Structure-function investigation of substrate-binding proteins from *Synechococcus sp*.

By Benjamin Ford, BSc.

A thesis submitted in partial fulfilment of the requirements for the degree
Masters of Research (MRes)

Department of Chemistry and Biomolecular Sciences (CBMS)

Macquarie University

9th October 2017

Table of contents

Acknowledgements

My heartfelt thanks must first go to my supervisor, A/Prof. Bridget Mabbutt for her expert guidance and mentorship throughout what has been a turbulent year. Bridget's encouragement has been paramount to me realising this thesis. I would also like to express my gratitude towards Prof. Ian Paulsen, Dr. Martin Ostrowski and Dr. Bhumika Shah for providing the direction for my project and their continued support in shaping it along the way, as well as to Heather for her patience and willingness to take me under her wing in the lab.

I would also like to thank Dylan, Patrick and Steph for the free therapy sessions, and their willingness to entertain both my awful sense of humour and propensity to speak in French without warning. I'd like to extend my deepest thanks to Josh for the many cheese and wine nights, and your continued support throughout this year. Finally, I would like to acknowledge the on-going contributions my family have made in helping me reach this milestone. Getting to the finish line would have been much more difficult without all of you. *Merci à tous!*

List of Abbreviations

Styles adopted in this work are according to the Instructions for Authors in the Journal of Molecular Biology

A_{280}	absorbance at 280 nm
ABC	ATP-binding cassette
dNTPs	deoxyribose nucleoside triphosphates
DSF	differential scanning fluorimetry
gDNA	genomic DNA
IMAC	immobilised metal affinity chromatography
K_{av}	distribution coefficient for size exclusion chromatography
LIC	ligation-independent cloning
LB	Luria-Bertani media
LGT	lateral gene transfer
MQ/SGC	Macquarie University/Structural Genomics Consortium
M_R	molecular radius within solution
OD_{600}	optical density at 600 nm
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein databank
SBP	substrate-binding protein
SDS	sodium dodecyl sulphate
SEC	size exclusion chromatography
TEMED	*n,n,n',n'*-tetramethyl-ethane-1,2-diamine
T_m	melting temperature
OPPF-UK	Oxford Protein Production Facility – United Kingdom
V_e	elution volume of a species of interest
V_o	void volume of chromatography matrix
V_t	total volume of chromatography matrix

Declaration

Where appropriate, work done in collaboration with other groups or individuals has been acknowledged. Aside from these contributions, the material contained in this thesis is entirely my own work, and to the best of my knowledge is original.

No part of this thesis has been submitted for a higher degree to any other university or institution. I consent to my thesis being made available for photocopy and loan.

Ethics approval was not required for this project.

This project was conducted under exempt dealings, approved by the Institutional Biosafety Committee (Ref# 5201400046).

Benjamin Ford
9 October 2017

Abstract

Synechococcus is a ubiquitous genus of photosynthetic marine bacteria, and vital for primary production and nutrient cycling in global oceans. Genomic annotation for this genus has progressed faster than the functional characterisation of predicted proteins. Consequently, the veracity of predicted functions for many genes is questionable. This is particularly so for the substrate-binding protein (SBP) family, integral to nutrient import *via* ATP-binding cassettes, which display unusually low sequence homology (despite a highly conserved structural fold). Lack of validation concerning the natural ligands of SBP proteins impacts on our depth of understanding of the function and utility of cellular import machinery in these highly prevalent marine microorganisms.

A panel of nineteen SBPs annotated to bind saccharides across representative *Synechococcus* species has been selected for *in vitro* characterisation. Two alternative formats for high-throughput production were trialled, and the yield and percentage of soluble protein products compared. Recombinant forms of three SBPs have been produced in sufficient quantities to enable ligand screening using differential scanning fluorimetry and biophysical characterisation in solution. The approach undertaken has demonstrated potential to complement genomic data for previously uncharacterised SBPs.

Chapter 1: Molecular Understanding of Bacterial Transport

1.1 – Cellular Transport in Prokaryotes

Vital to the survival of all living cells is the requirement to simultaneously import nutrients and to remove waste and toxic compounds [1-3]. Prokaryotes (unicellular organisms lacking membrane-bound organelles) perform this function through various transport machinery elements at the interface between their cellular membrane and the environment [4] (Fig. 1.1). Up to one-third of the entire proteome in prokaryotes corresponds to membrane proteins, many specifically dedicated to transport [1]. Comparative analyses of 141 prokaryotic genomes shows 16 % of all open reading frames (ORFs) are predicted to encode transport proteins [5].

From an evolutionary perspective, this abundance of molecular transporters reflects the underlying metabolic demand of cell growth. *Escherichia coli*, for example, requires the uptake of 10^6 glucose molecules per cell, per second during active growth [2, 6]. Not all natural environments can provide the rich nutrient source required for this high metabolic turnover. Oligotrophic oceans, for instance, are characterised by a dearth of nutrients [7, 8]. For microorganisms to survive, they must successfully compete for essential nutrients [9, 10]. Autotrophic bacteria, such as *Synechococcus sp.*, are able to meet their own energy requirements, yet still require growth-limiting nutrients such as metallic ions, sulphur-, and nitrogen-containing compounds [5, 6, 11]. In low-nutrient environments, cellular machinery facilitating high-affinity uptake of these nutrients would impart a distinct evolutionary advantage.
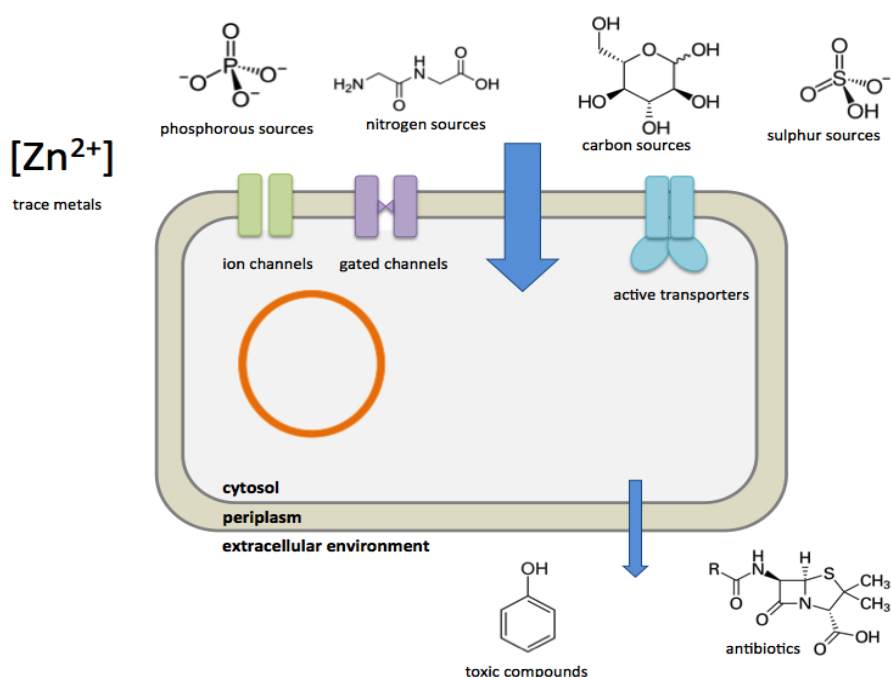

Fig 1.1 Chemical flux in prokaryotes. Stylised bacterial cell, showing essential nutrients are imported into cells using dedicated machinery at the cell membrane. Exporters contribute to the flux of chemicals in the cell by removing unwanted or toxic compounds.

1.2 – Cyanobacteria

Cyanobacteria are ubiquitous among all global ocean environments [12]. This collection of bacteria is an ancient phylum and, as the first organisms to utilise photosynthesis, are generally held to have been responsible for the 'Great Oxygenation Event' of 2.3 billion years ago [13, 14]. Cyanobacteria have, and continue to, shape our global oceans and atmosphere due to their vital roles in photosynthesis, carbon and nutrient cycling, and contribution to global ocean biogeochemical processes [10].

One subgroup, the picocyanobacteria, contains two of the most abundant primary producers of oxygen on earth: the Gram-negative bacteria *Synechococcus* and *Prochlorococcus*. Globally, these species contribute an estimated 16.7 % and 8.5 %, respectively, to net oxygen production in oceans [12, 15]. The long history of these two genera is demonstrative of their evolutionary success, with both bacteria inhabiting diverse environments including estuarine, coastal, and offshore waters across all latitudes, with the exception polar regions [10].

The genome organisations of *Prochlorococcus* and *Synechococcus* appear remarkable. For these genera, their genomes of 1.6-2.7, and 2.2-2.8 mega base pairs (Mbp), respectively, are notably smaller than the average genome size of other cyanobacteria (5.3 Mbp) [10]. This raises questions about the genetic basis for their global success across evolutionary history. For *Prochlorococcus*, this reduced genome size is coupled with a reduction in GC content, indicative of extensive genome streamlining (to remove superfluous genes) [10].

This is not true for *Synechococcus* which, rather, has been described as a "genomic mosaic" [16]. Comparison by Partensky's team of eleven *Synechococcus* isolates allowed the delineation of nearly 8,000 protein families into 'core' and 'accessory' genomes [16]. The number of unique genes was variable but correlated positively with genome size [17]. A large proportion of these unique genes were localised in genomic 'islands', *i.e.* syntenic blocks of genes acquired through lateral gene transfer (LGT) [18], suggesting the pangenome of *Synechococcus* to be ultimately larger and far more diverse than defined by any individual isolate [19, 20].

Prokaryotic organisms such as *Synechococcus sp.* are perhaps best classified according to ecotype, or populations that occupy distinctly different environments [21]. These are best analysed through the detection of conserved markers, such as 16s RNA [22]. For *Synechococcus* the use of the *petB* transcript, which encodes the conserved photosynthetic pigment phycoerythrin, has allowed the delineation of three main subclusters and nine separate clades [10].

The Cyanorak database (www.sb-roscoff.fr/cyanorak), maintained by Station Biologique – Roscoff, France, is a repository for genomic and transcriptomic data for the

marine picocyanobacteria *Prochlorococcus* and *Synechococcus* [23]. As at June 2017, the database contains 51 *Synechococcus* genomes arising from individual ecotypes successfully cultured axenically following global ocean sampling [10]. These ecotypes are assumed to be broadly representative of global distributions and incorporate variations in ocean, seasonality, and latitudinal distribution [10]. These are summarised in Table 1.1 with representative strains and typical environmental locations listed for each clade. The global origin of each clade and their representative strains are shown in Fig. 1.2. Whilst identifying the genetic basis for niche adaptation for *Synechococcus* has been an area of intensive research [10], the addition of transcriptomic information to help elucidate the complex community nature of marine microbial taxa allows functional information to be ascribed to observed genetic diversity. Recent microarray tools that identify strain-level heterogeneity [24] have now allowed previously unknown metabolic and cellular processes to be identified [25].

Table 1.1 Synechococcus clade numbers and their distributions

Clade [a]	Representative strain	Environmental conditions	Latitudinal band
I	CC9311	coastal/temperate mesotrophic open ocean waters.	above 30°N and below 30°S.
II	CC9605	open ocean, continental shelf and offshore, oligotrophic tropical and subtropical waters.	between 30°N and 30°S
III	WH8102	ultraoligotrophic open oceans.	
IV	CC9902	coastal/temperate mesotrophic open ocean waters.	above 30°N and below 30°S.
V/VI/ VII	WH7803/ WH7805/ RS9917	widely distributed particularly in areas of upwelling, low abundance.	
CRD1	MITS9220	high nutrient, low chlorophyll, iron limited.	equatorial upwelling region (all depths)

[a] Identified using petB marker [10]

Fig. 1.2 Map of *Synechococcus sp.* clade distributions (M. Ostrowski, personal communication).

Functional characterisation, rather than genetic, is also useful due to the propensity for phage-mediated LGT [16], particularly in populations occupying similar niches (for instance Clades I and IV, Fig 1.2) and provides a more holistic understanding of community composition and dynamics [24, 25]. Several investigations to date have focused on the acquisition of growth-limiting nutrients in related species, such as metal-binding proteins from *Synechocystis* [26-28].

1.3 – Cellular Solute Transport Mechanisms

1.3.1 – The ATP-binding cassette transporters

Despite the diverse range of membrane transport machinery available to bacteria for both passive and active nutrient uptake, evolutionary pressures can introduce a selection bias. Genomic annotation of *Synechococcus WH8102* indicates that while genes associated with transport machinery comprise 5-6 % of predicted ORFs (a value consistent with other prokaryotic genomes [6]), its transporter capability displays a heavy bias (60 %) towards ATP-binding cassettes (ABC transporters) [17].

ABC transporters comprise one of the largest superfamilies of transporters, and are highly conserved across all domains of life. As depicted in Fig. 1.3, these perform active translocation of small chemical substrates across membranes following hydrolysis of ATP [2, 3]. ABC importers are present solely within prokaryotes and some plants [6, 29, 30].

Multiple structures of this family of proteins have been solved, with the most recent systematic review of available ABC transporter structures covering fourteen examples [31].

This structural analysis of the ABC transporters showed that they could be grouped into 4 main classes: Type I importers; type II importers; energy-coupling factor/Type III importers; and exporters [31].

At the molecular level, the ABC transporter is comprised of four core domains: two trans-membrane spanning domains (TMD), and two highly conserved nucleotide-binding domains (NBD) [29, 32]. These elements are all transcribed together in a single operon [30].

In all ABC transporter types, the TMDs form the translocation pathway, with the cis- and trans-sides of the membrane alternatively accessible, enabling the translocation of substrates as shown in Fig. 1.3. These TMDs may be either homodimers or structurally similar (although this is class dependent) and the arrangement of the NBDs is dimeric. Binding of ATP between the two subunits facilitates the closure of the NBD dimer interface, forming a tight head-to-tail dimer sandwich [33, 34].

In addition to their overarching architecture, between the three importer types there is a certain degree of similarity with respect to substrate specificity. This raises the question for the evolutionary selection of three similar transporters. Whereas Type I importers generally translocate substances required in bulk, type II and energy-coupling factor transporters generally display much higher specificity for compounds, such as metal chelates, which may only be needed in small quantities [31]. Substrate recognition is conferred by the additional subunit, the substrate-binding protein (SBP), a molecular adapter unique to prokaryotes [32, 35]. For type I and type II importers, the involvement of the SBP is integral to their function. Docking of the SBP, however, does not induce sufficient conformational change to move the TMD between an outward- and inward conformation, with the power stroke instead being generated by ATP hydrolysis [2] at the NBDs.
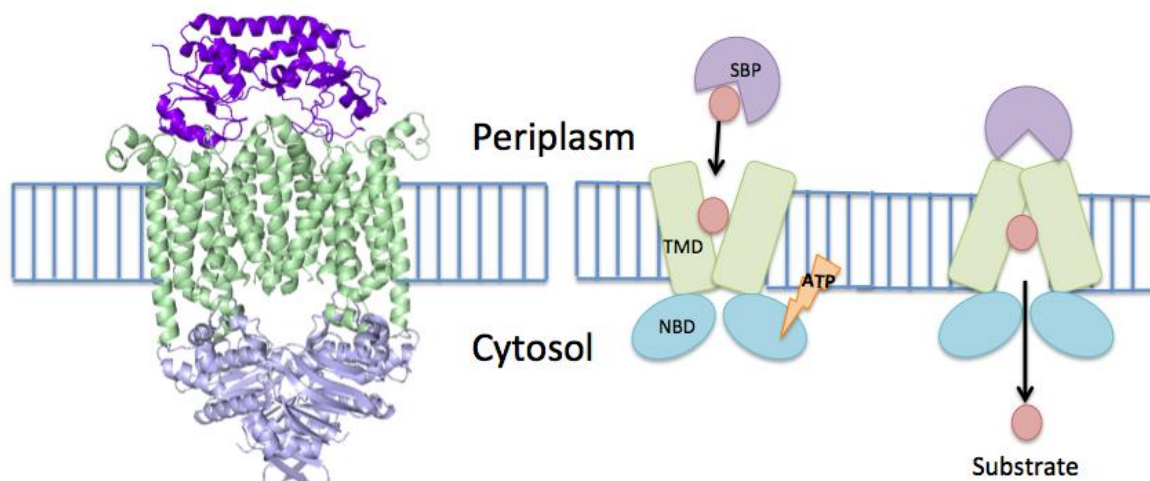
Fig 1.3 Molecular architecture of ABC transporters. Left, cartoon representation of vitamin B12 ABC transporter from *E. coli* (PDB ID: 2QI9) [36] showing the two nucleotide binding domains (blue), the two trans-membrane spanning domains (green), with the additional SBP shown (purple). Right, stylised representation of mechanism of substrate translocation.

1.3 – The Substrate-binding Protein

SBPs are associated with prokaryotic ABC transporters for the import of solutes. In Gram-negative bacteria, such as *Synechococcus sp.*, the SBP is a soluble protein that freely diffuses through the periplasm [37]. In Gram-positive bacteria (which lack a periplasmic space), the SBP is a lipoprotein tethered to the cell membrane [37]. SBPs bind to their cognate ligand with high specificity. High-affinity binding allows rapid responses even in low ligand concentrations, particularly in cases where these proteins also play a role in regulatory mechanisms (such as transcriptional regulation) [35].

The molecular architecture of these proteins is highly conserved, consisting of two globular α/β domains connected by a hinge region [38, 39]. Each domain comprises a core β-sheet of between four and six β-strands that is flanked by α-helices, and are reminiscent of a Rossmann-like fold [38, 39]. The hinge region generally provides the defining molecular architecture necessary to delineate SBPs into their structural classes (with the exceptions of Classes C and G) [38]. Whilst the open conformation appears to be highly flexible [40, 41], ligand binding induces a conformational change bringing its two separated domains together and capturing the ligand in the interdomain cleft. This 'Venus Fly-trap" mechanism was first described by Quicho [41-43] (Fig 1.4).
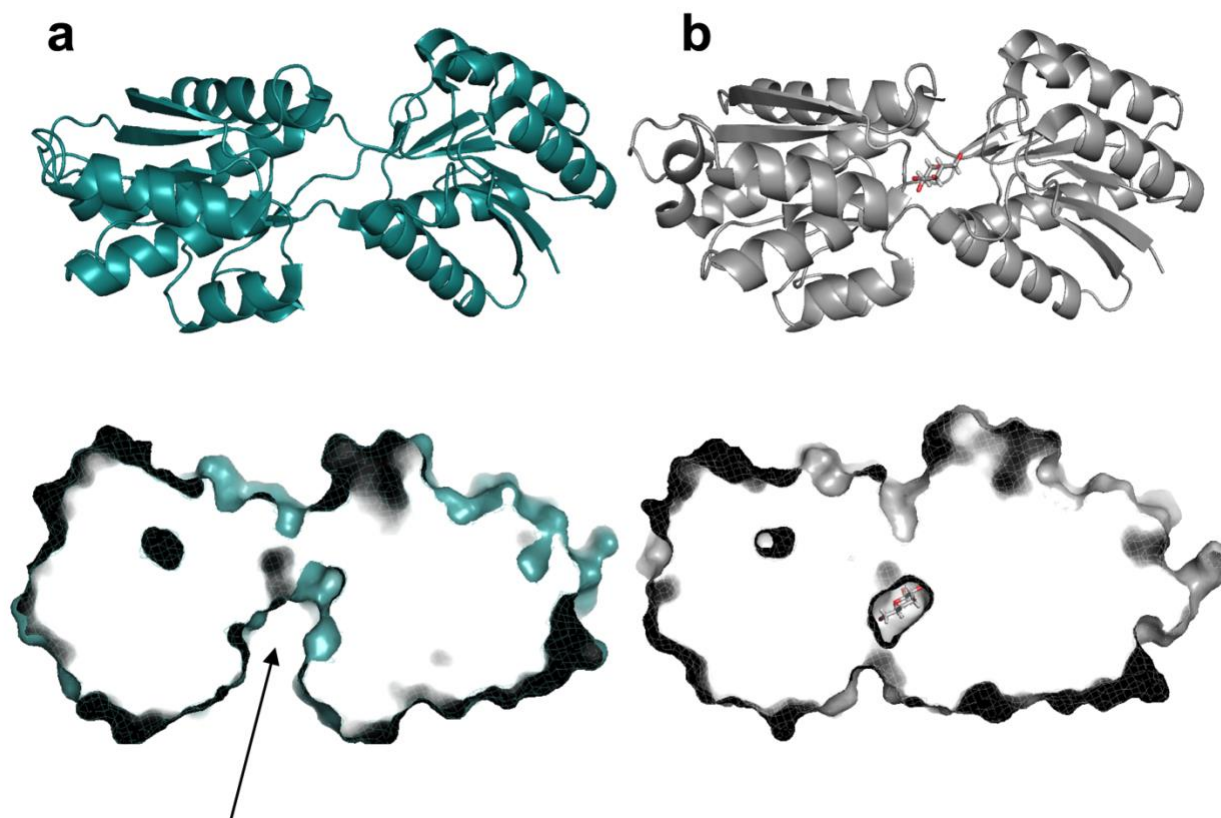
Fig. 1.4 Mechanism of ligand binding. Glucose/galactose-binding protein (ggBP) from *E. coli* [40] is shown in an **(a)** un-liganded conformation and **(b)** with bound glucose. The binding pocket in the un-liganded conformation is indicated with an arrow.

