Characterisation of putative glycan and drug binding proteins predicted using *in silico* screening methods

Jack Rowlatt
B.Sc

Institute for Glycomics
School of Medical Science
Griffith University, Gold Coast

Submitted in fulfilment of the requirements of the degree of
Master of Medical Research

May, 2020

Abstract

This thesis describes the characterisation of six targets for a novel antimicrobial drug 3,4-methylenedioxy-beta-nitropropene (BDM-I) shown to inhibit bacterial, protozoal and fungal infections and the characterisation putative carbohydrate binding proteins (CBP) BT_411 and BT_3781. Furthermore, described here is the experimental validation of bioinformatic programs SPOT-Ligand, of which identified six putative drug targets for BDM-I, and SPOT-Struc, that predicted BT_411 and BT_3781 as CBPs. Specifically I expressed and purified the six BDM-I targets predicted by SPOT-Ligand and expressed the two SPOT-Struc predicted putative CBPs BT_411 and BT_3781. The drug targets were then used to identify a potential mechanism of action for BDM-I while the putative CBPs were likewise characterised to identify their carbohydrate binding affinity. Lastly the characterisation results were then utilised to evaluate the predictive capabilities of SPOT-Ligand and SPOT-Struc and in turn help clarify the role of bioinformatics in experimental research.

The six BDM-I targets (Gene name: EF0414, ubiE, ftsZ5, Lebu_1328, acpD, and ubiH) were expressed in *E. coli* BL21 (DE3) cells with expressed recombinant soluble protein purified to homogeneity using HIS-select nickel affinity and size exclusion chromatography (SEC). Likewise, two putative CBPs BT_411 and BT_3781 as well as an additional novel mushroom lectin PSL-2 (used as a positive lectin control), were expressed in *E. coli* BL21 (DE3) cells and purified to homogeneity using affinity chromatography and SEC. All proteins were confirmed pure by SDS-PAGE and of those proteins containing a HIS-tag (all except PSL-2), western blot immunodetection. All purified proteins were then measured using circular dichroism (CD) spectroscopy as a quality control step to determine if they had denatured before proceeding with further characterisation. The six potential

drug targets were analysed for BDM-I affinity using surface plasmon resonance (SPR) with those showing affinity for BDM-I further characterised by computational docking analysis. Similarly, putative CBPs BT_411 and BT_3781 in conjunction with PSL-2 were analysed for carbohydrate affinity with a wide variety of glycans using SPR followed by computational docking analysis of the CBPs with some of the SPR defined glycan matches.

Results showed that BDM-I has high affinity for drug target AcpD from *Salmonella enterica, serovar Typhimurium*, a bacterial azoreductase protein responsible for the breakdown of azo dyes in *S. typhimurium* (*S. typhimurium* causes gastroenteritis in humans). SPR analysis of the binding of BDM-I to AcpD revealed an equilibrium dissociation constant (K_D) of 0.58 µM with an association constant (k_a) of $9.4 \times 10^4\ M^{-1}s^{-1}$ and dissociation constant (k_d) of $9.7 \times 10^{-4}\ s^{-1}$, suggesting that BDM-I binds rapidly and easily to acpD followed by a slow dissociation. Docking analysis of BDM-I in AcpD showed hydrogen bonds between a BDM-I nitro group and residues Ala115 and Asn98, with a binding energy of – 6.7 kcal/mol. AcpD structural alignment with an azoreductase that also shows nitroreductase activity, PaAzoR from *Pseudomonas aeruginosa*, showed a high enough structural similarity with AcpD to computationally infer that function is likely shared between the two proteins. As such, it was computationally determined that AcpD may act on BDM-I as a nitroreductase via the binding of BDM-I's nitro group in order to reduce it to chemical intermediates (similar to paAzoR nitroreductase activity). This reveals a possible mechanism of action for BDM-I; namely that bacterial nitroreductases may bind and reduce the nitro group on BDM-I resulting in the production of bacterially toxic intermediates.

SPR analysis of BT_411 revealed carbohydrate affinity for β1,4 linked *N*-acetylglucosamines (GlcNAc) and α2,3 linked sialic acids (Neu5Ac) with K_D's between

0.10 µM and 0.23 µM. Computational analysis of BT_411 further indicated preferential affinity towards GlcNAc (-4.6 kcal/mol) over Neu5Ac and indicated that BT_411 is likely a carbohydrate binding module (CBM) showing possible functional attributes of a βGNase (hydrolase) that catalyses the reduction of GlcNAc glycosidic bonds.

SPR analysis of BT_3781 indicated carbohydrate affinity for fucose (Fuc) and galactose (Gal) containing glycans; specifically blood group H, B and A showing high affinity with K_D's between 0.17 µM and 0.34 µM. Computational analysis of BT_3781 supported SPR analysis indicating affinity with Fuc, in particular blood group H disaccharide (Fucα1-2Gal) with a binding energy of -6.8 kcal/mol. Indicating that BT_3781 is also a CBM with functional similarities to that of a glycosidic hydrolase; namely a fucosidase that catalyses the reduction of Fucα1-2 linkages.

SPR analysis of the positive control PSL-2 confirmed crystallography data of the protein that showed preferential binding to Gal and Fuc residues. Specifically, SPR indicated PSL-2 has high affinity towards blood group B (Galα1-3(Fucα1-2)Galβ1-4Glc), implying that PSL-2 may be a blood group lectin. Moreover, successful characterisation of PSL-2 validates the experimental procedures used in the characterisation of BT_411 and BT_3781.

The characterisation results of BDM-I drug targets and CBPs BT_411 and BT_3781 validate the supportive role of bioinformatic programs SPOT-Ligand and SPOT-Struc in experimental research. Mainly in the structural predication of drug targets and identification of novel CBPs.

Statement of Originality

This work has not previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

(Signed)_____
Jack Rowlatt

Table of Contents **Page number**

Table of Contents **Page number**

List of Tables **Page number**

Chapter 1

Chapter 2

Chapter 3

List of Figures **Page number**

Chapter 1

Chapter 2

Chapter 3

List of Figures **Page number**

Appendix

List of Abbreviations

AcpD	Azoreductase from *Salmonella typhimurium LT2*
AfcA	1,2-a-L-fucosidase from *Bifidobacterium bifidum*
AgnC	Family 89 glycoside hydrolase from *C. perfringens*
A.I.	Artificial intelligence
Ala	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
AzoR	Azoreductase from *Escherichia coli*
BDM-I	3,4-methylenedioxy-β-nitropropene
BSA	Bovine serum albumin
BCA	Bicinchoninic acid assay
BT_3781	Hypothetical protein from *Bacteroides thetaiotaomicron*
BT_0441	Hypothetical protein from *Bacteroides thetaiotaomicron*
CBM	Carbohydrate binding molecule
CBP	Carbohydrate binding protein
CD	Circular dichroism
CPU	Central processing units
CRD	Carbohydrate recognition domain
CV	Column volume
Cys	Cysteine
EF0404	Azoreductase from *Enterococcus faecalis*
EDTA	Ethylenediaminetetraacetic acid
EF0414	DadA family oxidoreductase from *Enterococcus Faecalis*
ER	Endoplasmic reticulum
FC	Flow cell
FMN	Flavin mononucleotide
FPLC	Fast protein liquid chromatography
ftsZ5	Cell division protein from *Halobacterium sp. NRC-1*
Fuc	L-Fucose
GAG	Glycosaminoglycan
Gal	D-Galactose

GalNAc	*N*-acetyl-D-galactosamine
GHz	Gigahertz
Glc	D-Glucose
GlcNAc	*N*-acetylglucosamine
Glu	Glutamic acid
Gly	Glycine
GPU	Graphics processing units
GRFT	Griffithsin
HCl	Hydrochloric acid
HIS	Histidine
Ile	Isoleucine
IPTG	Isopropyl β- d-1-thiogalactopyranoside
K_a	Association constant
K_{av}	Phase distribution coefficient
K_D	Equilibrium dissociation constant
K_d	Dissociation constant
KCl	Potassium chloride
kDa	Kilo-Dalton
KH_2PO_4	Monopotassium phosphate
LAB	Lactic acid bacteria
Lac	Lactose
Lebu_1328	Tubulin/FtsZ GTPase from *Leptotrichia buccalis C-1013-b*
Leu	Leucine
Lys	Lysine
Man	D-Mannose
Met	Methionine
MIC	Minimum inhibitory concentrations
MS	Mass spectrometry
Mw	Molecular weight
NaCl	Sodium chloride
NADH	Nicotinamide adenine dinucleotide
Na_2HPO_4	Disodium phosphate
Neu5Ac	*N*-acetylneuraminic acid
NFZ	Nitrofurantoin

Ni^{+}NTA	Nickel-charged affinity resin
PBS	Phosphate buffered saline
paAzoR	Azoreductase from *Pseudomonas aeruginosa*
Phe	Phenylalanine
PSL-2	Lectin from *Phaeolus schweinitzii*
PTP	Protein tyrosine phosphatases
qubits	Quantum bits
RU	Response/Resonance units
SEC	Size exclusion chromatography
Ser	Serine
SD	Standard deviation
SDS-PAGE	Sodium dodecyl (lauryl) sulfate-polyacrylamide gel electrophoresis
SPI-1/2	Salmonella pathogenicity island 1/2
SPR	Surface plasmon resonance
TBS	Tris buffered saline
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
ubiE	Ubiquinone/menaquinone biosynthesis methylase from *Cytophaga hutchinsonii*
ubiH	2-octaprenyl-6-methoxyphenol hydroxylase from *Agrobacterium tumefaciens*
Val	Valine
WGA	Wheat germ agglutinin
$\bar{x}$	Mean

Acknowledgments

I would like to thank my supervisors A/Prof. Joe Tiralongo and Dr. Darren Grice for supporting this project and advising me throughout its progression with additional acknowledgment and thanks to A/Prof. Joe Tiralongo for always keeping an open door and advising me since before this project commenced.

As well I would like to thank Barbara Hadley for her advice regarding FPLC systems and Oren Cooper for his help with SPR and glycan comprehension. I am also thankful for the help of Dr. Christopher Day with SPR systems and SPR data analysis, Dr. Lauren Hartley with help in the array lab and Dr. Tom Litfin for providing docking images and structural alignment data for BT_411 and BT_3781 in this project. I also have great appreciation for Jenni Dyason and Fiona Crone for their invaluable aid with all things administrative.

I am thankful to the Institute for Glycomics for providing a safe and productive atmosphere to work and the Griffith Graduate Research School for allowing me the opportunity to study at Griffith University.

Chapter 1

Introduction

1.1. Bioinformatics

Bioinformatics is a blanket term encompassing any biology related field of study that makes use of computers (via computational modelling software) to analyse large amounts of data in order to solve large scale biological problems and/or develop predictive models relating to biological structures or systems [1]. The term bioinformatics was originally coined in the early 1970s to describe the study of information processes within biotic systems [2,3]; a complementary partner discipline to biochemistry (the study of chemical processes within biotic systems). The field of study evolved to become publicly associated with genomics [4,5] after the race to map the human genome in the 1990s garnered increased public interest in computational genetics [6]. The application of computational modelling as part of biochemical research grew in tandem with the evolution of computer science [7]. A mission statement sent out by Microsoft co-founder Bill Gates to, "put a computer on every desk in every home" began a surge of interest and development into more accessible computer systems [8]. The objective was fully supported by industry with IBM ™, Apple ™ and like manufacturers rushing to make smaller computers with user friendly software that everyone could access [9]. Likewise, advancements in the computational support of research and analysis grew as governments and higher education adopted the implementation of bioinformatic techniques and equipment into their work and research [10]. By the early 2000s the influence of bioinformatic-related applications could be seen across multiple research disciplines, but most notably in STEM related fields [11]. Sequencing and genomics, the disciplines that initiated the prevalence of bioinformatics [6], continued to push the boundaries of computational analysis in an attempt to better analyse the vast wealth of data being produced from sequencing [12]. This was to the advantage of other data heavy research objectives in proteomics, chemistry, statistics, physics, glycobiology and related life

sciences that gladly applied genomics-derived computational algorithms to organize and analyse their own specific data sets [13].

Improvements in computer hardware systems have led to multi-core gigahertz (GHz) central processing units (CPUs) and graphics processing units (GPUs) that are capable of processing billions of instructions per second [14] . Ultra-fast processing systems able to analyse immense quantities of data are now being implemented regularly in *in silico* (using computer software) prediction-based science [15,16]. *In silico* prediction utilizes machine learning and/or artificial intelligence (A.I.) to predict putative molecular interactions and configurations in all aspects of biology [17]. Prediction software is now being implemented across the life sciences with applications in chemical toxicity determination, cell-cell interactions, drug design, predicted protein structure and ligand-receptor affinity [18,19,20]. Recent breakthroughs in quantum computing involve quantum bits (qubits) where the binary computer language of ones and zeros can exist as both a one and zero at the same time, hence the quantum allocation in the name. Google's quantum processor is capable of processing a staggering 2^{53} bits of information per second; in benchmark tests it took their sycamore processor 200 seconds to sample data that would take a non-quantum modern supercomputer ~ 10,000 years to process [21]. Quantum theoretical approaches and quantum computers are a window into the fast-approaching future of bioinformatics, of which protocols with quantum computing in mind are already being planned for chemical and proteomic based predictions and design [22,23].

1.2. Drug discovery and bioinformatics

Drug discovery is the process of identifying, developing and validating drug candidates for the treatment of associated conditions in the fields of medicine, pharmacology and

biotechnology (Figure 1.1.) [24]. The Tufts Center for the Study of Drug Development, a non-profit academic organization that monitors the cost of drug production, released a report in 2016 that estimates the cost of a single drug from its conception to availability on the market, is 2.8 billion (USD) [25]. This estimate involves research, development, marketing and production but also accounts for expenses relating to the development of failed compounds in association with development of the approved drug.

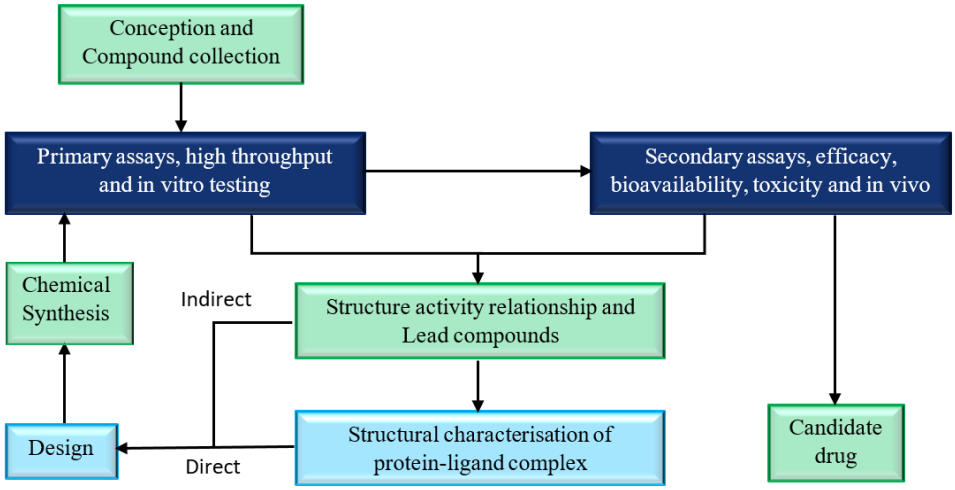


Figure 1.1. Flow chart depicting the drug discovery cycle from conception to approved candidate

Furthermore, the time it takes to get one new drug from concept to customer is on average twelve years, that in addition to ballooning costs equates to drug discovery being an extremely costly and time-consuming endeavour [26]. The drug discovery process is a system that needs improvement as we confront novel biological threats that require expeditious drug treatments at a reasonable price. This is no more evident than with the rising threat of drug resistant microbial pathogens, that require novel drugs and techniques for effective treatment [27, 28]. Drug-resistant microbe-related mortality is estimated at over 700,000 deaths a year due to infection from drug resistant microbes; an estimate that is projected to increase to 10 million by the year 2050 [29]. Furthermore, the pressing need for effective

antibiotics has produced a market in substandard or counterfeit antibiotics [30]. The sellers of these products take advantage of increased demand by manufacturing drug treatments with a greatly reduced amount of the active drug or no active drug at all. Some examples include counterfeit beta-lactams (antibiotic), artemisin and chloroquine (antimalarials) [30]. Thus, the sense of urgency within medical, pharmacological and biotech communities to develop cheap, novel broad-spectrum antimicrobial drugs increases with each new drug-resistant phenotype and counterfeit drug related mortality [31].

Fortunately, there are several paths of exploratory research attempting to resolve both the drug resistant pathogen threat and the ballooning time/cost expenses associated with drug discovery. Some of these include examining the need to provide antibiotic alternatives in poultry and other livestock using synbiotics, prebiotics, probiotics, enzymes, phytogenics, organic acids, antimicrobial peptides, hyperimmune egg antibodies, metals and clay [32,33]. Bioengineering studies into ribosomally synthesized bacteria (LAB) bacteriocins, known to inhibit the growth of closely related bacterial species, are being conducted to produce a human microbiota friendly antibiotic [34,35]. Yet another promising antimicrobial alternative is that of actinobacterial endophytes; a cross species inhibitor of protozoan, bacterial and fungal pathogens that does not harm the host organism it inhabits [36,37].

These and like studies within the greater scientific community show that research is being focused on the problem of finding novel treatments against drug resistant pathogens; however, the need to develop said treatments in a timely and cost-effective manner remains of equal urgency [38]. Enter bioinformatics, the putative saviour of the drug discovery industry [39, 40, 41, 42]. A dramatic statement to be sure but not inaccurate; *in silico* screening, analysis and prediction are beginning to become integrated components of drug discovery and design [43]. Many new and upcoming drug discovery projects are adopting computational

prediction, design and analysis to aid in compound structure analysis and drug interaction identification/prediction [44]. The implementation of machine learning and A.I. based algorithms into the drug design and analysis phase of drug discovery is reducing both the time and cost of developing a new drug (relative to the absence of *in silico* support) [45, 46]. In addition, the development of extensive databases to support bioinformatic tools is as significant as the *in silico* tools themselves [47]. One such example is DrugBank, a comprehensive online library of detailed chemical drug data and relevant biochemical interactions [48]. DrugBank contains over 8,000 drug entries and related targets providing an extensive chemical and pharmacological wealth of accessible data that includes sequence, pathway and structure details of known drug/target interactions. New bioinformatic supportive databases are being developed every year; another example being the Target-Pathogen database. This database is an online resource that integrates detailed protein information including function, metabolic role, structural properties and omic experiments into a searchable library of pathogen drug targets [49]. There are dozens of similar databases both new and old (RCSB PDB, UniProt, ChEMBL, BindingDB) [50], some more generalized and others more focused with regard to their content but all providing easily accessible libraries of data that can be used by bioinformatic screening, prediction and analysis software.

New *in silico* screening applications are developed almost every month with recent examples being FAF-Drugs4, MTiOpenScreen, AMMOS2 and the PyRx Virtual Screening Tool [51]. Some of these tools are more efficient than others creating a need to distinguish the high performers amongst them, that is to say those programs that have the highest track record of success. In order for a new bioinformatic software tool to build a reliable track record, laboratory validation of *in silico* prediction and/or analysis is needed, as predictions

made in virtual environments need to be confirmed in the real world [52]. Only after exhaustive laboratory testing and validation of *in silico* results can a new bioinformatic software tool be considered reliable [53, 54]. SPOT-Ligand is one such virtual sequence and structure based screening application. Developed by Professor Yaoqi Zhou and cohort at the Institute for Glycomics, SPOT-Ligand searches through a combination of databases (ChEMBL, BindingDB, DrugBank, UniProt, RCSB PDB) for structure and sequence similarities between a SPOT-Ligand derived template ligand-receptor pair and a query ligand-receptor pair (Figure 1.2.) [55]. For example, the physical structure and chemical sequence data for a new drug (for example DrugX) is entered into SPOT-Ligand, along with the same data for any known targets of DrugX. SPOT-Ligand will then identify unique chemical sequence and physical structural characteristics about DrugX and its known targets. This information is used by SPOT-Ligand to construct a template that is then used to screen relevant databases for new Drug X targets with similar characteristics as the SPOT-Ligand generated template. SPOT-Ligand target screening has the potential to save time and costs that would otherwise come with testing drug targets one after another, a slow and expensive task when there are thousands of target options [56].

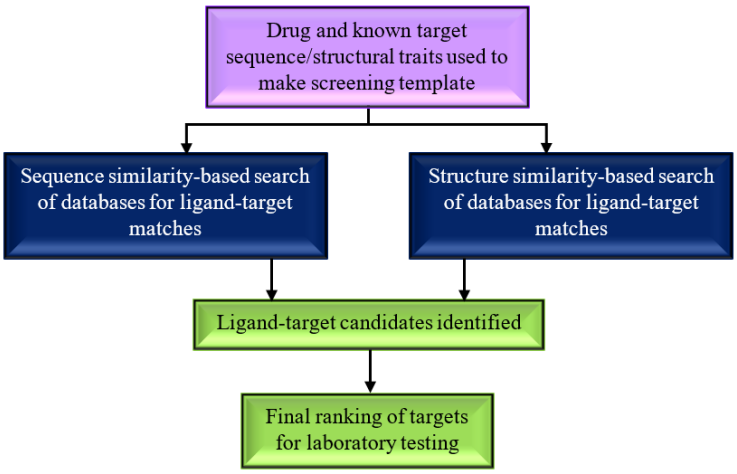


Figure 1.2. Flow diagram showing SPOT-Ligand screening process

It is with the novel antimicrobial drug BDM-I that we come full circle back to the start of this topic, drug discovery and bioinformatics. We emphasized at the beginning of this section that drug discovery is a slow and expensive process; moreover the rapid increase of drug resistant pathogens adds further pressure to the drug discovery cycle, as the demand for novel drugs and/or treatments against emerging drug resistant pathogens increases [57]. BDM-I (Figure 1.3.) is a novel antimicrobial drug created by Australian company BioDiem that displays a broad utility in combating bacterial, protozoal and fungal infections while encountering few resistant phenotypes [58]. BDM-I displayed evidence of protein tyrosine phosphatase (PTP) inhibition as it's mechanism of action; PTP enzymes are key regulatory compounds in signal transduction pathways and cell cycle control and are rarely seen as targets for antimicrobials [58]. However, BDM-I's broad range of activity may imply that inhibition of PTPs is not its only mechanism of action and as such further study to clarify this is warranted.

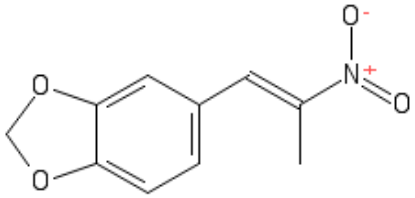


Figure 1.3. Molecular structure of BDM-I

Suitably, SPOT-Ligand is a closed access software local to Griffith university, of where my thesis study took place (allowing me access to the software), that requires further laboratory validation of its screening methods. For these reasons, we proposed in this research to experimentally validate SPOT-Ligand (as opposed to another software) predicted targets for BDM-I to clarify the BDM-I mechanism of action and identify potential new targets for the drug. Moreover, by doing so we will further validate SPOT-Ligand as a new

bioinformatic screening tool and further explore the role of bioinformatics in drug discovery research.

1.3. Protein-carbohydrate interactions and bioinformatics

All cells in multicellular organisms are coated with a dense, intricate layer of glycans collectively called the glycocalyx [59]. These include extra-cellular matrix cells down to enveloped viruses that adopt the glycosylation pattern of the infected cells they bud from [60]. The glycocalyx is essential to multicellular life due to its involvement in cell adhesion, signal transduction, endocytosis, cell protection, receptor activation and other cell-cell interactions [61,62,63]. The glycocalyx is comprised of glycoconjugates, a product of glycosylation; a post-translational (with the exception of co-translational *N*-glycosylation) modification that involves the enzymatic covalent attachment of sugar moieties to proteins and lipids. Glycoconjugates are constructed in the endoplasmic reticulum (ER) and the golgi apparatus, where they are then directed to their final destination within the glycome (the complete carbohydrate profile of the cell) [64]. Glycosylation is critical to life as it is involved in a number of biological processes such as protein folding and structure stability while comparably, abnormal glycosylation can also lead to multiple disorders including forms of congenital muscular dystrophy and the promotion of cancer cell metastasis [65.66.67]. There are several types of glycosylation however, proteins are not limited to one particular type (Table 1.1.). Depending on amino acid sequence, folding conformation (amino acid availability due to protein fold) and enzyme availability, proteins can be glycosylated at multiple sites [68].

Table 1.1. Types of glycosylation

Glycosylation type	Action
N-linked	Glycan binds to asparagine
O-linked	Monosaccharides bind to serine or threonine hydroxyl group
Glypiation	Glycan links a protein and a phospholipid
C-linked	Mannose binds to tryptophan indole ring
Phosphoglycosylation	Glycan binds via a phosphodiester bond to serine

Lectins are a class of carbohydrate binding proteins (CBPs) that reversibly bind to specific cell surface glycoconjugates; they perform a significant role in many cell-cell interactions, acting to initiate vital cellular processes [69]. They are a central field of study amongst glycobiologists given their ubiquitous presence in nature and integral participation in multiple biological systems [70, 71]. Lectins were identified by their cell agglutination and glycan-specific binding properties in the 1960's, when the term lectin became the primary identifier for this type of sugar-binding protein [72]. Subsequent years since have given rise to broad in-depth lectin studies with numerous classes of lectins being identified. The evolution of lectins has led to several distinct families defined by their carbohydrate recognition domains (CRD) [60]. These include *F*-type lectins that are recognizable by a fucose binding domain in a jellyroll fold, possessing calcium and carbohydrate binding motifs specific to this class [73]. *C*-type lectins that are characterised by a *C*-type fold that requires calcium $^{2+}$ to bind carbohydrates [74]. They are the largest and most diverse class of lectins and are found in all organisms with their *C*-like fold present even in proteins without carbohydrate affinity [60]. Galectins specifically bind galactose and are characterised as having a β-sandwich or jelly roll configuration and have been found in a number of biological systems [75]. *I*-type lectins also known as siglecs, specifically bind sialic acids that are often located on cell surfaces, involved with cell-cell interactions; siglec architecture is that of a β-sandwich immunoglobulin like fold [60]. Additionally there are carbohydrate

binding modules (CBMs); proteins with a CBM domain that are typically carbohydrate-activated enzymes. CBM's assist in the hydrolysis of complex sugars, most notable amongst them are the various hydrolases named for their carbohydrate affinity (Fucosidase, GNase, etc.) [60]. There are hundreds of individual lectins from these and other lectin families that have since been isolated and their structures determined [76].

Lectins play key roles in a wide range of disease states and as such, their medical significance is abundant. For example, a characterised lectin from the Indian Hyacinth plant was isolated to test its effects on human colon and cervical cancer cells. Results showed that it triggered apoptosis in these cell lines and may be a viable cytotoxic drug development candidate [77]. An additional study found that the capping (a modification) of *O*-glycan-specific, α2,6 sialyltransferases, can regulate cancer growth and metastasis via the galectin dependent binding of *O*-glycans on cancer cells [78]. A separate study further validated these findings by showing metastasis growth inhibition of prostate cancer cells through the binding of Galectin-4-*O*-Glycan to the cancer cell surface *O*-linked oligosaccharides [79]. Increased focus has also been generated towards lectin anti-viral properties with studies showing that *C*-type lectin receptors are able to identify glycans found on certain viral pathogens and initiate antiviral immune responses via phagocytosis and T-cell activation [80]. Studies into an algae-derived lectin named Griffithsin (GRFT) have shown broad-spectrum action against multiple enveloped viruses, including the inhibition of HIV, adding GRFT to a growing list of HIV lectin based microbicides [81,82]. Moreover, lectin binding preference to cell surface glycoconjugates has led to the development of lectins-based biosensors as analytical tools to identify specific cells by their unique cell surface glycoconjugates, aiding in the early diagnosis of various types of cancer [83,84].

The variety of possible lectin-based applications accordingly led to an increased popularity in lectin research and the identification and characterisation of novel lectins. However, the large amount of available protein data spread across a multitude of databases makes finding novel lectins a difficult and time-consuming process [85]. As with drug discovery, bioinformatic solutions are gaining traction with the use of *in silico* analysis and prediction programs like TargetScan and MiRanda that comb through databases to identify novel lectins based on taxonomic structural similarities with known CBPs [86,87,88]. SPOT-Struc (developed by Prof. Yaoqi Zhou) is one such predictive program that utilises machine learning to search through CBP structural databases to identify CBP physical characteristics including fold, class, topology and ligand affinity [89]. SPOT-Struc then generates a structural CBP template based on the collected data to perform a template-query comparison search against thousands of uncharacterized proteins to identify putative CBPs amongst them. SPOT-Struc has a 79% success rate with predicting positive CBPs, with the most recent reported laboratory validated SPOT-Struc CBP prediction YesU shown to be a novel fucose binding lectin [90]. As such, the combination of *in silico* predictive software like SPOT-Struc with experimental verification provides a time and cost-effective means of identifying and characterizing novel CBPs.

SPOT-Struc, like SPOT-Ligand, is a closed access software local and accessible to Griffith university only, of where my thesis study took place and as such was the tool of choice for this part of my thesis study. Mainly, SPOT-Struc has identified two hypothetical proteins, BT_411 and BT_3781 (hypothetical due to unknown function), from human gut bacterial family *Bacteroides thetaiotaomicron,* which I intended to experimentally validate in conjunction with the SPOT-Ligand predicted BDM-I targets. By doing so I will further explore SPOT-Struc potential as a new bioinformatic prediction tool and define the

supportive role of bioinformatics in CBP identification. In addition to experimentally identifying two novel lectins in BT_411 and BT_3781.

1.4. Project Significance

CBPs are highly specialised reversible binders that are ubiquitous in nature and proven to be instrumental in several biological processes [91,92], including cell-cell signaling, viral pathogen immune response, glycoprotein folding and as agents for novel cancer therapies and lectin-based biosensors [93 - 98]. Therefore, identifying and characterising such novel lectins as BT_411 and BT_3781 (Figure 1.4.) is not solely of academic interest but also raises the potential of identifying two novel lectins for drug research. BT_411 is classed as a putative chitobiase beta protein, a member of the galactose (Gal) binding like domain superfamily having a β-sandwich structure and jellyroll topology (RCSB PDB ID: 3HNM). However, its function is unknown as such BT_411s exact functional title cannot be stated. SPOT-Struc matched BT_411 to a human gastrointestinal pathogen *Clostridium perfringens* family 32 CBM protein with similar architecture to BT_411 and a known affinity for galactose (Gal), N-Acetyl-D-lactosamine (LacNAc) and type II blood group H-trisaccharide (RSCB PDB ID: 2j7m). As such we would expect to see similar carbohydrate affinity in BT_411 to the *C. perfringens* family 32 CBM protein . BT_3781 is classed as an α-protein with an α/α toroid fold, a member of the six-hairpin glycosidases superfamily possessing an α/α barrel structure with glycosyltransferase topology (RSCB PDB ID: 2p0v). Accordingly, SPOT-struc identified BT_3781 as a putative CBP and a likely match to *Hypocrea jecorina* protein Glycoside Hydrolase Family 15 Glucoamylase (RSCB PDB ID: 2vn4) which shows a similar architecture as BT_3781. The carbohydrate affinity of the

Glycoside Hydrolase Family 15 Glucoamylase is yet uncharacterized and therefore insight into BT_3781 may shed light into Glycoside Hydrolase Family 15 Glucoamylase activity.

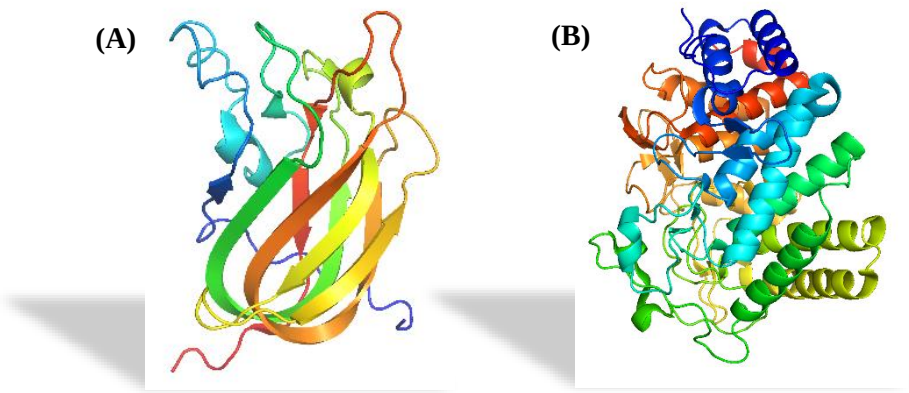


Figure 1.4. Crystal structure of **(A)** BT_411 (PDB ID: 3HNM) and **(B)** BT_3781 (PDB ID: 2p0v)

The need to develop novel broad-range antimicrobial drugs to treat multi-drug resistant pathogens has become imperative [99]. Bacterial infections are one of the leading causes of mortality worldwide, notably in developing countries and amongst infants, the elderly and those with compromised immune systems [100]. Increased drug resistance prolongs patient treatment and suffering, which in turn increases economic burden on both the individual and the society they are a member of [101]. As new pathogens continue to emerge the discovery of new antimicrobial treatments will become instrumental in preventing a global health crisis [102]. As such, characterisation of the new antimicrobial drug BDM-I using SPOT-Ligand screened targets will further elucidate BDM-I's mechanism of action and help progress it towards candidacy as a novel antimicrobial drug.

