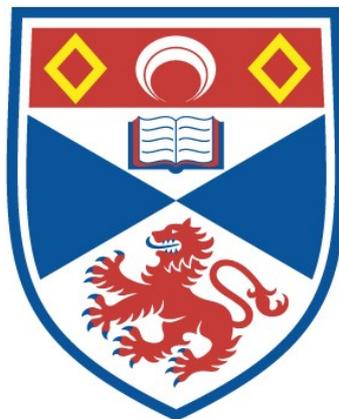


THE DEVELOPMENT OF SIALIDASE INHIBITORS  
USING STRUCTURE-BASED DRUG DESIGN

Graeme W. Rogers

A Thesis Submitted for the Degree of PhD  
at the  
University of St Andrews



2017

Full metadata for this item is available in  
St Andrews Research Repository  
at:

<http://research-repository.st-andrews.ac.uk/>

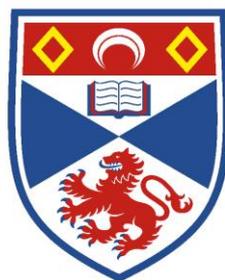
Please use this identifier to cite or link to this item:

<http://hdl.handle.net/10023/15516>

This item is protected by original copyright

# The Development of Sialidase Inhibitors using Structure-based Drug Design

Graeme W. Rogers



University of  
St Andrews

This thesis is submitted in fulfilment for the degree of PhD  
at the  
University of St Andrews

26<sup>th</sup> of March 2017

**1. Candidate's declarations:**

I, Graeme Rogers, hereby certify that this thesis, which is approximately 60,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September 2012 and as a candidate for the degree of Doctor of Philosophy (Ph.D.) in September 2012; the higher study for which this is a record was carried out in the University of St Andrews between 2012 and 2017.

Date signature of candidate

**2. Supervisor's declaration:**

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy (Ph.D.) in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date signature of supervisor

**3. Permission for publication: (to be signed by both candidate and supervisor)**

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

**No embargo on print copy**

ELECTRONIC COPY

**No embargo on electronic copy**

Date signature of candidate signature of supervisor

## *Acknowledgements*

There are many people I would like to thank that have made this thesis possible.

Firstly, I would like to thank my principle supervisor Professor Garry Taylor for his help, support and guidance throughout my PhD. I would like to thank my secondary supervisor, Professor Nick Westwood for providing me with the opportunity of starting a PhD in St. Andrews. I would also like to thank Dr Gordan Florence for providing space, equipment and materials to conduct some synthesis in his lab.

A big thank you goes to the Taylor and Westwood research groups (past and current members) who have provided some great times and long lasting friendships. Particularly, I would like to mention Dr Jane Potter, Dr Helen Connaris and Dr Fanny Tran who have given me great help and support throughout my PhD.

I would also like to thank all the research staff that contribute to the services provided in the School of Chemistry and the BSRC that many of us use; this includes Mass Spectrometry, X-ray crystallography, NMR, autoclaving and media services as well as the secretarial team.

Lastly a huge thank you goes to my wife, Katie for all her support and encouragement throughout my PhD. I would also like to acknowledge my son, Jamie who was born when I wrote this thesis. My wife and son are my motivation and I love them more than words can describe. My thanks also go out to my close family who have provided encouragement and advice over the years. Without the love and support of my wife, son and close family I would not have made it this far and I am forever grateful. This thesis is dedicated to them.

## ***Abstract***

The sialidases/neuraminidases represent a family of enzymes whose function is important in the pathogenicity of bacteria and the virulence of influenza. Relenza and Tamiflu represent two drugs that were developed using structure-based drug design (SBDD) and computational-assisted drug design (CADD). These drugs target the active site of the influenza neuraminidase A and B (GH-34 family). Sialidases in the GH-33 family could represent novel drug targets for the treatment of bacterial or parasitic infection. SBDD was employed to develop chemical tools of two GH-33 sialidases, NanB and TcTS.

NanB is a potential drug target for *S. pneumoniae*. The chemical tool developed for NanB follows on from work within the Taylor and Westwood research groups, in which a molecule of CHES and a glycerol were found serendipitously bound within a water channel at an allosteric site. Using this information as a basis for SBDD an allosteric inhibitor of NanB, Optactin was developed. Within this work, synthesis of this inhibitor was achieved and optimised. Optactin was then modified to improve potency. This proceeded through an amide analogue and addition of an arene resulting in a mid-micromolar inhibitor (IC<sub>50</sub>: 55.4±2.5 µM). Addition of polar substituents improved potency further resulting in a low micromolar inhibitor of NanB, Optactamide (IC<sub>50</sub>: 3.0±1.7 µM). Application of this tool *in vitro* demonstrated that NanB and NanA have a role in invasion of *S. pneumoniae* into lung epithelial cells.

TcTS is a potential drug target for the treatment of Chagas disease. A CADD approach using a fragment library was unsuccessful at identifying an allosteric inhibitor of TcTS despite structural similarity with NanB. A re-task of the CADD approach towards the active site was successful in identifying an inhibitor of TcTS and a fragment useful for further development. This work sets the groundwork for the development of a chemical tool targeting TcTS.

## ***Abbreviations***

2,3-difluoro-N-acetylneuraminic acid (2,3F-Neu5Ac)  
2,7-anhydro- $\alpha$ -N-acetylneuraminic acid (2,7-anhydro-Neu5Ac)  
2-deoxy-2,3,dehydro-N-trifluoro-acetylneuraminic acid (FANA)  
2-deoxy-2,3-didehydro-5-N-neuraminic acid (Neu5Ac2en/DANA)  
3-deoxy-D-glycero-D-galacto-2-nonulopyranos-1-onic acid (KDN)  
3-phosphoinositide-dependent kinase-1 (PDPK1)  
ABC transporter  
acquired immune deficiency syndrome (AIDS)  
adenosine monophosphate (AMP)  
ATP-binding cassette  
autolysin LytA  
BET bromodomain probe  
bromodomain and extra- terminal domain (BET)  
calcium-sensing receptor (CaSR)  
C-C chemokine receptor 5 (CCR5)  
Central nervous system (CNS)  
central nervous system (CNS)  
centre for drug evaluation and research (CDER)  
centre for drug evaluation and research (CDER)  
cerebrospinal fluid (CSF)  
Chronic myelogenous leukemia(CML)  
constrained approach (CA)  
constrained approach (CA)  
De-oxynucleic acid (DNA)

dihydrofolic acid (DHFA)

Drug Discovery Unit (DDU)

Electrocardiogram (ECG)

enzyme-linked immunosorbant assay (ELISA)

factor VIIa (FVIIa)

galactose (Gal)

Global disease burden (GDB)

glutathione-S-transferase (GST)

glutathionylspermidine (Gsp)

glutathionylspermidine synthetase (GSPS)

glycogen phosphorylase (GlyP)

glycoside hydrolase (GH)

glycosylphosphatidylinositol (GPI)

G-protein coupled receptors (GPCR, receptor family)

hepatitis C virus (HCV)

high throughput screens (HTS)

human immunodeficiency virus (HIV)

human neuraminidase enzymes (hNEU)

Innovative Medicines Initiative (IMI)

KNF model

*Leishmania infantum (L.infantum)*

LPXTG-anchored protein

methicillin-resistant *S. aureus* (MRSA)

mitogen-activated protein kinase (MEK)

Molecular dynamics (MD)

molecular mechanism of action (MMOA)

Monad-Wyman Changeux (MWC)

*N*-acetylgalactosamine (GalNAc)  
*N*-acetylneuraminic acid (Neu5Ac)  
Negative allosteric modulators (NAMs)  
Neglected disease (ND)  
neglected tropical diseases (NTDs)  
neuraminidase1 (Neu1)  
neuraminic acid (Neu)  
neuraminidase 2 (Neu2)  
neuraminidase 3 (Neu3)  
neuraminidase 4 (Neu4)  
neuraminidase inhibitors (NAIs)  
neuraminidases (NA)  
new molecular entities (NMEs)  
*N*-glycolylneuraminic acid (Neu5Gc)  
nuclear magnetic resonance (NMR)  
Office of Orphan Products Development (OOPD)  
Office of Orphan Products Development (OOPD)  
*P. falciparum* translocation elongation factor 2 (PfeEF2)  
penicillin-binding proteins (PBP)  
Pharmaceutical Research and Manufacturers of America (PhRMA)  
phosphoinositide 3-kinase (PI3K)  
platelet-activating factor (PAF)  
pneumococcal surface protein A (PspA)  
pneumococcal surface protein C (PspC)  
positive allosteric modulators (PAMs)  
Protein of interest (POI)  
Protein-protein interactions (PPIs)

Protein-protein interactions (PPIs)

Relaxed state (R)

relaxed approach (RA)

relaxed approach (RA)

research and development (R&D)

Sialic acids (Sia)

small and medium enterprises (SMEs)

staphylococcal cassette chromosome mec (SCCmec)

*Streptococcus pneumoniae* (*S. pneumoniae*)

*T. cruzi* Trans-sialidase (TCTS)

Tense state (T)

tetrahydrofolic acid (THFA)

transverse relaxation optimized spectroscopy (TROSY)

*Trypanosoma brucei* (*T. brucei*)

*Trypanosoma Cruzi* (*T. cruzi*)

Trypanothione (T(SH)<sub>2</sub>)

trypanothione synthetase (TRYS)

U.S Food and Drug Administration (FDA)

UDP-GlcNAc-2-epimerase/ManNAc kinase (GNE)

United States (U.S.)

Wild-type (WT)

World Health Organisation (WHO)

p-aminobenzoic acid (PABA)

γ-aminobutyric acid (GABA)

γ-aminobutyric acid receptor α (GABA<sub>A</sub>)

## **Amino Acids**

Ala, A Alanine

Arg, R Arginine

Asn, N Asparagine

Asp, D Aspartic acid

Cys, C Cysteine

Gln, Q Glutamine

Gly, G Glycine

His, H Histidine

Ile, I Isoleucine

Leu, L Leucine

Lys, K Lysine

Met, M Methionine

Phe, F Phenylalanine

Pro, P Proline

Ser, S Serine

Thr, Threonine

Trp, W Tryptophan

Tyr, Y Tyrosine

Val, V Valine

## ***Table of Contents***

Acknowledgements .....	1
Abstract .....	2
Abbreviations .....	3
Table of Contents .....	8
Table of Figures .....	14
1.0 Chemical Biology .....	19
1.10 Chemical genetics.....	19
1.11 Forward chemical genetics.....	20
1.12 Reverse chemical genetics.....	22
1.20 Biological Target Types and Regulatory Approvals .....	28
1.21 Target Identification .....	30
1.22 Target Validation .....	32
1.23 Chemical Tools/Probes .....	33
1.30 Protein modulation .....	34
1.31 Allosteric modulation .....	35
1.32 Discovery of Allosteric Inhibitors.....	40
1.33 Phage Display.....	40

1.34 Tethering approach .....	41
1.35 Fragment screening and serendipitous binders .....	42
1.36 The correct approach to allosteric discovery .....	44
1.40 Neglected disease (ND) and unmet need.....	45
1.41 Chagas disease.....	49
1.42 Antibiotics.....	52
1.43 Streptococcus pneumoniae.....	54
1.50 The sialidases and sialic acid as potential drug targets.....	56
1.51 Sia .....	56
1.52 The core structure of Sia .....	57
1.53 Sia linkage and presentation within the “Sialome” .....	58
1.54 Sialidases (EC 3.2.1.18) .....	60
1.55 Glycoside Hydrolase family (GH-33).....	61
1.56 Bacterial Sialidases .....	61
1.57 Trans-sialidase .....	65
1.58 Glycoside Hydrolase family (GH-34).....	67
1.59 Conservation.....	67
1.59.1 Active site .....	67

1.59.2 Active site mechanism.....	68
1.59.3 Water channel.....	72
1.59.4 “Asp boxes” and F/Y-R-I-P motif.....	74
1.60 Current Inhibitors of Sialidases.....	74
1.61 Current Inhibitors of NanB.....	78
1.62 Current Inhibitors of TcTS.....	80
1.70 Rational Design.....	83
1.71 Medicinal chemistry.....	84
1.72 Structure-based Design.....	86
1.73 Computational design.....	88
1.80 Thesis Aims.....	96
1.81 Development of a NanB “relaxed” chemical tool.....	96
1.82 Development of a TcTS “constrained” chemical tool.....	97
2.0 NanB and the development of an allosteric chemical tool.....	98
2.1 <i>The story so far</i> .....	98
2.2 <i>Resynthesis of Optactin</i> .....	103
2.3 <i>Production of NanB protein</i> .....	106
2.4 <i>Testing of Optactin</i> .....	110

2.5 Identification of a single point mutation .....	112
2.6 Biochemical comparison .....	114
2.7 Crystallisation of mutant D643G NanB and WT NanB .....	114
2.8 Retesting of Optactin .....	117
2.9 Summary .....	118
3.0 Development of alternative series .....	119
3.1 Previous Mutant studies .....	119
3.2 Repeated mutant studies on WT .....	122
3.3 Design and development of simplified analogues .....	122
3.3.1 Synthesis of analogues .....	123
3.3.2. Optimisation .....	126
3.3.3. Further optimisation .....	131
3.4 Optactamide binding position .....	133
3.5 Activity assays and synergistic inhibition .....	134
3.6 Control experiments .....	136
3.7 Cell and bacterial assay .....	137
3.8 Summary .....	139
4.1 Discussion and Future work .....	140

4.2. Experimental .....	147
4.2.1 Chemistry.....	147
4.2.2 Biology .....	168
5.0 The discovery of constrained tools for TcTS.....	175
5.1 The TcTS story so far.....	175
5.2 The water channel within TcTS.....	181
5.3 Fragment library generation .....	185
5.4 In Silico docking and “HIT” identification .....	186
5.5 TcTS construct, expression and purification .....	189
5.6 TcTS crystal generation and structures .....	189
5.7 Fragment activity and crystal soaking .....	191
5.8 Summary.....	192
6.0 Development of a TcTS constrained chemical tool that targets the active site.....	194
6.1 TcTS active site .....	194
6.1.1 Known binders of the TcTS active site.....	197
6.2 Library generation and docking.....	198
6.2.1 FlexX docking and validation .....	200
6.2.2 Faster docking using the KNIME workflow.....	201

6.2.3 Docking scores.....	202
6.2.4 Active site binding assessment.....	204
6.3 Biological testing of compound 81 and 82.....	204
6.4 SAR with analogues of 84.....	206
6.5 Summary.....	209
7.0 Discussion and Future work.....	210
7.1 Experimental.....	212
7.1.1 Computational modelling:.....	212
7.1.2 Protein expression:.....	213
7.1.3. Protein purification:.....	213
7.1.4. Kinetic analysis:.....	214
7.1.5. Protein crystallisation and X-ray crystallography:.....	214
8.0 Bibliography.....	217

## *Table of Figures*

Figure 1. A. Schematic of the forward and reverse genetic approach. ....	20
Figure 2. Schematic of the forward chemical-genetic examples discussed in this subchapter .....	21
Figure 3. A. Schematic of a reverse chemical genetic example in this subchapter showing the development and use of PD184352.....	24
Figure 4. Schematic of the reverse chemical genetic approach used to identify TRYS as an essential enzyme for <i>L. infantum</i> .....	25
Figure 5. Graph illustrating the percentage of NMEs by approach used for discovery for first-in-class drugs and follower drugs .....	27
Figure 6. A. Figure of the estimated number of exploitable drug targets from the genome i.e the ‘druggable’ genome .....	29
Figure 7. A schematic representing the different requirements of drugs and chemical tools/probes. The general approach applies to probes only whereas the constrained approach applies to both at the interface between probes and drugs.. .....	34
Figure 8. A simplistic diagram of the types of allosteric modulation on an enzyme. ....	35
Figure 9. A schematic diagram of the MWC model showing the transition between the T and R states upon ligand binding.....	36
Figure 10. Model of the GABAA receptor, a transmembrane ligand-gated ion channel found within the central nervous system (CNS).. .....	38
Figure 11. Graph showing the frequency of amino acids in allosteric sites compared to orthosteric sites .....	39
Figure 12. Diagram of the tethering screening approach adapted from Hardy and Wells, 2004 .....	41
Figure 13. Images of cyclohexyl-hexyl- $\beta$ -D-maltoside bound in the allosteric site of S70C SHV $\beta$ -lactamase (PDB: 4FD8).. .....	43
Figure 14. Graphs representing global spend on R&D for PhRMA (consisting of 39 members of the largest pharmaceutical and biotechnology companies) as a comparison against global neglected disease spend.. .....	47
Figure 15. Schematic of T.Cruzi life cycle within vector and mammalian host .....	50
Figure 16. A. Schematic of MMOA of various antimicrobial agents.....	55

Figure 17. A. Chemical structures of the “four core sialic acid” molecules. ....	58
Figure 18. Chemical structures of the various sialic acid linkages found within the “sialome”. ....	60
Figure 19. Structure of NanB. ....	64
Figure 20. Structure of TcTS ....	66
Figure 21. Active site overlay representation of sialidases.....	68
Figure 22. Glycosidic hydrolases reaction mechanism of the retaining hydrolase and the inverting hydrolase reaction.....	69
Figure 23. Reaction mechanism of the Streptococcus pneumoniae sialidases (NanA, NanB and NanC) .....	70
Figure 24. Reaction mechanism TcTS .....	71
Figure 25. NanB 3D structure alignment of SWISSPROT database.....	73
Figure 26. The chemical structure of different sialidase inhibitors .....	76
Figure 27. NanB surface representation with CHES bound within the active site cavity .....	79
Figure 28. A schematic of the Topliss Scheme .....	85
Figure 29. SBDD of Sildenafil.....	89
Figure 30. SBDD of Zanamivir.....	95
Figure 31. Identification of a secondary site within NanB .....	98
Figure 32. A profile of the active site and secondary site within NanB.....	100
Figure 33. Generation of Optactin.....	101
Figure 34. Binding analysis of Optactin .....	102
Figure 35. Grubbs 1st generation and Grubbs 2nd generation catalysts. ....	106
Figure 36. SDS page of NanB purified through a nickel column using different methods .....	107
Figure 37. SDS Page of NanB after anion exchange purification and size-exclusion chromatography.....	109
Figure 38. 4-Mu fluorescence, near-UV CD and DLS analysis of Optactin against NanB.....	111
Figure 39. The 2Fo-Fc (black mesh, contoured at 1.5 $\sigma$ ) and Fo-Fc (green (3.0 $\sigma$ ) and red(-3.0 $\sigma$ )) electron density maps for amino acid residue 643 of published structure 2VW2. ....	112
Figure 40. Schematic of the NanB domains and sequence of NanB WT.....	113
Figure 41. Diagram of the sitting drop method .....	115
Figure 42. Images taken of crystals grown in the initial hit and conditions when crystal seeding was performed .....	116

Figure 43. IC <sub>50</sub> curve of re-crystallised Optactin tested against WT NanB. Optactin tested against D643G and tested against fresh 4-Munana substrate. ....	117
Figure 44. Binding mode of Optactin within the Optactin-NanB <sub>D643G,K499G</sub> crystal structure and the binding mode of Optactin within the Optactin-NanB <sub>D643G</sub> (PDB: 4XJ9) crystal structure .....	120
Figure 45. Optactin-NanB <sub>WT</sub> complex .....	121
Figure 46. SDS page gel of WT and K499G after purification and the activity of NanB <sub>WT</sub> and NanB <sub>K499G</sub> using the 4-Munana assay including the IC <sub>50</sub> values of Optactin against NanB <sub>WT</sub> and NanB <sub>K499G</sub> .....	122
Figure 47. ORTEP plots showing 50% thermal ellipsoid probability of 40 and 41.....	124
Figure 48. The percentage of 4-Munana hydrolysis by NanB in the presence of amine 31, amide 40 and amide 41 including the binding position of amide 40 compared with Optactin within the allosteric site of crystal structures NanB <sub>WT</sub> -amide 40 and NanB <sub>WT</sub> -Optactin.....	125
Figure 49. CADD optimisation of 40 .....	127
Figure 50. Binding position of amide 56 within the allosteric site of NanB.....	128
Figure 51. Optactamide bound within the allosteric site of NanB with Optactin superimposed.....	133
Figure 52. Overlay of Optactamide-NanB crystal structure the identified displaced waters.....	134
Figure 53. Line weaver-Burke plot of Optactamide and IC <sub>50</sub> values of Optactamide when in the presence and absence of active site inhibitor compound 80 .....	135
Figure 54. A. The fluorescence of 4-Mu (100 μM) at various concentrations of Optactamide. B. CD near UV spectra of NanB in the absence (red line) and presence of 500 μM Optactamide (green line). CD near UV spectrum was also measured on a solution containing no protein/NanB (blue line). C. DLS showing size distribution by diameter of (i) NanB control and (ii) NanB in the presence of Optactamide at 2.5 mM.....	136
Figure 55. The number of colonies determined from the invasion assay (with the antibiotic addition step). Adhesion and invasion assay (without the addition of the antibiotics). A549 cell toxicity assay using MTT.....	138
Figure 56. Superimposed structure of NanA against Optactamide-NanB crystal structure (RMSD: 1.8 Å) ....	141
Figure 57. Superimposed structure of NanC against Optactamide-NanB .....	143
Figure 58. Examples of CHES serendipitously bound within crystal structures .....	145

Figure 59 Chemical structures of two similar anthraquinones and the TcTS active site with the proposed binding conformation of compound 10. ....	176
Figure 60. TcTS active site with residue Phe58 .....	179
Figure 61. The chemical structure of the TcTS inhibitors identified by Dr Telford. ....	180
Figure 62. Aligned structures of TcTS <sub>F58N</sub> and solved structures by Buschiazzo et al., 2002 .....	182
Figure 63 Overlaid structures of TcTS and NanB secondary site .....	183
Figure 64. Bar chart representing the hydrophobic and polar/charged residues as a percentage of the total for NanB (allosteric site and active site) and TcTS (mapped site and active site) .....	184
Figure 65. Examples of fragments from the ZINC <sup>12</sup> database. ....	186
Figure 66. Schematic of the ligand similarity results generated from JChem.....	187
Figure 67. Top ten hits and ZINC ID's from the FlexX screen of 1,015 fragments. ....	187
Figure 68. SDS Page of TcTS <sub>F58N</sub> after nickel-column chromatography (1) and size-exclusion chromatography (2).....	189
Figure 69. Images of TcTS crystals grown in conditions, A: 200 mM L-proline, 100 mM Hepes and 10% PEG 3350, B: 100 mM Tris pH 8.5 and 10% PEG 8000. ....	190
Figure 70. Proline found bound within the active site of TcTS.....	191
Figure 71. Graph showing the percentage activity of each fragment against TcTS using the 4-Munana assay .....	192
Figure 72. Binding position of lactose and DANA within the active site of TcTS .....	195
Figure 73. Three pockets identified within the TcTS binding site identified by Dr Telford. ....	196
Figure 74. Graph showing the percentage inhibition of known sialidase inhibitors against TcTS activity using the 4-Munana assay .....	198
Figure 75. A schematic of the ligand similarity selection applied to the large "clean drug like" ZINC <sup>12</sup> library. ....	199
Figure 76. Siastatin B binding position within the TcTS active site and FlexX prediction of siastatin B within TcTS.....	200
Figure 77. Schematic representing the KNIME workflow for FlexX docking within LeadIT. ....	201
Figure 78. Common chemical structure found within the top ten "hits" identified for the open TcTS active site structure by FlexX .....	203

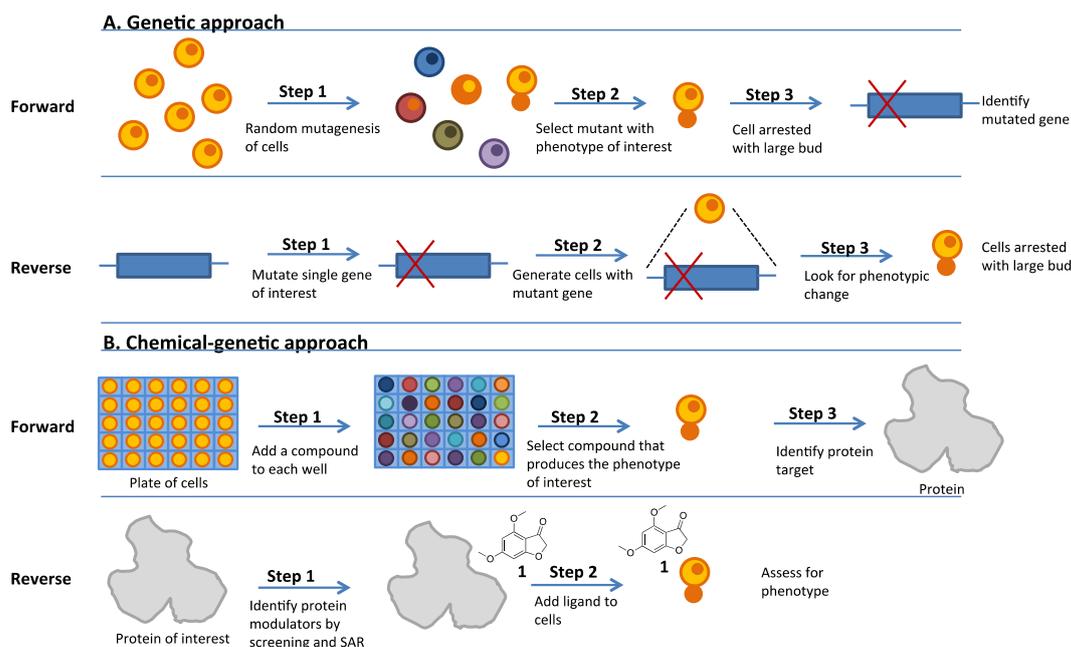
Figure 79. The binding pose generated from the FlexX screen for two examples of “hits” .....	204
Figure 80. Evaluation of “hits” using the 4-Munana assay. ....	205
Figure 81. Binding position of 84 within the active site of TcTS. ....	206
Figure 82. Binding position of <i>cis</i> -4-hydroxy proline within the active site of TcTS. ....	207
Figure 83. Binding position of <i>trans</i> -4-hydroxy proline within the active site of TcTS. ....	208
Figure 84. Overlaid binding positions of 84, <i>trans</i> -4-hydroxy proline, <i>cis</i> -4-hydroxy proline and L-proline within the TcTS active site and the % inhibition of TcTS activity observed for 84, <i>trans</i> -4-hydroxy proline, <i>cis</i> -4-hydroxy proline and L-proline (all tested at 1mM) using the 4-Munana assay. ....	208
Figure 85. Binding of indole-5-carboxylic acid (green) found bound on the surface of TcTS. ....	211

## ***1.0 Chemical Biology***

The term “chemical biology” came into wide use in the 1990s and is a discipline that encompasses a collaborative investigation between the subjects of chemistry and biology (Bucci et al., 2010). It is a varied field covering a wide range of subjects including, but not limited to, medicinal chemistry, proteomics, cell biology, biochemistry and structural biology (Ostler, 2007). The synergistic combination of biology and chemistry allows novel approaches to address problems. In particular, a chemist’s understanding of molecular structure, interaction and reactivity can be useful in the dissection of mechanisms and interactions in large biological molecules (Science, 1884). This understanding enables an alternative perspective on large macromolecular systems and scope for an explorative investigation of function using chemical tools (Stockwell, 2004, Arrowsmith et al., 2015). Biology, and the variety of synthetically challenging molecules identified as natural products, have not only inspired new chemical space (Paterson and Anderson, 2005) but also enabled the development of new synthetic methodology (Mohr et al., 2008). Biological catalysts, such as enzymes, are capable of complex regioselective and stereoselective transformations of molecules (Wong, 1989). Chemists are using biological catalysts and adapting them for use in chemical synthesis (Klibanov, 1990, Whitesides and Wong, 1985). In drug discovery, both disciplines are needed in synergy to understand the biological target, the mechanism of action and what constitutes a good drug candidate with an optimal pharmacokinetic and toxicological profile (Hughes et al., 2011). This thesis will have an emphasis on the combination of biology and chemistry to develop novel small biological modulators, which will enable the study of the biological system of interest. This multi-disciplinary speciality is called chemical genetics (Schreiber, 1998).

### ***1.10 Chemical genetics***

Chemical genetics uses small molecules to interrupt biological processes by inducing a phenotypic or physiological change (Schreiber, 1998). This approach is more advantageous over the conventional genetics method as the modification is dose dependent and usually reversible (Choi et al., 2014). Two chemical genetic approaches exist: a) a chemical modulator is identified from a phenotypic screen and methods are used to identify the target (forward chemical genetic approach), and b) a target is identified and a chemical modulator/ tool is developed to investigate the mode of action of the target (reverse chemical genetic approach). (Figure 1) (Stockwell, 2000).



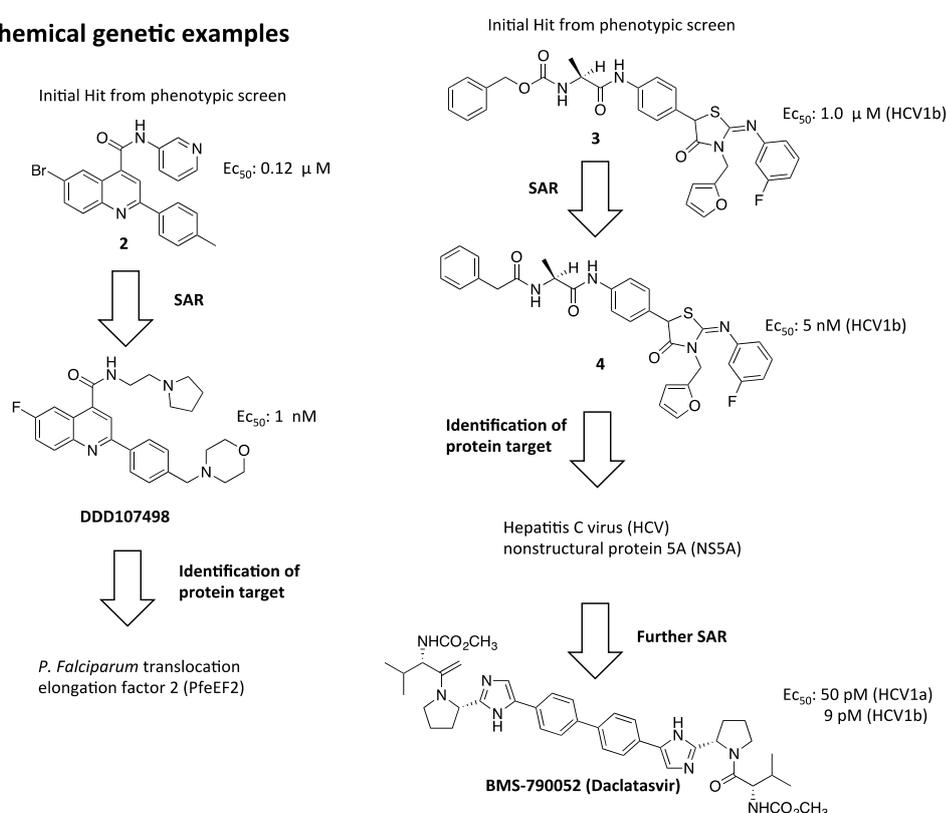
**Figure 1.** A. Schematic of the forward and reverse genetic approach. B. Schematic of the forward and reverse chemical genetics approach (adapted from Stockwell, 2000). Figure created using Microsoft® PowerPoint® 14.6 and ChemDraw® 15.0.

### 1.11 Forward chemical genetics

In the forward approach, phenotypic screens assess small molecule effects at the cellular, tissue or whole organism level (Murphey et al., 2006). The target is then identified through further investigation. The techniques used to identify the target can vary from affinity tags to full genomic-wide studies (Choi et al., 2014). A success story using forward chemical genetics is in the discovery of an anti-malarial published in the journal *Nature* (Baragana et al., 2015). In this recent publication, the Drug Discovery Unit (DDU) in Dundee used a phenotypic screen to identify a novel anti-malarial (DDD107498) that had activity against all stages of the parasite's life-cycle (Baragana et al., 2015). The target was identified by active promotion of drug-resistance through culturing of asexual blood-stage *P. falciparum* in the presence of DDD107498 (Baragana et al., 2015). Genomic deoxyribonucleic acid (DNA) was extracted and sequenced for resistance mutations (Baragana et al., 2015). By comparison with the wild-type (WT) genomic DNA, nine mutations in three clustered regions of *P. falciparum* translocation elongation factor 2 (PfeEF2) were identified, revealing PfeEF2 as a novel parasitic drug target (Baragana et al., 2015). Another literature example of forward chemical genetics was in the identification of a hepatitis C virus (HCV) NS5A inhibitor BMS-790052 (Daclatasvir) (Gao et al., 2010). HCV infection had an unmet need as, at this time, therapeutic strategies relied upon a combination of pegylated interferon- $\alpha$  and ribavirin (Fried et al., 2002). This combination therapy was associated with poor tolerance and a sustained response of less than 50%

(Palumbo, 2011). The study identified a novel small molecule with picomolar EC<sub>50</sub> towards a variety of HCV genotypes in an *in vitro* replicon assay (Gao et al., 2010). The target was identified in a similar approach to DDD107498, replicon cell resistance was generated with incubation of HCV replicon cells with compound **4** (Lemm et al., 2010). Resistant cells were isolated and mutations mapped. This identified HCV NS5A as the protein target (Lemm et al., 2010). Proof of concept and successive progression led to phase I clinical studies where a single 100mg dose reduced mean viral load by 3.3 log<sub>10</sub> after 24hrs within a sustained period of 120hrs (Gao et al., 2010). Daclatasvir (Figure 2) has successfully passed through phase III clinical studies and is now clinically approved as a combination therapy with Sofosbuvir (Sovaldi®). Sofosbuvir is an NS5A inhibitor approved in 2013 by the U.S Food and Drug Administration (FDA) for HCV infection (Chayama et al., 2016).

### Forward chemical genetic examples



**Figure 2.** Schematic of the forward chemical-genetic examples discussed in this subchapter (examples include the discovery and development of compound **2** and compound **3** through phenotypic screening) (Baragana et al., 2015, Gao et al., 2010, Lemm et al., 2010). Figure created using Microsoft® Powerpoint® 14.6 and ChemDraw® 15.0.

Phenotypic screening provides a direct translation into therapeutic impact and a transition to biological target identification (Swinney and Anthony, 2011). The use of phenotypic-screening has led to a variety of successful marketed drugs, such as: Rufinamide (Inovelon®) (Jain, 2000), Memantine (Nameida®) (Witt et al., 2004), Ezetimibe (Zetia®) (Clader, 2004), Cinacalcet (Sensipar®)

(Nemeth, 2006) and Sirolimus (Rapamune®) (Swinney and Anthony, 2011, Vezina et al., 1975). Prior knowledge or understanding of the molecular target or mechanism of action is not required (Swinney, 2013) for further drug development and often the process of forward genetics is not continued (Swinney, 2014). Rufinamide is a first in class central nervous system (CNS) therapeutic developed by Novartis and gained US Food and Drug Administration (FDA) approval in 2008 (Kluger et al., 2009) without prior knowledge of the target (Swinney, 2014).

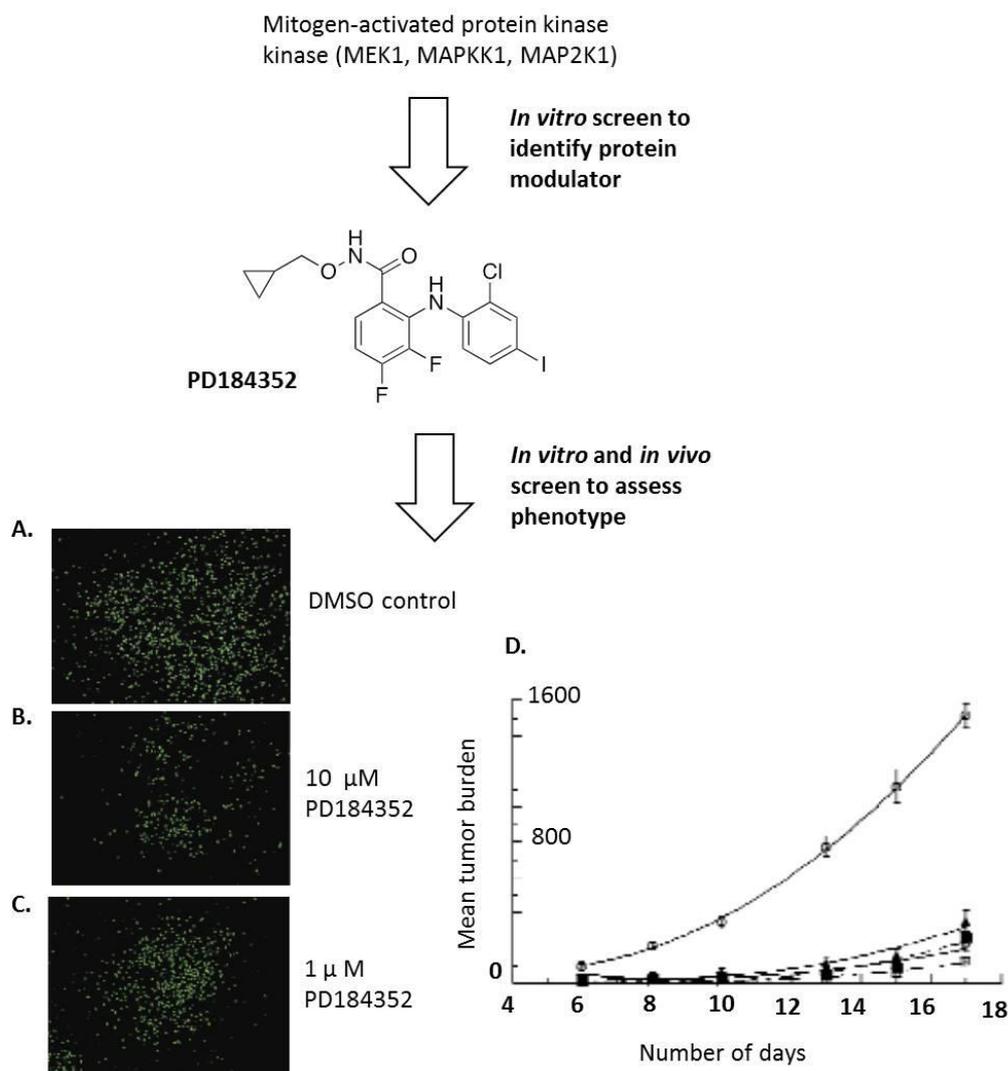
Phenotypic screening is a mainstay of drug discovery (Futamura et al., 2013) and although it was the greatest contributor to FDA approved drugs between 1999-2008 (56%) it has its limitations (Swinney and Anthony, 2011). The challenge of optimising a screening hit without prior knowledge of the molecular mechanism of action to guide design is one such limitation. The screening technology/platform also needs careful design, as development of new screening technology for effective modelling of disease state is problematic and expensive (Macarron et al., 2011). The success of the technology relies heavily on the quality and quantity of the chemical screen (Smith, 2002). Limited diversity in chemical libraries (Macarron, 2006) and the confidential nature of the pharmaceutical companies restricts the success of this approach (Kogej et al., 2013). In academia, a target-centric/reverse chemical genetics approach was more commonly seen due to limited resources and inexpertise in high-throughput screening (Frearson and Collie, 2009, Kawasumi and Nghiem, 2007, Stein, 2003). Improvement in collaboration and partnering between sectors of healthcare research has seen more access and expertise in this area with enterprises such as the Innovative Medicines Initiative (IMI) that enable academic departments and small and medium enterprises (SMEs) access to large chemical libraries (Kogej et al., 2013, Roy et al., 2010). Due to the challenges of discovering novel best-in-class therapeutics, more than one strategy is needed (Al-Ali, 2016). Reverse chemical genetics, which is target-centric, is an alternative to phenotypic screening. A target centric-approach is vital for future drug discovery, particularly in an age when tailored medicines now have an important role (Verweij et al., 2012).

### ***1.12 Reverse chemical genetics***

Target-based screening focuses entirely on a particular biological target. Chemical tools are designed to modulate the biological target and are used in reverse chemical genetics to determine the role and mode of action of the biological target (Spring, 2005). This provides prior knowledge of the molecular target and its relevance to a particular disease or diseases (Spring, 2005). This enables selection and design for superior therapeutic properties, improving the drug candidate's chance of success (Futamura et al., 2013). Target-based screening, which employs reverse chemical genetics has also led to a number of successfully marketed drugs. Examples include: Imatinib (Gleevec®)

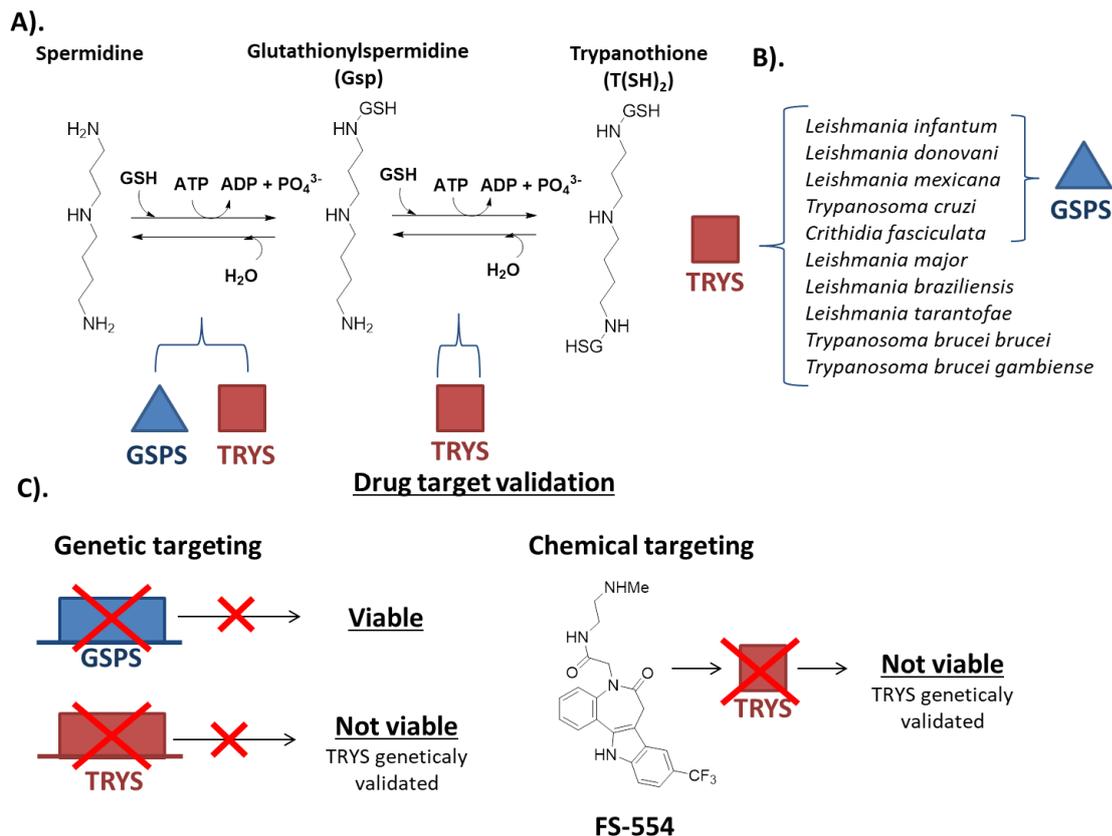
(Capdeville et al., 2002) for chronic myelogenous leukemia(CML), Zanamivir (Relenza®) for flu (von Itzstein et al., 1993), Raltegravir (ISentress®) for human immunodeficiency virus (HIV) (Summa et al., 2008) and Maraviroc (Celsentri®) for HIV (Armour et al., 2006). A small molecule (PD184352, Figure 3) that had an IC<sub>50</sub> of 17nM against MEK1 was identified and developed from an *in vitro* cascade-based assay using glutathione-S-transferase (GST) fusion proteins of MEK1 and MAPK (Sebolt-Leopold et al., 1999). MEK1 integrates signals into the MAPK pathway through phosphorylation of both tyrosine and threonine residues and was identified as a likely cancer target due to its activation leading to cellular transformation (Sebolt-Leopold et al., 1999). To identify the role of MEK1 in proliferative disease, further studies were undertaken with PD184352 (Sebolt-Leopold et al., 1999). In a whole cell assay, PD184352 inhibition of MEK1 decreased the activation of MAPK and reversed high levels of MAPK in human derived cancer cells (Sebolt-Leopold et al., 1999). *In vivo*, PD184352 dosing of tumour-grafted mice decreased tumour growth by as much as 80% (Sebolt-Leopold et al., 1999). This study showed MEK1 to have an important role in MAPK activity and classified it as a drug target for proliferative disease.