Despite their highly conserved structural fold, a characteristic feature among this family of proteins is a remarkably low degree of homology at the primary sequence level [38, 44]. As such, phylogeny-based alignments are generally unreliable for predicting binding partners *in silico*, with additional experimental validation (typically the generation of a structure) required to assess the veracity of these annotations.

The first reported structurally-based classification was performed by Fukami-Kobayashi in 1999 [45], and delineated SBPs into two classes, according to the topology of the β-sheet core of each domain, to which a third class was later added based on the presence of an α-helix in the hinge region. The most recent comprehensive structural classification (performed by the Poolman group [39]) incorporated 501 structures, identifying seven different structural groupings, named Clusters A to G, as depicted in Fig. 1.5. Representative structures of each class have been shown, and their defining structural elements highlighted. For each class, the number of representative proteins and their common ligand chemistries have been identified in Table 1.2.

Cluster A, for example, includes those structures containing a single α-helix within the inter-domain linker, shown in orange on Fig. 1.5. Examples of representative SBPs in this class include the zinc-binding proteins (ZnuA), and the vitamin B12-binding protein. The ligand chemistry of this group encompasses metal ions, and siderophores/hemes.

Cluster E, including the TakP protein function exclusively with tripartite ATP-independent (TRAP) transporters and have an extended helix forming the linker region, as indicated in Fig. 1.5. TRAP transporters utilise electrochemical gradients, rather than a nucleotide energy source to power active translocation. TakP, responsible for binding α-keto acid in *Rhodobacter sphaeroides* exists as a helix-swapped dimer [46] (Fig. 1.6) and is one of the few SBPs demonstrated to function as a dimer. Dimerisation with other SBPs has been observed within solution but the functional, physiological significance of this has not been explored.

Table 1.2 Substrate specificity for structural clusters of SBPs. Adapted from [39].

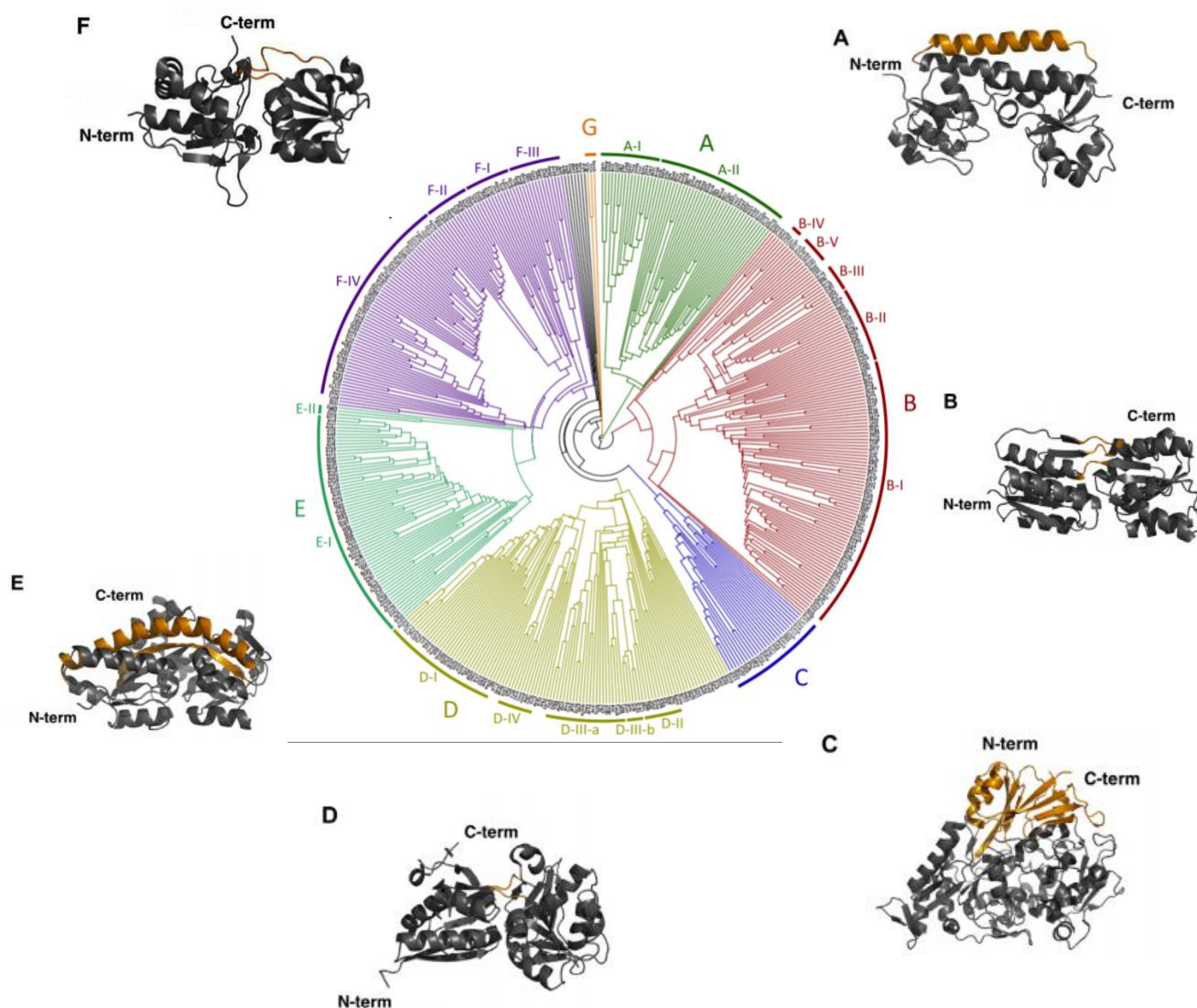| Cluster	Substrates	#
A	metal ions; including heme and siderophores, metal oxoanions	55
B	sugars and sugar alcohols, amino and aromatic acids	129
C	di- and oligopeptides, opine, nickel, cellobiose	30
D	sugars, sugar alcohols, spermidine, phosphate, metal oxoanion, iron (ionic)	108
E	organic acids, amino alcohols, amino acids, dipeptides	70
F	thiamine, pyrimidines, sulphonates, bicarbonates, amino acids, compatible solutes	106
G	indeterminate due to sample size	3

Figure 1.5: Structural classification of substrate binding proteins. 501 SBPs were clustered into eight classes based on their structural relatedness in the last comprehensive SBP review. Archetypal examples of molecular architecture are shown for each cluster with defining features highlighted. Adapted from [38].
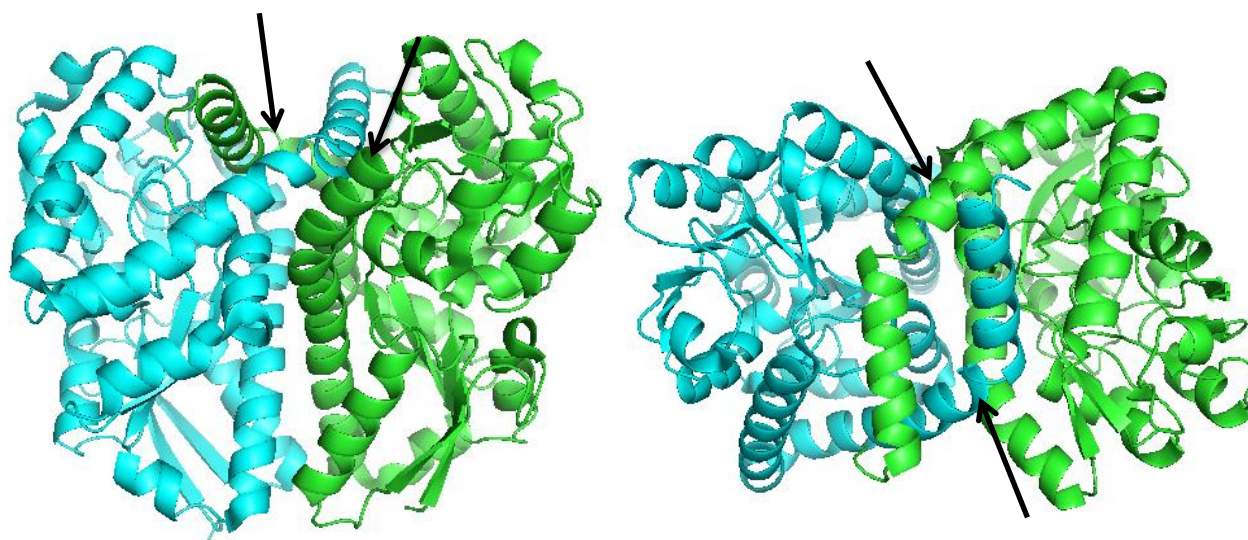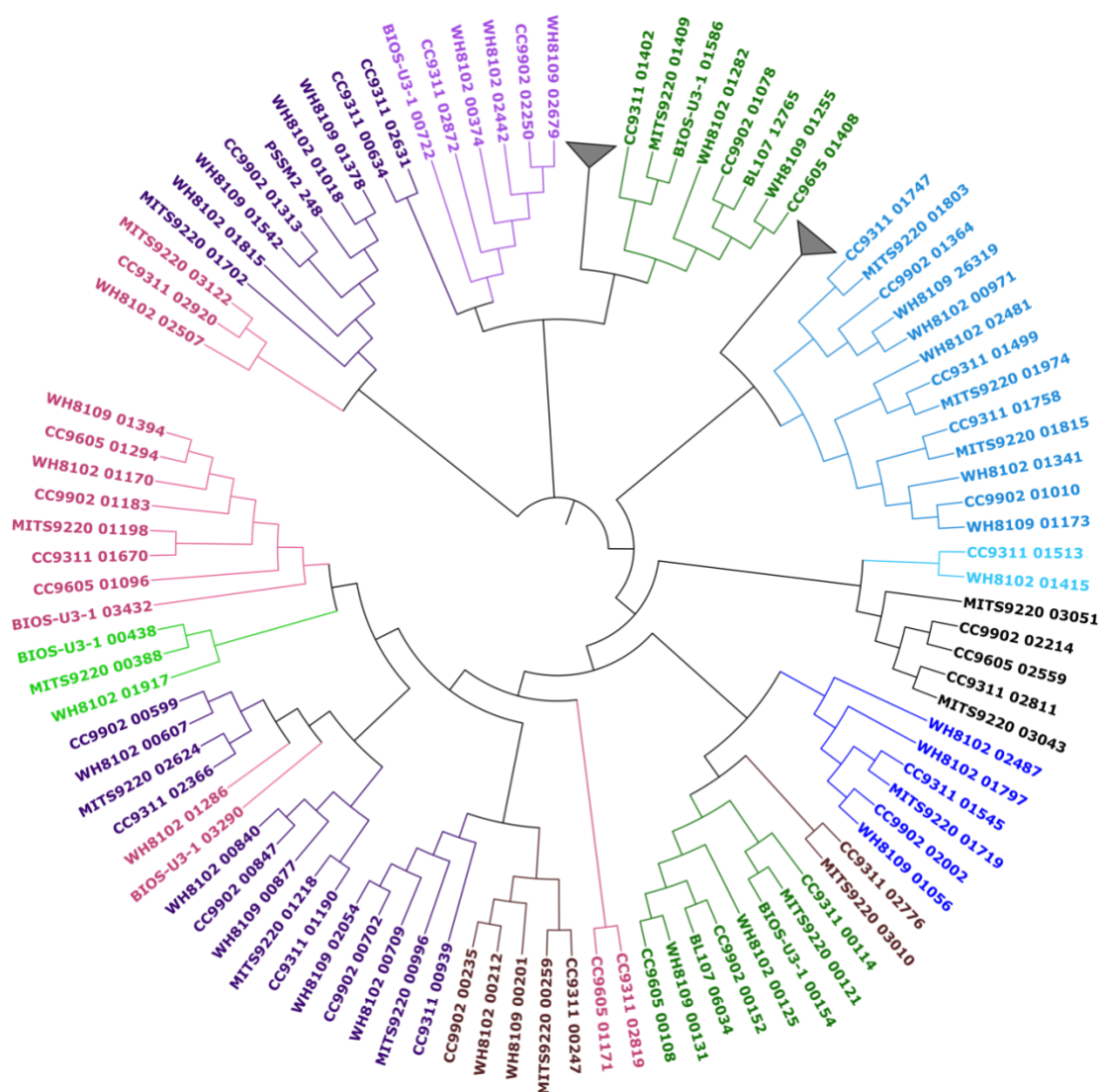
Fig 1.6 Dimerisation of TakP. The homodimer of TakP [46] has been coloured by each monomer and is shown from the front (left) and top (right) indicating the helix-swapped dimer interface involving the linker-region helices (indicated with arrows).

1.4 – Substrate-binding Proteins in *Synechococcus sp.*

Annotated SBP componentry in all known *Synechococcus* genomes was identified using text-based searching (S. Mazard, Macquarie University, personal communication). This identified 198 orthologous gene clusters corresponding to 6930 annotated genes for various subunits of ABC transport machinery. This was largely due to high sequence homology between the NBD components ABC transporters transcribed in single operons.

Eight representative strains sufficiently robust to be grown under laboratory conditions have been selected for further study of their ABC importer machinery by B. Shah (Macquarie University). From this, a phylogenetic tree was constructed specifically using evolutionary relationships between the gene sequences and to predict their ligand chemistry (Fig. 1.7). This sequence-based approach differs from the structurally-based method applied by Scheepers et al [39] to generate Fig. 1.5.

The phylogenetic tree depicted in Fig. 1.7 shows a diverse chemical space covered by *Synechococcus* transport machinery. The presence of growth-limiting nutrients, such as ferric iron [8], phosphorous [7, 47], and nitrogen sources [47], is obviously a consequence of their importance for cell size, growth, and morphology [48]. However, the biological rationale for other chemistries covered by this SBP repertoire, such as aromatic carbon compounds and saccharides, is less clear. Nineteen SBPs are predicted to bind mono- or di-saccharides. The presence of componentry for saccharide import in an organism capable of photosynthesis is paradoxical given their ability to synthesise these compounds *de novo*, making these SBPs attractive candidates for functional characterisation.

10

Fig 1.7 Phylogenetic tree of SBPs from representative strains under investigation at Macquarie University. Each leaf corresponds to a gene tag in an individual *Synechococcus* isolate. Ligands were predicted for these based on their genome-level annotation or PsiBlast searching for clusters without an annotated ligand. Following clustering, the tree has been coloured to indicate the chemistry of their putative ligands. (B. Shah, personal communication).

1.5 – Aims of My Work

Functional understanding of Picocyanobacteria, such as *Synechococcus sp.*, has failed to keep pace with their genomic annotation. Predictions made *in silico* regarding function can allow inferences to be made about their biochemistry, but must be underpinned by robust annotation at the genomic level. For some protein families, the validity of these annotations is questionable. The SBP family is a prime example of this, whose ligand partners cannot be reliably identified based on phylogeny alone. The question of their function must be investigated using alternative approaches, such as characterisation *in vitro*, or identification of biological importance through the use of knockouts and targeted mutagenesis studies *in vivo*.

SBPs can serve multiple functions, from molecular adaptors for greater nutrient import machinery, to roles in transcriptional regulation. Recent identification of previously unknown metabolism in related cyanobacteria species highlights that there is still much more to discover about these processes in microbial systems [9, 49]. Identifying the role played by SBPs in such processes may have profound implications for the current paradigms of fundamental biochemistry for *Synechococcus sp..*

Ascribing function to uncharacterised proteins is regularly achieved through the generation of structural information. SBPs have evolved to recognise diverse ligands, yet still retain a highly conserved fold. SBPs in TRAP systems have previously been subjected to large-scale structural genomics comparison [35]. While their solubility and stability make them attractive candidates for structural studies, this may not be the most effective means of investigating their functional role, which is instead dictated by their binding partner. Identification of ligand candidates *in vitro* can provide vital information for their biological role.

In vitro characterisation requires the production of sufficiently pure protein, meaning large-scale recombinant production and purification is essential. One strategy is to adopt a structural genomics pipeline, such as those developed at consortia like the Structural Genomics Consortium (SGC) at the University of Toronto [50, 51]. These pipelines rely on identifying robust and tractable candidates from a large panel of targets. There is currently no established method for the large-scale recombinant production of proteins from *Synechococcus sp.*, however the amenable properties of the SBP family make these attractive candidates for trialling the development of such a large-scale approach.

My study focuses on the *in vitro* functional characterisation for a panel of nineteen SBPs annotated to bind carbohydrates. These have been identified in the genomes of eight representative *Synechococcus* ecotypes, yet the biological rationale for their presence in an

autotrophic bacterium is unclear. This approach relies on the large-scale recombinant expression of protein material, and biophysical analysis of their behaviour within solution.

The Protein Structure Group at Macquarie University has previously developed methods for the recombinant production of mobile gene cassettes [52]. SBPs from *Synechococcus* are similar to these in many respects, being highly conserved and playing vital roles in microbial adaptation, yet largely uncharacterised. This method will form the basis for developing a strategy for heterologous expression, prior to characterising these proteins within solution.

Therefore, the specific aims of my project were to:
- Develop a strategy for the recombinant expression of *Synechococcus* SBPs.
- Optimise purification and handling of recombinantly expressed protein.
- Provide biophysical characterisation of these proteins within solution
- Interrogate their functional role by carrying out ligand screening to identify binding partners.

Chapter 2: Materials and Methods

2.1 Reagents

All reagents used, minimally of analytical grade, were obtained through mainstream suppliers (Astral Scientific, Sigma Aldrich and VWR) unless otherwise specified. DNA primers were obtained from Integrated DNA Technologies (Sydney, Australia). Common media and buffers used throughout experimental procedures are detailed in Table 2.1. Solutions were pH adjusted using HCl or NaOH (5 M) where indicated. Purified water from a MilliQ system (Millipore) was used for the preparation of buffers and reagents unless otherwise stated. All solutions for chromatography were filtered (0.2 μm) and degassed in a sonicator prior to use. Glassware used in culturing of *Synechococcus sp.* was acid-washed prior to use and polyethylene equipment was used to prepare all media to prevent trace metal contamination.

2.2 Bioinformatics Analyses

The nucleotide sequences for genes of interest were translated into primary protein sequences using the ExPASY translate tool. Protein candidates for recombinant preparation were analysed using GlobPlot 2 [53] to predict regions of globularity or disorder, TMHMM v2.0 [54] to predict trans-membrane regions, and SignalP 3.0 [55] and SOSUI [56] to identify signal sequences. Default parameters for each program were used for all analyses.

2.3 Molecular Biology

All steps were performed maintaining aseptic conditions (flame or biological safety flowhood). *Escherichia coli* (*E. coli*) were cultured using Luria-Bertani (LB) medium, unless otherwise stated. With the exception of transformation steps, LB media contained antibiotics for positive selection of propagation strain (ampicillin, 50 μg mL^{-1}) or expression strain containing the transformed plasmid (ampicillin, 50 μg ml^{-1}; chloramphenicol, 25 μg mL^{-1}).

2.3.1 Purification of genomic DNA from source organism

Single strain isolates of *Synechococcus sp.* previously obtained through flow cytometry sorting were kindly gifted by M. Ostrowski (Macquarie University). Strains and database (Cyanorak, www.sb-roscoff.fr/cyanorak) deposition details are provided in Table 2.2. Isolates (5 mL) were used to inoculate artificial seawater media (500 mL). Cultures were grown at 22 °C (2-3 weeks) under constant illumination (white LED) until the optical density at 750 nm (OD$_{750}$) reached ~ 0.2 Absorbance Units (AU). Genomic DNA (gDNA) was extracted using the cetyltrimethyl ammonium bromide (CTAB)-phenol-chloroform method [57], before being dried down and re-suspended in nuclease-free water.