Bioinformatics research is fast becoming an invaluable means of quickly identifying and predicting both drug targets and novel proteins [103]. The drug discovery cycle is a costly process in both time and resources however, the integration of computer predictive models into experimental research, as directors of that research, will not only save time and money,

but also identify targets that may be otherwise overlooked. By experimentally validating SPOT-Ligand screened BDM-I targets and SPOT-Struc predicted CBPs, we will further clarify the significant role of bioinformatics in research. In addition we will characterise a new and much needed antimicrobial drug BDM-I and experimentally identify two new lectins for future research.

1.5. Project Aims

The bioinformatic screening software SPOT-Ligand has identified six putative archaeal and bacterial targets for a novel antimicrobial drug candidate BDM-I. Furthermore, bioinformatic predictive software SPOT-Struc has identified two putative bacterial lectins from *B. thetaiotaomicron*, BT_411 and BT_3781. As such the aims of this Masters of Medical Research project are:

1. Express in *E. coli* and purify BT_411, BT_3781 and BDM-I targets.
2. Characterise BDM-I target protein affinity by SPR to potentially identify BDM-I's mechanism of action.
3. Characterise BT_411 and BT_3781 using SPR to identify their carbohydrate binding affinity.
4. Evaluate SPOT-Ligand and SPOT-Struc performance based on the results from Aims 2-3.

References

1. Can, T. (2014) Introduction to bioinformatics. *Methods in Molecular Biology*, 1107:51-71. https://doi.org/10.1007/978-1-62703-748-8_4.

2. Hogeweg, P. (2011) The roots of bioinformatics in theoretical biology. *PLoS Computational Biology*, 7(3):e1002021. https://doi.org/10.1371/journal.pcbi.1002021

3. Hogeweg, P. (1978) Simulating the growth of cellular forms. *Simulation*, 31(3):90-96. https://doi.org/10.1177/003754977803100305

4. Perco, P., Rapberger, R., Siehs, C., Lukas, A., Oberbauer, R., Mayer, G. & Mayer, B. (2006) Transforming omics data into context: bioinformatics on genomics and proteomics raw data. *Electrophoresis*, 27:2659-2675. https://doi.org/10.1002/elps.200600064.

5. Chen, J. & Coppola, G. (2018) Bioinformatics and genomic databases. *Handbook of Clinical Neurology*, 147:75-92. https://doi.org/10.1016/B978-0-444-63233-3.00007-5.

6. Hood, L. & Rowen, L. (2013) The human genome project: big science transforms biology and medicine. *Genome Medicine*, 5(9):79. https://doi.org/10.1186/gm483

7. Gauthier, J., Vincent, A.T., Charette, S.J. & Derome, N. (2019) A brief history of bioinformatics. *Briefings in Bioinformatics*, 20(6):1981–1996. https://doi.org/10.1093/bib/bby063

8. Gates, B. (2007) A robot in every home. *Scientific American*, 296(1):58-65. Available from: www.jstor.org/stable/26069115

9. Pfaffenberger, B. (1988) The social meaning of the personal computer: or, why the personal computer revolution was no revolution. *Anthropological Quarterly*, 61(1):39-47. https://doi.org/10.2307/3317870

10. Committee on Innovations in Computing and Communications: Lessons from History, Computer Science and Telecommunications Board, National Research Council, National Academy of Sciences (1999) Funding a revolution: government support for computing research. USA, *The National Academies Press.* Available from: https://www.nap.edu/read/6323/chapter/1

11. Kovarik, D. N. *et al.* (2013) Bioinformatics education in high school: implications for promoting science, technology, engineering, and mathematics careers. *CBE Life Sciences Education*, 12(3):441–459. https://doi.org/10.1187/cbe.12-11-0193

12. Bayat A. (2002) Science, medicine, and the future: bioinformatics. *British Medical Journal (Clinical research ed.)*, 324(7344):1018–1022. https://doi.org/10.1136/bmj.324.7344.1018

13. Yao, T. (2002) Bioinformatics for the genomic sciences and towards systems biology: Japanese activities in the post-genome era. *Progress in Biophysics and Molecular Biology*, 80(1–2):23-42. https://doi.org/10.1016/S0079-6107(02)00011-1

14. Thakur, S., Rai, H.M., Kumar, S. & Pawar, S. (2013) Factors determining the speed and efficiency of a micro-processor in a pc. *International Journal of Emerging Trends in Electrical and Electronics*, 9:61–64. Available from: https://www.researchgate.net/publication/282665356_Factors_Determining_the_Speed_and_Efficiency_of_a_Micro-Processor_in_a_PC/citations#fullTextFileContent

15. Li, F. (2019) Twenty years of bioinformatics research for protease-specific substrate and cleavage site prediction: a comprehensive revisit and benchmarking of existing methods, *Briefings in Bioinformatics*, 20(6):2150–2166. https://doi.org/10.1093/bib/bby077

16. Guan, N. N., Wang, C. C., Zhang, L., Huang, L., Li, J. Q., & Piao, X. (2020) In silico prediction of potential miRNA-disease association using an integrative bioinformatics approach based on kernel fusion. *Journal of cellular and molecular medicine*, 24(1):573–587. https://doi.org/10.1111/jcmm.14765

17. Staples, M., Chan, L., Si, D., Johnson, K., Whyte, C., & Cao, R. (2019) Artificial intelligence for bioinformatics: applications in protein folding prediction. *IEEE Technology & Engineering Management Conference (TEMSCON)*, Atlanta, GA, USA, 1-8. https://doi.org/10.1109/TEMSCON.2019.8813656.

18. Raies, A. B., & Bajic, V. B. (2016) In silico toxicology: computational methods for the prediction of chemical toxicity. Wiley interdisciplinary reviews. *Computational molecular science*, 6(2):147–172. https://doi.org/10.1002/wcms.1240

19. Ruas, F., & Guerra-Sá, R. (2020) In silico prediction of protein-protein interaction network induced by manganese ii in meyerozyma guilliermondii. *Frontiers in microbiology*, 11:236. https://doi.org/10.3389/fmicb.2020.00236

20. Cicaloni, V., Trezza, A., Pettini, F. & Spiga, O. (2019) Applications of in silico methods for design and development of drugs targeting protein-protein interactions. *Current Topics in Medicinal Chemistry*, 19(7):534-554. https://doi.org/10.2174/1568026619666190304153901.

21. Arute, F., Arya, K., Babbush, R. *et al.* (2019) Quantum supremacy using a programmable superconducting processor. *Nature*, 574:505–510. https://doi.org/10.1038/s41586-019-1666-5

22. Lu, L.H. & Li, Y.Q. (2019) Quantum approach to fast protein-folding time. *Chinese Physics Letters*, 36(8):080305. https://doi.org/10.1088/0256-307X/36/8/080305

23. Cao, Y. (2019) Quantum chemistry in the age of quantum computing. *Chemical Reviews*, 119(19):10856-10915. https://doi.org/10.1021/acs.chemrev.8b00803

24. Drews, J. (2000) Drug discovery: a historical perspective. *Science*, 287(5460):1960-1964. https://doi.org/10.1126/science.287.5460.1960

25. DiMasi, J.D., Grabowski, H.G. & Hansenc, R.W. (2016) Innovation in the pharmaceutical industry: new estimates of R&D costs. *Journal of Health Economics*, 47:20-33. https://doi.org/10.1016/j.jhealeco.2016.01.012

26. Mohs, R. C., & Greig, N. H. (2017) Drug discovery and development: role of basic biological research. *Alzheimer's & dementia*, 3(4):651–657. https://doi.org/10.1016/j.trci.2017.10.005

27. Aslam, B., *et al.* (2018) Antibiotic resistance: a rundown of a global crisis. *Infection and drug resistance*, 11:1645–1658. https://doi.org/10.2147/IDR.S173867

28. Ventola C. L. (2015) The antibiotic resistance crisis: part 1:causes and threats. *Pharmacy and Therapeutics*, 40(4):277–283. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4378521/

29. Hrvatin, V. (2017) Combating antibiotic resistance: new drugs or alternative therapies?. *Canadian Medical Association Journal*, 189(37):E1199. https://doi.org/10.1503/cmaj.109-5469

30. Kelesidis, T. & Falagas, ME. (2015) Substandard/counterfeit antimicrobial drugs. *Clinical Microbiology Reviews*, 28(2):443–464. https://doi.org/10.1128/CMR.00072-14

31. Penchovsky, R. & Traykovska, M. (2015) Designing drugs that overcome antibacterial resistance: where do we stand and what should we do? *Expert Opinion on Drug Discovery*, 10(6):631-50. https://doi.org/10.1517/17460441.2015.1048219.

32. Lillehoj, H., Liu, Y., Calsamiglia, S. *et al*. (2018) Phytochemicals as antibiotic alternatives to promote growth and enhance host health. *Veterinary Research*, 49(1):76. https://doi.org/10.1186/s13567-018-0562-6

33. Gadde, U., Kim, W.H., Oh, S.T. & Lillehoj, H.S. (2017) Alternatives to antibiotics for maximizing growth performance and feed efficiency in poultry: a review. *Animal Health Research Reviews*, 18(1):26-45. https://doi.org/10.1017/S1466252316000207

34. Ołdak, A. & Zielińska, D. (2017) Bacteriocins from lactic acid bacteria as an alternative to antibiotics. *Advances in Hygiene and Experimental Medicine (Online)*, 71(0):328-338. https://doi.org/10.5604/01.3001.0010.3817

35. Zhang, S., & Chen, D. C. (2019) Facing a new challenge: the adverse effects of antibiotics on gut microbiota and host immunity. *Chinese medical journal*, 132(10): 1135–1138. https://doi.org/10.1097/CM9.0000000000000245

36. Beiranvand, M., Amin, M., Hashemi-Shahraki, A., Romani, B., Yaghoubi, S., & Sadeghi, P. (2017) Antimicrobial activity of endophytic bacterial populations

isolated from medical plants of Iran. *Iranian journal of microbiology,* 9(1):11–18. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5533999/

37. Martinez-Klimova, E., Rodríguez-Peña, K. & Sánchez, S. (2017) Endophytes as sources of antibiotics. *Biochemical Pharmacology*, 134:1-17. https://doi.org/10.1016/j.bcp.2016.10.010

38. Álvarez-Machancoses, O. & Fernández-Martínez, J.L. (2019) Using artificial intelligence methods to speed up drug discovery. *Expert Opinion on Drug Discovery*, 14:8:769-777. https://doi.org/10.1080/17460441.2019.1621284

39. Xia X. (2017) Bioinformatics and drug discovery. *Current Topics in Medicinal Chemistry*, 17(15):1709–1726. https://doi.org/10.2174/1568026617666161116143440

40. Readhead, B. & Dudley, J. (2013) Translational bioinformatics approaches to drug development. *Advances in Wound Care*, 2(9):470–489. https://doi.org/10.1089/wound.2012.0422

41. Katara, P. (2013) Role of bioinformatics and pharmacogenomics in drug discovery and development process. *Network Modelling Analysis in Health Informatics and Bioinformatics*, 2:225–230. https://doi.org/10.1007/s13721-013-0039-5

42. Gill, S. K., Christopher, A. F., Gupta, V. & Bansal, P. (2016) Emerging role of bioinformatics tools and software in evolution of clinical research. *Perspectives in clinical research,* 7(3):115–122. https://doi.org/10.4103/2229-3485.184782

43. Chen, H. (2018) The rise of deep learning in drug discovery. *Drug Discovery Today*, 23(6):1241-1250. https://doi.org/10.1016/j.drudis.2018.01.039

44. Brown, D. K. & Tastan Bishop, Ö. (2017) Role of structural bioinformatics in drug discovery by computational SNP analysis: analysing variation at the protein level. *Global Heart,* 12(2):151–161. https://doi.org/10.1016/j.gheart.2017.01.009

45. Hingorani, A.D. *et al.*(2019) Improving the odds of drug development success through human genomics: modelling study. *Scientific Reports,* 9:18911. https://doi.org/10.1038/s41598-019-54849-w

46. Law, G.L., Tisoncik-Go, J., Korth, M.J. & Katze, M.G. (2013) Drug repurposing: a better approach for infectious disease drug discovery? *Current Opinion in Immunology*, 25(5):588-592. https://doi.org/10.1016/j.coi.2013.08.004

47. Chen, C., Huang, H. & Wu, C. H. (2017) Protein bioinformatics databases and resources. *Methods in Molecular Biology*, 1558:3–39. https://doi.org/10.1007/978-1-4939-6783-4_1

48. Wishart, D.S., Knox, C., Guo, A.C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z. & Woolsey, J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research,* 34:D668–D672. https://doi.org/10.1093/nar/gkj067

49. Sosa, E.J. *et al.* (2018) Target-Pathogen: a structural bioinformatic approach to prioritize drug targets in pathogens. *Nucleic Acids Research*, 46(1):D413–D418. https://doi.org/10.1093/nar/gkx1015

50. Xu, D., & Xu, Y. (2004) Protein databases on the internet. *Current Protocols in Molecular Biology*, Unit–19:4. https://doi.org/10.1002/0471142727.mb1904s68

51. Mangul S. *et al.* (2019) Challenges and recommendations to improve the installability and archival stability of omics computational tools. *PLoS Biology*, 17(6):e3000333. https://doi.org/10.1371/journal.pbio.3000333

52. Singh, D. *et al*. (2014) Revealing pharmacodynamics of medicinal plants using in silico approach: a case study with wet lab validation. *Computers in Biology and Medicine*, 47:1-6. https://doi.org/10.1016/j.compbiomed.2014.01.003

53. Kant, K., Lal, U. R., & Ghosh, M. (2018) In silico prediction and wet lab validation of arisaema tortuosum (wall.) schott extracts as antioxidant and anti-breast cancer source: a comparative study. *Pharmacognosy magazine*, 13(Suppl 4):S786–S790. https://doi.org/10.4103/pm.pm_69_17

54. Koren, E. *et al*. (2007) Clinical validation of the "in silico" prediction of immunogenicity of a human recombinant therapeutic protein. *Clinical Immunology*, 124(1):26-32. https://doi.org/10.1016/j.clim.2007.03.544

55. Yang, Y., Zhan. J, & Zhou, Y. (2016) SPOT-Ligand: fast and effective structure-based virtual screening by binding homology search according to ligand and receptor similarity. *Journal of Computational Chemistry*, 37(18):1734-9. https://doi.org/10.1002/jcc.24380.

56. Timo, G. O., Reis, R., Melo, A. F., Costa, T., Magalhães, P. O. & Homem-de-Mello, M. (2019) Predictive power of in silico approach to evaluate chemicals against m. tuberculosis: a systematic review. *Pharmaceuticals*, 12(3):135. https://doi.org/10.3390/ph12030135

57. Livermore, D.M. (2004) The need for new antibiotics. *Clinical Microbiology and Infection*, 10(4):1-9. https://doi.org/10.1111/j.1465-0691.2004.1004.x

58. White, K. (2008) Antimicrobial mechanism of action of 3,4-methylenedioxy-beta-nitropropene. *Doctor of Philosophy (PhD), Applied Sciences, RMIT University: Melbourne, Australia.* Available from: https://researchbank.rmit.edu.au/eserv/rmit:7872/White.pdf

59. Yarema, K.J. & Bertozzi, C.R. (2001) Characterizing glycosylation pathways. *Genome Biology 2*, reviews0004:1. https://doi.org/10.1186/gb-2001-2-5-reviews0004

60. Varki, A., Esko, JD. & Colley, KJ. (2009) Cellular organization of glycosylation. essentials of glycobiology: 2nd edition. Cold Spring Harbor (NY): *Cold Spring*

Harbor Laboratory Press. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28876808

61. Monne, H.M.S. *et al.* (2013) Structure of glycocalyx. *Biophysical Journal*, 104(2):251a. https://doi.org/10.1016/j.bpj.2012.11.1412

62. Freeze, H. (2006) Genetic defects in the human glycome. *Nature Reviews Genetics*, 7:537–551. https://doi.org/10.1038/nrg1894

63. Margraf-Schönfeld, S., Böhm, C. & Watzl, C. (2011) Glycosylation affects ligand binding and function of the activating natural killer cell receptor 2B4 (CD244) protein. *Journal of Biological Chemistry*, 286(27):24142–24149. https://doi.org/10.1074/jbc.M111.225334

64. Lodish, H. *et al.* (2000) Molecular cell biology: 4th edition. New York: *W. H. Freeman.* Available from: https://www.ncbi.nlm.nih.gov/books/NBK21744/

65. Reily, C. *et al.* (2019) Glycosylation in health and disease. *Nature Reviews Nephrology*, 15:346–366. https://doi.org/10.1038/s41581-019-0129-4

66. Shental-Bechor, D. & Levy, Y. (2008) Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proceedings of the National Academy of Sciences,* 105(24):8256-8261. https://doi.org/10.1073/pnas.0801340105

67. Ohtsubo, K. & Marth, JD. (2006) Glycosylation in cellular mechanisms of health and disease. *Cell,* 126(5):855-867. https://doi.org/10.1016/j.cell.2006.08.019

68. Varki, A., Esko, JD. & Colley, KJ. (2009) Cellular organization of glycosylation. essentials of glycobiology: 2nd edition. Cold Spring Harbor (NY): *Cold Spring Harbor Laboratory Press*. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28876808

69. Gabius, H.J. *et al.* (2011) From lectin structure to functional glycomics: principles of the sugar code. *Trends in Biochemical Sciences,* 36(6):298-313. https://doi.org/10.1016/j.tibs.2011.01.005

70. Manning, J.C. *et al.* (2017) Lectins: a primer for histochemists and cell biologists. *Histochemistry and Cell Biology*, 147:199–222. https://doi.org/10.1007/s00418-016-1524-6:

71. Coelho, L.C. *et al.* (2018) Lectins as antimicrobial agents. *Journal of Applied Microbiology*, 125(5):1238-1253. https://doi.org/10.1111/jam.14055

72. Sharon, N. & Lis, H. (2004) History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology,* 14(11):53R-62R. https://doi.org/10.1093/glycob/cwh122

73. Vasta, GR., Ahmed, H., Bianchet, M.A., Fernandez-Robledo, J.A. & Amzel, L.M. (2012) Diversity in recognition of glycans by F-type lectins and galectins:

molecular, structural, and biophysical aspects. *Annals of the New York Academy of Sciences*, 1253:E14-26. https://doi.org/10.1111/j.1749-6632.2012.06698.x

74. Dambuza, I.M. & Brown, G.D. (2015) C-type lectins in immunity: recent developments. *Current Opinion in Immunology*, 32:21-27. https://doi.org/10.1016/j.coi.2014.12.002

75. Nagae, M. *et al.* (2006) Crystal structure of the galectin-9 N-terminal carbohydrate recognition domain from mus musculus reveals the basic mechanism of carbohydrate recognition. *Journal of Biological Chemistry*, 281:35884-35893. https://doi.org/10.1074/jbc.M606648200

76. Lam, S. K. & Ng, T. B. (2011) Lectins: production and practical applications. *Applied Microbiology and Biotechnology*, 89(1):45–55. https://doi.org/10.1007/s00253-010-2892-9

77. Naika, S. *et al.* (2017) Biochemical characterisation of lectin from Indian hyacinth plant bulbs with potential inhibitory action against human cancer cells. *International Journal of Biological Macromolecules*, 105:1349–1356. https://doi.org/10.1016/j.ijbiomac.2017.07.170

78. Dimitroff, C.J. (2015) Galectin-binding o-glycosylations as regulators of malignancy. *Cancer Research*, 15:75(16):3195-202. https://doi.org/10.1158/0008-5472

79. Tsai, C.H. *et al.* (2016) Metastatic progression of prostate cancer is mediated by autonomous binding of galectin-4-o-glycan to cancer cells. *Cancer Research*, 76(19):5756-5767. https://doi.org/10.1158/0008-5472.CAN-16-0641

80. Monteiro, J.T. & Lepenies, B. (2017) Myeloid c-type lectin receptors in viral recognition and antiviral immunity. *Viruses*, 9(3):pii:E59. https://doi.org/10.3390/v9030059.

81. Lusvarghi, S. & Bewley, C.A. (2016) Griffithsin: an antiviral lectin with outstanding therapeutic potential. *Viruses*, 8(10):pii:E296. https://doi.org/10.3390/v8100296

82. Koharudin, L.M. & Gronenborn, A.M. (2014) Antiviral lectins as potential HIV microbicides. *Current Opinion in Virology*, 7:95-100. https://doi.org/10.1016/j.coviro.2014.05.006

83. Hendrickson, D. & Zherdev, A.V. (2018) Analytical application of lectins. *Critical Reviews in Analytical Chemistry*, 48(4):279-292. https://doi.org/10.1080/10408347.2017.1422965

84. Belický, Š., Katrlík, J. & Tkáč, J. (2016) Glycan and lectin biosensors. *Essays in Biochemistry*, 60(1):37-47. https://doi.org/10.1042/EBC20150005.

85. Nagasuma R. *et al.* (2006) Lectindb: a plant lectin database, *Glycobiology*, 16(10):938–946. https://doi.org/10.1093/glycob/cwl012

86. Liu, B. *et al.* (2009) Bioinformatics analyses of the mannose-binding lectins from Polygonatum cyrtonema, Ophiopogon japonicus and Liparis noversa with antiproliferative and apoptosis-inducing activities. *Phytomedicine*, 16(6-7):601-8. https://doi.org/10.1016/j.phymed.2008.12.010.

87. Shi, Z., An, N., Zhao, S., Li, X., Bao, J. K. & Yue, B.S. (2013) In silico analysis of molecular mechanisms of legume lectin-induced apoptosis in cancer cells. *Cell Proliferation*, 46(1):86–96. https://doi.org/10.1111/cpr.12009

88. Ismaya, W.T., Yunita, Damayanti, S., Wijaya, C., Tjandrawinata, R.R., Retnoningrum, D.S. & Rachmawati, H. (2016) In silico study to develop a lectin-like protein from mushroom Agaricus bisporus for pharmaceutical application. *Scientia Pharmaceutica*, 84(1):203–217. https://doi.org/10.3797/scipharm.ISP.2015.11

89. Zhao, H., Yang, Y., von Itzstein, M. & Zhou, Y. (2014) Carbohydrate-binding protein identification by coupling structural similarity searching with binding affinity prediction. *Journal of Computational Chemistry*, 35(30):2177-83. https://doi.org/10.1002/jcc.23730.

90. Tiralongo, J., Cooper, O., Litfin, T., Yang, Y., King, R., Zhan, J., Zhao, H., Bovin, N., Day, CJ. & Zhou, Y. (2018) YesU from Bacillus subtilis preferentially binds fucosylated glycans. *Scientific Reports*, 8(1):13139. https://doi.org/10.1038/s41598-018-31241-8.

91. Wu, A.M., Lisowska, E., Duk, M. & Yang, Z. (2009) Lectins as tools in glycoconjugate research. *Glycoconjugate Journal*, 26(8):899-913. https://doi.org/10.1007/s10719-008-9119-7.

92. Lam, S.K. & Ng, T.B. (2011) Lectins: production and practical applications. *Applied Microbiology and Biotechnology*, 89:45. https://doi.org/10.1007/s00253-010-2892-9

93. Caramelo, J.J. & Parodi, A.J. (2007) How sugars convey information on protein conformation in the endoplasmic reticulum. *Seminars in Cell & Developmental Biology*, 18(6):732-742. https://doi.org/10.1016/j.semcdb.2007.09.006

94. Drickamer, K. & Taylor, M.E. (2015) Recent insights into structures and functions of C-type lectins in the immune system. *Current Opinion in Structural Biology*, 34:26-34. https://doi.org/10.1016/j.sbi.2015.06.003.

95. Mason, C.P. & Tarr, A.W. (2015) Human lectins and their roles in viral infections. *Molecules*, 20(2):2229-71. https://doi.org/10.3390/molecules20022229.

96. Pearse, B.R. & Hebert, D.N. (2010) Lectin chaperones help direct the maturation of glycoproteins in the endoplasmic reticulum. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1803(6):684-693. https://doi.org/10.1016/j.bbamcr.2009.10.008

97. Rodriguez, E., Schetters, S.T.T. & Van Kooyk, Y. (2018) The tumour glyco-code as a novel immune checkpoint for immunotherapy. *Nature Reviews Immunology*, 18: 204–211. https://doi.org/10.1038/nri.2018.3

98. Dan, X., Liu, W. & Ng, T.B. (2016) Development and Applications of Lectins as Biological Tools in Biomedical Research. *Medicinal Research Reviews*, 36(2):221-47. https://doi.org/10.1002/med.21363.

99. Allen, H.K., Trachsel, J., Looft, T. & Casey, T.A. (2014) Finding alternatives to antibiotics. *Annals of the New York Academy of Sciences*, 1323:91-100. https://doi.org/10.1111/nyas.12468

100. Frieri, M., Kumar, K. & Boutin, A. (2017) Antibiotic resistance. *Journal of Infection and Public Health*, 10(4):369-378. https://doi.org/10.1016/j.jiph.2016.08.007

101. Gandra, S., Barter, D.M. & Laxminarayan, R. (2014) Economic burden of antibiotic resistance: how much do we really know? *Clinical Microbiology and Infection*, 20(10):973-80. https://doi.org/10.1111/1469-0691.12798.

102. Talebi Bezmin Abadi, A. *et al.* (2019) World health organization report: current crisis of antibiotic resistance. *BioNanoScience*, 9:778–788. https://doi.org/10.1007/s12668-019-00658-4

103. Chang, J.W., Zhou, Y.Q., Ul Qamar, M.T., Chen, L.L. & Ding, Y.D. (2016) Prediction of protein-protein interactions by evidence combining methods. *International Journal of Molecular Sciences*, 17(11):1946. https://doi.org/10.3390/ijms17111946

Chapter 2

Characterisation of putative BDM-I targets predicted using SPOT-Ligand

Abstract

The drug discovery cycle is a time consuming and costly process equating to billions in capital expenditure that takes a decade or more to produce a single drug. A consequence of this process is the slow and limited production of novel antimicrobial drugs that are needed to quell the global rise of drug resistant pathogens. One possible solution is the implementation of bioinformatic screening and analysis software into the drug discovery cycle with the intention of reducing failed drug prospects and quickly finding viable drug candidates. SPOT-Ligand is a bioinformatic screening tool that searches through multiple databases to find novel drug targets that can be used for the characterisation of novel drugs. Here we characterise the mechanism of action for the novel antimicrobial drug BDM-I by experimentally examining the binding affinity of six SPOT-Ligand screened drug targets. The drug targets were expressed in *E. coli*, purified using affinity chromatography and SEC; underwent quality control using CD spectroscopy and analysed using SPR for binding affinity with BDM-I. SPR and *in silico* docking analysis revealed high BDM-I affinity for *S. typhimurium* protein AcpD, validating the potential use of *in silico* screening in support of drug discovery research.

2.1. Introduction

Modern drug discovery remains a costly [1] and competitive industry that requires consistent refinement, innovation and the development of novel drug candidates to keep abreast of patient, market and regulatory demands [2]. These demands are now often achieved using *in silico* design [3], prediction [4] and analysis [5] in conjuncture with laboratory validation [6,7]. As is the case with the search for new antimicrobial agents [8]; a rising focus of drug discovery research aimed at combating the evolution of antimicrobial resistant pathogens [9].

SPOT-Ligand is a virtual sequence and structure-based screening method that searches through a combination of databases (ChEMBL, BindingDB, DrugBank, UniProt, RCSB PDB) for binding similarities between query ligand-receptor pairs and a template ligand-receptor pair [10]. This type of virtual screening can identify putative receptors for any particular ligand using structural (both known or sequence predicted) similarities with a known template ligand-receptor pair. The SPOT-Ligand screening technique was performed prior to my study by Prof. Yaoqi Zhou and cohort on the novel antimicrobial drug BDM-I developed by Australian biotech company BioDiem. BDM-I has been shown to have an inhibitory effect against a wide range of bacterial and fungal infections (Table 2.1.) as well as anti-protozoal activity against *Trichomonas vaginalis* and intracellular sporozoan pathogens *Plasmodium* and *Eimeria* [11].

Table 2.1. Minimum inhibitory concentrations (MIC) of BDM-I for bacterial and fungal targets [11]

Bacteria Gram (-)	µg/mL	Bacteria Gram (+)	µg/mL	Fungi	µg/mL
Haemophilus influenzae	0.3	*Streptococcus pyogenes*	2	*Trichophyton rubrum*	1
Pasteurella multocida	2	*Clostridium perfringens*	2	*Pseudoallescheria boydii*	2
Neisseria gonorrhoeae	5	*Staphylococcus aureus A*	6	*Fusarium graminearum*	2
Proteus vulgaris	8	*Corynebacterium xerosis*	8	*Aspergillus fumigatus*	3
Moraxella catarrhalis	16	*Enterococcus faecalis A*	16	*Rhizopus stolonifer*	4
Yersinia enterocolitica	*16*	*Lactobacillus casei*	64	*Rhodotorula rubra*	8

Long term exposure of BDM-I (12-16 weeks) to bacterial pathogens with known ability to develop multi-drug resistance (*Staphylococcus aureus* and *Enterococcus*) did not induce drug-resistant phenotypes. BDM-I appears to inhibit metabolic pathways involved with survival responses (swarming motility in *P. vulgaris* at 32 µg/mL, *Proteus mirabilis* at 64 µg/mL and pigment production in *Serratia marcescens*) by targeting protein tyrosine phosphatases (PTPs) [11]. The PTP family of enzymes catalyse the dephosphorylation of protein tyrosines and are key regulatory components of cell-cycle control and signal transduction pathways [12]. Moreover, BDM-I is able to enter the cell and target intracellular pathogens without affecting ATP production, the cytoplasmic membrane, cell wall and cell-free protein synthesis. PTP targeting anti-microbial agents that do not inhibit other cellular processes are rare [13] and given the broad range of activity BDM-I displays, it is unlikely that PTPs are its sole target of inhibition [11].

To better clarify BDM-I's mechanism of action and validate the use of *in silico* screening in drug discovery, SPOT-Ligand identified six putative BDM-I targets (Table 2.2.). Targets were chosen if they had a > 60% match to the SPOT-Ligand template, have already been purified by another group showing a primarily soluble recombinant protein product and having similar HIS-tagged expression vectors to facilitate simultaneous mass expression and purification of all proteins.

Table 2.2. BDM-I putative targets

UniProtKB ID - Gene Symbol - Name	Species	Domain
Q838M8 - EF0414 - DadA family oxidoreductase	*Enterococcus Faecalis*	Bacteria Gram (+)
Q11QM3 - ubiE - ubiquinone/menaquinone biosynthesis methylase	*Cytophaga hutchinsonii ATCC 33406*	Bacteria Gram (-)
Q9HHS0 - ftsZ5 - cell division protein	*Halobacterium sp. NRC-1*	Archaea
C7NAN5 - Lebu_1328 - Tubulin/FtsZ GTPase	*Leptotrichia buccalis C-1013-b*	Bacteria Gram (-)
P63462 - acpD - azoreductase	*S. typhimurium LT2*	Bacteria Gram (-)
A9CIM2 - ubiH - 2-octaprenyl-6-methoxyphenol hydroxylase	*Agrobacterium tumefaciens*	Bacteria Gram (-)

Here I describe the experimental validation of SPOT-Ligand screened putative targets for BDM-I using SPR to identify BDM-I protein target affinity. I then performed docking analysis on positive BDM-I drug candidates to determine possible BDM-I mechanism of action. I confirm that SPOT-ligand accurately predicted high binding to AcpD, an azoreductase from *Salmonella typhimurium* (strain LT2). Moreover, I identified a new potential BDM-I mechanism of action as a nitroreductase inhibitor seen by the predicted binding of BDM-I nitro-group to AcpD; which likewise reveals previously unknown nitroreductase activity in the known azoreductase AcpD.

2.2. Materials and Methods

2.2.1. Expression and purification of recombinant BDM-I targets EF0414, UbiE, FtsZ5, Lebu_1328, AcpD, and UbiH

Drug target vectors (Table 2.3.) were purchased from the DNASU plasmid repository (http://dnasu.org). Vectors were transformed into competent *E. coli* BL21 (DE3) cells to be used for recombinant protein expression.

Table 2.3. BDM-I drug target identification details; DNASU ID, Vector, Tag and promoter.

