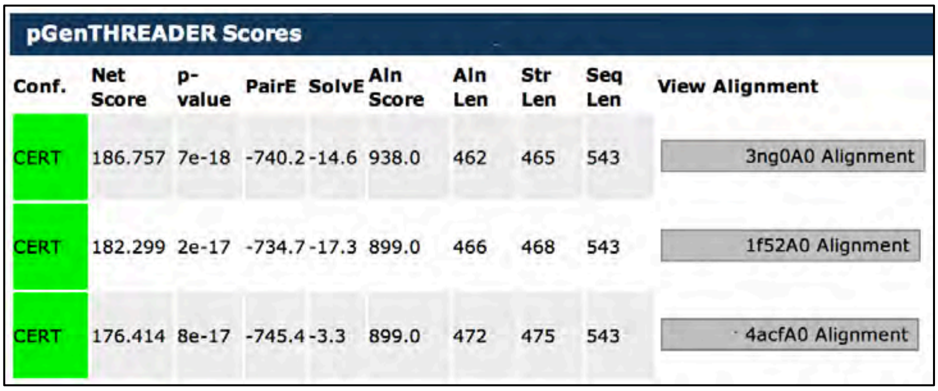
Figure 4.2 PSIPRED Protein fold predictions for *Pf*GS outputted using the BioSerf2 protocol. The confidence levels of the predicted fold are indicated by the blue bar above each amino acid (*Reproduced from BioSerf2 output*)

All fold predictions, as indicated in Figure 4.2, have a high level of confidence except for three zones between 299-306, 361-369 and 449-458.

4.3.2 Structural Homolog Predictions and Scoring

The outputs of the three highest scoring structural alignments from pGenThreader are indicted in Table 4.1. Reporting of p-value statistic in column 3 is compared to the traditional threshold value less than 0.05 as a significant result is indicated in column 1. P-

GenTHREADER classifies a p-value of less than 0.00001 as CERT., less than 0.001 as HIGH and less than 0.01 as MEDIUM.



pGenTHREADER Scores

Conf.	Net Score	p-value	PairE	SolvE	Aln Score	Aln Len	Str Len	Seq Len	View Alignment
CERT	186.757	7e-18	-740.2	-14.6	938.0	462	465	543	3ng0A0 Alignment
CERT	182.299	2e-17	-734.7	-17.3	899.0	466	468	543	1f52A0 Alignment
CERT	176.414	8e-17	-745.4	-3.3	899.0	472	475	543	4acfA0 Alignment

Table 4.1 Top three scoring organisms from pGenTHREADER blast. Namely, PDB:3NG0 (*Synechocystis sp*), PDB:1F52 (*Salmonella Typhimurium*) and PBD:4ACF (*Mycobacterium tuberculosis*) (*Reproduced from BioSerf2 output*)

The top three structures reported from the pGenTRHEADER scoring all have a p-value score less than 8 x 10^{-17} indicative of a good structural similarity. *Synechocystis sp* (PDB 3NGO) having the highest score at 7 x 10^{-18}, was used for visualisation and comparative purposes.

Pairwise and Solvation energy scores are indicated in column 4 and 5 respectively. Pairwise alignment scores, the length of the alignment and the length of the structural hit are shown in the next three columns. Column 9 indicates the query sequence length (*Pf*GS in this case). The button in the View Alignment column links out to enable viewing an annotated alignment.

4.3.3 Structure Prediction

The structure predicted from the BioSerf2 protocol for *Pf*GS is depicted in Figure 4.3 as visualized in CHIMERA.

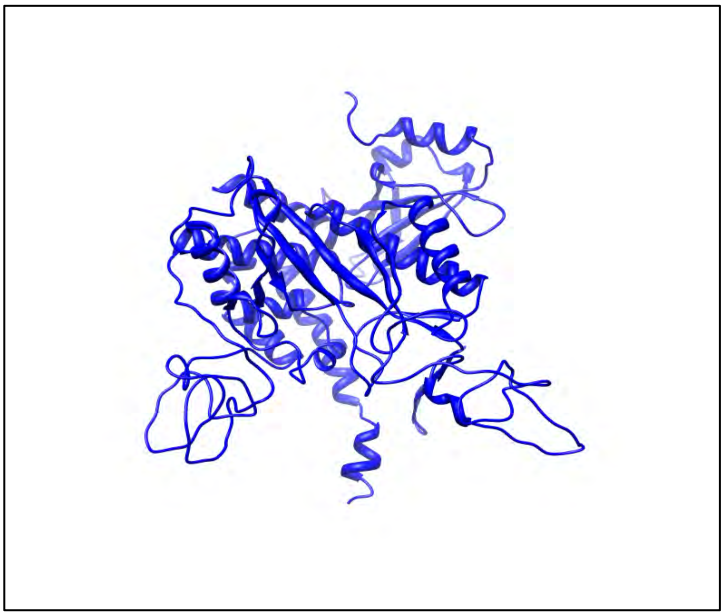


Figure 4.3 *Pf*GS predicted model visualized in CHIMERA.

Comparing the structure folds in the predicted model of the monomer and the expected secondary fold predictions of PSIPRED, it was found to be in agreement.

4.3.4 Predicted Structure Validation

Using ProSA the predicted model scores Z= -5.66, which falls within the range of scores derived from PDB structures of similar number of residues as shown in figure 4.4

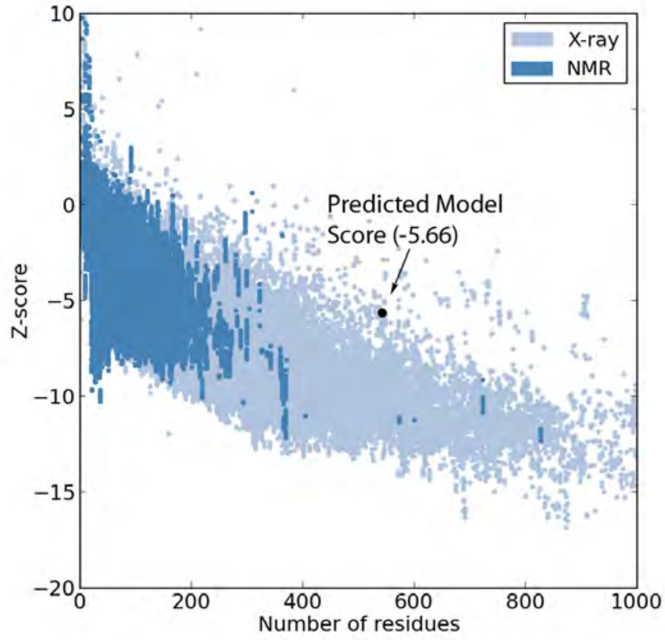


Figure 4.4 *Pf*GS predicted model Z-score indicated against PDB deposited structures derived from X-ray crystallography and NMR techniques.

The energy plot averaging 20 residues on either side of the sequence position, indicates

Figure 4.5 *Pf*GS predicted model energy plot (average of 40 residues) indicates two problematic zones as highlighted in the figure.

two problematic zones as highlighted in figure 4.5. These zones co-incide with the two large insertion loops as identified in the sequence alignment with type 1 glutamine synthetases' in figure 4.8.

Verify-3D scores for 3D-1D averaged scores below 0.2 are considered problematic.

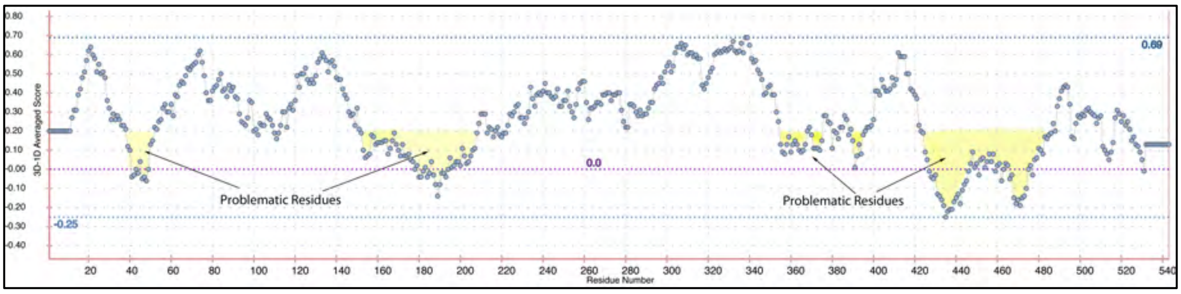


Figure 4.6 *Pf*GS predicted model 3D-1D Averaged scores plotted against residue sequence

Verify-3D results for the predicted *Pf*GS also confirms the two zones with the large insertions to be problematic. Two further zones centered at residue 44 and 380 using the Verify-3D methodology are also identified as problematic. The problematic zones are highlighted in figure 4.7.

The Ramachandran plot, as indicated in figure 4.8, shows bulk of the residue phi and psi angles were within the core area and more than 90% of all residues being within the

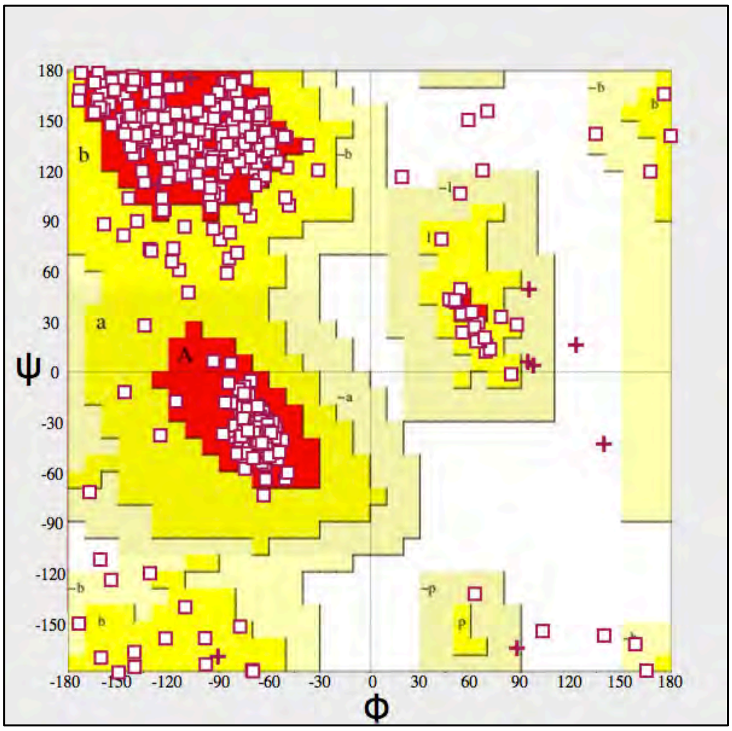


Figure 4.7 Ramachandran plot of *Pf*GS predicted model. Core areas indicated in red and acceptable areas in yellow. Glycines indicated as "+".

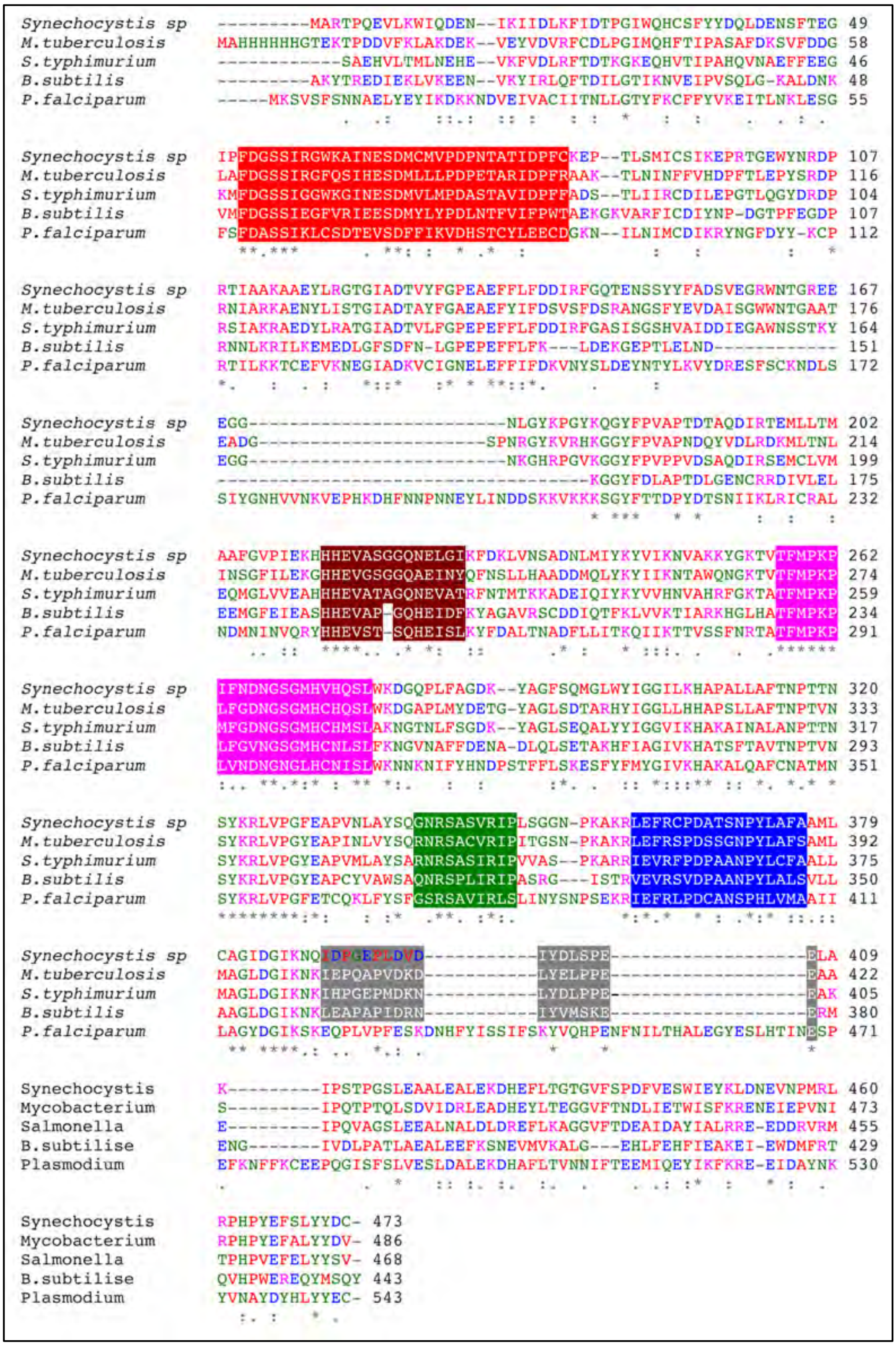
allowed area. Whilst this in itself does not prove that the predicted model is robust, it is however, indicative of a model with conformational stability.

4.3.5 Multiple Sequence Alignment scores of *Pf*GS against GS PBD Depositions

GS Type	Deposition	Chain Length	Organism	*Pf*GS
II	4BAX	344	*Streptomyces Coelicolor*	9
	4IS4	378	*Medicago truncatula*	6
	2OJW	384	*Human*	9
	2D3A	356	*Zea Maize*	5
	3FKY	370	*Saccharomyces cerevisiae*	10
	2UU7	381	*Canis Familiaris*	10
I	4LNI	443	*Bacillus subtilis*	30
	1F52	468	*Salmonella Typhimurium*	34
	3NG0	479	*Synechocystis sp*	34
	4ACF	486	*Mycobacterium tuberculosis*	32
?	n/a	543	*Plasmodium falciparum*	-
III	2O6X	729	*Bacteroides fraglis*	10

Table 4.2 Multiple alignment scores of *Pf*GS against other PDB depositions for glutamine synthetase using Clustral W.