Table 2.1 Common media and buffers utilised in this work

Media or Buffer [a]	Composition	Ref.
LB	tryptone (10 g), yeast extract (5 g), NaCl (5 g)	[57]
LB agar	tryptone (10 g), yeast extract (5 g), NaCl (5g), agar (bacteriological grade, 15 g).	[57]
artificial seawater	NaCl (25 g), NaNO$_3$ (0.75 g), MgCl$_2$ (2.0 g), KCl (0.50 g), CaCl$_2$ (0.50 g), MgSO$_4$ (3.5 g), Tris base (1.1 g), K$_2$HPO$_4$ (0.030 g).	[58]
trace metal stock (1000x)	H$_3$BO$_3$ (2.86 g), MnCl$_2$•4H$_2$0 (1.81 g), ZnSO$_4$•H$_2$O (0.222 g), Na$_2$MoO$_4$•2H$_2$O (0.390 g), CuSO$_4$•5H$_2$O (0.008 g), Co(NO$_3$)$_2$•6H$_2$O (0.0494 g), FeCl•6H$_2$O (3.0 g), EDTA(Na$_2$Mg) (0.5 g).	[59]
TE buffer	Tris (10 mM, pH 8.0), EDTA (0.1 mM)	[59]
TAE Buffer	Tris (80 mM), acetic acid (0.7% v/v), EDTA (2 mM)	[59]
SOC media	tryptone (20 g), yeast extract (5 g), NaCl (10 mM), KCl (2.5 mM), MgCl$_2$ (10 mM), MgSO$_4$ (10 mM), glucose (1.1 M, after autoclaving).	[59]
M9 salts	KH$_2$PO$_4$ (15 g), Na$_2$HPO$_4$•7H$_2$0 (64 g), NaCl (2.5 g), NH$_4$Cl (5 g)	[57]
ZYP-rich media	tryptone (1% w/v), yeast extract (0.5% w/v), NaCl (0.17 M), MgSO$_4$ (2 mM), Na2HPO4 (50 mM), KH2PO4 (50 mM), (NH4)2SO4 (25 mM), trace metal stock (1% v/v), glycerol (0.5% v/v), glucose (0.05% w/v), lactose (0.2% w/v).	[59]
Buffer A	HEPES (50 mM, pH 7.4), NaCl (300 mM), glycerol (5% v/v)	
Buffer B	HEPES (50 mM, pH 7.4), NaCl (300 mM), glycerol (5% v/v), imidazole (5mM), DNAse (100 mg), lysozyme (300 mg)	

[a] quantities specified per. 1 L

Table 2.2 *Synechococcus* strains utilised in experimental procedures

Strain [a]	Geographic origin	Ref.
BL107	Spanish coast; Mediterranean Sea	[16]
CC9311	California current; Pacific Ocean	[60]
CC9605	California current; Pacific Ocean	[16]
CC9902	California current; Pacific Ocean	[16]
BIOSU31 [b]	Chilean upwelling; Pacific Ocean	[23] [c]
MITS9220 [b]	Equatorial Pacific; Pacific Ocean	[23] [c]
WH8102	Caribbean Sea; Atlantic Ocean	[16]
WH8109 [b]	Sargasso Sea; Atlantic Ocean	[23] [c]

[a] Designated in Cyanorak database (www.sb-roscoff.fr/cyanorak) as at Jan 2017
[b] Genome not publicly available
[c] Unpublished data

2.3.2 PCR amplification of target genes

Primer sequences were designed to be complementary to the pET-15b vector before being evaluated using OligoCalc (basic.northwestern.edu/biotools/OligoCalc.html) and the virtual PCR tool of SerialCloner (serialbasics.free.fr/SerialCloner.html). gDNA was amplified by PCR according to reagents and quantities stipulated in Table 2.3, using primers as given in Table 2.4. The following thermal cycle was used: 94 °C (5 min); 94 °C (30 s) + 55 °C (30 s) + 72 °C (30 s), 30 cycles; final extension 72 °C (10 min).

Amplified genes were visualised using agarose gel electrophoresis. Agarose gel (1.2% w/v) was prepared in TAE buffer (Table 2.1), with the addition of 1 µL commercial dye (GelRed, Biotium). Electrophoresis was carried out at 100 V for 50 min using 1x TAE as the running buffer. Gels were imaged using a GelDoc imaging system (BioRad). Amplicons were purified using a commercial PCR cleanup kit (Qiagen) and DNA concentration measured using absorbance at 260 nm (A_{260}) on a Nanodrop spectrophotometer (ThermoScientific). Sample purity was also estimated using A_{260}/A_{280} ratio.

2.3.3 Ligation-independent cloning of target genes

Ligation-independent cloning (LIC) into the pET-15b vector (Fig 2.1b) [61] was carried out following the manufacturer's protocol (Clontech) with the following modifications: the gene for insertion was combined in a 3:1 molar ratio with 1 µL vector (Fig 2.1a) previously digested using *BamHI* and *NdeI*; following incubation (37 °C, 30 min), the reaction mixture (2.5 µL) was added to thawed *E. coli* Stellar competent cells (Clontech) (25 µL) and stored on ice (30 min); cells were heat-shocked (42 °C, 60 s) and transferred to ice (2 min) before the addition of SOC media (500 µL) at room temperature; cells were recovered with shaking (37 °C, 250 rpm, 2 h) before 100 µL of the cell suspension was plated onto pre-warmed LB agar containing ampicillin (antibiotic concentrations as given, Section 2.2). Single colonies were picked, re-plated, and colony PCR performed (Table 2.3) to verify gene insertion.

Table 2.3 Reagent mix for PCR steps

Components [a]	gDNA amplification	colony PCR
10x Buffer	5 µL	5 µL
MgCl$_2$ (25 mM)	2 µL	2 µL
dNTPs (10 mM)	1 µL	0.5 µL
MilliQ water	26.8 µL	15.3 µL
Taq polymerase	0.2 µL	0.2 µL
forward primer (5 µM µL^{-1}) [b]	5 µL	1 µL
reverse primer (5 µM µL^{-1}) [b]	5 µL	1 µL
template DNA (20 ng µL^{-1})	5 µL	Scraping from plate

[a] All reagents provided by QIAGEN (Taq Polymerase Kit)
[b] Gene specific primer used (forward) and vector specific used (reverse) to verify gene insertion

2.3.4 Establishment of glycerol stocks

Colonies containing inserted genes were grown overnight (37 °C, 250 rpm) in LB (10 mL supplemented with ampicillin, 50 mL tube) prior to plasmid extraction. Glycerol stocks were made by pelleting 5 mL of this culture (10 min, 2200 g) before re-suspending in M9 salts (750 μL) and glycerol (50 % v/v, 750 μL). Plasmids were purified by kit (Qiagen) from the culture remaining and sequenced (Macrogen, Seoul, Republic of Korea) to verify the integrity of the inserted gene.

Following sequencing, purified plasmids were transformed into *E. coli* Rosetta2 pLysS (Novagen) cells as described (Section 2.2.3). Cultures were plated on LB agar supplemented with ampicillin and chloramphenicol for positive selection. All colonies were assumed to contain successful transformants. Selected colonies were picked and grown overnight in LB (10 mL, supplemented with ampicillin and chloramphenicol) for the production of glycerol stocks.

2.4 Recombinant Expression and Isolation

2.4.1 Expression of target protein

Starter cultures were grown overnight (50 mL tube, 37 °C, 250 rpm) in LB (10 mL, supplemented with ampicillin and chloramphenicol). These were used to inoculate ZYP-rich media (500 mL with ampicillin and chloramphenicol in 2 L baffled Nalgene flasks) and shaken (25 °C, 250 rpm) for 24 h. Cells were harvested by centrifugation (40 min, 5,000 g). Cells were subjected to a freeze-thaw cycle, and then sonicated to effect complete cell lysis (S-2500 Branson Digital Sonifier, ampl. 60 %, 10 s on, 10 s off, 6 cycles). Supernatant containing the soluble protein fraction was centrifuged (40 min, 11,000 g), filtered on ice (0.2 μm syringe filter), and collected for purification.

Table 2.4 Primer sequences used for cloning procedures

ID	Source	Gene ID[a]	Forward primer (5'-3')[b]	Reverse primer (5'-3')[b]
SP1	BIOS-U3-1	BIOS-U3-1_00154	GCGCGGCAGCCAT**ATG**cccgagggcacactc	GTTAGCAGCCGGATC**CTC**tcagggccaacgagc
SP2	BL107	BL107_06034	GCGCGGCAGCCAT**ATG**ccagagttgaatttgtgg	GTTAGCAGCCGGATC**CTC**ttaagaattgcgttgac
SP3	CC9311	CC9311_00114	GCGCGGCAGCCAT**ATG**ccagaaggagctctc	GTTAGCAGCCGGATC**CTC**ttagggccatcgcgcc
SP4	CC9605	CC9605_00108	GCGCGGCAGCCAT**ATG**gagttgtggaccctg	GTTAGCAGCCGGATC**CTC**ttaagaattgccggcg
SP5	CC9902	CC9902_00152	GCGCGGCAGCCAT**ATG**ccagaactgaatttgtgg	GTTAGCAGCCGGATC**CTC**ttaagaattacttccatc
SP6	MITS9220	MITS9220_00121	GCGCGGCAGCCAT**ATG**cctgagggaactctc	GTTAGCAGCCGGATC**CTC**tcagggccatcgggc
SP7	WH8102	WH8102_00125	GCGCGGCAGCCAT**ATG**cagcggtctcttgag	GTTAGCAGCCGGATC**CTC**ttagggccagcgagcc
SP8	WH8109	WH8109_00131	GCGCGGCAGCCAT**ATG**gagttgtggaccctg	GTTAGCAGCCGGATC**CTC**ttaagaattgccgacaacctc
SP9	BIOS-U3-1	BIOS-U3-1_01586	GCGCGGCAGCCAT**ATG**gaccatgtctcgatattgatg	GTTAGCAGCCGGATC**CTC**tcagggcgctgctcc
SP10	BL107	BL107_12765	GCGCGGCAGCCAT**ATG**ccagtctcaattttgatgcctg	GTTAGCAGCCGGATC**CTC**tcatgcttcggcccc
SP11	CC9311	CC9311_01402	GCGCGGCAGCCAT**ATG**gtgattgaatcagtatctatattgatgc	GTTAGCAGCCGGATC**CTC**tcatggttcaactcctg
SP12	CC9605	CC9605_01408	GCGCGGCAGCCAT**ATG**gaggaggtgagcatc	GTTAGCAGCCGGATC**CTC**tcatggcgtggctcc
SP13	CC9902	CC9902_01078	GCGCGGCAGCCAT**ATG**cccgtctcaattttgatgcc	GTTAGCAGCCGGATC**CTC**tcatgcttctgcccctg
SP14	MITS9220	MITS9220_01409	GCGCGGCAGCCAT**ATG**ccagaccatatctcaatc	GTTAGCAGCCGGATC**CTC**ttagggagctgctcctg
SP15	WH8102	WH8102_01282	GCGCGGCAGCCAT**ATG**cagcagcagatgcag	GTTAGCAGCCGGATC**CTC**tcatgccttgcccccag
SP16	WH8109	WH8109_01255	GCGCGGCAGCCAT**ATG**gaggaggtggtgag	GTTAGCAGCCGGATC**CTC**tcatggcgtggctccc
SP17	BIOS-U3-1	BIOS-U3-1_00438	GCGCGGCAGCCAT**ATG**ggccaaagaggtgtg	GTTAGCAGCCGGATC**CTC**tcagggtgccgaaatc
SP18	MITS9220	MITS9220_00388	GCGCGGCAGCCAT**ATG**ggcaagggaggtgtg	GTTAGCAGCCGGATC**CTC**tcaagttgccgatatgtcac
SP19	WH8102	WH8102_01917	GCGCGGCAGCCAT**ATG**cctgaagccgacacatccg	GTTAGCAGCCGGATC**CTC**ctagatctctccggttg

[a] Gene ID as designated in Cyanorak database
[b] Start and stop codons are indicated in bold. Vector-specific sequences (pET-15b) are in uppercase, gene-specific sequences are in lowercase.
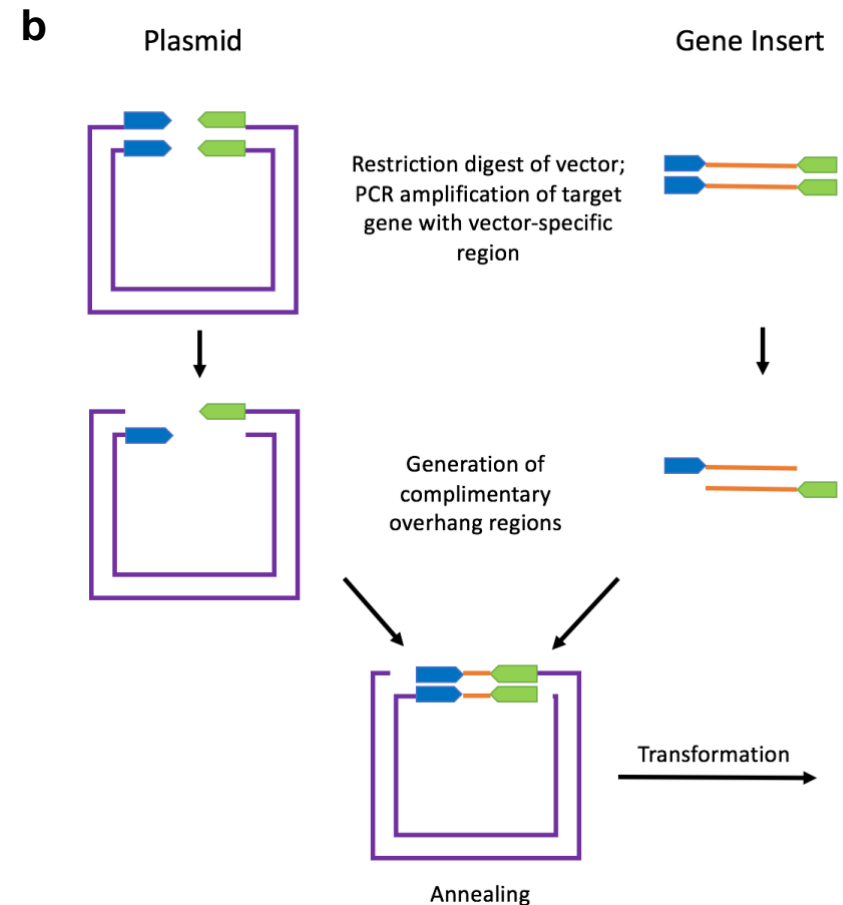
a

pET-15b
(5708bp)

Bpu1102 I(267)
BamH I(319)
Xho I(324)
Nde I(331)
Nco I(389)
Xba I(428)
Bgl II(494)
SgrA I(535)
Sph I(691)
EcoN I(751)
EcoR I(5706)
Cla I(24)
Hind III(29)
Aat II(5635)
Ssp I(5517)
Sca I(5193)
Pvu I(5083)
Pst I(4958)
Bsa I(4774)
Ahd I(4713)
Ap (4643-5500)
AlwN I(4236)
ori (3882)
BspLU11 I(3820)
Sap I(3704)
Bst1107 I(3591)
Acc I(3590)
BsaA I(3572)
Tth111 I(3565)
Bpu10 I(2926)
Msc I(2791)
Bsm I(2704)
BspM I(2399)
Nru I(2319)
Eag I(2284)
PshA I(2061)
lacI (866-1945)
Mlu I(1216)
Bcl I(1230)
BstE II(1397)
Apa I(1427)
BssH II(1627)
Hpa I(1722)

Bgl II
AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGA

T7 promoter primer #69348-3
T7 promoter
lac operator
Xba I
rbs

Nco I
TATACCATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCTCGAGGATCCGGCTGCTAACAAAGCCCGA
MetGlySerSerHisHisHisHisHisHisSerSerGlyLeuValProArgGlySerHisMetLeuGluAspProAlaAlaAsnLysAlaArg
His•Tag
Nde I
Xho I
BamH I
thrombin

Bpu1102 I
AAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG
LysGluAlaGluLeuAlaAlaAlaThrAlaGluGlnEnd
T7 terminator
T7 terminator primer #69337-3

pET-15b cloning/expression region

b

Plasmid   Gene Insert

Restriction digest of vector;
PCR amplification of target
gene with vector-specific
region

Generation of
complimentary
overhang regions
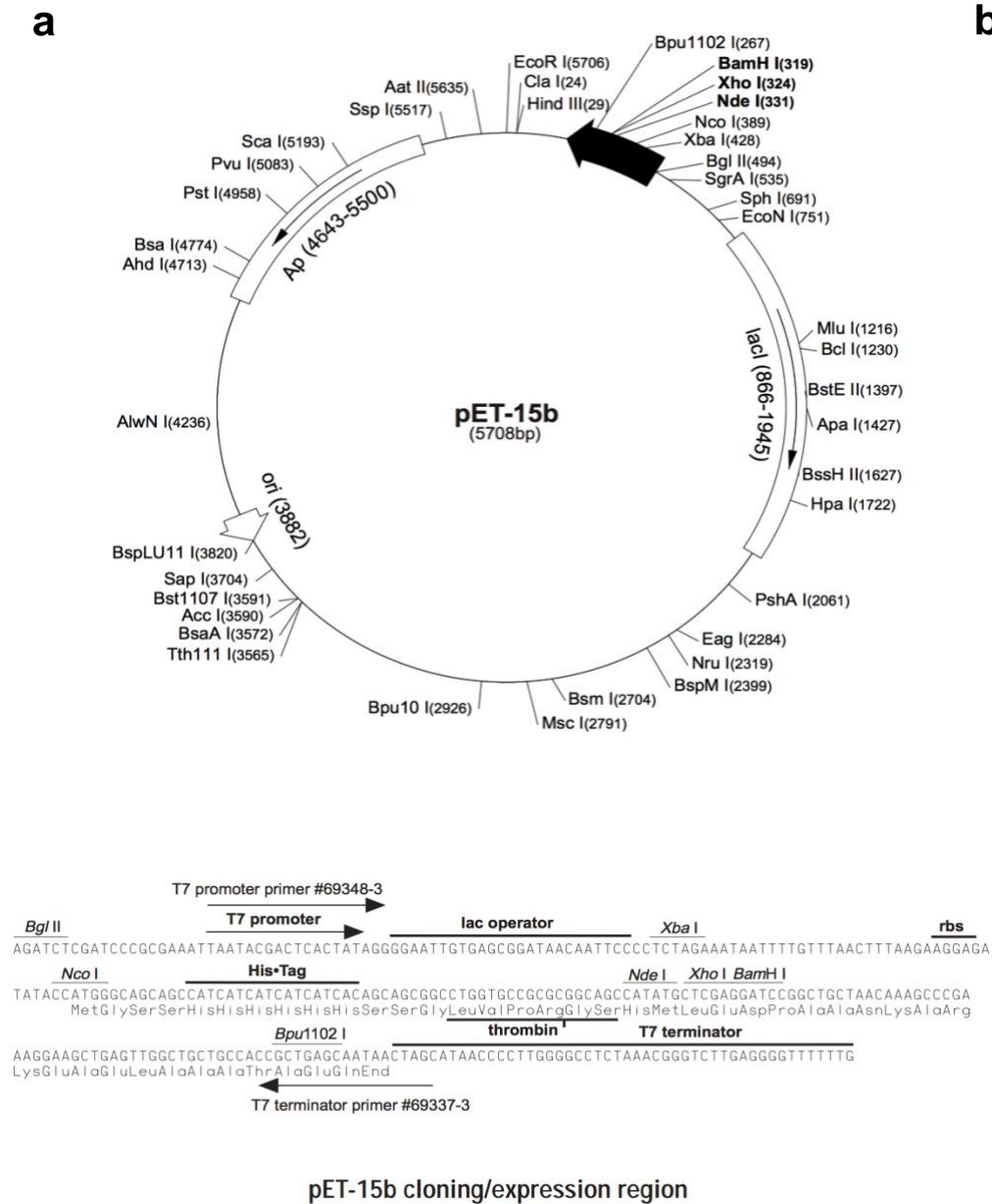
Transformation

Annealing

Fig 2.1 Ligation independent cloning of target genes. **(a)** the pET-15b vector (Novagen) utilised for gene expression with cloning site indicated and **(b)** a stylised representation of the LIC process [61].

19

2.4.2 Purification of recombinant products by IMAC

Pre-packed columns of Ni-Sepharose immobilised metal affinity chromatography (IMAC) media (His-Trap 1 mL, GE Healthcare) were washed with 10 column volumes (cv) each of water and Buffer A containing 5 mM imidazole. Filtered cell lysate (30 mL aliquots) was loaded under pressure provided by a bench-top peristaltic pump (1 mL min^{-1}). Following sample loading, each column was washed with 10 cv Buffer A containing 5 mM imidazole, then Buffer A with 75 mM imidazole (> 45 cv) to remove non-specifically adsorbed proteins.

Captured His-tagged (His$_6$) targets were next eluted from the IMAC columns in Buffer A containing 500 mM imidazole using an LC system (Äkta Start, GE-Healthcare) operating at 0.5 mL min^{-1}. Fractions (0.5 mL) recording an absorbance reading at 280 nm (A$_{280}$) significantly above baseline were pooled. Buffer exchange was carried out immediately using size exclusion chromatography (SEC). A Superdex 200 (HiLoad 16/600 column, GE-Healthcare) equilibrated in Buffer A operating at 1 mL min^{-1} on a LC system (Äkta Pure, GE-Healthcare) was loaded with pooled eluent from IMAC purification. Fractions (1 mL) with A$_{280}$ values significantly above baseline were collected and pooled before concentrating to ≥ 1 mg mL^{-1} in a spin device (Vivaspin spin 20kDA cutoff, GE-Healthcare). Purified protein was snap-frozen in thin-walled PCR tubes using liquid nitrogen and stored (-80 °C) for further analysis.

2.5 Protein Analysis

2.5.1 Size exclusion chromatography

SEC is an entropically-controlled method of separation based on the size, or hydrodynamic volume, occupied by a macromolecule in solution relative to the pore size of the matrix [62]. The evaluation of native protein mass within solution was carried out using analytical SEC procedures on a Superdex 200 10/300 GL column (GE-Healthcare) equilibrated in Buffer A.