## Reverse chemical genetic example



**Figure 3.** A. Schematic of a reverse chemical genetic example in this subchapter showing the development and use of PD184352. A, B and C. Images showing the inhibition of colon 26 cell invasion through matrigel. Cells stained with Calcein Am and then visualised with fluorescence imaging. A. Image of the control experiment with addition of DMSO only. B. Image of colon cancer cell invasion in the presence of 10  $\mu$ M PD184352. C. Image of colon cancer cell invasion in the presence of 1  $\mu$ M PD184352. The inhibition of colon cancer cell invasion in the presence of PD184352 suggests the importance of MEK1 in this cellular process. D. Graph showing the mean tumor burden in mice after tumor implantation in the presence and absence of PD184352. PD184352 was orally administered to mice in doses of 200 mg/kg ( $\diamond$ ), 124 mg/kg ( $\blacksquare$ ), 77 mg/kg ( $\bullet$ ) or 48 mg/kg ( $\blacktriangle$ ). A diluent control was also performed (O). The presence of PD184352 impairs growth of colon tumors. Figure created using Microsoft® PowerPoint® 14.6 and ChemDraw® 15.0. Figures A,B,C and D adapted from Sebolt-Leopold *et al.*, 1999.

A reverse chemical genetics approach combined with genetic analysis identified trypanothione synthetase (TRYS) as an essential enzyme for *Leishmania infantum* (*L.infantum*). Trypanothione (T(SH)<sub>2</sub>) is a redox metabolite of trypanosomatid parasites (Sousa et al., 2014). The synthesis of T(SH)<sub>2</sub> occurs *via* the conjugation of two glutathione molecules to spermidine in the presence of the catalyst TRYS (Sousa et al., 2014). TRYS is present in the redox pathway of trypanosomatid parasites and it is not known if TRYS is essential in all parasites, particularly in parasites expressing glutathionylspermidine synthetase (GSPS) (Sousa et al., 2014, Manta et al., 2013) (Figure 4).

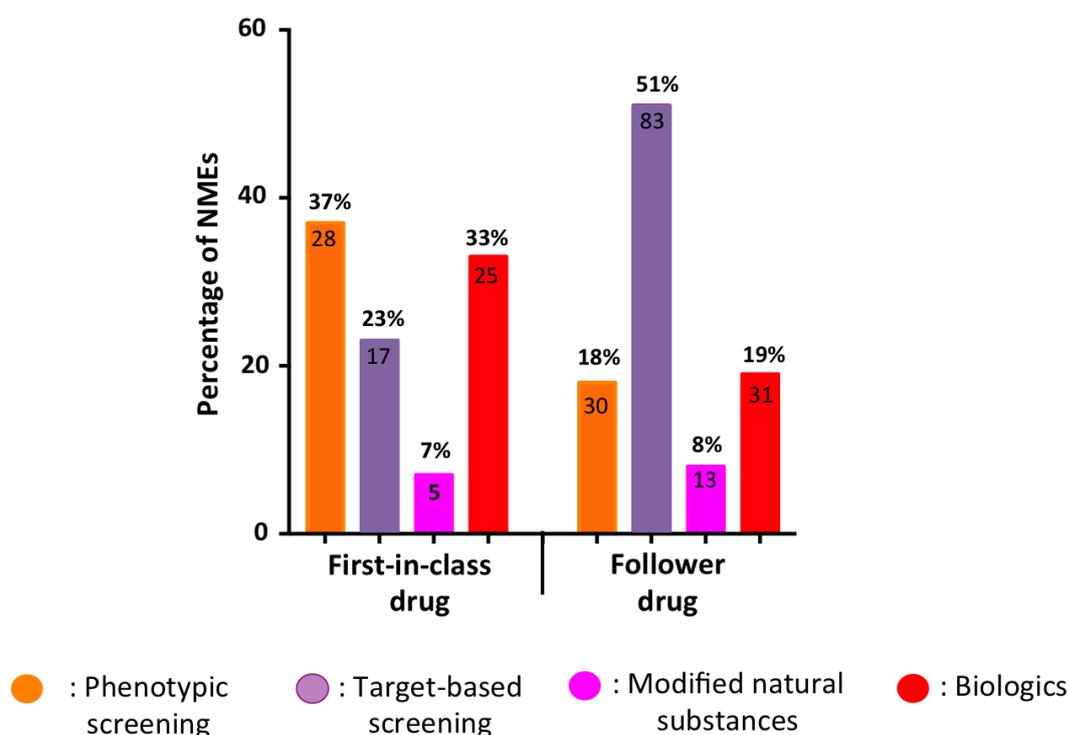


**Figure 4.** Schematic of the reverse chemical genetic approach used to identify TRYS as an essential enzyme for *L. infantum*. **A.** A schematic of the reaction pathway to Trypanothione (T(SH)<sub>2</sub>). TRYS and GSPS are enzymes present in the reaction pathway. **B.** TRYS coding sequences are located in the entire Trypanosomatidae lineage whereas GSPS is located in *L. infantum*, *L. Donovan*, *L. Mexican*, *T. cruzi* and *C. fasciculata*. **C.** Genetic targeting of GSPS and TRYS indicate that TRYS only is important for *L.infantum* survival. Chemical targeting of TRYS confirms that TRYS is critical for *L.infantum* survival.

GSPS is able to catalyse the first step of trypanothione synthesis to produce a monothiol, glutathionylspermidine (Gsp) (Shames et al., 1986). This monothiol can be substituted for T(SH)<sub>2</sub> in essential reactions. Similarly to T(SH)<sub>2</sub>, Gsp can reduce tryparedoxin and is recycled by trypanothione reductase (TR) (Arias et al., 2013). The physiological functions that require the oxidoreductase tryparedoxin include: DNA repair and replication, methionine sulfoxide reduction, iron-sulfur cluster biogenesis and degradation of hydroperoxides, peroxyxynitrite, xenobiotics and oxo-aldehydes (Manta et al., 2013, Comini and Flohé, 2013). It was possible that within parasites harbouring GSPS the enzyme TRYS was redundant (Sousa et al., 2014). A gene targeting approach was employed, using an *L.infantum gsp*<sup>-/-</sup> knockout that was non-lethal and capable of replicating (Sousa et al., 2014). A knockout of *trys*<sup>-/-</sup> was not possible without episomal copies of the gene (Sousa et al., 2014). Retention of this copy throughout several passages indicated TRYS as critical for *L.infantum* survival (Sousa et al., 2014). In further support of this chemical targeting of *L.infantum* TRYS, using a small drug-like compound of the paullone family (FS-554) resulted in parasitic death (Sousa et al., 2014) (Figure 4). Combining both genetic and reverse chemical genetics data indicated TRYS as a chemotherapeutic target for *L.infantum* and possibly the entire trypanosomatid lineage.

During 1999-2008, a target-centric approach (prior knowledge of the molecular target and mode of action) contributed 34% to all FDA approved first-in-class small molecules (Swinney and Anthony, 2011). This contribution is second only to phenotypic screening (Swinney and Anthony, 2011). Other contributors during this period include small molecule new molecular entities (NMEs) based on natural substrates (10%) (Swinney and Anthony, 2011). The relative success of one approach over the others depends upon the type of disease in question, its complexity and biological nature. Relevance to the disease area can give a division in the success of the approach used. In the same study between 1999-2008, the target-centric approach contributed 62.5% and the phenotypic approach contributed to 37.5% of all first-in-class NMEs for cancer (Swinney and Anthony, 2011). For infectious diseases however, the target centric approach and the phenotypic approach contributed 30% and 70% respectively for all first-in-class FDA approved new small molecule therapeutics (Swinney and Anthony, 2011). Phenotypic screening of infectious disease allows for the direct interrogation and impact of the NME on the pathogen/whole cell, already addressing issues with cellular uptake and efflux (Gilbert, 2013). Relative ease for counter screening against mammalian cells and subsequent filtering of compound showing general cytotoxicity leads to a more advantageous discovery set up (Gilbert, 2013). The malfunction of 'normal' cells results in the "hallmarks of cancer" including abnormal cell proliferation and metastatic capabilities (Hanahan and Weinberg, 2011). The challenge in cancer therapy is to eliminate or reprogram cancerous cells while minimizing its effects on 'normal' cells (Moffat et al., 2014). A diverse range of molecularly and

phenotypically distinct cancer types exist requiring a larger repertoire of drugs with different molecular mechanism of action (MMOA) (Moffat et al., 2014). Phenotypic screens for anti-cancer drugs rely predominately on cytotoxic assays that do not effectively model clinical cancer phenotypes or identify NMEs that would avoid targeting ‘normal’ cells (Moffat et al., 2014).

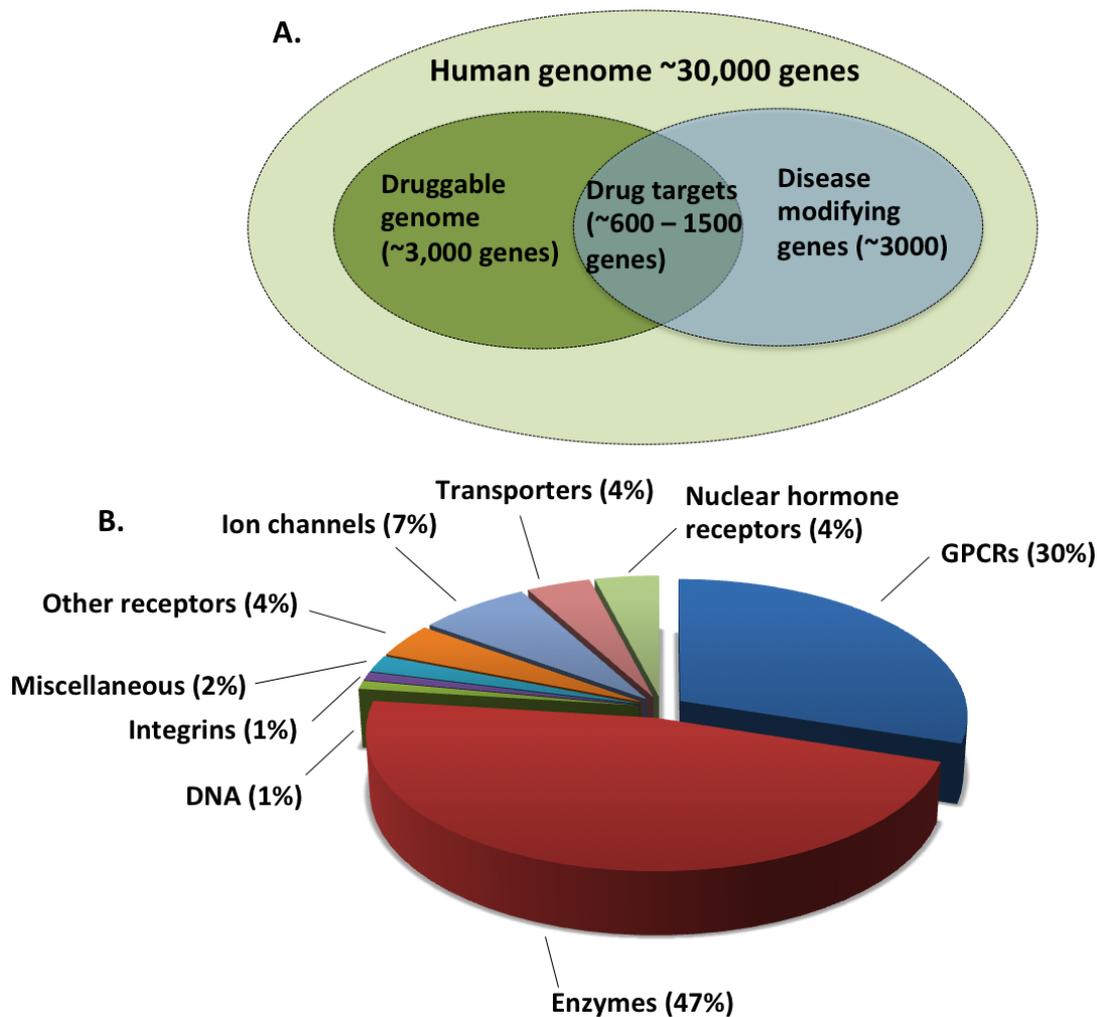


**Figure 5.** Graph illustrating the percentage of NMEs by approach used for discovery for first-in-class drugs and follower drugs (adapted from Swinney, 2013 and Swinney and Anthony, 2014).

In a target-centric approach limitations exist: a) the target may not be relevant to disease pathogenesis, b) modulation of the target may not have a sufficient therapeutic index in patients and c) it may not be clinically competitive (Swinney and Anthony, 2011). If target modulation is prioritized at the expense of understanding the MMOA then the target-based NME will most likely fail. In a study performed by Swinney and Anthony they suggested that the absence of MMOA is a key contributing factor to high attrition of target-based first-in-class NMEs (Swinney and Anthony, 2011). However, if the MMOA is known then target-based development will more likely supersede phenotypic assays. This is supported by follower drugs where target-based approaches account for 51% of NMEs whereas phenotypic approaches account for 18% (Swinney and Anthony, 2011). Chemical biology and good chemical tools are much needed to enable the interrogation of systems in the target-centric approach. This enables the identification and validation of the biological target leading to an improved understanding of the MMOA.

## ***1.20 Biological Target Types and Regulatory Approvals***

A biological target is a broad term for a macromolecular biological entity that can undergo modification/ modulation through binding of a molecule (Hughes et al., 2011). Proteins, polysaccharides, lipids and nucleic acids are deemed to be important macromolecules in normal cellular function and are the four types of macromolecule, which can be modulated by small molecules (Hopkins and Groom, 2002). Abnormalities in the expression of these macromolecules and/or function can result in disease. The difficulties in obtaining potent compounds with low toxicity and high specificity against polysaccharides, lipids and nucleic acids means that the vast majority of successful drugs work through binding to proteins (Hopkins and Groom, 2002). Last year the FDA approved 45 new drugs and 81% of these targeted proteins (FDA, 2016b). Proteins are a large class of drug target-containing families of enzymes, receptors and ion channels. Within these families the most prominent drug targets are hydrolases (enzyme family), G-protein coupled receptors (GPCR, receptor family) and voltage-gated  $\text{Ca}^{2+}$  channels (ion-channel family) (Imming et al., 2006).



**Figure 6. A.** Figure of the estimated number of exploitable drug targets from the genome i.e the ‘druggable’ genome. **B.** Graph of the marketed small-molecule drugs by target class. Figures created using Microsoft® PowerPoint® 14.6 and Microsoft® Excel® 14 (Figures adapted from Hopkins and Groom, 2002).

Excluding cosmetics and imaging agents, the FDA approved 353 NMEs from 2000 to 2012 (Mullard, 2016). This averages at 27 per year of which 65% are small molecules (Mullard, 2016). From 2012 to 2015, the FDA approved 152 new drugs averaging at 38 per year (Mullard, 2016). Last year, 71% of drugs were small molecules with 60% as oral formulations (FDA, 2016b). The number of drug approvals last year (45) is a 19-year high, beating the previous record from the year before of 41 (Owens, 2015). Targeting neglected disease has contributed to this increase in approvals with 47% in 2015 being orphan-designated drugs (FDA, 2016b). Orphan designation is a special status granted through the FDA (Sharma et al., 2010). Orphan drugs are pharmaceutical therapeutics developed to treat a rare disease or condition that affects less than 200,000 persons in the United States (U.S.) (Sharma et al., 2010). Designation of orphan status for a therapeutic that treats a disease that affects more than 200,000 people occurs when: 1) adequate drugs for the disease have

not yet been developed and 2) there is no reason to expect U.S. sales would recuperate the cost of developing the drug for FDA market approval (FDA, 2016c). Incentives for the development of orphan designated drugs include seven years of market exclusivity, waived fees for regulatory activities, tax credits (this can total half of the development costs), research and development grants, access to a single centralised authorisation and assignment of a priority review voucher (Wellman-Labadie and Zhou, 2010). Since 1983, more than 400 drugs and biologic therapeutics are registered with orphan status through the FDA Office of Orphan Products Development (OOPD) (FDA, 2016a). Regulatory rejection rates last year were low with the centre for drug evaluation and research (CDER) only issuing two denied approvals to would-be drugs (Mullard, 2016). This low rejection rate has been attributed to “more meetings with drug sponsors, better applications, and targeted and orphan drugs that offer a clearer benefit-risk balance” (Mullard, 2016). A clearer benefit-risk balance depends on the specificity of the new molecular entity and the identification of a suitable disease target.

### ***1.21 Target Identification***

Target identification is one of the most important steps in a drug discovery campaign, with past failures largely attributed to improper target selection (Hughes et al., 2011). The Genome Wide Project has rapidly expanded the number of potential drug targets (Broder and Venter, 2000, Chanda and Caldwell, 2003). Post-genomic research to associate polymorphisms with disease exacerbation or progression was an approach used for the identification of biological targets (Hughes et al., 2011). A review by Drews of an accumulated pharmaceutical industry portfolio identified 417 targets suggesting that, based on the number of disease-related genes, a potential further 3,000-10,000 could be identified (Drews and Ryser, 1997, Drews, 1996). With inclusion of ligand binding domains and number of potential points upon which therapeutic agents could act, this has been suggested to be even more than 10,000 (Bailey et al., 2001). In a comprehensive survey that followed nine years on from the seminal review by Drews, only 324 drug targets for all classes of approved therapeutic drugs have been identified (Overington et al., 2006). Of the 324 drug targets, 266 are classified as human genome targets of approved drugs (Overington et al., 2006). The number of targets identified post the genome wide project that has led to approved drugs has been disappointing (Clarke and Cooper, 2010). Identifying the targets to proceed with from such a large portfolio is challenging. This strategy linking disease with variation in human genetics and changes in protein expression relies heavily upon good clinical samples, supporting medical data and a good bioinformatics platform for data analysis and correlation (Lindsay, 2003). Bioinformatic approaches have been used to compile data on biological targets from a variety of sources including: publications, patents, gene expression, proteomics, transgenic phenotyping and compound profiling

(Yang et al., 2009). Data mining is then used to prioritise and select potential disease targets (Gershell and Atkins, 2003). The bioinformatics approach relies on the availability of data and conclusive evidence (Whittaker, 2004). Within the genome approximately half of the proteins expressed are functionally unclassified (Hopkins and Groom, 2002). The increase in a large number of uncharacterised potential drug targets from the genome project has resulted in the expansion of chemical biology and the further development of new screening technologies for drug development (Makley and Gestwicki, 2013).

As mentioned above (Chapter One, Section Two, Chemical Genetics), chemical genetics is an alternative approach to target identification and elucidation of target function (Schenone et al., 2013). High quality chemical probes are powerful research tools that have initiated the development of new medicines (Scott, 2016). Biological testing on cell or organisms *via* high-throughput screening has resulted in the discovery of biologically active small molecules (Swinney and Anthony, 2011). Follow up studies and target deconvolution aids the medicinal chemistry process of optimisation, selectivity and identification of promiscuity (Schenone et al., 2013). Using chemical tools on the target in question provides information on the role of that target in disease and a suitable starting point for small molecule drug development (Lipinski and Hopkins, 2004). This information is useful for the selection and narrowing of targets from the 'druggable' genome. Chemical genetics relies on chemical modulation of the target (Spring, 2005). However, not all biological targets are easy to modulate with small molecules. Protein-protein interactions (PPIs) are more acquiescent to antibodies, proteins or peptides due to the need to disrupt large interfaces (Hughes et al., 2011). However, there has been some success with the development of small molecules for the disruption of PPIs. Certain biological classes such as G-protein coupled receptors are more amenable to small molecule modulation (Hughes et al., 2011). This type of receptor contains more than one site of modulation (Hughes et al., 2011). Binding site homology is another complication for specificity. Enzymes within families often have conserved catalytic sites and so targeting a specific enzyme can be difficult (Nussinov and Tsai, 2012). Protein modulation can occur in a variety of different forms and this is explained below (Chapter One, Section Three, Protein Modulation). Validation is another key process in drug development and follows on from target identification.

## 1.22 Target Validation

To be deemed a “valid” biological target for drug development, the biological target needs to fulfil certain criteria including efficacy, safety, commercial and clinical need (Hughes et al., 2011). The biological target also needs to be ‘druggable’, which means it needs to be accessible to the putative drug molecule and this molecule needs to be able to elicit a response *in vivo* or *in vitro* upon binding (Hughes et al., 2011). Identifying and validating a target is a difficult process and relies heavily on the thorough evaluation of the target’s relevance to disease, efficacy and potential impact (Smith, 2003). To gain regulatory approval, sufficient efficacy and safety needs to be demonstrated (Hughes et al., 2011). This is difficult as more than 90% of compounds fail clinical trials (Plenge et al., 2013). The failure rate is highest in phase II with at least 50% due to a lack of efficacy and 25% due to unwanted toxicity. In a retrospective study performed by Pfizer of 44 drug programmes in phase II the lack of efficacy was identified as the major cause of failure (Morgan et al., 2012). To address this problem an emphasis has been placed on the ‘front-loading’ of research to reduce the failure from incorrect biological hypotheses (Lindsay, 2003). Early validation is important in this challenge. The techniques used to validate a target range from *in vitro* tools to *in vivo*/ whole animal disease models (Hughes et al., 2011).

*In vivo* knockout models are useful in determining if the target is disease relevant and produces a similar phenotype (Hughes et al., 2011). Of the 3000-10000 targets identified through disease-related genes, not all of these targets would be valid. Firstly, not all of these targets are ‘druggable’ and secondly, large scale mouse knockout studies reveal that approximately only 10% have the ability to be disease modifying (Gilbert, 2013). Gene knockouts are technically challenging and require avoidance of compensatory mechanisms and developmental phenotype (Hughes et al., 2011). Further complexity occurs when the knockout results in embryonic lethality (Hughes et al., 2011). Knock-in studies are an alternative method to knockout (Manis, 2007). This method replaces functional protein with non-functional protein (Manis, 2007). A key factor in the high failure rate of drug development is the limited ability of pre-clinical disease models to predict direct benefit to patients and so determining if a target is valid is sometimes only possible through clinical testing (Hughes et al., 2011). Confidence increases when a range of techniques is used and the target is multi-validated using different approaches (Hughes et al., 2011). *In vitro* tools provide additional support and validation to *in vivo* experiments. *In vitro* techniques for pre-clinical validation range from genetic engineering at the molecular, whole cell or tissue level to the use of chemical biology for the rigorous testing of targets. The importance of good chemical tools/probes is integral to this challenge as well as confidence in the data set. Chemical probes are necessary for the rigorous testing (Bunnage et al., 2013). In this type of testing a more robust chemical tool will be used.

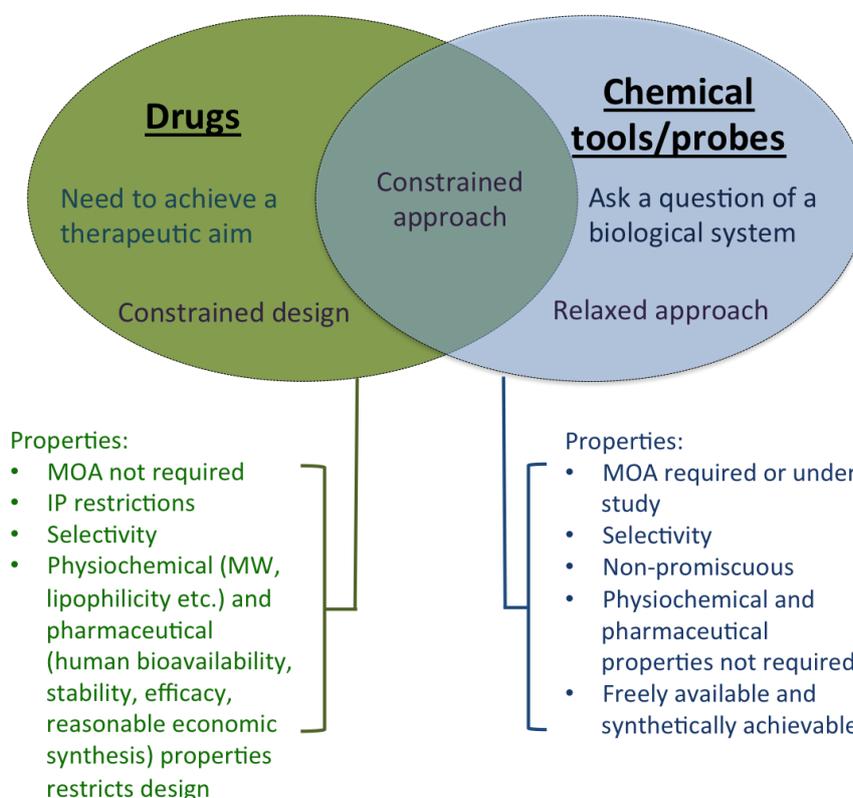
Avoiding promiscuity, achieving good affinity, good permeability, good bioavailability and disease modifying are all areas looked for in a robust chemical tool (Bunnage et al., 2013). Pfizer adopt a 'pillar' framework to target validation, which comprises exposure at site of action (1<sup>st</sup> Pillar), target engagement (2<sup>nd</sup> Pillar) and functional pharmacology (3<sup>rd</sup> Pillar) (Morgan et al., 2012). Within a retrospective study of the 44 drug programmes mentioned earlier all the successful programs achieved the first three 'pillars of survival' (Morgan et al., 2012). In the use of chemical probes a 'fourth pillar', relevant phenotype, is applied to determine if the target modulated is disease relevant (Bunnage et al., 2015). By applying these four pillars a more confident link can be made between target modulation and translation to the clinic (Bunnage et al., 2015, Bunnage et al., 2013). A successful example of this is a bromodomain and extra-terminal domain (BET) bromodomain probe produced by GlaxoSmithKline, GSK525762A (Bunnage et al., 2013, Filippakopoulos et al., 2010).

### **1.23 Chemical Tools/Probes**

Two different strategies can be used to design chemical tools/probes and these depend entirely upon the goal of the researcher. The first strategy is called the relaxed approach (RA). As chemical genetics is not primarily concerned with clinical relevance or therapeutic outcome it is used to determine the identity (forward chemical genetics) or the importance (reverse chemical genetics) of the target in question. Such chemical tools used for forward or reverse chemical genetics need not follow the drug-like property guidelines and so further interrogation of the biological system is unhindered by drug property considerations. In this situation, chemical tools need to only elicit the modulation required to provoke a response. They are required to have some solubility and lipid permeability, the latter depending on the *loci* of the target (Workman and Collins, 2010). Unhindered chemical design could likely feature chemical functionality associated with failure in drug discovery and promiscuous effects on the biological system. Despite the 'relaxed' criteria, chemical tools in this strategy have some similarity with drug-like compounds in that they need to be synthetically tractable, pure and selective for the target to avoid off-target effects and promiscuity (Workman and Collins, 2010).

The second strategy, called the constrained approach (CA), is more specialised. It is employed by drug discovery to develop a tool that would be a suitable starting point for a drug. The intention of this strategy is to avoid the pitfalls of poor *in vivo* chemical utility that the RA can lead to and is more likely to avoid promiscuous effects on the biological system (Workman and Collins, 2010). The scope of commercially available chemical space with the CA is smaller and frequently drug-like compounds are more expensive to purchase or difficult to synthesise (Lipinski and Hopkins, 2004). Ultimately,

both strategies are useful and the one employed for chemical tool development depends on the goals of the researcher. This thesis will look to develop tools for reverse chemical genetics in the elucidation of the function and disease relevance of biological targets.



**Figure 7.** A schematic representing the different requirements of drugs and chemical tools/probes. The general approach applies to probes only whereas the constrained approach applies to both at the interface between probes and drugs. (Adapted from Arrowsmith *et al.* 2015).

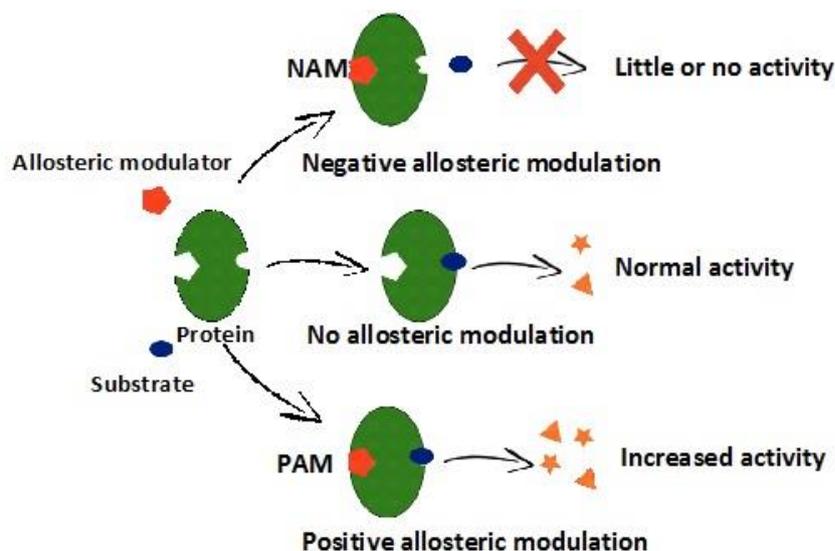
### 1.30 Protein modulation

Modulation of a protein can occur through activation or inhibition. The ‘classical’ mode of modulation is through a small molecule/ligand interaction at the orthosteric site (Nussinov and Tsai, 2012). This site is where the endogenous ligand binds and is often referred to as the active site in enzymes (Nussinov and Tsai, 2012). Agonist and antagonistic activity are two types of modulation that occur. An agonist is a ligand that binds to the protein and activates it to produce a response (Rang *et al.*, 2014). An antagonist is a ligand that binds to the protein and inhibits its response. Agonists or antagonists can have full or partial efficacy resulting in full activation/inhibition or submaximal activation/inhibition respectively (Rang *et al.*, 2014). Competitive or non-competitive agonist or antagonist modulation occurs. When the antagonist or agonist competes directly with the endogenous ligand it is a competitive agonist or antagonist (Rang *et al.*, 2014). If the antagonist or

agonist does not compete directly with the endogenous ligand this is non-competitive modulation and can happen at two sites: the orthosteric site or the allosteric site (Nussinov and Tsai, 2012). This alternative method of modulation is termed allostery.

### 1.31 Allosteric modulation

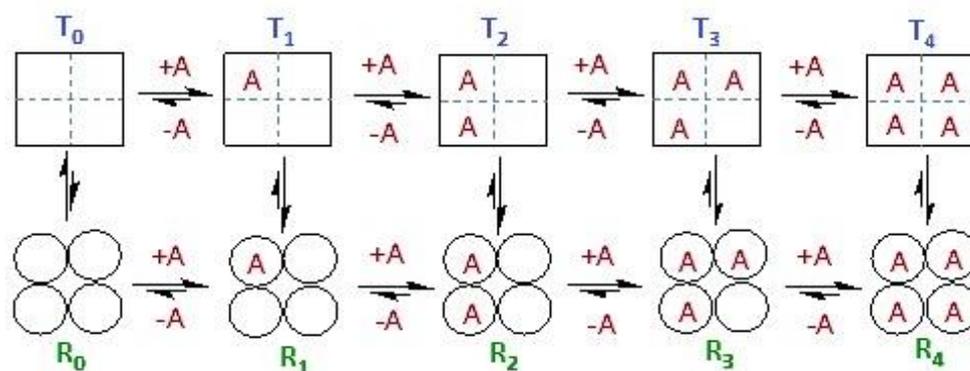
The modulation of a protein can occur through binding of a ligand to a protein on a site that is topographically distinct and remote from the orthosteric site (Changeux, 2013, Conn et al., 2009). This type of modulation is termed allostery (Nussinov and Tsai, 2012). Monod and Jacob introduced the term “allostery” in the 1960’s (Monod et al., 1963). This term was to account for ligand interactions they observed on the enzymatic action of over 24 allosteric enzyme systems (Monod et al., 1963). From this they determined that allostery occurs within proteins that have multiple ligand recognition sites, contain an axis of symmetry and are oligomeric in structure (Monod et al., 1963, Canals et al., 2011). Allosteric sites have been discovered in a variety of protein types including: G-protein coupled receptors, ion channels, enzymes and transcription factors (Conn et al., 2009). Allosteric regulation is present in biological systems and crucial for the control of metabolism. Allosteric regulation can occur through either “negative” or “positive” modulatory mechanisms (Gunasekaran et al., 2004, Conn et al., 2009). Negative allosteric modulators (NAMs) inhibit the protein’s function (this can also be referred to as antagonistic action). Alternatively, positive allosteric modulators (PAMs) enhance the protein’s activity (this can be referred to as agonist action) (Figure 8)(Conn et al., 2009, Gunasekaran et al., 2004).



**Figure 8.** A simplistic diagram of the types of allosteric modulation on an enzyme. The binding of a ligand to the allosteric site can result in a subtle conformational change in the protein conferring action on the orthosteric

site. Commonly appearing within biological systems, this phenomenon governs features of molecular control and cellular function.

The “Monod-Wyman Changeux (MWC)” model was formulated by Monod and Jacob to explain fundamental features of allosteric modulation on oligomeric proteins (Monod et al., 1965). Within this model the protein can exist between two states: the “tense” (T) state and the “relaxed” (R) state (Monod et al., 1965). When a ligand binds to one site it switches the site from T state to a R state causing a conformational change that promotes switching of the other binding sites from the T to the R state (cooperative binding). The R state favours ligand binding and promotes further ligand (A) binding (Figure 9) (Monod et al., 1965). This model was based on the conformational states observed in haemoglobin when oxygen binds to the deoxyhaemoglobin state (Gunasekaran et al., 2004). Cooperative binding of oxygen to the remaining deoxyhaemoglobin binding sites is observed and explained *via* this model (Monod et al., 1965, Milo et al., 2007).



**Figure 9.** A schematic diagram of the MWC model showing the transition between the T and R states upon ligand binding (adapted from Milo *et al.*, 2007).

Failures of this model to explain all features of allosteric regulation resulted in the development of other models. An example of an alternative model is the sequential model (also known as the KNF model) proposed by Koshland, Némethy and Filmer 1966. In this proposed model subunits change conformation between the two T and R states, however a hybrid form (the TR state) also exists between the two states (Levitzki and Koshland, 1969). Nonetheless, allostery and its mechanism are often complicated and might not always follow these proposed models (Changeux, 2012).

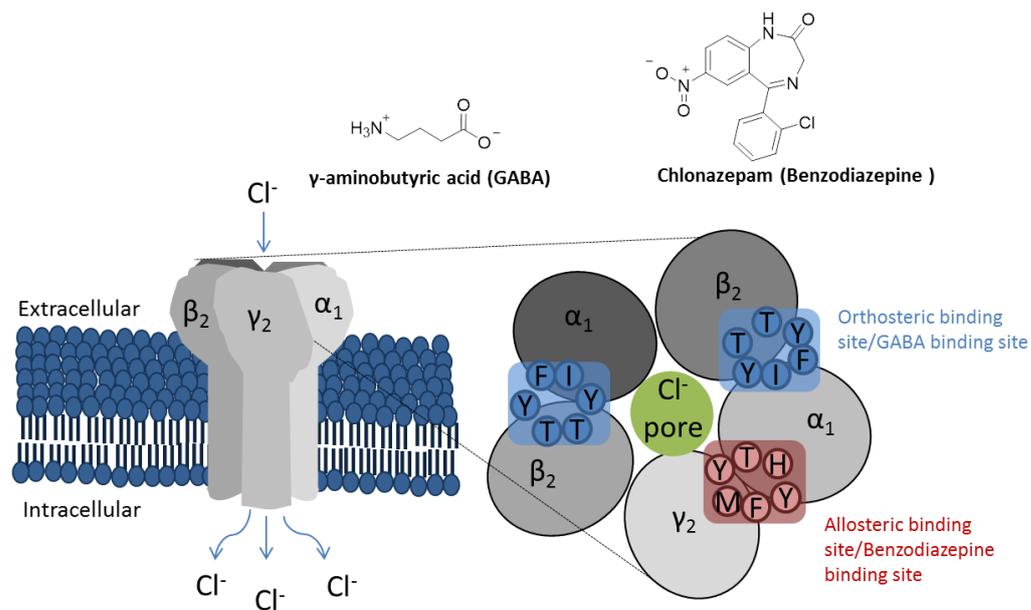
Communication between the allosteric site and the orthosteric site can happen over long distances and despite only a minor conformational change, a significant modification in the protein’s activity can be observed. Within the aspartate receptor, a small 1.0Å structural change in an

allosteric site causes a significant change in activity of the orthosteric site located 100Å away (Gunasekaran et al., 2004). The mechanism and reason for a profound effect on the protein's activity from a subtle change is not well understood.

Allosteric ligands can be split into two classes, these being homotropic (similar binding site to the orthosteric site) and heterotropic (different binding site from the orthosteric site) ligands (Hardy and Wells, 2004). Further subdivision of allosteric sites into orphan allosteric sites and serendipitous allosteric sites was made by Hardy and Wells, 2004. An orphan allosteric site is an "allosteric site used by an, as yet, undiscovered natural effector" and a serendipitous allosteric site is an "allosteric site that may not interact with any natural ligand and has only adventitiously been exploited by the small molecule that binds there" (Hardy and Wells, 2004).

A number of papers give extensive reviews of allosteric mechanisms (Cui and Karplus, 2008, Tsai et al., 2009). This thesis will focus more on the benefits of allostery, its scope and the discovery of allosteric regulators.

Allosteric regulation occurs in a range of protein types including cell surface receptors. Cell surface receptors contribute to more than 60% of current drugs targets (Christopoulos, 2002). Ion channels and G-protein coupled receptors (GPCRs) are two types of cell surface receptor (Christopoulos, 2002). Ion channels are transmembrane proteins possessing a central pore (Unwin, 1989). This pore allows the flux of ions through the membrane (Unwin, 1989). These channels can contain multiple subunits and have the potential for allosteric modulation.

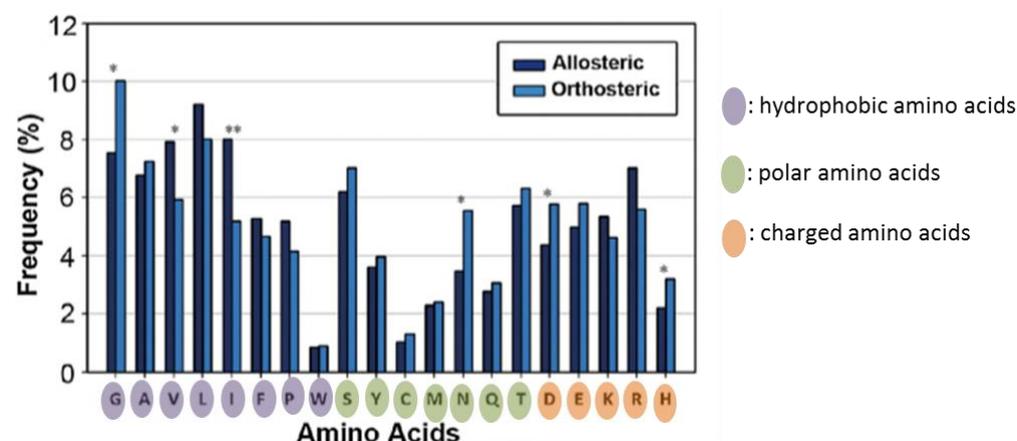


**Figure 10.** Model of the GABA<sub>A</sub> receptor, a transmembrane ligand-gated ion channel found within the central nervous system (CNS). Five subunits combine together to form the central cationic/Cl<sup>-</sup> pore. The orthosteric site is located between the  $\alpha$  and  $\beta$  subunits (highlighted in blue). The orthosteric site is where the endogenous ligand  $\gamma$ -aminobutyric acid (GABA) binds. The allosteric site (highlighted in red) is located between the  $\alpha$  and  $\gamma$  subunits. Benzodiazepines bind to this allosteric site and operate by PAM (Chlonazepam is an example of clinically approved benzodiazepine) adapted from Christopoulos, 2002.

Benzodiazepines are used clinically to treat anxiety and sleeping disorders as well as for the treatment of epilepsy. This class of drugs act allosterically on GABA<sub>A</sub> receptors to increase the ion channel's affinity for the endogenous ligand GABA (Christopoulos, 2002). When GABA is bound to the ion channel it causes a conformational change, which opens the channel and allow a flux of Cl<sup>-</sup> through the central cationic pore (Goetz et al., 2007).

For drug design and development, targeting allosteric sites can present several advantages over orthosteric sites (Hardy and Wells, 2004). Orthosteric sites where the endogenous ligand is a protein or peptide can have physicochemical and pharmacokinetic properties that tend to be unsuitable scaffolds for small molecule drug discovery e.g. for some GPCRs (Conn et al., 2009). Due to this limitation, allosteric sites can provide an alternative avenue for the development of modulators for this protein class (Conn et al., 2009, May et al., 2007). Across certain protein subfamilies, orthosteric sites can be highly conserved (Nussinov and Tsai, 2012). Achieving high selectivity within the protein subfamily by targeting a conserved orthosteric site is challenging (Conn et al., 2009). Examples of highly conserved orthosteric sites within protein subfamilies include: GPCRs and kinases (Conn et al., 2009, Wenthur et al., 2014). Low selectivity leads to "off-target" effects that are

unwanted in drug development (Leeson and Springthorpe, 2007). Allosteric sites are much less conserved across protein subtypes than orthosteric sites due to less severe evolutionary pressure (Wenthur et al., 2014).



**Figure 11.** Graph showing the frequency of amino acids in allosteric sites compared to orthosteric sites (adapted from Li *et al.*, 2013).

The physicochemical environments of allosteric sites are different to that of orthosteric sites (Figure 11). Li *et al.*, 2013 compiled a list of the amino acids and their frequency in allosteric and orthosteric sites (Li *et al.*, 2013). Within allosteric sites a higher proportion of hydrophobic groups were observed suggesting the importance of hydrophobic interactions rather than polar interactions in allosteric binding (Li *et al.*, 2013). The frequency of glycine in allosteric sites was lower than in orthosteric sites (Li *et al.*, 2013). Li *et al.* attributed this difference to restricted movement within allosteric sites as opposed to orthosteric sites (Li *et al.*, 2013). It is clear from this report that allosteric sites contain different frequencies of amino acids to orthosteric sites and consequently have different physicochemical properties to orthosteric sites. Allosteric sites provide an alternative option for drug development (Wenthur et al., 2014).

Achieving high affinity and selectivity for the allosteric site can prove challenging, which limits the success of this developing allosteric modulators within drug development programmes (Wang et al., 2012). This may be due to a lower frequency of polar amino acids within allosteric sites and the reliance on hydrophobicity for ligand to protein contacts (Wang et al., 2012). Another disadvantage is that resistance is more likely to arise within allosteric sites as these sites are functionally less important than active sites/orthosteric sites (Nussinov and Tsai, 2013). Despite this, a number of allosteric modulators have made it through to market in the past years, examples include: Trametinib a mitogen-activated protein kinase (MEK) inhibitor (FDA approved for metastatic melanoma in 2013) (Wu et al., 2016), Maraviroc a C-C chemokine receptor 5 (CCR5) antagonist (FDA

approved in 2007 for HIV) (Armour et al., 2006) and Cinacalcet a calcium-sensing receptor (CaSR) activator (FDA approved for hypercalcemia in hyperthyroidism in 2004) (Wenthur et al., 2014, Nemeth, 2006).