The multiple alignment scores in Table 4.2 indicate that *Pf*GS has greater than 30% similarity to GS I and less than 10% to the other two GS types. A multiple alignment of *Pf*GS with the four GS I sequences is indicated the Figure 4.4.



```
Synechocystis sp    ---------MARTPQEVLKWIQDEN--IKIIDLKFIDTPGIWQHCSFYYDQLDENSFTEG 49
M.tuberculosis      MAHHHHHHGTEKTPDDVFKLAKDEK--VEYVDVRFCDLPGIMQHFTIPASAFDKSVFDDG 58
S.typhimurium       ------------SAEHVLTMLNEHE--VKFVDLRFTDTKGKEQHVTIPAHQVNAEFFEEG 46
B.subtilis          ---------AKYTREDIEKLVKEEN--VKYIRLQFTDILGTIKNVEIPVSQLG-KALDNK 48
P.falciparum        -----MKSVSFSNNAELYEYIKDKKNDVEIVACIITNLLGTYFKCFFYVKEITLNKLESG 55
                                .  .:    ::.:  :: :   : :  *   :  :    .  . : .

Synechocystis sp    IPFDGSSIRGWKAINESDMCMVPDPNTATIDPFCKEP--TLSMICSIKEPRTGEWYNRDP 107
M.tuberculosis      LAFDGSSIRGFQSIHESDMLLLPDPETARIDPFRAAK--TLNINFFVHDPFTLEPYSRDP 116
S.typhimurium       KMFDGSSIGGWKGINESDMVLMPDASTAVIDPFFADS--TLIIRCDILEPGTLQGYDRDP 104
B.subtilis          VMFDGSSIEGFVRIEESDMYLYPDLNTFVIFPWTAEKGKVARFICDIYNP-DGTPFEGDP 107
P.falciparum        FSFDASSIKLCSDTEVSDFFIKVDHSTCYLEECDGKN--ILNIMCDIKRYNGFDYY-KCP 112
                      **.***      .  **: :  *  .*  :            :   :        :   *

Synechocystis sp    RTIAAKAAEYLRGTGIADTVYFGPEAEFFLFDDIRFGQTENSSYYFADSVEGRWNTGREE 167
M.tuberculosis      RNIARKAENYLISTGIADTAYFGAEAEFYIFDSVSFDSRANGSFYEVDAISGWWNTGAAT 176
S.typhimurium       RSIAKRAEDYLRATGIADTVLFGPEPEFFLFDDIRFGASISGSHVAIDDIEGAWNSSTKY 164
B.subtilis          RNNLKRILKEMEDLGFSDFN-LGPEPEFFLFK---LDEKGEPTLELND------------ 151
P.falciparum        RTILKKTCEFVKNEGIADKVCIGNELEFFIFDKVNYSLDEYNTYLKVYDRESFSCKNDLS 172
                    *.   :  . :   *::*   :* * **::*.    .     :

Synechocystis sp    EGG-------------------------NLGYKPGYKQGYFPVAPTDTAQDIRTEMLLTM 202
M.tuberculosis      EADG----------------------SPNRGYKVRHKGGYFPVAPNDQYVDLRDKMLTNL 214
S.typhimurium       EGG-------------------------NKGHRPGVKGGYFPVPPVDSAQDIRSEMCLVM 199
B.subtilis          ------------------------------------KGGYFDLAPTDLGENCRRDIVLEL 175
P.falciparum        SIYGNHVVNKVEPHKDHFNNPNNEYLINDDSKKVKKKSGYFTTDPYDTSNIIKLRICRAL 232
                                                        * ***   * *     :  :   :

Synechocystis sp    AAFGVPIEKHHHEVASGGQNELGIKFDKLVNSADNLMIYKYVIKNVAKKYGKTVTFMPKP 262
M.tuberculosis      INSGFILEKGHHEVGSGGQAEINYQFNSLLHAADDMQLYKYIIKNTAWQNGKTVTFMPKP 274
S.typhimurium       EQMGLVVEAHHHEVATAGQNEVATRFNTMTKKADEIQIYKYVVHNVAHRFGKTATFMPKP 259
B.subtilis          EEMGFEIEASHHEVAP-GQHEIDFKYAGAVRSCDDIQTFKLVVKTIARKHGLHATFMPKP 234
P.falciparum        NDMNINVQRYHHEVST-SQHEISLKYFDALTNADFLLITKQIIKTTVSSFNRTATFMPKP 291
                       .. ::  ****.. .* *:  ::      .* :   * :::. .   .  .******

Synechocystis sp    IFNDNGSGMHVHQSLWKDGQPLFAGDK--YAGFSQMGLWYIGGILKHAPALLAFTNPTTN 320
M.tuberculosis      LFGDNGSGMHCHQSLWKDGAPLMYDETG-YAGLSDTARHYIGGLLHHAPSLLAFTNPTVN 333
S.typhimurium       MFGDNGSGMHCHMSLAKNGTNLFSGDK--YAGLSEQALYYIGGVIKHAKAINALANPTTN 317
B.subtilis          LFGVNGSGMHCNLSLFKNGVNAFFDENA-DLQLSETAKHFIAGIVKHATSFTAVTNPTVN 293
P.falciparum        LVNDNGNGLHCNISLWKNNKNIFYHNDPSTFFLSKESFYFMYGIVKHAKALQAFCNATMN 351
                    :.. **.*:* : ** *:.   :  :      :*. .  :: *:::** :: *. *.* *

Synechocystis sp    SYKRLVPGFEAPVNLAYSQGNRSASVRIPLSGGN-PKAKRLEFRCPDATSNPYLAFAAML 379
M.tuberculosis      SYKRLVPGYEAPINLVYSQRNRSACVRIPITGSN-PKAKRLEFRSPDSSGNPYLAFSAML 392
S.typhimurium       SYKRLVPGYEAPVMLAYSARNRSASIRIPVVAS--PKARRIEVRFPDPAANPYLCFAALL 375
B.subtilis          SYKRLVPGYEAPCYVAWSAQNRSPLIRIPASRG---ISTRVEVRSVDPAANPYLALSVLL 350
P.falciparum        SYKRLVPGFETCQKLFYSFGSRSAVIRLSLINYSNPSEKRIEFRLPDCANSPHLVMAAII 411
                    ********:*:   : :*  .**. :*:.          *:*.*  * : .*:* ::.::

Synechocystis sp    CAGIDGIKNQIDPGEPLDVD-----------IYDLSPE-------------------ELA 409
M.tuberculosis      MAGLDGIKNKIEPQAPVDKD-----------LYELPPE-------------------EAA 422
S.typhimurium       MAGLDGIKNKIHPGEPMDKN-----------LYDLPPE-------------------EAK 405
B.subtilis          AAGLDGIKNKLEAPAPIDRN-----------IYVMSKE-------------------ERM 380
P.falciparum        LAGYDGIKSKEQPLVPFESKDNHFYISSIFSKYVQHPENFNILTHALEGYESLHTINESP 471
                     **  ****.:  ..  *.: .            *    *                   *

Synechocystis       K---------IPSTPGSLEAALEALEKDHEFLTGTGVFSPDFVESWIEYKLDNEVNPMRL 460
Mycobacterium       S---------IPQTPTQLSDVIDRLEADHEYLTEGGVFTNDLIETWISFKRENEIEPVNI 473
Salmonella          E---------IPQVAGSLEEALNALDLDREFLKAGGVFTDEAIDAYIALRRE-EDDRVRM 455
B.subtilise         ENG-------IVDLPATLAEALEEFKSNEVMVKALG---EHLFEHFIEAKEI-EWDMFRT 429
Plasmodium          EFKNFFKCEEPQGISFSLVESLDALEKDHAFLTVNNIFTEEMIQEYIKFKRE-EIDAYNK 530
                    .             .  *   :: :. :.  :.  .    ..:  :*  :   * :  .

Synechocystis       RPHPYEFSLYYDC- 473
Mycobacterium       RPHPYEFALYYDV- 486
Salmonella          TPHPVEFELYYSV- 468
B.subtilise         QVHPWEREQYMSQY 443
Plasmodium          YVNAYDYHLYYEC- 543
                      :. :   * .
```

Figure 4.8 *Pf*GS multiple sequence alignment with GS I PBD depositions. The five conserved regions are indicated according to the following colour key: Red=Region1 – Latch; Brown=Region2 – Enzyme Active Site; Magenta=Region3 ATP binding site; Green=Region4 Glutamate Binding Site; Blue=Region5 Enzyme Active Site and Grey= Adenylylation Loop (*adapted from Clustal W output*)

The multiple alignments in Figure 4.4 clearly indicate the two large insertion areas in *Pf*GS compared to the other GS I sequences. The second insertion area is in the vicinity of the adenylylation loop (residues 412 to 482) with *Pf*GS having three tyrosine residues in this vicinity. GS I with conserved adenylylation loop has a single tyrosine residue in this vicinity. This observation may indicate that *Pf*GS activity is not inhibited by adenylylation.

4.3.6 Rigid Docking of Predicted Model

The outcome of docking the predicted model in the 3-D reconstructed model is indicated in Figures 4.9 to 4.11.

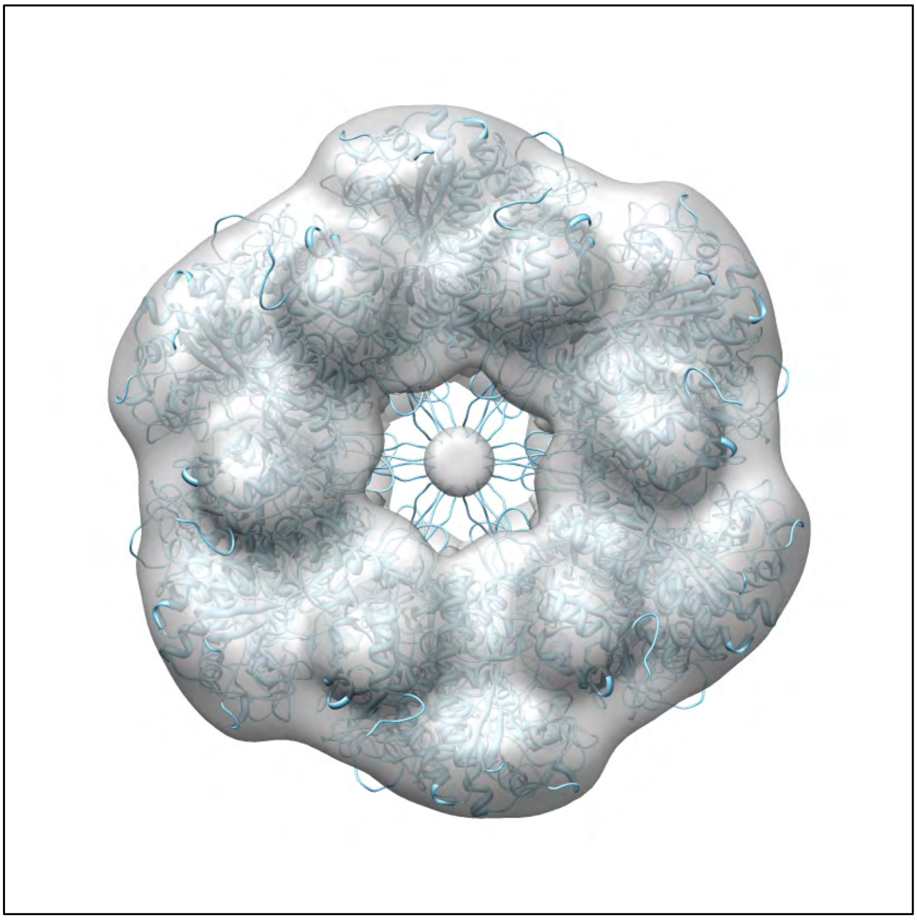


Figure 4.9 Top view of the rigid docking of predicted *Pf*GS model using Situs functions. The reconstructed density does not cover a substantial length of the inner loop of the predicted model.

The reconstructed 3-D model density, as can be seen in Figure 4.5, does not cover the inner loop of the predicted model fully. Unoccupied density cannot be located in the vicinity from this view. The rest of the visible predicted structure is contained within the reconstructed density.

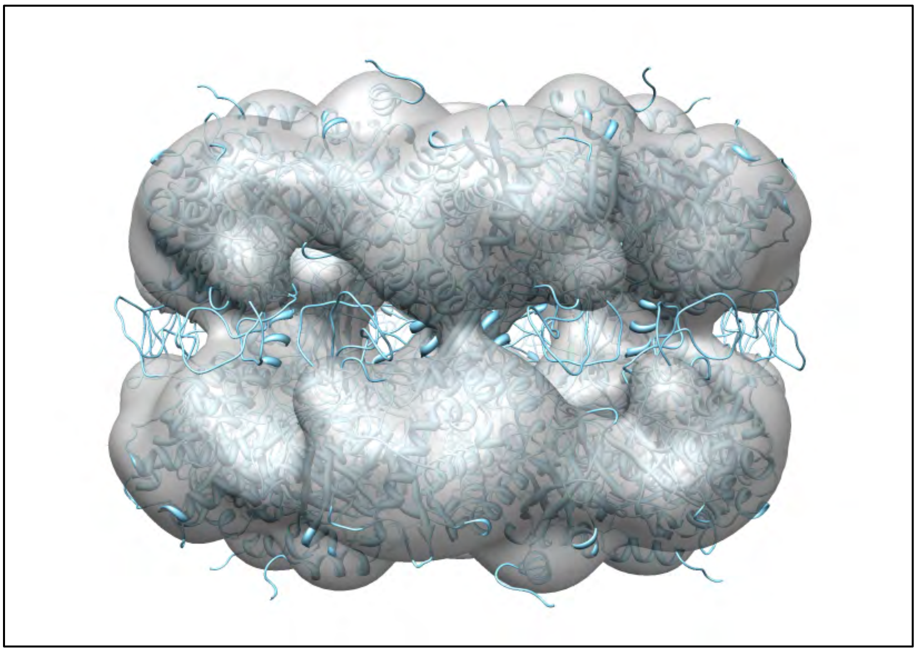


Figure 4.10 A side view of the rigid docking indicates the reconstructed density does not cover a substantial length of the outer loop of the predicted model.

The reconstructed 3-D model density, as can be seen in Figure 4.6, does not cover the outer loop of the predicted model fully. Unoccupied density from this view cannot be located in the vicinity of the loop. The rest of the visible predicted structure is contained within the reconstructed density.