Calibration of elution times was carried out using commercial size standards (GE-Healthcare): ferritin (440 kDa); aldolase (158 kDa); ovalbumin (43 kDa); RNase I (13.7 kDa); conalbumin (75 kDa); and carbonic anhydrase (26 kDa). Void volume (V$_o$) was estimated using blue dextran. Fig. 2.2a shows the relationship observed between the elution volume (V$_e$) of each species on utilized column and their associated distribution coefficients (K$_{av}$). A calibration plot of K$_{av}$ and the logarithm of the molecular radius (M$_R$) within solution, logM$_R$, values yielded a straight line (Fig. 2.2b), allowing the interpolation of native M$_R$ for unknown protein samples. Values for K$_{av}$ were calculated using the following formula:

$$K_{av} = \frac{V_e - V_O}{V_t - V_O}$$

where V_e represents the elution volume of a species of interest, V_O represents the empirically-determined void volume of the column (8.23 mL), and V_t represents the total column volume (24 mL)[63].
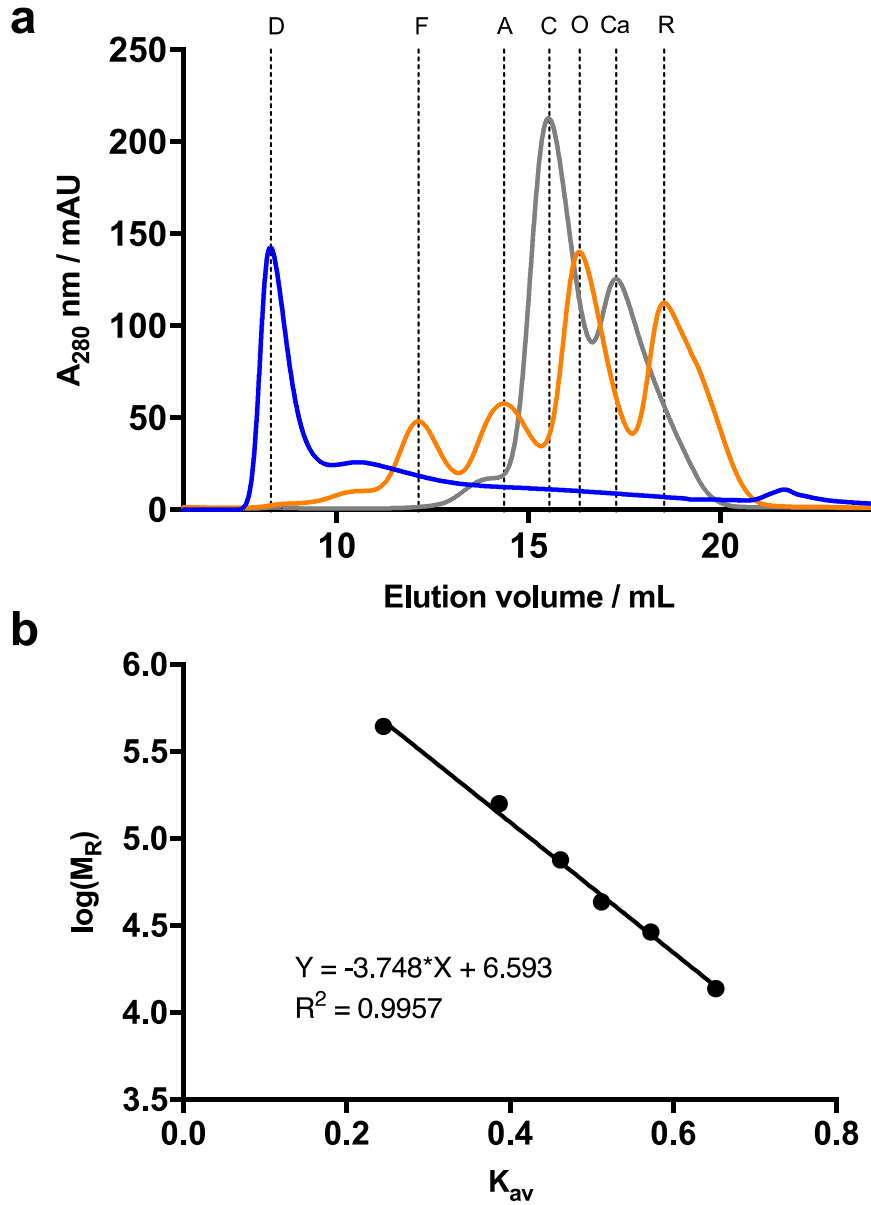

Fig 2.2 Calibration of analytical SEC. **(a)** Overlay of standard traces used to determine retention volumes for proteins of known M_R. From left to right: dextran (8.33 mL), ferritin (440 kDa), aldolase (158 kDa), conalbuin (75 kDa), ovalbumin (43 kDa), carbonic anhydrase (29 kDa), RNaseI (13.7 kDa). **(b)** Calibration plot with derived K_{av} vs. logarithm of M_R. The equation for the line of best fit is given as *y = -3.748x + 6.593*. The correlation co-efficient (R^2) of 0.9957 indicates a strong linear relationship between the two variables.

2.5.2 Protein electrophoresis

Purified proteins were subjected to SDS-PAGE analysis [64] (15 % separating and 5 % stacking gel) using a Tris/glycine buffer system and reagents listed in Table 2.4. Samples (25 μL aliquots) were diluted twofold in loading dye and boiled for 10 min prior to loading. Electrophoresis was carried out by first focusing the samples (100 V, 10 min) followed by separating (120 V, 1 h) in running buffer. Gels were then fixed (10 min), stained (7 min), and then destained until the background was clear. A commercial protein ladder (Novex Sharp Pre-stained, Invitrogen) allowed estimation of protein size on each gel.

Table 2.4 Reagents used in SDS-PAGE procedures

Reagent	Composition
Loading dye	SDS (4 % w/v), glycerol (20 % v/v), DTT (200 mM), Tris buffer (100 mM pH 6.8), bromophenol blue (0.2 % w/v)
Running buffer	Tris (25 mM), glycine (250 mM), SDS (10 % w/v)
Fixer	ethanol (50 % v/v), acetic acid (10 % v/v)
Stain	Coomassie Brilliant Blue™ (0.25 % w/v), ethanol (10 % v/v), acetic acid (10 % v/v)
Destain	acetic acid (10 % v/v)

2.5.3 Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) was used to monitor protein thermal stability and its response to discrete additives [65]. In this technique, a protein solution is heated in the presence of a fluorescent dye (SYPRO Orange, Invitrogen). As thermal denaturation occurs across a temperature gradient, increased hydrophobic contacts are made with the external dye causing it to fluoresce. The melting temperature (T_m) is monitored as a change in this observed fluorescence and corresponds to the point at which half of the solution population is found in the denatured state (Fig. 2.3).

DSF reactions were carried out in a 96-well plate format as per published protocols [66]. Protein samples were mixed with dye (50x, 2 μL) to a final concentration of at least 2 μg mL^{-1}. The protein-dye mix was delivered in 2 μL aliquots into a skirted 96-well PCR plate (Stratagene) and the reaction volume was made up to 20 μL with addition of Buffer A. All samples were prepared in triplicate. Following centrifugation (1 min, 1000 rpm) to ensure sample homogeneity, the plate was heated (25 °C - 95 °C) at 1 °C min^{-1} in a real-time qPCR instrument (Mx3005P, Sratagene). Excitation was induced at 485 nm, with the emission profile at 590 nm being measured.

For screening of small molecule cocktails to identify potential protein binding partners, aliquots from the Silver Bullets screen (Hampton Research) were diluted 1:10, and 10 μL added to each well containing the protein-dye mix as previously described. The volume of Buffer A added was adjusted such that the final reaction volume remained 20 μL. Observed changes in T_m were subsequently validated by repeating with protein-ligand combinations utilising individual components. All measurements were performed in triplicate.
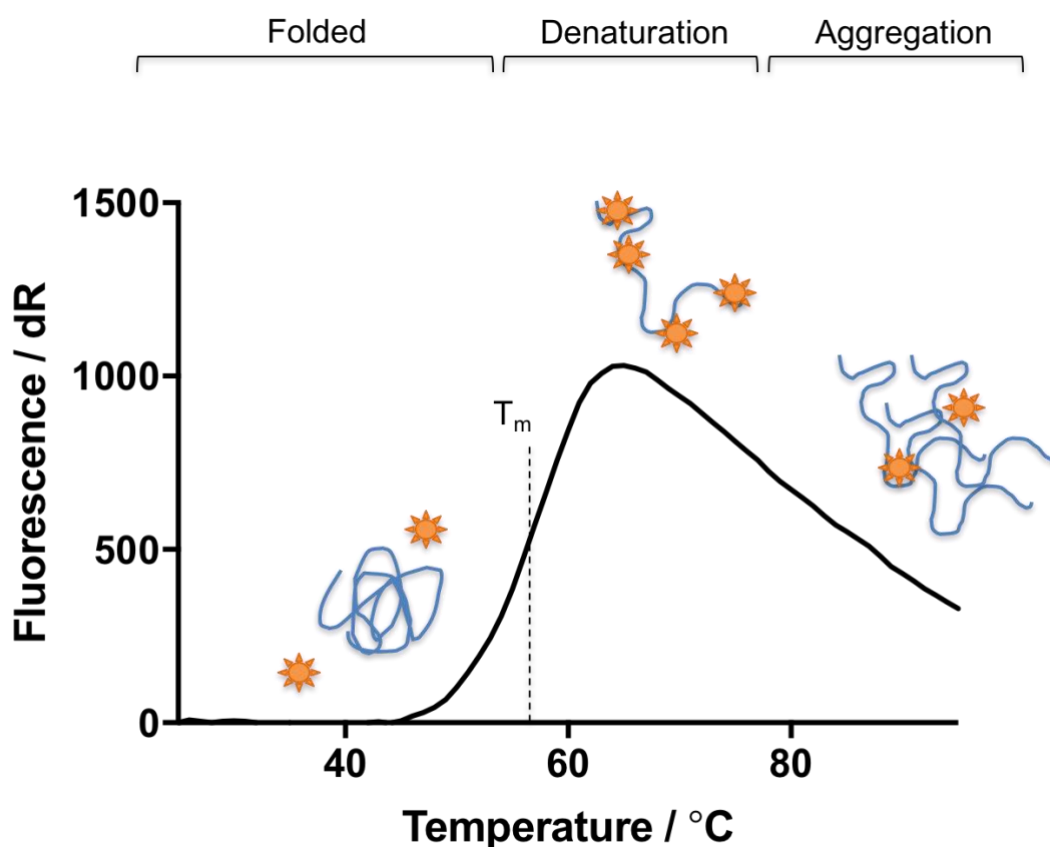
Figure 2.3 Typical DSF response curve obtained for target named SP6. A melting curve corresponding to a single denaturation event is shown. In this technique, a folded protein (blue) is heated in the presence of a fluorescent dye (orange) until denaturation (maxima). Aggregation sequesters hydrophobic residues, resulting in a reduction in observed fluorescence. The T_m value is indicated by dashed vertical at asymptote.

Chapter 3: Optimising Recombinant Production of SBPs

The availability of over 50 sequenced genomes for different *Synechococcus* ecotypes across a variety of global oceans [14, 15] provides a large dataset for understanding the lifestyle and biochemistry of this genus [23]. The validity of specific gene annotations requires more robust investigation, especially for elements evolved to respond to a wide variety of signals or substrates in diverse niche environments. The SBP family is one such example, a family of proteins displaying an inherently low degree of sequence homology, so casting doubt on biochemical inferences made using phylogenetic relationships alone. My study focuses on pursuing a selection of *Synechococcus* SBPs to explore as test cases for more refined functional characterisation at the protein level, with the advantage of an *in vitro* evaluation.

This will involve recombinant production of a set of *Synechococcus* SBPs in a heterologous *E. coli* expression system. Figure 3.1 outlines the general steps required in a production pipeline approach. High attrition rates are intrinsic to this method, however the large sample sizes that feed the production pipeline mean that robust and amenable targets ('low hanging fruit') are successful, and these are generally highly tractable for further analysis [50, 52].
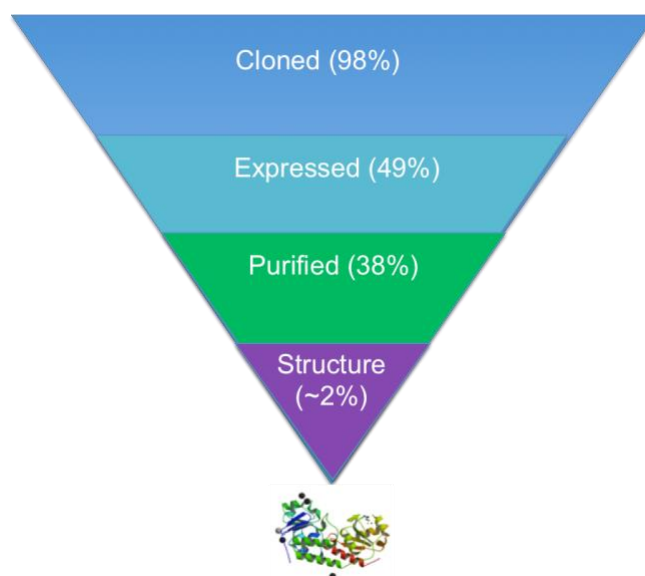

Fig 3.1 Production pipeline and associated attrition rates. Bacterial proteins expressed in recombinant systems experience high attrition rates moving through various stages, with only approximately 2 % amenable to structural analysis. Adapted from [52].

3.1 Selection of SBP Target Panel

Gram-negative bacteria, such as *Synechococcus,* are characterised by their periplasm: a redox-isolated compartment between the inner- and outer-extracellular membranes [67]. *Synechococcus sp.* additionally possess an internal network of thylakoid membranes for both photosynthesis and respiration [10] adding an additional complexity to their cellular biochemistry [68, 69]. In *Synechococus sp.* either the general secretory (*Sec*) pathway, or the *Tat* pathway [67, 69] direct nascent proteins to the appropriate membrane following ribosomal synthesis [68] (Fig. 3.2). How a specific gene product is targeted may depend on the chemistry of the signal peptide [70], chemical environments of the membranes themselves [71, 72] or post-translocational sorting [73, 74].

The essential difference between the two pathways is a difference in cargo conformation. The *Tat* pathway translocates fully-folded protein across the inner membrane [69, 75], contrasting with the *Sec* pathway which translocates protein prior to full folding.

For the purposes of this project, in which SBPs are to be produced as cytosolic products in the bacterial expression host, *E. coli*, direct production of fully-folded products is an important consideration. The presence of signal peptides for export *via* the *Tat* pathway indicates that the formation of a folded, soluble product is favoured in *Synechococcus* cytoplasm, and so the removal of the signal sequence is required for recombinant production of viable folded and bioactive product in a heterologous expression system. The *Tat* signal is identified by the presence of the twin-arginine motif, and in prokaryotes contains the canonical sequence proposed by Berks: S/ TRRXFLK [76], with some deviation from this identified in *Synechococcus* [68].
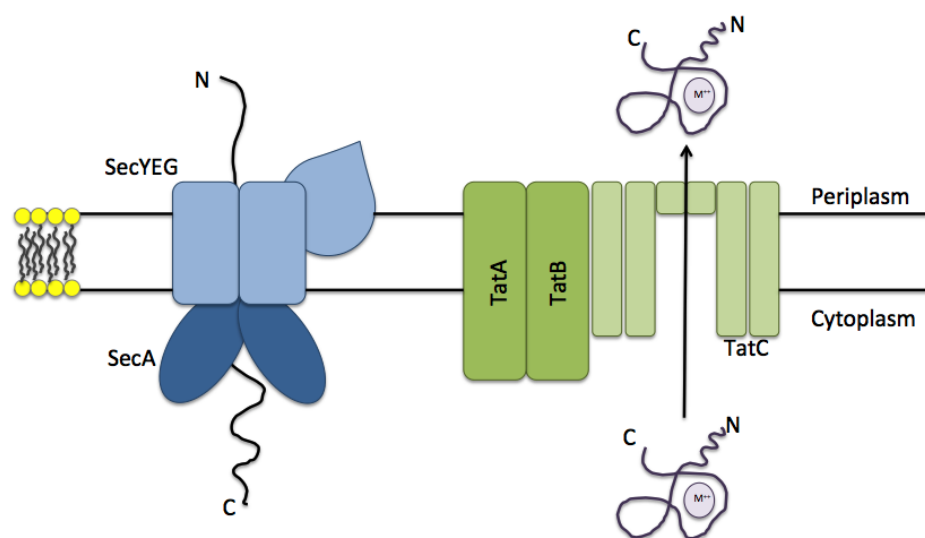

Fig 3.2 Translocation machinery in prokaryotes. The general secretory (Sec) pathway (left) utilises ATP hydrolysis at the SecA subunits to pass a nascent polypeptide through the narrow channel formed by the SecYEG complex. In the twin-arginine (Tat) pathway (right) fully folded proteins (generally with appropriate cofactor) are moved through the porin formed by the TatABC subunits by a proton-motive force. Adapted from [77].

3.1.1 *In silico* assessment of gene targets

In silico analysis of the gene sequence is commonly used in structural genomics to triage a panel of targets of interest to identify and remove non-tractable candidates [51, 52, 78]. Given the high attrition rates inherent in this approach, selecting those targets most amenable to rapid production in *E. coli* as stable, soluble proteins is advantageous. Straightforward algorithms are used to detect globularity/disorder, predict trans-membrane regions and signal peptide sequences (Section 2.2).

Canonically, the SBP fold consists of two globular domains connected by a flexible hinge region, with the structure of this latter region providing defining features according to its subclass [38, 39]. The selected sequences might therefore be expected to show evidence for two globular domains. An exemplar output from the GlobPlot program [53] is shown with the target sequence SP6 in Fig. 3.3. The sequence showed two regions of globularity (residues 1-283 and 300-433) connected by a disordered element. The sixteen residues forming this disordered region were found to be similar across all proteins in the CK_00001342 cluster. Ten targets, including all members of cluster CK_00001342, were identified to contain more than one globular domain (Table 3.1), with the remaining nine showing only one region of globularity. Only two targets were predicted to contain trans-membrane regions that did not arise from the signal peptide. These may reflect a binding pocket for a potential hydrophobic ligand at the hinge region.
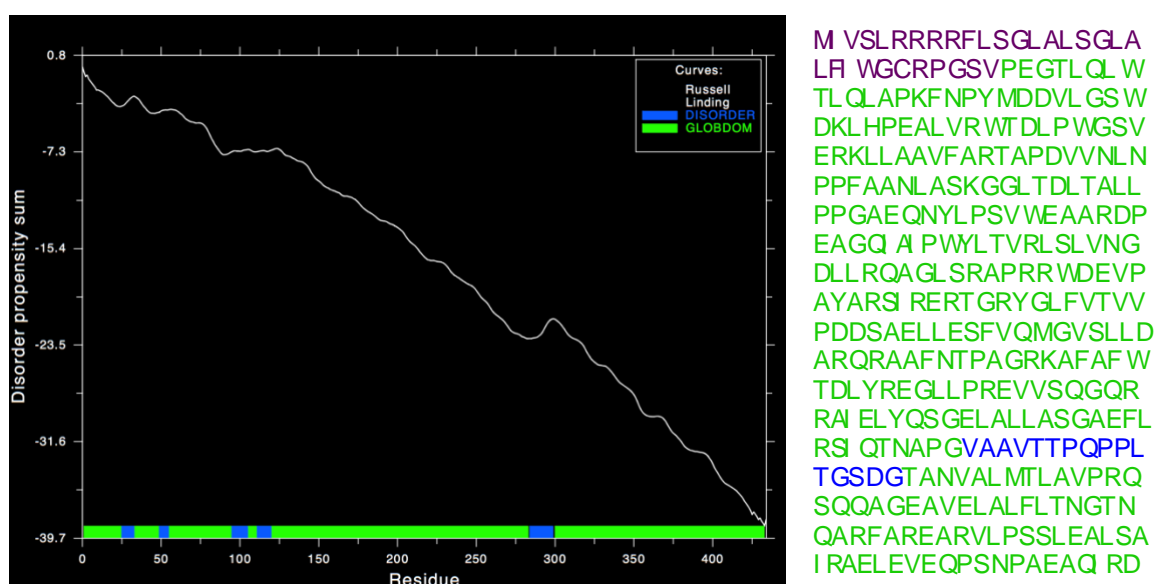

Fig 3.3 GlobPlot output for SBP target named SP6, from *Synechococcus sp. MITS9220*. The GlobPlot analysis indicated the presence of two large globular domains (green) (residues 1-283, and 300-433) with disordered region (blue) and signal peptide sequence (purple).

3.1.2 Signal sequences of *Synechococcus sp.* Targets

Verification of the *Tat* assignment was performed using the TatFIND 1.4 algorithm [79, 80], locating the *Tat*-related motif in eleven of the nineteen proteins (Table 3.1). Only targets within two specific gene clusters (CK_00001342 and CK_00001944) were predicted to possess a *Tat* motif, which contrasted with the manual assignment of *Tat* signal peptides. This represents a major difference between manually- and bioinformatically-predicted sequences, however may be attributable to the variation on the consensus known to occur within *Synechococcus* [68]. The choice of the pET-15b vector required the removal of the signal peptide to ensure incorporation of the polyhistidine tag into the mature protein product to facilitate purification. The summary of all predicted features for the panel of target proteins is presented in Table 3.1.

3.2 Comparison of High-Throughput Protocols

In contrast to genomic and transcriptomic stages of characterisation, which are largely automated and rapidly generate large amounts of data, recombinant expression of proteins still requires more time-consuming manipulation [81]. *E. coli* is the most popular host for the generation of bacterially-sourced recombinant proteins, due to low cost, rapid growth, ease of manipulation, and well-established protocols [81, 82].