Gene	Residues	DNASU ID	Vector	Tag	Promoter
EF0414	371	EfCD00339696	pET21_NESG	*C*-terminal 6xHis	T7 Lac
ubiE	227	ChCD00333821	pMCSG19	*C*-terminal 6xHis	T7 Lac
ftsZ5	225	HsCD00339215	pET21_NESG	*C*-terminal 6xHis	T7 Lac
Lebu_1328	269	LbCD00519810	pMCSG48	*N*-terminal 8xHis	T7 Lac
acpD	201	StCD00584746	pMCSG7	*C*-terminal 6xHis	T7 Lac
ubiH	402	AtCD00540476	pET21_NESG	*C*-terminal 6xHis	T7 Lac

Overnight cultures of BL21 (DE3)/EfCD00339696, BL21 (DE3)/ChCD00333821, BL21 (DE3)/HsCD00339215, BL21 (DE3)/LbCD00519810, BL21 (DE3)/StCD00584746 and BL21 (DE3)AtCD00540476 were utilized to inoculate LB broth (ThermoFisher) containing ampicillin (75 μg/mL). The inoculated LB broth was then incubated at 37^{o}C with aeration until $OD_{600\ nm}$ reached 0.5 to 0.7, at which point expression was induced using 1 mM IPTG. Temperature was reduced to 16^{o}C for 4 hours incubation for the targets EF0414, FtsZ5, UbiH while the targets UbiE, Lebu_1328 and AcpD remained at 37^{o}C for a further 4 hours incubation as determined by small scale expression I performed prior to the commencement of this study. The resulting cell pellets were resuspended in PBS lysis buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4; pH 7.5), containing lysozyme (2 mg/mL), DNaseI and protease inhibitor cocktail mix (20 μL). Cell lysis was then performed by sonication with cellular debris removed via centrifugation at 16 000 x g for 15 min. The refined supernatants containing EF0414, UbiE, FtsZ5, Lebu_1328, AcpD, and UbiH were mixed separately with 1 mL of HIS-select nickel affinity resin (Sigma). The slurry mix of each protein was then packed by gravity into separate 10 mL Bio-Rad chromatography columns. Columns were washed once with PBS lysis buffer, then washed with high salt buffer (1 M NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4; pH 7.5) followed by the final elution wash of the bound HIS-tagged protein using Imidazole elution buffer (300

mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 500 mM Imidazole; pH 7.5) in 1 mL volume. The eluted protein solutions were further purified, and Imidazole removed, using size SEC on the AKTA pure FPLC system (GE Healthcare) with a 24 mL Superdex 200 10/300 gel filtration column (GE Healthcare). Column was equilibrated with PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4; pH 7.5) and calibrated using blue dextran 2000 (to determine void volume), Conalbumin (75 kDa), Ovalbumin (43 kDa) and Ribonuclease-A (13.7 kDa). Calibration standard elution volumes were used to generate a calibration curve to estimate the molecular weight (Mw) of the target proteins in solution. Proteins were eluted using PBS (pH 7.5) and monitored at 280 nm with 0.75 mL fractions collected and pooled according to chromatogram elution peaks. Each pooled sample was tested for presence of target protein by dot blot using a primary mouse monoclonal anti-His6 (1:10,000 dilution, Cell Signaling Technologies) antibody and secondary goat anti-mouse horseradish peroxidase conjugated (1:10,000 dilution, Bio-Rad Laboratories) antibody. SEC protein elution volumes were used to calculate phase distribution coefficient (K_{av}) for each protein (equation shown in Figure 2.1.); K_{av} was then inputted into the calibration standards straight-line equation to calculate protein Mw.

$$\mathbf{K_{av}} = \frac{\textbf{Elution Volume (VE)} - \textbf{Void Volume (VO)}}{\textbf{Column Volume (Vc)} - \textbf{Void Volume (VO)}}$$

Figure 2.1. Equation used to calculate the phase distribution coefficient (K_{av}) with protein elution volume, void volume (determined using blue dextran) and column volume (24 mL)

Purified protein was concentrated using Amicron 10 K 50 mL centrifugal filter (2 mg/mL) with purity confirmed by SDS-PAGE using 4-20% gradient SDS-polyacrylamide gels (Bio-Rad) stained with Coomassie brilliant blue and western blot immunodetection using a primary mouse monoclonal anti-His6 (1:10,000 dilution, Cell Signaling Technologies)

antibody and secondary goat anti-mouse horseradish peroxidase conjugated (1:10,000 dilution, Bio-Rad Laboratories) antibody. SDS-PAGE protein samples were denatured chemically and by heat (95^{o}C for 5 min) with any potential disulfide bonds reduced using 2-mercaptoethanol. Figure 2.2. shows a summary of EF0414, UbiE, FtsZ5, Lebu_1328, AcpD, and UbiH expression and purification.

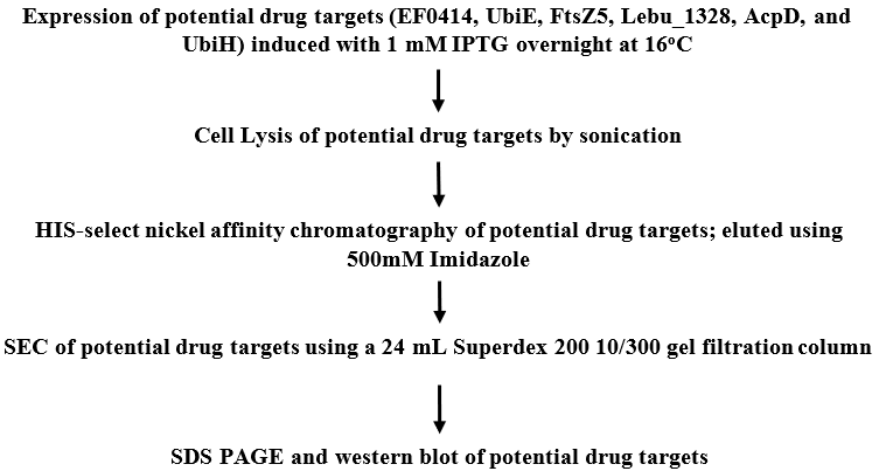


Figure 2.2. Flow chart depicting expression and purification of EF0414, UbiE, FtsZ5, Lebu_1328, AcpD, and UbiH

2.2.2. Protein estimation

Protein concentration was determined by colorimetric assay using a BCA protein assay kit (Thermo scientific) with 25 µL of BSA (concentration range between 2 mg/mL and 0 mg/mL) loaded in duplicate to each appropriate well of a 96-well microtiter flat bottom plate. Twenty-five µL of pure, 1/10 and 1/100 samples of EF0414, UbiE, FtsZ5, Lebu_1328, AcpD, and UbiH were loaded into appropriate wells on the same 96-well plate followed by the addition of 200 µL of working reagent (Thermo scientific) to each well containing BSA and protein sample which was then incubated at 37^{o}C for 40 min. Absorbance was measured at 575 nm using a Bio-Rad xMark™ Microplate Absorbance Spectrophotometer with concentrations calculated using Microplate Manager® Software

(based on the line equation generated from the BSA concentration range and sample dilution factor).

2.2.3. Quality control check of proteins using CD spectroscopy

CD spectroscopy was performed using the Jasco J-1500 Circular Dichroism Spectrophotometer as a quality control step to determine if a protein was denatured (as opposed to analyses of secondary structures, which was not performed) before committing to characterisation phase. This was done by comparing the CD spectra of a non-denatured protein versus the denatured equivalent and if they matched it was inferred the protein was denatured and not suitable for SPR analyses. Purified EF0414, UbiE, FtsZ5, Lebu_1328, AcpD, and UbiH (25 mg/mL) were added to a low conductivity buffer (100 mM KH_2PO_4) and placed in a 1 mm quartz cuvette. Protein measurement parameters were set at standard sensitivity between 260 nm and 190 nm with a 1 mm cell path using an accumulation of 9 (average of 9 repeated scans). Low conductivity buffer without protein was used as a baseline. Purified EF0414, UbiE, FtsZ5, Lebu_1328, AcpD, and UbiH (25 mg/mL) were then denatured using 20 % SDS (w/v), Tris-HCl (0.5 M), 2-mercaptoethanol and by heat (95^{o}C for 5 min) and measured under the same parameters in low conductivity buffer.

2.2.4. Characterisation of recombinant BDM-I targets EF0414, UbiE, FtsZ5, Lebu_1328, AcpD, and UbiH

SPR Detection

SPR detection was conducted using a BIAcore T200 Biosensor system (GE Healthcare) at 25^{o}C in 10 mM PBS (pH 7.4) at a flow rate of 20 µL/min (30 µL/min for regeneration).

Purified EF0414, UbiE, FtsZ5, Lebu_1328, AcpD, and UbiH were diluted to 2 μg/mL in PBS (pH 7.4) with 10% DMSO and immobilized onto flow cells (FC) 2, 3 and 4 (respectively) of two BIAcore CM5 sensor chips (3 proteins per chip) using amine covalent coupling with 10 min contact time at a flow rate of 5 μL/min. FC 1 on each chip contained no protein and was used as a blank reference. Triplicate cycles were performed for each drug target to establish validity of the SPR response. Five-fold serial dilutions (5 to 0.008 μM) of BDM-I were prepared in PBS (pH 7.4) with 10% DMSO and loaded onto the sensor chip to be assessed using single-cycle kinetics with the CM5 chip regenerated using EDTA after each series injection. Post-regeneration the chip was reloaded with drug targets prior to the injection of the next dilution series. Drug target binding affinity was recorded by the response signal difference between protein loaded FC's and the blank reference of FC 1. Injection of each analyte concentration was followed by 300 s of dissociation. T200 BIAcore Evaluation software was used to analyse SPR signals, equilibrium dissociation constants (K_D) as determined from steady state analysis. Results were considered reliable for affinity if all 3 runs showed a standard deviation (SD) < 1 of $\bar{x}\, K_D$ (μM).

Multicycle kinetics analysis performed on drug target AcpD under the same conditions as described above, with the exception of a change in analyte concentration consisting of five-fold serial dilutions (500 to 0.8 μM) of BDM-I prepared in PBS (pH 7.4) with 10% DMSO with association (k_a) and disassociation (k_d) kinetics measured.

BDM-I docking analysis with AcpD

Docking of BDM-I to AcpD was performed on the drug target in a 50 Å box around predicted binding pocket using Autodock Vina with default parameters. The putative binding site were chosen based on close proximity to the location of Flavin mononucleotide (FMN)

cofactor (FMN already defined in AcpD crystal structure, RCB PDB ID: 1T5B) that is required for substrate reduction activity [14].

2.3. Results

2.3.1. Expression and purification of recombinant BDM-I targets EF0414, UbiE, FtsZ5, Lebu_1328, AcpD, and UbiH

Recombinant drug targets EF0414, UbiE, FtsZ5, Lebu_1328, AcpD, and UbiH were expressed in *E. coli* BL21 (DE3) and purified to homogeneity as described in section 2.2.1. Figure 2.3. shows the successful purification to homogeneity of all proteins by HIS-select nickel affinity chromatography and SEC. HIS-tagged EF0414 is visible as a single ~ 41 kDa band on SDS-PAGE (lane 4, Coomassie stain and lane 5, western blot); UbiE visible as a single ~26 kDa band (lane 9, Coomassie stain and lane 10, western blot); FtsZ5 as a single ~ 23 kDa band (lane 13, Coomassie stain and lane 14, western blot); Lebu_1328 as a single ~ 30 kDa band (lane 18, Coomassie stain and lane 19, western blot); AcpD as a single ~ 22 kDa band (lane 22, Coomassie stain and lane 23, western blot); and UbiH visible as a single ~ 43 kDa band on SDS-PAGE (lane 26, Coomassie stain and lane 28, western blot). Total quantity of purified BDM-I drug target proteins per litre of *E.coli* culture following purification was 0.08 mg for EF0414, 1.34 mg of UbiE, 1.51 mg of FtsZ5, 4.37 mg of Lebu_1328, 2.42 mg of AcpD and 0.21 mg of UbiH.

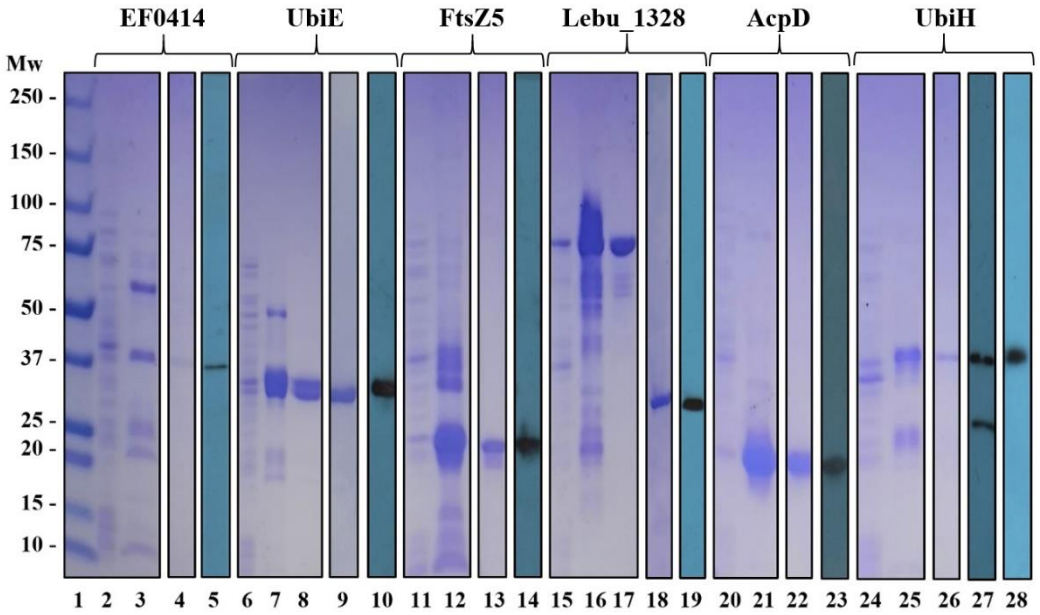


Figure 2.3. Coomassie stained gels and associated western blots showing the purification of recombinant BDM-I targets. Purified protein concentrated using Amicron 10 K 50 mL centrifugal filter to 2 mg/mL. Lane 1, Bio-Rad precision plus protein Mw standard; Lane 2, soluble recombinant EF0414; Lane 3, Ni^{+}NTA purified EF0414 elution; Lane 4, concentrated SEC purified EF0414 protein; Lane 5, western blot of purified EF0414 protein (371 residues) at the expected Mw of ~ 41 kDa; Lane 6, soluble recombinant UbiE; Lane 7, Ni^{+}NTA purified UbiE elution; Lane 8, concentrated SEC purified UbiE protein; Lane 9, rerun of concentrated SEC purified UbiE protein to eliminate doublet (due to possible gel inconstancies) seen in lane 8; Lane 10, western blot of purified UbiE protein (227 residues) at the expected Mw of ~ 26 kDa; Lane 11, soluble recombinant FtsZ5; Lane 12, Ni^{+}NTA purified FtsZ5 elution; Lane 13, concentrated SEC purified FtsZ5 protein; Lane 14, western blot of purified FtsZ5 protein (225 residues) at the expected Mw of ~ 23 kDa; Lane 15, soluble recombinant Lebu_1328 protein; Lane 16, Ni^{+}NTA purified Lebu_1328 elution; Lane 17, concentrated SEC purified Lebu_1328 protein; Lane 18, rerun of concentrated SEC purified Lebu_1328 protein under increased denaturing conditions to resolve non-denatured sample in lane 17; Lane 19, western blot of purified Lebu_1328 protein (268 residues) at the expected Mw of ~ 30 kDa. Lane 20, soluble recombinant AcpD; Lane 21, Ni^{+}NTA purified AcpD elution; Lane 22, concentrated SEC purified AcpD protein; Lane 23, western blot of purified AcpD protein (201 residues) at the expected Mw of ~ 22 kDa; Lane 24, soluble recombinant UbiH; Lane 25, Ni^{+}NTA purified UbiH elution; Lane 26, concentrated SEC purified UbiH protein; Lane 27, western blot of purified UbiH protein (402 residues) at the expected Mw of ~ 43 kDa with artefact at ~ 27 kDa; Lane 28, rerun of western blot of purified UbiH protein (402 residues) at the expected Mw of ~ 43 kDa to resolve artefact (from contamination of previous loading sample) in lane 27 (confirmed by elimination of said artefact using fresh sample).

CD spectroscopy measures changes in light absorption of left and right-handed circularly polarized light, when that light interacts with asymmetric molecules [15]. In the case of protein analysis, CD spectroscopy is generally used in the identification of secondary structures and folding properties, represented by waves on a CD spectra. However, this was not how CD spectroscopy was used in this study, but rather it was utilized as a quality control step to determine if a protein was denatured (as opposed to analyses of secondary structures, which was not performed) before committing to characterisation phase. This was done by comparing the CD spectras of a non-denatured protein versus the denatured equivalent and if they matched it was inferred the protein was denatured and not suitable for SPR analyses. All proteins were confirmed to be non-denatured and ready for SPR. (Figure A1.1.).

SEC purification was performed and used to derive a possible estimation of native Mw for BDM-I target recombinant proteins. Pooled fractions associated with visible peaks in protein chromatograms were tested for presence of pure protein using dot blot immunodetection, as described in section 2.2.1. Mw was estimated as described in section 2.2.1, using the straight-line equation generated from the Mw standard calibration curve (Figure 2.4.).

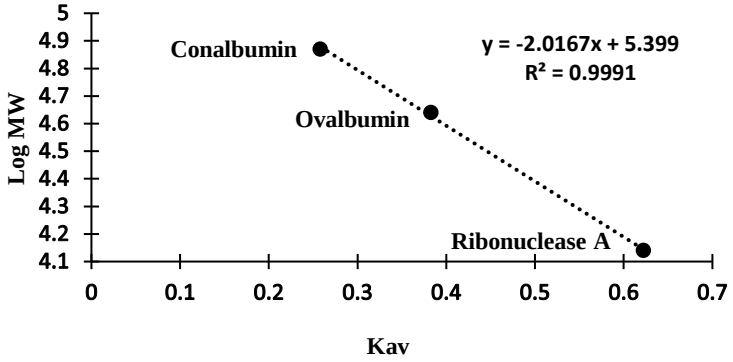


Figure 2.4. Standard Mw calibration curve and trendline. Protein standards Conalbumin (75 kDa), Ovalbumin (43 kDa) and Ribonuclease-A (13.7 kDa) eluted with 20 mM PBS (pH 7.4.) by SEC using a Superdex 200 Increase 10/300 GL column and monitored at 280 nm.

SEC is a further purification step but as described in section 2.2.1, it can also be used to estimate possible oligomeric state of a protein in solution by simply dividing the SEC-estimated Mw by the protein Mw as determined by amino-acid sequence calculated mass or to a less accurate degree, SDS-page inferred Mw. Figure 2.5. shows the SEC chromatograms for all BDM-I drug targets with Figure 2.5.A revealing EF0414 is possibly monomeric in solution, with a SEC Mw of 38.8 kDa (which correlates to the Mw inferred by the protein band visible in Figure 2.3., lane 4). Figure 2.5.B shows UbiE to also possibly be monomeric in solution with a SEC Mw of 29.7 kDa (which correlates to the Mw inferred by the protein band visible in Figure 2.3., lane 9). Figure 2.5.C suggests FtsZ5 may be dimeric in solution, with a SEC Mw of 59.7 kDa (correlating to the Mw inferred by the protein band visible in Figure 2.3., lane 13). Lebu_1328 has a SEC Mw of 94.9 kDa, which correlates to the Mw inferred by the protein band visible in Figure 2.3. lane 18, suggesting that it exists as a trimer in solution (Figure 2.5.D). Drug target AcpD is likely a dimer in solution (Figure 2.5.E), with a SEC Mw of 38.4 kDa, correlating to the Mw inferred by the protein band visible in Figure 2.3. lane 22. Lastly, UbiH has a SEC Mw of 43.6 kDa, correlating to the Mw inferred by the protein band visible in Figure 2.3. lane 26, revealing that ubiH is likely monomeric in solution (Figure 2.5.F). AcpD SEC estimated oligomeric state matches its crystal derived oligomeric state as a dimer (RCSB PDB ID: 1T5B); EF0414, UbiE, FtsZ5, Lebu_1328 and UbiE structures are sequence predicted and as such oligomeric state was not previously known.

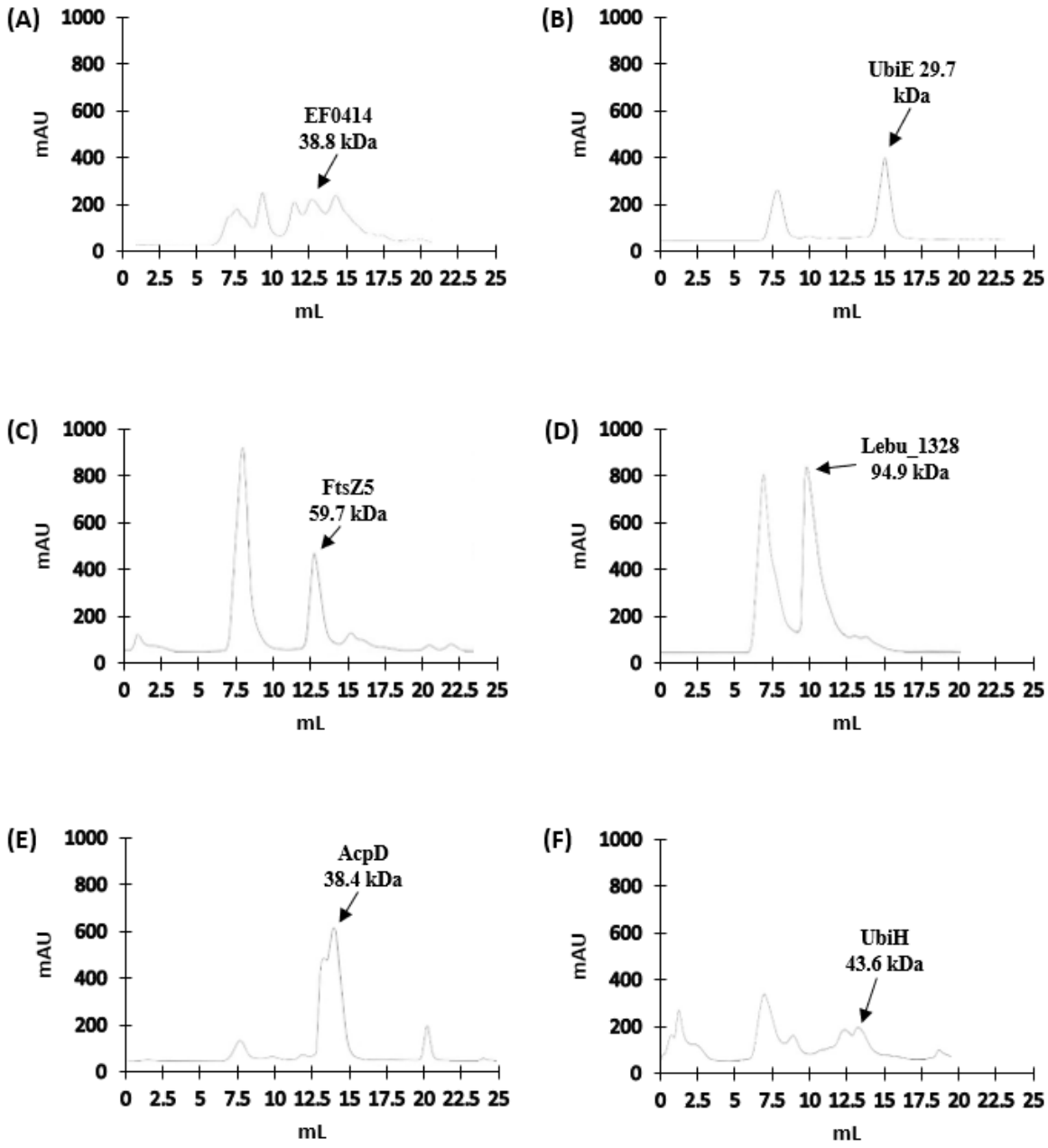


Figure 2.5. SEC chromatograms showing protein elution using a Superdex 200 Increase 10/300 GL column. Column was equilibrated with 10 mM PBS (pH 7.5) and chromatography monitored at 280 nm. **(A)** Purified EF0414 with a SEC Mw of 38.8 kDa; **(B)** Purified UbiE with a SEC Mw of 29.7 kDa; **(C)** Purified FtsZ5 with a SEC Mw of 59.7 kDa; **(D)** Purified Lebu_1328 with a SEC Mw of 94.9 kDa; **(E)** Purified AcpD with a SEC Mw of 38.4 kDa; **(F)** Purified UbiH with a SEC Mw of 43.6 kDa

Table 2.4. shows the theoretical, SDS PAGE and SEC Mw of pure EF0414, UbiE, FtsZ5, Lebu_1328, AcpD, and UbiH as well as oligomeric state and total protein yield.

Table 2.4. Drug targets theoretical & SEC estimated Mw, inferred oligomeric state and total yield

Gene	Theoretical Mw (kDa)	SDS PAGE Mw (kDa)	SEC Mw (kDa)	Oligomeric State*	Total Yield (mg/mL)
EF0414	40.97	37-40	38.82	Monomer	0.078
ubiE	25.82	27-32	29.69	Monomer	1.344
ftsZ5	23.34	21-24	59.70	Dimer	1.507
Lebu_1328	29.72	26-31	94.91	Trimer	4.372
acpD	21.62	21-24	38.37	Dimer**	2.421
ubiH	43.14	40-45	43.62	Monomer	0.210

* Estimated for recombinant protein in solution based on SEC

** Same as oligomeric state from known crystal structure

2.3.2. Characterisation of recombinant BDM-I targets EF0414, UbiE, FtsZ5, Lebu_1328, AcpD, and UbiH

SPR Analysis

SPR defines how strong the binding interaction is between two molecules through the value of K_D, which represents the concentration of analyte at which half of all binding sites are occupied. In regard to drug target analysis, a lower K_D is optimal, as it indicates that a "low" concentration of drug is needed to bind the target protein and inhibit its activity. SPR single-cycle affinity drug target analysis for BDM-I showed strong binding with *S. typhimurium* (strain LT2) protein AcpD (K_D: 0.58 µM ± 0.04). Drug targets EF0414 and UbiH did show initial weak binding (K_D > 15.00 µM; Figure A1.2.) however, SPR results were not reproducible in the third run; while drug targets UbiE, FtsZ5 and Lebu_1328 did not bind (Figure A1.3.).

Figure 2.6. shows the representative (best K_D of the triplicate runs) SPR single-cycle sensorgram (Figure 2.6.A) and affinity plot (Figure 2.6.B) for protein AcpD.

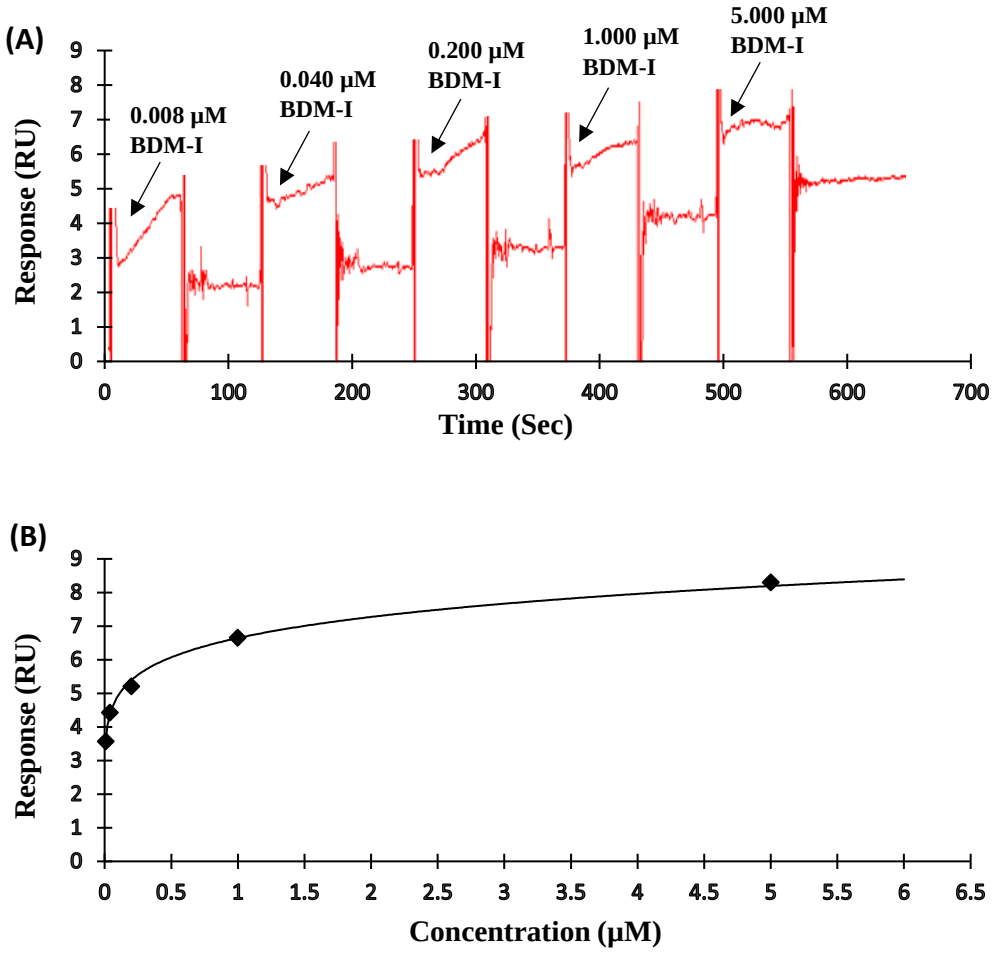


Figure 2.6. Representative (best K_D of the triplicate runs) **(A)** AcpD SPR single-cycle sensorgram with BDM-I concentration injection points (indicated by arrows) and **(B)** related affinity plot

SPR binding kinetics is another important measurement in the characterisation of a potential drug as it defines how fast the drug binds to its target (k_a) and how fast it disassociates (k_d). A large proportion of approved and clinically effective inhibitors of drugs, such as pain medications [16], bind quickly and are slow to release with the exception of drugs that have toxic side effects like anticancer agents, where fast dissociation is

preferred (in most cases) to limit toxicity [17]. SPR multicycle kinetics analysis of AcpD with BDM-I showed a K_D of 0.01 µM with a k_a of 9.4 x 10^4 $M^{-1}s^{-1}$ and a k_d of 9.7 x 10^{-4} s^{-1}, indicating that BDM-I bound to AcpD quite quickly and was slow to dissociate. Figure 2.7. shows the SPR multicycle sensorgram (Figure 2.7.A) and affinity plot (Figure 2.7.B) of BDM-I binding to AcpD at a concertation range from 500 µM to 0.8 µM. The kinetics K_D of 0.01 µM indicates strong binding at minimal concentrations, as was also seen with single single-cycle analysis (Figure 2.6.).

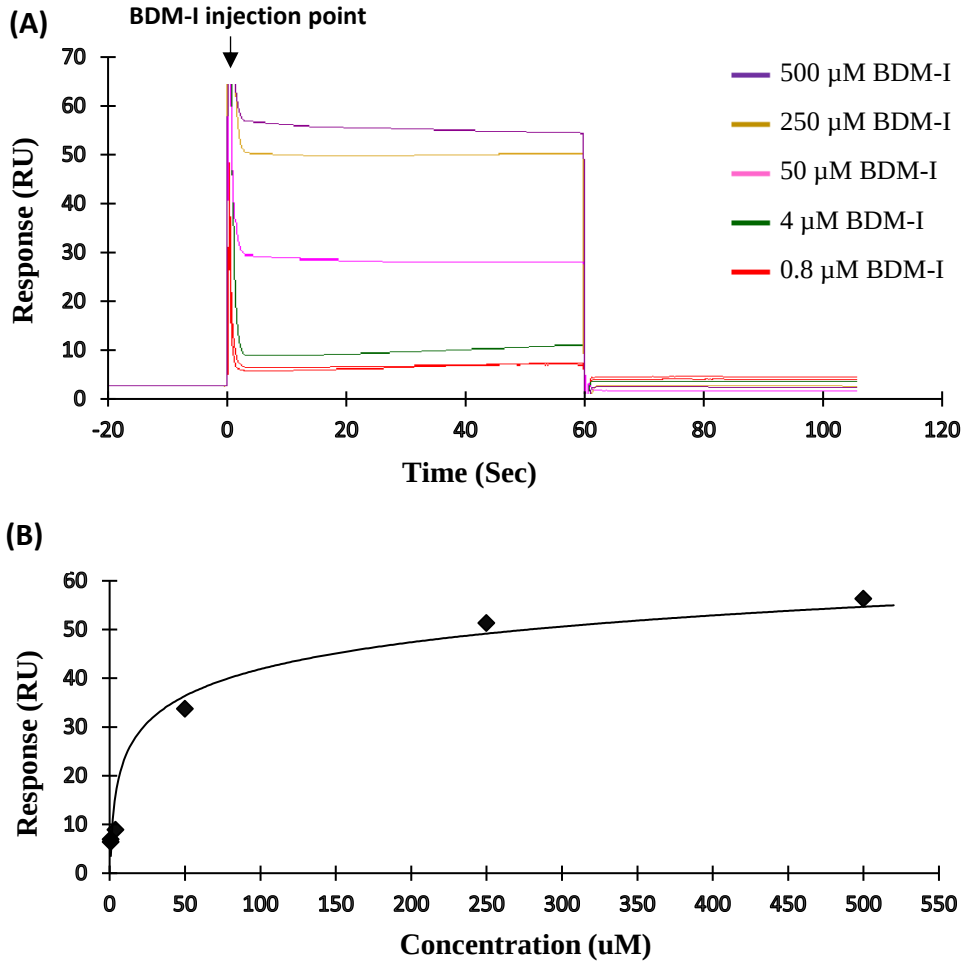


Figure 2.7. (A) AcpD SPR analysis of multicycle kinetics sensorgram showing BDM-I concentration (500 µM – 0.8 µM) response (RU) and **(B)** related affinity plot

BDM-I docking analysis

S. typhimurium protein AcpD is an alpha/beta classed protein with 3-layer (αβα) sandwich architecture and Rossmann fold topology [18] and is a member of the Flavoprotein super-family, identified by having a flavodoxin-like fold [19]. AcpD is a flavin dependent NADH preferring azoreductase; it catalyses the reduction of azo bonds (R-N=N-R) in aromatic azo compounds to their respective amines (Figure 2.8.) [20]. Enzymatic activity occurs when a Flavin mononucleotide (FMN) cofactor, NADH and substrate are all bound to AcpD; initiating an FMN facilitated two-stage donation of electrons from NADH to the azo substrate, thereby reducing any accessible azo bonds [21].