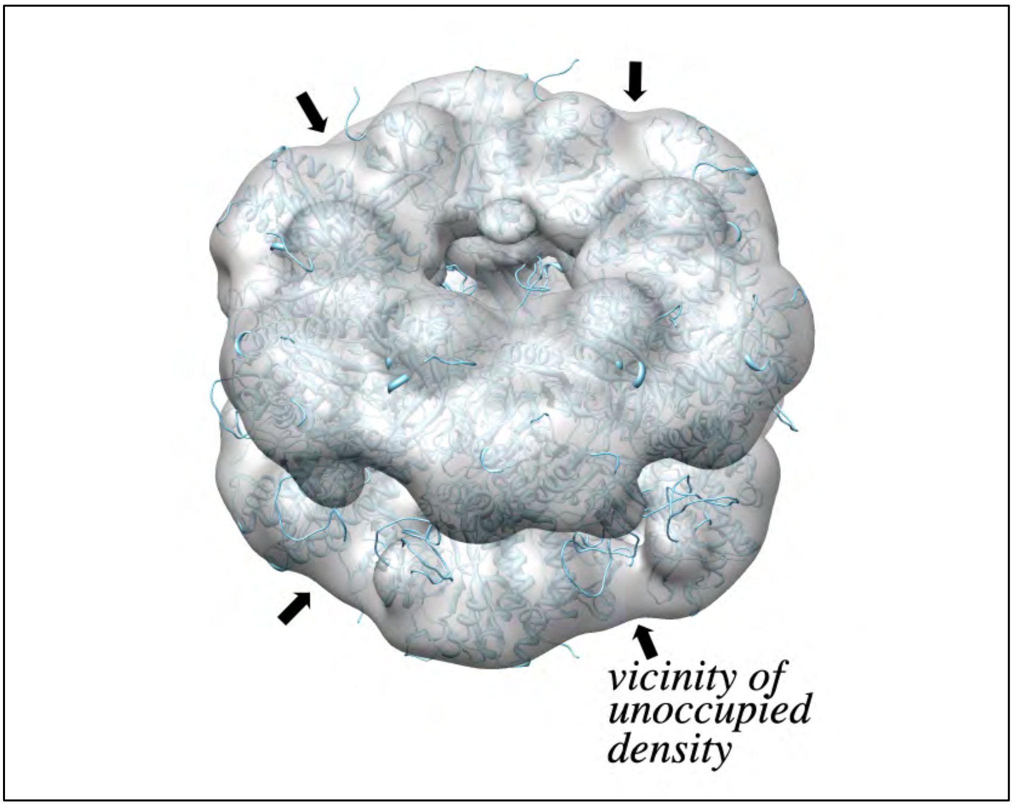


Figure 4.11 Another view of the rigid docking better indicating some unoccupied density in the vicinity of the outer loop.

A slightly different view of the docking in Figure 4.7 indicates some unoccupied density in the vicinity of the outer loop. However, this density is unlikely to account for the total visible length of the outer loop in the predicted model.

4.3.7 Surface interactions

Four subunits of the *Pf*GS predicted model are docked in Figure 4.12, indicate that the surface interactions that hold the macromolecules together are similar to that of GS I. The subunits are held together in a ring by heterologous contacts (red-blue and green-yellow sub-units in Figure 4.8) while the C-terminal helices make isologous contacts (green-red and yellow-blue in Figure 4.8) to bind the two rings together. In the case of the first crystallized GS III, van Rooyen, *et al*. (2011) surprisingly showed the N-terminal helices make the isologous contact between the rings.

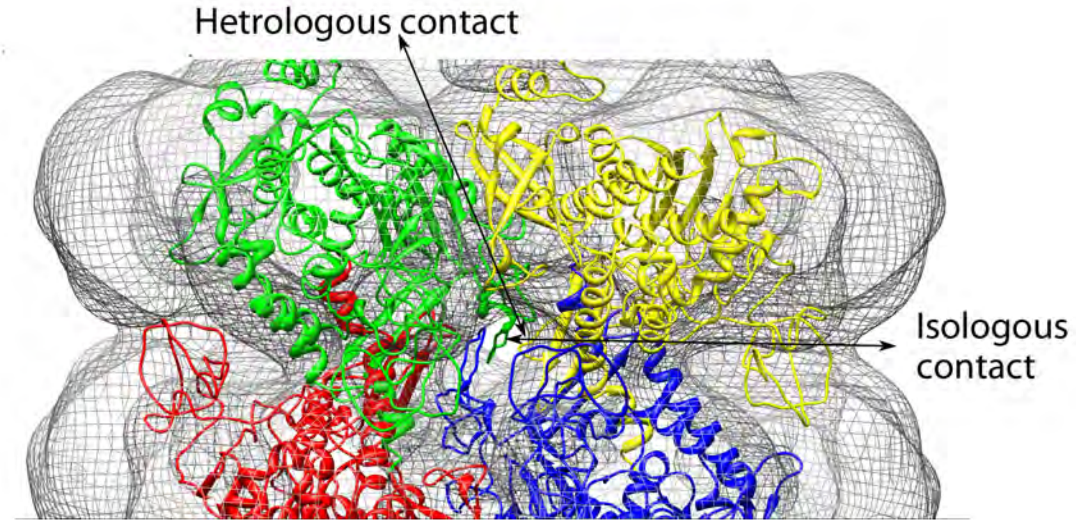


Figure 4.12 Four subunits docked in the 3-D reconstruction indicate that the macromolecule is held together in a similar manner as GS I.

4.3.8 Visualisation of Predicted model against PDB:3NGO

Visualisation in Chimera of the homolog PDB-3NG0 to the predicted *Pf*GS structure is shown in Figure 4.9.

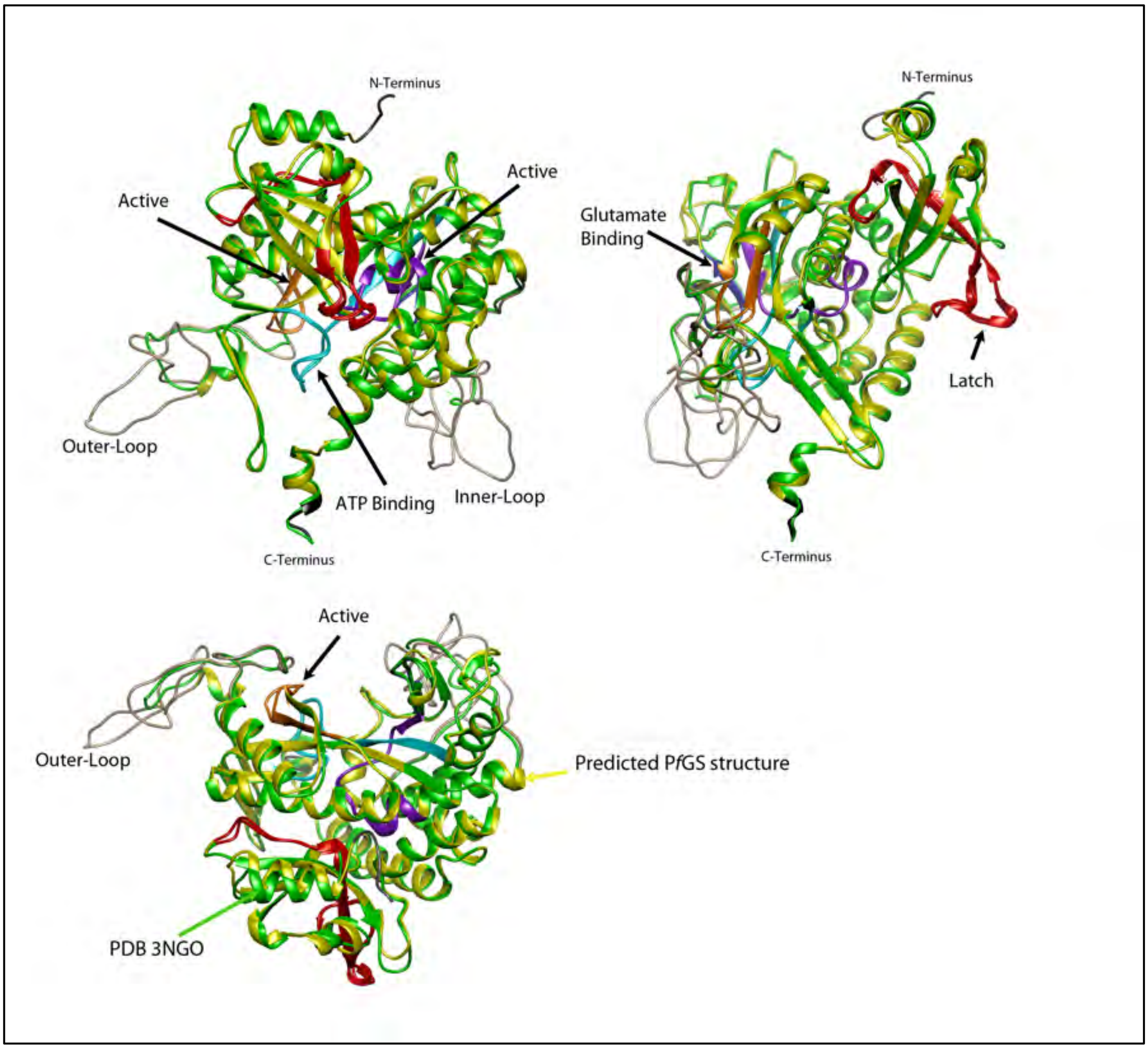


Figure 4.13 Visualisation of the *Pf*GS predicted model aligned to PDB-3NG0. The conserved regions have been highlighted in different colors. The significant larger inner and outer loops for *Pf*GS have also been indicated.

The inner loop shows itself up at the center of the macromolecule while the outer loop shows itself at the outer periphery of the macromolecule at the intersection of the two half rings. This is apparent in the docking images in Figure 4.5 to Figure 4.7.

The overall RMSD between the predicted *Pf*GS model, without the two insertions and the PDB:3NG0 structure is 0.6 Å, which is indicative of a very close structural alignment with the rest of the structure. Similar RMSD’s were obtained with the other two high similarity PDB structures, 1F52 and 4ACF. Whilst the homology model is not a substitute for a crystal structure, early indications from the docking of the homology model suggests that

*Pf*GS would be similar to a GS I structure, but with two larger flexible loops. In this regard, it can be expected that the structure function, as reviewed by Eisenberg *et al*. (2000), and summarized in section 1.2.5.3, would be applicable to *Pf*GS.

4.3.9 Active Site Positioning at the Heterologous Interface

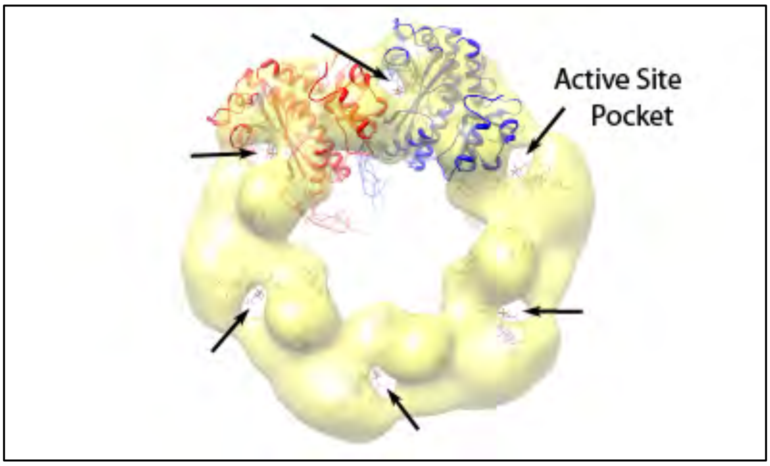


Figure 4.14 An increased threshold level of the 3-D volume density creates an opening in the vicinity of the two sub-unit interface.

In Figure 4.10, the threshold of the 3-D reconstructed volume density was increased, creating an opening on the volume map in the vicinity of the sub-unit interface, which also locates the active site entrance.

4.3.10 Visualisation of Active Site Relative to *Mycobacterium tuberculosis* GS and *Canis familiaris* GS (as proxy for Human GS)

The *Canis familiaris* was used as a proxy for the Human GS on the basis of a 96% sequence similarity score (Table 1.4) and *Mycobacterium tuberculosis,* PDB:2BVC was chosen to leverage the extensive functional studies undertaken by Eisenberg *et al*. (2000). Each sub-unit of GS from the respective organisms was aligned with the adjacent subunit assuming a resultant position (Fig 4.12) as positioned on the respective macromolecule.

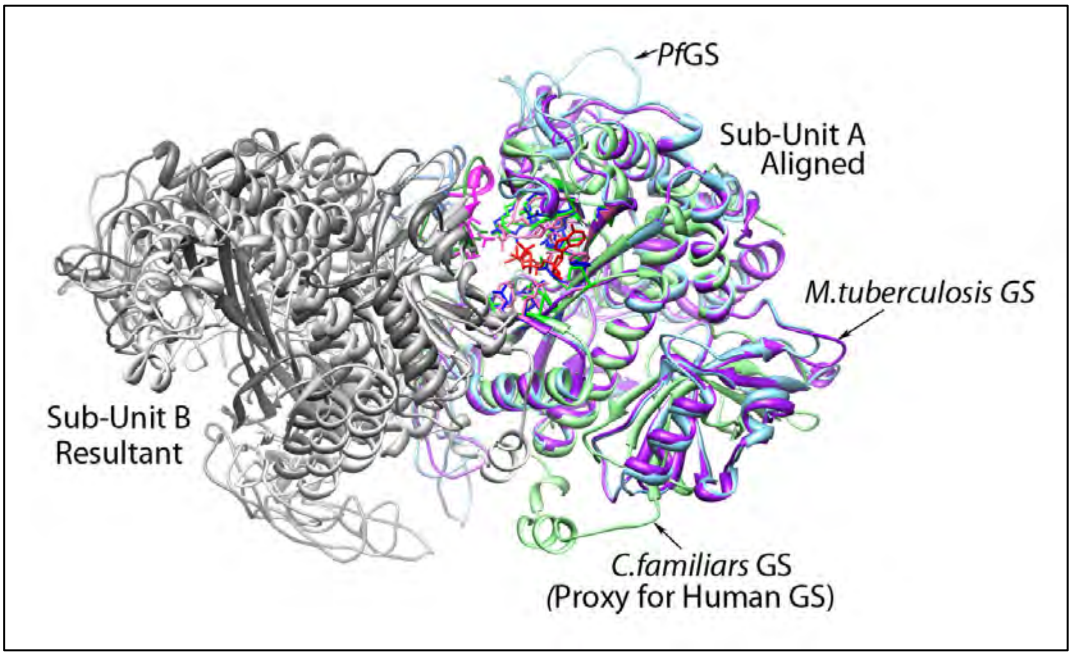


Figure 4.15 *Pf*GS homology model alignment to *M.tuberculosis* GS and *C.familiaris* GS. Sub-unit A is used for the alignment, making subunit B resultant based on the respective macromolecule position.