Selection of affinity tags or fusion partners are parameters that can have an appreciable impact on the solubility, stability, and recovery of recombinant product [81, 83]. However, trialling of all possible experimental conditions is prohibitive to achieving a reasonable target throughput in production. In the last decade, structural genomics consortia have pioneered standard approaches designed to balance these two parameters. In my project, I have implemented two contrasting approaches to preparing recombinant SBPs.

The first approach utilised the established Macquarie University/Structural Genomics Consortium (MQ/SGC) approach, a protocol successfully used by the Protein Structure Group at Macquarie [52, 84](Section 2.2-2.4).

The second approach, developed by the Oxford Protein Production Facility, United Kingdom (OPPF-UK) [83, 85], is a miniaturised method, designed to allow increased throughput in target screening. To facilitate this, OPPF-UK has developed a suite of vectors (pOPIN) for amplification of target genes using a single set of primers and compatible with a series of cloning vectors. By incorporating a choice of affinity tags and fusion partners, these products allow simultaneous comparison of multiple production conditions. A summary of the differences in scale between the approaches in terms of amplification, cloning, transformation, and screening steps is shown in Fig. 3.4.

Table 3.1 Sequence properties of annotated sugar-binding SBPs

Target	Source and Gene Tag	Cluster [a]	#AA	MW /kDa	pI	GRAVY	SigP length /AA	Tat Motif (N-term)	TM regions	Glob Domains
SP1	BIOS-U3-1_00154		433	47	9.4	-0.08	30	RRRRFL		2-92; 122-281; 298-431
SP2	BL107_06034		436	48	9.8	-0.12	24	TRRRLL		2-88; 112-433
SP3	CC9311_00114		447	49	9.9	-0.25	42	HRRRWL		1-106; 133-295; 313-445
SP4	CC9605_00108	1342	435	47	10.1	-0.08	30	NRRDLL		1-278; 293-431
SP5	CC9902_00152		436	47	9.1	-0.12	24	TRRRLL		2-88; 112-430
SP6	MITS9220_00121		435	47	8.7	-0.09	24	RRRRFL		1-283; 300-433
SP7	WH8102_00125		429	47	9.9	-0.15	26	SRRKLL		1-88; 113-427
SP8	WH8109_00131		435	47	10.2	-0.04	30	TRRDLL		1-278; 293-431
SP9	BIOS-U3-1_01586		413	45	5.4	-0.25	21			1-411
SP10	BL107_12765		433	46	5.1	-0.08	26			1-433
SP11	CC9311_01402		429	47	5.1	-0.05	36		84-106	2-427
SP12	CC9605_01408	1455	421	45	4.6	-0.07	29	not present		1-418
SP13	CC9902_01078		433	47	5.2	-0.13	23			1-433
SP14	MITS9220_01409		412	45	5.1	-0.22	18			2-194; 209-410
SP15	WH8102_01282		420	45	4.9	-0.17	24		73-95	1-204; 221-420
SP16	WH8109_01255		421	45	4.7	-0.10	28			1-418
SP17	BIOS-U3-1_00438		334	36	8.0	-0.07	50	RRRAVL		71-332
SP18	MITS9220_00388	1944	334	36	6.5	-0.09	47	RRRAVL		71-332
SP19	WH8102_01917		321	35	5.5	-0.16	29	RRRSVL		65-331

[a] identifier for an orthologous gene cluster as designated in Cyanorak database
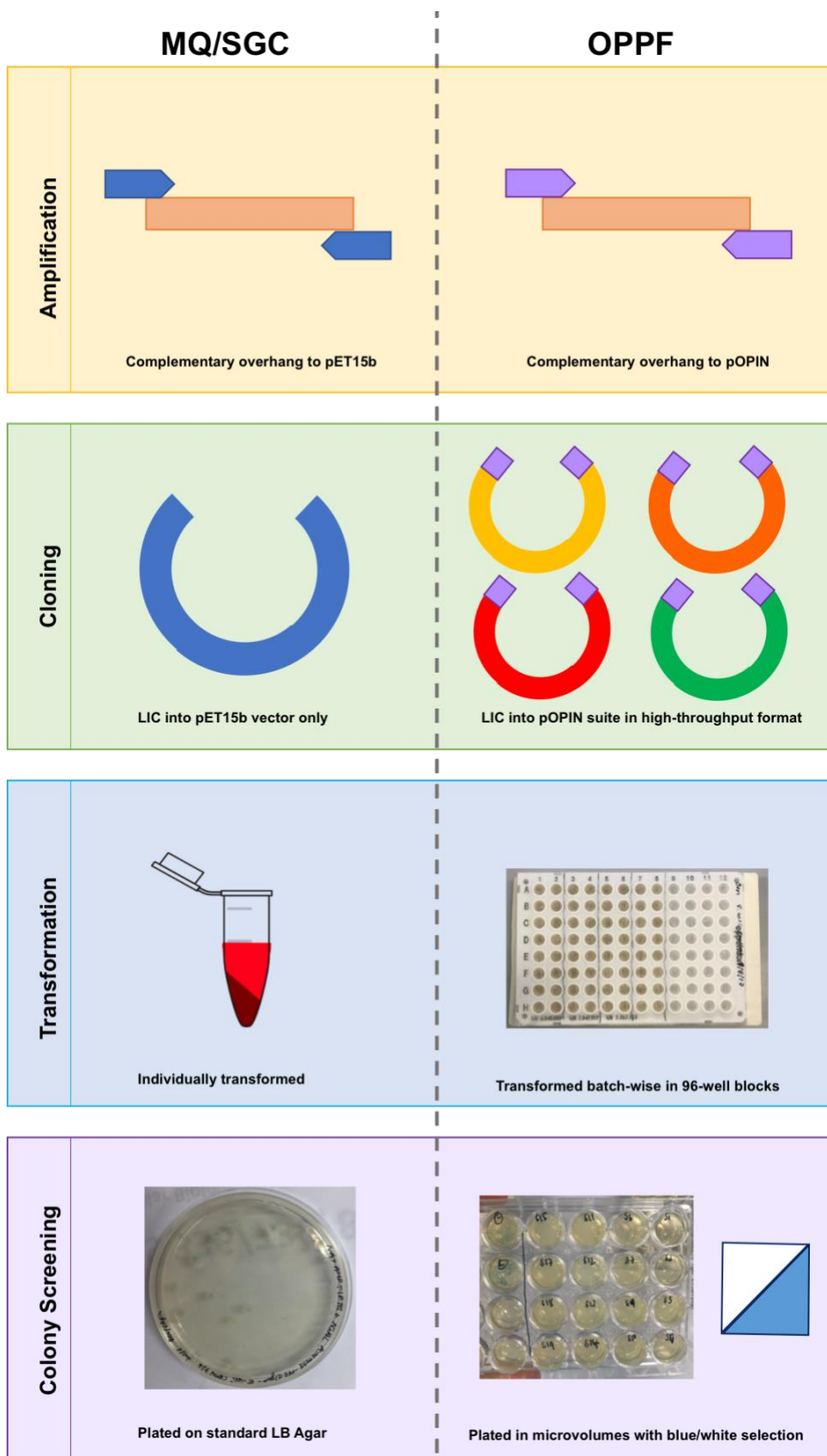
Fig 3.4 Comparison of two high-throughput approaches utilised in this experiment. Left, the MQ/SCG standard approach utilised by the Protein Structure Group at Macquarie University [52], and right the novel OPPF-UK method [85]. Cloning, transformation, and colony screening steps are different between these methods.

3.2.1 Macquarie University/Structural Genomics Consortium approach

I designed primers for gDNA amplification incorporating overhang regions complementary to the pET-15b vector system (Section 2.3.2). Following gDNA amplification, successful gene amplicons were visualised using agarose gel electrophoresis. This showed eighteen of the nineteen targets were successfully amplified, consistent with expected attrition rates (Fig 3.1). SP8 failed to yield a positive amplicon despite increasing extension time (by 1 min), and incorporating a thermal gradient to the extension temperature (± 2.5 °C) over the cycle (Section 2.3.2). This failure to amplify likely indicates a possible error in the original sequencing reaction for this strain, and thus SP8 was discarded from the working panel.

Remaining targets were each inserted into the pET-15b vector through the use of LIC (Section 2.3.3). At the conclusion of cloning and transformation, thirteen successful targets remained in the production pipeline. A notably poor efficiency was observed for this transformation step, with some targets requiring as many as eighteen individual colonies to be screened for a positive amplicon. This is shown for target SP6 in Fig. 3.5, with only one band corresponding to an amplicon detected with agarose gel electrophoresis following eighteen attempts at colony PCR.

Upon sequencing the thirteen successfully cloned targets, ten were carried on for expression trials. The ten targets without mutations were transformed into the expression cell line (Section 2.3.4). A high frequency of deleterious mutations were observed during sequencing and assumed to be as a result of using a low-fidelity polymerase (Taq) for the original amplification of gDNA. As a rescue strategy, the use of a high-fidelity polymerase (Kod) was trialled for gDNA amplification, however, this resulted in no target genes being successfully amplified.
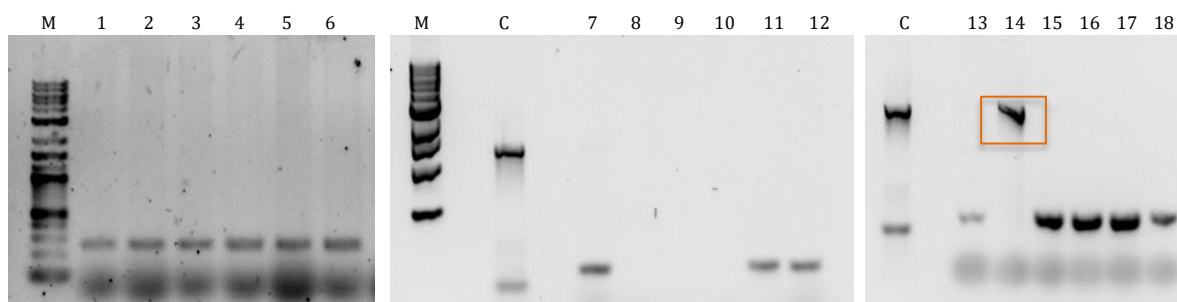

Fig 3.5 Colony PCR for SP6 transformants. Indicated on gels are molecular weight marker (M), amplified gene from gDNA as control (C), and individual colonies (numbered 1-18). Successful colony (#14) is indicated with a box.

For large-scale expression trials, target proteins were grown as described, before cells were harvested and lysed (Section 2.4.1). The fresh lysate was applied to IMAC columns in approximately 30 mL aliquots to saturate the column binding capacity and allow His_6 tagged proteins to outcompete endogenous host contaminants. A gradient elution was performed to identify the optimal imidazole concentration for column washing (Fig. 3.6) to further increase sample purity.
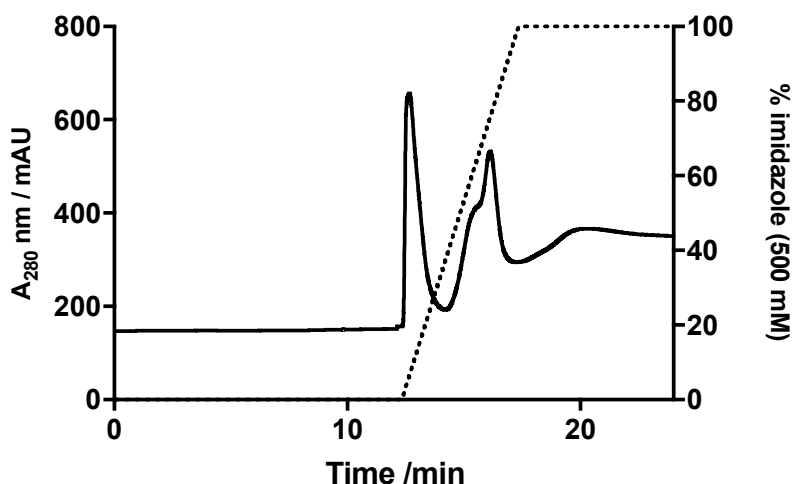

Fig 3.6 Gradient elution profile. Adding imidazole in gradient showed two distinct populations formed. The first population began eluting at approximately 10% addition, corresponding to a total imidazole concentration of ~75 mM, subsequently used for purification of protein (Section 2.4).

Inspection of the IMAC elution profile and SDS-PAGE gels allowed the targets to be evaluated for product yield and purity. As shown in Fig. 3.7, a successful target (SP5) was identified by a single peak in the elution profile, and a large band in SDS-PAGE analysis at the appropriate molecular weight (~42 kDa), with few contaminating proteins from the host-cell line. This contrasted with SP13, an unsuccessful target, which instead showed multiple elution peaks, a small band at the appropriate molecular weight (~40 kDa), and many contaminant proteins, indicating low expression yield. Large-scale expression screening resulted in three proteins being successfully expressed in high yield (typically in quantities > 2 mg mL^{-1}) and purified to homogeneity using this approach. A summary of target progress using this approach is shown in Fig. 3.8.
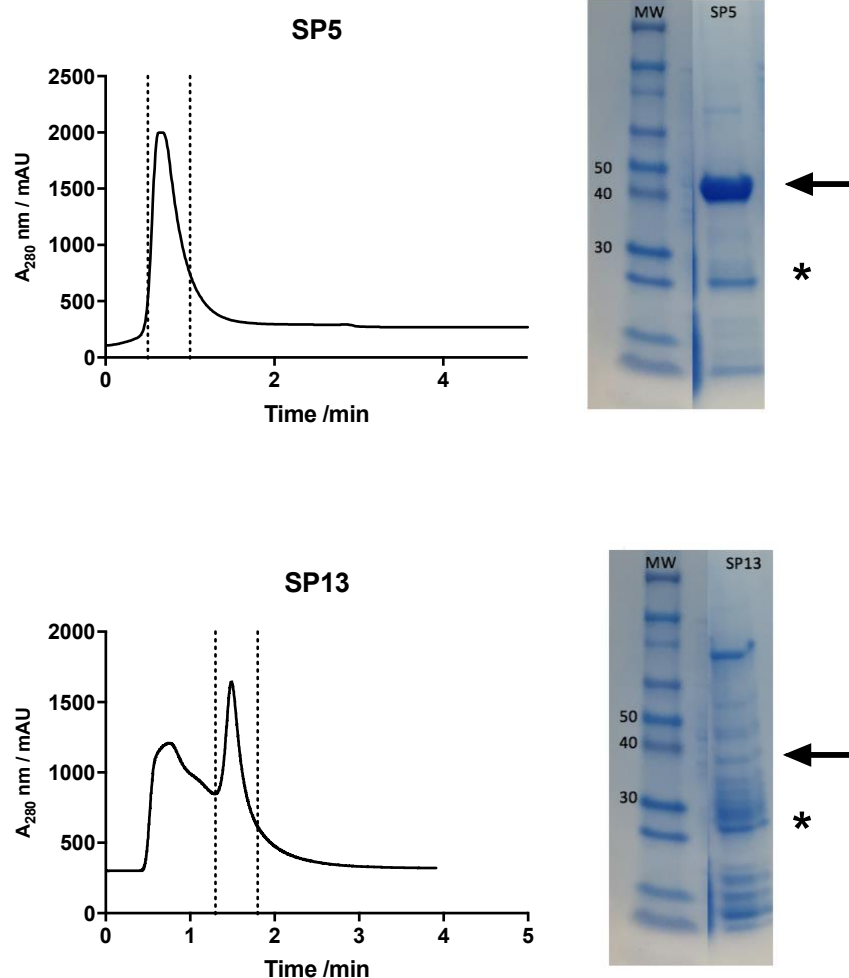
Fig 3.7 Purification of recombinant proteins SP5 and SP13. IMAC elution profiles are shown for two targets (SP5 and SP13). Elution from 1 mL His-Trap Ni-Sepharose columns was performed on an Äkta Start LC system operating at 0.5 mL min^{-1}. Elution began with addition of 500 mM imidazole. Fractions (0.5 mL) collected for SDS-PAGE analysis indicated with dashed line. SDS-PAGE for each target shown below corresponding elution profile. Expected target molecular weight indicated with an arrow. SlyD (common contaminant) band indicated with asterisk.
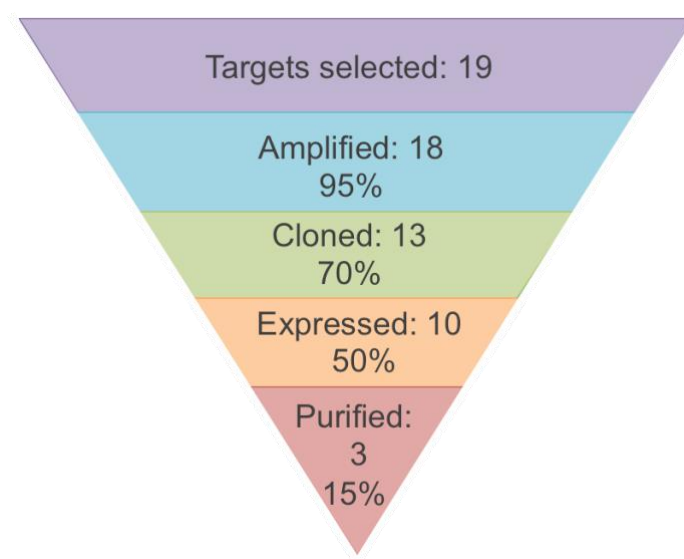

Fig 3.8 Attrition rate of targets using MQ/SGC approach. The final number of proteins purified from the original panel (3/19) is consistent with reported success rates [52].

3.2.2 Oxford Protein Production Facility approach

Primers for amplification of gDNA were designed as described (Section 2.3.2), however the overhang region was instead complementary for the pOPIN vectors [83], containing the sequences: 5'-AAGTTCTGTTTCAGGGCCCGATG-3' for the forward primer, and 5'-ATGGTCTAGAAAGCTTTA-3 for the reverse primer.

PCR amplification and visualisation of the target genes was performed as described (Section 2.3.2), resulting in fifteen successful amplicons. This was a smaller number than observed in the MQ/SGC approach, suggesting the use of primers complementary for a wide range of vectors may represent a trade-off for compatibility with a standard PCR cycle. Both approaches failed to amplify target SP8, further supporting the inference of an original sequencing error for this strain.

The miniaturised cloning format in the OPPF-UK method meant each LIC reaction was scaled down from 10 μL to 3.3 μL, resulting in a higher observed success rate, with all fifteen remaining targets in the pipeline cloned. Transformation was carried out using deep 96-well blocks, with a remarkably high transformation efficiency observed, indicating the reduced scale did not adversely effect this step. While scaling down can introduce false positives if ligation becomes inefficient, the pOPIN vectors incorporate an additional blue/white selection marker using the β-galactosidase enzyme [83, 85], allowing these to be filtered out during colony screening when plated on LB-agar containing X-gal (5-bromo-4-chloro-3-indolyl-D-galactopyranoside) (Fig. 3.9a).
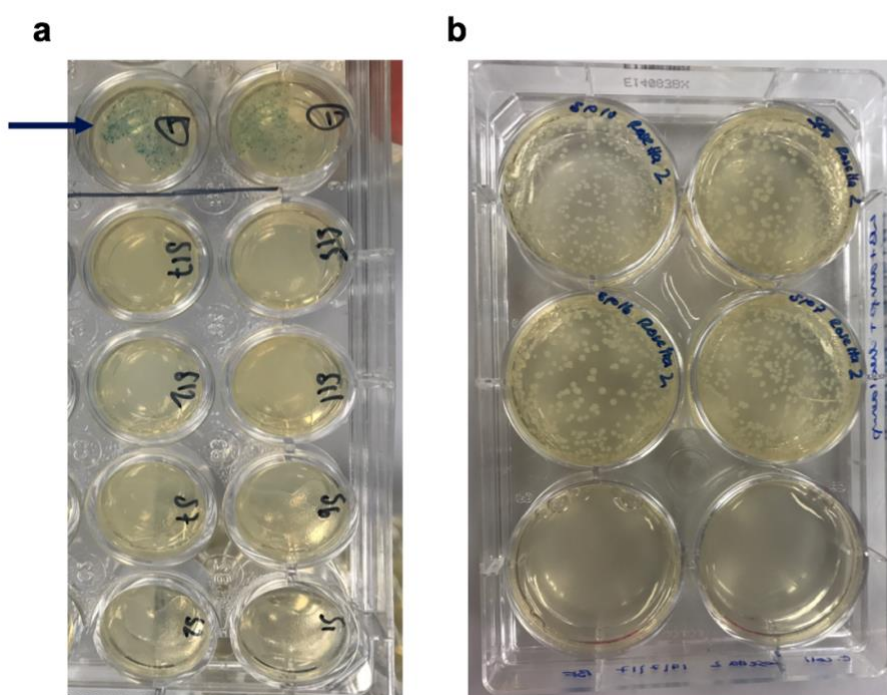

Fig 3.9 Colony screening of targets. **(a)** for OPPF-UK method in 24-well tissue culture plates. Top two wells indicating blue colour of negative control (arrow), and bottom eight wells showing *E. coli* lawns for various targets: SP1, 2, 6, 7, 11, 12, 15 and 17, **(b)** using MQ/SGC approach in 6-well plates showing single colonies.