### ***1.32 Discovery of Allosteric Inhibitors***

The discovery of allosteric sites is not simplistic (Wang et al., 2012). Orthosteric cavities are rich in functional groups suited for small molecule binding (Hardy and Wells, 2004). This renders selective screening for allostery difficult making the process laborious (due to the need to filter out and discount orthosteric binders) (Sadowsky et al., 2011). High throughput screening is the traditional approach to allosteric site discovery (Hardy and Wells, 2004). This method enables the testing of a large range of chemical entities for activity on biological screens (Macarron, 2006).

The identification of non-competitive inhibitors within high throughput screens (HTS) can be achieved (Hardy and Wells, 2004). However, many of the non-competitive inhibitors identified either act by nonsensical or non-drug like mechanisms including denaturation, high-stoichiometry binding and protein aggregation (McGovern et al., 2002, Rishton, 2003, McGovern et al., 2003). HTS therefore tends to be an inefficient and laborious process due to the high volume of promiscuous data (Hardy and Wells, 2004). Despite such difficulties, combining high throughput screening with X-ray crystallography has led to the successful discovery of several new allosteric sites (Hardy and Wells, 2004). Using a HTS screen of 300,000 compounds, Pfizer and the National Hellenic Research Foundation discovered an allosteric site within the human liver glycogen phosphorylase (GlyP) (Oikonomakos et al., 2000, Martin et al., 1998, Rath et al., 2000). GlyP is an enzyme involved in the control of blood glucose levels (Hardy and Wells, 2004). This enzyme is competitively inhibited by glucose and allosterically modulated (PAM) by adenosine monophosphate (AMP) (Hardy and Wells, 2004). Indole-based compounds were observed to bind in a site situated 30 Å away from the catalytic site and 10 Å away from the AMP site (Hardy and Wells, 2004). Indole-based compounds stabilised the enzyme in a T state, which promoted glucose binding and synergistic inhibition (Hardy and Wells, 2004).

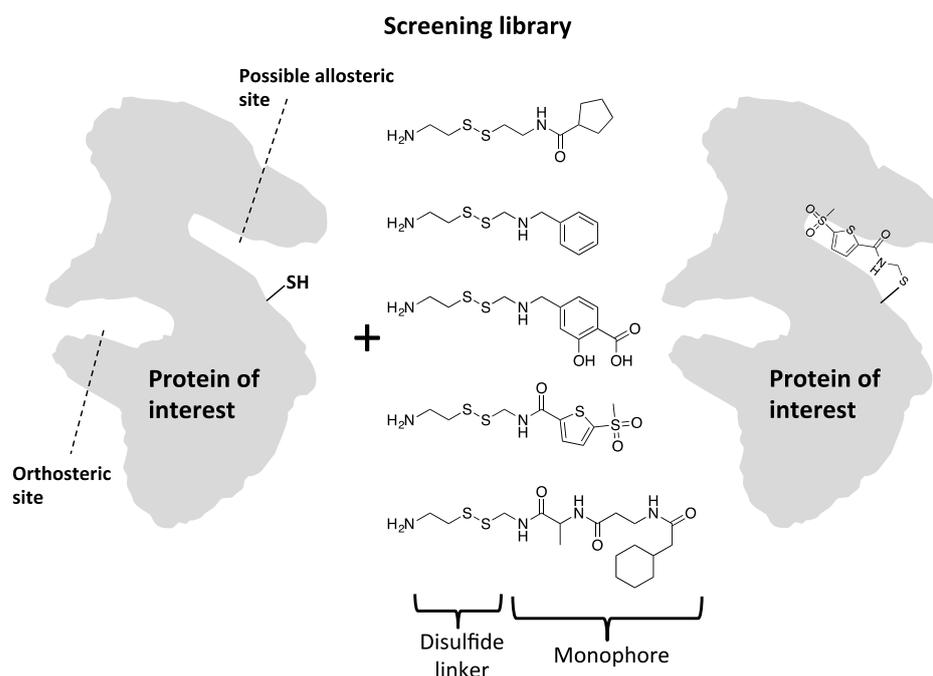
### ***1.33 Phage Display***

Libraries of peptides, proteins or antibodies can be generated through phage display technology (Terstappen et al., 2007, Rossenu et al., 1997). DNA fragments encoding certain peptide/protein sequences are inserted into a phage (Terstappen et al., 2007). The phage then expresses the peptide/protein on its surface (Terstappen et al., 2007). Screening of phage is performed against a given target and the phage that binds to the target is collected and amplified (Terstappen et al.,

2007). The protein, peptide or antibody on the phage coat is identified from the DNA sequence (Terstappen et al., 2007). A group at Genentech used peptide phage display to develop and find inhibitors of factor VIIa (FVIIa), which is an essential protease for blood clotting (Dennis et al., 2000). They found that two peptides were bound to this protease at two separate allosteric sites (Roberge et al., 2001, Maun et al., 2003).

### 1.34 Tethering approach

A different approach to discover allosteric sites uses cysteine tags as a means of covalently trapping bound small molecules in secondary sites (Hardy and Wells, 2004). Tethering involves screening a library of compounds that contain thiols, against proteins with native or introduced cysteines on their surface (Hardy and Wells, 2004). Small molecules that are tight binders have the residency time to form a disulfide bond between the thiol within the molecule and the cysteine present on the protein (Hardy and Wells, 2004). Combining this approach with peptide mapping can negate the need for X-ray structure determination providing a means by which to discover new allosteric sites (Hardy and Wells, 2004).



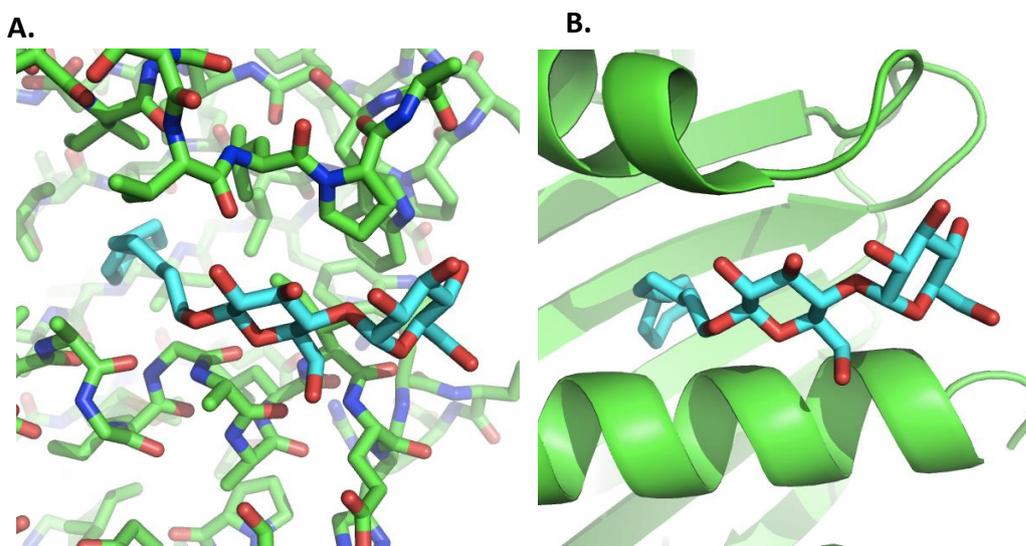
**Figure 12.** Diagram of the tethering screening approach adapted from Hardy and Wells, 2004. Protein of interest (POI) is incubated with a screening library. Each compound in the library has a unique molecular weight and contains a disulfide linker and a monophore. The level of reductant is adjusted so that only tightly binding monophores are present long enough to form disulfide linkers with a cysteine on the protein. Mass spectrometry identifies covalently bound monophores. Kinetic analysis or binding analysis of the POI with the monophore of interest investigates if the novel site is allosteric (adapted from Hardy and Wells, 2004).

A new allosteric site on the dimer interface of caspase-3 and caspase-7 was discovered using this approach (Hardy et al., 2004). Caspase-3 and caspase-7 are two apoptotic proteases involved in the caspase cascade resulting in the formation of the mitochondrial pore and controlled cell death (Fan et al., 2005).

### ***1.35 Fragment screening and serendipitous binders***

Fragments are small molecules of less than 250 Da (Zartler and Shapiro, 2005). Screening with fragments requires sensitive high-throughput screens due to low binding affinities (Carr et al., 2005). Fragment screens have advantages over the conventional high-throughput screens (Rees et al., 2004). Fragments lower the chances of unfavourable interactions and can identify the minimal binding unit necessary for affinity and potency (Rees et al., 2004). Consequently, this results in modulators that have high ligand efficiency (Murray et al., 2010). Another advantage of fragment screens is they tend to be more chemically diverse than conventional HTS libraries (Keserú et al., 2016). Allosteric regulators of 3-phosphoinositide-dependent kinase-1 (PDK1) in the phosphoinositide 3-kinase (PI3K) pathway was identified by Pfizer using nuclear magnetic resonance (NMR) spectroscopy combined with fragment-based screening (Stockman et al., 2009). By enriching PDK1 with  $^{15}\text{N}$  and acquiring  $^1\text{H}$ - $^{15}\text{N}$  transverse relaxation optimized spectroscopy (TROSY) spectra they identified binders in an allosteric PDK1 interacting pocket (Stockman et al., 2009).

An alternative method to allosteric site discovery was first described by Hardy and Wells, 2004. This approach uses information within X-ray crystal structures to find novel binding pockets from serendipitous binders. In this paper, a maltoside used in the crystallisation buffer was found serendipitously bound to the allosteric site of a  $\beta$ -lactamase crystal structure (Figure 13) (Hardy and Wells, 2004).



**Figure 13.** Images of cyclohexyl-hexyl- $\beta$ -D-maltoside bound in the allosteric site of S70C SHV  $\beta$ -lactamase (PDB: 4FD8). A. Stick representation of the maltoside and protein complex .B. The maltoside is located between the two alpha helices within the allosteric site of the protein. Figures created using Microsoft® PowerPoint® 14.6 and PyMOL®.

Horn and Schoichet were the first to identify this site as an allosteric site. They used thermodynamics, mutagenesis and crystal structure analysis to identify the allosteric site (Horn and Shoichet, 2004). Hardy and Wells discovered that maltoside (a crystal additive) was found bound to the same site on a similar  $\beta$ -lactamase (Hardy and Wells, 2004). In this instance, crystal additives serendipitously bound to the crystal structure would have identified the novel allosteric site without the need for laborious experimentation (Hardy and Wells, 2004). Within this paper, Wells *et al.* acknowledged that serendipitous binding of “crystallisation artefacts” (small molecules present in the crystallisation buffer) within protein cavities is not uncommon (Hardy and Wells, 2004). They further suggested that these “crystallisation artefacts” recorded in the protein data bank (PDB) could be the largest collection of information on serendipitous binders and undiscovered allosteric sites (Hardy and Wells, 2004).

The Hardy and Wells approach provides an advantage over the previous defined approaches of allosteric site discovery described in this section. Firstly, the protein crystal structure has been solved and conditions for purification, expression and crystallisation are published (reducing the time/workload needed to generate crystals for structure based drug design). Secondly, the small molecule has identified a suitable target without the need for high-throughput screening and laborious analysis. Finally, the secondary binding pocket and consequently residues located within this pocket have been identified for binding small compounds to. This then allows a more site

directed approach to computational or fragment based screening to aid the development of a lead compound.

Serendipitous binders are useful in identifying a novel-binding pocket suited for a small molecules, however they do not identify if this site is allosteric. The novel binding site would have to confer some negative or positive communication with the orthosteric site resulting in NAM or PAM. Extensive kinetic/functional analysis is still a requirement in order to define if this site is allosteric.

### ***1.36 The correct approach to allosteric discovery***

All approaches to allosteric discovery have disadvantages and no single approach is the correct approach. High-throughput screening either through large “lead-like”, fragment or phage-based libraries identify a large plethora of “hits”. Not all “hits” identified will be true modulators of the biological target in question and the true “hits” identified from the screens may not modulate *via* allosteric mechanisms. Further work is needed to establish if the modulator is allosteric and this is accomplished *via* structural binding and/or extensive kinetic analysis. It requires a skilled team to work through the results from a screen to design and conduct experiments to establish the method of modulation, which can be time consuming and difficult (particularly if a structure of the target has not been obtained previously). The advantage of HTS/fragment-based/phage-display screening is that the chemical modulator/modulators identified already have reasonable potency and are a good starting point for SBDD. Additionally, a number of “hits” identified from the screen may bind in similar positions providing useful structural information for further SBDD optimisation.

Approaches specifically designed to identify novel binding pockets such as the tethering approach also have their limitations. The monophore binding unit, the length of the linker and the position/s of the cysteine on the protein are factors that could limit the success of this screening approach. To improve the likelihood a successful allosteric “hit” would be identified through the tethering approach a large monomer-linker screening library is required with varied linker lengths and monophore binding units. The position of the cysteine residue is important as a novel binding pocket/allosteric site can only be identified if the cysteine residue is within close proximity for disulphide linker reactivity.

The other approach to allosteric site discovery includes the data mining approach identified by Hardy and Wells. Although serendipitous binders might identify novel pockets amenable to small molecule binding, this method provides no functional information relating small molecule binding to protein modulation. The major challenges for this approach include: a) identify if the bound serendipitous modulator causes functional modulation of the protein and b) establish affinity,

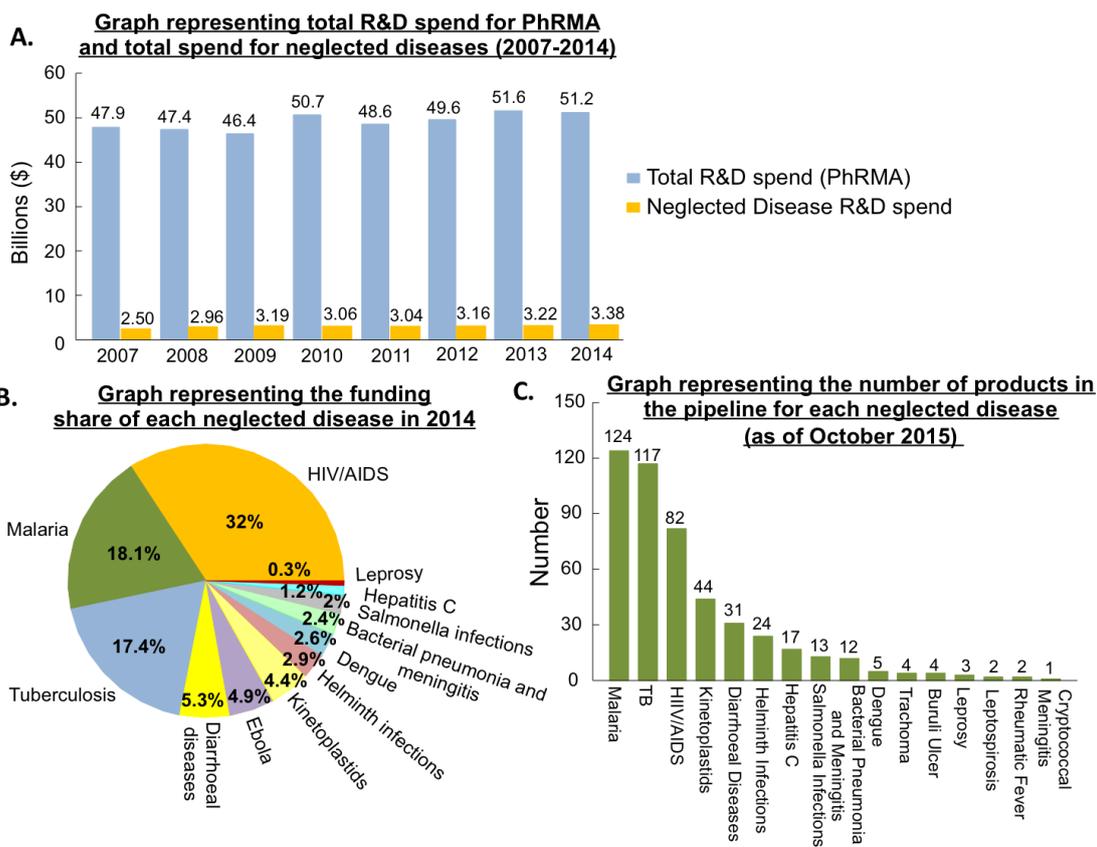
potency and selectivity at this site. A further challenge exists in sorting through PDB structures and the serendipitous binders within the PDB. Chemicals are commonly used in high concentrations in order to generate the conditions appropriate for crystal formation and cryoprotection. Consequently, a high proportion of the serendipitous binders found within these protein structures are likely to be surface bound and promiscuous. Mining for novel allosteric sites within the PDB using serendipitous binders as guides will be labourious and so serendipitous binders buried deep within pockets should be prioritised for screening. Additional fragments/small molecules found bound in close proximity to serendipitous binders could provide additional structural information for improved SBDD. This was the approach used within this thesis.

### ***1.40 Neglected disease (ND) and unmet need***

Neglected disease (ND) corresponds to disease that principally impacts on the world's poorest. Global Funding of Innovation for Neglected Diseases (G-FINDER) defines 34 diseases as ND or having an unmet need (POLICYCURES, 2015, POLICYCURES, 2016). Of the 35 diseases, 15 are classified as neglected tropical diseases (NTDs) and include: dengue fever, helminth infections (acariasis, ancylostomiasis, necatoriasis, trichuriasis, strongyloides, lymphatic filariasis, onchocerciasis, schistosomiasis and cysticercosis/taeniasis), kinetoplastids (human African sleeping sickness, leishmaniasis, Chagas disease), trachoma, Buruli ulcer and leprosy (Choffnes and Relman, 2011). NTD is a subset of ND that predominates in the tropics where poverty is found in the greatest concentration (Feasey et al., 2010). The World Health Organisation (WHO) estimates NTD to be a health burden with more than 1 billion people infected in 149 different countries (WHO, 2012, Armah et al., 2015). NDs have a lack of options for treatment or prevention and are generally disregarded by the pharmaceutical industry due to the likelihood of poor commercial revenue (Mrazek and Mossialos, 2003, Chokshi, 2006). As a consequence, the research and development (R&D) of new ND or NTD therapeutics is poorly funded (Choffnes and Relman, 2011). The most recent annual survey from G-FINDER (at the time of writing) is a report detailing ND R&D funding in 2014 (POLICYCURES, 2015). In this year \$3.4 billion was invested globally in ND R&D (POLICYCURES, 2015). This investment is up 4.9% from the previous year, however this is attributed to the Ebola epidemic (Ebola global funding totalled \$165 million in 2014) (POLICYCURES, 2015, POLICYCURES, 2014). Global total R&D spending in 2014 of the Pharmaceutical Research and Manufacturers of America (PhRMA) was estimated at \$51.2 billion (PhRMA, 2015). PhRMA has 58 members consisting of the largest biopharmaceutical companies and their subsidiary companies (PhRMA). ND R&D spending represents 6.6% of the total PhRMA R&D spending. This is a crude estimate of R&D PhRMA contribution funding for ND. It is likely this contribution is lower as this estimate does not account

for biotech, public/government or philanthropic contribution to the \$3.4 billion total ND fund. Although spending on ND R&D by the pharmaceutical industry is disproportionately low, a 28% rise in industry funding for ND R&D was observed in 2014 compared with 2013 (POLICYCURES, 2015, POLICYCURES, 2014). Combining this rise with a decrease of 0.8% in PhRMA total R&D spending for 2014 from the previous year suggests greater interest in the development of ND therapeutics (Figure 14). The introduction of incentive mechanisms has likely promoted this interest. The current incentives for ND R&D investment include: access to public funding and partnerships (enables the cost of development to be shared with publically funded research institutions), tax credits (in the U.S.) this would make pharmaceutical companies eligible for a 20% R&D tax credit or if the disease had orphan status in the U.S., e.g. tuberculosis, it could be as high as 50% and prizes (this includes FDA priority review vouchers and a payment to the researcher on the condition the therapeutic achieves a set outcome) (Mueller-Langer, 2013).

A large disparity is observed for funding between the different neglected diseases. The three 'top tier' diseases HIV/acquired immune deficiency syndrome (AIDS) (32%), tuberculosis (17%) and malaria (18) received over two thirds of the funding designated for neglected disease (67%) (POLICYCURES, 2015). Kinetoplastids, a 'second tier' disease, received 4.4% of the 2014 total R&D funding. Diseases designated 'second tier' include: kinetoplastids, helminth infections, dengue, diarrhoeal diseases, bacterial pneumonia and meningitis, salmonella infections, hepatitis C (genotypes 4,5 and 6) and Ebola (POLICYCURES, 2015). Buruli ulcer, a 'third tier' classified disease, only received 0.1% of the total 2014 R&D funding (the 'third tier' disease classification includes: trachoma, Buruli ulcer, leprosy, rheumatic fever, cryptococcal meningitis and leptospirosis) (POLICYCURES, 2015). ND represents one disease classification where there is an unmet need for the development of new therapeutics. It is clear that some diseases classified under ND receive greater support than others (Figure 14). Global disease burden (GDB) is likely the reason for the disparity seen in ND R&D funding. The observed GDB for HIV/AIDS is approximately 36.9 million patients with a mortality of 1.2 million in 2014 (UNAIDS, 2015, Lima et al., 2016). Malaria has a GDB of 214 million patients with a mortality of 438, 000 in 2015 (Coats, 2016). For tuberculosis the GDB is 9.6 million patients with 1.5 million mortalities in 2014 (WHO, 2016a). In contrast to 'top tier' diseases, kinetoplastids have a GDB of approximately 9.3 million patients and 31,000 mortalities in 2014 (Chagas disease independently has a GDB of 8 million with 10,600 mortalities in 2014) (Beaumier et al., 2016, WHO, WHO, 2016b). Buruli ulcer, a 'third tier' disease, has a GDB of 2,200 patients and unreported mortality in 2014 (WHO, 2016c). GDB is linked to commercial viability and so diseases with low global patient population and mortality are less lucrative (Pollastri, 2014). These diseases are subsequently reliant upon public and philanthropic funding (Moran et al., 2009).



**Figure 14.** A. Graph representing global spend on R&D for PhRMA (consisting of 39 members of the largest pharmaceutical and biotechnology companies) as a comparison against global neglected disease spend. B. Graph representing the difference in funding in 2014 for each neglected disease. Tier one: HIV/AIDS, Malaria and Tuberculosis receive the greatest proportion of funding, whereas tier two and tier three receive a much lower share. Tier three neglected diseases sharing less than 0.3% of total 2014 neglected disease fund are not included on the graph (this includes: Trachoma, Cryptococcal meningitis, Buruli Ulcer, Leptospirosis and Rheumatic fever). C. Graph representing the number of products in the pipeline for each neglected disease. Statistics are taken from G-FINDER and Reuters. Graphs generated using SLIDE® and manipulated for presentation using Microsoft® PowerPoint®.

Unmet need also corresponds to orphan disease (previously mentioned in 1.20 Biological Target Types and Regulatory Approvals), emerging disease and re-emerging disease (Morens et al., 2004). Emerging disease relates to a disease that has appeared within a population for the first time or existed previously but has rapidly increased in incidence or geographical range (Morens et al., 2004). The sudden recurrence of Ebola in 2014 is an example of an emerging disease due to its rapid incidence rate and geographical distribution (Morens et al., 2004). The first appearance of Ebola was in Central Africa in 1976 with a small local outbreak in a small town and neighbouring areas. The

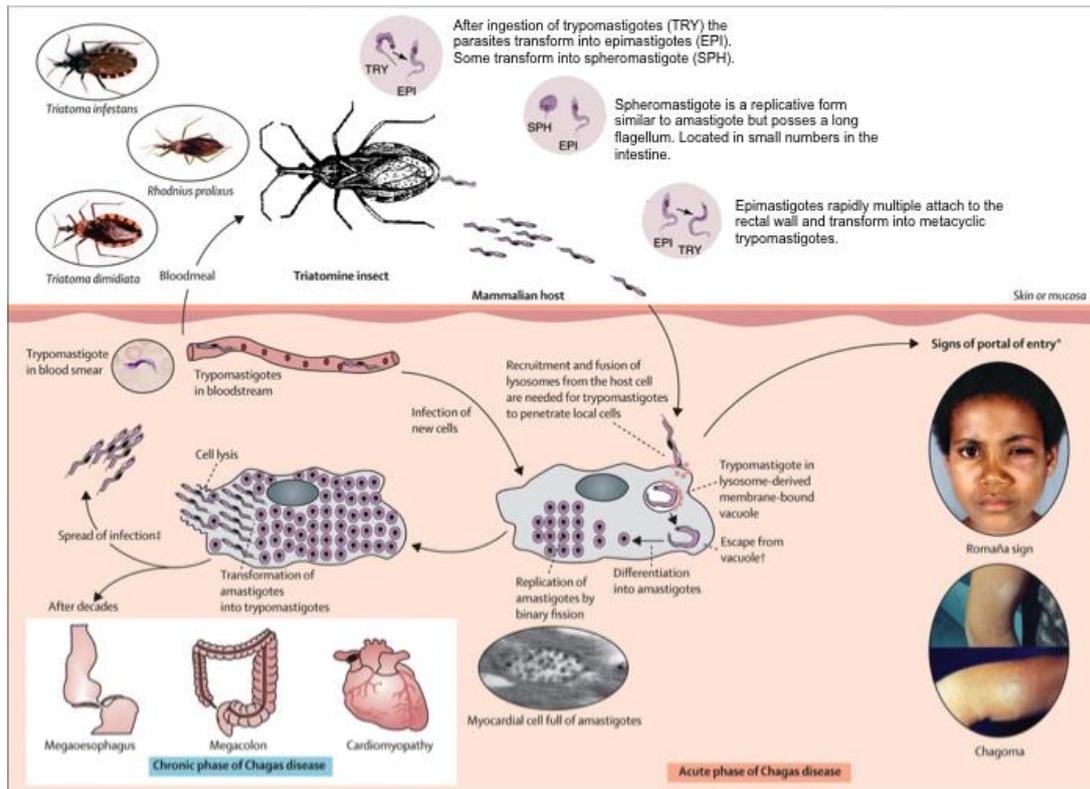
total number of infected individuals totalled 284 individuals resulting in 151 mortalities. Further outbreaks were reported in Democratic Republic of Congo (318 cases with 280 mortalities) and in the U.S. (identified in imported *Cynomolgus* monkeys) (Feldmann et al., 2003). Ebola became a threat to public health in 1995 when outbreaks were reported in Kikwit and Democratic Republic of Congo (293 cases with 233 deaths) (Feldmann et al., 2003). In 2014 the WHO declared a public health emergency in response to an outbreak of Ebola that resulted in more than 8,000 deaths (an underestimated figure), spread to eight countries and infected more than 20,000 individuals before containment (Sanchez et al., 1995).

Re-emerging disease relates to a disease that has emerged again with no sudden change in incidence (Morens et al., 2004). This occurs in the same geographical location or can occur in a different geographical location. Tuberculosis and malaria are two examples of re-emerging diseases (Morens et al., 2004). Tuberculosis was well managed with previously developed therapeutics in the past but has re-emerged in immune deficient patient populations (HIV/AIDS) (Morens et al., 2004). The evolution of *Mycobacterium tuberculosis* in HIV patients has contributed to the emergence of drug-resistant and multi-drug resistant strains (Shah et al., 2007). Drug resistance is a factor that causes viral and microbial re-emergence resulting from the mutation or bacterial acquisition of genes (Morens et al., 2004). Antibiotic resistance is a major health concern (see Antibiotics) and antibiotic resistant bacterial infection is an example of a re-emerging disease (Jones et al., 2008).

The impact of both emerging and re-emerging diseases on human populations is dependent on the rate and spread across geographical locations (Morens et al., 2004). Diseases emerging or re-emerging in populations within developing countries is likely to have a greater impact and GDB (Morens et al., 2004). Chemical biology can have a major role in identifying targets and understanding MMOA in neglected disease and diseases with unmet need. This research carried out by independent publically funded research institutions would reduce the costs associated with R&D (Nwaka and Ridley, 2003). This would make the development of therapeutics for some of the disregarded diseases more attractive and lucrative to the pharmaceutical industry (Nwaka and Ridley, 2003).

### **1.41 Chagas disease**

Chagas disease is named after a physician and researcher (Carlos Riberio Justiniano Chagas) who discovered the disease in the 1900s (Steverding, 2014). It is caused by the vector-borne protozoan parasite *Trypanosoma Cruzi* (*T. cruzi*). The vector is a triatomine bug (“kissing bug”) that ingests the parasite upon taking a blood meal from an infected mammalian host (Steverding, 2014). Within the vector the epimastigotes multiply in the midgut and transform to metacyclic trypomastigotes in the hindgut (Steverding, 2014). Disease transmission of the metacyclic trypomastigotes occurs when urine and/or faeces of the blood-sucking triatomine bug contact a wound or intact mucous membrane of the mammalian host enabling trypomastigote entry (Steverding, 2014). Other methods of transmission include: organ transplantation, blood transfusion and ingestion of contaminated food and drink (Steverding, 2014). Newborn infants are at risk of transmission from their mothers through breast milk or congenitally through the placenta (Steverding, 2014). Within the human body metacyclic trypomastigotes penetrate into cells and transform into amastigotes (Steverding, 2014). These amastigotes undergo binary fission in cells of the infected tissue. These intracellular amastigotes can transform into trypomastigotes and burst out of the cell. The circulating trypomastigotes can infect other cells and organs resulting in clinical manifestations (Figure 15) (Noireau et al., 2009).



**Figure 15.** Schematic of *T. Cruzi* life cycle within vector and mammalian host (adapted from Rassi and Marin-Neto, 2010). Localised inflammation at the site of entry (Chagoma and Romaña's sign) is a classical sign of Chagas disease. The Chronic phase of Chagas disease can occur after many years. During the acute phase the patient/animal can be asymptomatic. Figure 14. Schematic of *T. Cruzi* life cycle within vector and mammalian host (adapted from Rassi and Marin-Neto, 2010). Localised inflammation at the site of entry (Chagoma and Romaña's sign) is a classical sign of Chagas disease. The chronic phase of Chagas disease can occur after many years. During the acute phase the patient/animal can be asymptomatic.

In the initial/acute phase, which begins 6-10 days after infection and can last for about 4-8 weeks, symptoms are absent or mild and are unspecific/typical for many infections (Rassi and Marin-Neto, 2010). A specific symptom occurs at the entry point of infection. Inflammation at the *loci* of the bite or trypomastigote infection site results in a Chagoma (Rassi and Marin-Neto, 2010). If the eye is the entry point, a unilateral periorbital swelling, also known as a Romaña's sign, may occur (Rassi and Marin-Neto, 2010). Electrocardiogram (ECG) abnormalities can follow in about 50% of infected individuals but usually disappear. Following the acute phase, a prolonged asymptomatic chronic "indeterminate" phase can ensue and can last for a patient's lifetime (Steverding, 2014). Approximately 15-30% of people enter into the determinate/chronic phase and this can last for 10-25 years post infection (Steverding, 2014). In the chronic phase the disease exhibits diverse clinical manifestations including cardiomyopathy (30% of patients), neuropathies and dilations of the colon or oesophagus (10% of patients) (Steverding, 2014). Immune suppression can result in reactivation

of latent infection with chronic infection resulting and cardiac, neuropathic and oesophageal manifestations (Steverding, 2014). Low levels of circulating parasites at the “indeterminate” stage cause limited success from serological diagnostic tests (Sabino et al., 2013, Angheben et al., 2015).

An estimated 6 to 7 million people worldwide are infected by the parasite *T. cruzi* and approximately 25 million people live in risk areas (WHO, 2010). Traditionally thought of as a disease endemic to Latin America, the WHO found cases of Chagas disease in Latin American immigrants to Europe (WHO, 2010). Transmission through blood donation has occurred in North America, Europe, Japan and Australia resulting in its classification as a global health threat and new practices to limit transfusional transmission (Angheben et al., 2015). An enzyme-linked immunosorbant assay (ELISA) is used to screen blood donors for Chagas disease (FDA approved in 2006) (Kirchhoff and Pearson, 2007). This assay uses antibodies present for epimastigote lysate antigens to detect *T. cruzi*. Donor selection criteria and repeat testing is performed on blood donations deemed to be high risk (Kirchhoff and Pearson, 2007).

Chagas disease has only two therapeutics available for treatment: Nifurtimox and Benznidazole (Bermudez et al., 2016). Problems exist with these treatments as they have high toxicity issues and require prolonged administration (Bermudez et al., 2016). Additionally, these treatments are only useful in the acute phase of infection and are labile in chronic instances of disease (Bermudez et al., 2016). Currently no vaccine is yet available for the prevention of Chagas disease (Carrea and Diambra, 2016). Another approach to limiting the spread of Chagas disease is through vector control. Vector control is employed throughout Latin America and this is currently the most effective prevention strategy (Schofield and Dias, 1991). However secondary transmission, emergence of insecticide resistant vectors and transmission through infected food limit the effectiveness of this strategy (Sosa-Estani and Segura, 2015). Since the description of Chagas disease in 1909 extensive research has been carried out to identify the biological mechanisms of the parasites infection and survival in the host (Coura and Borges-Pereira, 2010).

## **1.42 Antibiotics**

Antibiotics represent one of the most crucial therapeutics in modern medicine (Aminov, 2009). They offer protection from infectious bacteria and are in use every day to limit septicaemia from invasive surgery to chemotherapy (Silvani, 1947, DeVita and Chu, 2008). The use of antibiotics has reduced childhood mortality and increased our life expectancy (Cohen, 2000, Ahmad et al., 2000). The MMOA of antibiotics includes inhibition of: bacterial cell wall synthesis, fatty acid synthesis, DNA replication and protein synthesis (Neu, 1992). For example, the antibiotic vancomycin binds to D-Ala-D-Ala peptides preventing peptidoglycan crosslinking thereby inhibiting cell wall synthesis (Walsh, 1999). The  $\beta$ -lactam antibiotics bind to penicillin-binding proteins (PBP), which is essential in peptidoglycan synthesis and a further example of a cell wall inhibitor (Spratt, 1983). Triclosan biotics target the enoyl-ACP reductase enzyme responsible for fatty acid synthesis used in cell function and growth (Heath et al., 1999). Bacteriostatic drugs target the 30S (tetracyclins and aminocyclitols) and 50S (macrolides and chloramphenicols) ribosome subunits inhibiting ribosome function (Neu, 1992). The quinolone antibiotics target DNA replication through binding with DNA gyrase complexed with DNA, promoting double-strand breaks and cell death (Kohanski et al., 2007). Antibiotics work by a variety of MMOA (Figure 16), but despite a plethora of antibiotics available multi-drug resistant strains of bacteria are emerging (Hashemi et al., 2013). Complacency in the use of antibiotics (overuse and inappropriate prescribing) has led to the rise in bacterial resistance (Ventola, 2015). Antibiotic use within livestock production contributes to this problem (Phillips et al., 2004). Global average consumption of antimicrobials in 2010 is estimated to be 45 mg/kg for cattle, 145 mg/kg for chickens and 172mg/kg for pigs. This could potentially rise 65% by 2030 due to an increase in consumer demands and shift to large-scale farming (Van Boeckel et al., 2015). A significant fraction of the antibiotics used in livestock consumption are antibiotics that are in clinical use and medically important for human health (Van Boeckel et al., 2015). Significant volumes are used prophylactically for healthy animals to promote growth and stop the development of infection (Angulo et al., 2005). Prolonged exposure of bacteria to antibiotics represents a risk of cultivating drug resistance through natural selection (Davies and Davies, 2010). Routes of drug-resistant bacterial transmission from animals can occur through: human consumption of contaminated meat, direct contact from animals to humans (i.e. farmers) and environmental contamination and exposure through animal faeces used for fertiliser (Van Boeckel et al., 2015). An extensive literature review of 280 published, peer-reviewed research papers observed that 72% of these research papers established a link between antibiotic consumption in animals and antibiotic resistance in humans (Resistance, 2015). Multi-drug resistance to antibiotics is a global health concern and is listed in the most recent World Economic Forum Global Risks report as one of the greatest threats to human health (Forum, 2016). Bacterial

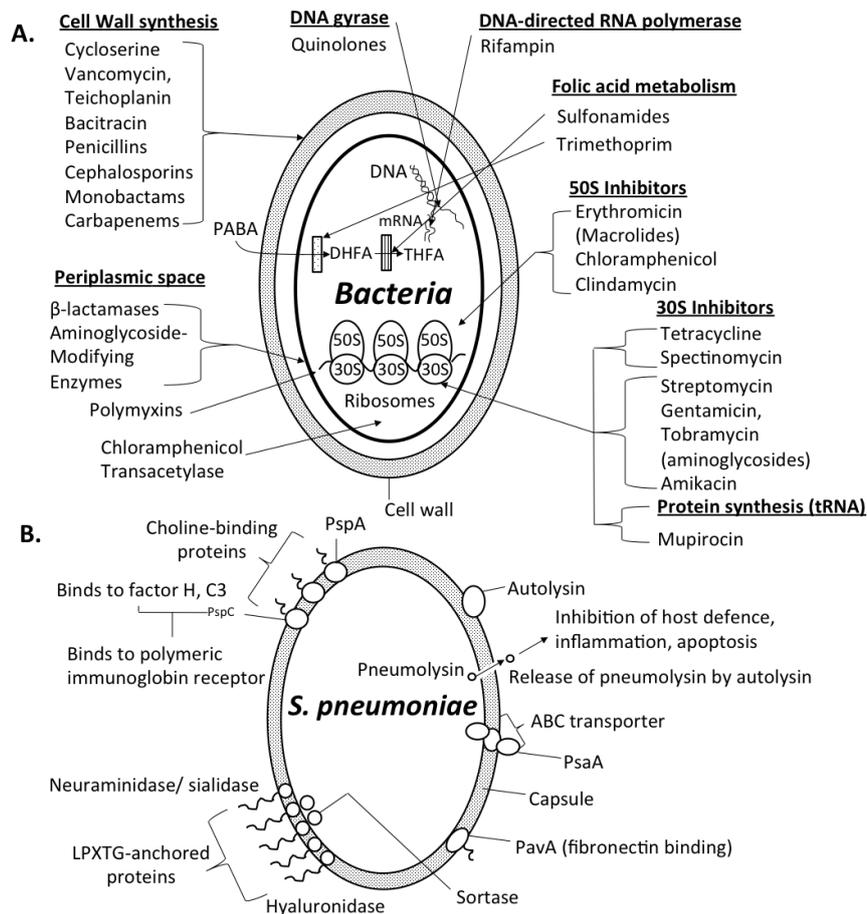
resistance to antibiotics costs the European Union economy \$1.5 billion annually (WHO, 2015). Within Europe, 25,000 mortalities occur annually as a result of antibiotic resistance and in the U.S. the corresponding figure is 23,000 mortalities annually (WHO, 2015).

Resistance is already found intrinsically within bacteria, as certain bacterial strains are less susceptible to certain types of antibiotic (Martínez, 2008). Additionally, resistance can be acquired through chromosomal changes (chromosomal mutation or inductive expression of a latent chromosomal gene) or exchange of genetic material using plasmids and transposons (transformation, transduction or conjugation) (Davies and Davies, 2010). More than 20,000 potential antibiotic resistance genes are in existence (Davies and Davies, 2010). Transposons have the ability to enter transmissible plasmids or chromosomes and contribute to the inter-genus transfer of resistance (Davies and Davies, 2010). Gram-positive bacteria can pass resistance to Gram-negative bacteria by inter-genus transmission (Gram-negative to Gram-positive transmission however is uncommon) (Davies and Davies, 2010). Resistance to antibiotics occurs by three main mechanisms: 1) prevention of access to the target, 2) inactivation of the antibiotic *via* destruction or modification or 3) alteration of the target site (Neu, 1992). Resistance against  $\beta$ -lactam antibiotics can occur by alteration of PBP, reduced permeability and  $\beta$ -lactamase activity (Neu, 1992). Mechanisms to limit access of the antibiotic to the target include: reducing porins expression to limit the diffusion of hydrophilic antibiotics through transmembrane proteins and increasing the efflux of antibiotics through expression of bacterial efflux pumps (Neu, 1992). Alteration of the target occurs through changes in the target structure limiting efficient antibiotic binding or acquisition of a homologous gene to the original target (Neu, 1992). An example of target alteration is found within methicillin-resistant *S. aureus* (MRSA) in which resistance to methicillin is generated through the acquisition of staphylococcal cassette chromosome mec (SCCmec) (Blair et al., 2015). This relays the *mecA* gene that encodes PBP2a (a  $\beta$ -lactam insensitive protein) enabling cell wall synthesis in the presence of a PBP inhibiting antibiotic (Blair et al., 2015). The effectiveness of the antibiotic and its clock relies heavily on its appropriate use and management within human and veterinary health. The antibiotic clock is the length of time an antibiotic will be useful before clinically significant resistance emerges (Clardy et al., 2006). Resistance generated in microbes to clinically used antibiotics is inevitable and so a continuing need for new biotics is required.

### 1.43 *Streptococcus pneumoniae*

The Gram-positive bacteria *Streptococcus pneumoniae* (*S. pneumoniae*) is a member of the commensal microflora present on human mucosal surfaces (Patterson, 1996). Also known as pneumococcus or Streptococci, it is a non-spore forming, non-motile,  $\alpha$ -hemolytic, catalase-negative facultative anaerobe (Patterson, 1996). Pneumococci have a lancet-shaped morphology encased by a polysaccharide capsule. This bacteria colonises the mucosal surfaces of the host's upper airway and nasopharynx and is often carried asymptotically in the nasopharynx (approximately 60% of the population carry pneumococcus asymptotically) (Kadioglu et al., 2008). To establish and cause infection the pathogen will need to compete with normal flora present on the mucosal surface, invade host cells, avoid the host immune system and replicate (Janeway Jr et al., 2001). Pneumococci cause a range of serious infections: meningitis (in children and adults), sinusitis, otitis media and community acquired pneumonia (Mitchell, 2003). Host factors (reduced immune function or lack of pre-existing antibodies against Streptococci) and/or bacterial factors contribute to the development of pneumococcal disease (Kadioglu et al., 2008). The polysaccharide capsule is a key bacterial factor, which protects against phagocytosis. Other factors that contribute to pathogenesis are: components of the cell wall, haemolysin (pneumolysin) and cell-surface proteins (LPXTG-anchored protein, choline-binding proteins, ABC transporter, ATP-binding cassette, fibronectin-binding protein and laminin-binding protein) (Mitchell, 2003). The cell wall components are important in mediating cell attachment to the host, an example of which is phosphorylcholine (within the cell wall) (Tuomanen et al., 1985). This molecule binds to platelet-activating factor (PAF) present on the host cells. PAF upregulation is an inflammatory response to infection or disease and could explain the increased occurrence of pneumonia post viral infection or at any stage of chronic obstructive pulmonary disease (Tuomanen et al., 1985). Pneumolysin is a pore-forming protein and is secreted when autolysin is host cell-bound (Mitchell, 2003). Pneumolysin is toxic to pulmonary, epithelial cells and causes damage to ependymal cilia of the brain inducing apoptosis (Spreer et al., 2003, Mitchell, 2003). As well as forming pores, pneumolysin stimulates the production of inflammatory mediators including: TNF- $\alpha$ , IL-1 $\beta$ , nitric oxide, IL-8, prostaglandins and leukotrienes (Mitchell, 2003). It also interferes with the complement pathway and inhibits phagocyte and lymphocyte function (Mitchell, 2003). Isogenic mutants of pneumococci that do not express pneumolysin or autolysin caused no or very mild pneumococcal disease when infected into the cerebrospinal fluid (CSF) of rats (Hirst et al., 2008). In contrast, wild type pneumococci infected into rat CSF caused meningitis within 26hrs (Hirst et al., 2008). The roles of autolysin and pneumolysin are essential for pneumococcal pathogenicity (Hirst et al., 2008). Cell surface proteins consist of a broad range of protein types (Maestro and Sanz, 2016). The choline binding proteins known to be

important virulence factors for pneumococci are anchored non-covalently to the cell wall *via* a repeat region of the C-terminal end of the protein (Maestro and Sanz, 2016). This protein class includes: pneumococcal surface protein A (PspA), pneumococcal surface protein C (PspC) and autolysin LytA (Paton, 1998). PspA is expressed by all clinically important serotypes of *S.pneumoniae* (Jedrzejewski, 2001, Crain et al., 1990). Mutant pneumococci unable to express PspA tested in models of systemic disease are less virulent and easily cleared from the bloodstream (Tu et al., 1999). This choline binding protein inhibits complement activation and complement receptor mediated clearance (Jedrzejewski, 2001). The MMOA of PspA on the complement pathway is unknown although it is known to interfere with C3b deposition and can inhibit the formation of alternative pathway C3 convertase (Tu et al., 1999) (Figure 16). The MMOA of a number of different virulence factors for *S.pneumoniae* has not yet been established and the use of chemical biology can aid our understanding (Blair et al., 2015).