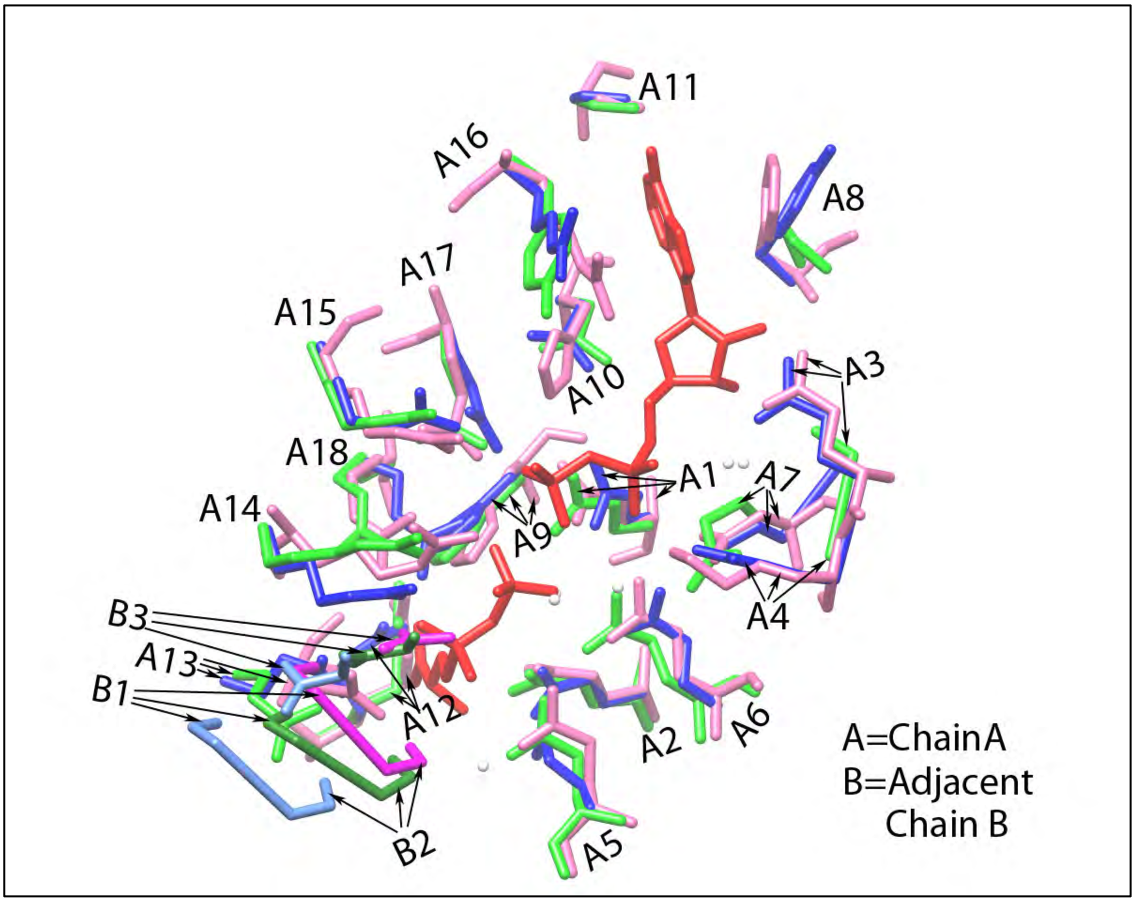


Figure 4.16 Active site *Pf*GSpredicted model against *Mycobacterium tuberculosis*, PDB:2BVC and *Canis familiaris* PDB:2UU7. The active site is completed by three amino acids from the adjacent monomer, indicated by B1,B2,B3. Colour Key: Red=Ligands; Green= *Canis familiaris GS*; Hot Pink= *Mycobacterium tuberculosis GS;* Blue=*Pf*GS. Ligands shown are ADP and P3S. The cations have been excluded for clarity.

The active site zoomed in to show the side chains for *Pf*GS are indicated in Figure 4.12, against the *Mycobacterium tuberculosis,* PDB:2BVC and *Canis familiaris* PDB:2UU7 structures.

Three amino acids, two serine and one aspartic acid from the adjacent sub-unit complete the active site. The RMSD for each of the conserved active site amino acids is shown in Table 4.3.

Conserved	***Plasmodium falciparum***	***Canis familiaris***	***Mycobacterium tuberculosis***	*(1) - (2)*	*(3) – (2)*	*(4) – (5)*
	(1)	*(2)*	*(3)*	*(4)* Å	*(5)* Å	Å
A1	E-137	E-134	E-133	0.642	0.525	0.1
A2	E-139	E-136	E-135	0.352	0.455	0.1
A3	Q-240	A-191	E-214	2.262	2.639	0.4
A4	Q-241	G-192	K-215	1.704	1.837	0.1
A5	E-245	E-196	E-219	0.369	0.452	0.1
A6	E-252	E-203	E-227	0.819	0.781	0.0
A7	S-254	Q-205	N-229	1.725	1.843	0.1
A8	Y-257	P-208	F-232	1.094	0.955	0.1
A9	H-301	H-253	H-276	0.307	0.507	0.2
A10	N-303	N-255	H-278	0.238	0.39	0.2
A11	S-305	S-257	S-280	0.354	0.112	0.2
A12	R-355	R-299	R-329	0.149	0.238	0.1
A13	E-361	E-305	E-335	0.819	2.264	1.4
A14	R-373	R-319	R-347	0.189	1.052	0.9
A15	R-378	R-324	R-352	0.51	0.788	0.3
A16	R-391	Y-336	R-364	0.408	0.504	0.1
A17	E-393	E-338	E-366	0.182	0.248	0.1
A18	R-395	R-340	R-368	0.108	0.287	0.2
B1	D-59	D-63	D-54	1.448	1.473	0.0
B2	S-61	S-65	S-56	3.544	1.839	1.7
B3	S-62	S-66	S-57	4.032	0.899	3.1

Table 4.3 RMSD of the conserved active site amino acids relative to *Canis familiaris.*

The *Pf*GS and *Mycobacterium tuberculosis* GS share very similar RMSD of the conserved active site amino acids with *Canis familiaris*. The *Pf*GS, has a significantly larger RMSD on two of the conserved active sites on the adjacent sub-unit than *Mycobacterium tuberculosis* GS by 1.7 and 3.1 Å .

4.4 Discussion

Using the BioSerf2 automated protocol provided a homology model with significantly high scores. Rigid docking of the homology model showed that two of the loop zones were not contained in the negative stain EM 3-D reconstructed volume. *Pf*GS has two large insertions not found in any of the GS I structures. Structurally, the two insertions are located in the vicinity of the loops also found in GS I structures, referred to as the inner and outer loop. The inner loop shows itself up at the center of the macromolecule while the outer loop shows itself at the outer periphery of the macromolecule at the intersection of the two half rings of the macromolecule. This is apparent in the docking images in Figure 4.6 to Figure 4.8.

The overall RMSD between the predicted *Pf*GS model, without the two insertions and the PDB:3NG0 structure is 0.6 Å, which is indicative of a very close structural alignment for the rest of the structure. Similar RMSD’s were obtained with the other two high similarity PDB structures, namely 1F52 (*Salmonella Typhimurium* GS) and 4ACF (*Mycobacterium tuberculosis* GS*)*. With such close RMSD’s it can be expected that the structure function as reviewed by Eisenberg *et al*. (2000), and summarized in section 1.2.5.3, would be applicable to *Pf*GS.

However, there was no unoccupied density of sufficient magnitude in the respective vicinities of the loops to occupy the full length of the predicted loops. The loci of these loops are likely to span a large volume and occupy random spatial orientations, thus increasing the possibility of not locating sufficient density to be seen on the 3-D reconstruction. More particles than used in this study for the 3-D reconstruction may locate the density for these loops. Homology modeling does not enable the structure of large inserted loops to be predicted as shown by the ProSA and Verify-3D algorithms that demonstrate poor structures in these parts of the molecule. Furthermore, the loops may not fold into a single structure because the density available in the 3D reconstruction is insufficient to accommodate the entire loops indicting the possibility of disorder.

The RMSD’s for the active site residues between the Human GS and the homology *Pf*GS ranges from 0.1 to 4.0 Å. The largest RMSD is found on the two residues of the adjacent chain, completing the active pocket of 3.5 and 4.0 Å. The marginally larger *Pf*GS macromolecule as indicated (Figure 4.13), probably results in the higher RMSD.

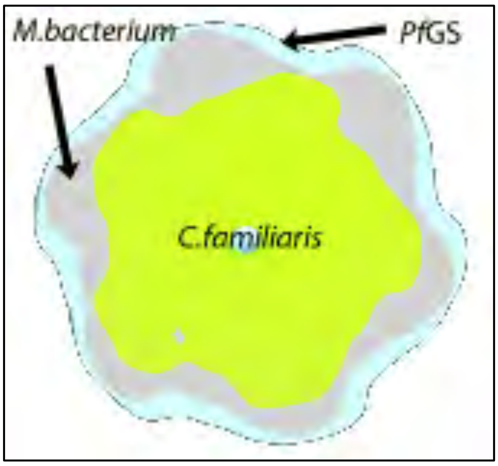


Figure 4.17 Relative size of the GS macromolecules, *Pf*GS reconstructed, M*ycobacterium tuberculosis*, PDB:2BVC and *Canis familiaris* PDB:2UU7. *Canis familiaris* has a D5 symmetry whilst *Mycobacterium tuberculosis* and *Pf*GS have a D6 symmetry.

for these two active site residues. This result could be negated if the extent of flattening of the specimen was significant due to the staining dye. Confirming the higher RMSD indicated by the homology model, with either Cryo-EM or X-Ray crystallography would provide information for targeting the glutamate analogues for *Pf*GS as a novel drug target, as opposed to most of the screening for new drug targets for *Mycobacterium tuberculosis* GS which is based on ATP analogues where the RMSD's of the active sites are the smallest when compared to the human GS.

Chapter 5 – Conclusions

5.1 Summary

The completion of the *Plasmodium falciparum* genome enabled synthesizing of glutamine synthetase. The expression of *Pf*GS was achieved by cloning the synthesized gene using a commercially available expression plasmid and transformed into an *Escherichia coli* strain. Purified *Pf*GS was eluted using chromatography and making use of competitive binding. *Pf*GS was stabilized using Mn^{++} cations before concentrating the macromolecule to 2 mg/ml.

Both side and top views of the *Pf*GS macromolecule on the negative stain electron microscope grids were achieved at 100 mM salt concentrations with the protein at 50 µg/ml concentration. Single particle reconstruction required the use of tight masking in the alignment step, to minimise the impact of adjacent particles that were in close proximity. Current approaches in bioinformatics provided the basis for recognizing the conservation of the active sites and a high confidence for the generation of an homology model.

5.2 Study Limitations

*Pf*GS crystal trials were conducted with Hampton standard screens HR2-110 & HR2-112 and samples were also sent to ICGEB, Delhi, India for use on their nano-litre pipetting robot (Mosquito) , enabling hi-speed screening. No *Pf*GS crystals resulted from over 3000 conditions that were tested.

The large loops on the negative stain 3-D reconstruction was not fully located and subsequently precluded flexible docking of the homology model. The 3-D reconstruction was computed with a limited number of particles (576) and with side views that were adjoined. This required tight masking for particle alignment, which may have introduced some artifacts at high resolution. In addition large loops are not reliably modeled by the tools available and have also been independently validated accordingly in the vicinity of the loops. Both side chain orientations and loop structures would be best handled with high resolution Cryo-EM 3D reconstruction with a large number of particles used in the reconstruction

Negative staining is known for flattening of the biological specimen due to the deposition of the stain onto the specimen. The impact of this may have resulted in the size of the molecule being overestimated and therefore having an impact of the RMSD’s that were calculated in chapter 4.

These limitations can be avoided by obtaining sub-atomic resolution of 3-D reconstruction of *Pf*GS.

5.3 Technical Achievements

Soluble *Pf*GS was expressed using a cold shock vector (pCold I (Takara) in *Escherichia coli* (strain BL21 (DE3)). Ni-affinity chromatography provided purified *Pf*GS in a single step due to the binding affinity of the attached histidine tag to the *Pf*GS. Screening of the fractions of the purified *Pf*GS indicated the presence of active GS.

Purified *Pf*GS was not stable at room temperature and was stabilized with the addition of manganese cations and forward reaction ligands, namely ATP and glutamate. A concentration of 8 mg/ml of *Pf*GS was achieved with the addition of 0.4 mM beta-octyl glucoside enabling crystal trial screening.

A 3-D reconstruction of *Pf*GS with a resolution of 16 Å was enabled using tight masking for particle alignment.

5.4 Scientific Achievements

The quartnernary structure of *Pf*GS was unequivocally demonstrated.. Twelve mononmers are arranged in two six-membered rings that are stacked back-to-back, resulting in a comple with D6 symmetry.

*Pf*GS has significant homology with GS I structures with the exception of two large insertions. One insertion is located in the central core of the rings – no unoccupied density was found for this insertion suggesting that it is disordered. The other insertion is externally located in a region of the map that has unoccupied density. The volume of the unoccupied density is insufficient to accommodate the full insertion indicating that part of it is disordered. The otherwise significant conservation indicates the presence of a common core folding and the conservation of the active sites. Both the multiple sequence alignment and structural alignment indicate that *Pf*GS is unlikely to have a conserved adenylylation loop.

With both the active sites in *Pf*GS found to be conserved structurally, the targeting of these active sites with their respective analogues is therefore not precluded.

5.5 Future Direction

The study has determined the conditions for the expression and purification of soluble active *Pf*GS. The protein prepared is suitable for structure determination by 3D-EM. Further structural function studies should employ high resolution cryo-EM using direct electron detectors and/or X-ray crystallography. In my opinion X-ray crystallography may not be the technique of choice for high resolution structure determination of *Pf*GS because of the presence of the apparently disordered, externally located loop. If further detail were obtained by cryo-EM studies a crystallizable clone might be prepared by excision of the loop.

Appendix A – Growth Media, Stock Solutions and Buffers

Water used was purified using the Milli-RP Plus (Millipore) system, abbreviated as dH_2O.

1. Growth Media – All media and solutions were autoclaved at 121°C for 20 minutes.

1.1.1 Luria broth (LB)

Tryptone	10	g
Yeast extract	5	g
NaCl	5	g
dH_2O to	1	L

1.1.2 Luria agar (LA)

Tryptone	10	g
Yeast extract	5	g
NaCl	5	g
Agar	15	g
dH_2O to	1	L

2. Stock solutions

2.1 Antibiotic solutions – Dissolve and filter sterilize mixture. Store in 1 ml eppendorfs at -20°C until required.

2.1.1 Ampicillin – 100 mg/ml

ampicillin	1	g
dH_2O to	10	ml

2.1.2 Kanamycin – 25 mg/ml

kanamycin	0.5	g
dH_2O to	10	ml

2.2 General Buffers

2.2.1 Na_2HPO_4 (20 mM)

Na_2HPO_4(anhydrous)	5.68	g
dH_2O to	2	L

2.2.2 Na_2HPO_4 (20 mM)

$NaH_2PO_4 \cdot H_2O$	5.52	g
dH_2O to	1	L

2.2.3 **Sodium Phosphate Buffer** 20 mM pH 7.5 - Check pH of mixture and adjust by addition of NaH_2PO_4 (20 mM) to increase pH and NaH_2PO_4 (20 mM) to decrease pH.

NaH_2PO_4 (20 mM)	810	ml
$NaH_2PO_4 \cdot H_2O$ (20 mM)	190	ml

2.2.4 **Cell re-suspension buffer** 50 mM NaCl

NaCl	0.29	G
Sodium Phosphate Buffer to	100	ml

2.2.5 **Protein buffer** 50 mM NaCl; 5 mM Mn^{++}

NaCl	0.29	g
$MnCl_2$	0.063	g
Sodium Phosphate Buffer to	100	ml

2.2.6 **Optional Protein buffer** 50 mM NaCl; 5 mM Mn^{++}; 5mM ATP; 5mM Glu

NaCl	0.29	g
$MnCl_2$	0.063	g
ATP	0.25	g
Glu	0.07	g
Sodium Phosphate Buffer to	100	ml

2.2.7 **EDTA** 0.5M - Dissolve both EDTA and NaCl in 400 ml OF dH_2O, adjust pH to 8.0 with HCl and adjust to final volume of 500 ml.

EDTA	93.05	g
NaCl	10	g
dH_2O to	500	ml

2.2.8 **Tris-HCl** 1M - Dissolve Tris in 80 ml dH_2O and adjust pH using HCl to required level and make up to final volume of 100 ml.

Tris	12.1	g
dH_2O to	100	ml

2.2.9 **TE Buffer**

Tris-HCl 1M pH 7.6	1	ml
EDTA 0.5M pH 8	200	µl
dH_2O to	2	ml

3. Agarose Gel buffers and Solutions

3.1.1 **TAE (Tris-Acetate) x 50**

Tris	242	g
EDTA	37.2	g
Glacial Acetic Acid	57.1	ml
dH_2O to	1	L

3.1.2 **EtBr (Ethidium Bromide) solution** - Work must be carried out in fume hood and stored in dark bottle.