For colony screening, the published protocol suggests the use of 24-well tissue culture plates is sufficient for isolating single colonies of transformed cells rather than standard agar plates. Even after significant dilution of the transformation reaction (1:500), and plating volumes as small as 2 μL, single colonies could not be isolated in this format, with lawns of *E. coli* predominating (Fig. 3.9a). This was compared with plating the transformation reaction from the standard protocol in a reduced 6-well format (Fig. 3.9b) following the procedure given (Section 2.3.3). The use of smaller 6-well plates produced successful single colonies in some instances. This indicated that with further optimisation, advantageous aspects of the OPPF-UK approach could be incorporated within the established MQ/SGC approach.

Despite the significant utility afforded by the OPPF-UK approach for screening multiple expression conditions and the method of filtering out false positives, the standard MQ/SGC approach led to a greater number of successful targets (Table 3.2). Therefore this approach was selected for use in this project due to its robust, established methodology.

Table 3.2 Comparison of attrition rates during genetic manipulation steps.

Target ID	MQ/SGC [a]		OPPF-UK [b]	
	Amplification	Cloning	Amplification	Cloning [c]
SP1		-		
SP2				
SP3			-	-
SP4		-		
SP5				
SP6			-	-
SP7			-	-
SP8	-	-	-	-
SP9		-		
SP10				
SP11				
SP12		-		
SP13				
SP14				
SP15				
SP16				
SP17		-		
SP18				
SP19				
#Successful	18	13	15	15

[a] MQ/SGC standard protocol used by the Protein Structure Group at Macquarie University [52]
[b] Novel protocol developed by the Oxford Protein Production Facility [85]
[c] single colonies unable to be isolated following cloning when transformed into propagation cell line when plated in 24-well format

Chapter 4: Biophysical Characterisation of Recombinant Products

4.1 Genomic Perspectives of Successful SBPs for Study

Three proteins (named SP5, SP6, and SP19) successfully passed through synthesis and production and were purified to homogeneity in high yield. These proteins, identified as SBPs within genomes of *Synechoccocus sp.* have never previously undergone characterisation. My analysis of these targets consisted of two main avenues: identification of their physical characteristics and validating these against canonical characteristics of the SBP family; and the assessment of their annotated function as saccharide binding partners.

SP5 originates from *Synechococcus CC9902*, an ecotype isolated from the oligotrophic surface waters of the California current in the Pacific Ocean (Section 1.2). Analysis of this isolate by Dufresne *et al.* [16] shows a genome size of 2.23 Mbp, at the smaller range for *Synechococcus sp.*, encoding 2,358 putative genes in one contig. This genome contains a low number of unique genes (< 5 %), with over half of its genome satisfied by the picocyanobacteria core genome (those common to both *Synechococcus* and *Prochlorococcus)* [17]. The GC content of the entire genome (54 %) is not indicative of genome streamlining, in contrast to *Prochlorococcus* in which a high proportion of their genome is also satisfied by the picocyanobacteria core genome [16].

SP6, a gene product from the same gene cluster as SP5 (Table 3.1), originates instead from *Synechococcus MITS9220*, an ecotype from the equatorial Pacific. Its genome size of 2.42 Mbp is in the median range for *Synechococcus sp.*, and is predicted to encode over 3,000 gene products. As yet, this strain has not undergone comparative analysis to assess the relative contributions of core, and accessory genomes, however, the trend for unique genes is loosely correlated with genome size [10, 16], and so it likely possesses a percentage of unique genes comparable to similar isolates ($\sim$ 15 %). Similarly, its GC content (56 %) is not indicative of genome streamlining.

SP19 originates from ecotype *Synechococcus WH8102*, indigenous to the Caribbean Sea: an oligotrophic region of the Tropical Atlantic. Analysis of this isolate by Palenik et al. [17] shows a genome of 2.43 Mbp, encoding 2,583 predicted gene products in one contig. Half of these belong to the picocyanobacteria core genome, with approximately 15 % unique to this strain [16]. It similarly does not appear to have undergone extensive genome streamlining.

Key physicochemical properties of each protein of interest are shown in Table 4.1. These are derived from sequence analysis (Section 3.1). SP19 is 36 % homologous to OpuA, the glycine betaine-binding protein from *Bacillus subtilis*, involved in the import of compatible

solutes under osmotic stress [86]. Validation through the identification of a ligand partner would allow this protein to be designated as OpuA and assigned to structural Class F. Homologues to SP5 and SP6 included the trehalose/maltose binding protein from *Thermococcus*, an SBP experimentally validated to bind two related disaccharides [87], however, the sequence identity below 25 % indicates this is an unreliable assignment.

Table 4.1 Results of sequence searching for successful targets

Target	MW /amu	# Cys	Closest homologue [a]	ID	E-val. [b]
SP5	45848	1	trehalose/maltose binding protein	22 %	6e-5
SP6	45061	0	trehalose/maltose binding protein	21 %	2e-3
SP19	32878	0	glycine betaine binding protein (OpuA)	36 %	2e-35

[a] From UniProtKB/SwissProt database, as at October 2017

[b] E-value obtained from PSI-BLAST tool using NCBI blast (blast.ncbi.nlm.nih.gov/Blast.cgi)

4.2 Analysis of Solution Properties

SEC is a powerful tool by which the compositional heterogeneity, or oligomeric profile, of a protein within solution can be analysed, and allows delineation of discrete oligomeric populations [88, 89]. Preparative SEC showed that each protein was purified as a monodisperse product. Exemplar traces are shown as Fig. 4.1, with each protein displaying a single, well-defined peak corresponding to a single oligomeric state. The absence of product within the void volume indicated no aggregation, showing these proteins to be stable within solution.
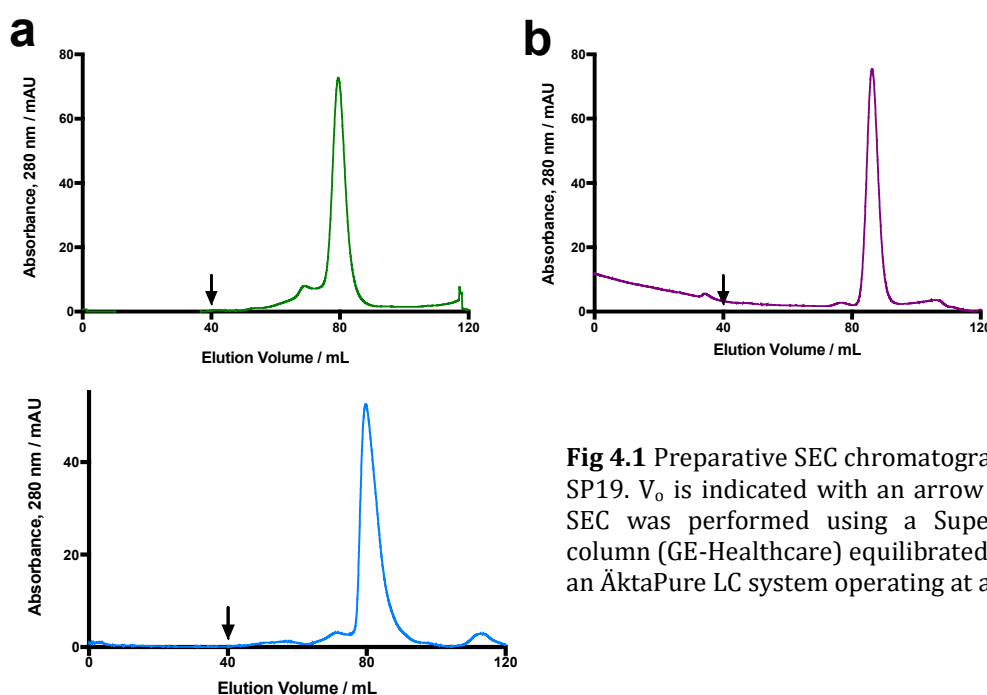

Fig 4.1 Preparative SEC chromatograms for **(a)** SP5, **(b)** SP6, **(c)** SP19. V_o is indicated with an arrow for each trace. Preparative SEC was performed using a Superdex 200 HiLoad 16/600 column (GE-Healthcare) equilibrated in Buffer A (Section 2.2) on an ÄktaPure LC system operating at a flow-rate of 1 mL min^{-1}.

To evaluate the native mass within solution, analytical SEC procedures were carried out as described (Section 2.5.1) and interpolation made from the standard trace obtained. Canonically, in Gram-negative prokaryotes, the substrate-binding protein functions as a monomer [38], thus the native mass determined experimentally was expected to reflect this. The chromatograms for each target are shown as Fig. 4.2 a-c. The observed V_e, derived K_{av} measurements and associated M_R values are summarised in Table 4.2.

For targets SP5 and SP6 (Fig. 4.2a and Fig. 4.2b), the derived M_R approximated the mass calculated based on their primary sequence, indicating these targets were both purified to homogeneity in their monomeric forms. The slight discrepancy is attributable to the expected 10 % error range inherent in analysis by this method [90]. The determination of mass was supported by SDS-PAGE analysis, which showed a single band at the appropriate molecular weight (approximately 45 kDa) (Fig. 4.2d).
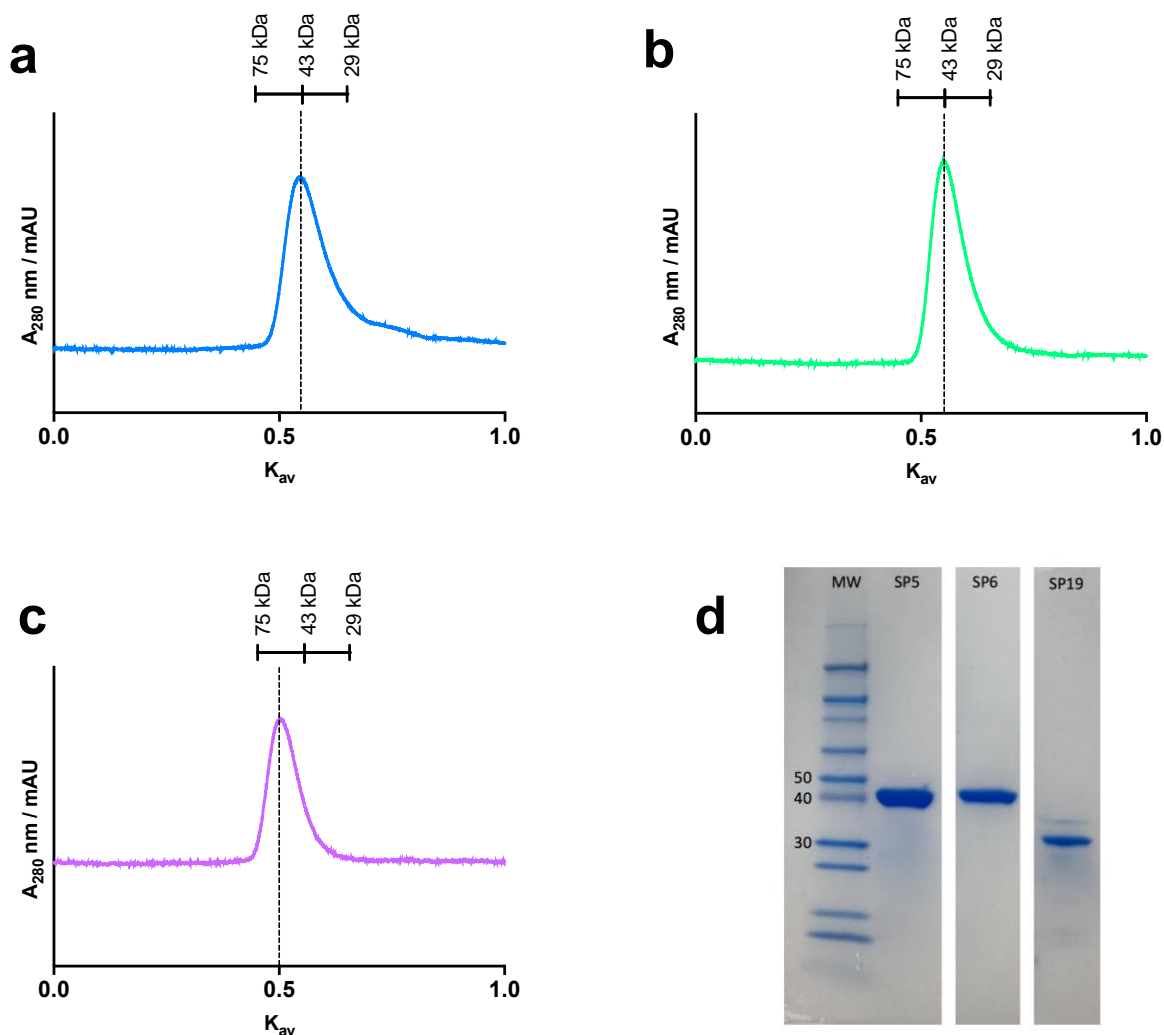

Fig 4.2 Determination of molecular size. The analytical SEC traces are shown for **(a)** SP5, **(b)** SP6, and **(c)** SP19. Peak position is indicated (dashed line). Corresponding M_R of calibration standards indicated (bar). **(d)** SDS-PAGE analysis showing molecular standard masses (MW) and each target (SP5, SP6 and SP19) with each displaying a single band at the appropriate calculated molecular weight.

Table 4.2 Analytical SEC parameters for targets

Target	V_e /mL	K_{av}	M_R /kDa [a]	Mass /kDa [b]
SP5	16.7	0.53	43	45.85
SP6	16.8	0.54	43	44.06
SP19	16.0	0.49	56	32.88

[a] Molecular radius within solution

[b] Molecular mass determined from primary sequence

SP19 exhibited a larger M_R than expected (Fig. 4.2c). The determined native mass of 56 kDa is consistent with that expected for a dimeric protein at pH 7.4. Analysis by SDS-PAGE under reducing conditions showed a single band at approximately 32 kDa, further implicating the influence of quaternary structure. The Gaussian peak shape observed during chromatographic procedures indicated the formation of the dimer was due to normal intermolecular contacts rather than non-specific aggregation. Similarly, the absence of cysteine residues in the primary sequence precludes redox-mediated oligomerisation. This observation represents an unexpected molecular organisation.

This may be due to the presence of a cognate ligand, with the conformational change required for ligand binding facilitating the formation of a dimer interface between two monomers. This has previously been observed for TakP, responsible for binding α-keto acid in *Rhodobacter sphaeroides* that exists as a helix-swapped dimer upon ligand binding [46] (Section 1.3). This is consistent with observations made during purification of this target *via* preparative SEC, which reproducibly showed the presence of a small molecule co-eluting with the protein of interest indicated in Fig. 4.3a.

Protein-ligand-solvent systems are thermodynamic in nature [91] however, despite extensive dilution [92] it is not always possible to remove a bound ligand due to the kinetic constraints of a system [93]. It was thus assumed this peak corresponds to an endogenous small molecule ligand scavenged from the growth media during purification as this peak eluted at the lowest size limit for the column. A crude calculation of molecular weight showed a MW in the region of approximately 150 amu, however this was outside the dynamic range of the column and so cannot be taken as an accurate determination of mass.

Fractions were collected from this peak during subsequent purifications and subjected to analysis with UV-spectroscopy. This analysis showed a strong absorbance band approaching 200 nm (Fig. 4.3b). This may be attributable to conjugated aromatic compounds [94], however was near the spectral limit of the quartz cuvette used. Further confirmation using other approaches such as infrared spectroscopy was unsuccessful due to the dilute sample.

Mechanisms of regulation of ABC transporter activity still remain largely unknown [2]. Reversible dimerisation driven by the presence of a ligand may be a means of exerting control over these transporters by sequestering the functional monomeric form [2, 32].
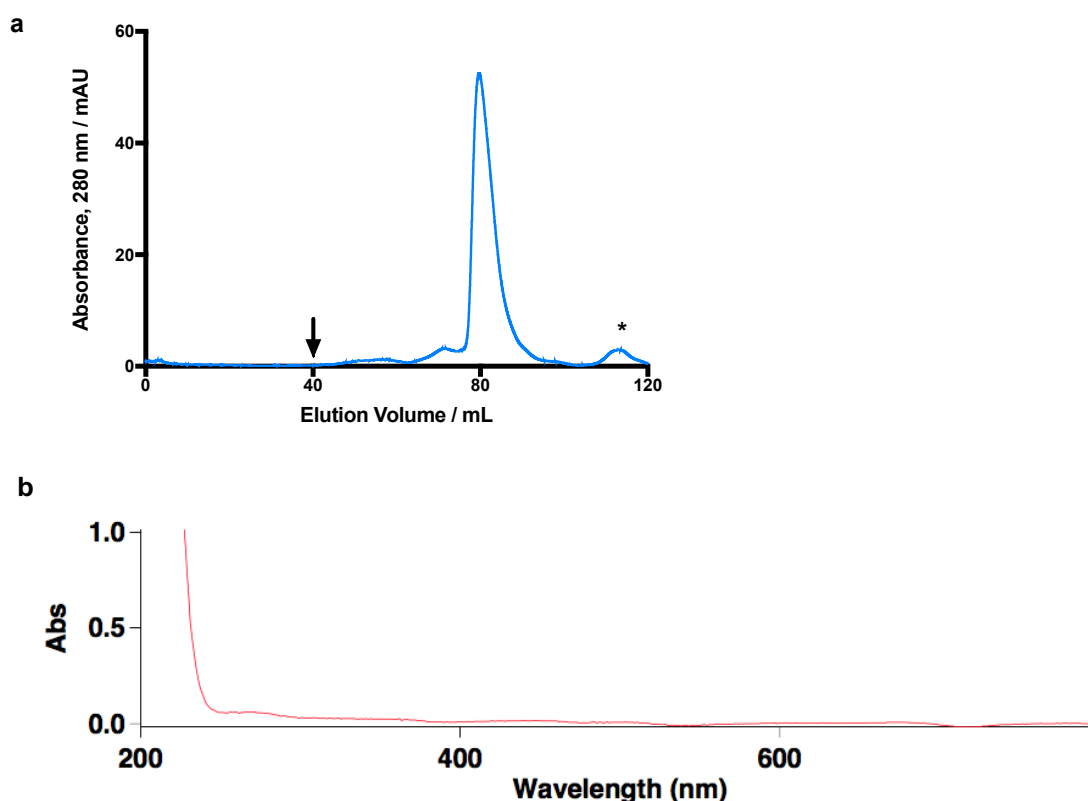

Fig 4.1 Evaluation of putative small molecule. **(a)** shows the preparative SEC for SP19, with V_o indicated by an arrow and proposed small molecule peak (*). **(b)** shows the results of UV-Visible spectroscopic analysis of collected small molecule peak. SEC was collected using a Superdex 200 HiLoad 16/600 column (GE-Healthcare) equilibrated in Buffer A (Section 2.2) on an ÄktaPure LC system operating at a flow-rate of 1 mL min^{-1}.

4.3 Identification of Binding Partners

In order to determine potential binding partners for each target, DSF was used to perform high-throughput screening of small molecule cocktails. This technique evaluates the stability of a protein fold conferred by being bound to its cognate ligand: for proteins in the presence of a cognate ligand, the increased intermolecular contacts stabilise tertiary structure and cause a positive shift in observed melting temperature, T_m [65, 95].