**Figure 16. A.** Schematic of MMOA of various antimicrobial agents (messenger ribonucleic acid (mRNA), p-aminobenzoic acid (PABA), dihydrofolic acid (DHFA), tetrahydrofolic acid (THFA)). Adapted from Neu et al. **B.** Schematic of *S.pneumoniae* and the identified virulence factors (adapted from Mitchell., 2003).

Transmission of *S.pneumoniae* occurs through aerosolised droplets (person to person transmission of pneumococci in a cough or a sneeze) and exacerbation occurs through aspiration of pneumococci into the lower respiratory tract (CDC, 2016, Kumar, 2012). Each year in the U.S., pneumococcal disease is responsible for 445,000 hospitalisations, 774,000 emergency department visits and 22,000 mortalities (Huang et al., 2011). This amounts to \$3.5 billion in direct medical costs (Huang et al., 2011). Currently, resistance of *S.pneumoniae* to antibiotics means penicillin is an ineffective antibiotic against pneumococcal meningitis. Previously in the 1940s, penicillin treatment would result in a cure from pneumococcal meningitis infection (Neu, 1992). Resistance of *S.pneumoniae* was first observed in the 1960s and since then *S.pneumoniae* antibiotic resistance has become a worldwide problem (Quagliarello and Scheld, 1997, Neu, 1992, Blair et al., 2015).

### ***1.50 The sialidases and sialic acid as potential drug targets***

Sialic acids (Sia) and sialidases (see description of both below) are involved in many cellular functions (Varki and Schauer, 2009). Disturbances in the biosynthesis or degradation of Sia can lead to disease (Varki and Schauer, 2009). Hereditary Sia disorders are an indication of the importance of Sia in human health (Varki and Schauer, 2009). Defects in the lysosomal Sia transporter Sialin causes Salla disease (Prolo et al., 2009). A missense mutation in the UDP-GlcNAc-2-epimerase/ManNAc kinase (GNE) gene causes hereditary inclusion body myopathy and a defect of GNE feedback inhibition by CMP-Neu5Ac causes sialuria (Weiss et al., 1989). Sia distribution and *loci* are observed to have a role in the exacerbation of cancer, atherosclerosis, diabetes and Alzheimer's disease (Varki and Varki, 2007, Dall'Olio and Chiricolo, 2001, Varki and Schauer, 2009, Lindberg et al., 1991). The external position of Sia on outer cell membranes implies it has a strong role in cell biology (Schauer, 2004). Consequently, this positioning of Sia makes them susceptible to action of enzymes such as esterases, sialidases and lyases (Varki and Schauer, 2009). This enzymatic action affects the amount of Sia present resulting in actions including: the stabilisation of membranes, masking of antigenic sites and immunological functions (Varki and Schauer, 2009). Pathogens use lectins, sialidases or trans-sialidases as virulence factors and many bacterial toxins and viruses bind to sialylated glycoconjugates (Taylor, 1996). Sia and the enzymes that regulate its distribution and concentration represent a potential route to target validation and the development of new therapeutics.

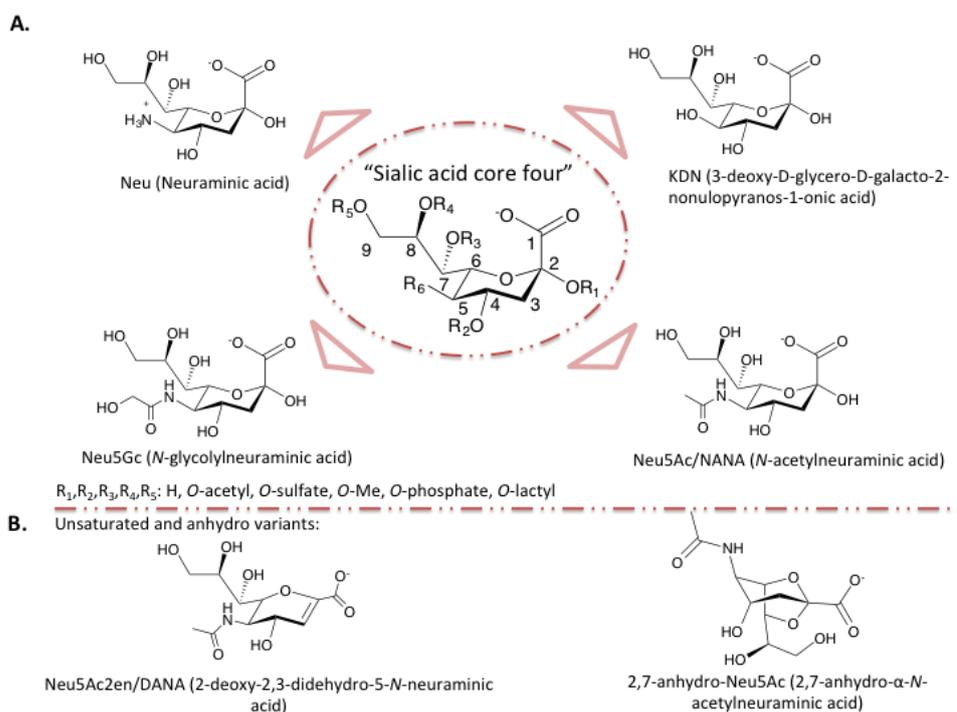
### ***1.51 Sia***

Sia are a diverse family of approximately 50 related negatively charged sugars usually present as terminal residues on cell surface glycoconjugates (Buschiazzi and Alzari, 2008a). Sialic acids have ubiquitous distribution and modulate a wide variety of physiological and pathological processes (Varki, 2008). Sia present on cell surfaces can mask underlying recognition sites and/or can act as

recognition sites (Varki and Gagneux, 2012). The physicochemical features of sialic acid include a net negative charge and hydrophilicity forming important components of binding sites enabling structural or modulatory roles for various pathogens or toxins (Varki, 2008). Typically, a cell exhibits tens of millions of Sia molecules on its surface with a local concentration approaching 100 mM (Varki and Gagneux, 2012). At blood physiological pH the carboxylic acid in Sia is deprotonated (Vimr et al., 2004). This negative charge affords charge repulsion of Sia coated erythrocytes avoiding unwanted cell-to-cell interactions in blood circulation (Varki and Gagneux, 2012). Apart from attracting or repulsing cells or molecules they are involved in the binding and transport of positively charged molecules (Lundblad, 2015). Sia coating on mucins contributes to its high viscosity and water solubility (Schauer, 2000). The anti-proteolytic effect of Sia in glycoproteins is also attributed to this negative charge (Schauer, 2000).

### ***1.52 The core structure of Sia***

The core structure of sialic acid consists of a nine-carbon polyhydroxylated  $\alpha$ -keto acid (Buschiazzo and Alzari, 2008b). Substitution at the fifth carbon generates sialic acid species including: *N*-acetylneuraminic acid (Neu5Ac), *N*-glycolylneuraminic acid (Neu5Gc) and 3-deoxy-D-glycero-D-galacto-2-nonulopyranos-1-onic acid (KDN) (Varki, 1992). A less common analogue is neuraminic acid (Neu) a non-*N*-acylated variant of Neu5Ac (Varki and Schauer, 2009). The hydroxyl groups on these four “core” Sia molecules (Neu, Neu5Gc, Neu5Ac and KDN) can be further substituted one or more times at the C4, C7, C8 or C9 position resulting in O-acetyl, O-sulfate, O-phosphate, O-lactyl or O-methyl moieties (Varki and Schauer, 2009). The C1 carboxylate, normally ionised under physiological conditions, can be condensed with free hydroxyl groups of adjacent saccharides resulting in lactones or with free amino groups resulting in lactams (Varki and Schauer, 2009). Unsaturated and anhydro variants of Sia can exist, examples include: 2-deoxy-2,3-didehydro-5-*N*-neuraminic acid (Neu5Ac2en/DANA) and 2,7-anhydro- $\alpha$ -*N*-acetylneuraminic acid (2,7-anhydro-Neu5Ac) (Varki and Schauer, 2009). The two main types of Sia found in mammals are Neu5Ac and Neu5Gc (Figure 17)(Varki and Schauer, 2009). The most abundant of these is Neu5Ac, which is the biosynthetic precursor to all other members of the Sia family (Varki and Schauer, 2009).



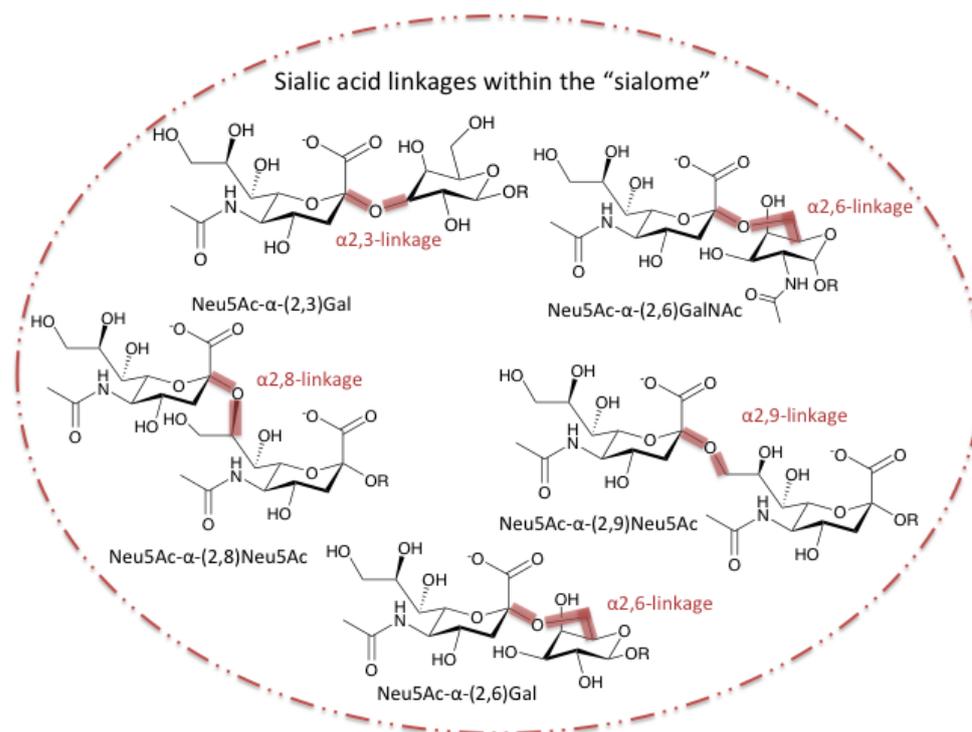
**Figure 17. A.** Chemical structures of the “four core sialic acid” molecules. **B.** Unsaturated and anhydro chemical structures of sialic acid.

### 1.53 Sia linkage and presentation within the “Sialome”

The “sialome” represents a subclass of the glycome focusing on particular Sia type, linkage and presentation on macromolecular structures (Cohen and Varki, 2010). The glycome denotes a vastly diverse set of monosaccharides and polysaccharides present as a repertoire within and on the cell, tissue and organism (Varki and Sharon, 2009). Further variation exists between saccharides through  $\alpha$  or  $\beta$  linkages at any of the several linking positions on each core saccharide unit (Schauer, 2000). Within the “Sialome”, Sia is typically found terminally linked on branches of *O*-glycans, *N*-glycans and glycosphingolipids (gangliosides) (Varki and Schauer, 2009). The more common linkages include: *N*-acetyl, *N*-glycolyl and *O*-acetyl (Stencel-Baerenwald et al., 2014). The less common linkages found are: *O*-lactyl, *O*-methyl and *O*-sulfate groups (Schauer, 2000). Sia sometimes contribute to branches on underlying glycans, for example Sia forms side chains on glycosylphosphatidylinositol (GPI) anchors (Cohen and Varki, 2010). Further complexity exists in Sia-to-sia linkages (Varki and Schauer, 2009). These linkages occur at the C2 position, referred to as  $\alpha$ -linkages, and result in different underlying sugars (Figure 18). The most common Sia  $\alpha$ -linkages observed are: the  $\alpha$ -2,3 and  $\alpha$ -2,6 to galactose (Gal) or the  $\alpha$ 2,6 to *N*-acetylgalactosamine (GalNAc) (Varki and Schauer, 2009). Other linkages observed include: the  $\alpha$ -2,8 and  $\alpha$ -2,9 linkages between Sias and the  $\alpha$ -2,4 linkage in repeating units of echinodermal glycans (Cohen and Varki, 2010). Sia-to-sia linkages form further

branched structures upon the underlying sugars/glycans (Cohen and Varki, 2010). These underlying glycans can be split into various classes. The *N*-Glycans are attached *via* an *N*-glycosidic covalent bond to asparagine residues in “sequons” on glycoproteins (Stanley et al., 2009). These “sequons” are a minimal amino acid sequence (Asn-X-Ser/Thr) to which an *N*-glycosidic bond can occur (X corresponds to any amino acid except proline) (Breitling and Aebi, 2013). All *N*-glycans contain a common core of Man $\alpha$ -1-6(Man $\alpha$ -1-3)Man $\beta$ -1-4GlcNAc $\beta$ -1-4GlcNAc $\beta$ -1-Asn-X-Ser/Thr (Stanley et al., 2009).

The *O*-glycans represent another class, archetypally consisting of GalNAc units covalently linked through an *O*-glycosidic bond to serine or threonine residues of mammalian glycoproteins (Bennett et al., 2012). Mucins are *O*-glycosylated glycoproteins expressed by various epithelial cell types either as mucous secretions or transmembrane glycoproteins (Brockhausen et al., 2009). In mucous secretions, mucins can be small monomeric (soluble mucins) or large polymeric (gel-forming mucins) complexes (Brockhausen et al., 2009). Mucins protect epithelial cells by the prevention of infection, dehydration, chemical and physical stress (Brockhausen et al., 2009). Mucins also facilitate the passage of materials through the respiratory, digestive or urogenital tracts (Behera et al., 2015). Other glycan classes include the glycosphingolipids and are predominantly found in the outer leaflet of the plasma membrane (Schnaar et al., 2009). Glycosphingolipids mediate cell signalling, act as microbial toxin receptors and are present in cell surface antigens (Schnaar et al., 2009). These particular glycans have an ability to form clusters through (*cis*) hydrogen bond interactions between OH and *N*-Ac functional groups resulting in cluster arrangements on cellular membranes (Todeschini and Hakomori, 2008a, Hakomori, 2004). The spatial organisation of glycans into clusters results in sialylated microdomains (Schnaar et al., 2014). Two examples of sialylated microdomains are the “saccharide patch” and the “glycosynapse” domains (Cohen and Varki, 2014). Presentation, location and density of Sia are important for macromolecular binding and cell signalling (Cohen and Varki, 2014). Individuals suffering from Miller Fisher syndrome (a rare acquired nerve disease resulting in paralysis and abnormal muscle coordination (Horton Jr et al., 2011)) generate antibodies that do not bind to single isolated gangliosides but bind specifically to ganglioside complexes: GQ1b/GM1 and GQ1b/GD1a (Kaida et al., 2006). The erythrocyte binding antigen generated by *Plasmodium falciparum* binds specifically to *O*-linked Neu5Ac $\alpha$ 2-3Gal tri- and tetra-saccharides on glycophorin A. Glycosynapse domains are known signalling mediators of cell growth, motility and adhesion (Todeschini and Hakomori, 2008b).



**Figure 18.** Chemical structures of the various sialic acid linkages found within the “sialome”.

Multicellular sialic-acid-deficient organisms that express sialic acid binding proteins include: plants, molluscs and arthropods (Cohen and Varki, 2010). Although important surface sugars in the lineage of animals Sia are found in other branches of life including viruses, bacteria and protozoa (Freire-de-Lima et al., 2015). These microorganisms exploit this commonality on host cells and use proteins and enzymes that bind or act on sialic acid for invasion, evasion and survival (Taylor, 1996). The sialidases are a family of proteins that cleave Sia (Traving and Schauer, 1998). Viral and microbial pathogens acquire these enzymes for nutrition and pathogenesis (Traving and Schauer, 1998).

### **1.54 Sialidases (EC 3.2.1.18)**

Sialidases (also known as neuraminidases or sialyl hydrolases) were first discovered in the 1940s (Varki and Schauer, 2009). They catalyse the removal of Sia within the “sialome” and can be split into two categories: exosialidases and endosialidases (Cohen and Varki, 2010, Chan and Bennet, 2012). This classification is dependent on the position of hydrolysis, as exosialidases remove terminal Sia and endosialidases hydrolyse internal glycosidic bonds within oligomers and polymers (Watts et al., 2006). Sialidases are widely distributed in vertebrates and microorganisms (viruses, bacteria, fungi, mycoplasma and protozoa) (Warren, 1963). Based upon amino acid sequence homology the sialidase superfamily has been classified into three glycoside hydrolase (GH) families: GH-33 (including most simple eukaryotic and bacterial sialidases as well as Trans-sialidases), GH-43

(includes viral sialidases) and GH-83 (includes hemagglutinin-neuraminidases) (CAZy, Lombard et al., 2014). An additional fourth classification for sialidases (GH-58) includes endosialidases present in bacteriophages (Kim et al., 2011). Despite their shared structural features (see 1.55 Conservation), the linkage and substrate specificities of the sialidases can vary (Kim et al., 2011).

### **1.55 Glycoside Hydrolase family (GH-33)**

Trans-sialidases, bacterial sialidases and eukaryotic sialidases make up this class of GH (Lombard et al., 2014, CAZy). Mammalian sialidases have been identified to modulate cell adhesion, proliferation/differentiation, metabolism and immunological functions (Fanzani et al., 2012). In vertebrates four sialidases have been characterised based upon subcellular location (Miyagi and Yamaguchi, 2012). These include an intra lysosomal neuraminidase 1 (Neu1), cytosolic neuraminidase 2 (Neu2), plasma membrane neuraminidase 3 (Neu3) and lysosomal or mitochondrial membrane neuraminidase 4 (Neu4) (Miyagi and Yamaguchi, 2012). The human neuraminidase enzymes (hNEU) are exosialidases cleaving terminal sialic acid linkages (Fanzani et al., 2012). Neu1 participates in lysosomal catabolism of sialoglycoconjugates and cellular immune response (Pshezhetsky and Hinek, 2011). Abnormalities in Neu1 activity results in the progressive lysosomal storage of oligosaccharides and sialylated glycopeptides (Fanzani et al., 2012). This causes Sialidosis, an autosomal recessive disease, characterised by firstly milder symptoms including visual defects, myoclonus syndrome, cherry-red macular spots, ataxia, hyperreflexia and seizures (Seyrantepe et al., 2003, Lowden and O'Brien, 1979). More severe symptoms can develop later that include: mental retardation, dysostosis multiplex, Hurler-like phenotype and hepatosplenomegaly (Seyrantepe et al., 2003). Neu2 hydrolyses GM3 gangliosides participating in myotube formation and regulation. Localisation of this enzyme is found at a low level within skeletal muscle, brain and liver cells (Chavas et al., 2005). Neu3 hydrolyses gangliosides involved in signal transduction including GM1 and GD1a (Seyrantepe et al., 2004). Neu3 has been implicated to play a role in signal transduction, cell proliferation and cell differentiation (Seyrantepe et al., 2003). Neu4 has broad specificity and participates in glycolipid catabolism (Yamaguchi et al., 2005). It has also been implicated in regulation of GD3 and mitochondrial apoptosis (Yamaguchi et al., 2005).

### **1.56 Bacterial Sialidases**

Bacteria use sialidases as virulence factors to recognise sialic acids exposed on host cell surfaces (Severi et al., 2007). Bacteria also use sialidases to scavenge sialic acid as a nutrient source for growth (Vimr et al., 2004). Sialidases have been implicated in many bacterial infections for example: septicaemia (*Pneumococcus*, *Streptococcus*, *Bacteroides*, *Corynebacterium*), gas gangrene (*Clostridium*), pneumonia (*Streptococcus*), peritonitis (*Clostridium*, *Bacteroides*), cholera (*Vibrio*

*cholerae*) and meningitis (*Streptococcus B*) (Gaskell et al., 1995). Most bacterial sialidases are secretory proteins containing signalling peptides cleaved upon secretion (Kim et al., 2011). These monomeric bacterial sialidases range in molecular weight from 40 to 150 kDa with several bacteria producing isoenzymes/multiple sialidases (Gaskell et al., 1995). The bacteria producing multiple sialidases include *Arthrobacter ureafaciens*, *Clostridium perfringens*, *Pasteurella multocida*, *S. pneumoniae* and *Arthrobacter nicotianae* (Kim et al., 2011). The regioselectivity and affinity of bacterial sialidases to various sialoglycoconjugate substrates varies (Kim et al., 2011). For example, *Corynebacterium diphtheriae* and *M. viridifaciens* sialidases show a specificity for  $\alpha(2,6)$  over  $\alpha(2,3)$  (Kim et al., 2011). The sialidase isoforms, L, M1, M2 and S from *A. ureafaciens* show a specificity for both  $\alpha(2,6)$  and  $\alpha(2,3)$  over  $\alpha(2,8)$  (Kim et al., 2011). Substrate type can impact on activity as *V.cholerae* sialidase has a higher hydrolase activity of sialic acid linked glycolipids to other sialic acid linked glycoconjugates (Kim et al., 2011). Despite substrate and linkage specificity, most bacterial sialidases have an ability to hydrolyse a broad range of sialoglycoconjugates containing  $\alpha(2,3)$ ,  $\alpha(2,6)$  and  $\alpha(2,8)$ -linked sialic acids (Juge et al., 2016, Nicholls et al., 2013). To this date, 11 structures of bacterial sialidases have been solved and published on the PDB (RCSB, Berman et al., 2000).

### **1.56.1 *Streptococcus pneumoniae* sialidases**

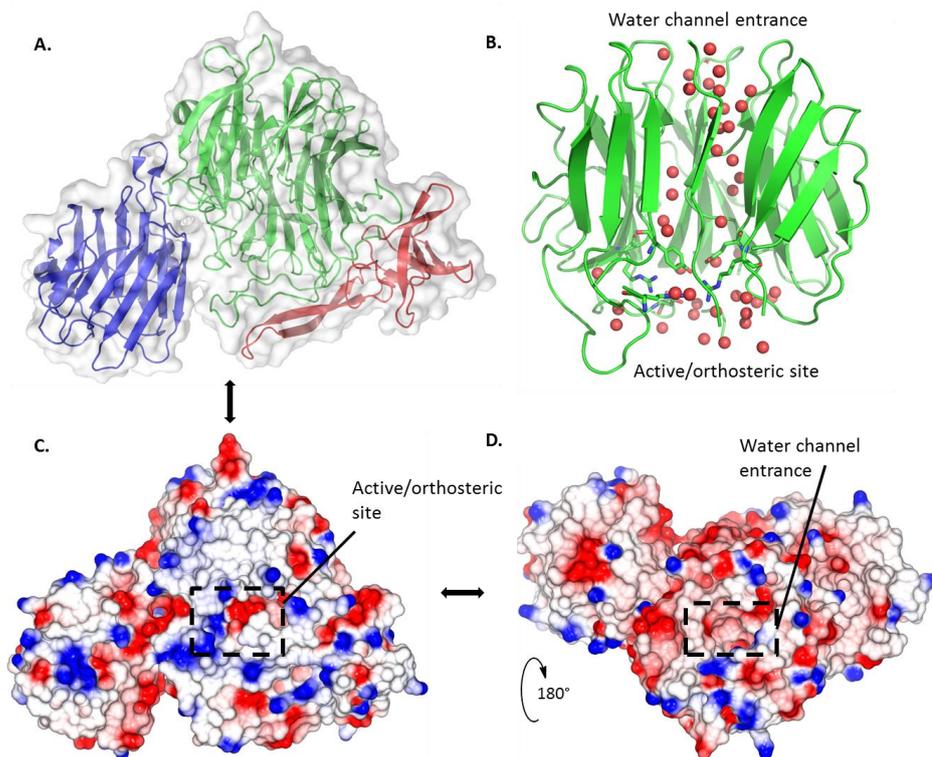
The genome of *S. pneumoniae* encodes three sialidases: NanA, NanB and NanC (Xu et al., 2011). Screening 342 clinical pneumococcal isolates for NanA, NanB and NanC genes found each to be present in 100%, 96% and 51% of these strains respectively (Pettigrew et al., 2006). Colonization of the upper respiratory tract during *S. pneumoniae* infection is promoted by the removal of sialic acid from the host's cell surface glycans (Brittan et al., 2012). NanA is a cell surface anchored sialidase (115 kDa) with broad substrate specificity towards  $\alpha(2,3)$ ,  $\alpha(2,6)$  and  $\alpha(2,8)$  linkages (Xu et al., 2008). Upregulation of NanA occurs in the presence of free sialic acid and by interaction of pneumococci with host cells (Hentrich et al., Parker et al., 2009). Bacterial survival is aided by NanA activity, as sialic acid hydrolysis provides a carbon source for *S. pneumoniae* (Marion et al., 2011, Manco et al., 2006). NanA also aids pathogenesis through surface modification of other bacterial surfaces, host defence function/modification and passage across the blood brain barrier (Chang et al., 2012). NanA activity has also been implicated in biofilm formation (Parker et al., 2009).

NanB (78 kDa) is an extracellular secreted sialidase (Xu et al., 2008). The structures of the sialidases provide an insight into the substrate selectivities of the sialidases (Xu et al., 2011). The structures of the catalytic NanA (Gut et al., 2011) and NanB (Xu et al., 2008) and NanC (Owen et al., 2015) have been solved within the Taylor group. The active site cavity of NanA is open and flat, allowing accommodation of the various types of sialic acid linkage (Xu et al., 2011). The NanB active

site cavity is narrow due to the presence of Trp647, sterically hindering  $\alpha(2,6)$  and  $\alpha(2,8)$  linkages (Xu et al., 2011). The active site cavity of NanC also features a Trp716 creating a narrow cleft for  $\alpha(2,3)$  specificity (Owen et al., 2015). Further structural differences of Phe396 and Trp581 are observed creating a more hydrophobic environment within the active site cavity (Owen et al., 2015). This difference enables NanC to convert  $\alpha(2,3)$ -sialyllactose into Neu5Ac2en/DANA, a general sialidase inhibitor (Owen et al., 2015). In the presence of a high concentration of DANA, NanC can perform the reverse reaction to form Neu5Ac (Owen et al., 2015). Although the structure and biochemical characterisation provide insight into its kinetic parameters, the function and role of NanC remains elusive and is the least studied of the three sialidases (Owen et al., 2015). NanA and NanB are well-studied sialidases due to their observed importance in pneumococcal infection.

### **1.56.2 NanB**

NanB is an intramolecular *trans*-sialidase that produces 2,7-anhydro-Neu5Ac from  $\alpha(2,3)$  sialoglyconjugates (Gut et al., 2008). Sequence homology with other pneumococcal sialidases is low (24% NanA and 46% NanC) (Gut et al., 2008). The pH optimum of NanB is also different to NanA at approximately pH 4.5 (NanA pH optima is approximately pH 6.5) (Berry et al., 1996). NanB exists as a monomer of 76kDa, consisting of three domains: a concanavalin-like lectin domain, a canonical  $\beta$ -propeller catalytic domain and an inserted third domain (Figure 19) (Xu et al., 2008). These structural features are conserved across sialidases (see conservation).



**Figure 19.** Structure of NanB (2VW2). **A.** Crystal structure of NanB consisting of three domains: catalytic site (green), carbohydrate binding domain (blue) and the inserted domain (red). **B.** Cartoon representation of the catalytic site and location of the water channel (waters in red). **C** and **D.** Electrostatic surface representation of NanB. Images created in PyMOL. Electrostatic surface representation created using CCP4MG.

NanB also plays an important role in pneumococcal infection. In 2006 *in vivo* experiments showed that NanB was an essential component in the colonisation of *Streptococcus pneumoniae* (Manco *et al.*, 2006). In this paper a NanB neuraminidase deficient mutant was introduced into mice by intranasal infection (Manco *et al.*, 2006). This mutant was unable to cause sepsis or persist in the blood for longer than 48hrs post infection due to rapid clearance (Manco *et al.*, 2006). In support of this, a study in 2012 on a NanB deficient pneumococci mutant D39 strain showed reduced total neuraminidase activity (52%) and decreased pneumococci adherence (50-80%) to four epithelial cell lines (A549, Hep-2, Detroit 562 and primary culture of nasopharyngeal cells) (Brittan *et al.*, 2012). Given the importance of NanB in the virulence of *Streptococcus pneumoniae* it is an identified target.

## 1.57 *Trans-sialidase*

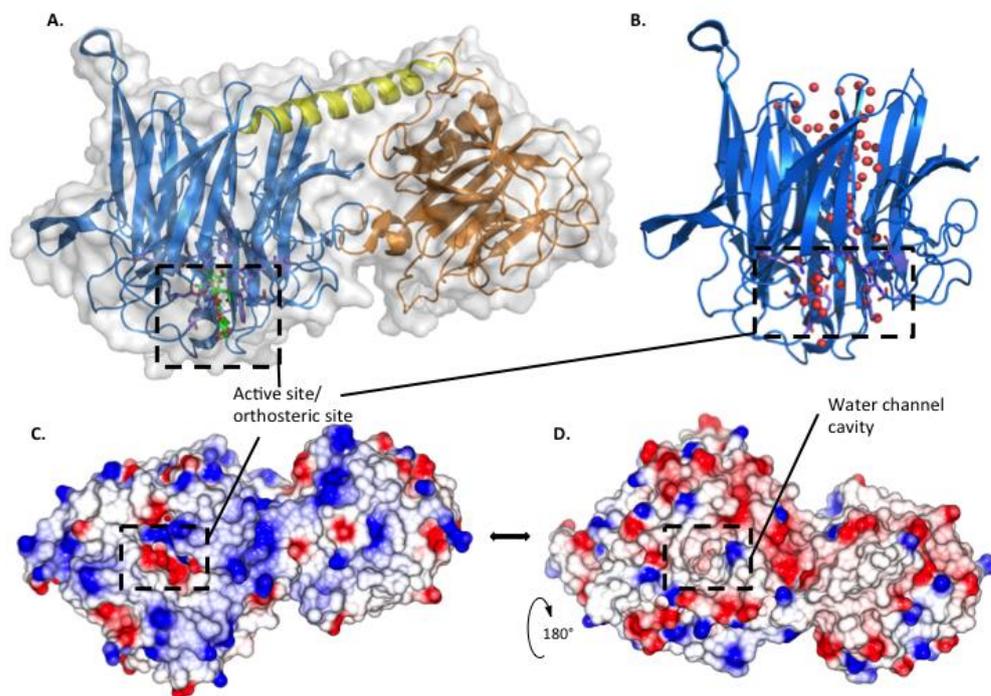
In contrast to sialidases that are strict hydrolases, trans-sialidases catalyse the transfer of sialic acids from sialoglycoconjugates to acceptor glycoconjugates (Juge et al., 2016). The trans-sialidases employ two catalytic functions; the first is hydrolysis of the donor substrate and second transfer of the donor to the acceptor. Trans-sialidases preferentially hydrolyse  $\alpha(2,3)$ -linked substrate donors (Tailford et al., 2015).

The trypanosomes *T. cruzi* and *T. brucei* contain sialidases with trans-sialidase activity (Montagna et al., 2002). In *T. brucei*, the trans-sialidase is expressed in the insect stages of its life cycle (Engstler et al., 1992). In *T. cruzi* the trans-sialidase is expressed in the mammalian stages of the life cycle (Schenkman and Eichinger, 1993, Taylor, 1996).

### 1.57.1 *TcTS Sialidase*

Sialidase activity was first reported in trypomastigotes of *T. cruzi* in 1983 (Pereira, 1983). This activity was found to be absent in amastigote forms of the parasite (Pereira, 1983). This activity was later associated with a membrane-anchored trans-sialidase known as the *T. cruzi* Trans-sialidase (TcTS) (Previato et al., 1985). *T. cruzi* cannot synthesise sialic acid and this trans-sialidase enables the parasite to evade the host immune response, preferentially catalysing the transfer of  $\alpha$ -2,3 linked sialic acid from host glycoconjugates to the terminal  $\beta$ -galactopyranosyl residues of mucin-like molecules on the parasite's cell surface (Freire-de-Lima et al., 2015). Donors are  $\beta$ -Gal residues linked with  $\alpha(2,3)$  sialic acid and not  $\alpha(2,6)$  or  $\alpha(2,8)$  (Vandekerckhove et al., 1992). The acceptors are disaccharides, oligosaccharides terminally linked with  $\beta$ -Gal (Vandekerckhove et al., 1992). The second step in the transfer is more efficient than the hydrolysis (Buschiazzo et al., 2000). The TcTS structure was solved by Buschiazzo *et al.*, 2002 (Figure 20). This structure provided insight into the reaction mechanism (Buschiazzo et al., 2002). TcTS has an *N*-terminal catalytic domain consisting of a canonical  $\beta$ -propeller fold (Buschiazzo et al., 2002). Connected by an  $\alpha$ -helix to the catalytic domain is a C-terminal lectin/CBM domain (Buschiazzo et al., 2002). Further structural evaluation with Michaelis complexes consisting of a point mutation at Asp59 to Ala resulted in structural complexes with uncleaved substrate bound (Amaya et al., 2004). Two key residues within the TcTS structure are Trp312 and Tyr119 and these form important stacking interactions with the aglycon moieties (Amaya et al., 2004, Buschiazzo et al., 2002). The aglycon has few protein hydrogen bond interactions but abundant water hydrogen bond interactions (Amaya et al., 2004). The greater affinity of TcTS to lactose than Sia corresponds to a hydrophobic driven binding affinity presumably from the Trp312 and Tyr119 stacking interaction (Amaya et al., 2004). Investigation of the mechanism *via* a transition state trapping intermediate, 2,3-difluoro-N-acetylneuraminic acid (2,3F-

Neu5Ac), identified Tyr342 as the catalytic nucleophile (Watts et al., 2003). The stereochemistry is retained *via* a ping pong mechanism (see active site mechanism) (Demir and Roitberg, 2009).



**Figure 20.** Structure of TcTS (1MS9, 1SOI). **A.** Crystal structure of TcTS consisting of three domains: catalytic site (blue), lectin-like domain (orange) and the  $\alpha$ -helix (yellow). **B.** Cartoon representation of the catalytic site and location of the water channel (waters in red). **C** and **D.** Electrostatic surface representation of TcTS. Images created in PyMOL. Electrostatic surface representation created using CCP4MG.

TcTS aids host cell recognition and attachment through active site binding to aglycon and sialic acid moieties (Neres et al., 2008). Cleavage of the GPI anchor releases TcTS into the bloodstream (Neres et al., 2008). The circulating TcTS removes sialic acid from the platelet surface causing thrombocytopenia, a symptom of the acute phase of Chagas disease (Neres et al., 2008). TcTS represents a potential target for the treatment of Chagas disease and disease progression (Neres et al., 2008). Treatment of trypomastigote parasites with sialidase renders them susceptible to complement alternative pathway, uptake by macrophages and abolishes lectin induced cell agglutination (Colli, 1993). Treatment of mice with trans-sialidase prior to trypomastigote infection in an *in vivo* mouse study increased mortality rates of mice over a 25 day period (Chuenkova and Pereira, 1995). In support of this, genetic immunization of mice with a pool of genes encoding for trans-sialidase resulted in protection from a normally lethal challenge of *T. cruzi* (Fralish and Tarleton, 2003)

### **1.58 Glycoside Hydrolase family (GH-34)**

The orthomyxoviridae family contains Influenza (Webster et al., 1992). The influenza viruses are split into three types: A, B and C (CDC, 2014). Influenza A is coated with haemagglutinin, sialidase and M2 ion channel proteins (Chlanda et al., 2015). Influenza A is classed into subtypes based upon the two surface proteins hemagglutinin and neuraminidase (Sui et al., 2009). Influenza B however, is classified under its genetic lineage (Jumat et al., 2014). The Victoria lineage (B/Victoria/2/87-like) and Yamagata lineage (B/Yamagata/16/88-like) have been circulating worldwide since 1983 (Rota et al., 1990). Both influenza A and B are clinically relevant as they cause human respiratory infections (Biere et al., 2010). Influenza B is less of a disease burden than influenza A (Paul Glezen et al., 2013).

The influenza sialidases/neuraminidases are exosialidases comprising 11 different subtypes (Shtyrya et al., 2009). The influenza virus sialidase is comprised of approximately 470 amino acid residues (Shtyrya et al., 2009). Homology between various subtypes is approximately 50%, but strict conservation of residues occurs within the active site as a requirement for catalytic function (Arg118, Asp151, Arg152, Arg224, Glu276, Arg292, Arg371 and Tyr406 are conserved) (Shtyrya et al., 2009).

The influenza sialidases are known to perform two critical steps in viral pathogenicity. This includes facilitation of virion progeny release and mobility of virus in the respiratory tract (Palese et al., 1974, Yang et al., 2014). The influenza neuraminidase is a validated drug target and a number of inhibitors of this sialidase have made it to market (see sialidase inhibitors) (Air, 2012).

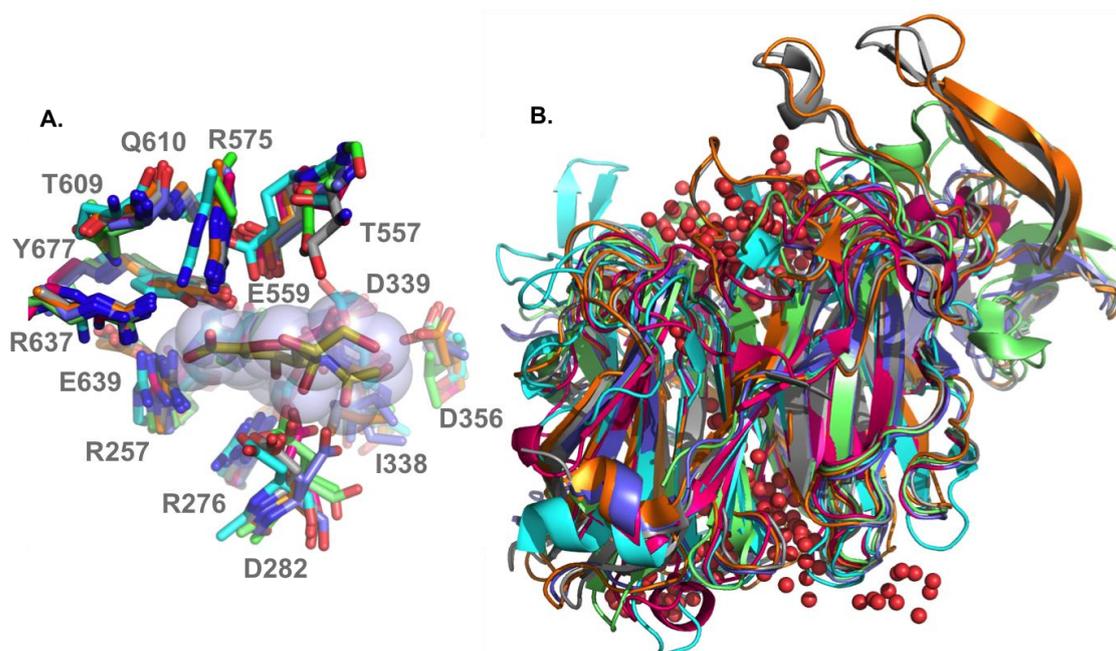
### **1.59 Conservation**

Despite low sequence homology across the four classes of sialidase, all sialidases have an overall fold similarity and a common catalytic domain (Figure 21 and Figure 25) (Buschiazzo and Alzari, 2008b).

#### **1.59.1 Active site**

Structural and sequence data of the sialidase superfamily have identified a number of conserved residues important for Sia recognition/binding and hydrolysis (Vimr et al., 2004, Vimr, 1994). The catalytic domain consists of a six-bladed  $\beta$  propeller topology with eight highly conserved residues in the active site (Gaskell et al., 1995). This domain also contains a water channel (Figure 21). The conserved residues responsible for Sia recognition and binding include an arginine triad (capable of electrostatic interactions directly with the carboxylate of Sia) (Gaskell et al., 1995). The presentation of a hydrophobic pocket made up of a multitude of amino acid residues accommodates the *N*-acetyl functionality of Sia (Owen et al., 2015). Both the arginine triad and the hydrophobic pocket are

responsible for the spatial orientation of Sia for catalytic hydrolysis (Owen et al., 2015). The glutamic and aspartic acid provide the acid catalyst for hydrolysis to occur (Newstead et al., 2005, Chan et al., 2011). The role of tyrosine in the hydrolysis is not so clear. The conserved tyrosine residue has two possible roles: 1) nucleophile attacking the anomeric carbon, or 2) as an electrostatic stabilising group for an oxonium ion intermediate (Vimr et al., 2004). Two proposed intermediates within sialidase hydrolysis of Sia have been suggested and the role of this tyrosine differs depending on the intermediate formed (Morley et al., 2009, Mendonça-Previato et al., 2010).

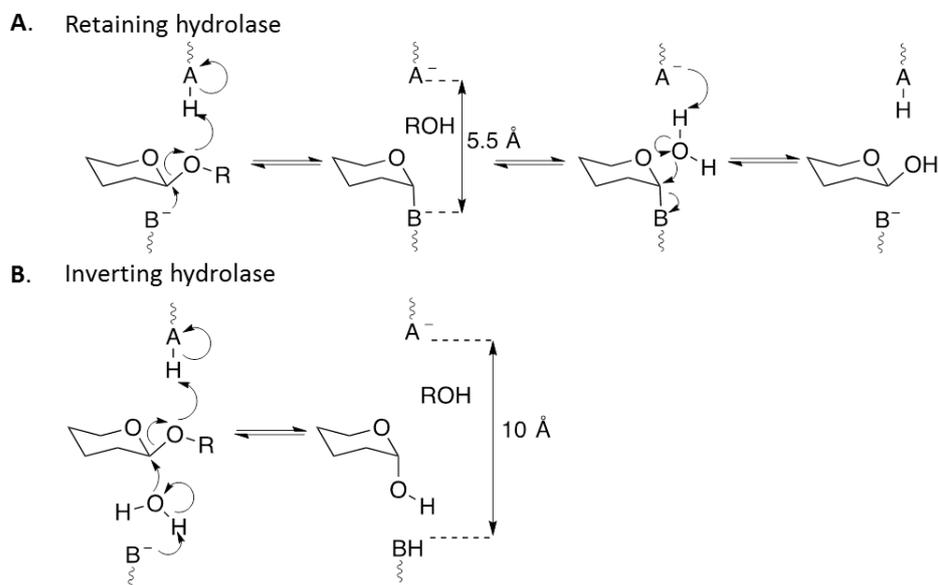


**Figure 21.** **A.** Active site overlay representation of sialidases (VCS (1KIT) in teal, NanI (2VK5) in violet, NanB (2VW0) in orange, NedA (1EUR) in light green, NanA (2W20) in dark green, BtSa (4BBW) in pink and RgNanH (4X47) in grey). DANA bound in the active site of RgNanH shown in yellow and van der Waals radius represented by spheres shown in light blue. Key residues of the RgNanH active site labelled in grey. **B.** Catalytic domain overlay representation of sialidases (VCS (1KIT) in teal, NanI (2VK5) in violet, NanB (2VW0) in orange, NedA (1EUR) in light green, NanA (2W20) in dark green, BtSa (4BBW) in pink and RgNanH (4X47) in grey). Structures aligned using CCP4MG Gesamt/SSM and figures created using PyMOL®.