EtBr	0.1	g
dH_2O to	10	ml

4. SDS-PAGE Buffers and Solutions

4.1.1 **SDS (Sodium Dodecycl Sulphate)** 10% - Dissolve SDS in 80 ml on hot plate without boiling, Make up to final volume of 100 ml.

SDS	10	g
dH_2O to	100	ml

4.1.2 **Resolving Gel Buffer -** Dissolve Tris in 80 ml dH20 and adjust to pH 8.8 with HCl. Make up to final volume of 100 ml.

Tris	18.17	g
dH_2O to	100	ml

4.1.3 **Resolving Gel Buffer** - Dissolve Tris in 40ml dH20 and adjust to pH 6.8 with HCl. Make up to final volume of 50 ml.

Tris	3.03	g
dH_2O to	50	ml

4.1.4 **Running Buffer**

Tris	3	g
Glycine	15	g
SDS	1	g
dH_2O to	1	L

4.1.5 **Sample application Buffer x5** – Adjust pH to 6.8 with HCl

Tris	3.8	g
SDS	3	g
Glycerol	10	ml
B-mercaptoethanol	5	ml
Bromophenol Blue	5	mg
dH_2O to	20	ml

4.1.6 **Coomassie blue staining solution**

Methanol	250	ml
Acetic acid	50	ml
Coomassie Blue R250	0.5	g
dH_2O to	500	ml

4.1.7 Destain solution

Ethanol	250	ml
Acetic acid	100	ml
dH_2O to	1	L

4.1.8 Resolving Gel 10% - Temed to be added just prior to loading gel slide

Resolving gel buffer	2.5	ml
Acrylamide mix (Promega)	3.33	ml
SDS 10%	100	µl
APS 10%	100	µl
Temed	10	µl
dH_2O to	10	ml

4.1.9 Stacking Gel – Temed to be added just prior to loading gel slide

Stacking gel buffer	1.25	ml
Acrylamide mix (Promega)	0.625	ml
SDS 10%	50	µl
APS 10%	50	µl
Temed	5	µl
dH_2O to	5	ml

5. Ni-Afffinity Elution Buffer - Elution buffers are vacuum filtered to 0.22 micron

5.1.1 Elution buffer A

NaCl	29.2	g
Protein buffer to	1	L

5.2.2 Elution buffer B (0.5 mM Imidazole)

Imidazole	34.04	g
NaCl	29.2	L
Protein buffer to	1	L

6. γ-glutamyl transferase assay (GGT)

6.1 1 Reaction Mix – Prepare stock solution of each reagent and make reaction mix for use within 12 hours and keep at 4°C.

Imidazole-HCl 1 M pH7.15	329.4	µl
Hydroxylamine hydrochloride 0.8 M	54.2	µl
$MgCl_2$ 0.1 M	6.6	µl
Potassium arsenate 0.28M pH 7.15	291.5	µl
ADP 40 mM	22.0	µl
dH2O	1321.3	µl

6.2.1 Glutamate 0.2 mM - Store in 1 ml eppendorfs and store at -18°C.

Glutamate	0.147	g
dH_2O	5	ml

6.3.1 Stop Reaction solution

$FeCl_3{\cdot}6H_2O$	11.0	g
Trichloroacetic acid	4.0	g
HCl concentrated	4.2	ml
dH2O	200	ml

Appendix B – Codon Optimised *Pf*GS Sequence

Codon optimised DNA sequence for *Pf*GS expression in *Escherichia coli*.

Optimised		ATGAAGAGCGTTAGCTTTAGCAATAATGCCGAGCTGTATGAATATATCAAAGACAAA
Native	1	ATGAAGTCCGTGAGTTTTTCAAATAATGCTGAGCTGTACGAGTACATAAAGGACAAA
Translation		M K S V S F S N N A E L Y E Y I K D K

Optimised		AAAAATGATGTGGAAATTGTGGCCTGCATTATTACCAATCTGCTGGGCACCTATTTC
Native	58	AAGAATGATGTAGAAATAGTAGCTTGTATTATTACGAATTTATTAGGAACCTATTTT
Translation		K N D V E I V A C I I T N L L G T Y F

Optimised		AAATGCTTTTTTTATGTGAAAGAAATTACCCTGAATAAACTGGAAAGCGGCTTTAGC
Native	115	AAGTGTTTTTTTTATGTAAAGGAGATTACATTAAATAAGTTAGAAAGTGGTTTTTCA
Translation		K C F F Y V K E I T L N K L E S G F S

Optimised		TTTGATGCAAGCAGCATCAAACTGTGTAGCGATACCGAAGTGAGCGATTTTTTTATT
Native	172	TTTGACGCGTCTTCTATTAAACTATGTTCAGATACCGAAGTAAGTGATTTTTTTATT
Translation		F D A S S I K L C S D T E V S D F F I

Optimised		AAAGTGGATCATAGCACGTGTTATCTGGAAGAATGTGATGGCAAAAATATTCTGAAT
Native	229	AAAGTAGATCATTCAACATGTTATTTGGAGGAATGTGATGGAAAGAATATATTAAAT
Translation		K V D H S T C Y L E E C D G K N I L N

Optimised		ATTATGTGCGATATTAAACGCTATAATGGCTTTGATTATTATAAATGCCCGCGTACC
Native	286	ATCATGTGTGATATAAAGAGATATAATGGTTTTGATTATTATAAATGTCCAAGAACA
Translation		I M C D I K R Y N G F D Y Y K C P R T

Optimised		ATTCTGAAAAAAACCTGTGAATTTGTGAAAAATGAAGGCATTGCCGATAAAGTGTGC
Native	343	ATATTAAAGAAAACATGTGAATTTGTTAAGAATGAAGGTATAGCAGATAAAGTTTGT
Translation		I L K K T C E F V K N E G I A D K V C

Optimised		ATTGGCAATGAACTGGAATTTTTTATCTTTGATAAAGTGAATTATTCCCTGGATGAA
Native	400	ATTGGAAATGAATTGGAATTTTTTATATTTGATAAAGTAAATTATAGTTTAGATGAA
Translation		I G N E L E F F I F D K V N Y S L D E

Optimised		TATAATACCTATCTGAAAGTGTATGATCGCGAAAGCTTTAGCTGCAAAAATGATCTG
Native	457	TATAATACATATTTAAAGGTTTATGATAGAGAATCATTTTCTTGTAAAAATGATTTA
Translation		Y N T Y L K V Y D R E S F S C K N D L

Optimised		AGCAGCATTTATGGTAATCATGTGGTGAATAAAGTGGAACCGCATAAAGATCATTTT
Native	514	TCGAGTATTTATGGTAATCATGTTGTAAATAAAGTTGAACCACATAAAGATCATTTT
Translation		S S I Y G N H V V N K V E P H K D H F

Optimised		AATAATCCGAATAACGAATATCTGATTAATGATGACAGCAAAAAAGTGAAAAAAAAA
Native	571	AATAATCCTAATAATGAATATTTAATTAATGATGATAGTAAAAAAGTAAAGAAAAAA
Translation		N N P N N E Y L I N D D S K K V K K K

Optimised		AGCGGCTATTTTACCACCGATCCGTATGATACCAGCAATATTATTAAACTGCGCATT
Native	628	TCCGGTTATTTTACTACAGATCCATATGATACTTCTAATATAATTAAACTTAGAATA
Translation		S G Y F T T D P Y D T S N I I K L R I

Optimised		TGTCGTGCACTGAATGATATGAATATTAATGTGCAGCGCTATCATCACGAAGTTAGC
Native	685	TGTAGAGCATTAAATGATATGAATATTAATGTACAAAGATATCATCATGAAGTTTCA
Translation		C R A L N D M N I N V Q R Y H H E V S

Optimised		ACCAGCCAGCATGAAATTAGCCTGAAATATTTTGATGCACTGACCAATGCAGATTTT
Native	742	ACAAGTCAACATGAAATTTCTTTAAAATATTTTGATGCTCTAACAAATGCTGATTTT
Translation		T S Q H E I S L K Y F D A L T N A D F

Optimised		CTGCTGATTACCAAACAAATTATTAAAACCACCGTGAGCAGCTTTAATCGTACCGCA
Native	799	TTACTTATTACAAAACAAATTATTAAAACAACAGTTAGCTCATTCAACAGAACAGCT
Translation		L L I T K Q I I K T T V S S F N R T A

Optimised		ACCTTTATGCCGAAACCGCTGGTTAATGATAATGGTAATGGCCTGCATTGTAATATT
Native	856	ACTTTTATGCCTAAACCTTTAGTTAATGATAATGGAAATGGTTTGCATTGTAATATA
Translation		T F M P K P L V N D N G N G L H C N I

Optimised		AGCCTGTGGAAAAATAACAAAAATATCTTTTATCATAATGATCCGAGCACCTTTTTT
Native	913	TCTTTATGGAAAAATAATAAAAATATATTCTATCATAATGATCCTTCTACTTTCTTT
Translation		S L W K N N K N I F Y H N D P S T F F

Optimised		CTGAGCAAAGAAAGCTTTTATTTTATGTATGGCATTGTGAAACATGCAAAAGCACTG
Native	970	TTATCAAAAGAATCTTTTTATTTTATGTATGGTATAGTTAAACATGCAAAAGCTTTA
Translation		L S K E S F Y F M Y G I V K H A K A L

Optimised		CAGGCATTTTGTAATGCAACCATGAATAGCTATAAACGTCTGGTTCCGGGTTTTGAA
Native	1027	CAAGCCTTTTGTAACGCAACCATGAATTCATATAAAAGATTAGTACCAGGTTTTGAA
Translation		Q A F C N A T M N S Y K R L V P G F E

Optimised		ACCTGTCAGAAACTGTTTTATAGCTTTGGTAGCCGTAGCGCAGTTATTCGTCTGAGC
Native	1084	ACTTGTCAAAAATTATTTTATTCATTTGGTTCAAGAAGTGCTGTTATCAGGTTATCT
Translation		T C Q K L F Y S F G S R S A V I R L S

Optimised		CTGATTAATTATAGCAATCCGAGCGAAAAACGTATTGAATTTCGTCTGCCGGATTGT
Native	1141	TTGATTAACTACAGTAATCCATCAGAAAAAAGAATTGAATTTAGATTACCTGATTGT
Translation		L I N Y S N P S E K R I E F R L P D C

Optimised		GCAAATTCTCCGCATCTGGTTATGGCAGCAATTATTCTGGCAGGCTATGATGGCATT
Native	1198	GCTAACTCACCACATCTAGTTATGGCTGCTATCATTTTGGCAGGTTATGATGGTATT
Translation		A N S P H L V M A A I I L A G Y D G I

Optimised		AAATCTAAAGAACAGCCGCTGGTTCCGTTTGAAAGCAAAGATAATCATTTTTATATC
Native	1255	AAATCAAAAGAACAACCGCTGGTTCCATTTGAAAGTAAAGATAATCATTTCTACATT
Translation		K S K E Q P L V P F E S K D N H F Y I

Optimised		AGCAGCATCTTTTCTAAATATGTGCAGCATCCGGAAAATTTTAATATTCTGACCCAT
Native	1312	TCAAGCATATTTTCGAAATATGTACAACACCCAGAAAATTTTAATATTCTTACTCAT
Translation		S S I F S K Y V Q H P E N F N I L T H

Optimised		GCCCTGGAAGGTTTTTCTAAATATGTGCAGCATCCGGAAAATTTTAATATTCTGACC
Native	1369	GCCTTGGAAGGTTTTTCGAAATATGTACAACACCCAGAAAATTTTAATATTCTTACT
Translation		A L E G Y E S L H T I N E S P E F K N

Optimised		TTTTTCAAATGCGAAGAACCGCAGGGTATTAGCTTTAGCCTGGTTGAAAGCCTGGAT
Native	1426	TTTTTTAAATGTGAAGAACCACAAGGTATATCTTTTTCTCTTGTCGAAAGTTTAGAT
Translation		F F K C E E P Q G I S F S L V E S L D

Optimised		GCACTGGAAAAAGATCATGCCTTTCTGACCGTGAATAATATTTTTACCGAAGAAATG
Native	1483	GCTTTGGAAAAGGATCACGCATTTTTAACTGTAAATAATATTTTCACGGAGGAAATG
Translation		A L E K D H A F L T V N N I F T E E M

Optimised		ATTCAAGAATACATTAAATTTAAACGCGAAGAAATTGATGCCTATAATAAATACGTG
Native	1540	ATACAAGAATATATAAAGTTCAAAAGGGAAGAAATTGACGCATATAATAAATATGTT
Translation		I Q E Y I K F K R E E I D A Y N K Y V

Optimised		AATGCCTATGACTATCATCTGTACTACGAATGCTAA
Native	1597	AACGCTTATGATTATCACTTATATTATGAATGTTAG (1632 bp)
Translation		N A Y D Y H L Y Y E C *

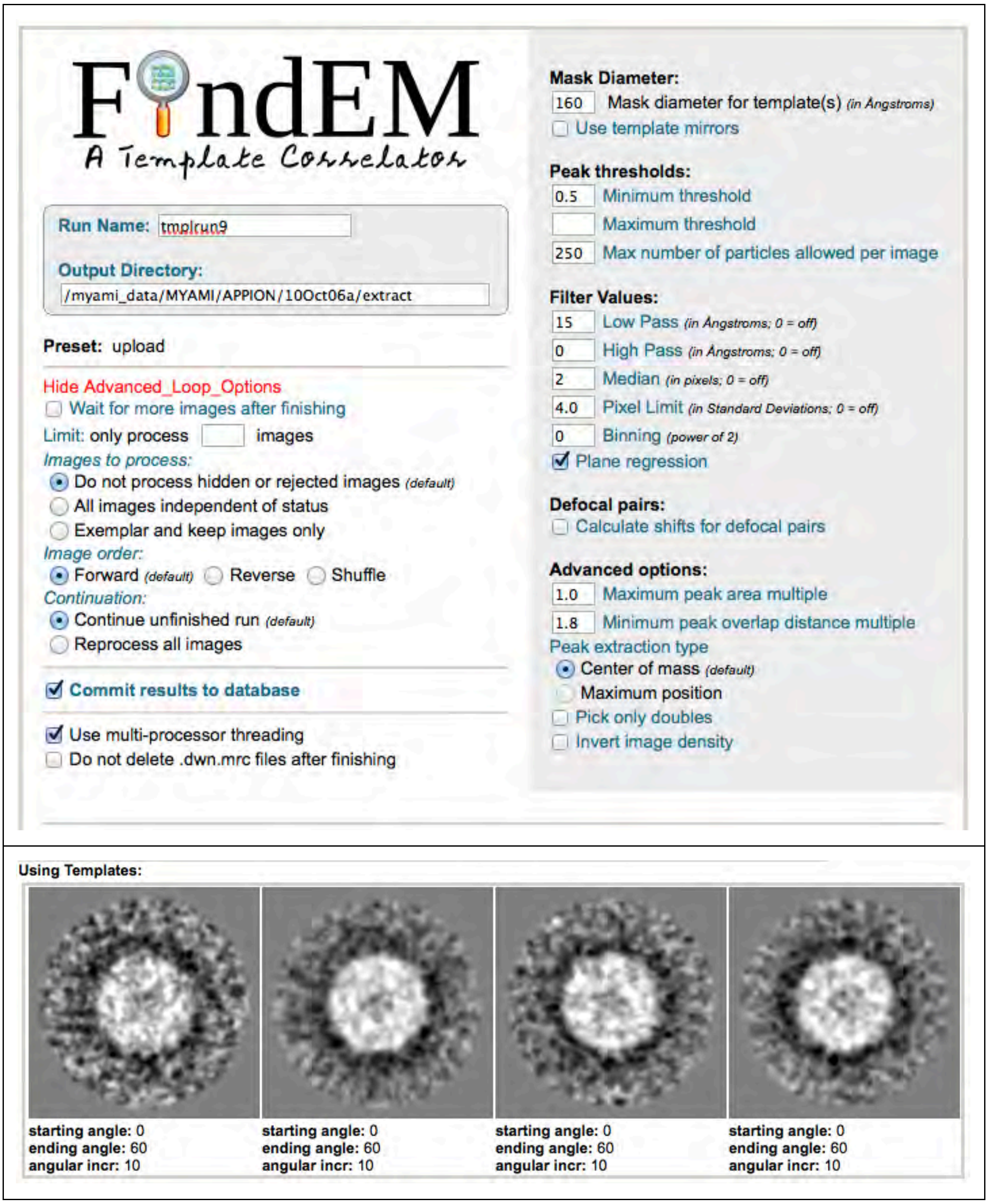
The DNA *Pf*GS sequence translates to 543 amino acid as indicated below.