Despite essentially functioning as high-affinity, high-specificity adaptors of molecular import machinery, SBPs also display a certain degree of flexibility for binding substrates chemically similar to their preferred ligand, with such promiscuous binding thought to be a vital part of the acquisition of new substrates and functions over their evolutionary history [96]. The diverse chemistry of small molecules present in crystallisation screens should be sufficient to generate a hit for either a cognate ligand or a moiety chemically similar to the natural ligand.
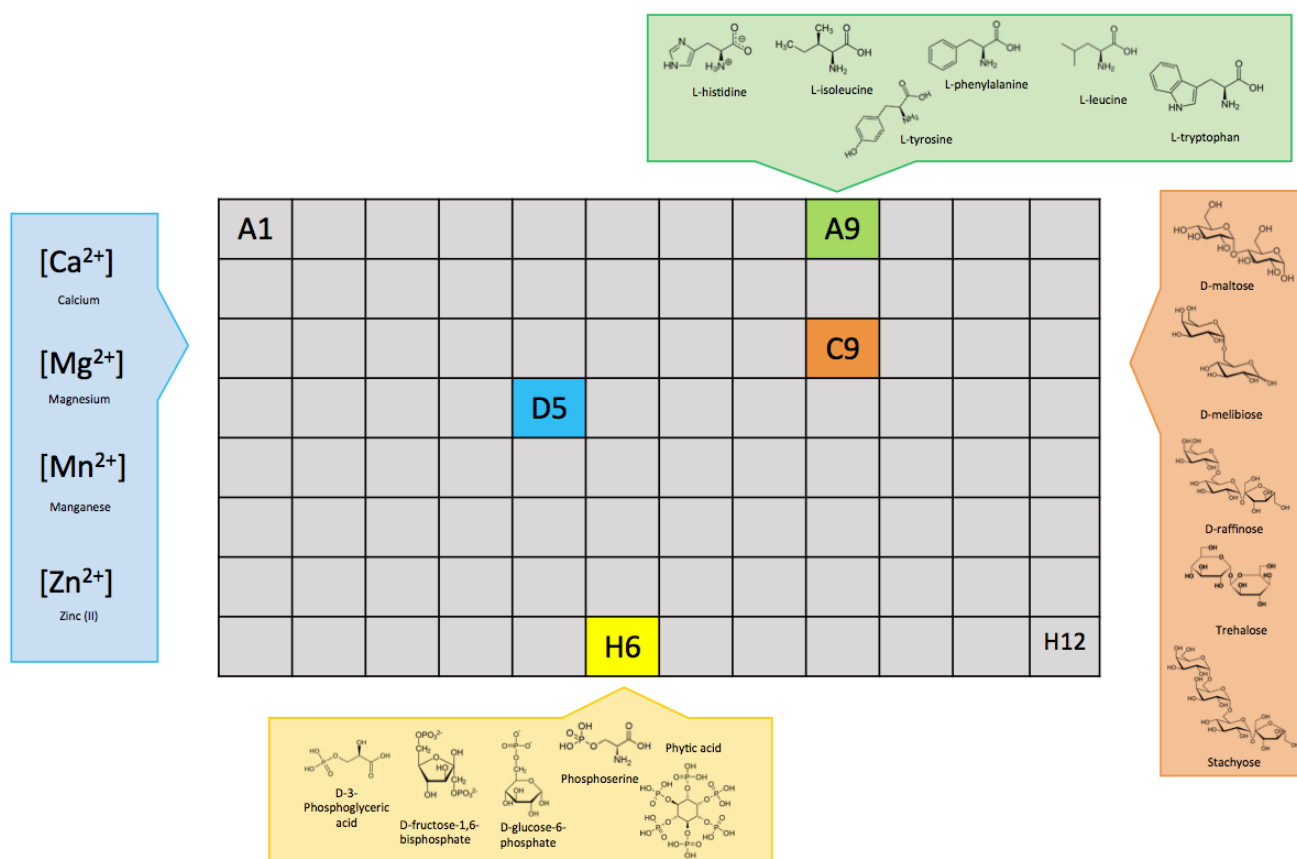
Fig 4.4 Determination of T_m values in in 96-well format. Constituent small molecule cocktails from the Silver Bullets screen [66] are indicated for four wells.

Each protein was pipetted into a 96-well plate with discrete additives from a small molecule cocktail as described (Section 2.5.3). Proteins were subjected to a thermal ramp, allowing indirect measurement of their thermal melting using an extrinsic fluorescent dye. The Silver Bullets screen (Hampton Research) was used as the small molecule screen as it contains 96 discrete conditions with nearly 300 different small molecules such as organic salts and acids, amino acids and peptides, biologically active molecules, and macromolecular digests (Fig 4.4) [66] allowing many potential small molecule binding partners to be screened for each target in a rapid, high-throughput manner..

For SP19, no specific condition resulted in a change in observed T_m from 52 °C. It may be that no condition screened contained a small molecule that interacts with this protein at a thermodynamic timescale amenable to this experiment. Also possible is the scenario that the system remains tightly engaged with a ligand endogenous to the growth media sequestered during recombinant production and purification. Therefore, for this target, identification of the bound ligand may instead require the use of mass spectrometry.
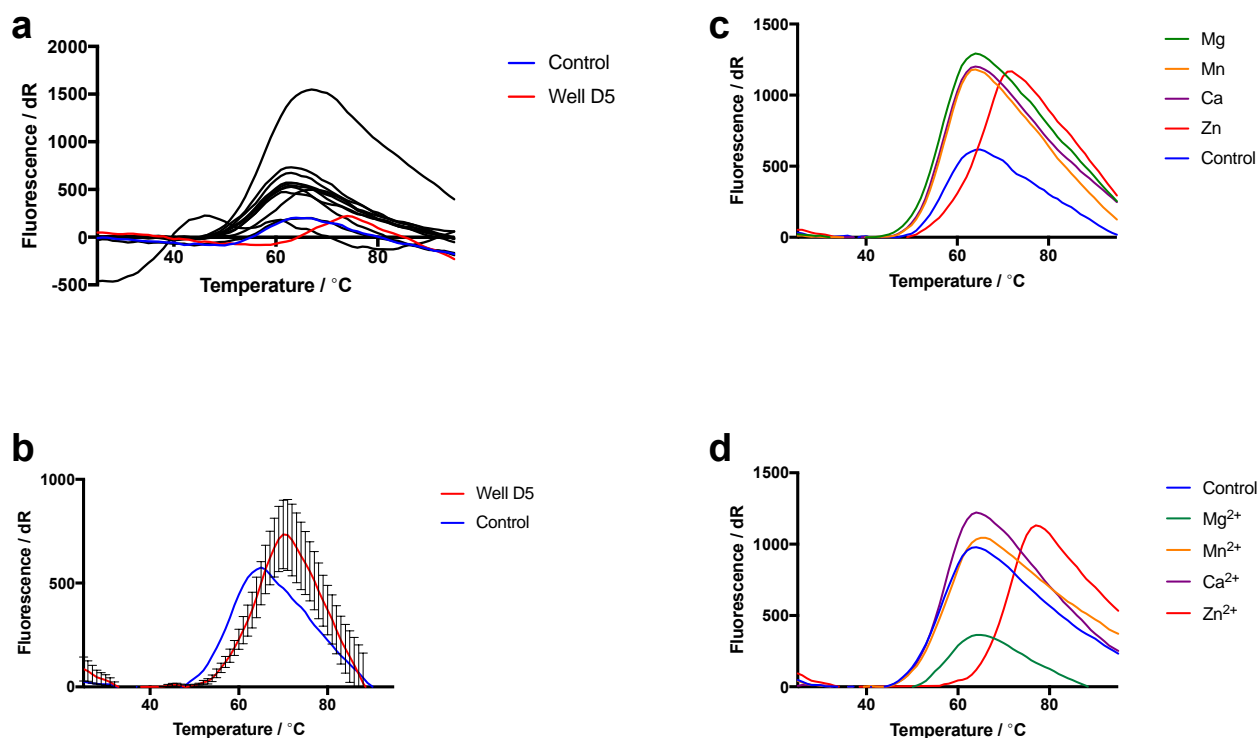
Figure 4.5 DSF analysis of small molecule screening. **(a)** The results of SP5 screened against the D-series wells showed condition D5 (red) produced a shift in observed T_m from the control (blue). **(b)** Repeating this in triplicate showed the change was significant and reproducible. The constituent divalent metal cations were individually screened for **(c)** SP5 and **(d)** SP6 with Zn^{2+} (red) showing a shift in T_m against the control (blue) in both cases.

For both SP5 and SP6, one condition produced a significant change in T_m (ΔT_m). The response for well D% observed for SP5 is shown in Fig. 4.5a (red), in comparison with other control conditions (black), and the T_m for the native protein (blue). These were validated in triplicate, as shown in Fig. 4.5b. The observed ΔT_m of 9 °C for SP5, and 15 °C for SP6, respectively, indicate a highly thermodynamically-favourable interaction with one (or more) of the constituent compounds in this condition composed of the following divalent metal chloride salts: $CaCl_2$, $MgCl_2$, $MnCl_2$, and $ZnCl_2$. No change in T_m was observed for any condition containing either mono- or disaccharides, with these conditions also having representative pentose and hexose members (e.g. well C9 as summarised in Table 4.3) [66].

To ascertain which cation conferred the observed change, the thermal shift was repeated using each constituent cation prepared as per the original screen (Fig. 4.5 c and d). This assay showed that Zn^{2+} was responsible for the observed increase in T_m for both targets, representing a different chemistry compared with the annotated disaccharide binding partner for this particular cluster. Thus, the annotation of these proteins to bind saccharides appears to be incorrect.

Table 4.3 Small molecule screening with DSF

Target	Protein T_m /°C	Cocktails [a]		Individual components (D5) [b]	
		C9 ΔT_m /°C	D5 ΔT_m /°C	Zn^{2+} ΔT_m /°C	Mg^{2+} ΔT_m /°C
SP5	57	+1	+8	+9	+1
SP6	58	+0	+13	+15	+1
SP19	52	+1	+0	-	-

[a] examples of two wells containing small molecule cocktails directly from the Silver Bullets screen. C9: D-(+)-maltose monohydrate, D-(+)-mellobiose monohydrate, D-(+)-raffinose pentahydrate, D-(+)- trehalose dihydrate, stachyose hydrate. D5: magnesium chloride hexahydrate, calcium chloride dihydrate, manganese (II) chloride tetrahydrate, zinc (II) chloride.
[b] Examples of two constituents from well D5 screened individually after an observed ΔT_m.

While engagement of the His-tag has previously been observed in erroneous zinc-binding events for the nickel-binding protein from *Thermatoga* (PDB ID 2ETV), the impact of the His-tag can be conclusively ruled out for SP5 and SP6. No response was observed for either protein in the presence of any other metallic ion, including the higher-affinity Cu^{2+} ion [97]. The His-tag is specifically designed to chelate metallic ions, a property exploited during the purification approach with Ni^{2+}. As no stability to the protein fold was conferred by the presence of Ni^{2+} or any other metallic ion during DSF screening, it is reasonable to conclude that the chelation chemistry is distinctly different from the ligand-binding event responsible for the increase in protein fold stability. DSF is now a well-established technique for screening small molecules against His-tagged proteins [65]. Within our laboratory, erroneous zinc binding has not previously been observed even for similar protein candidates [98] indicating that in this format a true binding event can be distinguished from a transient interaction.

DSF is a powerful analytical technique due to its sensitivity, compatibility with high-throughput formats, and small quantities of an analyte required [65, 99]. Despite being able to sensitively resolve thermal transitions, this technique does not allow measurement of kinetics of an interaction. Nevertheless, the use of this assay has rapidly allowed each protein to be screened against a chemical space incorporating 300 common small molecule binding partners including multiple representative saccharides (Table 4.4), allowing an assessment to be made of their annotation. For SP5 and SP6, their annotation as disaccharide-binding proteins appears incorrect as no response was observed for any saccharide compound screened. Disaccharide-binding proteins are found in structural Class D (Section 1.3), thus these proteins clearly do not fall into this structural grouping. Zinc-binding proteins, by comparison are found exclusively in Class A, allowing inferences about the tertiary structure of SP5 and SP6 to be made, with the single α-helix in the hinge region that defines this cluster

possibly corresponding to the disordered region previously observed for these proteins in bioinformatic analysis (Section 3.1). Further investigation, such as structural characterisation of these two proteins, may be necessary to validate this and ascribe their function.

Table 4.4 Saccharide components of Silver Bullets screen

Compound name	Compound class
D-(+)-trehalose dihydrate	disaccharide
D-fructose 1,6-bisphosphate trisodium salt hydrate	monosaccharide
D-glucose 6-phosphate sodium salt	monosaccharide
D-(+)-maltose monohydrate	disaccharide
D-(+)-melibiose monohydrate	disaccharide
D-(+)- raffinose pentahydrate	trisaccharide
D-(+)-trehalose dehydrate	disaccharide
stachyose hydrate	oligosaccharide
cyclodextrin	cyclic oligosaccharide
D-(+)-cellobiose	disaccharide
D-(+)-maltotriose	trisaccharide
D-(+)-melezitose hydrate	trisaccharide
digest of dextran	mixture of saccharides
D-sorbitol	sugar alcohol
myo-inositol	sugar alcohol

Chapter 5: Conclusions

5.1 Assessment of SBP Characterisation

In addressing the aims of my project, three novel SBPs from a panel of nineteen annotated saccharide-binding proteins from representative *Synechococcus sp.* were successfully produced in recombinant form. An alternative high-throughput strategy for this production was also explored, however, still requires further optimisation. Adoption of the OPPF-UK method shows great promise, particularly for advantageous sample handling and the screening of multiple conditions to potentially increase on the observed success rates using the MQ/SGC established method.

Following their production, each target was subjected to biophysical characterisation. This identified the formation of a dimeric complex for one target (SP19), which may have implications for regulation of the greater transporter activity, a process that is still poorly understood [2, 6, 100]. Reversible dimerisation may also play a role in the acquisition of compatible solutes under osmotic pressure, and conditions promoting the formation or dissolution of the dimeric complex could be further explored.

High-throughput screening of ligand candidates identified that two targets (SP5 and SP6) did not interact with their annotated saccharide binding chemistries, instead being shown to interact with divalent zinc in a highly thermodynamically favourable manner. This indicates their original bioinformatic annotation to be incorrect, casting doubt on functional and biochemical inferences made from these.

5.2 The Structural Basis for Zinc-binding in SBPs

Metals coordinated within proteins have their first coordination sphere satisfied largely by donor electrons from the side chains of individual amino acids and then exogenous waters of solvation, forming elaborate bonding networks [101]. Commonly, the required electrons are donated from nitrogen atoms (e.g. imidazolyl group of histidine), and sulphur atoms (e.g. thiolate group of cysteine), however, other sufficiently electronegative atoms may also conjugate this system. Biologically, zinc is found in an oxidation state of +2, influencing the overall molecular geometry of the metal-protein complex, requiring four co-ordinating residues to satisfy its tetrahedral molecular orbital arrangement [101].

Within the last comprehensive structural review of the SBP family, the structures of seven SBPs bound to zinc (Zn-SBPs) were noted [39], all containing zinc in its +2 oxidation state. Examples of the ligand-binding site from four representative Zn-SBPs are shown as Fig. 5.1. These show the coordination site is highly structurally conserved: these Zn-SBPs engage a

triad of histidine, with the fourth site filled by oxygen arising from exogenous water or an anionic side chain [102]. Therefore, it is surprising to find that neither SP5 or SP6 contain sufficient histidine residues to satisfy the histidine triad paradigm seen in other Zn-SBPs, with SP5 containing two histidine and one cysteine residues, and SP6 containing one histidine and no cysteine residues. The lack of conservation of the Zn-binding site seen in all examples of a Zn-SBP indicates zinc may not be the natural ligand for SP5 and SP6, or may highlight a new architecture for zinc binding not previously observed for this family of proteins. The group of SBPs designated to structural Class A (Section 1.3) [38, 39] and incorporating other metal ligand types are summarised in Table 5.1. Examination of the residues forming the coordination site, R_1-R_4, shows the regular involvement of histidine as a coordinating residue.

Metal-ligated SBPs found in other structural classes, such as the iron-binding proteins (Class D) contain different metal binding motifs, such as the tyrosine clamp, to satisfy alternative molecular orbital arrangements of their cognate metals. The example of the two iron binding proteins, FutA1 and FutA2 from *Synechocystis* highlights the involvement of a network of tyrosine residues forming the binding pocket [26, 27].

The presence of a *Tat* sequence motif for SP5 and SP6 can also be said to preclude the role of zinc as a co-factor required for protein folding prior to translocation (Section 3.1), due to the tightly regulated cytoplasmic concentrations of high-affinity metals, such as copper and zinc, to allow kinetically favourable interactions with lower affinity metals required during protein folding, for example manganese [28, 68].
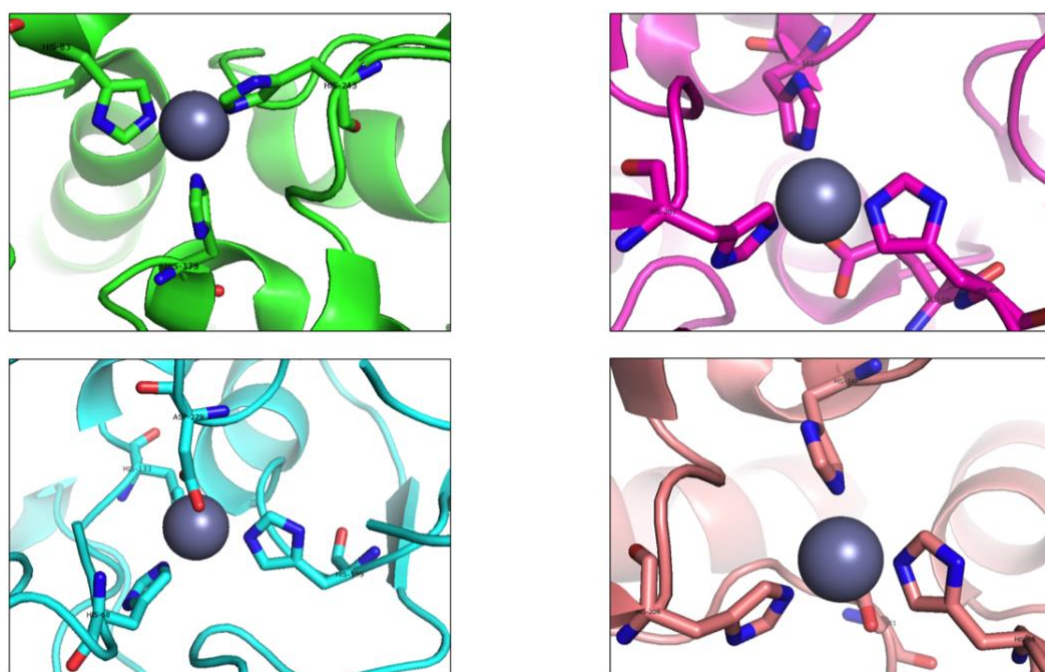

Fig 5.1 Examples of molecular architecture of the ligand-binding site for other Zn-SBPs. PDB IDs clockwise from top left: 1PQ4 [103], 1TOA [104], 2PRS [105], 3HJT [106].

45

Table 5.1 metal-binding motifs in other SBPs

PDB ID	Source organism	Ligand	Res. /Å	R_1	R_2	R_3	R_4
1PSZ	Streptococcus pneumoniae	Mn^{2+}	2.0	H67	H139	E205	D280
1XVL	Synechocystis sp. PCC 6803	Mn^{2+}	2.9	H89	H154	E220	H_2O
2ETV[a]	Thermotoga maritima	Ni^{2+}	1.7	H-1	H-3	E33	H_2O
3HH8	Streptococcus pyogenes	Fe^{3+}	1.8	H77	H139	E205	D280
4CL2	Candidatus liberibacter asiaticus	Mn^{2+}	1.6	H106	E172	D247	H_2O
4INP	Helicobacter pylori	Ni^{2+}	2.3	H103	E153	E94	R230
4K3V	Staphylococcus aureus	Mn^{2+}	2.2	H50	H123	E189	D264
4OXR	Staphylococcus pseudintermedius	Mn^{2+}	2.0	H64	H137	E203	D278

[a] metal binding for this protein required the engagement of the histidine tag

5.3 A Potential Functional Role for SP5 and SP6

In-depth analysis at the genome level shows gene cluster CK_00001342, from which both SP5 and SP6 originate, falls into the *Synechococcus* core genome. A pairwise alignment of the two primary sequences showed an identity score of 70 %. This, combined with their similar solution properties, and identical binding partner suggests the two to be orthologous. Similarly, despite the presence of pathogenicity islands in *Synechococcus* genomes that are indicative of phage-mediated LGT [16] (potentially explaining the cause of genome diversity and niche adaptation [10, 17]), the two source strains, CC9902 and WH8102, occupy distinct environments (Section 1.2) limiting the potential for LGT events. This demonstrates a potentially more fundamental role for these proteins in *Synechococcus* biology, indicating a vertical gene transfer event rather than niche adaptation explains their presence in these two different genomes.

Examination of the operons containing SP5 and SP6 (Fig. 5.2) shows a highly conserved operon structure, containing elements commonly associated with DNA binding (such as single-strand DNA-binding protein), and light-harvesting machinery (phycobilin activator). Contained in an operon downstream of this is the RNA-binding protein ssrA-binding protein (involved in recognition of ribosome stalling). The activation of light-harvesting apparatus may indicate a light-mediated response similar to previously observed mixotrophy (utilisation of light-generated ATP for nutrient uptake) seen in related species [9,

49]. Similarly, these SBPs are not transcribed together with their greater ABC machinery, indicating they are not constitutively expressed.

Biologically, zinc is an important catalytic and structural trace element vital to the functioning of many proteins, particularly in DNA replication, and the biosynthesis of amino acids and extracellular peptidoglycan [107]. Given the involvement of DNA-binding and RNA-binding elements, and the positioning of SP5 and SP6 within the wider genomic context, this indicates they may play a role in transcriptional activation under stress conditions.
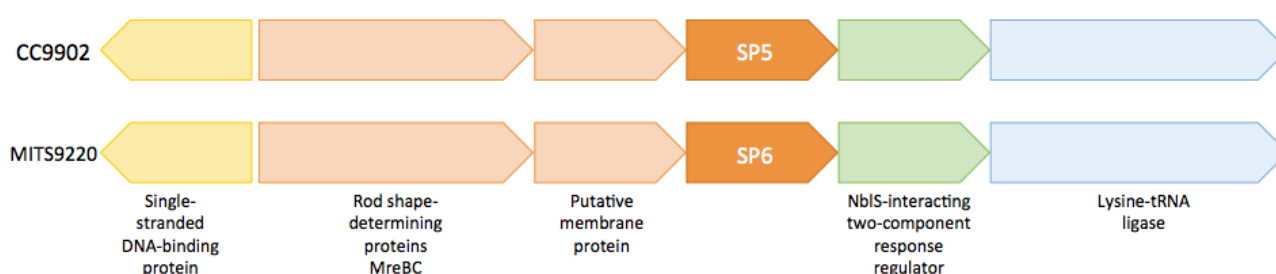

Fig 5.2 Stylised representation of genome contexts of SP5 and SP6. The operon containing SP5 and SP6 (orange) are shown with flanking operons (coloured individually). The orthologous gene for SP5 and SP5 is indicated in dark orange.