### 1.59.2 Active site mechanism

The enzymatic hydrolysis of the glycosidic bond requires a proton donor/acid and a nucleophile/base (Davies and Henrissat, 1995). Within this mechanism stereochemistry can be preserved (retention) or the stereochemistry can be changed (inversion) (Davies and Henrissat, 1995, Koshland, 1953). The position of the base within this reaction is key for the retention or inversion of stereochemistry at the anomeric carbon (Figure 22). In retaining enzymes the base is

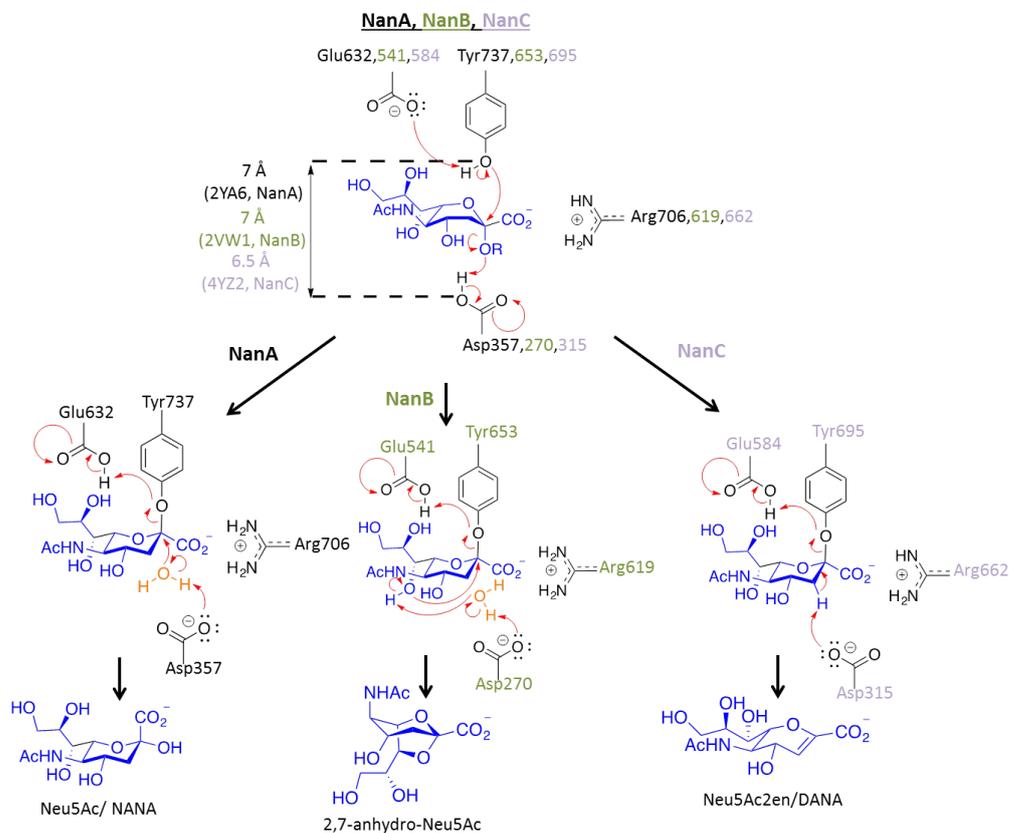
within a close proximity to the anomeric carbon (average distance 5.5 Å) whereas in inverting enzymes the base is more distant from the anomeric carbon allowing accommodation of a water molecule (average distance 10 Å) (Davies and Henrissat, 1995).



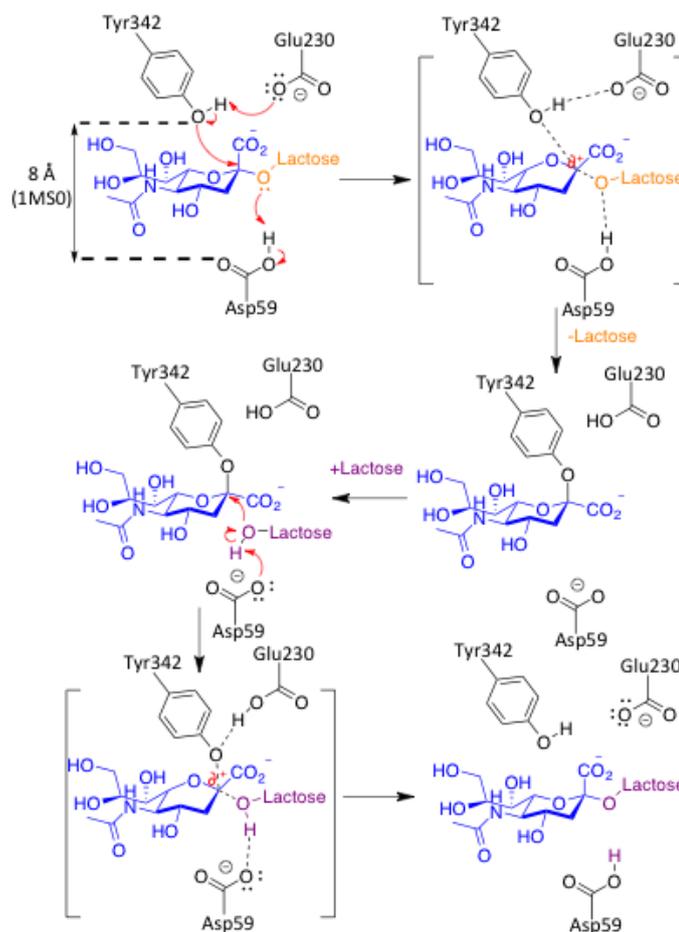
**Figure 22.** Glycosidic hydrolases reaction mechanism of **A.** retaining hydrolase reaction and **B.** inverting hydrolase reaction. Adapted from Davies and Henrissat., 1995. Schematic created using ChemDraw 12.0.

Most sialidases proceed with retention of the anomeric carbon. An example of an inverting sialidase is the bacteriophage K1F endo-sialidase (Morley et al., 2009). This sialidase lacks one of the two acid catalysts, one of the three arginines and the tyrosine nucleophile/base (Morley et al., 2009). As mentioned above two mechanisms for Sia hydrolysis have been postulated. The first mechanism proposed is through the formation of the half-chair oxocarbenium ion (Xu et al., 2011). The basic residue in this reaction mechanism is the deprotonated tyrosine and this acts to stabilise the oxonium ion formed in the transition state (Buschiazzo et al., 2002). The second is in the formation of a sialyl-enzyme intermediate with the tyrosine acting as a nucleophile (Chan et al., 2011). This was observed using a fluorinated Neu5Ac analogue (2,3-difluoro-N-acetylneuraminic acid (2,3F-Neu5Ac)). The position of the electron withdrawing fluorine on C-3 destabilises build-up of positive charge, stabilising and increasing the lifetime of the covalent intermediate (Vavricka et al., 2013). Following the formation of the substrate-covalent intermediate the reactions diverge forming different products (Figure 23) (Xu et al., 2011). In the hydrolytic sialidase,  $\alpha$ -Neu5Ac is formed from an activated water substitution at C-2 (Xu et al., 2011). In the trans-sialidase, sialyllactose is produced by the transfer of Sia to a new aglycon acceptor (Buschiazzo et al., 2002, Demir and

Roitberg, 2009). In the intramolecular sialidase, the nucleophile is the C-7 hydroxyl forming an intramolecular bond (Figure 23) (Xu et al., 2011).



**Figure 23.** Reaction mechanism of the *Streptococcus pneumoniae* sialidases (NanA, NanB and NanC). Adapted from Xu et al., 2011. Figure created using ChemDraw® and PowerPoint® 14.6.

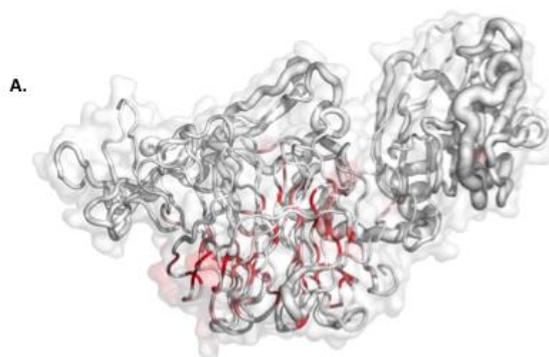


**Figure 24.** Reaction mechanism TcTS. TCTS trans-sialidase sialyl transfer with lactose donor and acceptor moieties. Hydrolysis reaction can occur and a water molecule takes the role of the acceptor lactose. Adapted from Demir and Roitberg, 2009. Figure created using ChemDraw® and PowerPoint® 14.6.

As hydrolysis of Sia by definition involves use of water for substitution, not all sialidases hydrolyse Sia (Owen et al., 2015). NanC cleaves Sia in the absence of water producing its primary reaction product Neu5Ac2en/DANA (Owen et al., 2015). The closer vicinity of Asp-315 (3.2Å) to the C-2 of substrate in NanC causes direct deprotonation of C-3 via Asp315 (Owen et al., 2015). NanC is observed to hydrolyse DANA (when in excess) into Neu5Ac2en showing mechanistic diversity (Owen et al., 2015).

### ***1.59.3 Water channel***

The conserved  $\beta$ -hairpin topology creates a centre pore in the catalytic domain (Figure 25). This central pore is comprised of a variety of solvent accessible amino acids leading to the formation of a water channel. This water channel leads from one face of the  $\beta$ -hairpin to the active site of the protein terminating at the conserved catalytic tyrosine (Xu et al., 2008). The importance of this central pore/water channel is undefined, yet the conservation of this feature within the catalytic domain of sialidases is indicative of its significance to its evolutionary function (Luo et al., 1998, Crennell et al., 1996). With some exceptions sialidases cleave Sia in the presence of a water molecule. A hydrophilic pocket and residues contributing to binding of a water molecule would likely increase the probability a water molecule would be in the correct space, time and orientation for hydrolysis/substitution to occur at the anomeric carbon (Buschiazzo et al., 2000).



**Table 1.**

	NedA (1EUR)	VCS (1KIT)	TcTS (1MRS)	TrSA (1MZS)	Neu2 (1SNT)	STNA (2SIL)	NanI (2VK5)	NanB (2VW0)	NanA (2W20)	PaNa (3H6J)	BtSa (4BBW)	BDI-2946 (4FJ6)	Baccac_01090 (4Q6K)	RgNanH (4X47)	NanC (4YZI)
NedA(1EUR)		0.26	0.19	0.18	0.30	0.23	0.35	0.23	0.31	0.17	0.34	0.32	0.33	0.28	0.23
VCS(1KIT)	1.64		0.27	0.24	0.23	0.22	0.24	0.21	0.25	0.13	0.27	0.26	0.26	0.25	0.22
TcTS(1MRS)	1.83	1.95		0.77	0.21	0.31	0.27	0.19	0.23	0.12	0.26	0.26	0.25	0.22	0.19
TrSA(1MZS)	1.78	1.94	0.54		0.19	0.32	0.26	0.18	0.22	0.11	0.24	0.26	0.25	0.23	0.19
Neu2(1SNT)	1.37	1.78	1.99	2.00		0.23	0.26	0.18	0.26	0.14	0.28	0.28	0.30	0.24	0.17
STNA(2SIL)	1.88	2.12	1.81	1.78	2.03		0.26	0.18	0.22	0.12	0.32	0.33	0.32	0.24	0.19
NanI(2VK5)	1.13	1.68	1.85	1.83	1.41	1.95		0.33	0.50	0.18	0.33	0.33	0.33	0.44	0.35
NanB(2VW0)	1.70	2.14	2.19	2.20	1.91	2.31	1.57		0.26	0.17	0.23	0.22	0.23	0.44	0.55
NanA(2W20)	1.24	1.67	1.83	1.84	1.39	1.87	0.89	1.69		0.17	0.30	0.30	0.30	0.35	0.30
PaNa(3H6J)	2.42	2.55	2.66	2.69	2.59	2.56	2.56	2.27	2.27		0.16	0.16	0.16	0.18	0.16
BtSa(4BBW)	1.45	1.87	1.93	1.90	1.72	1.88	1.41	1.85	1.44	2.53		0.83	0.89	0.31	0.27
BDI-2946(4FJ6)	1.39	1.84	1.90	1.87	1.64	1.87	1.40	1.88	1.44	2.50	0.52		0.83	0.28	0.26
Baccac_01090(4Q6K)	1.40	1.82	1.88	1.85	1.67	1.87	1.41	1.88	1.44	2.49	0.49	0.54		0.29	0.27
RgNanH(4X47)	1.34	1.88	2.05	2.02	1.69	2.13	1.31	1.44	1.48	2.58	1.61	1.62	1.61		0.49
NanC(4YZI)	1.66	2.04	2.15	2.13	1.88	2.30	1.50	1.32	1.71	2.75	1.87	1.88	1.88	1.37	

**Figure 25. A.** NanB 3D structure alignment of SWISSPROT database. Aligned against NanL, NanH and NanA. Sausage diagram representing homology between structures aligned. Red residues represent homologous residues in NanL, NanH and NanA. The grey residues represent the differences in residues between NanL, NanH and NanA. The RMSD is represented by the width of the area. Large structural deviation can be seen in the CBM domain and small structural deviation can be seen in the catalytic domain. Figure created using ENDscript 2.0, PyMOL® and PowerPoint®. **Table 1.** Multiple comparison of 15 published structures of GH-33 sialidases on the PDB. Pink cells represent the sequence similarity of the published GH-33 sialidase structures (values represented as a ratio). The grey cells represent the structural similarity in RMSD (values in Å) of the published GH-33 sialidase structures. All values calculated by PDBfold. Figures created using ENDscript® 2.0 and PowerPoint®.

In functionally related enzymes, glycosidases, a conserved water channel exists within the GH-1 family. Seven internal water molecules are conserved within 90% of known structures published on the PDB (Teze et al., 2013). Further analysis of *Thermus thermophilus*  $\beta$ -glycosidase through deuterium-exchange mass spectroscopy identified quick amide backbone hydrogen-deuterium exchange to occur in peptide L117-A125, a peptide indicated to be buried within the crystal structure (Teze et al., 2013). Residue W120 in this peptide forms contacts with catalytic residue E164 and solvation of this peptide may have functional significance (Teze et al., 2013). Molecular dynamics (MD) simulations identify three long and narrow water clusters of which one is speculated to perform a role in the function of the GH-1 family (Teze et al., 2013). The authors speculate that two functional possibilities exist for the water clusters; either they provide the catalytically added

water for substrate hydrolysis or, fast dehydration of the active site is enabled to accommodate incoming substrate due to nanosecond water exchange rates (Teze et al., 2013). Within the GH117 family a chain of seven water molecules located in the channel has been suggested to deliver the catalytically added water molecule (Rebuffet et al., 2011). However, the exact role of the water channel is yet to be determined. It is clear from molecular dynamic and crystallographic studies that water channels are important for the efficient proton transfer through chains of hydrogen bonds, hydration, dehydration or active site access for a number of enzymes, including serine proteases, kinases, cytochrome P450 and ATP-synthase (Teze et al., 2013, Knight et al., 2009, Meyer, 1992, Gohlke et al., 2012, Oprea et al., 1997).

#### **1.59.4 “Asp boxes” and F/Y-R-I-P motif**

The Asp-box is a recurring motif found within the sialidase six-bladed  $\beta$  propeller topology (Quistgaard and Thirup, 2009). The Asp-box sequence enables a  $\beta$ -hairpin loop stabilised by alternating hydrogen bonds (Quistgaard and Thirup, 2009). Despite low sequence homology between sialidases, these repeated “asp-boxes” are found in topologically identical positions in the  $\beta$ -hairpin loop (Copley et al., 2001). Clearly an important structural motif, but the function of the “asp-box” is still to be identified (Copley et al., 2001).

Asp-box motifs contribute to the formation of the water channel mentioned above. Conserved water molecules are found bound to two residues in Asp-box (serine and threonine) motif. These conserved water molecules form integral parts of the Asp-box motif forming hydrogen bonds between the  $\beta$ -strands (Copley et al., 2001).

Other conserved sections include an N-terminal motif: T/FYRI/VP. This motif contains an arginine residue (arginine triad) important in substrate recognition/ orientation within the active site (Giacopuzzi et al., 2012, Monti et al., 2010).

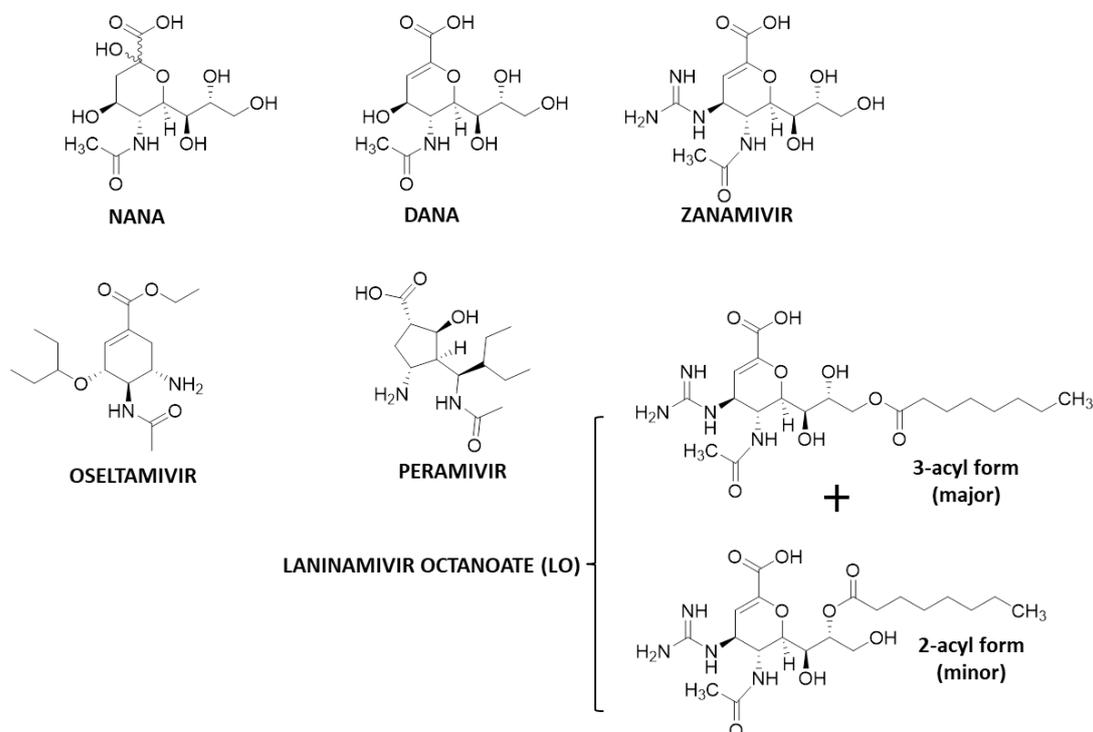
### **1.60 Current Inhibitors of Sialidases**

The first sialidase inhibitor reported was in the 1960s (Edmond et al., 1966, Rafelson Jr, 1962, Taylor, 2003). Rafelson published a review examining 85 compounds against sialidases (Rafelson Jr, 1962). Within this review, Rafelson noted NANA ( $C_{11}H_{19}NO_9$ ) was the most active competitive inhibitor and glutathione, cysteine and ascorbic acid acted as sialidase inhibitors (Rafelson Jr, 1962). Rafelson used NANA as a chemical tool to differentiate the kinetics between two strains of influenza virus (Rafelson Jr, 1962). In 1966, Edmond *et al.* published a paper attempting to find antivirals based around sialidase/neuraminidase inhibition (Edmond et al., 1966). Rather than synthesise

compounds structurally related to NANA a known competitive inhibitor at the time, they looked for another class identifying *N*-substituted oxamic acids as inhibitors (Edmond et al., 1966). Further studies and investigations around NANA led to the discovery of DANA (C<sub>11</sub>H<sub>17</sub>NO<sub>8</sub>), a general sialidase inhibitor mimicking the transition state of Sia hydrolysis (Edmond et al., 1966). The threat of a major human influenza pandemic is the key stimulus for the development of new sialidase inhibitors (Von Itzstein and Thomson, 2009). The influenza virus belonging to the orthomyxoviridae family is split into three serologically distinct types: A, B and C (Couch, 1996). Serological types A and B are responsible for annual incidences of human pathogenic disease whereas type C causes mild sporadic upper respiratory infections in children (Wong and Webby, 2013, Moriuchi et al., 1991). Influenza therapeutics, rimantidine and amantadine (adamantane-based M2 ion channel protein inhibitors), at that time were not extensively used due to CNS side effects, redundancy against influenza B infections and the rapid rise in drug resistant influenza A viral strains (Dong et al., 2015, Schnell and Chou, 2008, Zimmerman et al., 1997). New therapeutic agents were sought with influenza virion surface proteins hemagglutinin and influenza neuraminidases (NA) a major focus of drug development due to their role in virus infection. The published structures of hemagglutinin and NA provided opportunities for the development of new and specific inhibitors against the influenza A and B viral strains (von Itzstein, 2007). The unfavourable geometry of the sialic acid binding site in hemagglutinin resulted in failed attempts to generate potent selective inhibitors of this protein (Wade, 1997). Furthermore, the sporadic appearance of mutations within regions of this binding site reduced potency of designed inhibitors and added complexity to SBDD (Al-Majhdi, 2009). The influenza A and B NA are comparatively rigid with only minor structural changes in binding site conformation and geometry (Russell et al., 2006). The catalytic sites of these enzymes feature a high number of charged amino acids for hydrogen and electrostatic bonding to sialic acid and are invariant, indicating possible sub-type independent targeting (Zhang et al., 2008). The vital role played by the influenza sialidase in viral release and spread in the influenza A and B life cycle confirmed the validity of the target and provides a pharmacological strategy for drug development (von Itzstein, 2007).

Synthesis of DANA analogues led to a better inhibitor of influenza NA, 2-deoxy-2,3,dehydro-N-trifluoro-acetylneuraminic acid (C<sub>11</sub>H<sub>14</sub>NO<sub>8</sub>F<sub>3</sub>, FANA) with a K<sub>i</sub>=0.8 (Meindl et al., 1974b). FANA no activity *in vivo*, despite having the strongest potency of all previously developed sialidase inhibitors (von Itzstein, 2007). It was thought that metabolism and/or rapid clearance caused weak *in vivo* efficacy (von Itzstein, 2007). This data along with the recently published structure of DANA bound to influenza NA (solved in 1993 by P. Bossart-Whitaker *et al.*) enabled the development and advancement of better sialidase inhibitors through SBDD (Bossart-Whitaker et al., 1993, von Itzstein

et al., 1993). As a consequence, four neuraminidase inhibitors (NAIs) are currently clinically approved for flu: Oseltamivir ( $C_{16}H_{28}N_2O_4$ , Tamiflu<sup>®</sup>), zanamivir ( $C_{12}H_{20}N_4O_7$ , Relenza<sup>®</sup>), peramivir ( $C_{15}H_{28}N_4O_4$ , Rapivab<sup>®</sup>) and laninamivir octanoate ( $C_{21}H_{36}N_4O_8$ , Inavir<sup>®</sup>) (Ison, 2011) (Figure 26).



**Figure 26.** The chemical structure of different sialidase inhibitors.

Zanamivir (Relenza) was the first sialidase inhibitor rationally designed through SBDD (Taylor, 2003). Using DANA as a scaffold, GRID (a computational programme for *in situ* drug design) was used to optimise binding. From this SBDD a 4-guanidino group was identified as a functionality markedly favourable for influenza NA selectivity and potency (further discussed in 1.7.2 computer aided design) (von Itzstein et al., 1993). Important interactions of zanamivir include hydrogen bond and electrostatic interactions of the carboxylate with an arginine triad: Arg118, Arg292 and Arg371 (von Itzstein et al., 1993). Other hydrogen bond interactions include: the guanidino group with Glu227 and Glu119, the C8 – C9 hydroxyl groups with Glu276 (forming a bidentate bond), and the C5 acetamido group carbonyl oxygen and N-H with Arg152 and a buried water molecule respectively (von Itzstein et al., 1993). Zanamivir was approved for clinical use by the FDA in 1999 (Oxford, 2000). Zanamivir is a potent NA inhibitor, but suffers from poor oral bioavailability, rapid elimination and a small volume of distribution (Li et al., 1998, Hayden et al., 1997). Formulated as an inhalant for twice daily administration, zanamivir also has patient compliance concerns (Shanmugam, 2015). Based upon the discovery platform of zanamivir, oseltamivir was developed as a pro-drug for oral administration (von Itzstein, 2007). Core templates of cyclohexane, cyclohexane shikimic acid and

quinic acid were used in the discovery of oseltamivir (von Itzstein, 2007). Three key strategies were employed to develop an orally administered NA inhibitor: 1) positioning of the double bond, 2) replacement of the glycerol moiety with a lipophilic group and 3) pro-drug development to improve oral bioavailability (von Itzstein, 2007). The positioning of the double bond is critical to mimic the putative transition state sialosyl cation (von Itzstein, 2007). Movement of the double bond between C2 – C3 to C2 – C7 results in a >32 fold decrease in potency (von Itzstein, 2007). Replacement of the glycerol group was explored to improve lipophilicity while maintaining potency (von Itzstein, 2007). This led to the 3-pentyl ether side chain (von Itzstein, 2007). Structural information from a ligand-protein crystal complex showed an induced fit with changes in the active site upon binding, in particular an unpredicted movement of Glu276 towards Arg224 created a hydrophobic pocket (von Itzstein, 2007). This movement is essential for oseltamivir's potency allowing accommodation of the branched alkyl ether (Sriwilaijaroen et al., 2016). This change resulted in improved lipophilicity from LogP -4.1 for zanamivir to LogP -2.1 for oseltamivir carboxylate (Lindegardh et al., 2011, Bahrami et al., 2008, Oo et al., 2003). A prodrug strategy was employed, as the addition of this lipophilic group was not sufficient to improve bioavailability despite the improvement in lipophilicity (von Itzstein, 2007). The esterification of the carboxylate to the ethyl ester resulted in a significant change in lipophilicity (LogP = 0.36) (Oo et al., 2003). *In vivo* endogenous esterases hydrolyse the ester to form the active NA inhibitor, oseltamivir carboxylate (Rautio et al., 2008). Oseltamivir was FDA approved in 1999 and is currently marketed as Tamiflu® (Chand et al., 2005, Lindemann et al., 2010). Both oseltamivir and zanamivir are first generation NA inhibitors (Vavricka et al., 2013).

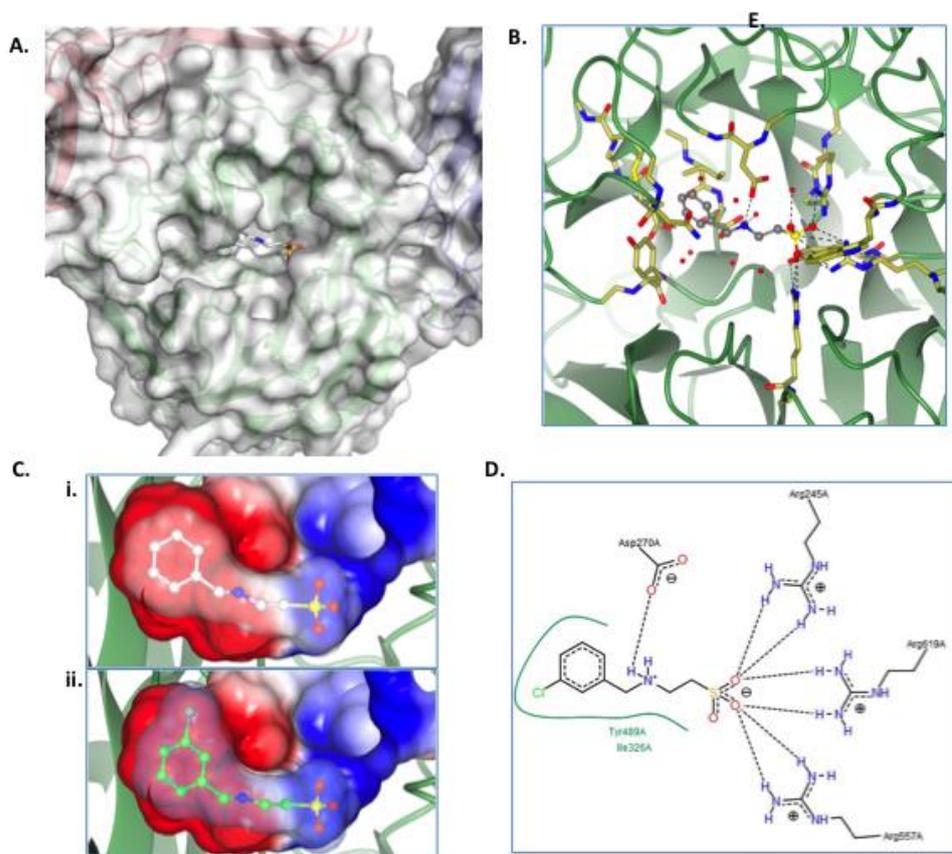
Second generation NA inhibitors followed on from the success of oseltamivir and zanamivir. Laninamivir is another example of a pro-drug where its active component contains a similar active structure to zanamivir but the C7 hydroxyl is methylated (De Clercq, 2013). The prodrug form of laninamivir is laninamivir octonate ester, with esterification on the C9 hydroxyl rather than on the carboxylic acid (De Clercq, 2013). This prodrug form is slow to eliminate lasting up to six days after a single inhalation in healthy volunteers. This long lasting pro-drug is also active against H274Y (H275Y in N1 numbering) mutated oseltamivir-resistant virus (Feng et al., 2012). This form of NA inhibitor marketed as Inavir® was approved in Japan in 2010 for the treatment of influenza (Feng et al., 2012, De Clercq, 2013). Another second generation NA inhibitor is peramivir (Król et al., 2014). Further structure based drug design on influenza NA led to the discovery of a series of cyclopentene derivatives (von Itzstein, 2007). From these derivatives (1S, 2S, 3S, 4R)-3-(1S)-1-acetylamino-2-hydroxycyclopentene carboxylic acid (Peramivir, BCX-1812) was developed as a highly selective inhibitor of influenza A and B (Smee and Sidwell, 2002). *In vitro* studies of peramivir ran in parallel with oseltamivir carboxylate and zanamivir found peramivir to be more or equipotent (Bantia et al.,

2006). Peramivir remains potent against the S247N influenza virus mutation but showed a decreased potency against the H274Y (H275Y in N1 numbering) influenza virus mutation (Abed et al., 2012). This Tyr side chain residue induces a shift in Glu276 resulting in weaker hydrophobic interactions with the pentanyl functional group of peramivir and oseltamivir (Steinmetzer et al., 2015). Peramivir has passed through clinical trials and was in use for emergency administration in 2009 during the H1N1 pandemic (Castillo et al., 2010). Peramivir marketed as Rapivab® was approved in 2014 as a single intravenous dose for the treatment of acute uncomplicated influenza (McLaughlin et al., 2015, Van Epps, 2016). Retrospective studies of hospitalized influenza patients observed that NAIs reduce viral replication and improve survival (Hernandez et al., 2011, Lee et al., 2010). Despite the success of these antivirals, new antivirals are needed to circumvent limitations of treatment-emergent antiviral resistance and variable resistance observed amongst differing strains.

Replacement of the carboxylate in sialic acid mimetics have shown to increase potency through improved electrostatic interactions with the arginine triad (Shie et al., 2011). Phosphonate derivatives of DANA (4-amino-1-phosphono-DANA) and zanamivir (phosphono-zanamivir) result in more extensive hydrogen bonding interactions and stronger electrostatic interaction with the arginine triad (Shie et al., 2011). Apart from increased potency, these phosphonate derivatives are nontoxic to human 293T cells and have nanomolar EC<sub>50</sub> values against oseltamivir-resistant strains (Feng et al., 2012).

### **1.61 Current Inhibitors of NanB**

To our knowledge the Westwood group is the only group to develop and characterise non-generalised sialidase inhibitors of *S. pneumoniae* sialidases. Within the Westwood group an active site targeted NanB inhibitor has been developed and characterised (Brear et al., 2012). This inhibitor was developed on the basis of a CHES molecule, which was found serendipitously bound in the protein crystal structure solved in the Taylor lab by Xu *et al.*, 2008. CHES was used as a minimal binding fragment and optimised based upon crystallographic structural information provided from ligand-protein complexes (Brear et al., 2012).



**Figure 27.** **A.** NanB surface representation with CHES bound within the active site cavity. **B.** Interactions of CHES (in grey) with residues in the active site (side chains in yellow and catalytic domain in green). **C.** i). Surface and van der Waals radius of CHES within the active site cavity. ii). Surface and van der Waals radius of 2-[(3-Chlorobenzyl)ammonio]ethanesulfonate within the active site cavity. **D.** Schematic representation of the binding of 2-[(3-Chlorobenzyl)ammonio]ethanesulfonate in the active site. Hydrogen bonds are represented by dashed lines and hydrophobic contacts are represented by a green line. Figures created using PyMOL<sup>®</sup>, CCP4 QTMG<sup>®</sup> and POSEVIEW<sup>®</sup>.

The most potent *N*-substituted analogue developed was 2-[(3-chlorobenzyl)ammonio]ethanesulfonate with an  $IC_{50}$  of  $38.9 \pm 0.8 \mu M$ . This CHES analogue at  $500 \mu M$  inhibited NanA at 0.3 ( $\pm 2.9$ ) exhibiting some selectivity between the *S. pneumoniae* sialidase family (Brear et al., 2012). Although this inhibitor appears to be selective between these sialidases it has not been tested against other intra-molecular sialidases and would likely be unselective against these enzymes due to the higher active site homology.

## 1.62 Current Inhibitors of TcTS

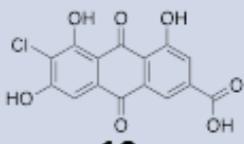
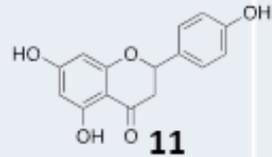
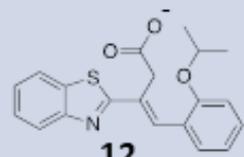
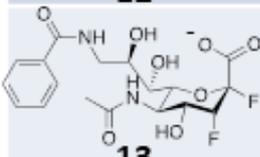
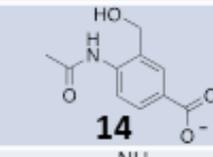
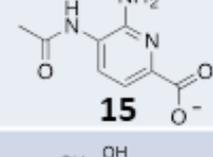
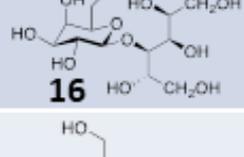
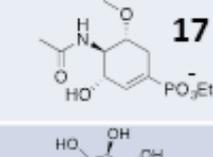
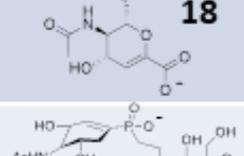
Following on the success of the Influenza neuraminidase inhibiting strategy, the current classes of inhibitors for TcTS have focused mainly on substrate mimetics (Neres et al., 2008). Despite sequence homology with other bacterial and viral sialidases slight variations in the potency of transition state mimetics is observed for TcTS (Paris et al., 2005, Neres et al., 2008). An oxonium ion transition state analogue 2,3-didehydro-2-deoxy-N-acetylneuraminic acid (Neu2en5Ac) is around 100 times less potent against TcTS than viral and bacterial sialidases (Todeschini et al., 2000, Ratier et al., 2008). The crystal structure of TcTS was solved by Buschiazzo *et al.*, 2002. In this structure sialic acid binds in a slightly tilted conformation as compared with other sialidases and this may explain the lack of potency of Neu2en5Ac against TcTS (Buschiazzo et al., 2002). Lactose binding to its pocket is mainly hydrophobic with one protein hydrogen bond contact (Asp52) (Amaya et al., 2004). Sialic acid mimetics have been observed as weak inhibitors of TcTS (Neres et al., 2009). Zanamivir and Peramivir do not inhibit TcTS at low mM concentrations (Neres et al., 2008). The covalent sialidase inhibitor, 2,3-F-Neu5Ac inhibits TcTS time dependently, but requires very high concentrations (20mM) for complete inactivation (Watts et al., 2003). The phosphonate derivatives of sialidase inhibitors do not inhibit TcTS despite improved potency against influenza neuraminidase (Neres et al., 2008).

A lactose binding site exists adjacent to the sialic acid binding site on TcTS (Buschiazzo et al., 2002). This is absent in other sialidases and enables more possibilities for exploitation of a chemical tool. This however may represent a difficulty in inhibitor development as strong potency may require binding to both pockets in the binding site. This may explain the reason why weak inhibitors have only been identified so far as a two pocket binding approach presents additional optimisation difficulties. Binding within the lactose pocket is important for potency against TcTS. Lactitol, a lactose derivative, is not a conventional TcTS inhibitor, but rather a competitive acceptor of Sia. Lactitol does not inhibit TcTS enzymatic activity, but was determined to have an  $IC_{50}$  of 0.57mM (Neres et al., 2008). Lactitol prevents sialylation of parasitic mucins and diminished *T. cruzi* infection of cultured mammalian Vero cells by 20-27% in a cellular infection assay (Neres et al., 2008). Hydrophobic interactions and potential for aromatic stacking interactions within the lactose binding pocket have led to the discovery of pyridoxal phosphate, benzoic acid and pyridine-2-carboxylic acid inhibitor scaffolds as TcTS inhibitors (Table 2).

Important residues located within the active site include the arginine triad (Arg314, Arg245, Arg35), Asp96, Asp59, Tyr312 and Tyr119 (Buschiazzo et al., 2002). These residues are important for ligand interactions in both the lactose and Sia binding pocket (Buschiazzo et al., 2002). A

computational and chemical library screening approach used these residues as a selection strategy/application filter to identify a benzothiazoyl as a novel inhibitor scaffold against TcTS. The Evotec in house supplier database was used containing approximately 2.5 million chemical entities narrowed down to 1.5 million by drug-likeness filters (Neres et al., 2009)(Lipinski's rules see Medicinal Chemistry). Rescoring was performed to give enrichment and ligands considered with high scores from both scoring functions (Neres et al., 2009). Ligands also found to hydrogen bond with Arg314, Arg245, Arg35, Asp96, Asp59 and hydrophobic interactions with Tyr312 and Tyr119 were considered as these interactions are observed in co-crystal structures of TcTS (Neres et al., 2009). 23 compounds were selected, purchased and assayed. 3-benzothiazol-2-yl-4-phenyl-but-3-enoic acid was identified as the most potent scaffold, however further kinetic analysis identified this inhibitor as non-competitive or a mixed inhibitor (Neres et al., 2009)(Table 2). The authors did not examine selectivity and it is likely that this compound suffers from promiscuity.

**Table 2.** Table of current published inhibitors of TcTS.

Structure	IC <sub>50</sub>	Method of Discovery	Reported
 <p><b>10</b></p>	0.58μM	Natural product library of 2,283 compounds	(Arioka et al., 2010)
 <p><b>11</b></p>	78μM	Natural product library of 2,283 compounds	(Arioka et al., 2010)
 <p><b>12</b></p>	150μM	In silico screen using Evotec database (approximately 2.5 million compounds) and GOLD molecular docking.	(Neres et al., 2009)
 <p><b>13</b></p>	N.R. (approximately mM activity)	SBDD based upon covalent general sialidase inhibitor.	(Buchini et al., 2008)
 <p><b>14</b></p>	540μM	SBDD using GRID and DOCK4.0	(Neres et al., 2007)
 <p><b>15</b></p>	440μM	SBDD using GRID and DOCK4.0	(Neres et al., 2007)
 <p><b>16</b></p>	570μM	Known as a sialic acid acceptor	(Agusti et al., 2004)
 <p><b>17</b></p>	4.7mM	Library synthesis and screen for general sialidase inhibition	(Streicher and Busse, 2006)
 <p><b>18</b></p>	12.3mM	Known sialidase inhibitor (DANA).	(Paris et al., 2005)
 <p><b>19</b></p>	1.5mM	SBDD	(Busse et al., 2007)

Further difficulties in developing inhibitors are presented by conformational changes observed within *apo* and *holo* crystal structures (Buschiazzo et al., 2002). In the *holo* enzyme, Tyr119 flips up and into an open conformation for lactose binding (Buschiazzo et al., 2002). In the *apo* enzyme Tyr119 flips down filling the pocket and forming the closed conformation (Buschiazzo et al., 2002). Another conformational switch between active and inactive states occurs with Tyr342 (Buschiazzo et al., 2002). In the active state, a hydrogen bonding interaction exists between Tyr342 and the carboxylate oxygen of Glu230, orientating the tyrosine underneath the scissile glycosidic linkage (Buschiazzo et al., 2002). This is thought to stabilise the substrate and/or the intermediate formed in the reaction (Buschiazzo et al., 2002). In unliganded crystal structures, Tyr342 does not form a hydrogen bond with Glu230 and is suggested to be in an inactive state (Buschiazzo et al., 2002). The importance of these conformational changes is unclear in terms of inhibitor design, i.e. is it important to design an inhibitor that stabilises the open or closed conformation for lactose binding.

A list of currently published inhibitors of TcTS has been compiled (Table 2.). Currently the most potent inhibitor to date is an anthraquinone. This scaffold was identified from a natural product library. SAR analysis led to 6-chloro-9,10-dihydro-4,5,7-trihydroxy-9,10-dioxo-2-anthracenecarboxylic acid (Arioka et al., 2010). To this date no inhibitor-protein crystal structure complex exists on the PDB. Poor inhibition observed for current TcTS inhibitors likely limits the success of these tools for *in vivo* and *in vitro* *T. cruzi* infection assays. It is clear new potent chemical tools are needed for TcTS to validate the sialidase as a drug target.

### **1.70 Rational Design**

Rational design is the major component of reverse chemical genetics (Spring, 2005). A rational approach to drug discovery only emerged 50-60 years ago when the behaviour of a drug could be correlated to its intrinsic physiochemical properties (Adam, 2005). In a target centric approach modern chemical tool/drug design is a use of medicinal chemistry, structural biology, computational chemistry and pharmacology (Satyanarayanajois and Hill, 2011). This combined knowledge allows the design of molecules with an appreciation of the binding and physiochemical properties required of an optimal modulator. This sub-chapter will discuss the uses of these disciplines that enable intuitive approaches to chemical tool/drug design.