MKSVSFSNNAELYEYIKDKKNDVEIVACIITNLLGTYFKCFFYVKEITLNKLESGFSFDASSIKLCSDTEVSDFFIK
VDHSTCYLEECDGKNILNIMCDIKRYNGFDYYKCPRTILKKTCEFVKNEGIADKVCIGNELEFFIFDKVNYSLDEY
NTYLKVYDRESFSCKNDLSSIYGNHVVNKVEPHKDHFNNPNNEYLINDDSKKVKKKSGYFTTDPYDTSNIIKLRI
CRALNDMNINVQRYHHEVSTSQHEISLKYFDALTNADFLLITKQIIKTTVSSFNRTATFMPKPLVNDNGNGLHC
NISLWKNNKNIFYHNDPSTFFLSKESFYFMYGIVKHAKALQAFCNATMNSYKRLVPGFETCQKLFYSFGSRSAVI
RLSLINYSNPSEKRIEFRLPDCANSPHLVMAAIILAGYDGIKSKEQPLVPFESKDNHFYISSIFSKYVQHPENFNIL
THALEGYESLHTINESPEFKNFFKCEEPQGISFSLVESLDALEKDHAFLTVNNIFTEEMIQEYIKFKREEIDAYNK
YVNAYDYHLYYEC* (543aa)

Appendix C - Appion Graphical User Interface Screens

The following Appion GUI screens were used for computational processing of the data.

1. Template Particle Picking (First Negative Stain Data Set)



Run Name: tmplrun9

Output Directory: /myami_data/MYAMI/APPION/10Oct06a/extract

Preset: upload

Hide Advanced_Loop_Options

☐ Wait for more images after finishing

Limit: only process ☐ images

Images to process:

● Do not process hidden or rejected images *(default)*

○ All images independent of status

○ Exemplar and keep images only

Image order:

● Forward *(default)* ○ Reverse ○ Shuffle

Continuation:

● Continue unfinished run *(default)*

○ Reprocess all images

☑ **Commit results to database**

☑ Use multi-processor threading

☐ Do not delete .dwn.mrc files after finishing

Mask Diameter:

160 Mask diameter for template(s) *(in Angstroms)*

☐ Use template mirrors

Peak thresholds:

0.5 Minimum threshold

Maximum threshold

250 Max number of particles allowed per image

Filter Values:

15 Low Pass *(in Angstroms; 0 = off)*

0 High Pass *(in Angstroms; 0 = off)*

2 Median *(in pixels; 0 = off)*

4.0 Pixel Limit *(in Standard Deviations; 0 = off)*

0 Binning *(power of 2)*

☑ Plane regression

Defocal pairs:

☐ Calculate shifts for defocal pairs

Advanced options:

1.0 Maximum peak area multiple

1.8 Minimum peak overlap distance multiple

Peak extraction type

● Center of mass *(default)*

○ Maximum position

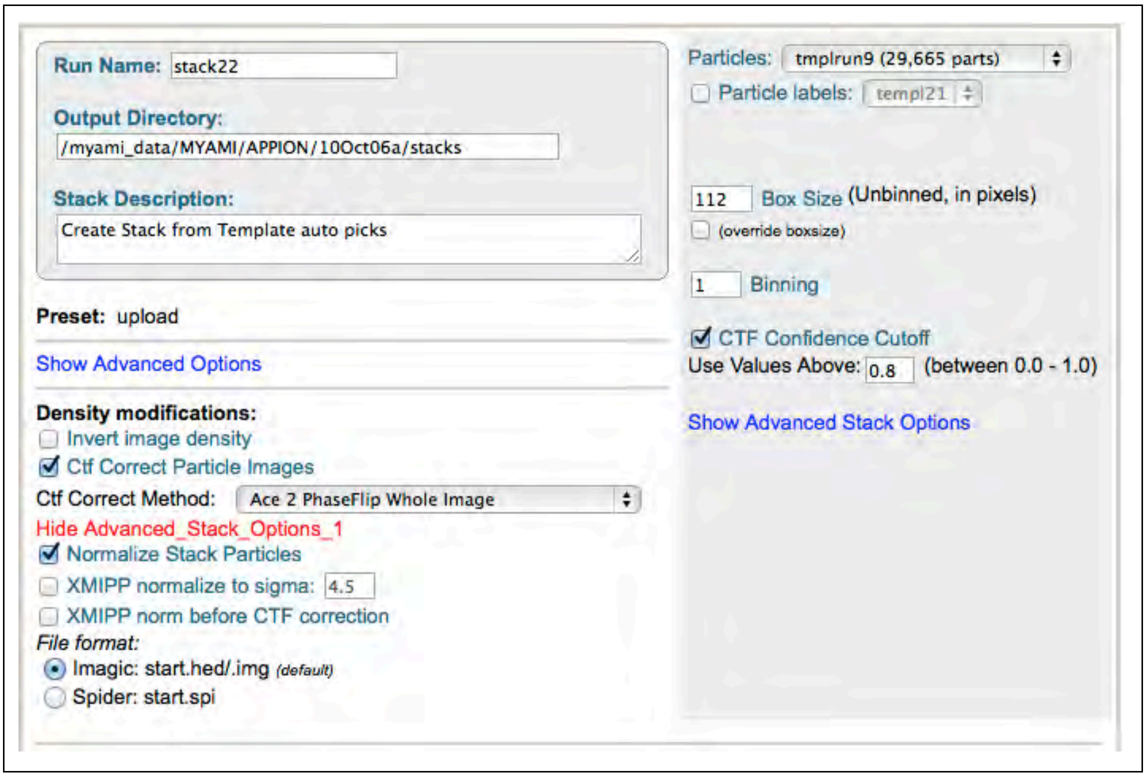
☐ Pick only doubles

☐ Invert image density

Using Templates:

starting angle: 0 **ending angle:** 60 **angular incr:** 10	**starting angle:** 0 **ending angle:** 60 **angular incr:** 10	**starting angle:** 0 **ending angle:** 60 **angular incr:** 10	**starting angle:** 0 **ending angle:** 60 **angular incr:** 10

2. Stack Creation (First Negative Stain Data Set)



Run Name: stack22

Output Directory: /myami_data/MYAMI/APPION/10Oct06a/stacks

Stack Description: Create Stack from Template auto picks

Preset: upload

Show Advanced Options

Density modifications:
☐ Invert image density
☑ Ctf Correct Particle Images
Ctf Correct Method: Ace 2 PhaseFlip Whole Image

Hide Advanced_Stack_Options_1
☑ Normalize Stack Particles
☐ XMIPP normalize to sigma: 4.5
☐ XMIPP norm before CTF correction

File format:
● Imagic: start.hed/.img *(default)*
○ Spider: start.spi

Particles: tmplrun9 (29,665 parts)
☐ Particle labels: templ21

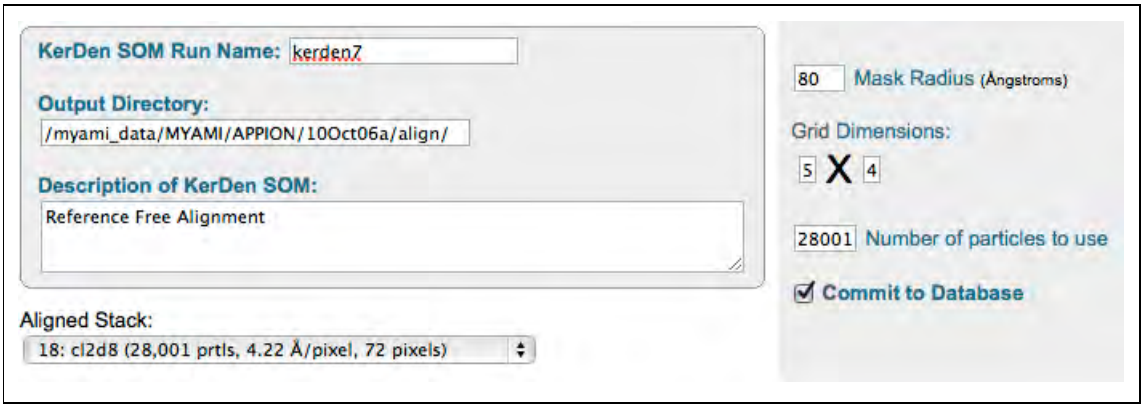
112 Box Size (Unbinned, in pixels)
☐ (override boxsize)

1 Binning

☑ CTF Confidence Cutoff
Use Values Above: 0.8 (between 0.0 - 1.0)

Show Advanced Stack Options

3. Reference Free Alignment (First Negative Stain Data Set)



KerDen SOM Run Name: kerden7

Output Directory: /myami_data/MYAMI/APPION/10Oct06a/align/

Description of KerDen SOM: Reference Free Alignment

Aligned Stack: 18: cl2d8 (28,001 prtls, 4.22 Å/pixel, 72 pixels)

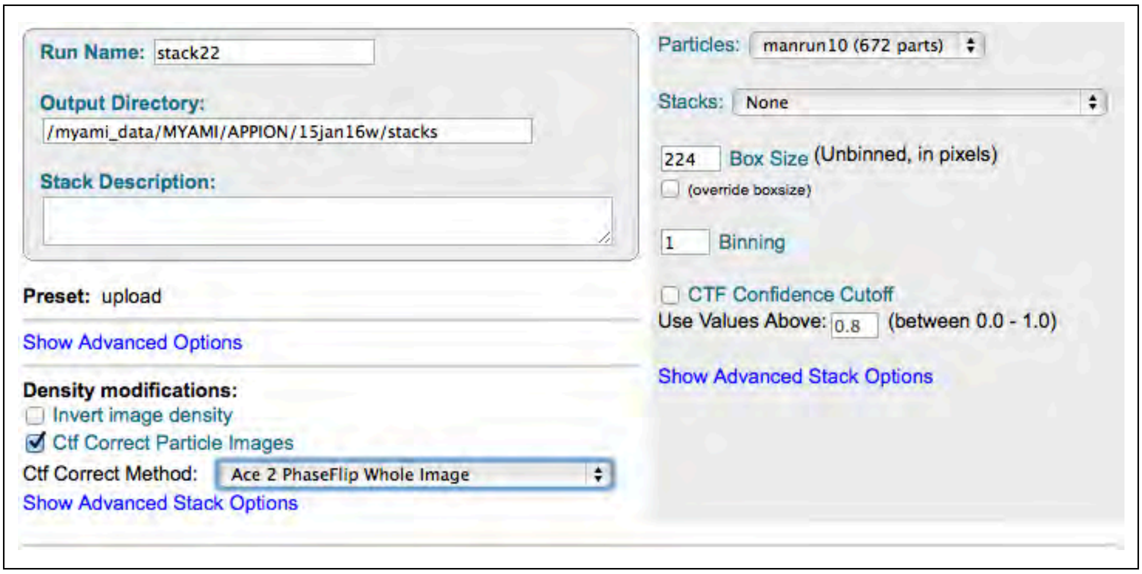
80 Mask Radius (Angstroms)

Grid Dimensions:
5 X 4

28001 Number of particles to use

☑ **Commit to Database**

4. Particle Stack Creation (Second Negative Stain Data Set)



Run Name: stack22

Output Directory: /myami_data/MYAMI/APPION/15jan16w/stacks

Stack Description:

Preset: upload

Show Advanced Options

Density modifications:

☐ Invert image density

☑ Ctf Correct Particle Images

Ctf Correct Method: Ace 2 PhaseFlip Whole Image

Show Advanced Stack Options

Particles: manrun10 (672 parts)

Stacks: None

224 Box Size (Unbinned, in pixels)

☐ (override boxsize)

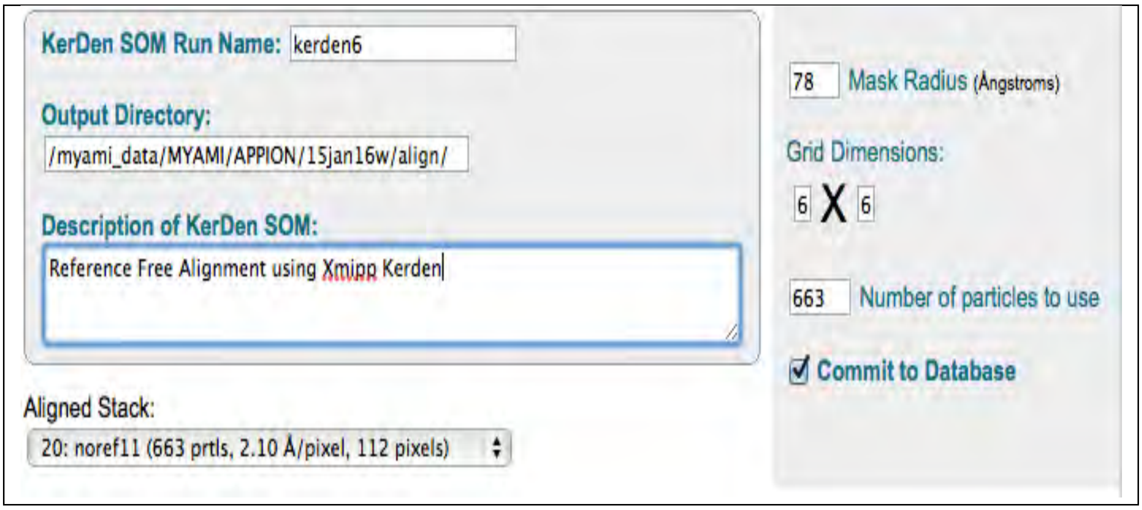
1 Binning

☐ CTF Confidence Cutoff

Use Values Above: 0.8 (between 0.0 - 1.0)

Show Advanced Stack Options

5. Reference Free Alignment (Second Negative Stain Data Set)



KerDen SOM Run Name: kerden6

Output Directory: /myami_data/MYAMI/APPION/15jan16w/align/

Description of KerDen SOM: Reference Free Alignment using Xmipp Kerden

Aligned Stack: 20: noref11 (663 prtls, 2.10 Å/pixel, 112 pixels)