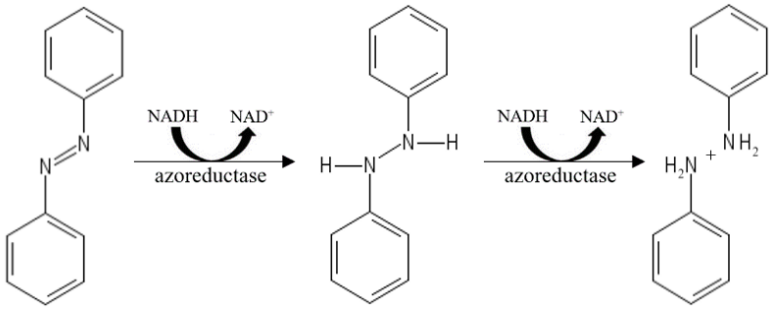


Figure 2.8. Two-stage reduction of benzidine azo bonds by azoreductase. Cleavage of azo bond (-N=N-) involves the transfer of two electrons (reducing equivalents) per stage in a bi-bi ping-pong mechanism from NADH to FMN to the substrate amine acceptor (electron acceptor)

Using Autodock Vina, analysis consisted of a 50 Å box surrounding the putative AcpD binding site containing a bound FMN cofactor (which is required for AcpD binding activity with a substrate [14]); NADH was also bound to AcpD however was not within the area of analysis and is thus not shown. Docking analysis predicts BDM-I binding to the active site composed of residues Tyr121, Phe163, Ala115, Ala113, Arg114, Arg60, Met59, Val56 and Asn98 with a binding energy of – 6.7 kcal/mol (Figure 2.9.). Hydrogen bonds are predicted between Ala115 and Asn98 to the nitro group on BDM-I, suggesting that the BDM-I nitro

group may be targeted for reduction. Tyr121 hydroxyl group likewise shows hydrogen bonding to the dioxyl ring on BDM-I, perhaps providing conformational stability during reduction activity, though this would require further testing to confirm. The remaining five potential BDM-I binding proteins did not have defined crystal structures for docking analysis, rather they were identified by SPOT-ligand on the basis of sequence predicted structures.

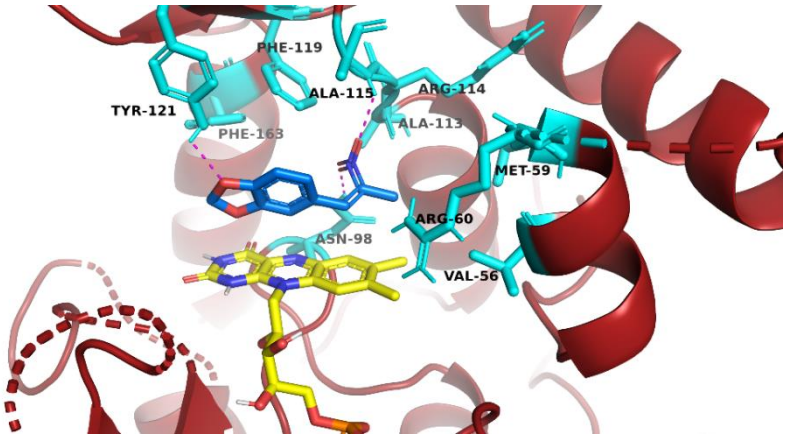


Figure 2.9. Docking of BDM-I (blue) to AcpD using Autodock Vina with default parameters, showing possible BDM-I binding orientation with a predicted binding energy of -6.7 kcal/mol. Predicted binding residues (cyan) are within 5 Å of docked BDM-I with hydrogen bonding predicted to Ala115, Asn98 and Tyr121 (dashed red line); FMN cofactor is highlighted in yellow. NADH was also bound to AcpD however, was not within the area of analysis and is thus not shown.

2.4. Discussion

As previously stated innovative techniques in drug discovery research are required, particularly given the challenges associated with the drug discovery industry; namely a need for refined drug candidates that can meet the concerns of patients, markets and regulatory agencies in a timely manner [2,22,23]. SPOT-Ligand is one such innovative technique that involves *in silico* screening of sequence and/or structural similarities

between a ligand-receptor template and related queries in order to identify novel drug targets [10,24]. *In silico* screening of putative antimicrobial targets for testing in a laboratory aids in the characterisation of novel antimicrobial drugs and indirectly in combating the increased rise of antimicrobial agent-resistant pathogens [25,26,27]. In this study, SPOT-Ligand predicted drug targets were carried through to experimentation if they had a > 60% structural or sequence match to known targets, were confirmed to have been previously purified with a primarily soluble recombinant protein product and having a HIS-tagged expression vector. These statistical and mechanical cut-off requirements were implemented to facilitate the mass expression of multiple proteins with high BDM-I predicted affinity, within the limited time frame available to complete this research. Six SPOT-Ligand predicted targets met the requirements for experimentation as novel targets for BDM-I, of which we validated using SPR. Docking analysis for AcpD identified it as a high affinity BDM-I target and provided clarification for the BDM-I mechanism of action. This in turn provides validation for the use of SPOT-Ligand as a drug target screening technique if screening is performed using a structural comparison and not sequence comparisons alone. To elaborate, AcpD was matched based on a known crystal structure whereas the other 5 targets did not have known crystal structures but were matched on sequence alone, indicating that successful predictions by SPOT-Ligand are based on structure and not solely sequence.

Prior to the experimental findings, the six SPOT-Ligand screened targets all had relatively equal chance of success based on cut-off requirements for experimentation however as mentioned above, we identified AcpD (RCSB PDB ID: 1T5B), an azoreductase protein from the species *S. typhimurium* (strain LT2), as the only novel target for BDM-I with a binding K_D of 0.58 µM. *S. typhimurium* is 1 of 2600 known *Salmonella* serovars that are all

Gram-negative flagellated anaerobic bacilli featuring O, H, and Vi antigens [28,29]. *Salmonella* is responsible for over 93 million annual cases of food born illnesses worldwide and approximately 155 thousand annual deaths [30]. *S. typhimurium* strain LT2 is one of the earliest identified serovars of *Salmonella* [31] and is often used in mouse models for human typhoid fever [32]. Primarily however, *S. typhimurium* strain LT2 is known as the leading cause of human enterocolitis; a severe inflammation of the intestinal mucosal epithelium, leading to excessive diarrhea and in extreme cases bowel necrotization [33]. *S. typhimurium* induces disease by means of two type III secretion systems (T_3SS) encoded by two bacterial chromosomal regions; Salmonella pathogenicity island 1 and Salmonella pathogenicity island 2 (SPI-1 and SPI-2). Following contact with the mucosal epithelium, SPI-1 secrets effector proteins that translocate into epithelial cells to promote bacterial entry and inflammation, while SPI-2 secretes effector proteins that in essence allow the intracellular survival of *S. typhimurium* following epithelial macropinocytosis of the infected cell [34]. Furthermore, a noticeable increase of multi-drug resistant strains of *S. typhimurium* has been observed between 1996 and 2016, with 43% of present-day *S. typhimurium* strains showing resistance against tetracycline, ampicillin, streptomycin, chloramphenicol and sulfonamides [35].

S. typhimurium protein AcpD is a flavin dependent NADH preferred azoreductase; an α/β classed protein with 3-layer (αβα) sandwich architecture and Rossmann fold topology [18] and is a member of the Flavoprotein super family, identified by having a flavodoxin-like fold [19]. AcpD like bacterial azoreductases are mainly known as biocatalysts for the degradation of azo dyes and azoic drugs [36]. Industrial wastewater containing azo-dyes is a leading contributor to global contamination of natural water sources [37]; an issue that has led to the identification of several bacterial azoreductases able to catalyse the degradation of

these harmful azo-dyes [38]. Moreover, human gut microbiota contain bacterial azoreductases that modulate the breakdown of azo-dyes during digestion [39], aiding in the degradation of food additives and the activation of azo-polymer coated drugs [40] or azo-based drugs such the analgesic Phenazopyridine hydrochloride [41].

Bacterial azoreductases have also shown quinone reductase activity, in some instances revealing greater affinity to quinone substrates over azo-substrates [42]. In knockout studies of the *E. coli* azoreductase AzoR (RCSB PDB ID: 2D5I), cell growth was inhibited in the ΔazoR mutant following the introduction of electrophilic quinone compounds. Electrophiles can be extremely damaging to nucleophiles and as such will target thiol-based proteins such as glutathione, a major thiol-based antioxidant found in bacteria that is responsible for the reduction of toxic dicarbonyl compounds. Knockout of AzoR stopped the reduction of intracellular quinones and directly resulted in an increased accumulation of electrophilic quinones that in turn caused a depletion of intracellular glutathione by 87% (compared to native *E. coli*). The depleted levels of glutathione led to increased cellular toxicity from quinones and inhibition of cellular growth [42]. As such, AzoR is involved in intracellular protection against thiol-specific stress as a quinone reductase; a more reasonable function in regards to maintaining organism health as the reduction of azo-dyes often result in bacterially toxic aromatic amines [42].

Structural alignment analysis between AcpD and AzoR using PDBeFold returned a Q score (quality of alignment) of 0.94 (1= exact match) with a 91% ID alignment (percentage of matched secondary structures) and 198 aligned residue pairs with a root-mean-squared deviation (RMSD) of 0.62 Å. These results indicate with high probability that AcpD would have a functionally similar intracellular protective role in *S. typhimurium* against forms of oxidative stress as seen with AzoR (based on structural similarities). Consequently,

inhibition of AcpD in *S. typhimurium* by BDM-I would result in the increase of intracellular toxicity from quinones and limitation of cellular growth.

Interestingly, there are several examples of bacterial azoreductase displaying nitroreductase activity via the catalytic reduction of nitro-compounds ($R\text{-}NO_2$), including an azoreductase from *Lysinibacillus sphaericus* [43], AZR azoreductase from *Rhodobacter sphaeroides* [44], *Enterococcus faecalis* azoreductase EF0404 [45] and paAzoR1 azoreductase from *P. aeruginosa* [46]. BDM-I has a nitro-compound with a nitro functional group attached to a propene (Figure 1.3.), showing predicted strong binding within the putative AcpD active site, with hydrogen bonds between the BDM-I nitro group and Ala115 and Asn98 (Figure 2.9.). This may indicate that AcpD has both nitroreductase and azoreductase activity, a finding that could elaborate the mechanism of action of BDM-I binding to AcpD. To further explore this, a structural alignment analysis was performed between AcpD and *P. aeruginosa* azoreductase PaAzoR; a known azoreductase with nitroreductase activity and a defined crystal structure with bound FMN cofactor and nitro compound substrate, Nitrofurantoin (NFZ) (RCSB PDB ID: 3R6W). Using PDBeFold AcpD showed a fairly high structural match with paAzoR1 (Figure 2.10.) with a Q score of 0.66 (1= exact match), 86% ID alignment with 184 aligned residue pairs and RMSD of 1.41 Å. This is a high enough global structure similarity to infer shared function between PaAzoR1 and AcpD [7, 47]. Furthermore, PaAzoR1 residues involved with binding of NFZ (Asn99, Phe100, Phe 173 and Tyr131) are likewise conserved in AcpD, with an Asn residue responsible for binding of the respective substrate nitro-group in both proteins (as seen in AcpD/BDM-I docking analysis and crystal structure of PaAzoR1 bound to NFZ; RCSB PDB ID: 3R6W) .

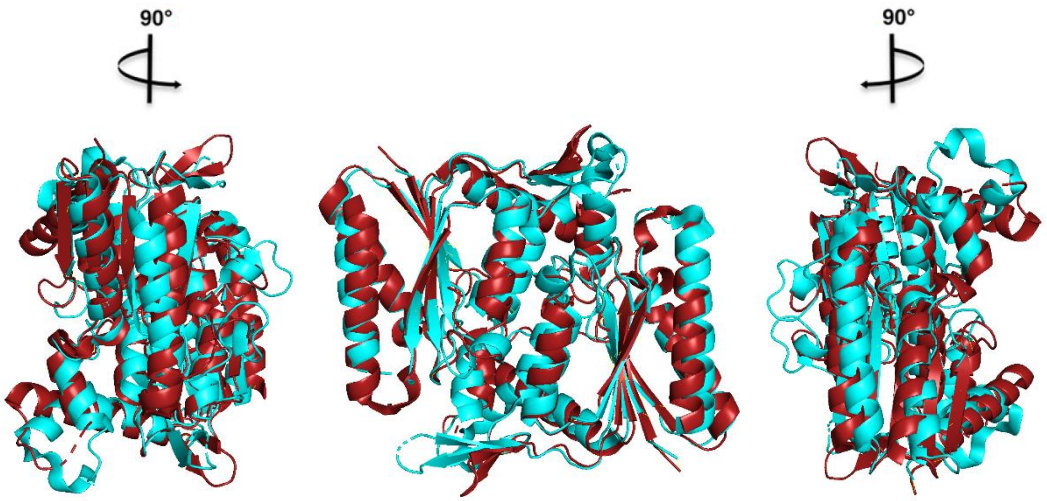


Figure 2.10. AcpD (red) structurally aligned to PaAzoR1 (cyan) using PyMOL v2.3.4

NFZ is a nitro-antibiotic used to treat urinary tract infections [48]. The reduction of NFZ nitroaromatic group results in the creation of extremely reactive electrophilic aniline intermediates that attack multiple intracellular sites, a mechanism of action that prohibits bacterial evolutionary resistance [49]. A large majority of nitro-based drugs in production are nitro-aromatics, that is to say they contain a nitro groups directly attached to an aromatic ring [50]. As such, possible consideration could be placed into the chemical synthesis of a new compound involving the addition of an aromatic ring to BDM-I to produce a nitro-aromatic [51] (Figure 2.11.).

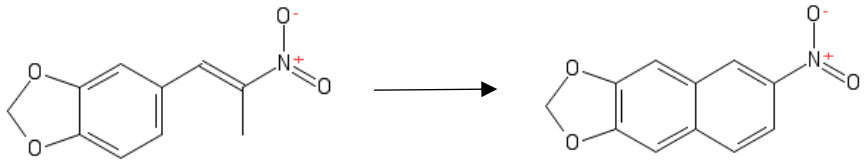


Figure 2.11. Possible chemical synthesis of BDM-I to 3,4-methylenedioxy-4-nitronaphthalene

If BDM-I does in fact prove to be reduced by a nitroreductase then any potential bacterial toxicity would need to be experimentally tested further as the presumed intermediates would be amines and not anilines however, accumulation of amines may also have indirect adverse effects on the bacterial cell through the raising of system pH [52]. The possible production of an intracellular toxic compound upon reduction of BDM-I would not be an ideal mechanism of action should that cellular toxicity transfer to humans in some way, given the quick association and slow disassociation (k_a of 9.4 x 10^4 $M^{-1}s^{-1}$ and a k_d of 9.7 x 10^{-4} s^{-1}). As mentioned in section 2.3.2. (SPR results) a large proportion of approved and clinically effective drug inhibitors bind quickly and are slow to release [16] with the exception of drugs that have toxic side effects like anticancer agents, where fast dissociation is preferred (in most cases) to limit toxicity [17]. If the potential of BDM-I to produce harmful (to bacteria) intermediates upon reduction is proven to be in some way toxic to humans as well [53], then quick dissociation would have been preferable to limit harmful side effects.

Clearly additional research is needed to determine BDM-I's viability as a human safe antimicrobial drug. This should involve experimentally testing BDM-I with other nitroreductase targets to affirm nitroreductase affinity, as well as crystallography or mass spectrometry (MS) of BDM-I bound to AcpD to confirm the docking results of this experiment. Furthermore, cytotoxicity assays using human cells would be prudent to determine toxicity of reduced BDM-I [54]. Should BDM-I prove to be toxic upon reduction of its nitro-group then this could indicate a multi-action attack mechanism (BDM-I is also believed to inhibit PTP [11]) that may account for its long-term effectiveness on *S. aureus* and *enterococci* [11]; two bacterial species known to develop drug-resistant phenotypes.

Of the five BDM-I non-binding targets, EF_0414 (UniProt ID: Q838M8), is predicted to be an oxidoreductase that catalyses oxidation-reduction reactions. UbiE (UniProt ID:

Q11QM3) is a methyltransferase that catalyses the transfer of a methyl-groups to an acceptor molecule and ftsZ5 (UniProt ID: W0JX85) modulates cell division and binds GTP. Lebu_1328 (UniProt ID: C7NAN5) is also predicted to catalyse the conversion of GTP to GDP, and UbiH (UniProt ID: F7U920) needs a flavin-adenine dinucleotide (FAD) cofactor to catalyse oxidation-reduction reactions. Some initial weak binding was seen (K_D > 15.00 µM) to EF_0414 and UbiH, which are both predicted to be oxidoreductase, with ubiH also predicted to be a flavoprotein; this is comparable to AcpD taxonomy (also an oxidoreductase) and as such these taxonomic similarities may account for the initial weak interactions. But the results were not reproducible in triplicate and duplicates K_D were too widely distributed, so future effort would be better spent testing further nitroreductase targets to validate the strong results with AcpD. The remaining non-binding outcomes indicated by SPR suggest BDM-I does not interact with GTPase (FtsZ5 and Lebu_1328), and though BDM-I does have a free methyl group it similarly did not interact with UbiE.

It is important to note that all of the functional characteristics of the non-binding targets were predicted or inferred based on amino acid sequence and have not been experimentally confirmed; as such these five targets may not truly do as inferred and require experimental validation of function. This also highlights a discrepancy between SPOT-Ligand sequence predicted structure screening and known crystal structure screening, showing that sequence-based structure prediction is a more difficult challenge [47]. Often dissimilar sequences of structurally similar proteins are the result of evolutionary mutations that greatly reduce sequence similarities between structurally comparative proteins, which accordingly renders sequence-based structure prediction a problematic prospect [55].

There are several limitations of this study that should be accounted for in replicate testing and analysis of BDM-I. Firstly, SEC can be used to estimate protein Mw however using

SEC alone for Mw can be unreliable do to inherent limitations of SEC including premature elution of asymmetric proteins and possible interaction of the protein with resin or column walls, lending to delayed elution [56]. As such other methods could be used to corroborate SEC Mw estimations in this study like multi-angle light scattering (MALS) or native mass spectrometry. Additionally the artificial recombinant structure of the proteins tested may not truly represent their structure in nature due to a lack of post-translation modifications found in the bacteria's natural host organism that may not be found in the expression host, as well as the artificial addition of HIS-tags which further changes the protein from the native variant. To elaborate, a protein changed from the native version of that protein may not behave as the native version would and as such the results found here may not be conclusive, at least without further testing. Additionally, the act of immobilising the protein onto the SPR sensor chip may change protein structure and in turn binding potential to the analyte, as well SPR is not the native environment of the molecules tested and so it may not perfectly represent how it would bind in nature. This may account for the lack of binding or lack of repeatable binding results found in the other 5 targets. As well, SPR kinetics does not necessarily represent the true drug-target time kinetics that may occur in the human body; a much more complex system with time dependent change in concentrations of both target and drug [57]. As such should BDM-I prove safe for human administration then association/disassociation rates may differ *in vivo*.

It would have been ideal to perform docking analysis on one of the non-binding BDM-I targets as a negative control against AcpD. However, docking analysis is only feasible with an actual crystal structure (which was only available for AcpD), as opposed to those predicted from sequence [47]. Therefore, future ranking of *in silico* drug targets for validation should involve an even mix of targets with actual crystal structures in addition to targets

with sequence predicted structures. Moreover, *in silico* docking does have its limitations, specifically binding scores that are only approximate and often time may not correlate perfectly with experimental results. Also *in silico* docking can only generate a limited number of conformational predictions of how the ligand will bind to the receptor and not necessarily all possible ligand-receptor configurations, as these predictions are dependent on the power of the computer processor and complexity of the docking software used [58]. As such, a repeat of this experiment could include additional molecular analysis techniques in support of SPR and *in silico* docking, such as nuclear magnetic resonance (NMR) spectroscopy of BDM-I bound to AcpD or isothermal titration calorimetry (ITC) of the same; so as to provide a more comprehensive binding analysis.

We achieved the experimental aims of this project, that is to say all six drug targets were successfully expressed and purified, then their affinity to BDM-I characterized using SPR analysis. Moreover, we may have identified a new mechanism of action for BDM-I through the binding of the BDM-I nitro-group to AcpD, as shown in docking analysis. Unexpectedly, we may have also revealed a novel nitroreductase in the known azoreductase AcpD, as demonstrated by its predicted binding of BDM-I nitro-group. Finally, we demonstrated the potential of SPOT-Ligand *in silico* structure-based screening to predict viable drug targets; validating bioinformatic drug target screening as a supportive technique in the characterisation of novel antimicrobials like BDM-I.

References

1. Prasad, V. & Mailankody, S. (2017) Research and development spending to bring a single cancer drug to market and revenues after approval. *JAMA Internal Medicine*, 177(11):1569–1575. https://doi.org/10.1001/jamainternmed.2017.3601

2. Kennedy, J.P. *et al.* (2008) Application of combinatorial chemistry science on modern drug discovery. *Journal of Combinatorial Chemistry*, 10(3):345-354. https://doi.org/10.1021/cc700187t

3. Macalino, S.J.Y. *et al*. (2015) Role of computer-aided drug design in modern drug discovery. *Archives of Pharmacal Research*, 38:1686–1701. https://doi.org/10.1007/s12272-015-0640-5

4. Li, Z. *et al.* (2017) In silico prediction of drug-target interaction networks based on drug chemical structure and protein sequences. *Scientific Reports*, 7:11174. https://doi.org/10.1038/s41598-017-10724-0

5. Yoo, S. *et al.* (2018) In silico profiling of systemic effects of drugs to predict unexpected interactions. *Scientific Reports*, 8:1612. https://doi.org/10.1038/s41598-018-19614-5

6. Kant, K., Lal, U. R. & Ghosh, M. (2018) In silico prediction and wet lab validation of arisaema tortuosum (wall.) schott extracts as antioxidant and anti-breast cancer source: a comparative study. *Pharmacognosy Magazine*, 13(4):S786–S790. https://doi.org/10.4103/pm.pm_69_17

7. Tiralongo, J., Cooper, O., Litfin, T. *et al.* (2018) YesU from Bacillus subtilis preferentially binds fucosylated glycans. *Scientific Reports*, 8:13139. https://doi.org/10.1038/s41598-018-31241-8

8. Schmid, T.F. *et al.* (2013) In silico prediction and experimental evaluation of Furanoheliangolide sesquiterpene lactones as potent agents against Trypanosoma brucei rhodesiense. *Antimicrobial Agents and Chemotherapy*, 58(1):325-332. https://doi.org/10.1128/AAC.01263-13

9. Prestinaci, F., Pezzotti, P. & Pantosti, A. (2015) Antimicrobial resistance: a global multifaceted phenomenon. *Pathogens and Global Health*, 109(7):309–318. https://doi.org/10.1179/2047773215Y.0000000030

10. Yang, Y., Zhan. J. & Zhou, Y. (2016) SPOT-Ligand: fast and effective structure-based virtual screening by binding homology search according to ligand and receptor similarity. *Journal of Computational Chemistry*, 37(18):1734-9. https://doi.org/10.1002/jcc.24380.

11. White, K. (2008). Antimicrobial mechanism of action of 3,4-methylenedioxy-beta-nitropropene. *Doctor of Philosophy (PhD), Applied Sciences, RMIT University: Melbourne, Australia.* Available from: https://researchbank.rmit.edu.au/eserv/rmit:7872/White.pdf

12. Tonks, N. (2006) Protein tyrosine phosphatases: from genes, to function, to disease. *Nature Reviews Molecular Cell Biology*, 7:833–846. https://doi.org/10.1038/nrm2039

13. White, K. S., Nicoletti, G. & Borland, R. (2014) Nitropropenyl benzodioxole, an anti-infective agent with action as a protein tyrosine phosphatase inhibitor. *The Open Medicinal Chemistry Journal*, 8:1–16. https://doi.org/10.2174/1874104501408010001

14. Verma, K., Kundu, D., Kundu, L.M., Singh, A.K. & Dubey V.K. (2019) Folding and stability of recombinant azoreductase enzyme from Chromobacterium violaceum. *Enzyme and Microbial Technology*, 131:109433. https://doi.org/10.1016/j.enzmictec.2019.109433.

15. Greenfield N. J. (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols*, 1(6):2876–2890. https://doi.org/10.1038/nprot.2006.202

16. Swinney, D. (2004) Biochemical mechanisms of drug action: what does it take for success? *Nature Reviews Drug Discovery*, 3:801–808. https://doi.org/10.1038/nrd1500

17. Folmer, R.H. (2018) Drug target residence time: a misleading concept. *Drug Discovery Today*, 23(1):12-16. https://doi.org/10.1016/j.drudis.2017.07.016.

18. Knudsen, M. & Wiuf, C. (2010) The CATH database. *Human genomics*, 4(3):207–212. https://doi.org/10.1186/1479-7364-4-3-207

19. Fox, N. K., Brenner, S. E. & Chandonia, J. M. (2014) SCOPe: structural classification of proteins--extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Research, 42(Database issue)*, D304–D309. https://doi.org/10.1093/nar/gkt1240

20. Hong, Y. G. & Gu, J. D. (2010) Physiology and biochemistry of reduction of azo compounds by Shewanella strains relevant to electron transport chain. *Applied Microbiology and Biotechnology*, 88(3):637–643. https://doi.org/10.1007/s00253-010-2820-z

21. Feng, J., Cerniglia, C. E. & Chen, H. (2012) Toxicological significance of azo dye metabolism by human intestinal microbiota. *Frontiers in Bioscience (Elite edition)*, 4:568–586. https://doi.org/10.2741/400

22. Kneller, R. (2010) The importance of new companies for drug discovery: origins of a decade of new drugs. *Nature Reviews Drug Discovery*, 9:867–882. https://doi.org/10.1038/nrd3251

23. Simpkin, V. *et al.* (2017) Incentivising innovation in antibiotic drug discovery and development: progress, challenges and next steps. *The Journal of Antibiotics*, 70:1087–1096. https://doi.org/10.1038/ja.2017.124

24. Litfin, T., Zhou, Y. & Yang, Y. (2017) SPOT-Ligand 2: improving structure-based virtual screening by binding-homology search on an expanded structural template library. *Bioinformatics*, 33(8):1238–1240. https://doi.org/10.1093/bioinformatics/btw829

25. Torrent, M., Nogues, V. & Boix, E. (2012) Discovering New *in Silico* Tools for Antimicrobial Peptide Prediction. *Current Drug Targets*, 13:1148. https://doi.org/10.2174/138945012802002311

26. Mhondoro, M. *et al.* (2019) Trends in antimicrobial resistance of bacterial pathogens in Harare, Zimbabwe, 2012–2017: a secondary dataset analysis. *BMC Infectious Diseases*, 19:746. https://doi.org/10.1186/s12879-019-4295-6

27. Hope, D. *et al.* (2019) Antimicrobial resistance in pathogenic aerobic bacteria causing surgical site infections in Mbarara regional referral hospital, Southwestern Uganda. *Scientific Reports*, 9:17299. https://doi.org/10.1038/s41598-019-53712-2

28. Giannella, R.A. (1996) Salmonella. In: *Medical Microbiology. 4th edition*. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 21. Available from: https://www.ncbi.nlm.nih.gov/books/NBK8435

29. Gal-Mor, O., Boyle, E.C. & Grassl, G.A. (2014) Same species, different diseases: how and why typhoidal and non-typhoidal Salmonella enterica serovars differ. *Frontiers in Microbiology*, 5:391. https://doi.org/10.3389/fmicb.2014.00391

30. Eng, S.K. *et al.* (2015) Salmonella: a review on pathogenesis, epidemiology and antibiotic resistance. *Frontiers in Life Science*, 8(3):284-293. https://doi.org/10.1080/21553769.2015.1051243

31. Hayden, H.S. *et al.* (2016) Genomic analysis of salmonella enterica serovar typhimurium characterizes strain diversity for recent U.S. salmonellosis cases and identifies mutations linked to loss of fitness under nitrosative and oxidative stress. *mBio*, 7(2):e00154-16. https://doi.org/10.1128/mBio.00154-16

32. Mathur, R. *et al.* (2012) A mouse model of Salmonella typhi infection. *Cell*, 151(3): 590–602. https://doi.org/10.1016/j.cell.2012.08.042

33. Anderson, C. J. & Kendall, M.M. (2017) Salmonella enterica Serovar Typhimurium Strategies for Host Adaptation. *Frontiers in Microbiology*, 8:1983. https://doi.org/10.3389/fmicb.2017.01983

34. Patel, S. & McCormick, B.A. (2014) Mucosal inflammatory response to Salmonella typhimurium Infection. *Frontiers in Immunology*, 5:311. https://doi.org/10.3389/fimmu.2014.00311

35. Wang, X., Biswas, S., Paudyal, N., Pan, H., Li, X., Fang, W. & Yue, M. (2019) Antibiotic resistance in Salmonella typhimurium isolates recovered from the food chain through national antimicrobial resistance monitoring system between 1996 and 2016. *Frontiers in Microbiology*, 10:985. https://doi.org/10.3389/fmicb.2019.00985

36. Misal, S.A. & Gawai, K.R. (2018) Azoreductase: a key player of xenobiotic metabolism. *Bioresources and Bioprocessing*, 5:17. https://doi.org/10.1186/s40643-018-0206-8

37. Kant, R. (2012) Textile dyeing industry an environmental hazard. *Natural Science*, 4:22-26. https://doi.org/10.4236/ns.2012.41004.