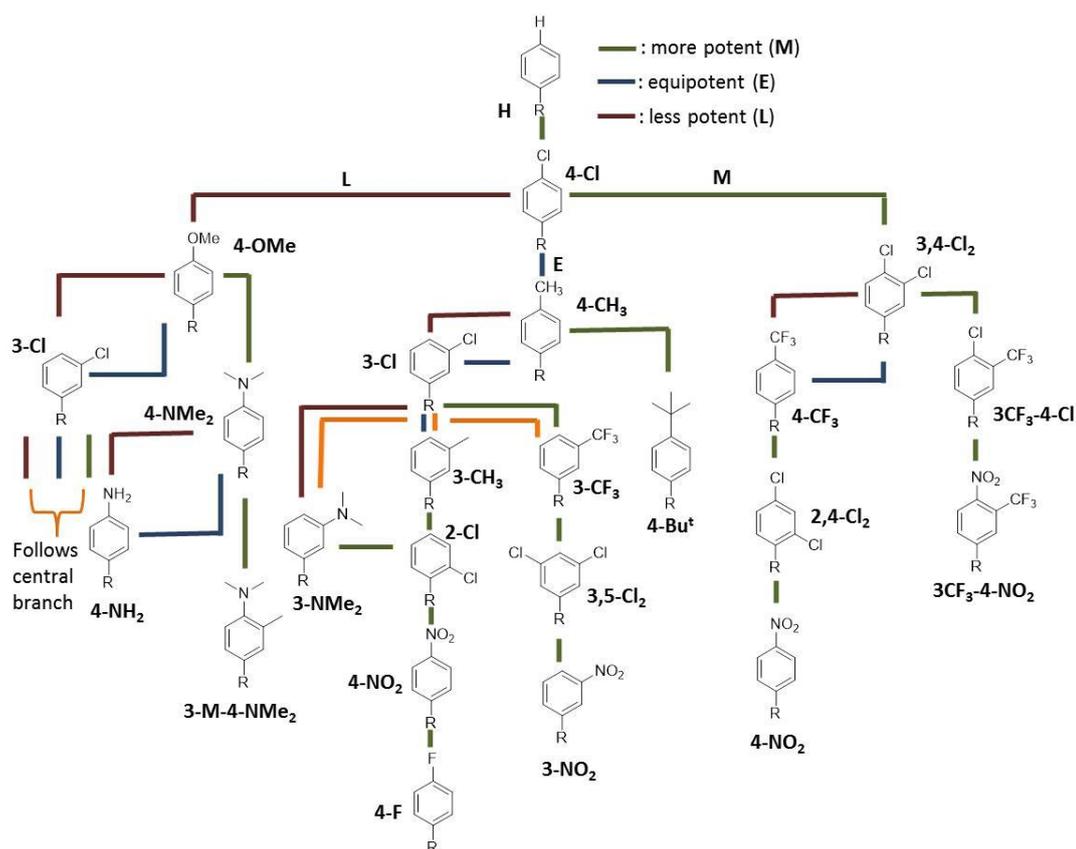
## **1.71 Medicinal chemistry**

Medicinal chemistry is the core of small molecule drug development (Hann and Keserü, 2012). Medicinal chemists prepare appropriate compounds for biological study (Lombardino and Lowe, 2004). The use of medicinal chemistry includes: the evaluation of structure activity relationships, chemical synthesis, purification, the disease requirement and the identification of the key physicochemical properties appropriate for good bioactivity (Lombardino and Lowe, 2004). In particular, oral ingestion is the preferred route of drug administration (Verma and Garg, 2001). Post analysis of drug success and failure has led to a series of guidelines for obtaining this goal. As such these guidelines have impacted on small molecule drug design (Cumming et al., 2013). Lipinski and co-workers in 1997 published a 'rule of five' (RO5) based upon observations of drugs in phase II and later development (Lipinski et al., 2012, Lipinski et al., 1997). They proposed that drug permeability and adsorption were more likely when the molecular weight is <500, LogP is <5, hydrogen bond donors <5 and hydrogen bond acceptors <10 (Livingstone and Davis, 2011).

The RO5 is useful if the drug makes use of passive diffusion (Livingstone and Davis, 2011). In instances where membrane transporters are used in permeation these guidelines no longer apply (Lipinski et al., 1997). In these observations failing one of these RO5 conditions is considered acceptable, however failing two conditions is not (Livingstone and Davis, 2011). Drugs that have failed two or more are rarely observed to pass through clinical trials (Livingstone and Davis, 2011). As a consequence this concept has gained wide acceptance as an approach to reduce attrition in drug discovery (Leeson and Springthorpe, 2007). Critical properties in defining compound quality include: log P, pKa, log D, solubility and hydrogen bonding descriptors (Gleeson et al., 2015). Log P is a measure of lipophilicity that represents the affinity of a molecule for a lipophilic environment (Chen and Weber, 2007). It is an important feature for solubility, binding and permeability (Faubert et al., 2014). Many compounds have had their log Ps measured experimentally at 20°C in an *n*-octanol/water partition system (Livingstone and Davis, 2011). These measurements are compiled in a database and used for a basis of prediction algorithms for log P calculations, the most popular being CLOGP (Leo et al., 1975). When a compound contains an ionisable centre the distribution within *n*-octanol and water differ depending on the pH of the aqueous phase (Livingstone and Davis, 2011). LogD is the 1-octanol-water coefficient at various pH values (Leeson and Springthorpe, 2007). Ionisation of the molecule is an important feature for membrane diffusion (Kerns and Di, 2003).

Drug design strategies include quantitative structure-activity relationships (QSAR) (Livingstone and Davis, 2011). Regression analysis is used to correlate physicochemical parameters with activity values. The Topliss scheme/series is one such QSAR approach aimed at identifying electrostatic

properties that impact on potency (Cherkasov et al., 2014). The Topliss series is an operational scheme used for maximising analogue potency as early in the series as possible (Topliss, 1972). Two Topliss schemes exist in practice: one for aliphatic substituents and one for aromatic substituents (Figure 28) (Austel, 1983). These schemes consider both electronic and hydrophobic features and the impact on biological activity (Topliss, 1972).



**Figure 28.** A schematic of the Topliss Scheme. The activity of the next analogue can be more (M), equal (E) or less (L) active. Sterics, hydrophobicity and/ or electron withdrawing properties impact on the activity and the result indicates the route followed for substitution (Topliss, 1972).

An assumption is made that the “lead” compound possesses a single mono-substituted alkyl chain or aromatic ring. The 4-Cl possesses positive  $\pi$  and  $\sigma$  values relating to more hydrophobic and electron-withdrawing properties (Jorge et al., 2011). The synthesis of the next analogue depends on the impact these properties have on biological activity (Jorge et al., 2011). If potency improves then increasing these properties would likely improve biological activity (Jorge et al., 2011, Topliss, 1972).

Ligand efficiency metrics are measures of potency correlated to physical properties (Hopkins et al., 2014). These metrics can be correlated to any physical property, but the most frequently used are heavy atom count (HA) and lipophilicity (Hopkins et al., 2014). Further considerations include

toxicology (Venkatesh and Lipper, 2000). The human Ether-à-go-go Related Gene (hERG) is the inner pore forming region/ $\alpha$  subunit of a potassium channel responsible for repolarisation of cardiac muscle (Sanguinetti and Tristani-Firouzi, 2006). When a compound binds in this channel it blocks cardiac muscle repolarisation causing a QT prolongation that can induce cardiac arrhythmia (Sanguinetti and Tristani-Firouzi, 2006). Pre-clinical candidates that have been evaluated to bind to hERG can undergo clinical trials (Redfern et al., 2003). However, clinical trials of drugs known to bind to this channel are expensive, longer and more extensive (Witchel, 2011, Fermini and Fossa, 2003). Early detection and avoidance of hERG inhibition is promoted throughout early drug development (Hefti, 2008). General features of hERG blockers include: a basic amine, hydrophobicity (clog P >3.7), absence of a negatively ionisable functionality and a low number of oxygen hydrogen bond acceptors (Aronov, 2005). Removing any of these features would likely decrease affinity of a compound for the hERG channel (Aronov, 2005).

### **1.72 Structure-based Design**

Structure-based drug design (SBDD) aims to identify and optimise interactions between ligands and the target molecule (Lounnas et al., 2013). Once a target has been identified, X-ray crystallography, NMR and homology modelling are the primary methods for obtaining structural information (Anderson, 2003). In structure-based design multiple iterative cycles of analysis and further development are performed before a 'lead' is generated (Anderson, 2003). The first cycle involves structure determination (including: cloning, purification and structural determination of the target) and identification of an initial 'hit' (fragment or full compound HTS). Additional cycles start when a promising 'hit' has been identified. Determination of the hit-target complex and further optimisation for potency is performed (Anderson, 2003). The optimised ligand is then subjected to a repeat of ligand-target structure determination and further optimisation (Anderson, 2003). A further cycle may exist in this process for optimisation of medicinal chemistry parameters including improving bioavailability and reducing toxicity issues. *In silico* docking is another approach to hit identification and optimisation (Kitchen et al., 2004) (See Computational Design).

In the optimisation of a ligand a number of energy considerations are needed in its design (Fenley et al., 2012). The Gibbs free energy ( $\Delta G$ ) of binding dictates binding affinity and is the sum of two energy terms, enthalpy ( $\Delta E$ ) and entropy ( $\Delta S$ ) (Du et al., 2016). In order to achieve high binding affinity both energy terms need to contribute favourably (Freire, 2008). The first energy term enthalpy is a difficult parameter to optimise (Freire, 2008). The formation of favourable new hydrogen bonds and van der Waals contacts with the protein would improve enthalpy, but conflict can occur with the unfavourable penalty of polar group desolvation (Freire, 2008). Solvation of the

ligand and binding site is vital in the energetics of binding (de Beer et al., 2010). Desolvation of polar groups carries an enthalpic cost in the order of 8 kcalmol<sup>-1</sup> at 25°C (Freire, 2008). Desolvation of non-polar groups is one order of magnitude lower (Freire, 2008). Hydrogen bonds and electrostatic interactions have a narrow tolerance of approximately 0.2 Å for both angle and distance (Anderson, 2003). If the distance and angle are sub-optimal in its binding the enthalpic contribution becomes unfavourable due to the desolvation penalty (Freire, 2008). Neutral hydrogen bond distances range from 2.7-3.2 Å and charged hydrogen bonds with carboxylates range from 2.6-3.0 Å (Davis and Ward, 2014). The hydrogen bond is oriented to approach the lone pair of the acceptor and is frequently observed to adopt a donor-hydrogen-acceptor angle of >150° (Davis and Ward, 2014).

A perfect conformational fit of the ligand with the protein maximises van der Waals contacts improving potency. Furthermore ligand geometry and conformation is equally important as energy minima or low energy conformations improve potency (Bissantz et al., 2010). The development of human carbonic anhydrase II (HCA II) inhibitor is an example of the importance of correct ligand conformation (Baldwin et al., 1989). Similar binding modes of two enantiomers of the prototype lead MK-927 within the X-ray complex structure could not explain the 100 fold difference in potency (Talele et al., 2010, Bissantz et al., 2010). Quantum mechanics calculations identified the *R* conformation to be suboptimal with the N-S-C-S dihedral angle to have a calculated strain of 1kcalmol<sup>-1</sup> (Talele et al., 2010). Further conformational difference occurs at the 4-isobutylamino substituent (Talele et al., 2010). In the *S* enantiomer the side chain is *trans* whereas in the *R* enantiomer it is *cis* (Talele et al., 2010). *Ab initio* calculations identified the *trans* geometry to be preferred by 1kcalmol<sup>-1</sup> (Talele et al., 2010). These two features explain the marked improvement in potency between two enantiomers and the importance of ligand conformation (Talele et al., 2010). Preferred conformations can be exploited through the use of hydrogen bond donor and acceptor regions within close proximity creating a temporary ring system (Kuhn et al., 2010). A thermodynamic equilibrium exists between the closed and open conformations where in the open form polar substituents are exposed making the molecule more soluble (Kuhn et al., 2010). Intermolecular hydrogen bonds can restrict the ligand in the favourable ligand-protein conformation shifting the equilibrium (Kuhn et al., 2010). In the closed form polar substituents are hidden increasing lipophilicity (Kuhn et al., 2010). As such the formation of intramolecular hydrogen bonds can impact on solubility, permeability and the lipophilicity of molecules (Kuhn et al., 2010). Intermolecular hydrogen bonds preferentially have angles greater than 150° and are close to linear (Kuhn et al., 2010). These intermolecular bonds are favoured when five or six membered rings are formed (Kuhn et al., 2010). The size of the ring system can impact on angle geometry and bond distances. The most common intermolecular hydrogen bond formation was observed in six-

membered rings (Kuhn et al., 2010). The constraints of the six membered rings however resulted in less optimal angles for intermolecular hydrogen bonds of 130° to 140° (Kuhn et al., 2010). Five membered rings have smaller angles and longer distances and should not be classed as classical hydrogen bonds but as favourable electrostatic interactions (Kuhn et al., 2010). The introduction of an intramolecular hydrogen bond increased brain penetration for an NK1 antagonist (Ashwood et al., 2001). By the same intramolecular hydrogen bond inclusion oral absorption was increased for a luteinizing hormone-releasing hormone receptor antagonist developed by Takeda (Sasaki et al., 2003).

Entropy the second energy term, is easier to optimise (Freire, 2008). Replacing well-ordered water molecules in the binding site will increase the entropy of the system (entropy driven ligand binding) (Ladbury, 1996, Michel et al., 2009). Ordered water molecules can be observed within high-resolution crystal structures (Wlodawer et al., 2008). Ordered water molecules can be treated as bound ligands and the contacts with the binding site can be used in the optimisation of the ligand (increasing enthalpy driven ligand binding), but only if the favourable enthalpy of the protein-ligand contacts outweighs the unfavourable penalty of desolvation (Bissantz et al., 2010).

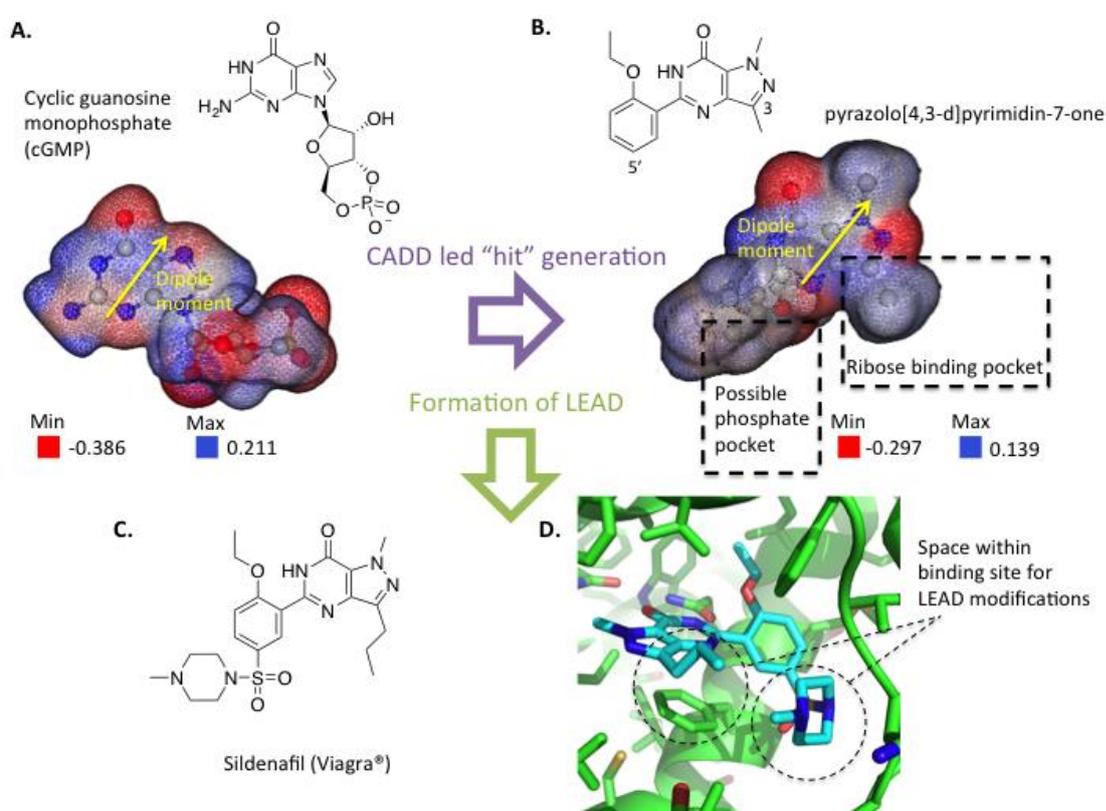
Misleading SBDD conclusions can be drawn from: 1) coordinate error in a crystal structure due to low-resolution data or uncertainty in ligand identity and/or position and 2) changes in structure and protein/ligand protonation states from crystallisation conditions and crystal packing (Davis et al., 2003). Despite this, structural information has been successfully used in target-based approaches. For drugs approved until the end of 2009, 31 known examples have had significant contributions from X-ray and NMR structural data for SBDD (Livingstone and Davis, 2011). It is likely that more approved drugs had a contribution from X-ray and NMR structural data, but this was not published in the literature (Livingstone and Davis, 2011).

### ***1.73 Computational design***

Computational design has been aided by the increasing availability, size and speed of computational resources as well as the increasing structural information of endogenous ligand and protein binding sites (Livingstone and Davis, 2011). Key developments in the algorithms used to calculate atomistic detail and the physicochemical properties of the ligand or the protein have led to more accurate solutions and an improvement in computational modelling and predictions. Two principal types of screening exist: ligand similarity screening and molecular docking.

In the past, the computational aided design was limited to modelling leads based upon similarities to endogenous ligand. A key example of this would be in the discovery of Sildenafil

(Viagra<sup>TM</sup>). Sildenafil was designed from informed computational lead optimisation (Campbell, 2000). Pfizer's R&D used pyrazolopyrimidinone as a core structure for optimisation based upon similar shape, size and dipole moment calculations as the endogenous cGMP ligand (Terrett et al., 1996). Derivatives of pyrazolo(4,3-d)pyrimidin-7-one were found to be potent inhibitors of phosphodiesterases (PDE) type 5 and type 1 (Terrett et al., 1996). Further optimisation of this core unit was carried out to increase potency and selectivity for PDE type 5 (Terrett et al., 1996). Using cGMP as a model it was determined that altering substituents on the 3-position would likely occupy the ribose binding site and substituents on the 5'-position could occupy the phosphate binding site (Terrett et al., 1996)(Figure 29).



**Figure 29.** **A.** Structure and an electrostatic representation of cGMP. **B.** Structure and an electrostatic surface representation of pyrazole[4,3-d]pyrimidin-7-one ("hit" led by computational aided drug design (CADD)). **C.** Structure of Sildenafil. Functionalization and substitution on the 3 and 5' position led to the LEAD, Sildenafil. **D.** The binding site of Sildenafil (in teal) in PDE5 (in green) (PDB: 2H42) published in 2006 after Sildenafil was approved (ligand-protein structure was not available during development). Binding site confirms the functionalization hypothesis with space available within the 3 and 5' position of the "hit". Structures created using ChemDraw<sup>®</sup> Professional 15. Electrostatic surface representations created using ChemAxon<sup>®</sup> software MarvinView<sup>®</sup>. Image of Sildenafil bound to PDE5 created using PyMOL<sup>®</sup>.

Structural modification with computational aid improved PDE affinity of the initial lead by 500 fold leading to the discovery of sildenafil (Viagra®) (Terrett et al., 1996). Ligand similarity screening is a useful tool currently used within drug design today (Lengauer et al., 2004). The ranking of a database using ligand similarity virtual screening was first described in two papers published in the 1980s (Willett, 2006). Ligand similarity aims to measure and quantify the similarity of the reference structure against a set of structures within a database (Willett et al., 1998). This similarity measure is performed by three components: 1) the representation used to characterise the molecules for comparison, 2) the weighting scheme used to assign degrees of importance to these representations and 3) the coefficient used to determine the similarity between these representations (Willett, 2006). The most popular similarity metric in use is the Tanimoto similarity metric (Willett, 2006). The Tanimoto similarity coefficient  $\tau$ , consists of:

$$\tau = N_{AB}/N_A + N_B - N_{AB}$$

$N_A$  consists of the features/representations in A (the reference) and  $N_B$  consists of the features/representations in B (the query) (Flower, 1998).  $N_{AB}$  consists of the features/representations in common between A and B (Flower, 1998). The size of a screening library can be drastically reduced using chemical similarity filtering criteria (Yadav and Singh, 2013). Select chemical subsets of the library can be prioritised and screened rather than the whole library saving time and money (Yadav and Singh, 2013). This ensures that molecules that have likely functionality for binding and activity are selected, thereby reducing the likelihood of finding a non-promiscuous hit within an activity screen. Problems do exist with similarity metrics. The Tanimoto similarity metric typically yields low similarity values when the search is just a few representations (Flower, 1998). Despite the size dependency of the search, the Tanimoto metric is efficient and has been demonstrated to be effective in computational studies. A virtual screen of a compound library against known melanin-concentrating hormone-1 receptor (MCH-1R) ligands with a  $\leq 0.7$  Tanimoto similarity threshold resulted in a 14% hit rate and 10 novel potent MCH-1R antagonists, two with nanomolar potency (Heifetz et al., 2013).

Virtual screening is a cost effective approach to "hit" discovery (Lionta et al., 2014). It was in 1982 that Kuntz et al. paved the way for *in silico* screening programmes with the development of the DOCK algorithm (Kuntz et al., 1982). Since then a variety of different docking programmes are in use today (Ferreira et al., 2015). These docking programmes aim to give an accurate ligand-protein binding pose with a prediction of the free energy of binding (Meng et al., 2011). This provides the researcher with insight into the chemical functionalities important for ligand binding. In its simplistic term docking programmes are all-similar in that they use an algorithm to predict the coordinates of

a complex from the coordinates of each of the individual molecular entities. Within this algorithm an exhaustive approach is used to generate all possible coordinates for the complex and a scoring function is applied to weight the intermolecular interaction energy. Despite this similar concept differences occur in the approach used by the algorithm in coordinate generation. Differences occur also in the influence of steric overlap and types of intermolecular interaction in its scoring. Two main approaches to ligand coordinate mapping exist and these are: 1) a systematic/direct approach and 2) random/stochastic method (Pujadas et al., 2008). In the systematic direct approach the degree of freedom of the ligand is explored and this is performed in one of three ways: conformation search method, fragmentation/incremental construction or the database method (Huang and Zou, 2010). Conformational search algorithms generate possible ligand conformations by rotating any ligand bonds that can be rotated  $360^\circ$  by increments then docking those conformations (Schwab, 2011). This generates multiple ligand conformations that increase exponentially with each additional rotatable bond (Schwab, 2011). The fragmentation construction method divides the ligand into several rigid fragments, these fragments are then docked and then rebuilt/joined with the flexible linkers (the place-and-join approach) (Pujadas et al., 2008). The incremental approach first docks the core fragment from the ligand and then the rest of the ligand is sequentially added (Kroemer, 2007). FlexX is an example of a programme that uses this incremental approach (Rarey et al., 1996). An alternative technique is the database method that uses libraries of pre-generated conformations (conformational ensembles) and docks each one into a rigid body (Kitchen et al., 2004).

The random/stochastic method samples the conformational space of the ligand generating different conformations (Kitchen et al., 2004). The new conformation is rejected or accepted based upon predefined probabilities (Reddy et al., 2007, Chang et al., 1989). When a new conformer is accepted a next random conformation is generated and this is accepted or rejected and the cycle repeats (Chang et al., 1989). Three subtypes exist: Monte Carlo (MC) Method, Genetic Algorithm (GA) method and the tabu search (TS) method (Kitchen et al., 2004). The MC method randomly places the ligand in the receptor and generates new conformations by random changes of the ligands position or rotatable bonds (Brooijmans and Kuntz, 2003). After each change the ligand is minimised and scored (Brooijmans and Kuntz, 2003). If the score is higher than the previous change it is accepted (Brooijmans and Kuntz, 2003). However if the new pose is not a new minimum it is subjected to a Boltzman-based probability function and accepted only if it succeeds this test (Brooijmans and Kuntz, 2003). The GA method is an approach that starts from an initial population of different conformations (Ru et al., 2016). This method uses the theory of evolution and the favourable “genes” conformation passed onto the next generation eliminating the unfavourable (Meng et al., 2011). Degrees of freedom are encoded into each gene and assigned a fitness function

(Meng et al., 2011). GOLD is a programme that uses this GA method that includes full ligand and receptor hydrogen flexibility (Schneider and Böhm, 2002, Verdonk et al., 2003). The TS is a docking algorithm starting from an initial random ligand conformation (Taylor et al., 2002). Upon this initial ligand conformation, random moves are applied and scored generating typically a population of 100 solutions (Taylor et al., 2002). The highest-ranking solution is then accepted as the current solution and a new population is generated by random moves and scored. The process is then repeated by a user-defined number of iterations (Taylor et al., 2002). Within these cycles a Tabu list is created of 25 former conformations. If the current solution is the lowest energy it is accepted, but if it is not the lowest energy the best non-tabu solution is used (Taylor et al., 2002).

The main limitation of any of these docking approaches is that the observed conformation determined experimentally may not be achieved within the theoretical conformation determined by the docking approach (Taylor et al., 2002). Validation tests are performed on computational docking approaches to evaluate effectiveness (Kitchen et al., 2004). A comprehensive validation test was performed on GOLD using a test set of 100 different protein complexes (Jones et al., 1997). Within this validation set, the success rate of a correctly predicted binding pose was 71% with 66 complexes having an RMSD of 2.0 Å and 71 having an RMSD of 3.0 Å (Jones et al., 1997). A validation test of FlexX performed on a dataset of 200 protein-ligand complexes observed a success rate of 46.5% for an RMSD of less than 2.0 Å (Kramer et al., 1999). However, within the generated solution set the correct binding pose is predicted in 70% of the solutions (Kramer et al., 1999). Another limitation of computational docking is the length of time a vast library takes to run (Lionta et al., 2014). Differences in docking speed are observed for each docking approach. FlexX has a short running time compared to GOLD (Hioual et al., 2012). GOLD on the other hand requires more CPU time and for large database screening may not be the programme of choice despite being considered a more accurate docking programme (Kellenberger et al., 2004, Hioual et al., 2012). Automated workflows and parallel docking runs (on multiple processors) can speed up the output process, however this will still generate a large dataset that will need to be analysed and evaluated (Raicu, 2009). The laborious evaluation of multiple “hits” and identification of a valid output requires the skills of a medicinal and computational chemist. Analysis of large libraries of results can be accomplished through large data set mining and profiling (Hughes et al., 2011, Good et al., 2000).

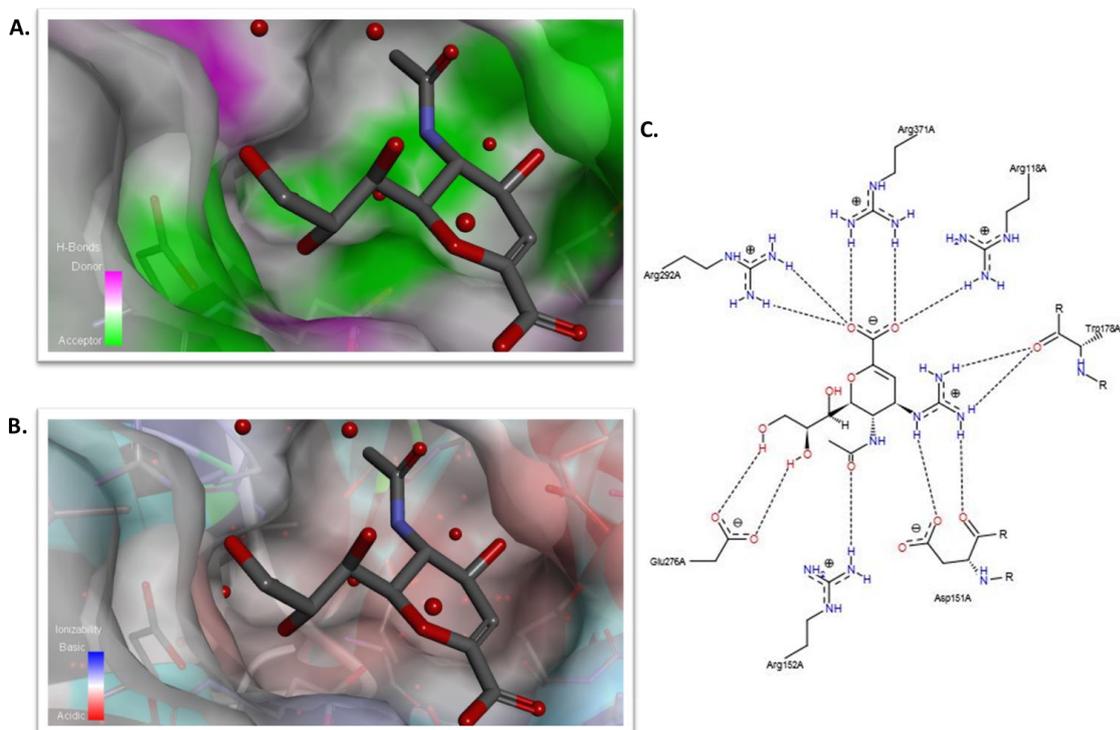
Scoring functions are implemented to make assumptions and simplifications of the binding interaction between ligand and protein and a prediction of binding affinity (Meng et al., 2011). These assumptions and simplifications do not account for all the physical aspects of molecular recognition due to the complexity and the time it would take to run (Kitchen et al., 2004). Three types of classes

of scoring function currently exist: force-field based scoring, knowledge based and empirical based scoring (Sousa et al., 2006). Force-field scoring quantifies the sum of two energies of the receptor-ligand interaction and the internal ligand energy (Sousa et al., 2006). Internal energy of the protein is usually omitted, simplifying scoring (Kitchen et al., 2004). The force-field scoring interactions described between ligand and proteins are van der Waals and electrostatic energy terms (Kitchen et al., 2004). Knowledge based functions are based on simple atomic interaction pair potentials (Kitchen et al., 2004). These pair potentials are derived from crystal data from the Cambridge Structural Database attempting to score binding effects difficult to model (Velec et al., 2005). Further adaptations to knowledge based scoring includes the use of derived protein-ligand interactions from the PDB retrieved using ReLiBase (Gohlke et al., 2000). This combined with terms considering burial of solvent-accessible surfaces of both ligand and protein are used to generate improved scoring (Gohlke et al., 2000). Empirical scoring functions are similar to knowledge based scoring functions in that they use experimental data in scoring (Gohlke et al., 2000). Experimental data sets are used for regression analysis and fitting to generate the coefficients used in the scoring function (Sliwoski et al., 2014). This class of scoring function can also include entropic terms. An additional scoring method includes a consensus scoring function, which combines multiple classes of scoring function described above (Charifson et al., 1999). This is used to balance errors in single scores and provide a more comprehensive score identifying “true” ligands (Kitchen et al., 2004).

Other computational resources for rational drug design include *de novo* ligand generation, toxicology and computational approaches to improve ADME properties in logP, logD calculations and P450 metabolism prediction (Sliwoski et al., 2014, Xing and Glen, 2002, Kirchmair et al., 2015). With crystal structures of the various CYP isoforms available for protein-ligand docking, this has improved structure-based predictions of metabolism and CYP inhibition (Kirchmair et al., 2012).

Despite the difficulties in accurately predicting binding conformations and scoring binding interactions as an estimate of affinity, computational design is at the core of modern drug discovery (Charifson and Kuntz, 1997). Success has been observed in SBDD and computer aided drug design (CADD) (Kapetanovic, 2008). Captopril (marketed as Capoten® by Bristol Myers-Squibb) was the first angiotensin-converting enzyme (ACE) inhibitor to enter clinical use in the U.S. (approved by the FDA in 1981) (Talele et al., 2010, Thiel, 2004). Captopril is one of the first successful endeavors of structure-based design (Thiel, 2004, Talele et al., 2010). The structure of a different zinc protease (carboxypeptidase A) in complex with its inhibitor (L-2-benzylsuccinic acid) provided the basis for a model of the ACE binding site guiding the generation of a LEAD ACE inhibitor N-succinoyl-L-proline

( $IC_{50} = 330\mu\text{M}$ ) (Kubinyi, 2006, Cushman et al., 1977). SAR on N-succinoyl-L-proline resulted in the generation of captopril (Kubinyi, 2006, Cushman et al., 1977). A more recent and topical example for this thesis is zanamivir. Zanamivir is a potent inhibitor of influenza neuraminidase marketed by GlaxoSmithKline under the name Relenza®. The GRID program developed by Peter Goodford was used to develop Relenza® (von Itzstein et al., 1993, Goodford, 1985). GRID determines energetically favourable binding sites enabling the identification of optimal ligand functionality for favourable ligand to protein interaction (English et al., 2001, Goodford, 1985). With this programme the transition state mimetic DANA (inhibits the protein with a  $K_i$  of approximately  $4\mu\text{M}$ ) was used as a scaffold for development of neuraminidase inhibitor analogues (Taylor and Russell, 2010, Meindl et al., 1974a). The influenza neuraminidase DANA crystal structure complex was solved and the bound conformation of inhibitor was used as a template for GRID SBDD (von Itzstein et al., 1996, von Itzstein et al., 1993). GRID identified that a positively charged group within the vicinity of the 4-OH group of DANA would make favourable van der Waals interactions to neighbouring Glu119 and Glu227 (von Itzstein et al., 1996, von Itzstein et al., 1993). Replacement of the 4-OH with an aliphatic amino group improved hydrogen bond interactions with the protein and improved affinity into the nanomolar range (von Itzstein et al., 1996) (Figure 30). Modification of the aliphatic amine to a guanidine group resulted in direct hydrogen bond interactions with both Glu119 and Glu227 residues improving the  $K_i$  to 0.2nM (from 50nM against A/Tokyo/3/67 strain) (Taylor and Russell, 2010, von Itzstein et al., 1996, Varghese, 1999). Despite improved interactions zanamivir suffers from poor bioavailability due to its high polarity (Varghese Gupta et al., 2011).



**Figure 30. A.** Hydrogen bond donor and acceptor regions within the influenza A neuraminidase binding site (PDB: 1F8B). DANA (in grey) is bound within the binding site, a strong hydrogen bond acceptor region is located near the 4-OH. **B.** Ionization surface representation of the influenza A neuraminidase binding site. An acidic region is located near the 4-OH of DANA. Both surface representations show agreement with GRID in placing a large basic, hydrogen bond donor group in the vicinity of the 4-OH of DANA. **C.** A diagram of the interactions between Zanamivir and the influenza A neuraminidase. Discovery studios<sup>®</sup> was used to construct ionization and hydrogen bonding surface representation of the influenza A neuraminidase binding site. PoseView<sup>®</sup> was used to construct a 2D image and interactions of the Zanamivir-protein complex.

Currently, computer aided drug design is an essential tool for academia and the pharmaceutical industry (Dalkas et al., 2012). It is estimated that computer aided design will account for approximately 20% of all pharmaceutical R&D expenditure by 2016 (Kapetanovic, 2008). With the increasing development and improvement in software, user interface design and availability of faster and cheaper computers, it is likely that computer aided design will contribute even more to R&D in the future (Liao et al., 2011).

## **1.80 Thesis Aims**

The main aim of this thesis is to develop chemical tools targeting specific sialidases for use in chemical biology experiments. Within the lab, the sialidase enzymes of current focus are the *S. pneumoniae* sialidases (NanA, NanB and NanC) and TcTS. The primary objective of this PhD is to develop novel sialidase inhibitors/chemical tools of NanB and TcTS.

## **1.81 Development of a NanB “relaxed” chemical tool**

An allosteric site was discovered previously within the group. A “relaxed” chemical tool called Optactin was designed, synthesized, tested and characterised previously by Dr Brear. The aim of the work relating to the development of a NanB chemical tool follows on from the work performed by Dr Brear and is as follows:

- 1) Firstly, the groundwork for testing and synthesis of new analogues needs to be established. The aim of this work was to synthesise, express and purify the materials required for testing new analogues as well as confirming previously observed results. Synthesis of Optactin was required for use as an experimental comparison and to confirm previous results. In Chapter 2, Optactin was re-synthesised and the synthetic route optimised. NanB was sequenced, expressed and purified. The expression, purification and activity of NanB were evaluated. Optactin was tested against NanB.
- 2) Secondly, modification of Optactin based on the identified area of optimisation was performed. The aim of this work in Chapter 3 was to improve on the current inhibitor and assess the function of the water channel on enzymatic activity including its overall impact on pneumococcal infection. In this chapter, an alternative series was developed based upon mutant studies. Optimisation of this alternative series was aided by rational design. These analogues were synthesized and tested against NanB<sub>WT</sub>. The lead inhibitor was evaluated for promiscuity/off target effects and activity. This included kinetic and binding analysis. The role of the catalytic domain in pneumococcal adherence and invasion was tested using an *in vitro* assay.

### **1.82 Development of a TcTS “constrained” chemical tool.**

The aim of this project was to develop or provide the ground work for novel TcTS chemical tool discovery. The tools developed followed the constrained approach to chemical tool design. Preliminary work was performed by Dr Telford. In this work the Maybridge fragment library was screened against TcTS. This screen did not discover any valuable “hits” and so a different approach to TcTS inhibitor discovery was pursued. Based on the identification of a novel allosteric site within NanB and using the structure solved by Dr Telford, a CADD approach was used to determine if an allosteric site on TcTS could be identified.

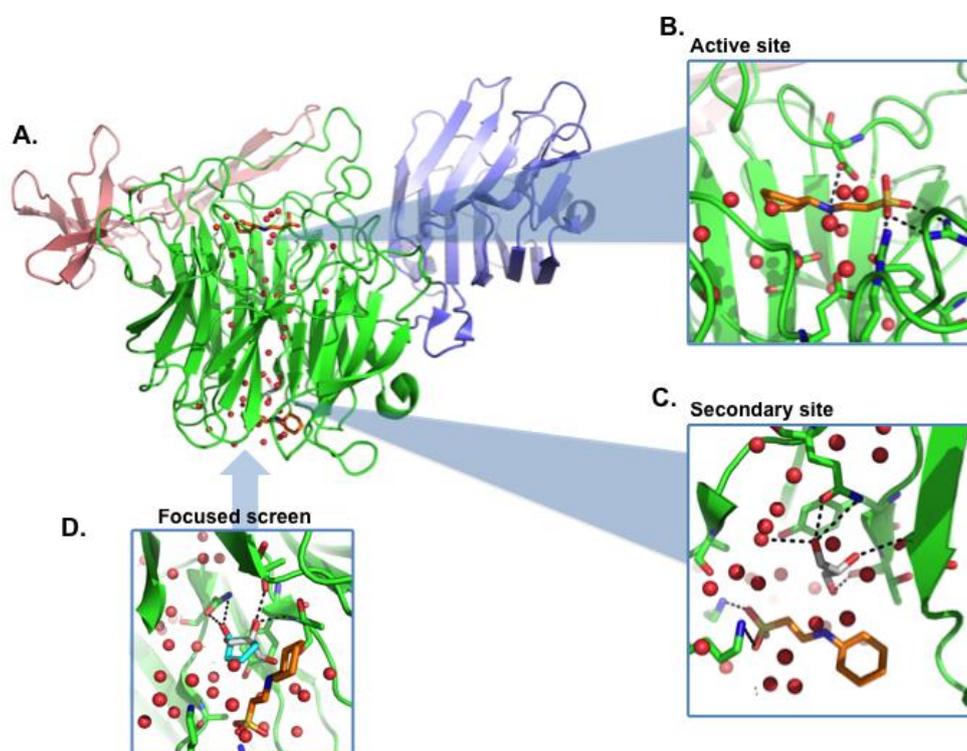
- 1) In chapter 5, the water channel in TcTS is compared to that of NanB. The residues within this site are profiled against NanB and a potential allosteric site is modelled for use in *in silico* screening. A CADD approach and a fragment library were used to identify potential fragments that would bind to this site. To decrease computational load and enrich the virtual library a ligand similarity screen was performed. The top three fragments were then screened using the 4-Munana assay against TcTS.
- 2) In chapter 6, following on from the work by Dr Telford, a CADD approach with a large virtual library from a free resource was used to identify novel active site binders. The speed of the docking process was improved with the use of a workflow environment (Knime), which was a new technique recently promoted with the BioSolveIT software. A multi-docking approach was used as a validation of the docking. This approach identified chemical moieties important for binding and inhibition of the active site and will lead the ground work for the generation of a potent constrained TcTS chemical tool.

## 2.0 NanB and the development of an allosteric chemical tool

This chapter details the discovery and synthesis of a negative allosteric modulator of NanB, called Optactin. Optactin was designed by Dr Brear using the relaxed approach to chemical tool design. Within this work the synthesis of Optactin was achieved and the synthetic route was optimised. Optactin was proved to inhibit NanB, but at a much lower potency than previously reported. Attempts to determine the reasoning behind this discrepancy were attempted within this chapter.

### 2.1 The story so far

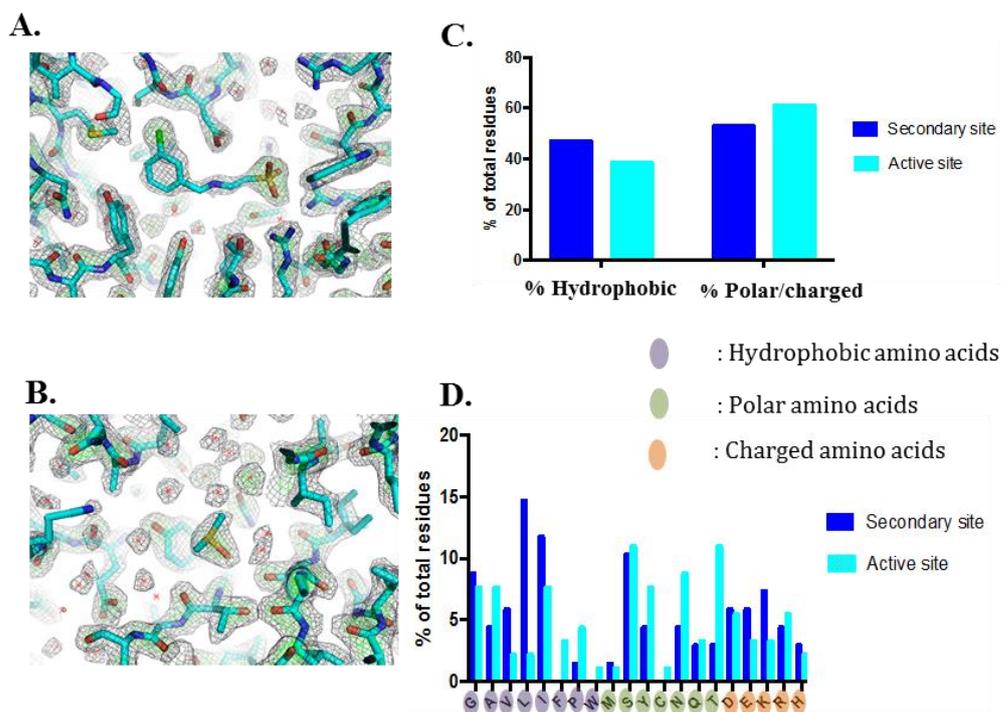
Within the crystal structure of NanB, solved by the Xu *et al.*, 2008 (2VW2), two molecules of CHES were found bound serendipitously (Xu et al., 2008). One molecule of CHES was observed within the active site cavity. Another molecule of CHES was located within the water channel (Figure 31). This water channel is around 39Å long with a diameter of 11Å at the opening. The CHES molecule is buried approximately 4.0Å within the channel opening located 24.4Å from the catalytic tyrosine (Tyr653) in the active site (distances measured within PyMOL using PDB:2VW2).



**Figure 31.** A. Structure of NanB consisting of three domains (catalytic in green, CBM domain in blue and inserted domain in red). Water molecules (red spheres) and CHES (orange) within the catalytic domain are

shown (PDB: 2VW2). **B.** Active site of NanB (PDB: 2VW2) with CHES bound and hydrogen bond contacts highlighted (black dashed line). **C.** The secondary site of NanB with CHES (orange) and glycerol (white) bound with hydrogen bond contacts shown as black dashed lines (PDB: 2VW2). **D.** The secondary site of NanB with *cis*-cyclopentane diol (PDB: 4XHB) overlaid with glycerol and CHES from PDB: 2VW2. Figures created with PyMOL and Microsoft® Powerpoint® 14.6.