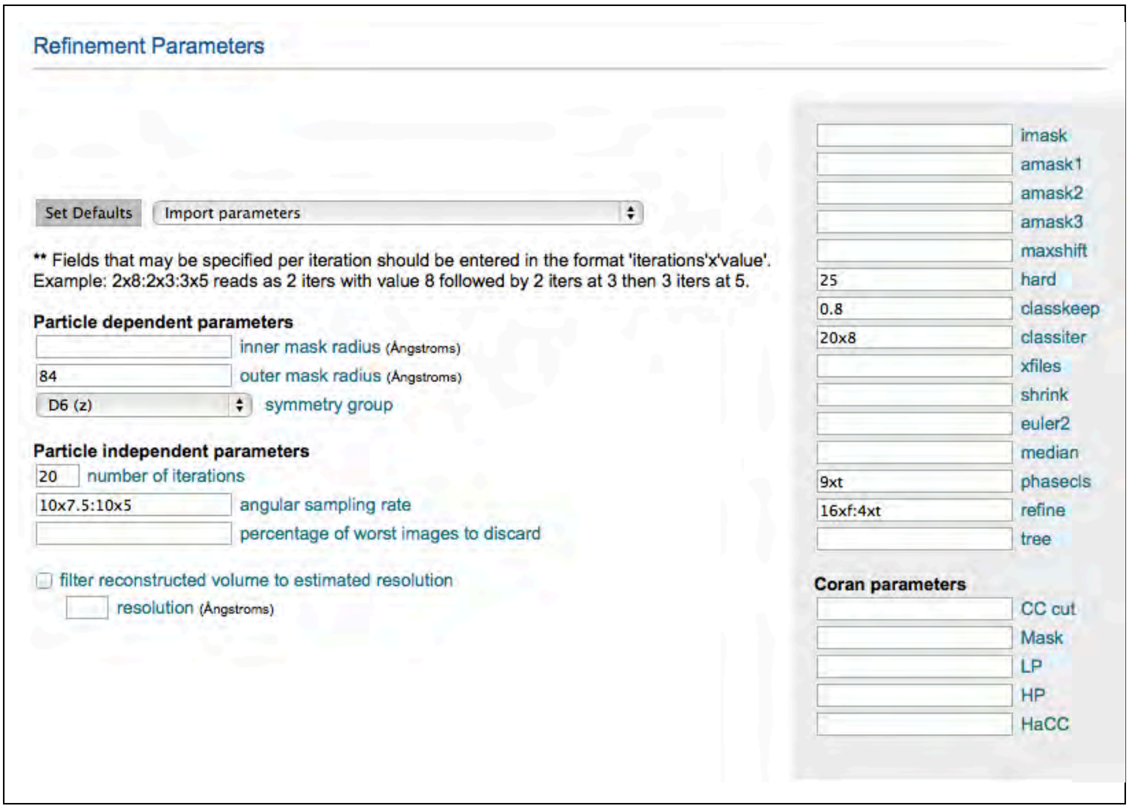
78 Mask Radius (Angstroms)

Grid Dimensions: 6 X 6

663 Number of particles to use

☑ Commit to Database

6. Model Refinement (Second Negative Stain Data Set)



Refinement Parameters

** Fields that may be specified per iteration should be entered in the format 'iterations'x'value'.
Example: 2x8:2x3:3x5 reads as 2 iters with value 8 followed by 2 iters at 3 then 3 iters at 5.

Particle dependent parameters

	inner mask radius (Ångstroms)
84	outer mask radius (Ångstroms)
D6 (z)	symmetry group

Particle independent parameters

20	number of iterations
10x7.5:10x5	angular sampling rate
	percentage of worst images to discard

☐ filter reconstructed volume to estimated resolution

	resolution (Ångstroms)

	imask
	amask1
	amask2
	amask3
	maxshift
25	hard
0.8	classkeep
20x8	classiter
	xfiles
	shrink
	euler2
	median
9xt	phasecls
16xf:4xt	refine
	tree

Coran parameters

	CC cut
	Mask
	LP
	HP
	HaCC

Appendix D – Glutamine Synthetase Sequences

The following sequences were used in multiple alignments to *Pf*GS sequences. Conserved regions are highlighted.

<table>
<tr><th>Organisms</th><th>PDB depositions and associated sequences</th></tr>
<tr><td><i>Salmonella typhimurium</i></td><td>Depositions: 1989 -2GLS; 1994 -2LGS, 1LGR; 2000 -1F1H, 1F52, 1FPY</td></tr>
<tr><td colspan="2">SAEHVLTMLNEHEVKFVDLRFTDTKGKEQHVTIPAHQVNAEFFEEGKMFDGSSIGGWKGINESDMVLMPDASTAVIDPFFADSTLIIRCDILEPGTLQGYDRDPRSIAKRAEDYLRATGIADTVLFGPEPEFFLFDDIRFGASISGSHVAIDDIEGAWNSSTKYEGGNKGHRPGVKGGYFPVPPVDSAQDIRSEMCLVMEQMGLVVEAHHHEVATAGQNEVATRFNTMTKKADEIQIYKYVVHNVAHRFGKTATFMPKPMFGDNGSGMHCHMSLAKNGTNLFSGDKYAGLSEQALYYIGGVIKHAKAINALANPTTNSYKRLVPGYEAPVMLAYSARNRSASIRIPVVASPKARRIEVRFPDPAANPYLCFAALLMAGLDGIKNKIHPGEPMDKNLYDLPPEEAKEIPQVAGSLEEALNALDLDREFLKAGGVFTDEAIDAYIALRREEDDRVRMTPHPVEFELYYSV (468 Amino Acids)</td></tr>
<tr><td><i>Mycobacterium tuberculosis</i></td><td>Depositions: 2001 -1HTO, 1HTQ; 2005 -2BVC; 2009 -2WGS, 2WHI; 2011 -3ZXR, 3ZXV</td></tr>
<tr><td colspan="2">TEKTPDDVFKLAKDEKVEYVDVRFCDLPGIMQHFTIPASAFDKSVFDDGLAFDGSSIRGFQSIHESDMLLLPDPETARIDPFRAAKTLNINFFVHDPFTLEPYSRDPRNIARKAENYLISTGIADTAYFGAEAEFYIFDSVSFDSRANGSFYEVDAISGWWNTGAATEADGSPNRGYKVRHKGGYFPVAPNDQYVDLRDKMLTNLINSGFILEKGHHEVGSGGQAEINYQFNSLLHAADDMQLYKYIIKNTAWQNGKTVTFMPKPLFGDNGSGMHCHQSLWKDGAPLMYDETGYAGLSDTARHYIGGLLHHAPSLLAFTNPTVNSYKRLVPGYEAPINLVYSQRNRSACVRIPITGSNPKAKRLEFRSPDSSGNPYLAFSAMLMAGLDGIKNKIEPQAPVDKDLYELPPEEAASIPQTPTQLSDVIDRLEADHEYLTEGGVFTNDLIETWISFKRENEIEPVNIRPHPYEFALYYDV (477Amino Acids)</td></tr>
<tr><td colspan="2"><i>Note: Mycobacterium tuberculosis</i> depositions since 2005 includes an histidine tag ; MAHHHHHHG……. followed by above sequence (486 Amino Acids)</td></tr>
<tr><td><i>Zea mays</i></td><td>Depositions: 2005 -2D3A, 2D3B, 2D3C</td></tr>
<tr><td colspan="2">MACLTDLVNLNLSDTTEKIIAEYIWIGGSGMDLRSKARTLPGPVTDPSKLPKWNYDGSSTGQAPGEDSEVILYPQAIFKDPFRRGNNILVMCDCYTPAGEPIPTNKRYSAAKIFSSPEVAAEEPWYGIEQEYTLLQKDTNWPLGWPIGGFPGPQGPYYCGIGAEKSFGRDIVDAHYKACLYAGINISGINGEVMPGQWEFQVGPSVGISSGDQVWVARYILERITEIAGVVVTFDPKPIPGDWNGAGAHTNYSTESMRKEGGYEVIKAAIEKLKLRHKEHIAAYGEGNERRLTGRHETADINTFSWGVANRGASVRVGRETEQNGKGYFEDRRPASNMDPYVVTSMIAETTIVWKP (356 Amino Acids)</td></tr>
<tr><td><i>Homo sapiens</i></td><td>Depositions: 2007 -2OJW, 2QC8</td></tr>
<tr><td colspan="2">MHHHHHHSSGVDLGTENLYFQSMASSHLNKGIKQVYMSLPQGEKVQAMYIWIDGTGEGLRCKTRTLDSEPKCVEELPEWNFDGSSTLQSEGSNSDMYLVPAAMFRDPFRKDPNKLVLCEVFKYNRRPAETNLRHTCKRIMDMVSNQHPWFGMEQEYTLMGTDGHPFGWPSNGFPGPQGPYYCGVGADRAYGRDIVEAHYRACLYAGVKIAGTNAEVMPAQWEFQIGPCEGISMGDHLWVARFILHRVCEDFGVIATFDPKPIPGNWNGAGCHTNFSTKAMREENGLKYIEEAIEKLSKRHQYHIRAYDPKGGLDNARRLTGFHETSNINDFSAGVANRSASIRIPRTVGQEKKGYFEDRRPSANCDPFSVTEALIRTCLLNETG (384 Amino Acids)</td></tr>
<tr><td><i>Canis familiaris</i></td><td>Depositions: 2007 -2UU7</td></tr>
<tr><td colspan="2">MATSASSHLNKGIKQVYMSLPQGEKVQAMYIWIDGTGEGLRCKTRTLDSEPKGVEELPEWNFDGSSTFQSEGSNSDMYLVPAAMFRDPFRKDPNKLVFCEVFKYNRKPAETNLRHTCKRIMDMVSNQHPWFGMEQEYTLMGTDGHPFGWPSNGFPGPQGPYYCGVGADKAYGRDIVEAHYRACLYAGIKIAGTNAEVMPAQWEFQIGPCEGIDMGDHLWVARFILHRVCEDFGVIATFDPKPIPGNWNGAGCHTNFSTKAMREENGLKYIEESIEKLSKRHQYHIRAYDPKGGLDNARRLTGFHETSNINDFSAGVANRGASIRIPRTVGQEKKGYFEDRRPSANCDPFSVTEALIRTCLLNETGDEPFQYKNLEHHHHHH (381 Amino Acids)</td></tr>
<tr><td><i>Saccharomyces cerevisiae</i></td><td>Depositions: 2008 -3FKY</td></tr>
<tr><td colspan="2">MAEASIEKTQILQKYLELDQRGRIIAEYVWIDGTGNLRSKGRTLKKRITSIDQLPEWNFDGSSTNQAPGHDSDIYLKPVAYYPDPFRRGDNIVVLAACYNNDGTPNKFNHRHEAAKLFAAHKDEEIWFGLEQEYTLFDMYDDVYGWPKGGYPAPQGPYYCGVGAGKVYARDMIEAHYRACLYAGLEISGINAEVMPSQWEFQVGPCTGIDMGDQLWMARYFLHRVAEEFGIKISFHPKPLKGDWNGAGCHANVSTKEMRQPGGTKYIEQAIEKLSKRHAEHIKLYGSDNDMRLTGRHETASMTAFSSGVANRGSSIRIPRSVAKEGYGYFEDRRPASNIDPYLVTGIMCETVCGAIDNADMTKEFERESS (370 Amino Acids)</td></tr>
<tr><td></td><td></td></tr>
</table>

<table>
<tr><td><i>Synechocystis sp</i></td><td>Depositions: 2010 -3NG0</td></tr>
<tr><td colspan="2">MARTPQEVLKWIQDENIKIIDLKFIDTPGIWQHCSFYYDQLDENSFTEGIPFDGSSIRGWKAINESDMCMVPDPNTATIDPFCKEPTLSMICSIKEPRTGEWYNRDPRTIAAKAAEYLRGTGIADTVYFGPEAEFFLFDDIRFGQTENSSYYFADSVEGRWNTGREEEGGNLGYKPGYKQGYFPVAPTDTAQDIRTEMLLTMAAFGVPIEKHHHEVASGGQNELGIKFDKLVNSADNLMIYKYVIKNVAKKYGKTVTFMPKPIFNDNGSGMHVHQSLWKDGQPLFAGDKYAGFSQMGLWYIGGILKHAPALLAFTNPTTNSYKRLVPGFEAPVNLAYSQGNRSASVRIPLSGGNPKAKRLEFRCPDATSNPYLAFAAMLCAGIDGIKNQIDPGEPLDVDIYDLSPEELAKIPSTPGSLEAALEALEKDHEFLTGTGVFSPDFVESWIEYKLDNEVNPMRLRPHPYEFSLYYDC (473 Amino Acids)</td></tr>
<tr><td><i>Bacteroides fragilis</i></td><td>Depositions: 2010 -3O6X</td></tr>
<tr><td colspan="2">MSKMRFFALQELSNRKPLEITTPSNKLSDYYASHVFDRKKMQEYLPKEAYKAVVDATEKGTPISREMADLIANGMKSWAKSLNVTHYTHWFQPLTDGTAEKHDGFIEFGEDGEVIERFSGKLLIQQEPDASSFPNGGIRNTFEARGYTAWDGSSPAFVVDTTLCIPTIFISYTGEALDYKTPLLKALAAVDKAATEVCQLFDKNITRVFTNLGWEQEYFLVDTSLYNARPDLRLTGRTLMGHSSAKDQQLEDHYFGSIPPRVTAFMKELEIECHKLGIPVKTRHNEVAPNQFELAPIFENCNLANDHNQLVMDLMKRIARKHHFAVLFHEKPYNGVNGSGKHNNWSLCTDTGINLFAPGKNPKGNMLFLTFLVNVLMMVHKNQDLLRASIMSAGNSHRLGANEAPPAILSIFLGSQLSATLDEIVRQVTNSKMTPEEKTTLKLGIGRIPEILLDTTDRNRTSPFAFTGNRFEFRAAGSSANCAAAMIAINAAMANQLNEFKASVDKLMEEGIGKDEAIFRILKENIIASELIRFEGDGYSEEWKQEAARRGLTNICHVPEALMHYMDNQSRAVLIGERIFNETELACRLEVELEKYTMKVQIESRVLGDLAINHIVPIAVSYQNRLLENLCRMKEIFSEEEYEVMSADRKELIKEISHRVSAIKVLVRDMTEARKVANHKENFKEKAFAYEETVRPYLESIRDHIDHLEMEIDDEIWPLPKYRELLFTK (729 Amino Acids)</td></tr>
<tr><td colspan="2"><i>Plasmodium falciparum</i> – gene of interest translated sequence from PlasmoDB genome website</td></tr>
<tr><td colspan="2">MKSVSFSNNAELYEYIKDKKNDVEIVACIITNLLGTYFKCFFYVKEITLNKLESGFSFDASSIKLCSDTEVSDFFIKVDHSTCYLEECDGKNILNIMCDIKRYNGFDYYKCPRTILKKTCEFVKNEGIADKVCIGNELEFFIFDKVNYSLDEYNTYLKVYDRESFSCKNDLSSIYGNHVVNKVEPHKDHFNNPNNEYLINDDSKKVKKKSGYFTTDPYDTSNIIKLRICRALNDMNINVQRYHHEVSTSQHEISLKYFDALTNADFLLITKQIIKTTVSSFNRTATFMPKPLVNDNGNGLHCNISLWKNNKNIFYHNDPSTFFLSKESFYFMYGIVKHAKALQAFCNATMNSYKRLVPGFETCQKLFYSFGSRSAVIRLSLINYSNPSEKRIEFRLPDCANSPHLVMAAIILAGYDGIKSKEQPLVPFESKDNHFYISSIFSKYVQHPENFNILTHALEGYESLHTINESPEFKNFFKCEEPQGISFSLVESLDALEKDHAFLTVNNIFTEEMIQEYIKFKREEIDAYNKYVNAYDYHLYYEC (543 amino Acids)</td></tr>
</table>

Bibliography

Books

Reitzer, L.J. & Magasnik, B., 1987. Ammonia Assimilation and the Biosynthesis of Glutamine, Gluamate, Aspartate, Asparagine, L-Alanine and D-Alanine. In C. Neidhardt, Frederick *et al.*, eds. *Escherichia coli and Salmonella Cellular and Molecular Biology*. Washington D.C.: American Society for Microbiology (ASM), pp. 305–311.