38. Jamee, R., & Siddique, R. (2019) Biodegradation of synthetic dyes of textile effluent by microorganisms: an environmentally and economically sustainable approach. *European Journal of Microbiology & Immunology*, 9(4):114–118. https://doi.org/10.1556/1886.2019.00018

39. Zahran, S.A. *et al*. (2019) Azoreductase activity of dye-decolorizing bacteria isolated from the human gut microbiota. *Scientific Reports*, 9:5508. https://doi.org/10.1038/s41598-019-41894-8

40. Roldo, M., Barbu, E., Brown, J.F., Laight, D.W., Smart, J.D. & Tsibouklis J. (2007) Azo compounds in colon-specific drug delivery. *Expert Opinion on Drug Delivery*, 4(5):547-60. https://doi.org/10.1517/17425247.4.5.547

41. Gaines, K.K. (2002) Phenazopyridine hydrochloride: the use and abuse of an old standby for UTI. *Urologic Nursing*, 24(3):207-9. https://www.suna.org/download/members/unjarticles/2004/04jun/207.pdf

42. Liu, G., Zhou, J., Fu, Q. S. & Wang, J. (2009) The Escherichia coli azoreductase AzoR Is involved in resistance to thiol-specific stress caused by electrophilic quinones. *Journal of Bacteriology*, 191(20):6394–6400. https://doi.org/10.1128/JB.00552-09

43. Misal, S.A. *et al.* (2014) Enzymatic transformation of nitro-aromatic compounds by a flavin-free NADH azoreductase from Lysinibacillus sphaericus. *Biotechnology Letters*, 36:127–131. https://doi.org/10.1007/s10529-013-1338-8

44. Liu, G. *et al.* (2007) Azoreductase from Rhodobacter sphaeroides AS1.1737 is a flavodoxin that also functions as nitroreductase and flavin mononucleotide reductase. *Applied Microbiology and Biotechnology*, 76:1271–1279. https://doi.org/10.1007/s00253-007-1087-5

45. Chalansonnet, V. *et al.* (2017) Identification of Enterococcus faecalis enzymes with azoreductases and/or nitroreductase activity. *BMC Microbiology*, 17:126. https://doi.org/10.1186/s12866-017-1033-3

46. Ryan, A. *et al*. (2011) Activation of nitrofurazone by azoreductases: multiple activities in one enzyme. *Scientific Reports*, 1:63. https://doi.org/10.1038/srep00063

47. Deng, H., Jia, Y. & Zhang, Y. (2018) Protein structure prediction. *International Journal of Modern Physics*, 32(18):1840009. https://doi.org/10.1142/S021797921840009X

48. Ju, K. S. & Parales, R. E. (2010) Nitroaromatic compounds, from synthesis to biodegradation. *Microbiology and Molecular Biology Reviews*, 74(2):250–272. https://doi.org/10.1128/MMBR.00006-10

49. McOsker, C.C. & Fitzpatrick, P.M. (1994) Nitrofurantoin: mechanism of action and implications for resistance development in common uropathogens. *Journal of Antimicrobial Chemotherapy*, 33:23–30. https://doi.org/10.1093/jac/33.suppl_A.23

50. Nepali, K., Lee, H.Y. & Liou, J.P. (2019) Nitro-group-containing drugs. *Journal of Medicinal Chemistry*, 62(6):2851-2893. https://doi.org/10.1021/acs.jmedchem.8b00147

51. Blakemore, D.C. *et al.* (2018) Organic synthesis provides opportunities to transform drug discovery. *Nature Chemistry*, 10:383–394. https://doi.org/10.1038/s41557-018-0021-z

52. Dean, R.T., Jessup, W. & Roberts, C.R. (1984) Effects of exogenous amines on mammalian cells, with particular reference to membrane flow. *The Biochemical Journal*, 217(1):27–40. https://doi.org/10.1042/bj2170027

53. Adams, G.E. *et al.* (1980) Toxicity of nitro compounds toward hypoxic mammalian cells in vitro: dependence on reduction potential. *Journal of the National Cancer Institute*, (64)3:555–560, https://doi.org/10.1093/jnci/64.3.555

54. Riss T. *et al.* (2019) Cytotoxicity assays: in vitro methods to measure dead cells. *Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences.* Available from: https://www.ncbi.nlm.nih.gov/books/NBK540958/

55. Miklós, I. *et al.* (2008) How reliably can we predict the reliability of protein structure predictions? *BMC Bioinformatics*, 9:137. https://doi.org/10.1186/1471-2105-9-137

56. Burgess, R.R. (2018) A brief practical review of size exclusion chromatography: rules of thumb, limitations, and troubleshooting. *Protein Expression and Purification*, 150:81-85. https://doi.org/10.1016/j.pep.2018.05.007

57. Tonge, P. J. (2018) Drug-target kinetics in drug discovery. *ACS Chemical Neuroscience*, 9(1):29–39. https://doi.org/10.1021/acschemneuro.7b00185

58. Pinzi, L. & Rastelli, G. (2019) Molecular docking: shifting paradigms in drug discovery. *International Journal of Molecular Sciences*, 20(18):4331. https://doi.org/10.3390/ijms20184331

Chapter 3

Characterisation of *B. thetaiotaomicron* hypothetical proteins BT_411 and BT_3781 predicted using SPOT-Struc

Abstract

The inclusion of *in silico* prediction as a director of biochemical research is fast becoming a necessary step in the design of cost and time effective experimentation. One such example being the characterisation of protein-ligand interactions, where navigating the wide variety of complex protein structures for specific subsets can be challenging and time consuming by traditional manual methods alone. Most notably, identifying the interactions between carbohydrate-binding proteins (CBPs) and their distinct glycan ligands is especially demanding given the high flexibility, low binding affinity and lack of a linear sequence exhibited in glycans. SPOT-Struc is one such *in silico* structure-based function prediction technique, designed to aid in the identification of novel CBPs. Here we experimentally explore the putative carbohydrate affinity and binding affinity of two SPOT-Struc predictions; *B. thetaiotaomicron* hypothetical proteins BT_411 (RCSB PDB ID: 3HNM) and BT_3781 (RCSB PDB ID: 2P0V). Both proteins are ideal candidates for this study given their functions remain unknown and as such, if proven to be CBPs, will add to the library of known CBPs and validate SPOT-Strucs predictive capabilities. Surface plasmon resonance (SPR) and *in silico* docking analysis revealed BT_411 features preferential binding to β1,4 linked *N*-acetylglucosamines (GlcNAc), as well as potentially to α2,3 linked sialic acids glycan moieties. Moreover, BT_3781 was shown to have high affinity for fucosylated blood group antigens. As a result, this study validates SPOT-Struc predicted CBPs and in turn the use of *in silco* prediction as a director of experimental research into novel CBPs.

3.1. Introduction

Lectins, a distinct subset of CBPs, are identified by a variety of structural and functional classifications including Galectins [1], *C*-type [2], *F*-type [3], *I*-type (siglecs) [4], ficolins [5] and carbohydrate binding modules (CBMs) involved in the hydrolysis of glycosidic linkages [6]. Found ubiquitously across all organisms [7], lectins play an integral part in a multitude of biological processes including cell to cell communication [8], glycoprotein folding and assembly [9], pathogen pattern recognition [10] and immunomodulation [11,12]. Furthermore, lectins display preferential and reversible binding to the monosaccharide branching motifs of their corresponding homo- and hetero-polymers; resulting in structurally complex lectin-oligosaccharide interactions [13,14] that regulate a variety of biologically essential cellular exchanges [15].

Characterisation of CBP interactions with the rich variety of cell surface glycoproteins and glycolipids on glycosylated proteins and lipids [16,17,18] has proven to be instrumental in critical research related to cell development and associated cellular disorders [19,20,21], viral vaccines [22,23], tumour metastasis/treatment [24] and biomarker discovery [25]. Thus, the necessity to identify novel CBPs and define their carbohydrate affinity has become increasingly relevant and vital to related medical research. As such, *in silico* predictive modelling of protein function, used as a director of experimental focus, has the potential to significantly improve the identification of novel CBPs in a timely and cost-effective manner [26].

At present, *in silico* prediction methods of novel CBPs involves machine-learning to identify similar physicochemical properties between a CBP template and possible CBP queries; properties that include protein structure, sequence and chemical composition in

addition to protein evolutionary data [27]. SPOT-Struc (created by Prof. Yaoqi Zhou), a structure-based function prediction technique, uses a combination of the structural alignment program SPalign and the measurement of free energy changes associated with residue interactions, also known as distance-scaled finite-ideal gas reference state (DFIRE), to statistically identify if proteins with unknown functions are in fact CBPs [28,29]. In essence, a SPOT-Struc generated CBP template is compared against several query proteins for structural similarities where a high match indicates a putative CBP. The actual functional annotation of the putative CBP can potentially be inferred based on the function of the matched template however, at this point in time SPOT-Struc mainly identifies if a protein is a CBP or not and is less specific as to the putative CBPs particular carbohydrate preference. Prior to my study, Prof. Yaoqi Zhou and cohort used SPOT-Struc to identify hypothetical proteins BT_411 and BT_3781 (hypothetical due to unknown function) from *B. thetaiotaomicron* as CBPs (RCSB PDB ID: 3HNM and 2P0V, respectively) [28].

The CATH Protein Structure Classification database classifies BT_411 as a putative chitobiase beta protein and a member of the galactose (Gal) binding like domain superfamily [30] having a β-sandwich structure comprising 6 to 9 strands organized as 2 sheets to create a jelly roll topology [31]. SPOT-Struc analysis determined that BT_411 was a match to the human gastrointestinal pathogen *Clostridium perfringens* family 32 CBM protein (RCSB PDB ID: 2j7m) [28], a known beta CBM with a β-sandwich and jellyroll fold [30] and an affinity for galactose (Gal), *N*-Acetyl-D-lactosamine (LacNAc) and type II blood group H-trisaccharide [32].

Structural Classification of Proteins–extended (SCOPe) classed BT_3781 as an α-protein with an α/α toroid fold and a member of the six-hairpin glycosidases superfamily possessing an α/α barrel structure with glycosyltransferase topology [33]. SPOT-Struc

analysis matched BT_3781 with fungal *Hypocrea jecorina* protein Glycoside Hydrolase Family 15 Glucoamylase (RCSB PDB ID: 2vn4) [28], a CBM with an α/α barrel structure and glycosyltransferase topology with unknown glycan affinity [34].

Recently, the novel lectin PSL-2 was isolated and purified from the fruiting body of the Australian macrofungi *Phaeolus schweinitzii* [35] and found by various techniques to bind lactose (Galβ1-4Glc). In collaboration with Dr Annabelle Varrot (CERMAV-CNRS, Grenoble, France), we have determined the crystal structure of the recombinant PSL-2 to 1.55 Å resolution. The PSL-2 crystal structure indicates that PSL-2 may form a tetramer with a galectin-like fold and possesses a well conserved Gal binding site and a secondary fucose (Fuc) binding site. The observation of a secondary Fuc binding site was unexpected and the carbohydrate binding affinity of the recombinant PSL-2 has not yet been fully characterized. As such PSL-2, a known lectin, would be an ideal positive control lectin to be further characterized alongside the recombinant hypothetical lectins BT_411 and BT_3781.

Here I describe the experimental characterisation of recombinant PSL-2 as a Fuc and Gal binding lectin and we validate BT_411 and BT_3781 as CBPs. Using SPR to establish detailed glycan affinity profiles of each protein, we confirm PSL-2 as having high affinity for Fuc and Gal containing glycans. I also reveal the new putative functional annotation of BT_411 as a βGNase CBM with affinity for β1,4 linked *N*-acetylglucosamine (GlcNAc) and putative affinity for α2,3 linked sialic acids. Moreover, I show that BT_3781 is a novel fucosidase CBM with binding affinity for fucosylated Blood group antigens. The identification of BT_411 and BT_3781 as new CBMs support the application of *in silico* structure-based function prediction as a guide for experimental research into novel CBPs.

3.2. Materials and Methods

3.2.1. Expression and purification of recombinant *B. thetaiotaomicron* hypothetical proteins BT_411, BT_3781 and recombinant *P. schweinitzii* lectin PSL-2

Separate vectors of pET21_NESG incorporating *B. thetaiotaomicron* hypothetical proteins BT_411 and BT_3781 (clone ID's BtCD00339868 and BtCD00342887 respectively) with C-terminal 6xHis tags under the control of a T7 lac promoter were purchased from the DNASU plasmid repository (http://dnasu.org). Vector pET-25B incorporating *P. schweinitzii* protein PSL-2 under the control of T7 lac promoter was obtained from Dr Annabelle Varrot (CERMAV-CNRS, Grenoble, France). Each vector was transformed into competent *E. coli* BL21 (DE3) cells to be used for recombinant protein expression. Overnight cultures of BL21 (DE3)/BtCD00339868, BL21 (DE3)/BtCD00342887 and BL21 (DE3)/PSL-2 were used to inoculate LB broth (ThermoFisher) containing ampicillin (75 μg/mL). The inoculated LB broth was then incubated at 37°C with aeration until $OD_{600\ nm}$ reached 0.5 to 0.8, at which point expression was induced using 1 mM IPTG and temperature reduced to 16°C for overnight incubation. The resulting cell pellets were resuspended in PBS lysis buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4; pH 7.5), lysozyme (2 mg/mL), DNaseI and protease inhibitor cocktail mix (20 μL). Cell lysis was then performed by sonication with cellular debris removed via centrifugation at 16 000 x g for 15 min. The refined supernatants containing BT_411 and BT_3781 were mixed separately with 1 mL of HIS-select nickel affinity resin (Sigma); PSL-2 was mixed with 1 mL Lactose-Sepharose affinity resin (GALAB). The slurry mix of each protein was then packed by gravity into separate 10 mL Bio-Rad chromatography columns. Columns containing nickel affinity resin slurry mix were washed once with PBS lysis buffer, then washed with high salt buffer (1 M NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4,

1.8 mM KH_2PO_4; pH 7.5) followed by the final elution wash of the bound HIS-tagged protein using Imidazole elution buffer (300 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 500 mM Imidazole; pH 7.5) in 1 mL volume. The PSL-2/Lactose-Sepharose resin slurry mix was washed once with PBS lysis buffer, then washed with high salt buffer followed by the final wash with lactose elution buffer (300 mM NaCl, 2.7 mM KCl, 20 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 200 mM lactose; pH 7.5) in 1 mL volume.

The removal of Imidazole from BT_411 and BT_3781 elutions and further purification of all target proteins was performed using SEC on the AKTA pure FPLC system (GE healthcare) with a 24 mL Superdex 200 10/300 gel filtration column (GE healthcare). Column was equilibrated with PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4; pH 7.5) and calibrated using blue dextran 2000 (to determine void volume), Conalbumin (75 kDa), Ovalbumin (43 kDa) and Ribonuclease-A (13.7 kDa). Calibration standards elution volumes were used to generate a calibration curve to calculate the Mw of the target proteins in solution. Proteins were eluted using PBS (pH 7.5) and monitored at 280 nm with 0.75 mL fractions collected and pooled according to chromatogram elution peaks. Each pooled sample was tested for presence of target protein by dot blot using a primary mouse monoclonal anti-His6 (1:10,000 dilution, Cell Signaling Technologies) antibody and secondary goat anti-mouse horseradish peroxidase conjugated (1:10,000 dilution, Bio-Rad Laboratories) antibody. SEC protein elution volumes were used to calculate phase distribution coefficient (K_{av}) for each protein (equation shown in Figure 3.1.); K_{av} was then inputted into the calibration standards straight-line equation to calculate estimated protein Mw.

$$K_{av} = \frac{\text{Elution Volume (VE)} - \text{Void Volume (VO)}}{\text{Column Volume (Vc)} - \text{Void Volume (VO)}}$$

Figure 3.1. Equation used to calculate the phase distribution coefficient (K_{av}) with protein elution volume, void volume (determined using blue dextran) and column volume (24 mL)

Purified protein was concentrated using Amicron 10 K 50 mL centrifugal filter (2 mg/mL) and purity was confirmed by SDS-PAGE using 4-20% gradient SDS-polyacrylamide gels (Bio-Rad) stained with Coomassie brilliant blue. HIS-tagged BT_411 and BT_3781 were further identified by western blot immunodetection using a primary mouse monoclonal anti-His6 (1:10,000 dilution, Cell Signaling Technologies) antibody and secondary goat anti-mouse horseradish peroxidase conjugated (1:10,000 dilution, Bio-Rad Laboratories) antibody. SDS-PAGE protein samples were denatured chemically and by heat (95ºC for 5 min) and any potential disulfide bonds were reduced using 2-mercaptoethanol. Figure 3.2. shows the summary of expression and purification for BT_411, BT_3781 and PSL-2.

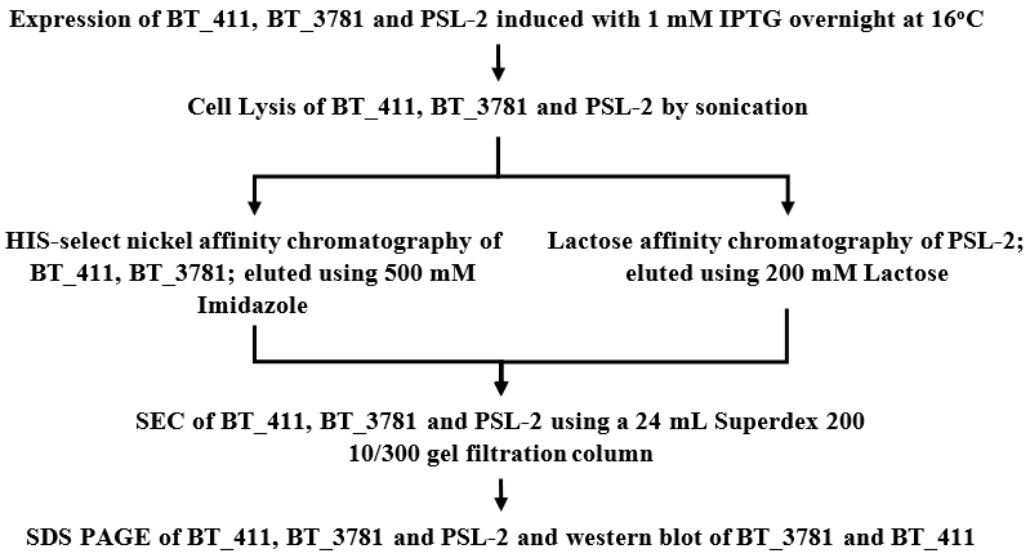


Figure 3.2. Flow chart depicting expression and purification of BT_411, BT_3781 and PSL-2

3.2.2. Protein estimation

Protein concentration was determined by colorimetric assay using a BCA protein assay kit (Thermo scientific) with 25 µL of BSA (concentration range between 0 and 2 mg/mL) loaded in duplicate to each appropriate well of a 96 well microtiter flat bottom plate. 25 µL of pure, 1/10 and 1/100 samples of BT_411, BT_3781 and PSL-2 were loaded into appropriate wells on the same 96 well plate followed by the addition of 200 µL of working reagent (Thermo scientific) to each well containing BSA and protein sample and then incubated at 37^{o}C for 40 min. Absorbance was measured at 575 nm using a Bio-Rad xMark™ Microplate Absorbance Spectrophotometer with concentrations calculated using Microplate Manager® Software (based on the line equation generated from BSA concentration range and sample dilution factor).

3.2.3. Quality control check of proteins using CD spectroscopy

CD spectroscopy was performed using the Jasco J-1500 Circular Dichroism Spectrophotometer as a quality control step to determine if a protein was denatured (not to analyse secondary structures) before committing to characterisation using SPR. This was done by visually comparing the CD spectra of non-denatured protein versus the denatured equivalent and if the wave patterns of the two spectra were different, it was inferred the protein was not denatured and thus suitable for characterisation. Purified BT_411, BT_3781 and PSL-2 (25 mg/mL) was added to a low conductivity buffer (100 mM KH_2PO_4) and placed in a 1 mm quartz cuvette. Protein measurement parameters were set at standard sensitivity between 260 nm and 190 nm with a 1 mm cell path using an accumulation of 9 (average of 9 repeated scans). Low conductivity buffer without protein was used as a baseline. Purified BT_411, BT_3781 and PSL-2 (25 mg/mL) was then

denatured using 20 % SDS (w/v), Tris-HCl (0.5 M), 2-mercaptoethanol and by heat (95ºC for 5 min) and measured under the same parameters in low conductivity buffer.

3.2.4. Characterisation of recombinant *B. thetaiotaomicron* hypothetical proteins BT_411, BT_3781 and recombinant *P. schweinitzii* lectin PSL-2

SPR Detection

SPR detection was conducted using a BIAcore T200 Biosensor system (GE Healthcare) at 25ºC in 10 mM PBS (pH 7.4) at a flow rate of 20 μL/min (30 μL/min for regeneration). Purified BT_411, BT_3781 and PSL-2 was diluted to 2 μg/mL in PBS (pH 7.4) and immobilized onto flow cells (FC) 2, 3 and 4 (respectively) of a BIAcore CM5 sensor chip using amine covalent coupling with 10 min contact time at a flow rate of 5 μL/min. FC 1 contained no protein and was used as a blank reference. Triplicate cycles were performed for each protein to establish validity of the SPR response. Five-fold serial dilutions (50 to 0.008 μM) of free glycans (supplied and passed quality control for structure and purity by the glycan array facility of the Institute for Glycomics) were prepared in PBS (pH 7.4) and individually loaded onto the sensor chip from a 96 well plate to be assessed using single cycle kinetics, with the CM5 chip regenerated using EDTA after each glycan dilution series injection. Glycans were chosen based on supply availability, cost and group (a few from each group for variation). Post regeneration the chip was reloaded with protein prior to the injection of the next dilution series. Glycan binding affinity was recorded by the response signal difference between protein loaded FC's and the blank reference of FC 1. Injection of each analyte concentration was followed by 300 s of dissociation. T200 BIAcore Evaluation software was used to analyze SPR signals and equilibrium dissociation constants (K_D) as determined from steady state analysis.

3.2.5 Docking analysis

A structural alignment search was performed using SPOT-Struc to identify the highest matched targets with a known ligand affinity similar to the ligand affinity of BT_411 and BT_3781, as identified by SPR affinity results. Docking analysis was performed using Autodock Vina in a 40 Å box surrounding the putative binding site of BT_411 and BT_3781. The binding site was inferred by superposition with the SPOT-Struc matched templates with characterised binding sites. The ligand structure of the matched target was then docked into the putative binding site of BT_411 and BT_3781 to determine the binding energies and residues involved.

3.3. Results

3.3.1. Expression and purification of recombinant *B. thetaiotaomicron* hypothetical proteins BT_411, BT_3781 and recombinant *P. schweinitzii* lectin PSL-2

Recombinant *B. thetaiotaomicron* hypothetical proteins BT_411 and BT_3781 as well as recombinant *P. schweinitzii* lectin PSL-2 were expressed in *E. coli* BL21 (DE3) cells and subsequently purified to homogeneity as described in section 3.2.1. Figure 3.3. shows the successful purification to homogeneity of HIS-tagged BT_411 as a single ~18 kDa band (lane 4, Coomassie stain and lane 5, western blot), and BT_3781 as a single ~54 kDa band (lane 8, Coomassie stain and lane 9, western blot), by HIS-select nickel affinity chromatography and SEC. Furthermore, we successfully purified PSL-2, visible as a single ~16 kDa band (lane 12 Coomassie stain) using Lactose-Sepharose affinity chromatography and SEC. The total quantity of purified BT_411, BT_3781 and PSL-2 proteins per litre of *E. coli* culture following purification was 2.44 mg, 1.65 mg and 6.73 mg, respectively.

As previously mentioned in chapter 2, CD spec is generally used in the identification of secondary structures and folding properties, represented by waves on a CD spectra. However, this was not how CD spec was used in this study, but rather it was utilized as a quality control step to determine if a protein was denatured (as opposed to analyses of secondary structures, which was not performed) before committing to characterisation by SPR. This was done by comparing the CD spectra of the non-denatured protein versus the denatured equivalent and if they matched it was inferred the protein was denatured and not suitable for SPR analyses. All proteins were confirmed to be non-denatured and ready for SPR. (Figure A1.4.).

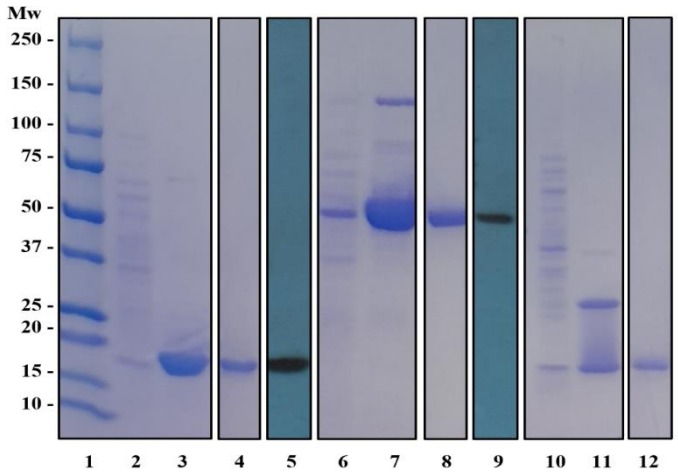


Figure 3.3. Coomassie stained gel and associated western blot (for BT_411 and BT_3781) showing the purification of recombinant *B. thetaiotaomicron* hypothetical proteins BT_411, BT_3781 and recombinant *P. schweinitzii* lectin PSL-2. Purified protein concentrated using Amicron 10 K 50 mL centrifugal filter to 2 mg/mL. Lane 1, Bio-Rad precision plus protein Mw standard; Lane 2, soluble recombinant BT_411; Lane 3, Ni^{+}NTA purified BT_411 elution; Lane 4, concentrated SEC purified BT_411 protein; Lane 5, western blot of purified BT_411 protein (165 residues) at the expected Mw of ~18 kDa; Lane 6, soluble recombinant BT_3781; Lane 7, Ni^{+}NTA purified BT_3781 elution; Lane 8, concentrated SEC purified BT_3781 protein; Lane 9, western blot of purified BT_3781 protein (481 residues) at the expected Mw of ~54 kDa; Lane 10, soluble recombinant PSL-2; Lane 11, lactose purified PSL-2 elution; Lane 12, concentrated SEC purified PSL-2 protein (166 residues) at the expected Mw of ~16 kDa.

SEC was performed as a final purification step and as a means to estimate the native Mw of recombinant BT_411, BT_3781 and PSL-2 protein. Pooled fractions associated with visible peaks on protein chromatograms were tested for presence of pure BT_411, BT_3781 and PSL-2 protein, using dot blot immunodetection as described in section 3.2.1. Mw was determined as described in section 3.2.1, using the straight-line equation generated from the Mw standard calibration curve (Figure 3.4.).

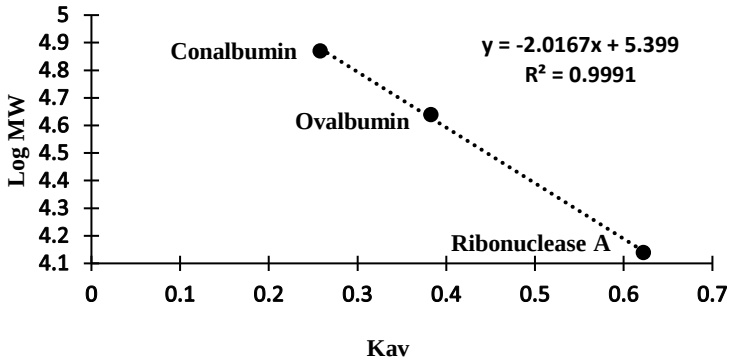


Figure 3.4. Standard Mw calibration curve and trendline. Protein standards Conalbumin (75 kDa), Ovalbumin (43 kDa) and Ribonuclease-A (13.7 kDa) eluted with 20 mM PBS (pH 7.4.) by SEC using a Superdex 200 Increase 10/300 GL column and monitored at 280 nm

SEC is a further purification step but as described in section 3.2.1, SEC may also be used to estimate oligomeric state of a protein in solution by simply dividing the SEC-indicated Mw by amino acid sequence derived Mw or to a lesser degree of accuracy Mw as shown by SDS PAGE. BT_411 (Figure 3.5.A) has a SEC Mw of 18.7 kDa, which correlates to the Mw inferred by the protein band visible in Figure 3.3. lane 4, also at ~18 kDa , suggesting that BT_411 may exist as a monomer in solution. BT_3781 (Figure 3.5.B) has a SEC Mw of 52.0 kDa, also correlating to the Mw inferred by the protein band visible in Figure 3.3., lane 8, at ~ 54 kDa, suggesting it may be a monomer however, this is contrary to the crystal structure of BT_3871 that implies it is a homodimer. PSL-2 (Figure 3.5.C) was

found to have a SEC Mw of 43.7 kDa, suggesting that PSL-2 exists as trimer in solution correlating to the Mw inferred by the protein band in Figure 3.3., lane 8 at ~16 kDa. This is also contrary to PSL-2s presumed oligomeric state which suggests it is a tetramer.

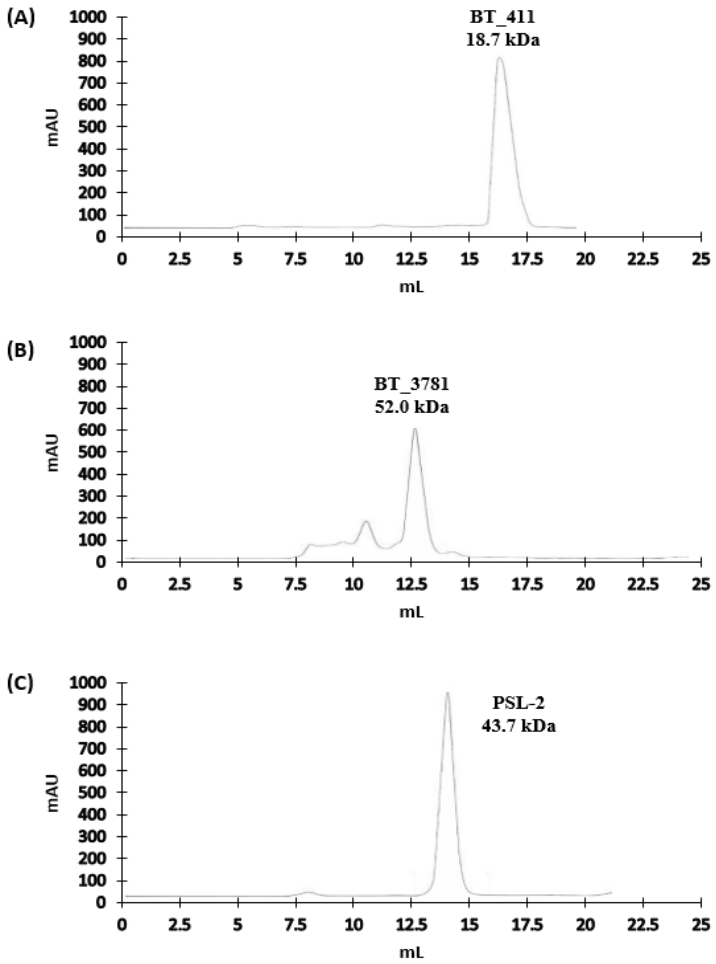


Figure 3.5. SEC chromatograms showing protein elution using a Superdex 200 Increase 10/300 GL column. Column was equilibrated with 10 mM PBS (pH 7.5) and chromatography monitored at 280 nm. **(A)** Purified BT_411 with SEC Mw of 18.7 kDa; **(B)** Purified BT_3781 with SEC Mw of 52.0 kDa; **(C)** Purified PSL-2 with SEC Mw of 43.7 kDa

Table 3.1. summarizes the estimated (based on amino acid sequence), SDS PAGE and SEC Mw of pure BT_411, BT_3781 and PSL-2, as well as the possible oligomeric state and total protein yield per litre of *E. coli* culture.

Table 3.1. Protein theoretical & SEC estimated Mw, inferred oligomeric state and total yield

Gene	Theoretical Mw (kDa)	SDS PAGE Mw (kDa)	SEC Mw (kDa)	Oligomeric State*	Total Yield (mg/mL)
BT_411	18.87	16-19	18.71	Monomer	2.44
BT_3781	54.71	51-55	52.00	Monomer	1.65
PSL-2	16.55	15-18	43.74	Trimer	6.73

* Estimated for recombinant protein in solution based on SEC

3.3.2. Characterisation of recombinant *B. thetaiotaomicron* hypothetical proteins BT_411, BT_3781 and recombinant *P. schweinitzii* lectin PSL-2

SPR Analysis

BT_411

SPR analysis of BT_411 confirmed its classification as a CBP by SPOT-Struc, indicating preferential binding to α2,3 linked sialic acid (*N*-acetylneuraminic acid, Neu5Ac) and terminal GlcNAc on chitin-based glycan moieties (Table 3.2. and Figure A1.5.). Comparison of BT_411 affinities across related structures shows near equal affinity across all terminal Neu5Ac-containing glycans tested, showing high affinity for Sialyl-Lewis^x tetraose (10B, Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAc, K_D: 0.10 μM), Sialyl-Lewis^a tetraose (10A, Neu5Acα2-3Galβ1-3(Fucα1-4)GlcNAc, K_D: 0.11 μM) and Sialylated TF antigen (292, Neu5Acα2-3Galβ1-3GalNAc, K_D: 0.11 μM). BT_411 likewise showed near equal affinity between chitotriose glycan moieties with a terminal GlcNAc (K_D: 0.13 μM to 0.19 μM). As mentioned in Materials and Methods section 3.3.4. each glycan was run three times and if the 3 runs had a SD > 1 then results were deemed to unevenly distributed from the mean K_D and not reliable enough to be considered a positive binding event; likewise applied to BT_3781 and PSL-2.

Table 3.2. BT_411 SPR analysis showing the mean K_D (μM) and SD of triplicate runs for each glycan. If SD was > 1 then K_D results was deemed unreliable and not included.