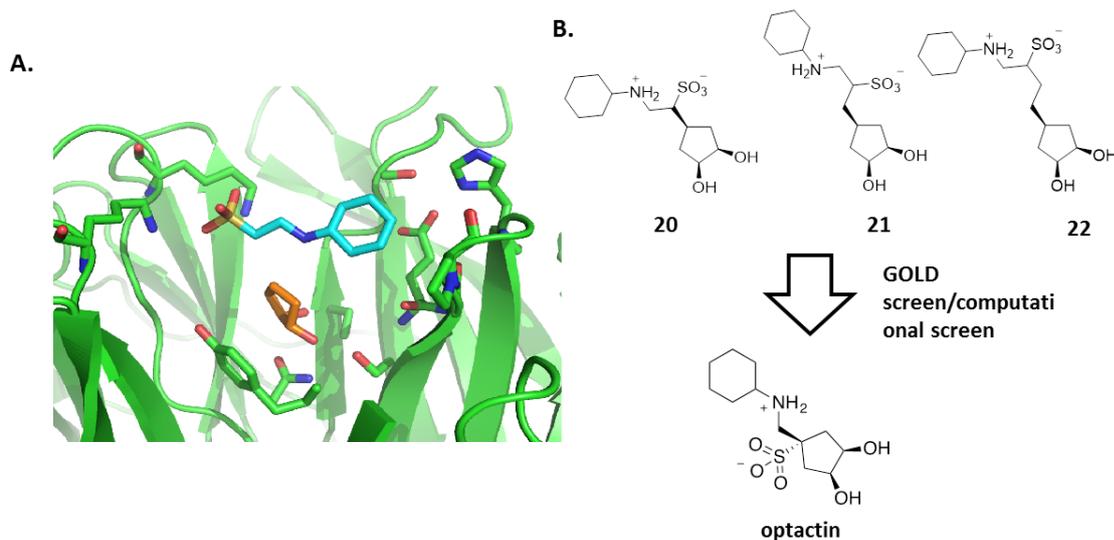
Using this structure, Dr Brear applied two approaches to the discovery of novel modulators of NanB. Firstly, the CHES molecule located within the active site was used as a basis for SAR optimization to generate a novel competitive inhibitor of NanB (NanB inhibitors) (Brear et al., 2012). A second approach was to adapt the CHES molecule bound within the secondary site to generate a secondary site-specific small-molecule binder. Water channels appear in structures of several of the sialidase superfamily and are thought to have important functional roles in substrate binding, hydrolysis and protein stability. To our knowledge, no report of targeting the water channel within the glycoside hydrolase family with a chemical tool exists. Targeting this water channel may disrupt enzymatic activity, not only confirming the importance of this channel for function but also identifying a novel allosteric site on this class of enzyme. Moreover, the binding of CHES indicates that this pocket is amenable to small molecule binding and provides a starting point for small molecule optimization. As CHES binds to both the active site and a secondary site, the aim was to design a ligand specific for the secondary site. It was thought that selectivity solely for the secondary site would be achievable as the sites (active site and secondary site) contain very different physical environments. The secondary site within NanB contains a higher proportion of hydrophobic amino acids (47%) than the active site (38%) and follows a similar trend to that of the allosteric sites reported by Li *et al.*, 2013. Furthermore, CHES has been previously adapted to produce an inhibitor 2-((3-chlorobenzyl)ammonio)ethane-1-sulfonate specific for the active site (PDB:4FPF). No observed density for this ligand was seen within the secondary site (Figure 30).



**Figure 32** **A.** 2Fo-Fc (black,  $\sigma = 1.5$ ) and Fo-Fc ( $\sigma = 3.0$  (green),  $\sigma = -3.0$  (red)) electron density maps of NanB active site with selective inhibitor, 2-[(3-chlorobenzyl)amino]ethanesulfonic acid bound (PDB: 4FPF). **B.** Secondary site of NanB with observed density for DMSO and water molecules. No electron density for 2-[(3-chlorobenzyl)amino]ethanesulfonic acid was observed. **C.** Bar chart of the percentage of hydrophobic and polar/charged amino acid residues located in the secondary site (blue) and active site (teal). The secondary site contains a similar number of hydrophobic amino acids as polar/charged amino acids. **D.** Bar chart of each amino acid residue located in the secondary site (blue) and active site (teal) as a percentage of the total amino acid residues within each site. Figures created within PyMOL® and Microsoft® Powerpoint® 14.6. A radius of 12 Å from CHES within the active site and the secondary site was used to isolate the amino acids within each site. The amino acids were tallied and percentages of each calculated.

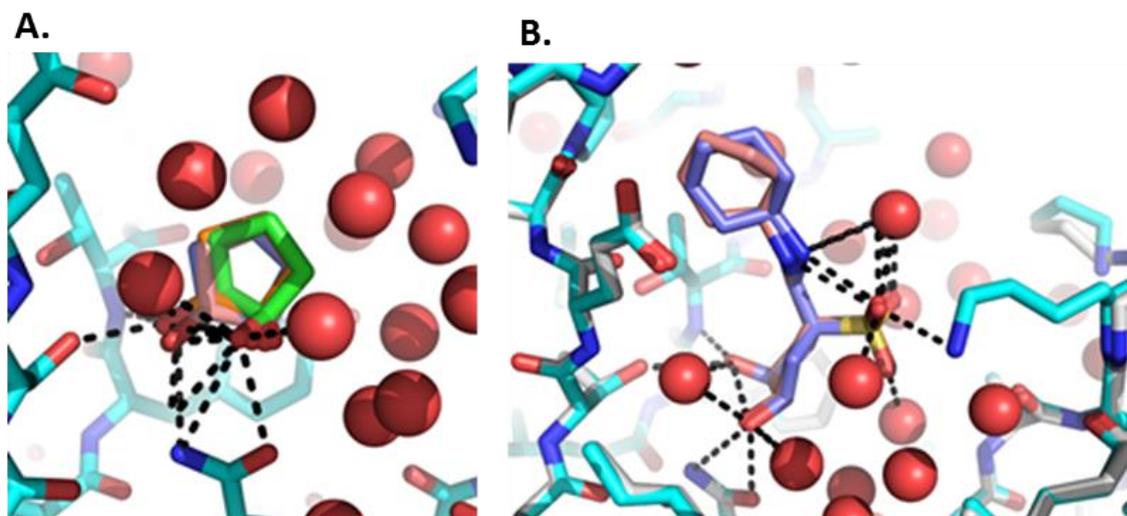
A molecule of glycerol was located within the crystal structure of NanB (4.0Å from the secondary site CHES). No molecule of glycerol was observed within the active site. Upon scouting other published crystal structures of NanB, ethylene glycol was found by Dr Brear to bind within a similar region to glycerol. Low inhibition of NanB was observed in the 4-Munana assay by glycerol and ethylene glycol at 5mM (NanB activity was inhibited  $11.3 \pm 5.9\%$  and  $2.1 \pm 3.7\%$  respectively) (Brear, 2012). Thus indicating that, firstly, this site might be allosteric and, secondly, this environment favours interactions with ligands containing diols and triols. Using this information for generating selectivity, a small focused library of 13 diol and triol compounds was screened (Brear, 2012). Of these 13 alcohols, only *cis*-cyclopentane-1,2-diol was observed to bind within this secondary site through crystal

soaking. In this crystal structure, *cis*-cyclopentane diol makes hydrophobic interactions with a Tyr250 as well as hydrogen bond interactions with Gln494, Thr657 and Thr251.



**Figure 33.** **A.** Crystal structure of overlaid structures of *cis*-cyclopentane diol (orange) and CHES (teal) bound to the allosteric site. **B.** Examples of structures computationally screened against the allosteric site. Optactin was identified as the best binder compound from this computational screen. Figures created within PyMOL®, Microsoft® Powerpoint® 14.6 and ChemDraw®.

Using the relaxed approach to chemical tool design a computational screen was applied to generate a secondary site-specific inhibitor. CHES and *cis*-cyclopentane diol were used as scaffolds (Figure 34). Combinations of various compounds were created *in silico* and docked against the secondary site. GOLD (Jones et al., 1997) was used as the program to identify the compound most likely to conform to the ligand-protein contacts of CHES and glycerol. To validate this program as an appropriate docking approach, *cis*-cyclopentane diol was docked and compared against the experimentally observed binding conformation. The docking pose was observed to be similar to the ligand-binding pose within the crystal structure (Figure 34). The top 3 GOLD solutions out of 7 had RMSD values within 1.5 Å. The best solution had an RMSD of 0.75 Å.

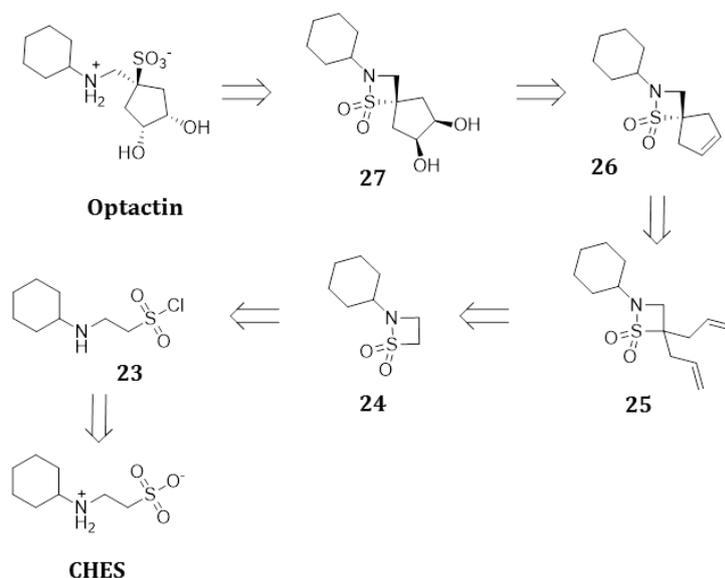


**Figure 34** **A.** Experimental binding pose of cis-cyclopentane diol from crystal structure (PDB: 4XHB) overlaid with the best pose from a GOLD computational screen (RMSD: 0.75 Å) **B.** Optactin (purple) bound within the secondary site of NanB (teal)(PDB: 4XHX) overlaid with Optactin (pink) bound within the secondary site of K499G NanB (grey) (PDB: 4XMA). Hydrogen bonds are shown by black dashed lines and water molecules represented by red spheres. Figures created within PyMOL® and Microsoft® Powerpoint® 14.6.

Screening of the small ‘virtual library’ identified 1-((cyclohexylamino)methyl-*cis*-3,4-dihydroxycyclopentanesulfonate (Optactin) to have the most optimal binding position. Synthesis of Optactin proceeded *via* a six step route. Optactin was reported to have an  $IC_{50}$  of  $79.5 \pm 6.5 \mu M$ . Crystal soaking of NanB with Optactin resulted in a ligand-protein complex. Optactin was observed to bind within this site confirming it as a novel allosteric site. The sulfonic acid within this site was buried 3.1Å deeper than the sulfonic acid of CHES and appears to form a salt bridge with Lys499. Optactin was observed to be equipotent against NanB and a K499G NanB mutant confirming the sulfonic acid as redundant and unnecessary for activity. Dr Brear carried out this work in both the Westwood and Taylor labs.

## 2.2 Resynthesis of Optactin

The synthetic scheme developed by Paul Brear was used and adapted to improve on the overall synthetic yield over six steps. The retrosynthetic approach to Optactin is shown in scheme 1.

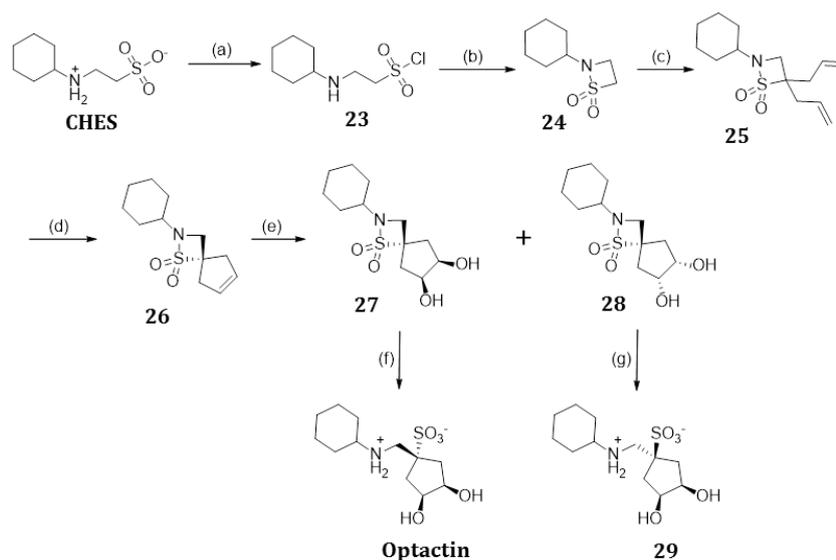


**Scheme 1.** Retrosynthesis to Optactin from starting material CHES.

CHES was used as the starting material in the synthetic approach to Optactin. Although CHES is a cheap starting material, generating aminoalkylsulfonic acid intermediates en route to Optactin could pose a synthetic challenge. Aminoalkylsulfonic acids in solution can exist in multiple ionization states (Long et al., 2010). CHES, an aminoalkylsulfonic acid is zwitterionic possessing both an anionic sulfonic acid (deprotonated in aqueous solution) and a cationic secondary amine (pKa of 9.5) (Fuguet et al., 2008). CHES and generated aminoalkylsulfonic acid molecules will be hydrophilic molecules, limiting the use of hydrophobic solvents in the synthetic scheme. It was decided therefore that the sulfonic acid of CHES would be masked by conversion into a  $\beta$ -sultam. This  $\beta$ -sultam could then be ring-opened generating the sulfonic acid in the last step.

In the first step of the synthesis, CHES was reacted with  $\text{PCl}_5$  to form the sulfonyl chloride in excellent yield. The sulfonyl chloride was then cyclized in the presence of a base to form the  $\beta$ -sultam. Deprotonation on the most acidic carbon occurs with LDA and this is functionalised upon addition of allyl bromide to produce a diallylated species. Ring closing metathesis was then performed on **25** to produce a spiro product, **26** in moderate yield. Dihydroxylation of **26** in the presence of osmium tetroxide formed two meso compounds **27** and **28**. Separation of **27** and **28** was

achieved by flash column chromatography in moderate (60%) to low (8%) yields respectively. Hydrolysis and ring opening of the  $\beta$ -sultam compound **27** was achieved by microwave irradiation with water. Recrystallisation afforded Optactin in moderate yield (45%). Successful synthesis of 16 mg of Optactin in an overall yield of 8.5% over 6 steps prompted us to optimize the synthetic route (previously reported overall yield by Dr Brear for Optactin using this route was 2.9% over the 6-steps).

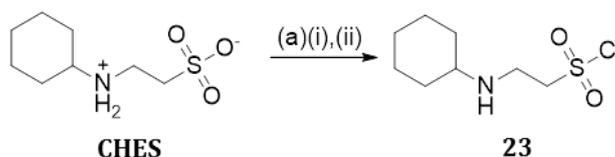


**Scheme 2.** Synthesis of Optactin: Reagents and conditions (a)  $\text{PCl}_5$ , DCE, reflux, 2hrs, 98%; (b)  $\text{Na}_2\text{CO}_3$ , EtOAc, RT, 48hrs, 84%; (c) LDA (2.5 eq), allyl bromide (4.0 eq), THF, 2hrs,  $-78^\circ\text{C}$ , 80%; (d) Grubbs 1<sup>st</sup> generation catalyst (0.05%mol), DCM, RT, 3hrs, 51%; (e) cat.  $\text{OsO}_4$ , NMO, THF:H<sub>2</sub>O (9:1), RT, 16hrs ((27:60%), (28:8%)); (f) H<sub>2</sub>O, microwave irradiation,  $140^\circ\text{C}$ , 8 minutes, 42%; (g) H<sub>2</sub>O, microwave irradiation,  $140^\circ\text{C}$ , 8 minutes, 45%.

The use of 1,2-dichloroethane (DCE) in reaction step 1 seemed unnecessary and was the first step identified for modification/optimization. DCE is toxic and can cause acute effects on the human nervous system, liver and kidneys resulting in cardiac arrhythmia, pulmonary edema, respiratory depression, nausea and narcosis.

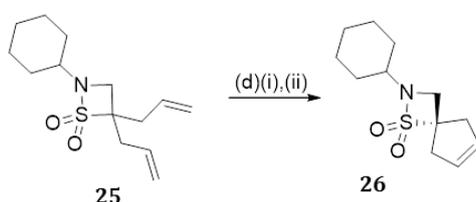
Classified as a group B2 probable carcinogen, DCE has a long-term exposure limit (LTEL) of 5ppm (Service, 2001). Considerations for solvent replacement in a reflux experiment include solubility and boiling point. Dichloromethane (DCM), a commonly used solvent in chemistry laboratories is less toxic than DCE with a LTEL of 100ppm (England, 2016). Similar polarity and dipole moments of DCM (0.309, 1.60 D) and DCE (0.327, 1.80 D) would indicate that the solubility of CHES within these two solvents would be similar (PubChem, PubChem). The boiling point of DCM is slightly lower than DCE ( $57.3^\circ\text{C}$ ) at  $39.6^\circ\text{C}$  which might impact on reaction time and/or yield (PubChem, PubChem). Reflux of

CHES with  $\text{PCl}_5$  in DCM for the same duration resulted in a lower yield (90%). However, increasing the reaction time by an hour produced a similar yield to the original conditions of 95% (Scheme 3).



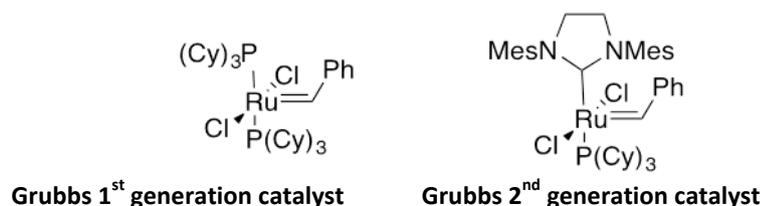
**Scheme 3.** Step 1 of Optactin synthesis: a,i)  $\text{PCl}_5$ , DCE, reflux, 2hrs, 98% (original conditions); a,ii)  $\text{PCl}_5$ , DCM, reflux, 3hrs, 95% (optimised conditions).

Reaction steps 2 and 3 required no optimization as they proceeded with excellent yield. Step 4 was the next step identified for optimization (Scheme 4).



**Scheme 4.** Step 4 of Optactin synthesis: (d,i) Grubbs 1<sup>st</sup> generation catalyst (Figure 35) (0.05 mol%), DCM, RT, 3hrs, 51% (original conditions); d,ii) Grubbs 1<sup>st</sup> generation catalyst (0.05 mol%), toluene, 40°C, 5hrs, 71% (optimised conditions).

Under the original conditions a moderate yield (51%) of spiro compound **26** was obtained. Ring closing metathesis is a carbon-carbon bond formation from two alkenes generating a cycloalkene (Grubbs et al., 1995). A number of ruthenium carbene complexes have been developed to catalyze olefin metathesis (Grubbs, 2006). The first air stable and water-soluble metathesis catalyst to be used widely in organic synthesis was the Grubbs 1<sup>st</sup> generation catalyst (Schrodi and Pederson, 2007, Schwab et al., 1995). Since then a variety of catalysts have been generated for olefin metathesis (Vougioukalakis and Grubbs, 2009). The Grubbs 2<sup>nd</sup> generation catalyst is an example of another catalyst developed for improved olefin catalysis; it is reported as more active than the first generation catalyst with a broader substrate scope (Schrodi and Pederson, 2007, Trnka and Grubbs, 2001, Scholl et al., 1999)(Figure 35).



**Figure 35.** Grubbs 1st generation and Grubbs 2nd generation catalysts.

Different temperatures, reaction time and catalyst were investigated (Table 3.). Increase in reaction time resulted in only a slight increase in yield from 51% to 54%. Further improvement in yield (64%) was obtained by increasing the temperature from 20°C to 40°C. Increase in temperature to 40°C combined with an increased reaction time improved the yield further to 71%. Changing the catalyst from the first generation to the second generation catalyst at room temperature (RT) resulted in a yield similar to the best condition identified, but due to cost this was not selected for use.

**Table 3.** Reaction conditions screened to optimize the ring closing metathesis reaction.

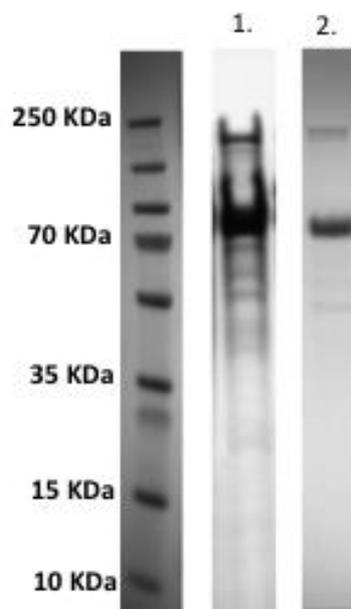
Grubbs Catalyst	Loading (mol%)	Time (hrs)	Temperature (°C)	Solvent	Yield (%)
1 <sup>st</sup> generation	0.05	3	RT	DCM	51
1 <sup>st</sup> generation	0.05	5	RT	DCM	54
1 <sup>st</sup> generation	0.05	3	40	Toluene	64
1 <sup>st</sup> generation	0.05	5	40	Toluene	71
2 <sup>nd</sup> generation	0.05	3	RT	DCM	69

A total of 487mg of Optactin was synthesized when incorporating both optimized conditions into the synthetic scheme. This resulted in a total overall yield of 11.8%, which is a 4-fold improvement from the overall yield reported by Dr Brear (2.9%).

### 2.3 Production of NanB protein

Xu *et al.*, 2008 developed the procedure used within the lab to produce NanB. This procedure used Luria Broth (LB) and isopropyl β-D-1-thiogalactopyranoside (IPTG) to produce NanB protein from a PET23b (ampicillin-resistant) plasmid (for full procedure see Experimental). LB is the most commonly used media for *E. coli* as it is easy to make, rich in nutrients and contains the optimal osmolarity for growth in early log phase (Rosano and Ceccarelli, 2014). Despite these features it is not the best media for achieving high cellular density due to lack of carbohydrates and divalent cations (Rosano

and Ceccarelli, 2014). IPTG is one of the most commonly used chemical inducers of the *lac* operon within biochemical labs. This thio-galactoside is a mimic of allolactose and inhibits the *lac* repressor *lacI* resulting in *lac* operon transcription (Politz et al., 2013). This chemical inducer is added during the mid-logarithmic phase of growth usually at an optical density (OD) at 600nm of 0.6 (Koopmans, 2009). Using this procedure 21 mg of NanB protein was expressed and purified from 4 liters of *E. coli* LB bacterial culture.

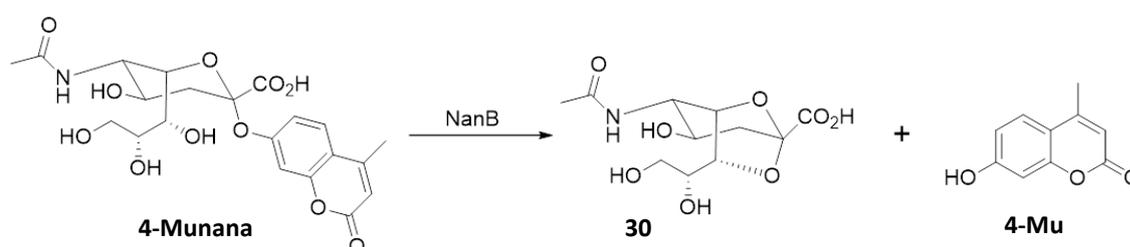


**Figure 36.** SDS page of NanB purified through a nickel column. 1. Original expression method (IPTG induction) and 2. First optimization attempt (autoinduction).

To optimize this procedure it was reasoned that the autoinduction technique commonly used within the Taylor group would result in an improved yield of NanB. Autoinduction media was introduced in 2005 and is an optimized blend of glucose, lactose and glycerol. Glucose is metabolized first as the preferred carbon source, and then glycerol and lactose (an inducer of the *lac* operon) metabolism follows in mid to late-logarithmic phase. This negates the need for biomass monitoring, culture manipulation and manual addition of the inducer (avoiding human error) (Rosano and Ceccarelli, 2014). This enriched media can produce a yield of target protein several-fold higher than IPTG induction (Studier, 2005). Protein purity was evaluated after the first purification process (a nickel column) using an SDS-PAGE gel. The autoinduction procedure produced a cleaner batch of protein compared with expression in LB after the first purification process (Figure 36). The anion exchange purification was omitted from the procedure and the protein fractions were directly subjected to a final purification step (size exclusion chromatography). Using this optimized procedure 52 mg of protein was expressed and purified from 4 liters of *E. coli* autoinduction media bacterial culture. This represents a 2.5 times increase in yield compared to the IPTG and LB broth procedure (21 mg). An

increase of 1.2g in biomass was observed for the autoinduction method for each 1L of bacterial culture (this represents an increase of 27%).

In order to determine if the protein produced using this optimised method was useful for the kinetic analysis of inhibitors, it was compared against the NanB previously produced using the original method. Activities of neuraminidases/sialidases are tested using the 2'-(4-methylumbelliferyl)- $\alpha$ -D-N-acetylneuraminic acid (4-munana) assay. This assay is described by Kongkamnerd *et al.*, 2011 and used to probe the activity of neuraminidases/sialidases. In the presence of a neuraminidase/sialidase, the substrate is cleaved into **30** and 4-methylumbelliferyl (4-Mu) (Kongkamnerd *et al.*, 2011)(scheme 5). The fluorescence intensity of the fluorophore 4-Mu generated over time is measured giving the rate of the enzyme.



**Scheme 5.** The substrate 4-munana cleaved into compounds **30** and **4-Mu**.

The activity of NanB purified and expressed from each approach was tested using the 4-munana assay at the optimum conditions of 37°C and pH 5.0. The activity of NanB differed between the two methods of expression and purification. Autoinduction produced a less active batch of protein compared with the IPTG LB method (Table 4).

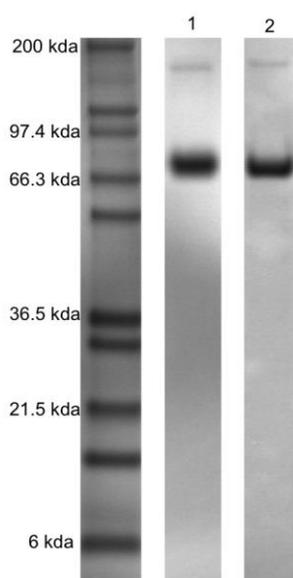
**Table 4.** Table of the enzymatic activity of NanB expressed and purified using the two approaches: IPTG (original method) and autoinduction (first attempt at optimization). Protein assayed using the 4-munana assay with 4-munana at a final concentration of 200 $\mu$ M and NanB at a final concentration of 60 ng/mL.

Protein Expression and Purification method	Mean activity (AFU/sec)
Autoinduction	250 $\pm$ 18
IPTG induction	1100 $\pm$ 25

Based upon these results protein produced using the autoinduction method was used only for crystallization. For kinetic analysis, the IPTG method was chosen for NanB expression.

An adaption to the purification method suggested by Professor Terry Smith involved purification of the protein without the use of cocktail inhibitors as this had shown to inhibit protein activity in the past. The purification method in the presence and absence of protease cocktail inhibitors was trialed. It was found that protein purified without protease cocktail inhibitors was more active ( $1440 \pm 120$  AFU/second).

Excluding protease cocktail inhibitors from the method resulted in an increased activity of 340 AFU/sec. The next step was to assess carefully the purification steps. Protein activity was assessed at each stage of the purification procedure and observed to be the most active after the anion exchange purification step ( $1886 \pm 84$  AFU/second) The purity of the protein after the anion exchange column by SDS page gel was observed to be similar to that after gel size exclusion chromatography (Figure 37). It was decided to avoid gel filtration and use protein directly after the anion exchange column.



**Figure 37.** SDS Page of NanB after anion exchange purification (1) and size-exclusion chromatography (2).

**Table 5.** Table of the enzymatic activity of NanB after each purification step.

<b>Purification Step</b>	<b>Mean activity (AFU/ second)</b>
<b>Nickel Column (step 1)</b>	281±30
<b>Desalt Column (step 2)</b>	476±16
<b>Anion Exchange Column (step 3)</b>	1886±84
<b>Gel Filtration Column (step 4)</b>	1440±120

The stability of the enzyme was measured over time. Left at 4°C the enzymatic activity remains stable for one week and left at -80°C the enzymatic activity remains stable for a month.

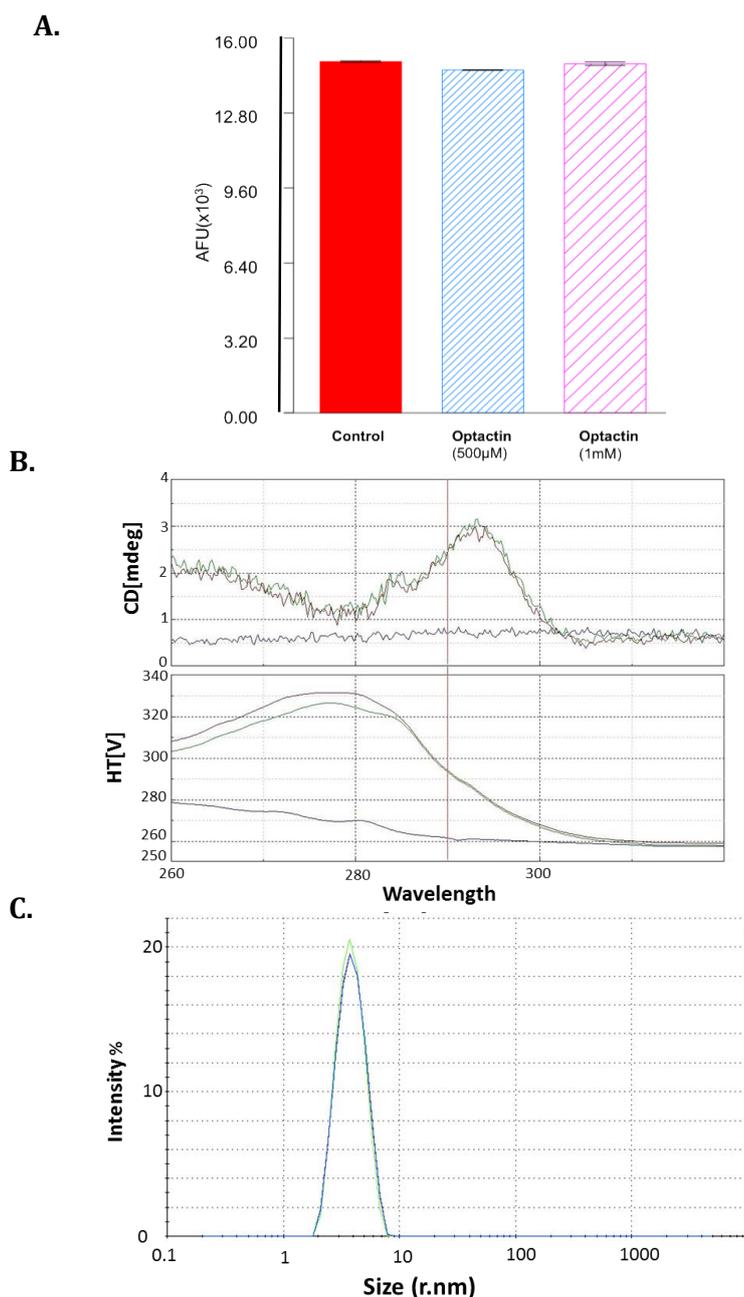
Having established an activity and stability profile of NanB, testing of Optactin against NanB was initiated.

## **2.4 Testing of Optactin**

The 4-munana assay for NanB was conducted in the presence and absence of Optactin to give the percentage inhibition. Different concentrations of Optactin were run enabling an estimation of IC<sub>50</sub>.

Unfortunately, it was not possible to reproduce the previously reported activity of Optactin in house (reported by Dr Brear to be 79.5±6.5 μM).

Repeated measurements of the IC<sub>50</sub> of Optactin against active NanB protein consistently gave values in the region of 537±45μM. The IC<sub>50</sub> curve was observed to be steep, an indication of promiscuity/"off-target" effects (Shoichet, 2006). Due to this observation and the inconsistency in the reported and obtained potency of Optactin, the promiscuity of Optactin was evaluated. Kongkamnerd *et al.*, 2011 described a method for the evaluation of quenching and promiscuity in the 4-munana screen. Compounds are assayed against 4-Mu at varying concentrations of ligand and the fluorescence measured (Kongkamnerd et al., 2011). No fluorescence quenching of 4-Mu was observed in the presence of Optactin at 500 μM or 1 mM (Figure 38).



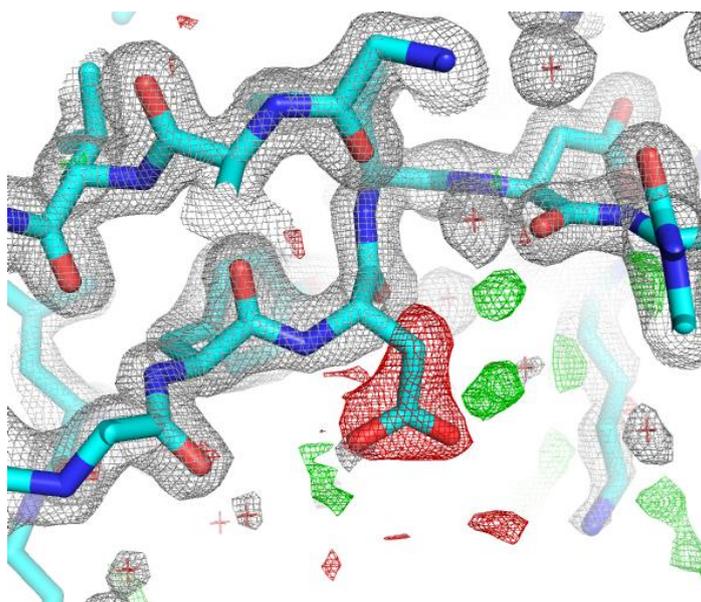
**Figure 38. A.** Graph of 4-Mu fluorescence in the presence and absence of Optactin (500  $\mu$ M and 1mM). **B.** CD near-UV trace of NanB in the presence (green) and absence (red) of 1mM Optactin. Blank (no protein/buffer only) is shown in blue. **C.** DLS chart of size distribution by intensity of NanB in the presence of Optactin (1mM). Only one peak is seen at estimated MW  $74.9 \pm 23.2$  KDa

Dynamic light scattering (DLS) is a technique used to monitor particle size in solution and is often used to determine the presence of protein aggregation in solution (Li et al., 2011). Circular dichroism measures the absorption of right hand and left hand polarised light. Spectra at near UV (250-350 nm) arise from aromatic residues and the size and intensity differs depending on the environment of

the aromatic residues (Kelly et al., 2005). These spectral features correspond to the tertiary structure (Kelly et al., 2005). Good near UV signals indicate that the protein is folded and in a well-defined structure. The spectra is also sensitive to protein-protein interactions (Greenfield, 2004). DLS and CD confirmed that the presence of Optactin did not cause protein aggregation or unfolding as a single peak/one population of particle is observed (corresponding to the molecular weight of NanB) and no change in near UV CD spectra between control and 1mM Optactin treated condition. Although the protein-ligand complex structure of Optactin-NanB shows that Optactin binds within the secondary site, we wanted further proof that Optactin is non-promiscuous/had no “off-target” effects. The results obtained here confirm that despite a steep inhibitory curve and discrepancies in potency, Optactin is a non-promiscuous and a valid allosteric inhibitor of NanB (Figure 38).

## 2.5 Identification of a single point mutation

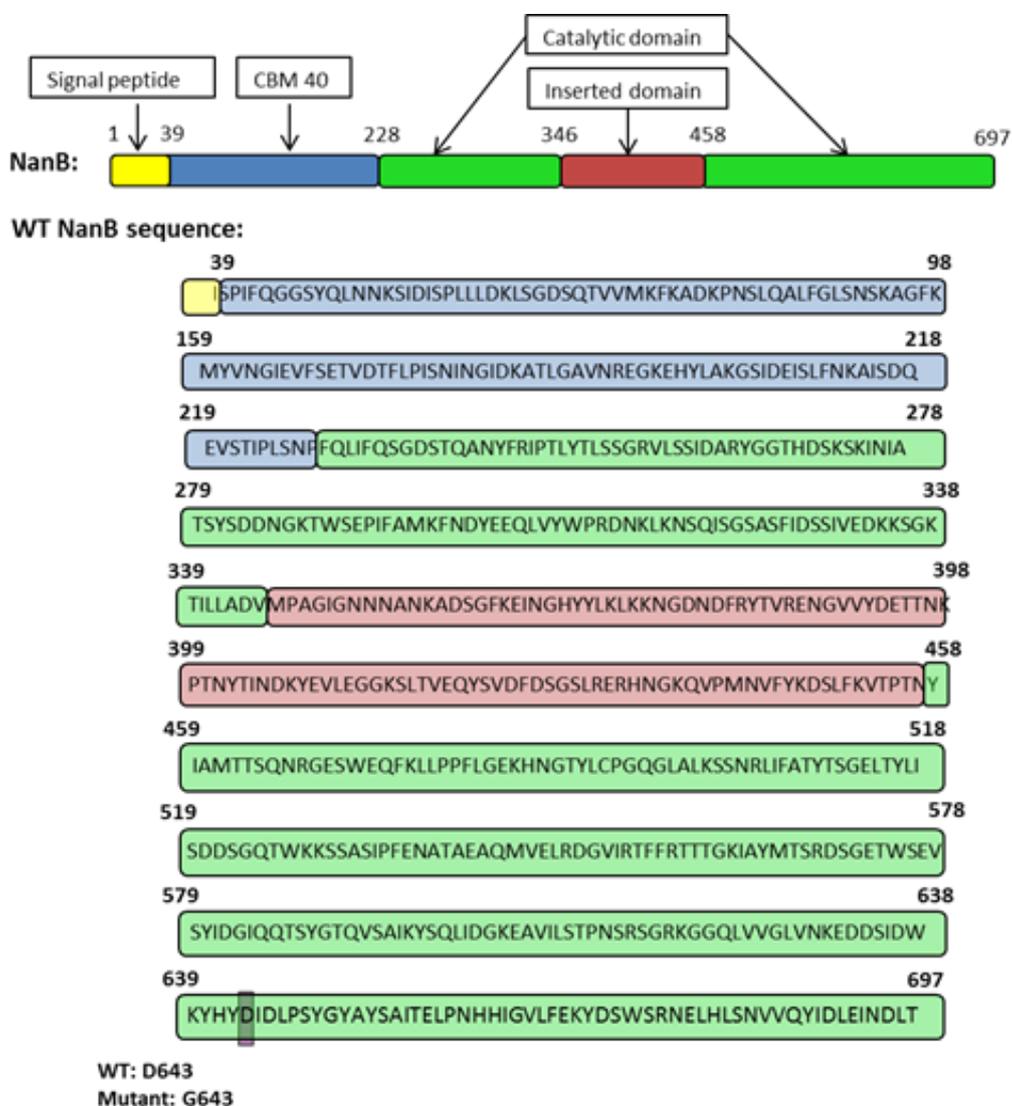
To evaluate reasons for the difference in the observed activity of Optactin the sequence of the protein was re-checked. The pET23b NanB DNA vector construct created within the Taylor lab (used previously by Xu *et al.*, 2008 for the structural identification of NanB and by Brear *et al.*, 2012 for the structural evaluation of inhibitors) was sent for sequencing. Upon evaluation of the sequence and inspection of previously published crystal structures, a single point mutation was identified at position D643G that differs to the WT sequence reported on the UniProtKB database (Figure 39).



**Figure 39.** The 2Fo-Fc (black mesh, contoured at  $1.5\sigma$ ) and Fo-Fc (green ( $3.0\sigma$ ) and red( $-3.0\sigma$ )) electron density maps for amino acid residue 643 of published structure 2VW2.

This amino acid is located on the surface of the catalytic domain 26 Å from the active site and 24 Å from the secondary site. Due to its positioning it is unlikely this will impact upon the activity of the protein or potency of the inhibitor, but this would need to be verified.

Primers were designed and a single point mutation was performed by site directed mutagenesis. The mutated construct was transformed into competent cells and colonies grown on agar plates containing antibiotic. A miniprep to purify plasmid DNA was performed on selected colonies and the new DNA construct was sent to GATC Biotech for full sequencing. The sequencing results confirmed this new construct as the WT sequence reported on the UniProtKB database (Figure 40).



**Figure 40. A.** Schematic of the NanB domains. and sequence of NanB WT with the mutation highlighted (purple box). Catalytic (green), inserted (red) and CBM40 (blue) domains highlighted. Figure created within Microsoft® PowerPoint® 14.6 and Esprint 3.0 (Robert and Gouet, 2014, Hann and Keserü, 2012, ESprint).

## 2.6 Biochemical comparison

WT NanB was expressed and purified using the optimized method for generating active D643G NanB and activity assessed and compared with the D643G mutant. The mutation had no effect on the activity of the enzyme (D643G mutant 1886±84 AFU/second and WT 1841±88 AFU/second) (Table 6).

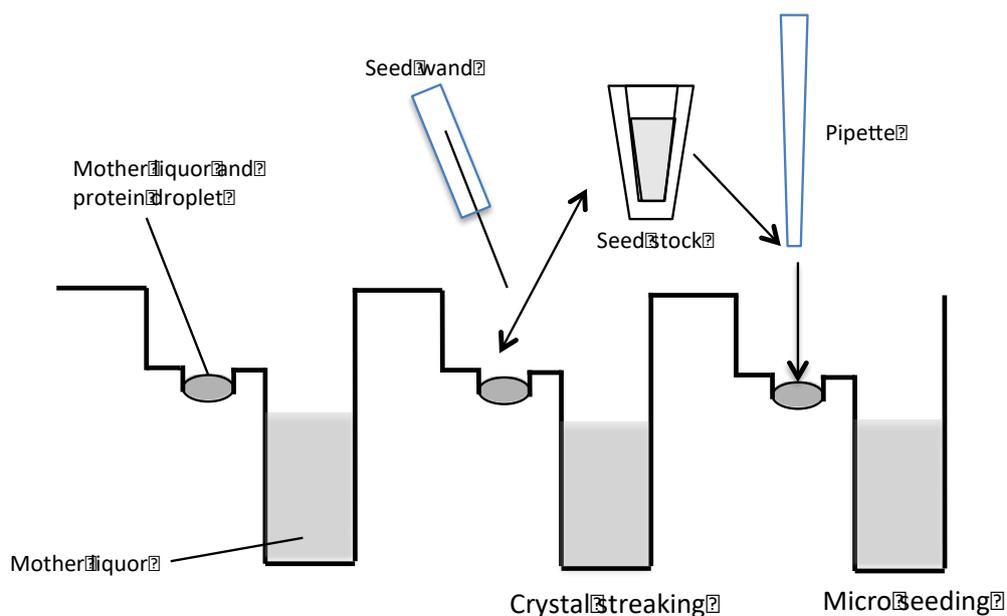
**Table 6.** Table of NanB activity for WT and D643G mutant.

Protein Type	Activity (AFU/second)
Mutant (D643G)	1886±84
Wild type	1841±88

However, the mutation was observed to have profound effects on the stability of the protein over time, with the activity of the wild type lasting for a long period after storage in -80°C, whereas the mutant would lose activity after a month. The activity of Optactin was observed to be similar when assayed against the wild type protein ( $IC_{50}$  517±12µM).

## 2.7 Crystallisation of mutant D643G NanB and WT NanB

X-ray crystallography, a technique for the structural determination of proteins and their binding sites at atomic resolution, provides valuable information for structure-based design. X-ray crystallography requires protein crystals, which are grown by bringing the macromolecule to supersaturation (Dessau and Modis, 2011). Vapor diffusion is one technique to obtain protein crystals. This involves placing a mix of protein solution and precipitant/mother liquor in a sealed chamber with pure precipitant/mother liquor (Dessau and Modis, 2011). Optimal crystal formation is influenced by a range of variables that are tested *via* crystallization trials. These trials occur in plates of 24 to 96 well formats set up to test a range of conditions chemical parameters (pH, type of buffer, additives and precipitants) (Benvenuti and Mangani, 2007). The sitting drop method is one such procedure used for protein crystallization trials (Figure 41).

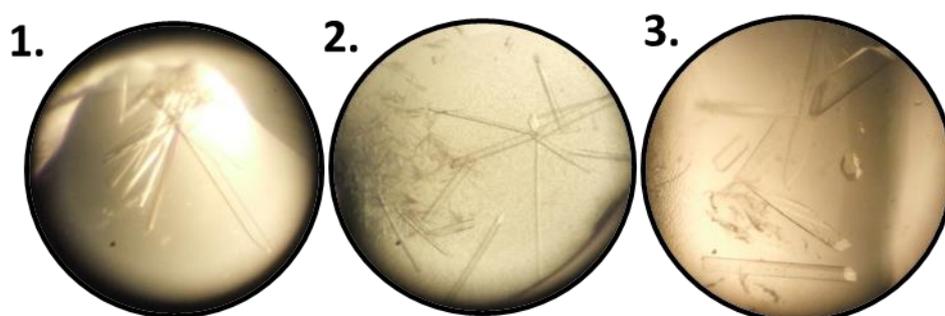


**Figure 41.** Diagram of the sitting drop method. (4c) Mother liquor typically consists of buffer and a precipitant such as polyethylene glycol (PEG). Mother liquor and protein droplet can be combined in varying ratios. Streaking of crystal seeds or pipetting of small volumes into the mother liquor and protein droplet is performed for crystal seeding.