5.4 Concluding Remarks

Exploring the annotation 'grey space' for the SBP family has demonstrated ligand predictions made through phylogenetic relationships require experimental validation. Identification of novel ligands may have profound implications for the current paradigms of fundamental biochemistry for *Synechococcus sp.*, and the models of global ecosystems made from these.

Despite genome-level evidence pointing to the involvement of SP5 and SP6 in transcriptional regulation or stress response mechanism, the evidence for zinc binding requires further validation. Despite the clear increase in thermal stability conferred by the presence of zinc, the canonical evidence of a highly conserved Zn-binding motif is contradictory. This may suggest a novel architecture for zinc binding not previously observed for this family. Conversely, it may be that as thermal denaturation occurs a lower affinity metal – either a co-factor required for folding, or a cognate ligand – may instead become solvent exposed and displaced by zinc, due to the greater stability imparted. Therefore, for this protein family, the question of ligand specificity, and thus biochemical function may only be answerable from a structural perspective.

Reference List:

[1] Padan E. Bacterial Membrane Transport: Superfamilies of Transport Proteins. 2009.

[2] Rees DC, Johnson E, Lewinson O. ABC transporters: the power to change. Nature Reviews Molecular and Cellular Biology. 2009;10:218-27.

[3] Saier Jnr. MH. A Functional-Phylogenetic Classification System for Transmembrane Solute Transporters. Microbiol Mol Biol Rev. 2000;64:354-411.

[4] Busch W, Saier MH, Jr. The transporter classification (TC) system, 2002. Crit Rev Biochem Mol Biol. 2002;37:287-337.

[5] Ren Q, Paulsen IT. Comparative analyses of fundamental differences in membrane transport capabilities in prokaryotes and eukaryotes. PLoS Comput Biol. 2005;1:e27.

[6] Cui J, Davidson AL. ABC solute importers in bacteria. Essays Biochem. 2011;50:85-99.

[7] Kretz CB, Bell DW, Lomas DA, Lomas MW, Martiny AC. Influence of growth rate on the physiological response of marine Synechococcus to phosphate limitation. Front Microbiol. 2015;6:85.

[8] Sunda WG, Huntsman SA. High iron requirement for growth, photosynthesis, and low-light acclimation in the coastal cyanobacterium Synechococcus bacillaris. Front Microbiol. 2015;6:561.

[9] Moore LR. More mixotrophy in the marine microbial mix. Proc Natl Acad Sci U S A. 2013;110:8323-4.

[10] Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, et al. Ecological genomics of marine picocyanobacteria. Microbiol Mol Biol Rev. 2009;73:249-99.

[11] Ren Q, Paulsen IT. Large-scale comparative genomic analyses of cytoplasmic membrane transport systems in prokaryotes. J Mol Microbiol Biotechnol. 2007;12:165-79.

[12] Flombaum P, Gallegos JL, Gordillo RA, Rincón J, Zabala LL, Jiao N, et al. Present and future distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. PNAS. 2013;110.

[13] Hagemann M. Molecular biology of cyanobacterial salt acclimation. FEMS Microbiol Rev. 2011;35:87-123.

[14] Mazard S, Penesyan A, Ostrowski M, Paulsen IT, Egan S. Tiny Microbes with a Big Impact: The Role of Cyanobacteria and Their Metabolites in Shaping Our Future. Mar Drugs. 2016;14.

[15] Jardillier L, Zubkov MV, Pearman J, Scanlan DJ. Significant CO2 fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. ISME J. 2010;4:1180-92.

[16] Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, et al. Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. Genome Biology. 2008;9:1-16.

[17] Palenik B, Brahamsha B, Larimer FW, Land M, Hauser L, Chain P, et al. The genome of a motile marine *Synechococcus*. Nature. 2003;424:1037-42.

[18] Juhas M, van der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol Rev. 2009;33:376-93.

[19] Baumdicker F, Hess WR, Pfaffelhuber P. The infinitely many genes model for the distributed genome of bacteria. Genome Biol Evol. 2012;4:443-56.

[20] Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. The microbial pan-genome. Curr Opin Genet Dev. 2005;15:589-94.

[21] Palinska KA, Surosz W. Taxonomy of cyanobacteria: a contribution to consensus approach. Hydrobiologia. 2014;740:1-11.

[22] Kirkham AR, Lepere C, Jardillier LE, Not F, Bouman H, Mead A, et al. A global perspective on marine photosynthetic picoeukaryote community structure. ISME J. 2013;7:922-36.

[23] Roscoff SBd. Cyanorak Database. 13/03/2012 ed2012.

[24] Shilova IN, Robidart JC, James Tripp H, Turk-Kubo K, Wawrik B, Post AF, et al. A microarray for assessing transcription from pelagic marine microbial taxa. ISME J. 2014;8:1476-91.

[25] Bombar D, Turk-Kubo KA, Robidart J, Carter BJ, Zehr JP. Non-cyanobacterial nifH phylotypes in the North Pacific Subtropical Gyre detected by flow-cytometry cell sorting. Environ Microbiol Rep. 2013;5:705-15.

[26] Koropatkin N, Randich AM, Bhattacharyya-Pakrasi M, Pakrasi HB, Smith TJ. The Structure of the Iron-binding Protein, FutA1, from *Synechocystis* 6803. Journal of Biological Chemistry. 2007;282:27468-77.

[27] Badarau A, Firbank SJ, Waldron KJ, Yanagisawa S, Robinson NJ, Banfield MJ, et al. FutA2 Is a Ferric Binding Protein from *Synechocystis* PCC 6803. Journal of Biological Chemistry. 2008;283:12520-7.

[28] Barnett JP, Millard A, Ksibe AZ, Scanlan DJ, Schmid R, Blindauer CA. Mining genomes of marine cyanobacteria for elements of zinc homeostasis. Front Microbiol. 2012;3:142.

[29] Lewinson O, Livnat-Levanon N. Mechanism of Action of ABC Importers: Conservation, Divergence, and Physiological Adaptations. J Mol Biol. 2017;429:606-19.

[30] Rice AJ, Park A, Pinkett HW. Diversity in ABC transporters: type I, II and III importers. Crit Rev Biochem Mol Biol. 2014;49:426-37.

[31] ter Beek J, Guskov A, Slotboom DJ. Structural diversity of ABC transporters. J Gen Physiol. 2014;143:419-35.

[32] Locher KP. Mechanistic diversity in ATP-binding cassette (ABC) transporters. Nat Struct Mol Biol. 2016;23:487-93.

[33] Smith PC, Karpowich N, Millen L, Moody JE, Rosen J, Thomas PJ, et al. ATP Binding to the Motor Domain from an ABC Transporter Drives Formation of a Nucleotide Sandwich Dimer. Molecular Cell. 2002;10:139-49.

[34] Locher KP, Lee AT, Rees DC. The *E. coli* BtuCD Structure: A Framework for ABC Transporter Architecture and Mechanism. Science. 2002;296:1091-8

[35] Maqbool A, Horler RS, Muller A, Wilkinson AJ, Wilson KS, Thomas GH. The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity. Biochem Soc Trans. 2015;43:1011-7.

[36] Korkhov VM, Mireku SA, Veprintsev DB, Locher KP. Structure of AMP-PNP-bound BtuCD and mechanism of ATP-powered vitamin B12 transport by BtuCD-F. Nat Struct Mol Biol. 2014;21:1097-9.

[37] Dwyer MA, Hellinga HW. Periplasmic binding proteins: a versatile superfamily for protein engineering. Curr Opin Struct Biol. 2004;14:495-504.

[38] Berntsson RP, Smits SH, Schmitt L, Slotboom DJ, Poolman B. A structural classification of substrate-binding proteins. FEBS Lett. 2010;584:2606-17.

[39] Scheepers GH, Lycklama ANJA, Poolman B. An updated structural classification of substrate-binding proteins. FEBS Lett. 2016;590:4393-401.

[40] Borrok MJ, Kiessling LL, Forest KT. Conformational changes of glucose/galactose-binding protein illuminated by open, unliganded, and ultra-high-resolution ligand-bound structures. Protein Sci. 2007;16:1032-41.

[41] Mao B, Pear MR, McCammon JA, Quiocho FA. Hinge-bending in L-Arabinose-binding Protein. Journal of Biological Chemistry. 1982;257:1131-3.

[42] Phillips GNJ, Mahajan VK, Siu AKQ, Quiocho FA. Structure of L-arabinose-binding protein from *Escherichia coli* at 5 Å resolution and preliminary results at 3.5 Å. PNAS. 1976;73:2186-90.

[43] Quiocho FA, Meador WE, Pflugrath JW. Preliminary Crystallographic Data of Receptors for Transport and Chemotaxis in Eschericia coli: D-Galactose and Maltose-binding Protein. J Mol Biol. 1979;133:181-4.

[44] Tam R, Saier Jnr. MH. Structural, Functional, and Evolutionary Relationships among Extracellular Solute-Binding Receptors of Bacteria. Microbiological Reviews. 1993;57:320-246.

[45] Fukami-Kobayashi K, Tateno Y, Nishikawa K. Domain Dislocation: a Change of Core Structure in Periplasmic Binding Proteins in their Evolutionary History. JMB. 1999;286:279-90.

[46] Gonin S, Arnoux P, Pierru B, Lavergne J, Alonso B, Sabaty M, et al. Crystal structures of an Extracytoplasmic Solute Receptor from a TRAP transporter in its open and closed forms reveal a helix-swapped dimer requiring a cation for alpha-keto acid binding. BMC Struct Biol. 2007;7:11.

[47] Mills MM, Moore CM, Langlois R, Milne A, Achterberg E, Nachtigall K, et al. Nitrogen and phosphorous co-limitation of bacterial productivity and growth in the oligotrophic subtropical North Atlantic. Limnol Oceanogr. 2008;53:824-34.

[48] Garcia NS, Bonachela JA, Martiny AC. Interactions between growth-dependent changes in cell size, nutrient supply and cellular elemental stoichiometry of marine Synechococcus. ISME J. 2016;10:2715-24.

[49] Eiler A. Evidence for the ubiquity of mixotrophic bacteria in the upper ocean: implications and consequences. Appl Environ Microbiol. 2006;72:7431-7.

[50] Christendat D, Yee A, Dharamsi A, Kluger Y, Gerstein M, Arrowsmith CH, et al. Structural proteomics: prospects for high throughput sample preparation. Progress in Biophysics and Molecular Biology. 2000;73:339-45.

[51] Structural Genomics Consortium, China Structural Genomics Consortium, Northeast Structural Genomics Consortium, Graslund S, Nordlund P, Weigelt J, et al. Protein production and purification. Nat Methods. 2008;5:135-46.

[52] Robinson A, Guilfoyle AP, Sureshan V, Howell M, Harrop SJ, Boucher Y, et al. Structural Genomics of the Bacterial Mobile Metagenome: an Overview. In: Kobe B, Guss M, Huber T, editors. Structural Proteomics: High-throughput Methods. Totowa, NJ: Humana Press; 2008.

[53] Linding R. GlobPlot: exploring protein sequences for globularity and disorder. Nucleic Acids Research. 2003;31:3701-8.

[54] Viklund H, Elofsson A. Best alpha-helical transmembrane protein topology predictions are achieved using hidden Markov models and evolutionary information. Protein Sci. 2004;13:1908-17.

[55] Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol. 2004;340:783-95.

[56] Imai K, Asakawa N, Tsuji T, Akazawa F, Ino A, Sonoyama M, et al. SOSUI-GramN: high performance prediction for sub-cellular localization of proteins in Gram-negative bacteria. Bioinformation. 2008;2:417-21.

[57] Sambrook J, Russell DW. Molecular Cloning 3rd ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 2001.

[58] L. R. Moore C, A., Zinser, E. R., Saito, M. A., Sullivan, M. B., Lindell, D., Frois-Moniz, K., Waterbury, J., and Chisholm, S. W. . Culturing the marine cyanobacterium Prochlorococcus. Limnol Oceanogr: Methods. 2007;5.

[59] Studier FW. Protein production by auto-induction in high-density shaking cultures. Protein Expression and Purification. 2005;41:207-34.

[60] Palenik B, Ren Q, Dupont CL, Myers GS, Heidelberg JF, Badger JH, et al. Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment. PNAS. 2006;103:13555-9.

[61] Aslanidia C, de Jong PJ. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Research. 1990;18:6069-74.

[62] Barth HG, Boyes BE, Jackson CJ. Size Exclusion Chromatography. Anal Chem. 1996;68:445-66.

[63] Harris DC. Quantitative Chemical Analysis. 8 ed. New York, USA: W. H. Freeman and Company; 2010.

[64] BioRad. A Guide to Polyacrylamide Gel Electrophoresis and Detection.

[65] Niesen FH, Berglund H, Vedadi M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc. 2007;2:2212-21.

[66] Silver Bullets User Guide. Hampton Research; 2017.

[67] Silhavy TJ, Kahne D, Walker S. The bacterial cell envelope. Cold Spring Harb Perspect Biol. 2010;2:a000414.

[68] Barnett JP, Robinson C, Scanlan DJ, Blindauer CA. The Tat protein export pathway and its role in cyanobacterial metalloprotein synthesis. FEMS Microbiol Lett. 2011;325:1-9.

[69] Driessen AJ, Nouwen N. Protein translocation across the bacterial cytoplasmic membrane. Annu Rev Biochem. 2008;77:643-67.

[70] Rajalahti T, Huang F, Klement MR, Pisareva T, Edman M, Sjöström M, et al. Proteins in Different Synechocystis Compartments Have Distinguishing N-Terminal Features: A Combined Proteomics and Multivariate Sequence Analysis. Journal of Proteome Research. 2007;6:2420-34.

[71] Hou B, Frielingsdorf S, Klösgen RB. Unassisted Membrane Insertion as the Initial Step in ∆pH/Tat-dependent Protein Transport. Journal of Molecular Biology. 2006;355:957-67.

[72] Bageshwar UK, Whitaker N, Liang F-C, Musser SM. Interconvertibility of Lipid- and Translocon-bound Forms of the Bacterial Tat Precursor pre-SufI. Molecular microbiology. 2009;74:209-26.

[73] Zak E, Norling B, Maitra R, Huang F, Andersson B, Pakrasi HB. The initial steps of biogenesis of cyanobacterial photosystems occur in plasma membranes. Proceedings of the National Academy of Sciences of the United States of America. 2001;98:13443-8.

[74] Nevo R, Charuvi D, Shimoni E, Schwarz R, Kaplan A, Ohad I, et al. Thylakoid membrane perforations and connectivity enable intracellular traffic in cyanobacteria. The EMBO Journal. 2007;26:1467-73.

[75] Matos CF, Branston SD, Albiniak A, Dhanoya A, Freedman RB, Keshavarz-Moore E, et al. High-yield export of a native heterologous protein to the periplasm by the tat translocation pathway in Escherichia coli. Biotechnol Bioeng. 2012;109:2533-42.

[76] Berks BC. A common export pathway for proteins binding complex redox factors? Mol Microbiol. 1996;22:393-404.

[77] Natale P, Bruser T, Driessen AJ. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane--distinct translocases and mechanisms. Biochim Biophys Acta. 2008;1778:1735-56.

[78] Ragionieri L, Vitorino R, Frommlet J, Oliveira JL, Gaspar P, Ribas de Pouplana L, et al. Improving the accuracy of recombinant protein production through integration of bioinformatics, statistical and mass spectrometry methodologies. FEBS J. 2015;282:769-87.

[79] Rose RW, Brüser T, Kissinger JC, Pohlschröder M. Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway. Mol Microbiol. 2002;45:943-50.

[80] Dilks K, Rose RW, Hartmann E, Pohlschroder M. Prokaryotic Utilization of the Twin-Arginine Translocation Pathway: a Genomic Survey. Journal of Bacteriology. 2003;185:1478-83.

[81] Jia B, Jeon CO. High-throughput recombinant protein expression in *Escherichia coli*: current status and future perspectives. Open Biol. 2016;6.

[82] Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: advances and challenges. Front Microbiol. 2014;5:172.

[83] Bird LE. High throughput construction and small scale expression screening of multi-tag vectors in Escherichia coli. Methods. 2011;55:29-37.

[84] Sureshan V, Deshpande CN, Boucher Y, Koenig JE, Midwest Center for Structural G, Stokes HW, et al. Integron gene cassettes: a repository of novel protein folds with distinct interaction sites. PLoS One. 2013;8:e52934.

[85] Bird LE, Rada H, Flanagan J, Diprose JM, Gilbert RJ, Owens RJ. Application of In-Fusion cloning for the parallel construction of E. coli expression vectors. Methods Mol Biol. 2014;1116:209-34.

[86] Horn C, Sohn-Bosser L, Breed J, Welte W, Schmitt L, Bremer E. Molecular determinants for substrate specificity of the ligand-binding protein OpuAC from Bacillus subtilis for the compatible solutes glycine betaine and proline betaine. J Mol Biol. 2006;357:592-606.

[87] Diez J, Diederichs K, Greller G, Horlacher R, Boos W, Welte W. The crystal structure of a liganded trehalose/maltose-binding protein from the hyperthermophilic Archaeon Thermococcus litoralis at 1.85 A. J Mol Biol. 2001;305:905-15.

[88] Sheehan D. Physical Biochemistry: Principles and Applications. 2 ed. Hoboken, NJ: John Wiley and Sons; 2009.

[89] Scopes RK. Protein Purification - Principles and Practice. 3 ed. New York, USA: Springer-Verlag; 1994.

[90] Tayyab S, Qamar S, Islam M. Size Exclusion Chromatography and Size Exclusion HPLC of Proteins. Biochemical Education. 1991;19:149-52.

[91] Du X, Li Y, Xia YL, Ai SM, Liang J, Sang P, et al. Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. Int J Mol Sci. 2016;17.

[92] Muller I. Guidelines for the successful generation of protein-ligand complex crystals. Acta Crystallogr D Struct Biol. 2017;73:79-92.

[93] Miklos AC, Sumpter M, Zhou HX. Competitive interactions of ligands and macromolecular crowders with maltose binding protein. PLoS One. 2013;8:e74969.

[94] McMurry J. Organic Chemistry. 8 ed. Belmont, California: Brooks/Cole; 2012.

[95] Vetting MW, Al-Obaidi N, Zhao S, San Francisco B, Kim J, Wichelecki DJ, et al. Experimental strategies for functional annotation and metabolism discovery: targeted screening of solute binding proteins and unbiased panning of metabolomes. Biochemistry. 2015;54:909-31.

[96] Clifton BE, Jackson CJ. Ancestral Protein Reconstruction Yields Insights into Adaptive Evolution of Binding Specificity in Solute-Binding Proteins. Cell Chem Biol. 2016;23:236-45.

[97] Irving H, Williams RJP. Order of Stability of Metal Complexes. Nature. 1948;162:746-7.

[98] McLeish AG. Molecular basis for nutrient uptake in marine bacteria. Macquarie University; 2017 (unpublished work).

[99] Vollrath F, Hawkins N, Porter D, Holland C, Boulet-Audet M. Differential Scanning Fluorimetry provides high throughput data on silk protein transitions. Sci Rep. 2014;4:5625.

[100] Locher KP. Mechanistic diversity in ATP-binding cassette (ABC) transporters. Nature Structural and Molecular Biology. 2016;23:487-93.

[101] Holm RH, Kennepohl P, Solomon EI. Structural and Functional Aspects of Metal Sites in Biology. Chem Rev. 1996;96:2239-314.

[102] Ilari A, Alaleona F, Petrarca P, Battistoni A, Chiancone E. The X-ray Structure of the Zinc Transporter ZnuA from Salmonella enterica Discloses a Unique Triad of Zinc-Coordinating Histidines. Journal of Molecular Biology. 2011;409:630-41.

[103] Banerjee S, Wei B, Bhattacharyya-Pakrasi M, Pakrasi HB, Smith TJ. Structural Determinants of Metal Specificity in the Zinc Transport Protein ZnuA from *Synechocystis 6803*. JMB. 2003;333:1061-9.

[104] Lee YH, Deka RK, M.V. N, Radolf JD, Hasemann CA. *Treponema pallidum* TroA is a periplasmic zinc-binding protein with a helical backbone. Nat Struct Biol. 1999;6:628-33.
[105] Yatsunyk LA, Easton JA, Kim LR, Sugarbaker SA, Bennett B, Breece RM, et al. Structure and metal binding properties of ZnuA, a periplasmic zinc transporter from Escherichia coli. Journal of biological inorganic chemistry : JBIC : a publication of the Society of Biological Inorganic Chemistry. 2008;13:271-88.
[106] Ragunathan P, Spellerberg B, Ponnuraj K. Structure of laminin-binding adhesin (Lmb) from Streptococcus agalactiae. Acta Crystallogr D Biol Crystallogr. 2009;65:1262-9.
[107] Porcheron G, Garenaux A, Proulx J, Sabri M, Dozois CM. Iron, copper, zinc, and manganese transport and regulation in pathogenic Enterobacteria: correlations between strains, site of infection and the relative importance of the different metal transport systems for virulence. Front Cell Infect Microbiol. 2013;3:90.