ID	Name	Structure	$\bar{x}$ K_D (μM) ± SD
10B	Sialyl-Lewis^x tetraose	Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAc	0.10 ± 0.04
292	Sialylated TF antigen	Neu5Acα2-3Galβ1-3GalNAc	0.11 ± 0.08
10A	Sialyl-Lewis^a tetraose	Neu5Acα2-3Galβ1-3(Fucα1-4)GlcNAc	0.11 ± 0.07
4A	N,N'-Diacetyl chitobiose	GlcNAcβ1-4GlcNAc	0.13 ± 0.05
4C	N,N',N'',N'''-Tetraacetyl chitotetraose	GlcNAcβ1-4GlcNAcβ1-4GlcNAcβ1-4GlcNAc	0.15 ± 0.03
4D	N,N',N'',N''',N'''',N''''-Hexaacetyl chitohexaose	GlcNAcβ1-4GlcNAcβ1-4GlcNAcβ1-4GlcNAcβ1-4GlcNAcβ1-4GlcNAc	0.19 ± 0.13
4B	N,N',N''-Triacetyl chitotriose	GlcNAcβ1-4GlcNAcβ1-4GlcNAc	0.23 ± 0.07
1C	β-1-4-galactosyl-galactose	Galβ1-4Gal	-
1F	asialo GM1	Galβ1-3GalNAcβ1-4Galβ1-4Glc	-
1H	Lacto-N-neotetraose	Galβ1-4GlcNAcβ1-3Galβ1-4Glc	-
12N	Chondroitin-4-sulfate	ΔUA-GalNAc-4S	-
13K	Chondroitin sulfate	(GlcA/IdoAβ1-3(±4/6S)GalNAcβ1-4)n (n<250)	-
2E	Xeno-antigen	Gala1-4Galβ1-4GlcNAc	-
1I	Lacto-N-neohexaose	Galβ1-4GlcNAcβ1-6(Galβ1-4GlcNAcβ1-3)Galβ1-4Glc	-
1N	α1-3 Galactobiose	Gala1-3Gal	-
1P	Linear B-6 Trisaccharide	Gala1-3Galβ1-4Glc	-
5A	β1-2 N-Acetylglucosamine-mannose	GlcNAcβ1-2Man	-
5B	Bianntennary N-linked core pentasaccharide	GlcNAcβ1-2Manα1-6(GlcNAcβ1-2Manα1-3)Man	-
5C	α1-2-Mannobiose	Manα1-2Man	-
5F	α1-6-Mannobiose	Manα1-6Man	-
7J	Lewis^a	Galβ1-3(Fucα1-4)GlcNAc	-
19J	Lewis^x tetrasaccharide	Galβ1-4(Fucα1-3)GlcNAcβ1-3Gal	-
7K	Blood Group A trisaccharide	GalNAcα1-3(Fucα1-2)Gal	-
7L	Lactodifucotetraose (LDFT)	Fucα1-2Galβ1-4(Fucα1-3)Glc	-
18D	Blood group B (Tetraose type 5)	Galα1-3(Fucα1-2)Galβ1-4Glc	-
7F	Blood group H disaccharide	Fucα1-2Gal	-
15E	Hyaluronan fragment 50kDa	NA	-
15G	Hyaluronan fragment 70kDa	NA	-
89	TF antigen	Galβ1-3GalNAc	-

BT_3781

SPR analysis of BT_3781 confirmed its classification as a CBP by SPOT-Struc, indicating high preferential binding to fucosylated glycans with both terminal and non-terminal Fuc as well as moderate affinity to glycan moieties with terminal Galactose (Gal) (Table 3.3. and Figure A1.6.). A comparison of BT_3781's glycan affinity reveals an almost 2-fold greater affinity towards Blood group H Disaccharide (7F, Fucα1-2Gal, K_D: 0.17 µM) and Blood group B (18D, Galα1-3(Fucα1-2)Galβ1-4Glc, K_D: 0.21 µM) compared to Lewis^x (19J, Galβ1-4(Fucα1-3)GlcNAcβ1-3Gal, K_D: 0.39 µM) and Lewis^a (7J, Galβ1-3(Fucα1-4)GlcNAc, K_D: 0.39 µM) signifying a higher preference for α1-2 Fuc over α1-3 or α1-4 linkages, irrespective of their location on the chain. Blood Group A (7K, GalNAcα1-3(Fucα1-2)Gal, K_D: 0.34 µM) showed a lower affinity compared to Blood group B possibly due to the presence of a terminal *N*-acetyl-D-galactosamine (GalNAc) whereas Lactodifucotetraose (7L, Fucα1-2Galβ1-4(Fucα1-3)Glc, K_D: 0.57 µM) displayed a 3-fold lower affinity compared to both Blood group H disaccharide or Blood group B, possibly due to binding occurring at the non-terminal α1-3 Fuc instead of the preferred α1-2 Fuc, though this would need to be further explored.

SPR of BT_3781 also revealed moderate binding (K_D range of 2.66 µM to 50 µM) to terminal Gal across all terminal Gal structures measured with the exception of TF antigen (89, Galβ1-3GalNAc), which did not bind. Terminal Gal affinity though was to a much lower degree than Fuc containing glycans indicating that the absence of a Fuc residue may greatly reduce the affinity of BT_411 to glycans with a terminal Gal. Standard deviations did vary somewhat between results indicating that the degree to which some glycans bound (their K_D) varied between repeat runs, however all are still within the acceptable margin of distribution.

Table 3.3. BT_3781 SPR analysis showing the mean K_D (μM) and SD of triplicate runs for each glycan. If SD was > 1 then K_D results was deemed unreliable and not included.

ID	Name	Structure	$\bar{x}$ K_D (μM) ± SD
7F	Blood group H disaccharide	Fucα1-2Gal	0.17 ± 0.01
18D	Blood group B (Tetraose type 5)	Galα1-3(Fucα1-2)Galβ1-4Glc	0.21 ± 0.07
7K	Blood Group A trisaccharide	GalNAcα1-3(Fucα1-2)Gal	0.34 ± 0.12
19J	Lewis^x tetrasaccharide	Galβ1-4(Fucα1-3)GlcNAcβ1-3Gal	0.39 ± 0.04
7J	Lewis^a	Galβ1-3(Fucα1-4)GlcNAc	0.39 ± 0.22
2E	Xeno-antigen	Gala1-4Galβ1-4GlcNAc	0.49 ± 0.01
7L	Lactodifucotetraose (LDFT)	Fucα1-2Galβ1-4(Fucα1-3)Glc	0.57 ± 0.27
1N	α1-3 Galactobiose	Gala1-3Gal	0.79 ± 0.03
1H	Lacto-N-neotetraose	Galβ1-4GlcNAcβ1-3Galβ1-4Glc	0.82 ± 0.17
1F	asialo GM1	Galβ1-3GalNAcβ1-4Galβ1-4Glc	0.86 ± 0.28
1P	Linear B-6 Trisaccharide	Gala1-3Galβ1-4Glc	1.10 ± 0.95
1C	β-1-4-galactosyl-galactose	Galβ1-4Gal	1.46 ± 0.86
1I	Lacto-N-neohexaose	Galβ1-4GlcNAcβ1-6(Galβ1-4GlcNAcβ1-3)Galβ1-4Glc	2.66 ± 0.85
4B	N,N',N''-Triacetyl chitotriose	GlcNAcβ1-4GlcNAcβ1-4GlcNAc	-
4D	N,N',N'',N''',N'''',N''''-Hexaacetyl chitohexaose	GlcNAcβ1-4GlcNAcβ1-4GlcNAcβ1-4GlcNAcβ1-4GlcNAcβ1-4GlcNAc	-
4A	N,N'-Diacetyl chitobiose	GlcNAcβ1-4GlcNAc	-
4C	N,N',N'',N'''-Tetraacetyl chitotetraose	GlcNAcβ1-4GlcNAcβ1-4GlcNAcβ1-4GlcNAc	-
13K	Chondroitin sulfate	(GlcA/IdoAβ1-3(±4/6S)GalNAcβ1-4)n (n<250)	-
12N	Chondroitin-4-sulfate	ΔUA-GalNAc-4S	-
5B	Bianntennary N-linked core pentasaccharide	GlcNAcβ1-2Manα1-6(GlcNAcβ1-2Manα1-3)Man	-
5A	β1-2 N-Acetylglucosamine-mannose	GlcNAcβ1-2Man	-
5C	α1-2-Mannobiose	Manα1-2Man	-
5F	α1-6-Mannobiose	Manα1-6Man	-
15E	Hyaluronan fragment 50kDa	NA	-
15G	Hyaluronan fragment 70kDa	NA	-
10A	Sialyl-Lewis^a tetraose	Neu5Acα2-3Galβ1-3(Fucα1-4)GlcNAc	-
10B	Sialyl-Lewis^x tetraose	Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAc	-
89	TF antigen	Galβ1-3GalNAc	-
292	Sialylated TF antigen	Neu5Acα2-3Galβ1-3GalNAc	-

PSL-2

SPR analysis of PSL-2 indicated high affinity to both Fuc containing glycans and glycans with a terminal Gal (Table 3.4. and Figure A1.7.). Binding structure comparison analyses indicates that PSL-2 has a strong affinity towards Blood group B (18D, Galα1-3(Fucα1-2)Galβ1-4Glc, K_D: 0.18 µM) with a non-terminal α1-2 Fuc and a terminal α1-3 Gal; a roughly 2-fold higher affinity as compared to Lactodifucotetraose (7L, Fucα1-2Galβ1-4(Fucα1-3)Glc, K_D: 0.39 µM) and Blood group H disaccharide (7F, Fucα1-2Gal, K_D: 0.41 µM). Moreover, PSL-2's affinity to Blood group B is 3-fold greater than Lewis group x (19J, Galβ1-4(Fucα1-3)GlcNAcβ1-3Gal, K_D: 0.54 µM) and Lewis group a (7J, Galβ1-3(Fucα1-4)GlcNAc, K_D: 0.62 µM) and 4-fold greater then Blood group A (7K, GalNAcα1-3(Fucα1-2)Gal, K_D: 0.72 µM). This indicates a preference towards glycan moieties with non-terminal α1-2 Fuc residues and a terminal Gal; further substantiated by blood group H disaccharide and Lactodifucotetraose of which both contain terminal Fuc residues rather than non-terminal. Likewise this is also seen with Lewis groups x and a, both with α1-3 and α1-4 non-terminal Fuc (respectively) rather than the PSL-2 preferred α1-2 Fuc linkage. Blood group A does contain a non-terminal α1-2 Fuc however, its terminal GalNAc may cause instability. PSL-2 also shows affinity for terminal Gal moieties with best recognition of β-1-4-galactosyl-galactose (1C, Galβ1-4Gal, K_D: 0.32 µM). Interestingly, β-1-4-galactosyl-galactose showed approximately 3-fold greater affinity compared to α1-3 Galactobiose (1N, Gala1-3Gal K_D: 0.84 µM) and Linear B-6 Trisaccharide (1P, Gala1-3Galβ1-4Glc, K_D: 0.93 µM). This potentially indicates a preference towards terminal β1-4 Gal linkages over α1-3 Gal; unless a non-terminal Fuc is present as was the case with Blood group B. SPR analysis also showed binding to chitin based terminal GlcNAc structures, however with a 3 to 4-fold reduced affinity compared to Blood group B.

Table 3.4. PSL-2 SPR analysis showing the mean K_D (μM) and SD of triplicate runs for each glycan. If SD was > 1 then K_D results was deemed unreliable and not included.

ID	Name	Structure	$\bar{x}$ K_D (μM) ± SD
18D	Blood group B (Tetraose type 5)	Galα1-3(Fucα1-2)Galβ1-4Glc	0.18 ± 0.18
1C	β-1-4-galactosyl-galactose	Galβ1-4Gal	0.32 ± 0.14
7L	Lactodifucotetraose (LDFT)	Fucα1-2Galβ1-4(Fucα1-3)Glc	0.39 ± 0.12
7F	Blood group H disaccharide	Fucα1-2Gal	0.41 ± 0.34
1F	asialo GM1	Galβ1-3GalNAcβ1-4Galβ1-4Glc	0.45 ± 0.38
2E	Xeno-antigen	Gala1-4Galβ1-4GlcNAc	0.45 ± 0.56
19J	Lewis^x tetrasaccharide	Galβ1-4(Fucα1-3)GlcNAcβ1-3Gal	0.54 ± 0.41
7J	Lewis^a	Galβ1-3(Fucα1-4)GlcNAc	0.62 ± 0.76
4A	N,N'-Diacetyl chitobiose	GlcNAcβ1-4GlcNAc	0.63 ± 0.57
4C	N,N',N'',N''-Tetraacetyl chitotetraose	GlcNAcβ1-4GlcNAcβ1-4GlcNAcβ1-4GlcNAc	0.70 ± 0.52
7K	Blood Group A trisaccharide	GalNAcα1-3(Fucα1-2)Gal	0.72 ± 0.03
4B	N,N',N''-Triacetyl chitotriose	GlcNAcβ1-4GlcNAcβ1-4GlcNAc	0.80 ± 0.80
1N	α1-3 Galactobiose	Gala1-3Gal	0.84 ± 0.65
4D	N,N',N'',N''',N'''',N''''-Hexaacetyl chitohexaose	GlcNAcβ1-4GlcNAcβ1-4GlcNAcβ1-4GlcNAcβ1-4GlcNAcβ1-4GlcNAc	0.86 ± 0.59
1P	Linear B-6 Trisaccharide	Gala1-3Galβ1-4Glc	0.93 ± 0.83
1H	Lacto-N-neotetraose	Galβ1-4GlcNAcβ1-3Galβ1-4Glc	-
15G	Hyaluronan fragment 70kDa	NA	-
15E	Hyaluronan fragment 50kDa	NA	-
13K	Chondroitin sulfate	(GlcA/IdoAβ1-3(±4/6S)GalNAcβ1-4)n (n<250)	-
12N	Chondroitin-4-sulfate	ΔUA-GalNAc-4S	-
1I	Lacto-N-neohexaose	Galβ1-4GlcNAcβ1-6(Galβ1-4GlcNAcβ1-3)Galβ1-4Glc	-
5A	β1-2 N-Acetylglucosamine-mannose	GlcNAcβ1-2Man	-
5B	Bianntennary N-linked core pentasaccharide	GlcNAcβ1-2Manα1-6(GlcNAcβ1-2Manα1-3)Man	-
5C	α1-2-Mannobiose	Manα1-2Man	-
5F	α1-6-Mannobiose	Manα1-6Man	-
10A	Sialyl-Lewis^a tetraose	Neu5Acα2-3Galβ1-3(Fucα1-4)GlcNAc	-
10B	Sialyl-Lewis^x tetraose	Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAc	-
89	TF antigen	Galβ1-3GalNAc	-
292	Sialylated TF antigen	Neu5Acα2-3Galβ1-3GalNAc	-

2.3.4. Docking analysis

BT_411

A structure-based alignment search was performed of the RCSB PDB database using SPOT-Struc to identify the highest matched proteins with known ligand affinity matching that of the SPR results for BT_411. The highest match was a Family 89 glycoside hydrolase from *Clostridium perfringens* that retuned an SP-align score of 0.75 (1 = exact match) with a 15% ID alignment, 132 aligned residue pairs at a RMSD of 2.10 Å (in comparison with BT_411). The crystal structure of Family 89 glycoside hydrolase contained a bound GlcNAc. A superimposed structural alignment with Family 89 glycoside hydrolase and BT_411 was performed to identify a possible active binding site on BT_411. Using Autodock Vina, analysis consisted of a 40 Å box surrounding the predicted BT_411 active site with a GlcNAc ligand used to inspect binding affinity. Docking analysis shows GlcNAc bound within an active site composed of residues Gly152, Asp151, Thr149, Leu153, Arg154, Val135, Lys60, Trp57 and Ala94 (all within 5 Å of ligand) with a predicted binding affinity of -4.6 kcal/mol (Figure 3.6.).

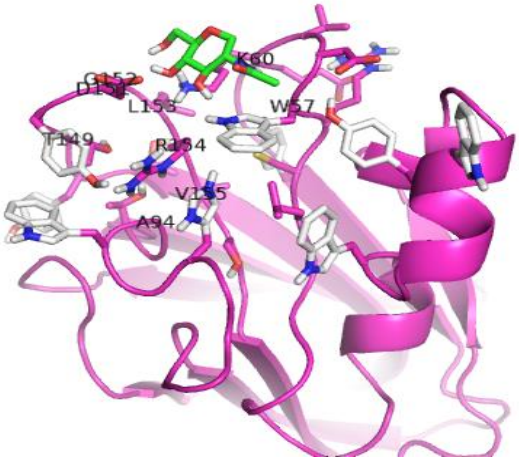


Figure 3.6. Docking of GlcNAc (green) to BT_411 using Autodock Vina with default parameters, showing possible GlcNAc binding orientation with a predicted binding affinity of -4.6 kcal/mol. Predicted binding residues (grey) are within 5 Å of docked GlcNAc.

BT_3781

Using SPOT-Struc, a structure-based alignment comparison search of the RCSB PDB database was performed between BT_3781 and any CBPs with a bound ligand similar to any of the ligands shown to have affinity with BT_3781 (as was identified by SPR). A 1,2-α-L-fucosidase from *Bifidobacterium bifidum* returned an SP-align score of 0.62 (1 = exact match) with a 5.9% ID alignment, 325 aligned residue pairs and a RMSD of 3.17 Å (in comparison with BT_3781). The 1,2-α-L-fucosidase crystal structure contains a bound Fucα1-2Gal. A superimposed structural alignment of 1,2-α-L-fucosidase and BT_3781 was performed to identify a possible binding site on BT_3781. Using Autodock Vina, analysis consisted of a 40 Å box surrounding the predicted BT_3781 active site with a Fucα1-2Gal ligand used to inspect binding affinity. Docking analysis shows Fucα1-2Gal bound within an active site composed of residues His396, Asn348, Ser265, Asp266, Asp267, Arg119, Glu439, Arg451, Phe454, Trp416, Ile103,Thr100 and Asp110 (all within 5 Å of ligand) with a predicted binding affinity of -6.8 kcal/mol (Figure 3.7.).

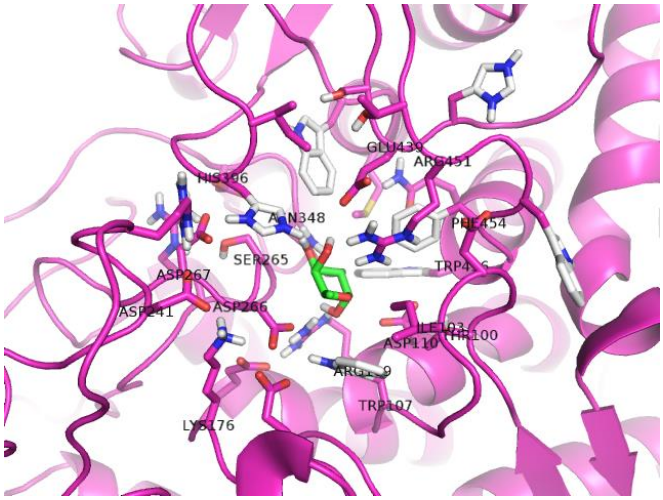


Figure 3.7. Docking of Fucα1-2Gal (green) to BT_3781 using Autodock Vina with default parameters, showing possible Fucα1-2Gal binding orientation with a predicted binding affinity of -6.8 kcal/mol. Predicted binding residues (grey) are within 5 Å of docked Fucα1-2Gal.

3.4. Discussion

Utilising computational prediction as a director of experimental research is rapidly emerging as a cost effective approach to gaining insights into complex biological process in a fast and reliable manner [37, 38]. However, the evolution of modern *in silico* guided experimentation relies on testing and retesting *in silico* predictions in the laboratory, so that the learning aspect of these novel algorithms, the predictive element in which lies their value, can produce predictions with a level of reliability that will establish *in silico* prediction as a trusted staple of biochemical research [39]. SPOT-Struc, a relatively new predictive software, identified two hypothetical proteins BT_411 and BT_3781 as putative CBPs, which we experimentally validated in this study. Moreover, I characterised the recombinant variant of a recently isolated novel mushroom lectin PSL-2, that not only provided a positive control lectin for BT_411 and BT_3781, but also allowed us to validate the glycan binding preferences as identified by PSL-2 crystal structure.

BT_411

Amino acid derived Mw, SDS-PAGE and SEC inferred that recombinant BT_411 had a Mw of 18 kDa and was revealed to be monomeric in solution, which is consistent with its crystal structure derived oligomeric state (RCSB PDB ID: 3HNM). SPR data indicated that BT_411 is a CBP, as predicted by SPOT-Struc however, its glycan affinity profile is different to the SPOT-Struc matched template, the *C. Perfringens* family 32 CBM protein. The family 32 CBM protein has a known preference towards galactose (Gal), *N*-Acetyl-D-lactosamine (LacNAc) and type II blood group H-trisaccharide. Even though BT_411 has been identified as having a Gal binding like domain, we found that it does not bind to Gal but rather has a high preference for terminal GlcNAc and sialylated glycan moieties with a terminal α2,3-linked Neu5Ac. Interestingly, additional research into the *C. Perfringens*

family 32 CBM protein did indicate affinity towards terminal GlcNAcs [40], unusual for Gal recognizing CBMs with the β-sandwich scaffold common to this family. Nevertheless, at the time of the SPOT-Struc predicted match the structural information for the GlcNAc binding family 32 CBM was available as a comparison template (RCSB PDB ID: 2WDB) but was not chosen as the template match for BT_411. This indicates that structural alignment by SPOT-Struc was able to identify BT_411 as a CBP but matching exact ligand affinity proves to be a more challenging task. A similar discrepancy occurred with the SPOT-Struc CBP prediction of novel lectin YesU that was correctly identified as a CBP but showed dissimilar ligand affinity to that of the SPOT-Struc predicted ligand affinity [29].

A SPOT-Struc structural alignment search of structurally similar templates with either a bound GlcNAc or Neu5Ac returned no statistically significant matches with a bound Neu5Ac (SPalign score < 0.3). A separate structural alignment comparison was performed with Wheat Germ Agglutinin Isolectin (WGA) from species *Triticum aestivum*, a known dual sialic acid and GlcNAc binding lectin [41] but homology was low (Q score < 0.20; 1 = exact match). The highest SPOT-Struc matched template was to that of a Family 89 glycoside hydrolase from *C. perfringens* (AgnC) bound to GlcNAc; a mainly beta class protein with sandwich architecture and a Jelly Roll topology similar to BT_411. BT_411 residues Leu153 and Trp57 involved with binding of GlcNAc (as shown by docking analysis), were likewise conserved in the GlcNAc binding site of AgnC (RCSB PDB ID: 2VCA). AgnC has been shown to specifically cleave GlcNAc from GlcNAcα1,4Galβ1 on O-glycans indicting α-*N*-Acetylglucosaminidases (αGNases) activity [42]. This may indicate similar *N*-Acetylglucosaminidases (GNase) activity with BT_411, though SPR indicates BT_411 may target β1-4 linked GlcNAc instead of α linkages as seen in AgnC. An indication perhaps of BT_411 displaying β-N-Acetylglucosaminidase (βGNase) activity.

βGNase are known to catalyse the hydrolysis of terminal GlcNAc [43] and would account for the strong BT_411 affinity with terminal GlcNAc chitin-based oligosaccharides seen in SPR. BT_411 may also target non-terminal GlcNAc β1-4 and β1-3 linkages which would account for affinity shown to sialic acids with non-terminal GlcNAc β1-4 and β1-3 linkages. A terminal Neu5Ac may account for stability in the cases of binding to a non-terminal β linked GlcNAc, though much further testing would be needed to confirm this. Furthermore, BT_411 showed binding with sialylated TF antigen (Neu5Acα2-3Galβ1-3GalNAc) which does not contain a GlcNAc. There is precedence for GlcNAc/GalNAc preferred lectins [44] however a structural alignment with the known GlcNAc/GalNAc preferred lectin Tachylectin-2 From *Tachypleus Tridentatus* (Japanese Horseshoe Crab) (RCSB PDB ID: 1TL2) did not show a high similarity (Q score < 30). As such further study is needed to better clarify BT_411's binding of sialic acids as observed by SPR.

BT_411 did not bind to any other terminal GlcNAc structures tested including β1-2 *N*-acetylglucosamine-mannose (5A; GlcNAcβ1-2Man) and bianntennary *N*-linked core pentasaccharide (5B; GlcNAcβ1-2Manα1-6(GlcNAcβ1-2Manα1-3)Man), perhaps indicating that mannose (Man) may interfere with binding and that a non-terminal GlcNAc is required in addition to a terminal GlcNAc or Neu5Ac for stable binding. BT_411 also displayed a lack of affinity for structures that only contained non-terminal GlcNAc residues, again supporting the need for a terminal GlcNAc or Neu5Ac for binding stability. Furthermore, BT_411 showed no affinity towards glycan structures with terminal Gal, Fuc and glycosaminoglycan (GAGs) including larger hyaluronan fragments of 50 kDa and 70 kDa.

Many invasive fungi and parasitic insects contain chitin in their cell walls and exoskeletons (respectively); a vulnerability that chitin-specific CBM based drugs can exploit as anti-

fungal/parasitic treatments [45]. Therefore, further research into BT_411 as βGNase would be of great benefit to aid in the treatment of parasitic or fungal related infections. Moreover, the data reported here now provides additional inputs and parameters that can refine SPOT-Struc predicted template options for ligand affinity. Such as comparing local sequence similarities of amino acid sequence fragments, amongst proteins already matched by phylogenetic taxonomy and structural similarity [46].

BT_3781

Amino acid derived Mw, SDS-PAGE and SEC indicated that recombinant BT_3781 had a Mw of 52 kDa and was primarily monomeric in solution; a result contrary to its crystal structure derived oligomeric state as a homo-dimer (RCSB PDB ID: 2P0V). However crystal structure derived from oligomeric state is not always reliable in that artificial oligomers can form during the crystallisation process and as such oligomeric state based on crystal structure alone may not truly represent native oligomeric state of the protein [47]. Further to this the crystal structure for BT_3781 also showed residue Cys322 is responsible for a single disulfide bond between the two BT_3781 homodimer chains however, this may be a result of subsequent oxidation during crystallisation and may be inaccurate. Though if hypothetically the crystal structure is accurate, then it is possible that the disulphide bond may only be partially responsible for the homodimer state, as docking analysis shows that the ligand binding pocket is located between chain A and B of the BT_3781 homodimer, perhaps indicating that a bound ligand is needed for homodimer oligomeric stability. As such, BT_3781 may retain in an inactive monomeric state until a ligand is present at which point changing to an active dimeric state through binding of the ligand [48].

SPOT-Struc matched BT_3781 with protein Glycoside Hydrolase Family 15 Glucoamylase, a starch binding CBM whose specific glycan affinity profile is yet to be defined. SPR analysis of BT_3781 confirmed its classification as a CBP by SPOT-Struc, indicating it has high preferential binding to fucosylated glycans with both terminal and non-terminal Fuc residues as well as moderate affinity to glycan moieties with terminal Gal. BT_3781 specifically appears to be a strong blood group lectin with a 2-fold (or greater) affinity for Blood group H disaccharide (7F; Fucα1-2Gal) and Blood group B - Tetraose type 5 (18D; Galα1-3(Fucα1-2)Galβ1-4Glc) over other Fuc and Gal containing glycans that returned positive for affinity. BT_3781 showed a distinct preference for α1-2 Fuc linkages over α1-3 or α1-4, regardless of their position on the chain (terminal or non-terminal). However the presence of a terminal *N*-Acetyl-D-galactosamine (GalNAc) in Blood group A trisaccharide (7K; GalNAcα1-3(Fucα1-2)Gal) may cause instability despite the presence of a non-terminal Fucα1-2, as BT_3781 showed a 2-fold decreased affinity in comparison to Blood group H disaccharide. In addition, BT_3781 showed preference for terminal Gal glycan moieties, though to a far lesser extent then Fuc (2 to 13-fold less), perhaps indicating that a terminal Gal may provide binding stability when in addition to a Fuc (terminal or non-terminal) or weak binding in the absence of a non-terminal Fuc. Interestingly, BT_3781 did not bind to TF antigen (89; Galβ1-3GalNAc) with a terminal Gal, perhaps due to the presence of a GalNAc that may interfere with binding of the shorter chained structure. BT_3781 showed no affinity towards glycan structures with GlcNAc (terminal), GAGs, Man, Sialylated glycans or larger Hyaluronan fragments of 50 kDa and 70 kDa.

A SPOT-Struc structural alignment search of structurally similar templates with either a bound Fuc or Gal returned a high structural similarity match with an 1,2-α-L-fucosidase

(AfcA) from *B. bifidum* bound to Blood group H disaccharide (Fucα1-2Gal). AfcA is a mainly α-class, α/α barrel glycosyltransferase similar to BT_3781. BT_3781 docking analysis predicted binding residues His396, Asn348, Arg119, Glu439, Arg451 and Trp416 were likewise conserved in the binding site for AfcA (RCSB PDB ID: 2EAE). AfcA is a glycoside hydrolase that targets the glycosidic linkage of Fucα1-2Gal [49]. Glycoside hydrolases catalyse the hydrolysis of glycosidic bonds in a variety of sugars and are fairly common enzymes [50]. Given the similar conserved binding residues and concurrent structural data shown between AfcA and BT_3781, a shared function is likely, indicating that BT_3781 is a newly defined fucosidase CBM with blood group preferentiality.

As such research into BT_3781 could be directed towards studies examining the breakdown of blood type antigens to create other blood types for transfusion; a means of essentially making all blood universal [51]. Accordingly, future study into BT_3781 as a blood group fucosidase should proceed involving further BT_3781 bound ligand structure crystallography to clarify mechanism of action as well as human cytotoxicity assays to determine toxicity in humans.

PSL-2

SDS-PAGE indicated that recombinant PSL-2 has a subunit Mw of 16 kDa, while the SEC Mw weight was 43.74 kDa, suggesting PSL-2 exists primarily as a trimer in solution; a result contrary to the presumed oligomeric state for native PSL-2 as a tetramer [35]. Computational analysis of PSL-2 crystal structure using PyMOL indicates no cysteine residues are involved with PSL-2's oligomeric state and that bonding between chains instead is a result of non-covalent interactions such as van der Waals interactions and hydrogen bonding [52]. As such the discrepancy in oligomeric state may instead be due to a

lack of environmental factors responsible for correct folding present in the native organism that were not present in this experiment. These may include differences in temperature or pH of the expression and/or purification medium/buffers compared to that found in the native organism *P. schweinitzii* [53]. Further to this, MS of PSL-2 with and without a bound ligand could be performed to determine if oligomeric state is also ligand dependent.

Previous studies of PSL-2 indicated binding pockets for Gal and Fuc, of which SPR analysis confirmed. PSL-2 showed high preferential binding for blood group B (18D; Galα1-3(Fucα1-2)Galβ1-4Glc) with an approximate 2-fold greater affinity then the next highest binder of β-1-4-galactosyl-galactose (Galβ1-4Gal). The inclusion of a non-terminal α1-2 Fuc linkage on a chain with a terminal α1-3 Gal, as with blood group B, appears to be a significant requirement for strong binding as all other Fuc/Gal combinations (or Gal alone) tested showed a significantly weaker interaction. PSL-2 also showed an affinity for chitin based glycans with repeating GlcNAcs which is an unexpected result due to PSL-2s preference for Gal and Fuc residues; a finding that may warrant further study. A Galectin that binds Fucs is rare as such a pairwise structural alignment was performed using PDBeFold with Fungal Galectin CGL2 known to bind Fuc, specifically Blood group H (RSCB PDB ID: 1ULD). Results showed a high structural similarity match with a Q score of 0.69 (1= exact match) with 92% ID alignment and 142 aligned residue pairs at an average root-mean-squared distance (RMSD) of 1.29 Å, supporting the Fuc binding functionality of PSL-2. Furthermore, CGL2 also exhibited affinity for Thomsen-Friedenreich antigen, a known tumor antigen on the surface of cancer cells [54]. Though PSL-2 did not show initial binding to TF antigen in our SPR analysis it may warrant further exploration given the relative medical applications of a TF antigen binding lectin, as a biosensor or means of targeting relevant cancer cells [55]. As well, PSL-2's affinity for blood

group B antigen and Histo-blood group antigens (Lewis$^{x/a}$) should be investigated further as it may prove valuable in research relating to competitive binding of blood group antigens as a treatment of blood antigen specific viruses, such as norovirus that result in severe gastroenteritis [56].

SPOT-Struc successfully predicted two novel bacterial CBPs; BT_411 and BT_3781 from *B. thetaiotaomicron*, as was inferred by SPR and docking analysis. Furthermore, successful characterisation of the known lectin PSL-2 by SPR confirmed its ligand affinity (based on previous crystallography data) and provided a positive control for the characterisation of BT_411 and BT_3781. There were several limitations that could be addressed in future iterations of this study identical to what was defined in Chapter 2 (page 50) regarding SEC, variations in recombinant protein structure compared to native, and SPR. In addition to this a glycan array would have provided a broader spectrum of potential glycans to test in SPR [29] but as we did not have access to functioning array equipment at the time of this study this could be explored in future experimentation. The challenging nature of correctly identifying ligand affinity using *in silico* prediction reveals an aspect of improvement for SPOT-Struc that the data from this experiment may assist with. Furthermore, future iterations of this experiment may warrant the inclusion of a negative control as determined by SPOT-Struc, to see if SPOT-Struc non-CBP predictions are as accurate. This notwithstanding, the successful prediction of CBPs by SPOT-Struc endorses the application of *in silico* prediction as a guide for experimental CBP research.