Wohlheuter, R.M., Schutt, H. & Holzer, H., 1973. Regulation of Glutamine synthesis in vivo in *E.coli*. In P. Stanley & R. Stadtman, Earl, eds. *The Enzymes of Glutamine metabolism*. New York: Academic Press, pp. 45–64.

Reports / Thesis

Krajewski, W.W., 2008. *Structural Studies of Glutamine Synthetases: Towards the Development of Novel Antitubercular Agents*. Acta Universitatis Upsaliensis.

World Health Organisation, 2011. Global Plan for Artemisinin Resistance Containment (GPARC), Available at: *http://www.who.int/malaria/publications/atoz/artemisinin_resistance_containment_2011.pdf*.

Journals

Abell, L.M., Schineller, J., Keck, P.J. & Villafranca, J.J., 1995. Effect of Metal-Ligand Mutations on Phosphoryl Transfer Reactions Catalyzed by. *Biochemistry*, 34(5), pp.16695–16702.

Abratt, V.R., Zappe, H. & Woods, D.R., 1993. A reporter gene vector to investigate the regulation of glutamine synthetase in Bacteroides fragilis Bf1. *Journal of general …*, 139(1), pp.59–65.

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

Backman, K., Chen, Y. & Magasanik, Boris, 1981. Physical and genetic characterization of the ginA-ginG region of the *Escherichia coli* chromosome. *Proceedings of the National Academy of Sciences of the United States of America*, 78(6), pp.3743–3747.

Bahl, A., Brunk, B., Crabtree, J., Fraunholz, M.J., Gajria, B., Grant, G.R., Ginsburg, H., Gupta, D., Kissinger, J.C., Labo, P., Li, Li., Mailman, M.D., Milgram, A.J., Pearson, D.S., Roos, D.S., Schug, J., Jr, Christian, S.,J. & Whetzel, P., 2003. PlasmoDB: the Plasmodium genome resource. A database integrating experimental and computational data. *Nucleic Acids Research*, 31(1), pp.212–215.

Bartesaghi, A., Matthies, D., Banerjee, S., Merk, A. & Subramaniam S., 2014. Structure of β-galactosidase at 3.2-Å resolution obtained by cryo-electron microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 111(32), pp.11709–14.

Bender, R.A., Janssen, K.A., Resnick, A.D., Blumberg, M., Foor, F. & Magasnik, B., 1977. Biochemical parameters of glutamine synthetase from Klebsiella aerogenes. *Journal of bacteriology*, 129(2), pp.1001–9.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. & Bourne,P,E., 2000. The Protein Data Bank. *Nucleic Acids Research*, 28(1), pp.235–242.

Bhatnagar, L., Zeikus, J.G. & Aubert, J., 1986. Purification and Characterization of Glutamine Synthetase from the Archaebacterium Methanobacterium ivanovi. *Journal of bacteriology*, 165(2), pp.638–643.

Bohannon, D.E. & Sonenshein, A.L., 1989. Positive Regulation of Glutamate Biosynthesis in Bacillus subtilis. *Journal of bacteriology*, 171(9), pp.4718–4727.

Bonekamp, F. & Jensen, K.F., 1988. The AGG codon is translated slowly in *E.coli* even at vey low expression levels. *Nucleic Acids Research*, 16(7), pp.3013–3024.

Bowie, J.U., Luethy, R. & Eisenberg, D., 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, 253(5016), pp.164–70.

Brana, A.F., Paiva, N. & Demain, A.L., 1986. Pathways and Regulation of Ammonium Assimilation in Streptomyces clavuligerus. *Journal of General Microbiology*, 132, pp.1305–1317.

Brenchley, J.E., Baker, C.A. & Patil, L.G., 1975. Regulation of the Ammonia Assimilatory Enzymes in *Salmonella typhimurium*. *Journal of bacteriology*, 124(1), pp.182–189.

Buchan, D.W.A., Minneci, F., Nugent, T.C.O., Bryson, K. & Jones D.T., 2013. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic acids research*, 41(Web Server issue), pp.W349–57.

Burkovski, A., 2003. Ammonium assimilation and nitrogen control in Corynebacterium glutamicum and its relatives: an example for new regulatory mechanisms in actinomycetes. *FEMS Microbiology Reviews*, 27(5), pp.617–628.

Chen, G. & Russell, J.B., 1989. Transport of Glutamine by Streptococcus bovis and Conversion of Glutamine to Pyroglutamic Acid and Ammonia. *Journal of bacteriology*, 171(6), pp.2981–2985.

Dainty, R.H., 1972. Glutamate Biosynthesis in Clostridium pasteurianum and its Significance in Nitrogen Metobolism. *Journal of Biochemistry*, 126, pp.1055–1056.

Eisenberg, D., Gill, H.S., Pfluegl, G.M. & Rotstein, S.H., 2000. Structure-function relationships of glutamine synthetases. *Biochimica et biophysica acta*, 1477(1-2), pp.122–45.

Fernández, J.J., Luque, D., Caston, J.R. & Carrascosa, J.L., 2008. Sharpening high resolution information in single particle electron cryomicroscopy. *Journal of structural biology*, 164(1), pp.170–5.

Fierro-monti, I.V.O.P., Reid, S.J. & Woods, D.R., 1992. Differential Expression of a Clostridium acetobutylicum Antisense RNA : Implications for Regulation of Glutamine Synthetase. *Journal of bacteriology*, 174(23), pp.7642–7647.

Fisher, S.H. & Wray, L.V.J., 1989. Regulation of Glutamine Synthetase in Streptomyces coelicolor. *Journal of bacteriology*, 171(5), pp.2378–2383.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadi, M. & Leith, A., 1996. SPIDER and WEB : Processing and Visualization of Images in 3D Electron Microscopy and Related Fields. *Journal of Strural Biology*, 116(1), pp.190–199.

Fuchs, R.L. & Keitster, D.L., 1980. Glutamine Synthetases. *Journal of bacteriology*, 141(2), pp.996–998.

Gising, J., Nilsson, M.T., Odell, L.R., Yahiaoui, S., Lindh, M., Iyer, H., Sinha, A.M., Srinivasa, B.R., Larhed, M., Mowbray, S.L. & Karle, A., 2012. Trisubstituted Imidazoles as *Mycobacterium tuberculosis* Glutamine Synthetase Inhibitors. *Journal of Medicinal Chemistry*.

Goodstadt, L. & Ponting, C.P., 2001. CHROMA : consensus-based colouring of multiple alignments for publication. *Bioinformatics*, 17(9), pp.845–846.

Grigorieff, N., 2007. FREALIGN: high-resolution refinement of single particle structures. *Journal of structural biology*, 157(1), pp.117–25.

Harth, B.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. *Journal of Experimental Medicine*, 189(9), pp.1425–1435.

Heel, M. van., Harauz, G., Orlova, E.V., Schmidt, R. & Schatz, M., 1996. A New Generation of the IMAGIC Image Processing System. *Journal of Structural Biology*, 116(1), pp.17–24.

Hohn, M., Tang, G., Goodyear, G., Baldwin, P.R., Huang, Z., Penczek, P.A., Yang, C., Glaeser, R.M., Adams,P.D. & Ludkte, S.J, 2007. SPARX, a new environment for Cryo-EM image processing. *Journal of structural biology*, 157(1), pp.47–55.

Ikemura, T., 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *Journal of Molecular Biology*, 146(1), pp.1–21.

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. *Journal of bacteriology*, 170(1), pp.400–408.

Jones, D.T., 1999. Protein Secondary Structure Prediction Based on Position-specific Scoring Matrices. *Journal of molecular biology*, 292, pp.195–202.

Koster, A.J., Grimm, R., Typke, D., Hegerl, R., Stoschek, A., Walz, J. & Baumeister, W., 1997. Perpectives of Molecular and Cellular Electron Tomography. *Journal of Structural Biology*, 120(3), pp.276–308.

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

Lander, G.C., Stagg, S.M., Voss, N.R., Cheng, A., Fellmann, D., Pulokas, J., Yoshioka, C., Irving, C., Mulder, A., Lau, P-W., Lyumkis, D., Potter, C.S. & Carragher, B., 2009. Appion: An integrated, database-driven pipeline to facilitate EM image processing. *Journal of Structural Biology*, 166(1), pp.95–102.

Larkin, M. A., Blackshields, G., Brown, N.P., Chenna, R., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopex, R., Thompson, J.D., Gibson, T.J. & Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, 23(21), pp.2947–8.

Liao, H.Y. & Frank, J., 2010. Definition and estimation of resolution in single-particle. *Structure*, 18(7), pp.768–775.

Liaw, S. & Eisenberg, D., 1994. Structural Model for the Reaction Mechanism of Glutamine Synthetase , Based on Five Crystal Structures of Enzyme-Substrate Complexes. *Biochemistry*, 33(3), pp.675–681.

Liaw, S., Jun, G. & Eisenberg, D., 1994. Interactions of Nucleotides with Fully Unadenylylated Glutamine Synthetase from *Salmonella typhimurium* A. *Biochemistry*, 33(37), pp.11184–11188.

Liaw, S., Pan, C. & Eisenberg, D., 1993a. Feedback inhibition of fully unadenylylated glutamine synthetase from *Salmonella typhimurium* by glycine, alanine, and serine. *Proceedings of the National Academy of Sciences of the United States of America*, 90(June), pp.4996–5000.

Liaw, S., Villafranca, J.J. & Eisenberg, D., 1993b. A Model for Oxidative Modification of Glutamine Synthetase, Based on Crystal Structures of Mutant H269N and the Oxidized Enzyme. *Biochemistry*, 32(31), pp.7999–8003.

Liaw, Sh.-H., Kuo, I. & Eisenberg, D., 1995. Discovery of the ammonium substrate site on glutamine synthetase , a third cation binding site. *Protein Science*, 4, pp.2358–2365.

Liu, J., Istvan, E.S., Gluzman, I.Y., Gross, J. & Goldberg, D.E., 2006. Plasmodium falciparum ensures its amino acid supply with multiple acquisition pathways and redundant proteolytic enzyme systems. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23), pp.8840–5.

Lobley, A., Sadowski, M.I. & Jones, D.T., 2009. pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. *Bioinformatics (Oxford, England)*, 25(14), pp.1761–7.

Logusch, E.W., Walker, D.M., McDonald, J.F. & Franz, J.E., 1991. Inhibition of Plant Glutamine Synthetases by Substituted Phosphinothricins. *Plant Physiology*, 95, pp.1057–1062.

Ludtke, S.J., Baldwin, Philip R & Chiu, W., 1999. EMAN : Semiautomated Software for High-Resolution Single-Particle Reconstructions. *Journal of Structural Biology*, 97(128), pp.82–97.

Luethy, R., Bowie, J.U. & Eisenberg, D., 1992. Assessment of protein models with three-dimensional profiles. *Nature*, 356(6364), pp.83–85.

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

Merrick, M.J. & Edwards, R.A., 1995. Nitrogen Control in Bacteria. *Microbiological Reviews*, 59(4), pp.604–622.

Mindell, J. A. & Grigorieff, N., 2003. Accurate determination of local defocus and specimen tilt in electron microscopy. *Journal of Structural Biology*, 142(3), pp.334–347.

Mowbray, S.L., Kathiravan, M.K., Pandet, A.A., & Odell, L.R., 2014. Inhibition of Glutamate Synthetase: A Potential Drug Target in *Mycobacterium tuberculosis*. *Molecules*, 19), pp.13161–13176.

Niesen, F.H., Berglund, H. & Vedadi, M., 2007. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protocols*, 2(9), pp.2212–2221.

Occhipinti, A., Berlicki, Ł. & Giberti, S., 2009. Effectiveness and mode of action of phosphonate inhibitors of plant glutamine synthetase. *Pest Management Science*, 66(2009), pp.51–58.

Olszewski, K.L., Mather, M.W., Mossisey, J.M., Garcia, B.A., Vaidya, A.B., Rabinowitz, J.D. & Llinas, M., 2010. Branched tricarboxylic acid metabolism in Plasmodium falciparum. *Nature*, 466(7307), pp.774–778.

Pascual-Montano, A., Donate, L.E., Valle, M., Barcena, M., Pascual-Marqui, R.D. & Carazo, J.M., 2001. A novel neural network technique for analysis and classification of EM single-particle images. *Journal of structural biology*, 133(2-3), pp.233–45.

Payne, S.H. & Loomis, W.F., 2006. Retention and loss of amino acid biosynthetic pathways based on analysis of whole-genome sequences. *Eukaryotic cell*, 5(2), pp.272–6.

Pesole, G & Lanave, C, 1991. Glutamine synthetase gene evolution: A good molecular clock. *Proceedings of the National Academy of Sciences of the United States of America*, 88(January), pp.522–526.

Pesole, G., Gissi, C., Lanave, C. & Saccone, C., 1995. Glutamine Synthetase Gene Evolution in Bacteria. *journal of Molecular Biology Evolution*, 12(2), pp.189–197.

Radchenko, M.V., Thornton, J. & Merrick, M., 2013. P II signal transduction proteins are ATPases whose activity is regulated by 2-oxoglutarate. *Proceedings of the National Academy of Sciences of the United States of America*, 10(32), pp.12948–12953.

Reitzer, L., 2003. Nitrogen assimilation and global regulation in *Escherichia coli*. *Annual review of microbiology*, 57, pp.155–76.

Robinson, M., Lilley, R., Little, S., Emtage, J.S., Yarranton, G., Stephens, P., Millical, A., Eaton, M. & Humphreys, G., 1984. Codon usage can affect efficiency of translation of genes in *Escherichia coli*. *Nucleic Acids Research*, 12(17), pp.6663–6671.

Rooyen, J.M.van., Abratt, V.R., Belrhali, H. & Sewell, T., 2011. Crystal Structure of Type III Glutamine Synthetase : Surprising Reversal of the Inter-Ring Interface. *Structure/Folding and Design*, 19(4), pp.471–483.

de la Rosa-Trevín, J.M., Oton, J., Marabini, R., Zaldivar, A., Vargas, J., Carazo, J.M. Sorzano, C.O.S., 2013. Xmipp 3.0: an improved software suite for image processing in electron microscopy. *Journal of structural biology*, 184(2), pp.321–8.

Roseman, A, 2004. FindEM—a fast, efficient program for automatic selection of particles from electron micrographs. *Journal of Structural Biology*, 145(1-2), pp.91–99.

Scheres, S.H.W., 2012. A Bayesian view on cryo-EM structure determination. *Journal of molecular biology*, 415(2), pp.406–18.

Sippi, M.J., 1993. Recognition of Errors in Three-Dimensional Structures of Proteins. *Proteins*, 17), pp.355–362.

Stadtman, E.R., 2001. The story of glutamine synthetase regulation. *The Journal of biological chemistry*, 276(48), pp.44357–44364.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I. & Ludtke S.J., 2007. EMAN2: an extensible image processing suite for electron microscopy. *Journal of structural biology*, 157(1), pp.38–46.

Wiederstein, M. & Sippi, M.J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of protein. *Nucleic Acids Research*, 35, pp.407–410.

Woods, D.R. & Reid, S.J., 1993. Recent developments on the regulation and structure of glutamine synthetase enzymes from selected bacterial groups. *FEMS Microbiology Reviews*, 11(4), pp.273–283.

Wriggers, W., 2010. Using Situs for the integration of multi-resolution structures. *Biophysical reviews*, 2(1), pp.21–27.