References

1. Brinchmann, M.F., Patel, D.M. & Iversen, M.H. (2018) The role of galectins as modulators of metabolism and inflammation. *Mediators of Inflammation*, 2018: 9186940. https://doi.org/10.1155/2018/9186940

2. Brown, G.D., Willment, J.A. & Whitehead, (2018) L. C-type lectins in immunity and homeostasis. *Nature Reviews Immunology*, 18:374–389. https://doi.org/10.1038/s41577-018-0004-8

3. Vasta, G.R., Amzel, L.M., Bianchet, M.A., Cammarata, M., Feng, C. & Saito, K. (2017) F-type lectins: a highly diversified family of fucose-binding proteins with a unique sequence motif and structural fold, involved in self/non-self-recognition. *Frontiers in Immunology*, 8:1648. https://doi.org/10.3389/fimmu.2017.01648

4. Lehmann, F., Tiralongo, E. & Tiralongo, J. (2006) Sialic acid-specific lectins: occurrence, specificity and function. *Cellular and Molecular Life Sciences*, 63:1331–1354. https://doi.org/10.1007/s00018-005-5589-y

5. Endo, Y., Matsushita, M. & Fujita, T. (2011) The role of ficolins in the lectin pathway of innate immunity. *The International Journal of Biochemistry & Cell Biology*, 43(5):705-12. https://doi.org/10.1016/j.biocel.2011.02.003

6. Boraston, A.B., Bolam, D.N., Gilbert, H.J. & Davies, G.J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochemical Journal*, 382(3):769–781. https://doi.org/10.1042/BJ20040892

7. Broeker, N.K. *et al.* (2017) Complex carbohydrate recognition by proteins: Fundamental insights from bacteriophage cell adhesion systems. *Perspectives in Science*, 11:45-52. https://doi.org/10.1016/j.pisc.2016.10.001

8. Gabius, H.J., André, S., Kaltner, H. & Siebert, H.C. (2002) The sugar code: functional lectinomics. *Biochimica et Biophysica Acta*, 1572(2-3):165-77. https://doi.org/10.1016/S0304-4165(02)00306-9

9. Caramelo, J.J. & Parodi, A.J. (2007) How sugars convey information on protein conformation in the endoplasmic reticulum. *Seminars in Cell and Developmental Biology*, 18:732–742. https://doi.org/10.1016/j.semcdb.2007.09.006

10. Cambi, A., Figdor, C.G. (2005) Levels of complexity in pathogen recognition by C-type lectins. *Current Opinion in Immunology*, 17(4):345-351. https://doi.org/10.1016/j.coi.2005.05.011

11. Mayer, S., Raulf, M. & Lepenies, B. (2017) C-type lectins: their network and roles in pathogen recognition and immunity. *Histochemistry and Cell Biology*, 147:223–237. https://doi.org/10.1007/s00418-016-1523-7

12. Ricci-Azevedo, R., Roque-Barreira, M.C. & Gay, N.J. (2017) Targeting and recognition of toll-like receptors by plant and pathogen lectins. *Frontiers in Immunology*, 8:1820. https://doi.org/10.3389/fimmu.2017.01820

13. Berg, J.M., Tymoczko, J.L. & Stryer, L. (2002) Biochemistry: 5th edition, section 11.4, lectins are specific carbohydrate-binding proteins. *New York: W H Freeman.* Available from*:* https://www.ncbi.nlm.nih.gov/books/NBK22545/

14. Christiansen, M.N., Chik, J., Lee, L., Anugraham, M., Abrahams, J.L. & Packer, N.H. (2014) Cell surface protein glycosylation in cancer. *Proteomics*, 14(4-5):525-46. https://doi.org/10.1002/pmic.201300387

15. Otto, D.M.E. *et al.* (2011) An expression system for screening of proteins for glycan and protein interaction. *Analytical Biochemistry*, 411:261–270. https://doi.org/10.1016/j.ab.2010.12.036

16. James, A.W., Ravi, C., Srinivasan, M. & Nachiappan, V. (2019) Crosstalk between protein N-glycosylation and lipid metabolism in Saccharomyces cerevisiae. *Scientific Reports*, 9:14485. https://doi.org/10.1038/s41598-019-51054-7

17. Spiro, R.G. (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*, 12(4):43R–56R. https://doi.org/10.1093/glycob/12.4.43R

18. Stanley, P. (2011) Golgi glycosylation. *Cold Spring Harbor Perspectives in Biology*, 3(4):pii:a005199. https://doi.org/10.1101/cshperspect.a005199.

19. Breen, K.C., Coughlan, C.M. & Hayes, F.D. (1998) The role of glycoproteins in neural development, function, and disease. *Molecular Neurobiology*, 16:163–220. https://doi.org/10.1007/BF02740643

20. Gennarini, G. *et al.* (2017) The role of Gpi-anchored axonal glycoproteins in neural development and neurological disorders. *Molecular and Cellular Neuroscience*, 81:49-63. https://doi.org/10.1016/j.mcn.2016.11.006

21. Robajac, D. *et al.* (2019) Glycoanalysis of the placental membrane glycoproteins throughout placental development. *Mechanisms of Ageing and Development*, 183:111151. https://doi.org/10.1016/j.mad.2019.111151

22. Racine, T., Kobinger, G.P. & Arts, E.J. (2017) Development of an HIV vaccine using a vesicular stomatitis virus vector expressing designer HIV-1 envelope

glycoproteins to enhance humoral responses. *AIDS Research and Therapy*, 14:55. https://doi.org/10.1186/s12981-017-0179-2

23. Melero, J.A., Mas, V. & McLellan, J.S. (2017) Structural, antigenic and immunogenic features of respiratory syncytial virus glycoproteins relevant for vaccine development. *Vaccine*, 35(3):461-468. https://doi.org/10.1016/j.vaccine.2016.09.045

24. Azevedo, R. *et al.* (2018) CD44 glycoprotein in cancer: a molecular conundrum hampering clinical applications. Clinical Proteomics, 15(22). https://doi.org/10.1186/s12014-018-9198-9

25. Song, E. & Mechref, Y. (2015) Defining glycoprotein cancer biomarkers by MS in conjunction with glycoprotein enrichment. *Biomarkers in Medicine*, 9(9):835–844. https://doi.org/10.2217/bmm.15.55

26. Ferrero, E., Dunham, I. & Sanseau, P. (2017) In silico prediction of novel therapeutic targets using gene–disease association data. *Journal of Translational Medicine*, 15:182. https://doi.org/10.1186/s12967-017-1285-6

27. Moumbock, A., Li, J., Mishra, P., Gao, M. & Günther, S. (2019) Current computational methods for predicting protein interactions of natural products. *Computational and Structural Biotechnology Journal*, 17:1367–1376. https://doi.org/10.1016/j.csbj.2019.08.008

28. Zhao, H., Yang, Y., von Itzstein, M. & Zhou, Y. (2014) Carbohydrate-binding protein identification by coupling structural similarity searching with binding affinity prediction. *Journal of Computational Chemistry*, 35(30):2177-83. https://doi.org/10.1002/jcc.23730

29. Tiralongo, J. *et al.* (2018) YesU from Bacillus subtilis preferentially binds fucosylated glycans. *Scientific Reports*, 8:13139. https://doi.org/10.1038/s41598-018-31241-8

30. Knudsen, M. & Wiuf, C. (2010) The CATH database. *Human Genomics*, 4(3):207–212. https://doi.org/10.1186/1479-7364-4-3-207

31. Sujatha, M.S. & Balaji, P.V. (2004) Identification of common structural features of binding sites in galactose-specific proteins. *Proteins*, 55(1):44-65. https://doi.org/10.1002/prot.10612

32. Ficko-Blean, E. & Boraston, A.B. (2006) The Interaction of a carbohydrate-binding module from a Clostridium perfringens N-Acetyl-β-hexosaminidase with its carbohydrate receptor. *The Journal of Biological Chemistry*, 281(49):37748 – 37757. https://doi.org/10.1074/jbc.M606126200

33. Fox, N.K., Brenner, S.E. & Chandonia, J.M. (2014) SCOPe: structural classification of proteins--extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Research*, 42:D304–D309. https://doi.org/10.1093/nar/gkt1240

34. Bott, R. *et al.* (2008) Three-dimensional structure of an intact glycoside hydrolase family 15 glucoamylase from Hypocrea jecorina. *Biochemistry,* 47(21):5746-5754. https://doi.org/10.1021/bi702413k

35. Rouf, R. (2015) Isolation and characterisation of novel lectins from Australian macrofungi. *Doctoral thesis, Griffith University, Gold Coast, Australia*. Available from: http://hdl.handle.net/10072/366328

36. Greenfield, N.J. (2006) Using circular dichroism spectra to estimate protein secondary. *Nature Protocols,* 1(6):2876–2890. https://doi.org/10.1038/nprot.2006.202

37. Li, F. *et al.* (2015) GlycoMine: a machine learning-based approach for predicting N-, C- and O-linked glycosylation in the human proteome. *Bioinformatics*, 31(9):1411–1419. https://doi.org/10.1093/bioinformatics/btu852

38. Larrañaga, P. *et al.* (2006) Machine learning in bioinformatics. *Briefings in Bioinformatics,* 7(1):86–112. https://doi.org/10.1093/bib/bbk007

39. Houk, K.N. & Liu, F. (2017) Holy grails for computational organic chemistry and biochemistry. *Accounts of Chemical Research*, 50(3):539-543. https://doi.org/10.1021/acs.accounts.6b00532

40. Ficko-Blean, E., & Boraston, A.B. (2009). N-acetylglucosamine recognition by a family 32 carbohydrate-binding module from Clostridium perfringens NagH. *Journal of Molecular Biology*, 390(2):208–220. https://doi.org/10.1016/j.jmb.2009.04.066

41. Peters, BP. *et al.* (1979) Interaction of wheat germ agglutinin with sialic acid. *Biochemistry,* (18)24:5505-5511. https://doi.org/10.1021/bi00591a038

42. Fujita, M., *et al.* (2011) Glycoside hydrolase family 89 alpha-N-acetylglucosaminidase from Clostridium perfringens specifically acts on GlcNAc alpha1,4Gal beta1R at the non-reducing terminus of O-glycans in gastric mucin. *The Journal of Biological Chemistry,* 286(8):6479–6489. https://doi.org/10.1074/jbc.M110.206722

43. Garrido, D. *et al.* (2012) Endo-β-N-acetylglucosaminidases from infant gut-associated bifidobacteria release complex N-glycans from human milk

glycoproteins. *Molecular & Cellular Proteomics*, 11(9):775–785. https://doi.org/10.1074/mcp.M112.018119

44. Beisel, H. G., Kawabata, S., Iwanaga, S., Huber, R. & Bode, W. (1999) Tachylectin-2: crystal structure of a specific GlcNAc/GalNAc-binding lectin involved in the innate immunity host defense of the Japanese horseshoe crab Tachypleus tridentatus. *The EMBO Journal*, 18(9):2313–2322. https://doi.org/10.1093/emboj/18.9.2313

45. Vega, K. & Kalkum, M. (2012) Chitin, chitinase responses, and invasive fungal infections. *International Journal of Microbiology*, 2012:920459. https://doi.org/10.1155/2012/920459

46. Karasev, D., Sobolev, B., Lagunin, A., Filimonov, D. & Poroikov, V. (2019) Prediction of protein-ligand interaction based on the positional similarity scores derived from amino acid sequences. *International Journal of Molecular Sciences*, 21(1):24. https://doi.org/10.3390/ijms21010024

47. Acharya, K.R. & Lloyd, M.D. (2004) The advantages and limitations of protein crystal structures. *Trends in Pharmacological Sciences*, 26(1):10-14. https://doi.org/10.1016/j.tips.2004.10.011

48. Rovira, X., Vivó, M., Serra, J., Roche, D., Strange, P. G. & Giraldo, J. (2009) Modelling the interdependence between the stoichiometry of receptor oligomerization and ligand binding for a coexisting dimer/tetramer receptor system. *British Journal of Pharmacology*, 156(1):28–35. https://doi.org/10.1111/j.1476-5381.2008.00031.x

49. Nagae, M. *et al.* (2007) Structural basis of the catalytic reaction mechanism of novel 1,2-alpha-L-fucosidase from Bifidobacterium bifidum. *Journal of Biological Chemistry*, 282(25):18497-509. https://doi.org/10.1074/jbc.M702246200

50. Bourne, Y. & Henrissat, B. (2001) Glycoside hydrolases and glycosyltransferases: families and functional modules. *Current Opinion in Structural Biology*, 11(5):593–600. https://doi.org/10.1016/s0959-440x(00)00253-0.

51. Olsson, M.L. & Clausen, H. (2008) Modifying the red cell surface: towards an ABO-universal blood supply. *British Journal of Haematology*, 140:3-12. https://doi.org/10.1111/j.1365-2141.2007.06839.x

52. Lodish, H. *et al.* (2000) Molecular Cell Biology: 4th edition. New York: *W. H. Freeman.* Available from: https://www.ncbi.nlm.nih.gov/books/NBK21726/

53. Chen, W., Xiao, H., Rizzo, A. N., Zhang, W., Mai, Y. & Ye, M. (2014) Endothelial nitric oxide synthase dimerization is regulated by heat shock protein 90 rather than

by phosphorylation. *PloS One*, 9(8):e105479. https://doi.org/10.1371/journal.pone.0105479

54. Dippold, W., Steinborn, A. & Meyer zum Büschenfelde, K.H. (1990) The role of the Thomsen-Friedenreich antigen as a tumor-associated molecule. *Environmental Health Perspectives*, 88:255–257. https://doi.org/10.1289/ehp.9088255

55. Gondim, A.C.S. *et al.* (2017) The potent anti-cancer activity of Dioclea Lasiocarpa lectin. *Journal of Inorganic Biochemistry*, 175:179-189. https://doi.org/10.1016/j.jinorgbio.2017.07.011.

56. Huang, P., Farkas, T., Zhong, W., Tan, M., Thornton, S., Morrow, A. L. & Jiang, X. (2005) Norovirus and histo-blood group antigens: demonstration of a wide spectrum of strain specificities and classification of two major binding groups among multiple binding patterns. *Journal of Virology*, 79(11):6714–6722. https://doi.org/10.1128/JVI.79.11.6714-6722.2005

Chapter 4

Conclusions and Future Directions

4.1. Conclusions

The aims of this study were to first express and purify SPOT-Ligand identified potential drug targets (Gene names: EF0414, ubiE, ftsZ5, Lebu_1328, acpD, and ubiH) for novel antimicrobial drug BDM-I, and SPOT-Struc predicted putative CBPs BT_411 and BT_3781. Following this the determination of a BDM-I possible mechanism of action through the characterisation of BDM-I's affinity with screened targets, as well as validate SPOT-Struc predicted CBPs by characterisation of their carbohydrate affinity. Finally, we intended to evaluate SPOT-Ligand and SPOT-Struc performance as bioinformatic aids in experimental research based on our experimental results.

To meet these objectives we expressed targets EF0414, UbiE, FtsZ5, Lebu_1328, AcpD, and UbiH, as well as putative CBPs BT_411, BT_3781 and novel mushroom lectin PSL-2 (as a positive control) in *E.coli* BL21 (DE3). We then purified them all to homogeneity using affinity chromatography and SEC with a resulting yield of between 0.08 mg and 6.73 mg per litre of *E. coli.* BDM-I protein target affinity was determined using SPR and BDM-I mechanism of action defined using *in silico* docking analysis with Autodock vina. Putative CBPs BT_411, BT_3781 and PSL-2 control lectin were analysed using SPR to define carbohydrate affinity and modelling analysis performed on BT_411 and BT_3781 to identify possible active site.

Using SPR we validated *S. typhimurium* azoreductase AcpD as a potential target for BDM-I. We showed that BDM-I binds easily (K_D of 0.58 µM) and quickly (k_a of 9.4 x 10^4 $M^{-1}s^{-1}$) to AcpD with slow dissociation (k_d of 9.7 x 10^{-4} s^{-1}). Docking analysis determined that BDM-I bound in the active site of AcpD via residues Tyr121, Phe163, Ala115, Ala113, Arg114, Arg60, Met59, Val56 and Asn98 with a binding energy of – 6.7 kcal/mol. A

structural alignment was performed between *P. aeruginosa* azoreductase PaAzoR, known to also have nitroreductase activity, and AcpD that showed a high structural alignment between the two proteins with conserved active site residues. As such, we identify a potential mechanism of action for BDM-I on nitroreductases, through the binding and possible reduction of BDM-I nitro group into toxic intermediates. In addition, we also identified a potential new functional designation of AcpD as a dual azoreductase and nitroreductase, something not seen before in AcpD. Our data demonstrates that SPOT-Ligand successfully screened positive BDM-I targets by structural comparison with a template (but not sequence predicted structures), validating a supportive role of bioinformatics in drug discovery.

SPR analysis confirmed SPOT-Struc predicted BT_411 and BT_3781 are CBPs. SPR analysis of BT_411 showed high affinity for GlcNAc and α2,3 linked Neu5Ac. A structure alignment search matched BT_411 to a Family 89 glycoside hydrolase (AgnC) from *C. perfringens*, known to bind GlcNAc. This in addition to docking analysis indicates that BT_411 is a putative CBM βGNase that may catalyse the hydrolysis of βGlcNAc linkages and that a terminal GlcNAc or Neu5Ac may be responsible for binding stability. As such BT_411 may be a potential new anti-parasitic and anti-fungal research candidate, given its predicted classification as a βGNase.

SPR analysis of BT_3781 revealed it to bind preferentially to Fuc and Gal residues, with high affinity for blood group antigens. A structural alignment search was performed on BT_3781 that matched it to 1,2-α-L-fucosidase (AfcA) from *B. bifidum* that binds blood group H disaccharide (Fucα1-2Gal). Furthermore, BT_3781 docking analysis indicated that BT_3781 bound Fucα1-2Gal with a predicted affinity of -6.8 kcal/mol. This suggests that there is a high probability that BT_3781 is a CBM, and putatively a blood group specific

fucosidase. Fucosidases are specific hydrolase that may have useful applications in blood antigen degradation research, such as seen in the development of a universal blood source.

Finally, prior PSL-2 analysis had shown affinity with Fuc and Gals which was confirmed by the SPR analysis, with PSL-2 showing high affinity for both Fuc and Gal containing oligosaccharides, in particular blood groups B, H and Histo blood groups (Lewis $^{x/a}$). PSL-2 also showed a structurally high match with the fungal lectin CGL2 which is known to bind cancer associated TF antigen, revealing a potential future study path for PSL-2 as a TF antigen binder. Moreover, successful characterisation of PSL-2 validates its role as a positive control for BT_411 and BT_3781 and in turn validates the experimental characterisation methods used to study these two novel lectins. We also further validated the support role that bioinformatics has in experimental research as shown by the SPOT-Struc successful prediction of two novel CBPs in BT_411 and BT_3781.

All project aims were achieved as all six drug targets and two predicted CBPs were expressed and purified to homogeneity for characterisation. A possible BDM-I mechanism of action via the binding of BDM-I nitro group to nitroreductases was defined. BT_411 and BT_3781 were both experimentally validated as novel CBPs (specifically CBMs) and additionally were indicated to be a putative βGNase and CBM fucosidase (respectively) by computational analysis. As such, the results of this study endorses both SPOT-Ligand and SPOT-Struc as valuable supportive aids in experimental research. Moreover, PSL-2 carbohydrate preference was experimentally characterised and nitroreductase activity not before seen in azoreductase acpD was identified.

4.2. Future directions

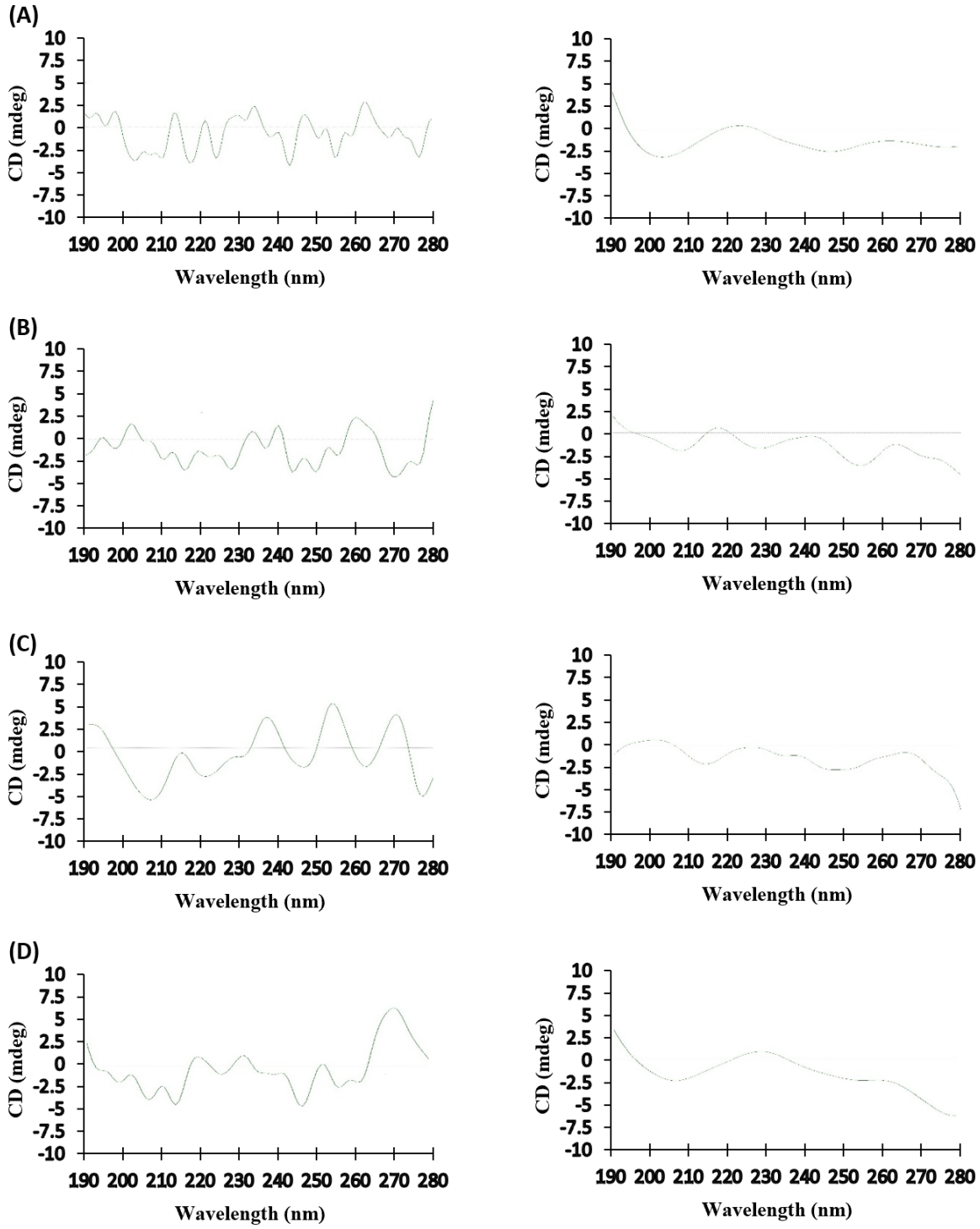
Future studies will first be focused on experimentally validating whether BDM-I is reduced by nitroreductase as possibly indicated from computational analysis. If proven the case then human cytotoxicity assays on BDM-I with human nitroreductases could be performed to determine toxicity levels of intermediate products resulting from reduction of BDM-I by a nitroreductase. As well further computational comparison analysis could be performed on AcpD with human nitroreductase, say for example found in the human microbiota, to determine if BDM-I may prove toxic to helpful gut bacteria.

In regard to BT_411, BT_3781 and PSL-2 glycan array analysis is warranted as I did not have access to a functioning array at the time this study was performed. Glycan arrays can help to define a broader range of carbohydrate affinity through the testing of hundreds of glycans immobilised on array slides. In addition, follow-up SPR analysis of any newly identified glycans by glycan array to analyse kinetics and affinity. As well crystallography and MS of BT_411 and BT_3781 with bound glycans identified in this study should be performed to further validate active site and binding residues predicted by docking analysis, as well as confirming specific type of CBM function (which is currently putative). Moreover, human cytotoxicity analysis on all lectins is worthwhile to test viability as potential drug candidates for the treatment of related human medical conditions.

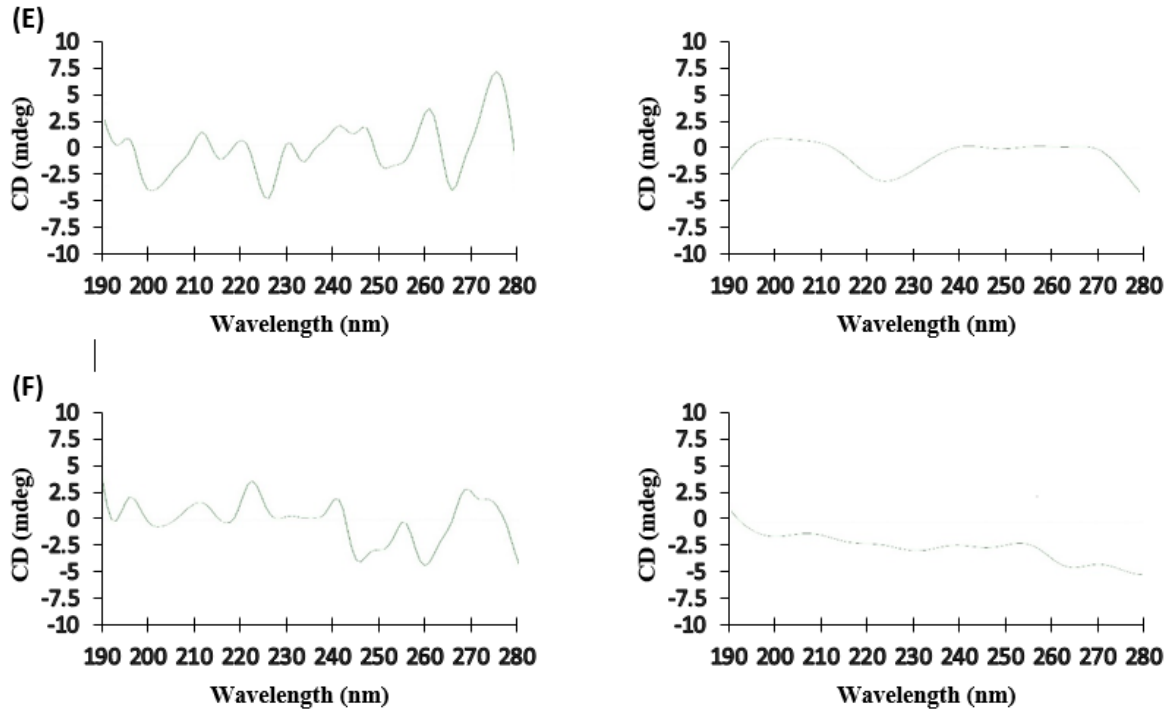
Based on the results of this study, refinement of SPOT-Ligand and SPOT-Struc should be explored by the authors/programmers of these two applications so as to improve prediction accuracy by SPOT-Struc of specific CBP carbohydrate binding and drug target sequence-based matching by SPOT-Ligand.

APPENDIX

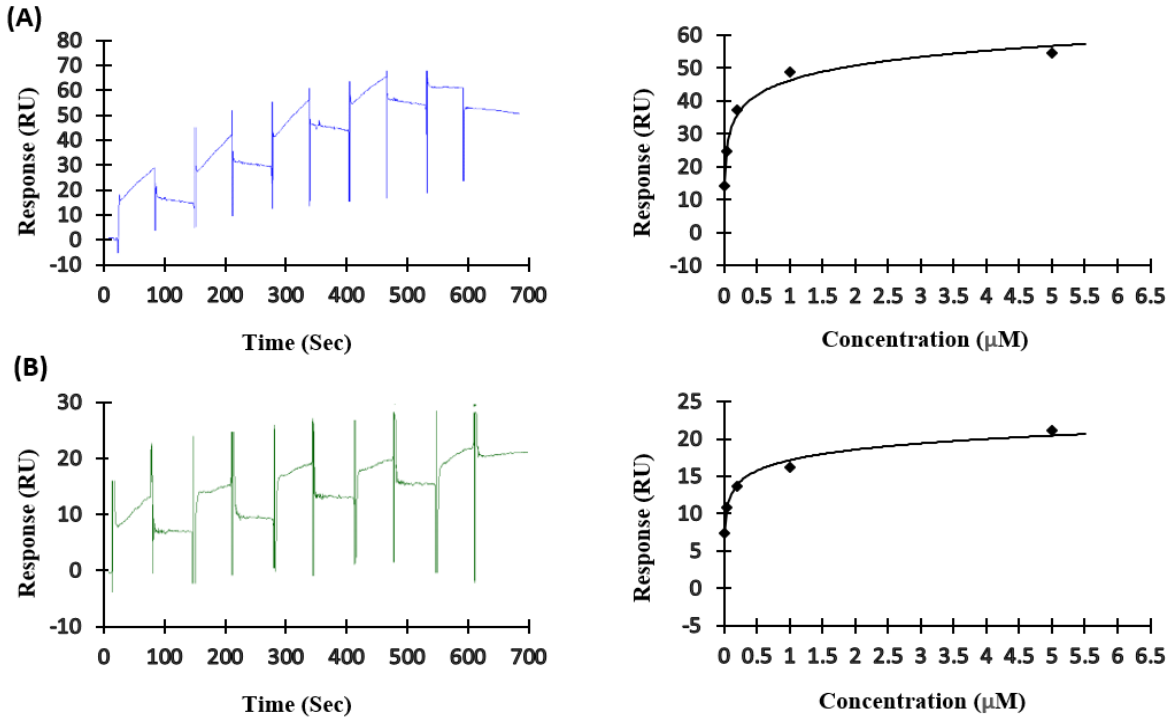
A1. Additional figures



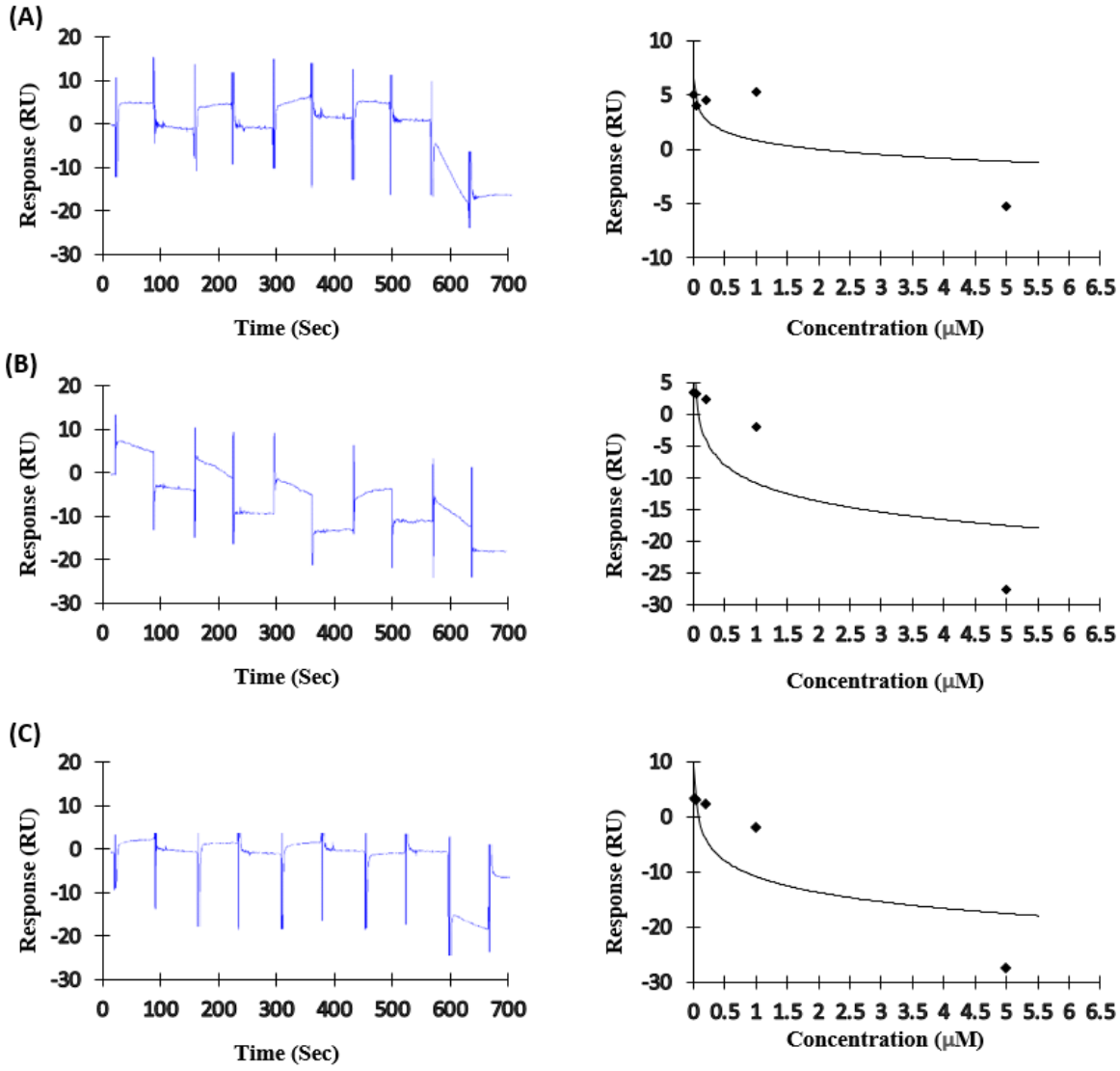
A1.1. CD spectra of recombinant BDM-I targets. Analysis was done using JASCO spectra manager software with default smoothing applied. Non-denatured protein spectra (left) and denatured protein spectra (right). **(A)** EF0414; **(B)** UbiE; **(C)** FtsZ5; **(D)** Lebu_1328



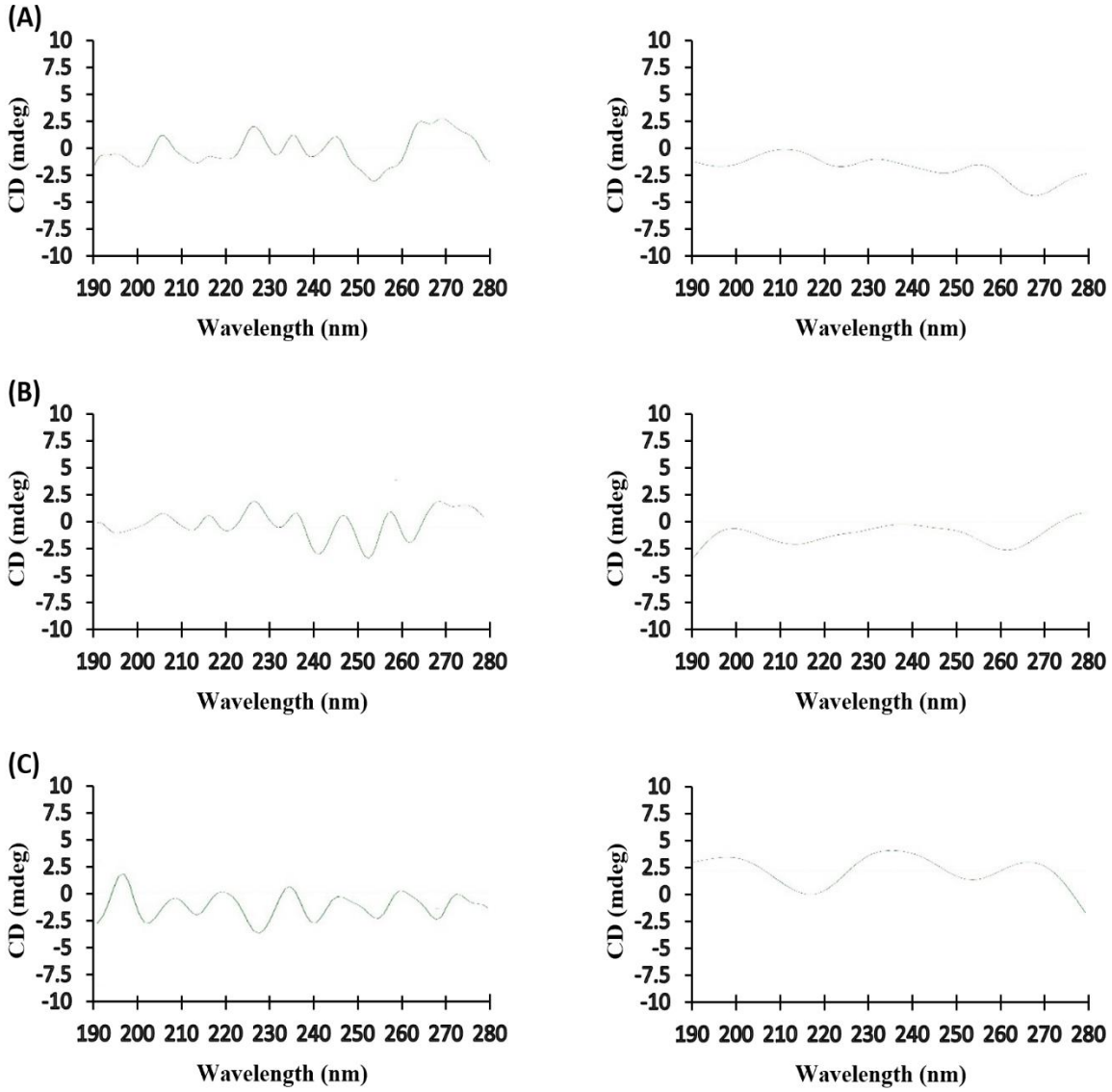
A1.1. continued. CD spectra of recombinant BDM-I targets. Non-denatured protein (left) and denatured protein (right). **(E)** AcpD; **(F)** UbiH



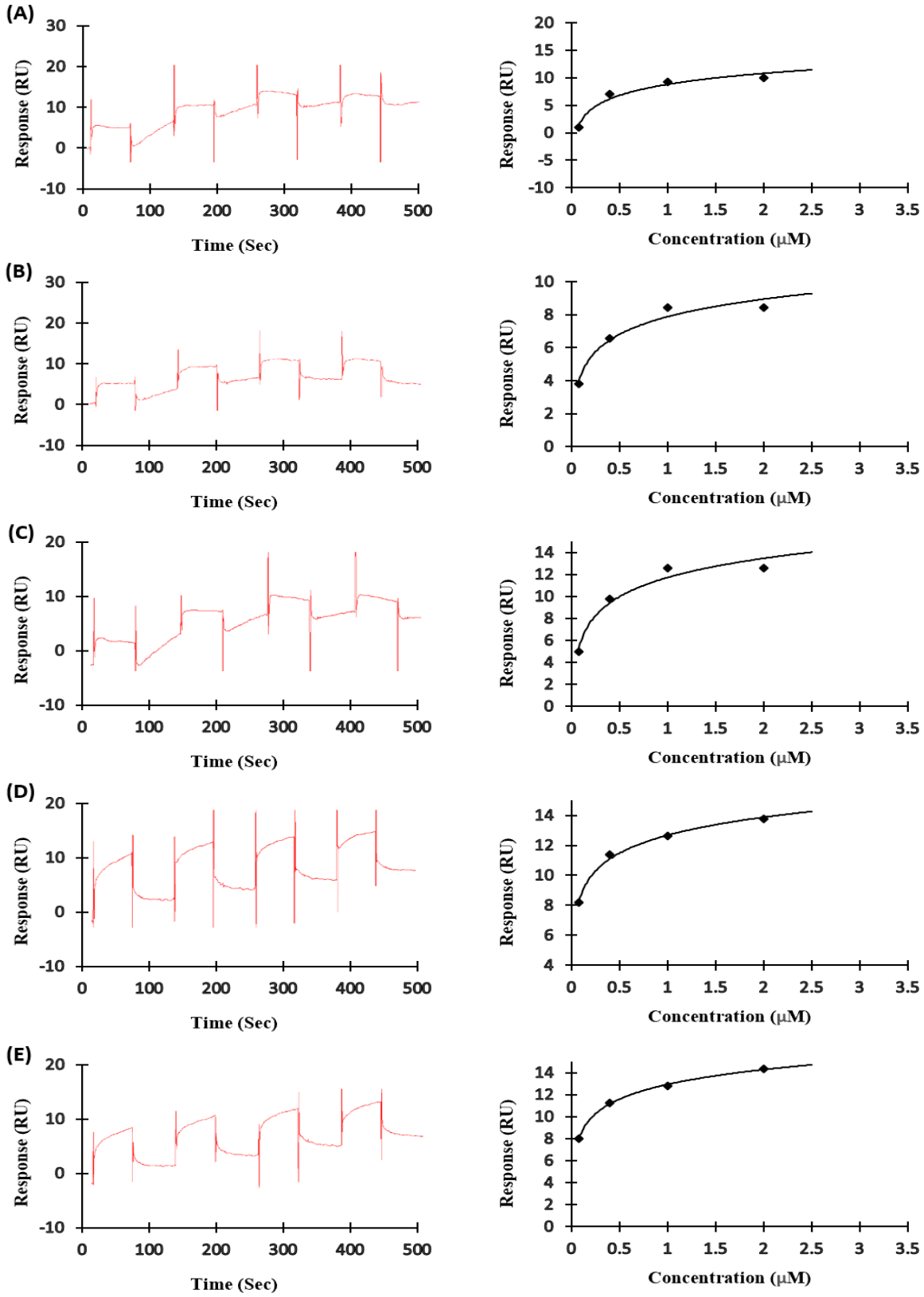
A1.2. SPR sensorgrams (left) and affinity plots (right) showing initial weak binding of BDM-I targets **(A)** EF0414; **(B)** UbiH



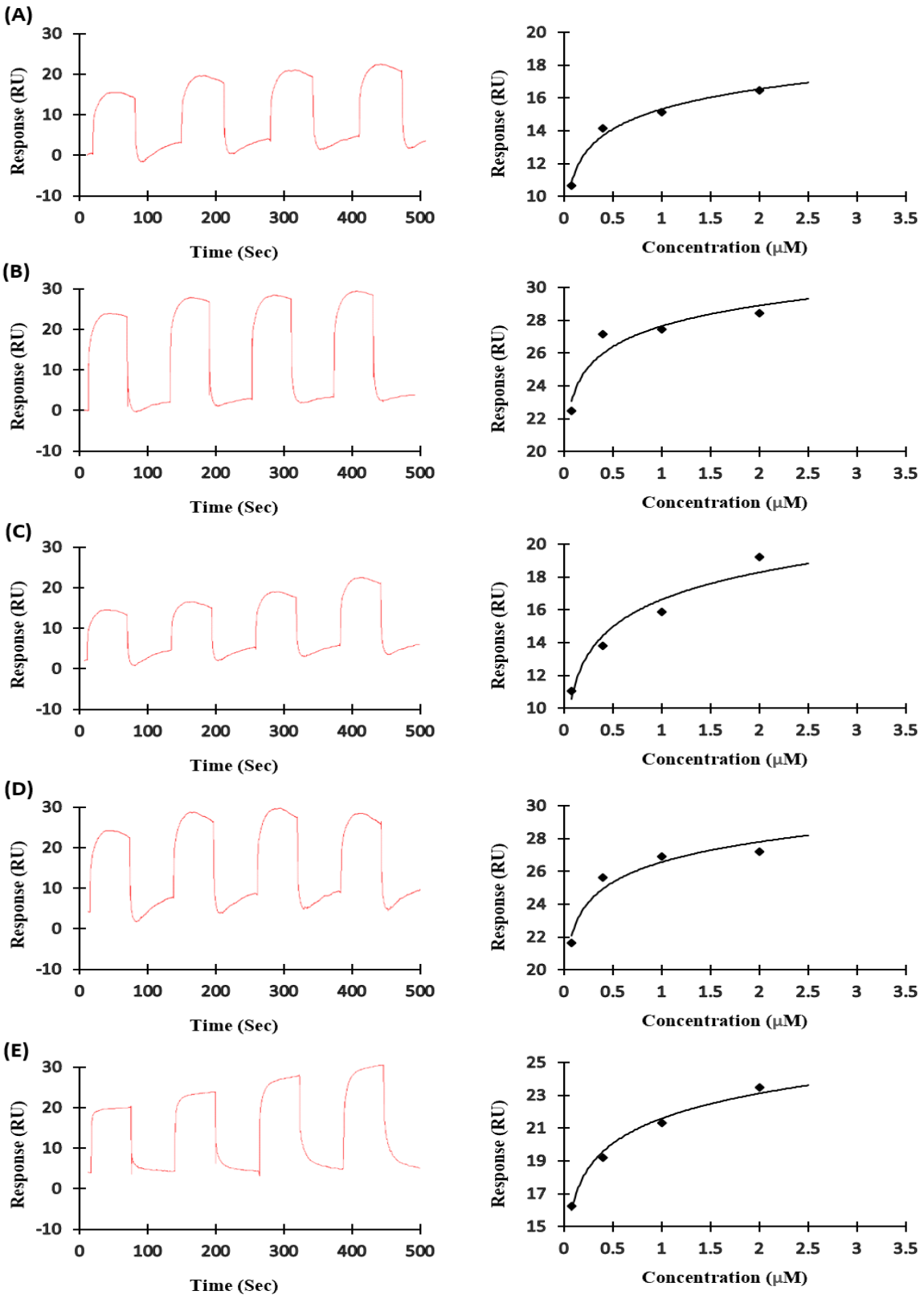
A1.3. SPR sensorgrams (left) and affinity plots (right) showing non-binding of BDM-I targets **(A)** UbiE; **(B)** FtsZ5**; (C)** Lebu_1328



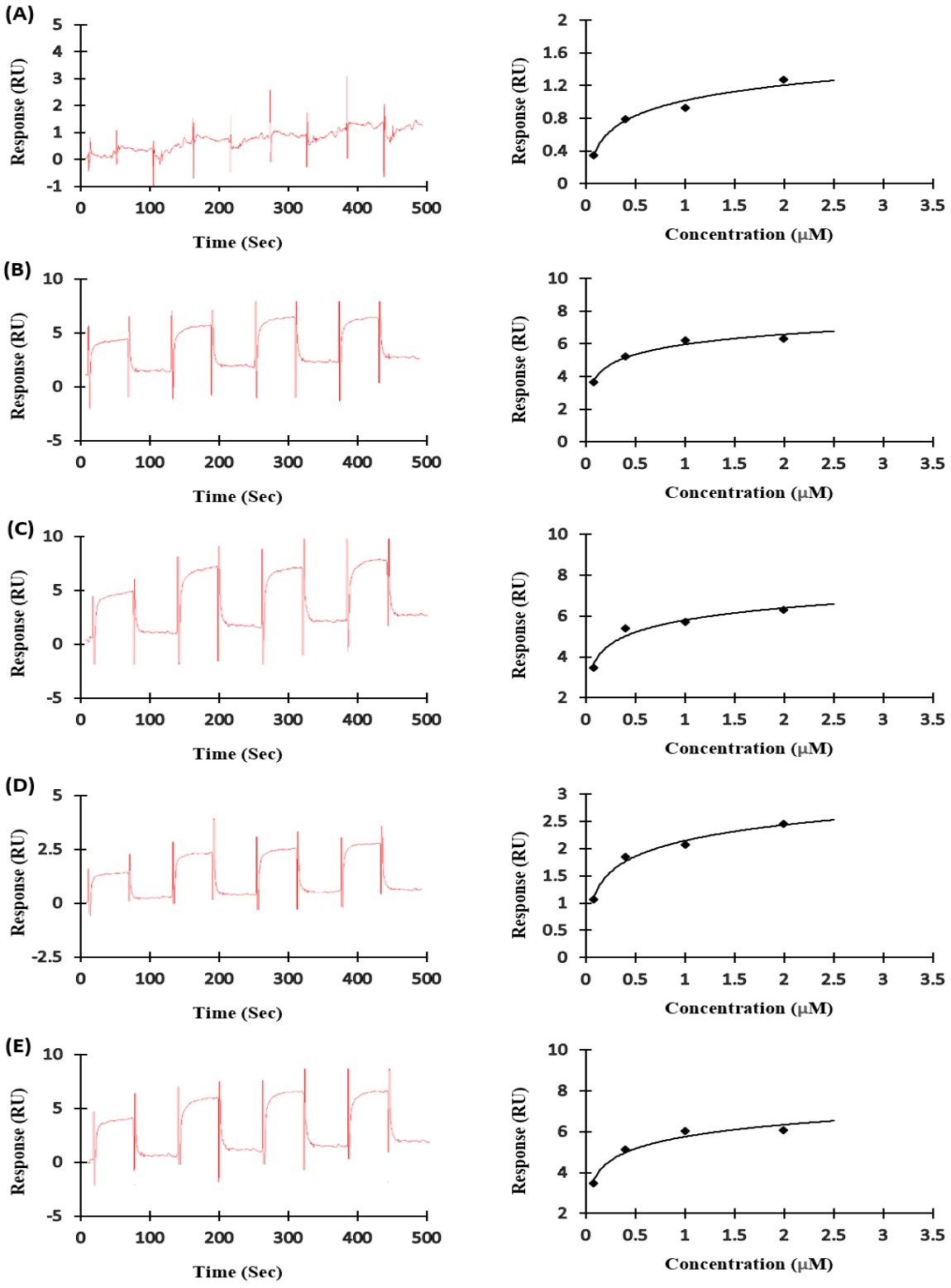
A1.4. CD spectra of recombinant BT_411, BT_3781 and PSL-2. Analysis was done using JASCO spectra manager software with default smoothing applied. Non-denatured protein spectra (left) and denatured protein spectra (right). **(A)** BT_411; **(B)** BT_3781; **(C)** PSL-2



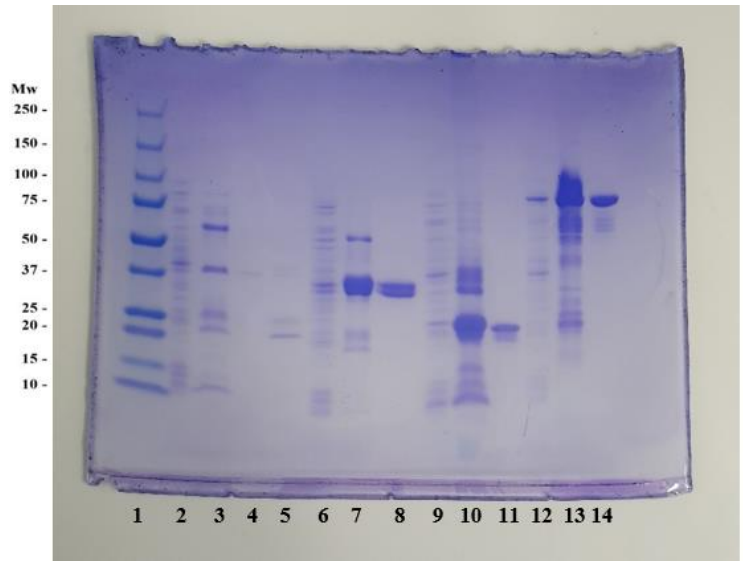
A1.5. BT_411 SPR sensorgrams (left) and affinity plots (right) of top five bound glycans (showing 1 of the 3 runs performed on each glycan). **(A)** Sialyl-Lewis^{x} tetraose; **(B)** Sialylated TF antigen**;** **(C)** Sialyl-Lewis^{a} tetraose; **(D)** N,N'-Diacetyl chitobiose; **(E)** N,N',N",N"'-Tetraacetyl chitotetraose



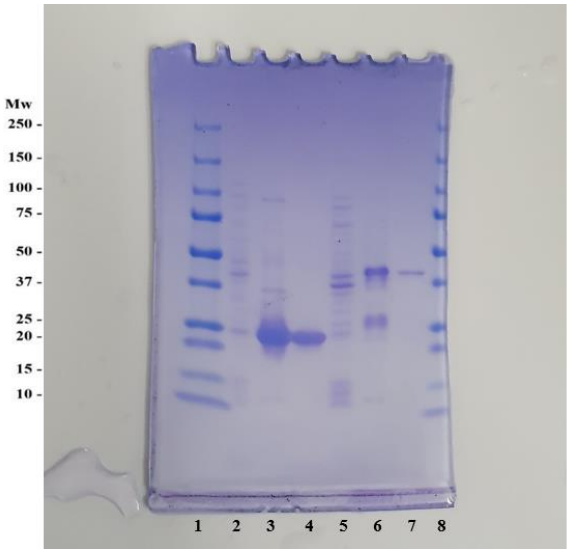
A1.6. BT_3781 SPR sensorgrams (left) and affinity plots (right) of top five bound glycans (showing 1 of the 3 runs performed on each glycan). **(A)** blood group H; **(B)** blood group B; **(C)** blood group A; **(D)** Lewis^{x}; **(E)** Lewis^{a}



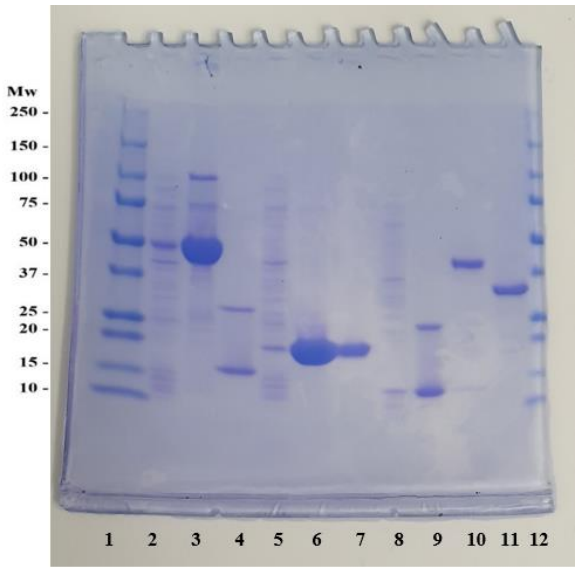
A1.7. PSL-2 SPR sensorgrams (left) and affinity plots (right) of top five bound glycans (showing 1 of the 3 runs performed on each glycan). **(A)** blood group B; **(B)** β-1-4-galactosyl-galactose; **(C)** Lactodifucotetraose; **(D)** blood group H; **(E)** asialo GM1



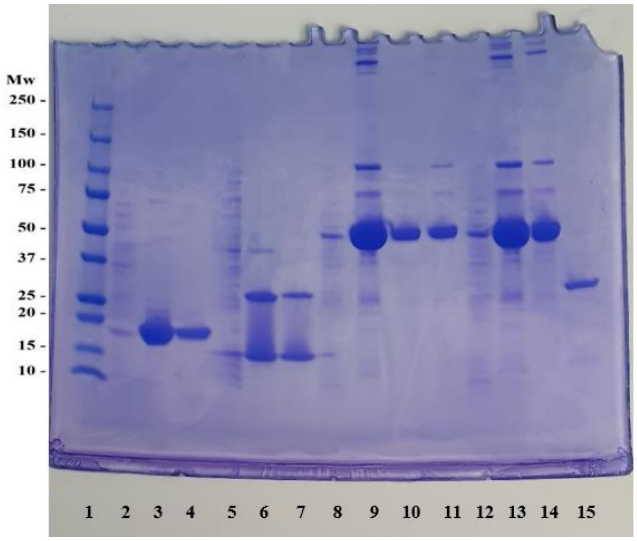
A1.8. Unedited Coomassie stain gel of EF0414, UbiE, FtsZ5 and Lebu_1328. Lane 1, Mw standard; lane 2, soluble EF0414; lane 3, Ni+NTA purified EF0414; lane 4, SEC purified EF0414; lane 5, checking another SEC peak for purified EF0414; lane 6, soluble UbiE; lane 7, Ni+NTA purified UbiE; lane 8, SEC purified UbiE; lane 9, soluble FtsZ5; lane 10, Ni+NTA purified FtsZ5; lane 11, SEC purified FtsZ5; lane 12, soluble Lebu_1328; lane 13, Ni+NTA purified Lebu_1328; lane 14, SEC purified Lebu_1328



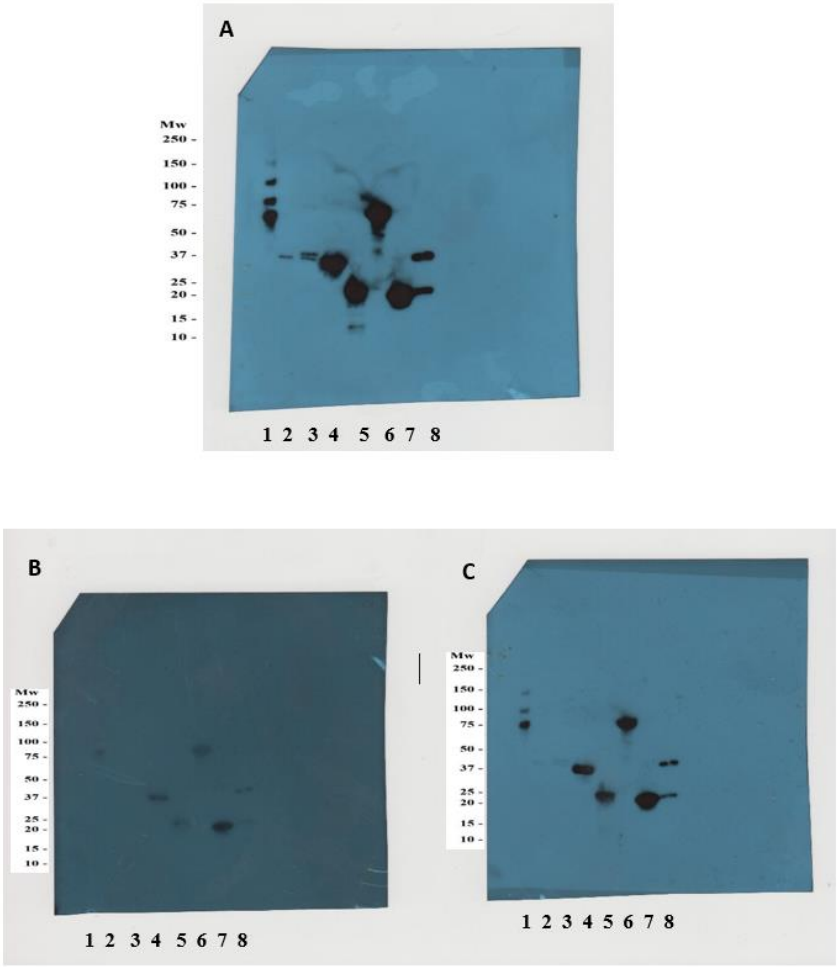
A1.9. Unedited Coomassie stain gel of AcpD and UbiH. Lane 1, Mw standard; lane 2, soluble AcpD; lane 3, Ni+NTA purified AcpD; lane 4, SEC purified AcpD; lane 5, soluble UbiH; lane 6, Ni+NTA purified UbiH; lane 7, SEC purified UbiH; lane 8, Mw standard



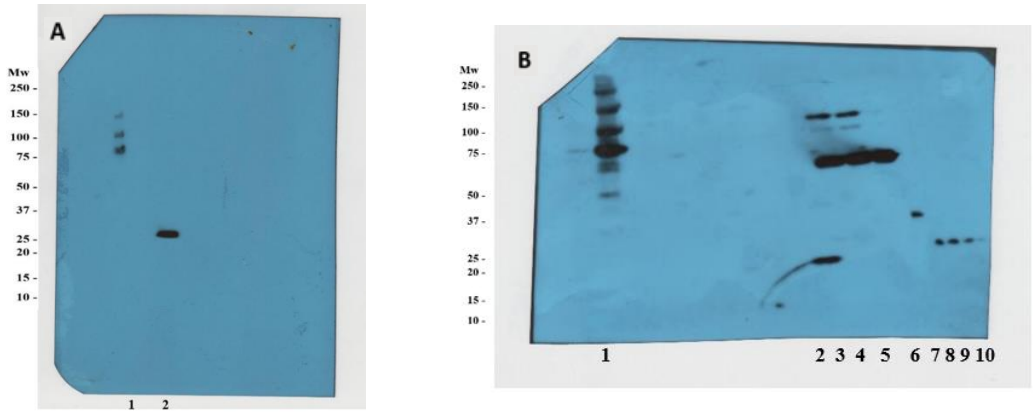
A1.10. Unedited Coomassie rerun of SEC purified UbiE to eliminate doublet seen in previous gel (Figure A1.8.) and various lectin samples to try and get less pronounced bands (did not use in results). Lane 1, Mw standard; lanes 2 – 10, rerun of various lectin samples; lane 11, SEC purified UbiE; lane 12, Mw ladder



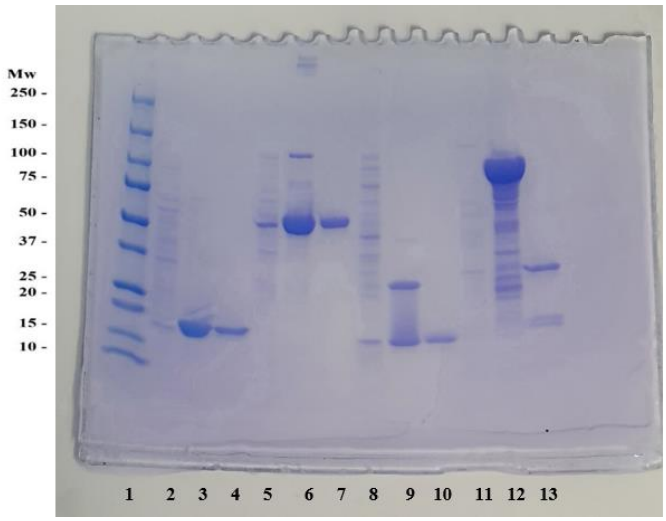
A1.11. Unedited Coomassie rerun of SEC purified Lebu_1328 with increased SDS to fix non-denatured band in previous gel (Figure A1.8.) and various lectin samples to try and get less pronounced bands (did not use in results). Lane 1, Mw standard; lanes 2 – 14, rerun of various lectin samples; lane 15, SEC purified Lebu_1328



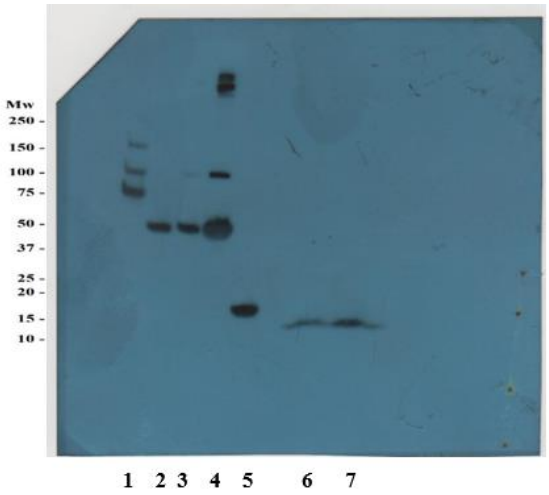
A1.12. Unedited western blots of drug targets – 3 layers of developed film from a single gel/transfer membrane containing drug targets. Bands used in thesis: **(A)** lane 2, EF0414; **(B)** lane 7, AcpD; **(C)** lane 4, UbiE; lane 5, FtsZ5; lane 8, UbiH



A1.13. Unedited rerun western blots of drug targets showing **(A)** lane 1, Mw standard; lane 2, SEC purified Lebu_1328 under increased denaturing conditions and **(B)** lane 1, Mw standard; lane 6 SEC purified UbiH with artefact resolved; lane 2-5 & 7-10, samples not relevant to this study



A1.14. Unedited Coomassie stained gel of BT_411, BT_3781 and PSL-2 and rerun of drug target Lebu_1328. Lane 1, Mw standard; lane 2, soluble BT_411; lane 3, Ni+NTA purified BT_411; lane 4, SEC purified BT_411; lane 5, soluble BT_3781; lane 6, Ni+NTA purified BT_3781; lane 7, SEC purified BT_3781; lane 8, soluble PSL-2; lane 9, lactose purified PSL-2; lane 10, SEC purified PSL-2; lane 11-13, Lebu_1328 samples (not used in final results)



A1.15. Unedited western blot of BT_411 and BT_3781. Lane 2, SEC purified BT_3781; lane 5, SEC purified BT_411; lane 1,3,4,6,7, samples not relevant to this study

A2. Reagent preparations

Buffer solutions that have not been defined in the methods section of chapters 2 or 3.

A2.1. Running buffer for SDS-PAGE

Tris-base	3.0 g
Glycine	14.50 g
20% SDS	5 mL

A2.2. Sample loading buffer for SDS-PAGE

Milli-Q water	4.6 mL
Tris-HCl (0.5 M)	1.00 mL
SDS (20% w/v)	0.80 mL
Bromophenol blue (1% w/v)	0.40 mL
Glycerol	0.80 mL
2-mercaptoethanol	0.40 mL

All ingredients except 2-mercaptoethanol were mixed together and then divided into100 µL aliquots to then be stored at -20 ºC until needed. 5.26 µL of 2-mercaptoethanol was added to 100 uL aliquot upon use of sample loading buffer to make a 4X reduction buffer, diluted down to 1X using protein and Milli-Q water.

A2.3. Coomassie staining solution

Coomassie Brilliant Blue R-250	0.25 g
Methanol	400.00 mL
Glacial acetic acid	100.00 mL

Used Milli-Q to make up to 1 litre of solution. De-stained using Milli-Q water and heat (microwave gel/solution for 20 seconds then let cool for 20 minutes and repeat).

A2.4. Western blot 10X transfer buffer

Tris base	15.2 g
Glycine	72.1 g
SDS (20% w/v)	5.0 g

Brought up to 500 mL with Mill-Q water

A2.5. Western Blot 1X Transfer buffer

10X Transfer buffer	50 mL
Methanol	100 mL
Milli-Q water	350 mL

A2.6. Western Blot TBS-T (Tris Buffered Saline/ Tween 20, pH 8.0)

Tris	6.10 g
NaCl	8.68 g
Tween 20	5.00 mL

Brought up to 1 L with Milli-Q water, adjusted pH with HCL.

A2.7. Western Blot blocking buffer

1 % skim milk powder	1 g
TBS-T	50 mL

Note. Buffers A2.4. to A2.7. also apply to dot blot.