The previous conditions reported by Xu *et al.*, 2008 (0.1 M imidazole pH 8.0 and 7% (w/v) PEG 8K) were used for the D643G NanB protein crystal trials. Trialing conditions of 0.1M imidazole pH 8.0 with PEG 8K at 3%, 4%, 5%, 6%, 7% and 8% resulted in crystals appearing in 5%, 6% and 7% after 5 minutes. These crystals were too small for use in X-ray structure determination. Three months were needed for a crystal to grow to an appropriate size for X-ray crystallography. This procedure was too long and needed to be optimised to shorten the length of time required to grow crystals of diffraction quality. To achieve this, crystal seeding was performed by pulverising a larger crystal into smaller pieces and then transferring these crystalline particles into a new drop by the use of a seeding wand (Stura and Wilson, 1991). These crystalline particles provide nucleation points for crystallisation and potentiated crystal growth. Seeding was trialed at 5%, 7% and 10% PEG 8K at pH 8.0. In the 5% and 7% conditions seeding produced diffraction quality crystals in a period of two weeks. The 7% PEG 8K trials produced appropriate size crystals in most wells and proved to be the best condition.

In order to generate a WT NanB X-ray structure, new crystallization conditions were required, as the WT protein did not crystallize in the conditions used for D643G NanB. Four commercial screens (JCSG+, Anion Suite, Salt Rx and Cryo I and II screen) were tested in a 96 well format.

For each condition, 0.1  $\mu\text{L}$  volumes of WT NanB protein were dispensed into separate drops of 0.1  $\mu\text{L}$  and 0.2  $\mu\text{L}$  of mother liquor. The JCSG+ screen was the most successful after two weeks, generating crystals in five conditions. Of these five conditions bicine pH 9.0 was found to regularly produce crystals of suitable size (0.1 M bicine pH 9.0, 20% PEG 6K). However, these crystals diffracted poorly (only to 4  $\text{\AA}$ ). In order to generate crystals with improved diffraction quality, these crystals were used for seeding experiments. A 96 well screen with a range of concentrations of bicine (0.02-0.30M at pH 9.5) and PEG 6K (8–30% w/v) were set up with protein concentrations ranging from 5 mg/mL to 15 mg/mL. Unfortunately, this did not lead to crystals with improved diffraction. The JGSC+ commercial screen was then re-trialed with micro seeding. Micro seeding is a similar process to crystal seeding except a volume of seed stock is added directly to the protein droplet instead of using a seeding wand (Bergfors, 2003).

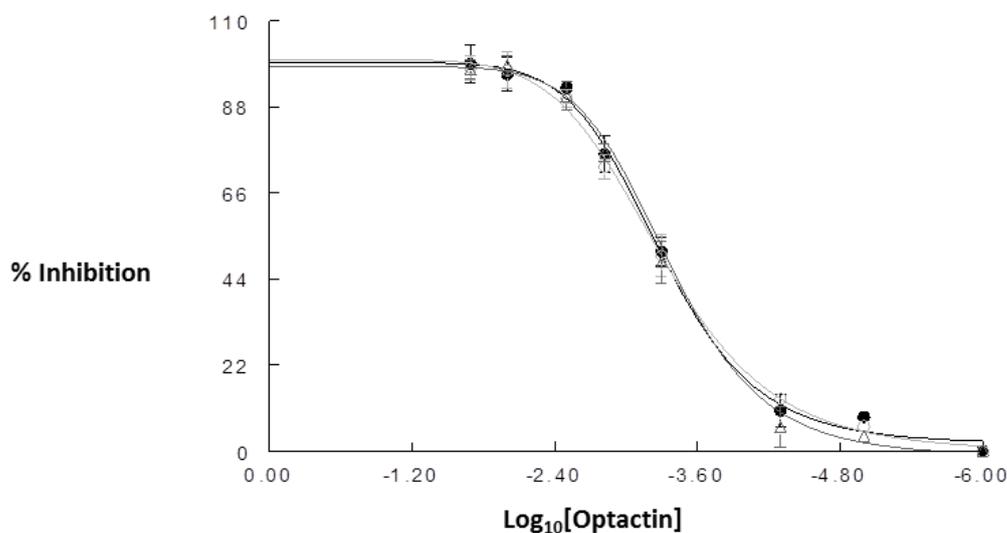


**Figure 42.** Images taken of crystals grown in the initial hit 1: 0.1 M bicine pH 9.0 20% PEG 6K, and conditions when crystal seeding was performed 2: 0.1 M CHES pH 9.5, 20% PEG 8K and 3: 0.3 M imidazole pH 8.0 12% PEG 8K.

The number of hit conditions increased to 17. The best crystals grew in 0.1 M CHES pH 9.5, 20% PEG 8K and diffracted to 2  $\text{\AA}$ . However, this crystal condition resulted in CHES within the active site. These crystals were used to generate a seed stock and used in crystal seeding experiments. With a good crystal seed stock the imidazole and PEG 8K conditions were revisited. A screen of ranging imidazole pH 8.0 (0.02-0.5 mM) and PEG 8K (3%-10%) concentrations were set up. Crystals of an appropriate size grew in 0.3M imidazole pH 8.0, 12% PEG 8K after two weeks and diffracted to 1.8  $\text{\AA}$ . The WT structure was solved and observed to be identical to the D643G structure apart from observed density for the aspartic acid. Both proteins were used for X-ray crystallography and generation of ligand-protein complexes.

## 2.8 Retesting of Optactin

Further attempts to explain the discrepancy in the measured activities of the inhibitor prompted a new evaluation of the purity of Optactin. This was achieved by repeated re-crystallisations followed by testing against NanB. Another approach was to evaluate the substrate. The substrate 4-munana is known to breakdown over time and therefore fresh substrate 4-munana was ordered from Carbosynth and tested.



**Figure 43.**  $\text{IC}_{50}$  curve of re-crystallised Optactin (O) tested against WT NanB ( $\text{IC}_{50} = 517 \pm 12 \mu\text{M}$ ). Optactin tested against D643G (●) ( $\text{IC}_{50} = 537 \pm 45 \mu\text{M}$ ) and tested against fresh 4-Munana substrate (Δ) ( $\text{IC}_{50} = 550 \pm 11 \mu\text{M}$ ). Data points and error bars represent the mean and standard error of the mean of 4 repeats. Graph created using SLIDE®.

The activity of Optactin was not observed to change under any of the re-evaluated conditions (Figure 43). Despite the inconsistency in Optactin's potency reported and observed here, it is evident that Optactin is a negative allosteric modulator of NanB and the first reported allosteric modulator of this protein class. This chemical tool will be further developed through rational design, hopefully leading to an improved understanding of the role of the water channel and sialidase impact on pneumococcal infection.

## **2.9 Summary**

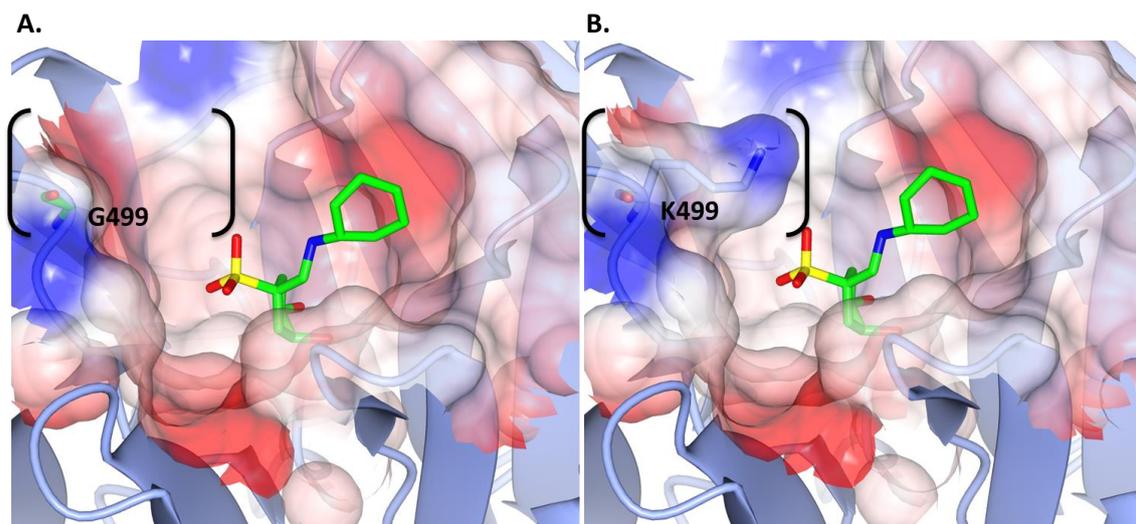
Dr Brear identified a novel allosteric site within NanB. However, this initial work was performed on a mutant variant of NanB. A spontaneous mutation was observed within the sequence at position 643 where an aspartic acid is mutated to a glycine. Although the positioning of this mutation is within the catalytic domain, it is located 26 Å away from the catalytic site and 24 Å away from the allosteric site. The position of this mutation is unlikely to impact on the results observed by Dr Brear. However, to confirm an allosteric site exists within NanB and that the mutation has no impact on the results found, the experiments were repeated on NanB<sub>WT</sub>. A point mutation was performed and after confirmation of the correct sequence, the NanB<sub>WT</sub> protein was expressed, purified and crystallised. The originally developed inhibitor, Optactin, was re-synthesised and the synthesis was optimized to improve overall yield (an overall improvement of 4 times from 2.9 % to 11.8 % over 6 steps). Analysis of the activity of Optactin using the 4-Munana assay determined the potency of Optactin to be weaker than by Dr Brear despite repeated attempts.

### **3.0 Development of alternative series**

This chapter is on the development and optimisation of the NanB allosteric modulator, Optactin. The aim of this work was to develop a potent chemical tool for the elucidation of NanB's role in *S.pneumoniae* pathogenicity. Within this chapter a NanB<sub>WT</sub>-Optactin crystal structure has been solved and used for structure based design to improve potency and selectivity of a "relaxed" chemical tool against NanB. Protein mutagenesis of a lysine (thought to form a salt bridge with the SO<sub>3</sub><sup>-</sup> of Optactin) to a glycine did not reduce the potency of Optactin suggesting that the SO<sub>3</sub><sup>-</sup> is redundant and has no role in binding or activity. The SO<sub>3</sub><sup>-</sup> functional group of Optactin was the first point identified for optimisation. Synthesis of an amine non-sulfonic acid analogue resulted in improved ligand efficiency and the fortuitous discovery of another analogue, an amide that had greater potency than the amine equivalent. CADD and SAR further optimized this amide to generate a low micromolar inhibitor, Optactamide. Control experiments were performed to test the selectivity and promiscuity of Optactamide, validating this molecule as a useful chemical tool. Use of this chemical tool in an *S.pneumoniae* adhesion and invasion assay identified NanA and NanB sialidase activity to be important in the invasion of *S.pneumoniae* into lung epithelial cells, but the exact role of NanB in *S.pneumoniae* pathogenicity still remains unclear.

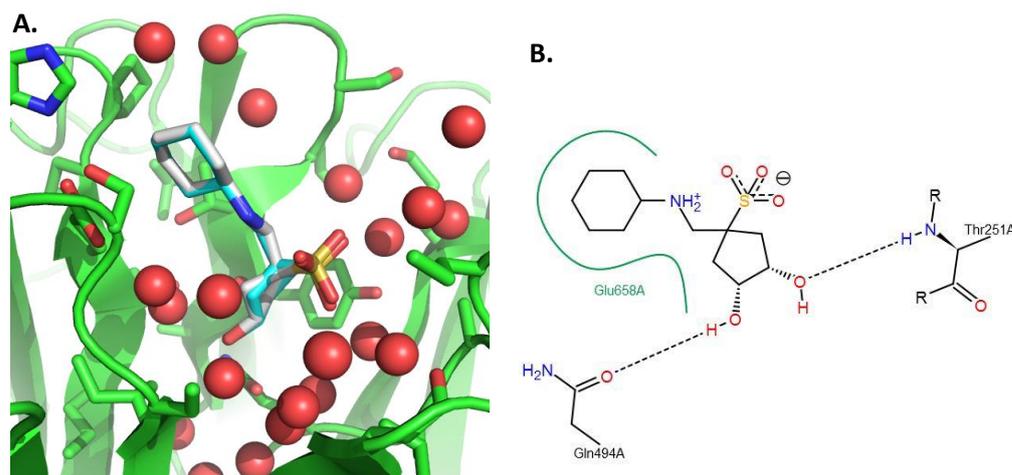
#### **3.1 Previous Mutant studies**

Dr Brear in his initial analysis of the NanB-Optactin crystal structure identified that three residues formed key van der Waals contacts with Optactin. One of the van der Waals contacts was identified as obsolete from mutagenesis studies of NanB<sub>D643G</sub>. The lysine at position 499 within NanB<sub>D643G</sub> was mutated to a glycine and activity of Optactin was tested against this mutant using the 4-Munana assay. From this assay it was observed that Optactin had similar activity against NanB<sub>D643G</sub> and its K499G mutant. A crystal structure of the K449G mutant of NanB<sub>D643G</sub> generated by Dr Brear identified that the binding mode of Optactin is identical in both mutants (Figure 44). The RMSD of NanB<sub>D643G</sub> (PDB: 4XJ9) against NanB<sub>D643G,K499G</sub> (PDB: 4XMA) is calculated to be 0.35 Å.



**Figure 44.** **A.** Binding mode of Optactin (green) within the Optactin-NanB<sub>D643G,K499G</sub> (PDB: 4XMA) crystal structure. Electrostatic surface potential is shown calculated within CCP4MG QTMG. **B.** Binding mode of Optactin within the Optactin-NanB<sub>D643G</sub>(PDB: 4XJ9) crystal structure. Electrostatic surface potential is shown calculated within CCP4MG QTMG. Within both images the side chain of residue 499 is labelled and highlighted by parentheses.

The approximate position of the <sup>+</sup>NH<sub>3</sub> from K499 side chain is 3.6 Å from the SO<sub>3</sub><sup>-</sup> of Optactin (PDB; 4XJ9). Salt bridge contacts occur over distances of less than 4 Å (Kumar and Nussinov, 2002) and so is thought to form as an interaction between the SO<sub>3</sub><sup>-</sup> of Optactin and the <sup>+</sup>NH<sub>3</sub> of K499. It is unlikely given the similarity of the WT structure to the NanB<sub>D643G</sub> mutant that the spontaneous mutation at position 643 would impact on this result. However, the crystal structure of Optactin-NanB<sub>WT</sub> was solved to confirm the binding position of Optactin and the proximity of K499 to the sulfonic acid of Optactin within the WT structure. Within this Optactin-NanB<sub>WT</sub> complex structure, the B-factor for the lysine residue is high relative to the surrounding residues indicating flexibility and movement of the residue. No complete electron density was observed for the K499 side chain indicative of no fixed position for the <sup>+</sup>NH<sub>3</sub> of K499. It is therefore unlikely that an electrostatic interaction occurs between the <sup>+</sup>NH<sub>3</sub> of K499 and the SO<sub>3</sub><sup>-</sup> of Optactin (Figure 45). The RMSD of the Optactin-NanB<sub>WT</sub> complex solved here against the Optactin-NanB<sub>D643G</sub> mutant is 0.342 Å (RMSD of the full NanB structure) (calculated using CCP4MG superpose).

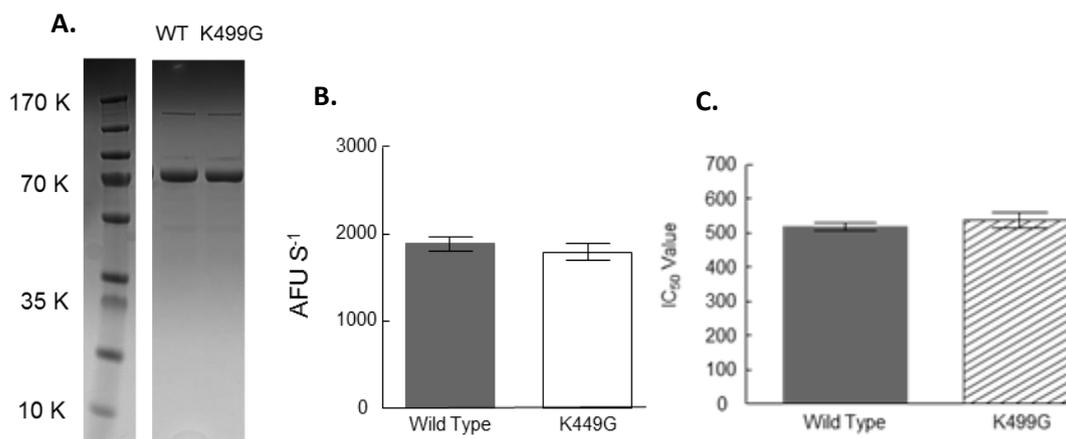


**Figure 45.** **A.** Optactin-NanB<sub>WT</sub> complex (grey) with an identical binding pose as Optactin-NanB<sub>D643</sub> crystal complex (teal, PDB:4XJ9). **B.** A 2D binding image of Optactin generated by PoseView using the Optactin-NanB<sub>WT</sub> crystal structure (Stierand et al., 2006).

PoseView (Stierand et al., 2006) was used to predict the van der Waals forces between Optactin and the interacting residues of NanB. PoseView is an interaction prediction software that uses atom coordinates from ligand-protein crystal structures to predict hydrogen bonding (black dashed lines), salt bridges (black dashed lines), metal interactions (black dashed lines), hydrophobic interactions (green solid lines),  $\pi$ - $\pi$  (green dashed lines) and  $\pi$ -cation (green dashed lines) interactions between ligands and interacting protein surfaces (Stierand et al., 2006). Using this software, hydrogen bond interactions were predicted to occur between the diol of Optactin and the backbone amide N-H of Thr251 and the side chain carbonyl of Gln494. A hydrophobic interaction was predicted to occur between the cyclohexyl ring of Optactin and NanB. A negatively charged Glu658 residue occupies the pocket that the cyclohexyl ring binds in. This indicates this interaction maybe suboptimum. PoseView did not identify a salt bridge between  $\text{SO}_3^-$  of Optactin and the  $^+\text{NH}_3$  of K499 from the Optactin-NanB<sub>WT</sub> complex. PoseView evaluates interactions based on geometric criteria and so this interaction must not fulfil this criteria. To further confirm these structure based results and the redundancy of the  $\text{SO}_3^-$  group within Optactin, mutagenesis of residue 499 was performed.

### 3.2 Repeated mutant studies on WT

Primers were designed and a single point mutation of the K499 was accomplished using site directed mutagenesis. The K499G mutant was expressed and purified using the same L-B method for NanB<sub>WT</sub> (Chapter 3.). NanB<sub>K499G</sub> had a slightly lower activity than NanB<sub>WT</sub> at 1653±130 and 1841±88 AFU S<sup>-1</sup> respectively.



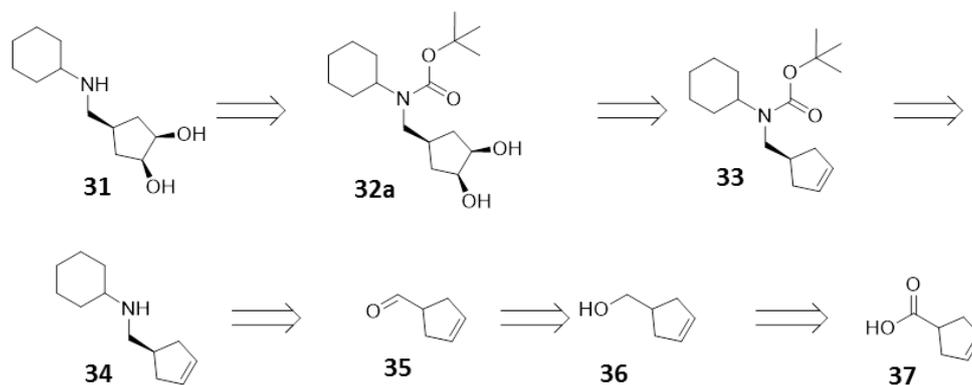
**Figure 46.** A. SDS page gel of WT and K499G after purification. B. The activity of NanB<sub>WT</sub> and NanB<sub>K499G</sub> using the 4-Munana assay. C. The IC<sub>50</sub> values of Optactin against NanB<sub>WT</sub> (IC<sub>50</sub> value = 517±12 μM) and NanB<sub>K499G</sub> (IC<sub>50</sub> value = 534±22μM).

Analysis of the K499G mutant using the 4-Munana assay identified that Optactin has a similar potency to that of NanB<sub>WT</sub> (Figure 46). The NanB<sub>K499G</sub> mutant studies performed here suggest that the sulfonic acid did not contribute to the potency of Optactin. Based on the result from this mutagenesis experiment and the structure based design observations from the crystal structure, the sulfonic group is an unnecessary functionality for activity and binding and should be the first point of modification for inhibitor optimisation.

### 3.3 Design and development of simplified analogues

The relaxed approach to chemical tool development does not need to follow medicinal chemistry guidelines for bioavailability. Achieving high potency and selectivity are the major design criteria for relaxed chemical tool development. Removing functional groups that do not contribute to binding is essential to improve selectivity and identify areas for optimisation for improved potency. Ligand efficiency (LE) is an *in vitro* based method for the comparison of the binding energy or potency and atom count of small molecules. This metric quantifies the molecular properties required to obtain binding affinity and/or potency. It has expanded to include other parameters such

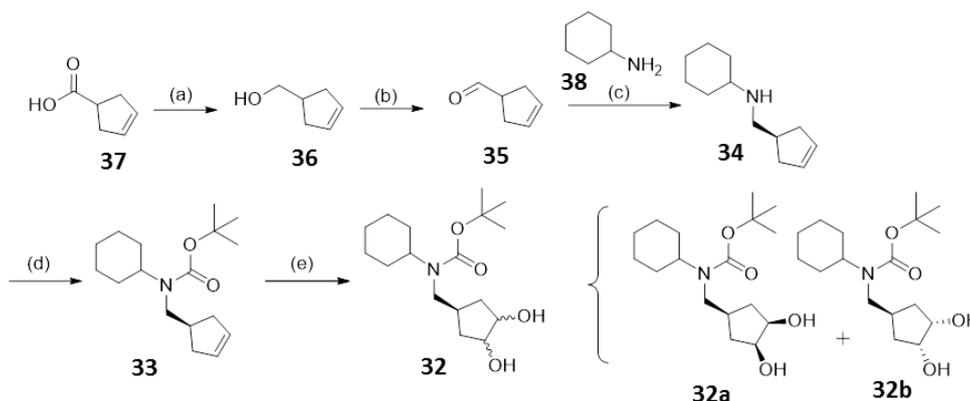
as molecular mass, lipophilicity, polar surface area and physicochemical properties. LE is useful in SBDD to clarify if removal or replacement of certain functional groups is beneficial for optimisation. Currently the LE for Optactin is 0.095 (using the LE equation:  $LE = 1.4(-\log IC_{50})/N$  (N is the number of non-hydrogen atoms)). Substitution of the  $SO_3^-$  to a hydrogen atom would likely improve ligand efficiency. To generate this non- $SO_3^-$  analogue, it was thought that the synthesis should proceed *via* a Boc-protected amine (Scheme 6) to enable purification through silica gel column chromatography prior to removal of the Boc-protecting group.



**Scheme 6.** Retrosynthetic scheme to amine 31.

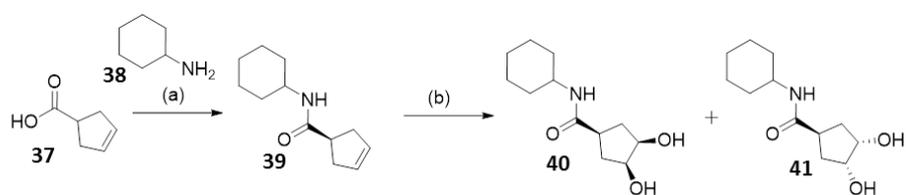
### 3.3.1 Synthesis of analogues

The first attempted synthesis of a non-sulfonic acid proceeded *via* the proposed retrosynthetic route in Scheme 6 using a Boc-protected amine (Scheme 7) as discussed in the retrosynthetic scheme.



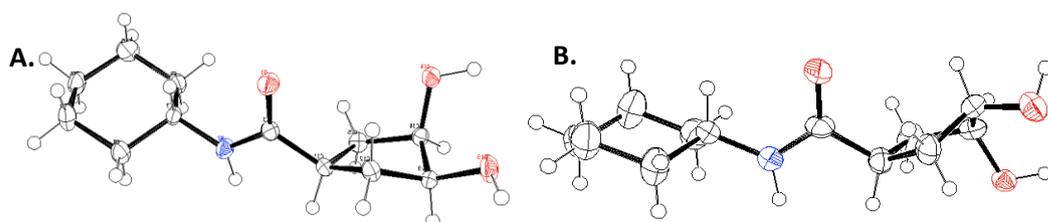
**Scheme 7.** synthetic route to **37**: (a)  $LiAlH_4$ , THF, RT, 12hrs; (b) Dess Martin Periodinane, DCM, RT, 5hrs; (c)  $NaBH_4$ , DCM, RT, 2hrs; (d) Di-*tert*-butyl dicarbonate, DIPEA, DCM, RT, 30mins, 14% (over 5 steps); (e)  $OsO_4$ , NMO, THF,  $H_2O$  (9:1), 52%.

The 3-cyclopentene carboxylic acid (**37**) was subjected to a reduction to the alcohol **36** in the presence of excess lithium aluminium hydride. The alcohol was then oxidized *in situ* to the aldehyde **35** using Dess-Martin periodinane. The amine **34** was generated by combination of cyclohexylamine (**38**) with the aldehyde **35**. Due to the volatility of **36** and **35** and the polarity of **34** these were not isolated and the synthesis was continued with crude products from steps 1, 2 and 3. Isolation and purification of the Boc-protected amine **33** was achieved at step 4. The Boc-protected amine **33**, was subjected to osmium tetroxide dihydroxylation to yield **32**, as a mix of two meso isomers **32a** and **32b**. Unfortunately, it was not possible to separate out the meso isomers **32a** and **32b** and so this route was abandoned. A new synthetic scheme was postulated based upon synthetic ease and potential for column separation. This new route proceeds *via* an amide (Scheme 8) intermediate.



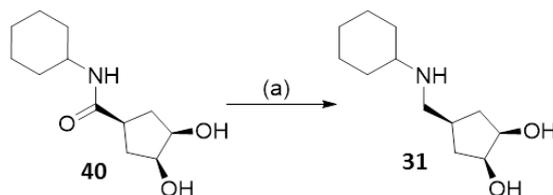
**Scheme 8.** Synthetic route to **40** and **41**. (a) CDI, DCM, RT, 6hrs, 82%; (b) OsO<sub>4</sub>, NMO, THF, H<sub>2</sub>O (9:1), RT, 12hrs (**20**: 46%, **21**: 32%).

The 3-cyclopentene carboxylic acid was coupled to the amine using carbonyl-diimidazole. This reaction proceeded for 6 hours. The amide **39** produced from this reaction was then subjected to a dihydroxylation using osmium tetroxide in the presence of NMO. This generated the two meso isomers **40** (“syn”) and **41** (“anti”), which were separable by flash column chromatography. To confirm the stereochemistry of each ligand, small-molecule crystals structures were generated (Figure 47). Additionally, NOESY 3D NMR was observed to distinguish between each ligand.



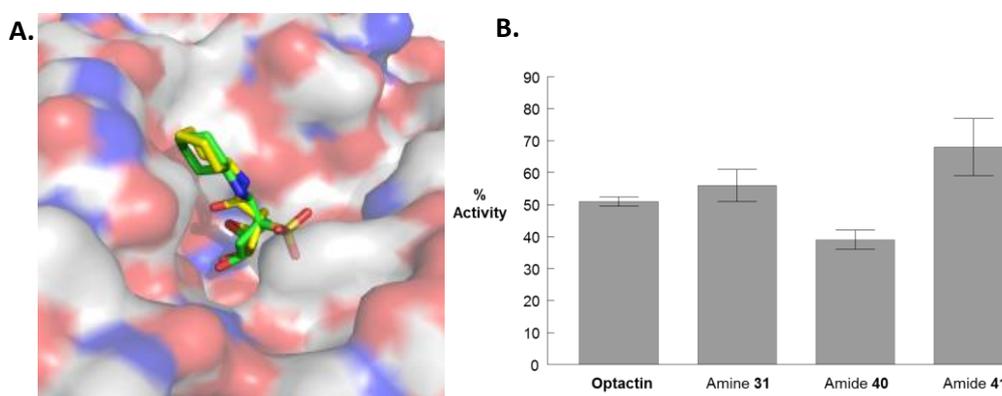
**Figure 47.** ORTEP plots showing 50% thermal ellipsoid probability of A. (**40**) and B. (**41**).

Reduction of the amide to the amine **31** proceeded in the presence of a strong reducing agent ( $\text{LiAlH}_4$ ) (Scheme 9).



**Scheme 9.** Synthetic route to **31**. (a)  $\text{LiAlH}_4$ , THF, 12hrs, 10%.

Silica gel column chromatography pre-treated with trimethylamine (TEA) was used to purify amine **31** using DCM and MeOH in a ratio of 9:1 as the mobile phase. The potency of amine **31** and the amides **40** and **41** were all evaluated using the 4-Munana assay. Surprisingly, the amide analogue **40** had a slightly higher potency than the amine **31** (Figure 48) tested at 500  $\mu\text{M}$ . The  $\text{IC}_{50}$  obtained for amide **40** was slightly more potent than that obtained for Optactin ( $\text{IC}_{50}$ :  $347 \pm 44 \mu\text{M}$ ). The calculated LE for amide **40** is 0.128, which is an improvement over the LE of 0.095 for Optactin. The binding of the amide was determined through X-ray crystallography and observed to bind into the allosteric site of NanB (Figure 48).



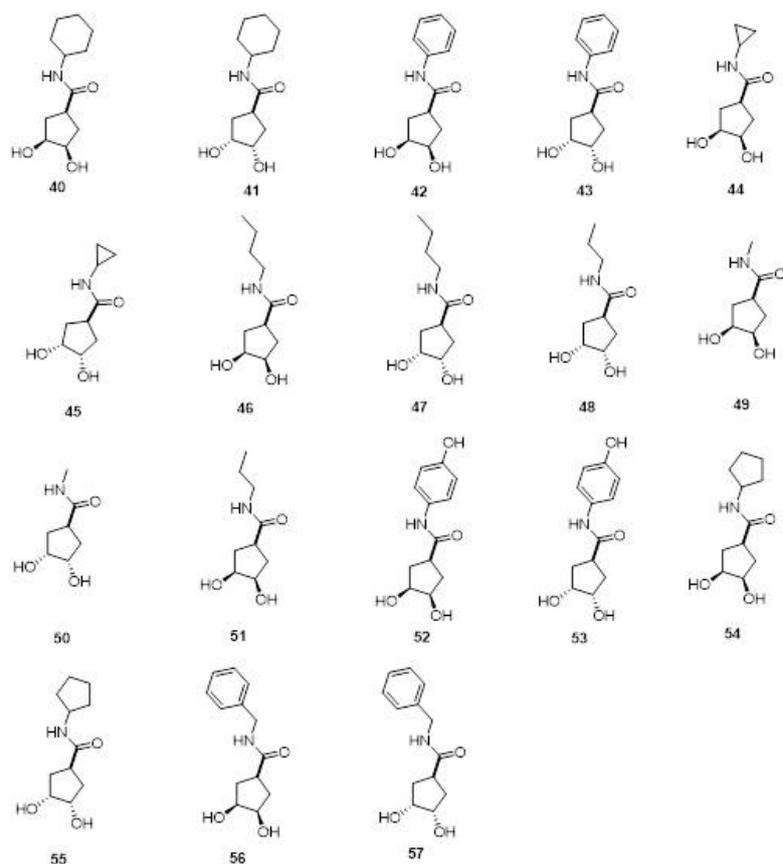
**Figure 48. A.** The percentage of 4-Munana hydrolysis by NanB in the presence of amine **31**, amide **40** and amide **41** (at 500  $\mu\text{M}$  as compared to the control). **B.** The binding of amide **40** (yellow) compared with Optactin (green) within the allosteric site of crystal structures  $\text{NanB}_{\text{WT}}$ -amide **40** and  $\text{NanB}_{\text{WT}}$ -Optactin.

The binding site of amide **40** was observed to bind in a similar fashion to that of Optactin within a  $\text{NanB}_{\text{WT}}$ -amide **40** crystal structure complex. The binding of amide **41** to the secondary site was difficult to determine due to weaker electron density observed for this amide despite similar potency to its meso isomer, amide **40**.

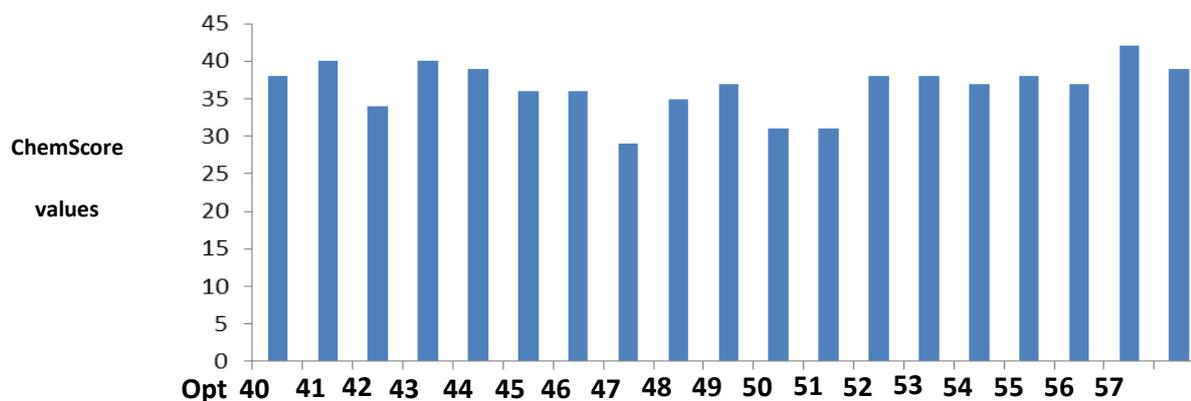
### **3.3.2. Optimisation**

The next point of modification identified from the amide-NanB complex was the cyclohexyl group. The positioning of the cyclohexyl ring within close proximity to E658, Y598 and H662 is likely a suboptimal functionality for binding affinity and a suitable point for the next modification. To identify a moiety as a suitable alternative at this position, a CADD approach was used. The GOLD docking program (Jones et al., 1997) was used to evaluate suitable moieties. A “virtual” library was created containing a population of analogues with different functionality at this identified point for modification (Figure 46.). These analogues were designed to evaluate the size of the hydrophobic pocket that the cyclohexyl group is adjudged to bind to (Figure 42) and the need for lipophilic, hydrophobic and/or electrostatic properties for improved binding. Before the docking run of the “virtual” library was initiated a validation run was performed. Within the top three binding poses generated of amide 40-NanB the correct binding pose was predicted, validating GOLD as a useful docking program for CADD.

**A.**



**B.**



**Figure 49. A.** The chemical structures of each analogue within the virtual library used for CADD assisted structure-based design. GOLD was the docking software used to dock the “virtual” library against the mapped. **B.** Graph of the best CHEMScores values output by GOLD for each compound within the “virtual” library (Opt = Optactin).

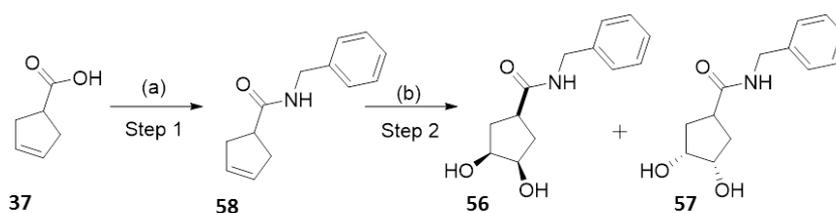
The CADD docking run identified that small ring systems made up of five or less carbons were suboptimum and had lower calculated binding scores. Replacing the ring system to an alkane chain also resulted in a lower binding score. However, replacing the cyclohexane ring with a phenyl ring caused no change in the calculated binding score. Adding a polar functionality (OH) to this arene in the para position decreased the calculated binding score suggesting that hydrophobicity within this

pocket is important for improved binding. The addition of a CH<sub>2</sub> to make a benzyl ring increased the binding score further by increasing hydrophobicity and/or positioning the arene further into the hydrophobic pocket.



**Figure 50.** Binding position of amide **56** within the allosteric site of NanB. The calculated aromatic receptor surface using Discovery Studio 4.0 is shown in the image. Positioning of the arene ring in amide **56** is in a possible edge-to-edge interaction.

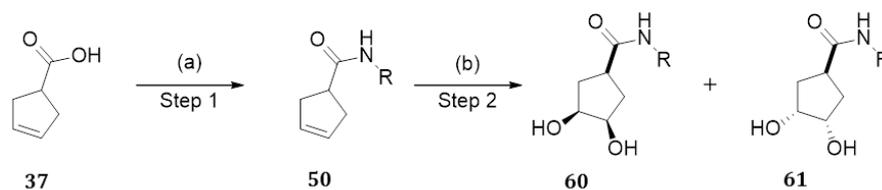
The benzyl group was identified from the docking run as the most promising candidate (compound **56** was the highest scoring analogue identified from the virtual library) and from visualisation within Discovery Studio 4.0 the arene of amide **56** contributes to an edge-to-edge aromatic interaction with histidine residue 662 (Figure 50). Amide **56** was chosen as the chemical candidate for synthesis and biological testing. This molecule was synthesized as described in scheme 9.



**Scheme 10.** (a) CDI, DCM, r.t. 6hrs, 75% (b) cat. OsO<sub>4</sub>, NMO, THF/H<sub>2</sub>O (9:1) r.t. 16hrs (56: 45%, 57: 40%).

The stereochemistry of the amides **56** and **57** were assigned using NOE. The potency of amide **56** and **57** was checked using the 4-Munana assay. With the presence of an aromatic ring in the new inhibitor, it is possible further improvements could be made using rational design and QSAR. The Topliss scheme (Topliss, 1972) discussed earlier (Chapter: Medicinal Chemistry 1.71) is an operational scheme suited for aiding further optimisation and was used to guide the synthesis of a

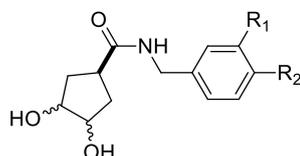
more potent analogue. The generation of further analogues was accomplished using the same route (Scheme 11).



**Scheme 11.** (a) CDI, DCM, r.t. 6hrs. (b) cat. OsO<sub>4</sub>, NMO, THF/H<sub>2</sub>O (9:1) r.t. 16hrs.

The stereochemistry of all further analogues of amides **56** and **57** were assigned based on their separation by flash chromatography as subtle (non-polar) changes on the aromatic ring was deemed unlikely to alter the fractional position of the two meso conformers on a silica column. For the final LEAD compound a protein-ligand complex crystal structure was solved, which confirms the assumption above to be true and the stereochemistry of the LEAD as “syn” ( 3.3.3. Further Optimisation).

**Table 7.** Table showing a selection of compounds assayed against NanB using the Topliss scheme. \* compounds synthesised and purified by Yernur Syzdykov (an undergraduate project student under my guidance).



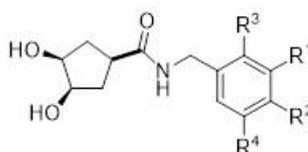
Compound	R <sub>1</sub>	R <sub>2</sub>	Stereochemistry	% Inhibition		
				500 μM	100 μM	IC <sub>50</sub>
56	H	H	"syn"	69±5	26±3	
57	H	H	"anti"	50±3		
62	H	Cl	"syn"	98±1	51±2	
63	H	Cl	"anti"	92±2	38±4	
64*	Cl	Cl	"syn"	92±1	90±2	55.4±2.5
65*	Cl	Cl	"anti"	81±2	16±3	
66*	CF <sub>3</sub>	Cl	"syn"	78±2	35±3	
67*	CF <sub>3</sub>	Cl	"anti"	65±4	20±5	

A project student under my guidance assisted with the synthesis and purification of some of these developed analogues. The most potent analogue generated using the Topliss series was compound **64**. It was evident from testing of both the "syn" and "anti" isomers of each compound that the "syn" isomers were more potent. Therefore further optimisation would only be carried out on the "syn" isomers. Compound **64** was used as a basis for further optimisation to improve potency.

### 3.3.3. Further optimisation

Further optimisation was implemented on the data generated from the Topliss series. Different electron withdrawing groups were tested to identify whether decreasing the electron density of the aromatic ring resulted in improved electrostatic or hydrophobic interactions with the allosteric site. Additionally, non-polar groups were tested to determine whether hydrophobicity was key for potency. The R<sup>1</sup> and the R<sup>3</sup> positions were investigated for functionalities that would improve potency, as these positions were not covered extensively within the Topliss series (Table 8).

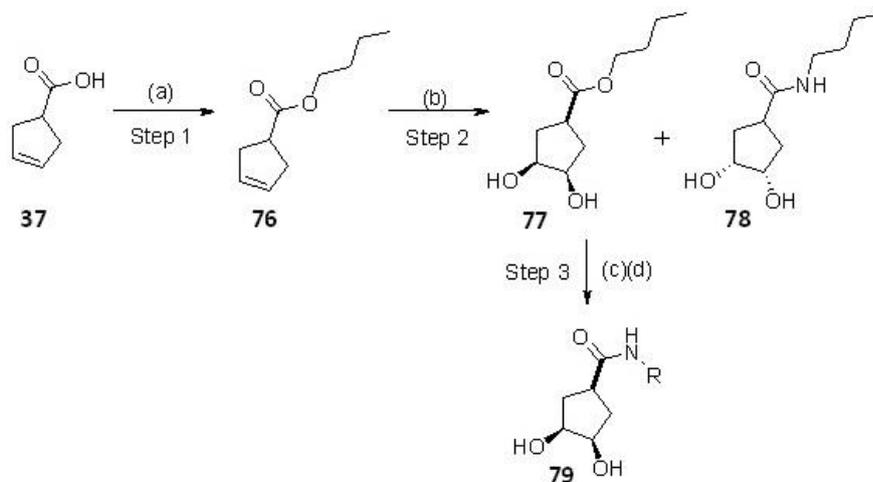
**Table 8.** Table showing compounds assayed against NanB using further SAR.



% Inhibition								
Compound	R1	R2	R3	R4	500 $\mu$ M	100 $\mu$ M	50 $\mu$ M	IC <sub>50</sub>
68	Br	H	H	H	20 $\pm$ 2	5 $\pm$ 3		
69	Me	H	H	H	45 $\pm$ 1	15 $\pm$ 2		
70	F	H	H	H	98 $\pm$ 1	93 $\pm$ 2	90 $\pm$ 1	8.0 $\pm$ 1.2
71	Cl	H	H	H	98 $\pm$ 2	97 $\pm$ 4	94 $\pm$ 1	6.2 $\pm$ 1.3
72	F	Cl	H	H	97 $\pm$ 2	95 $\pm$ 3	43 $\pm$ 3	
73	Cl	H	H	F	97 $\pm$ 3	97 $\pm$ 1	49 $\pm$ 2	
74	Cl	H	H	Cl	98 $\pm$ 1	41 $\pm$ 5		
75	F	H	Cl	H	95 $\pm$ 2	96 $\pm$ 1		3.0 $\pm$ 1.7

The first synthetic route to these amides involved a CDI coupling and then an osmium tetroxide dihydroxylation (Scheme 11). Due to the problems with the separation of the two-meso products a second synthetic route to these analogues was pursued. This route was designed so that separation

could be achieved and the functionality changed late in the synthesis providing an efficient means for analogue synthesis. This route proceeded through an ester (Scheme 12).



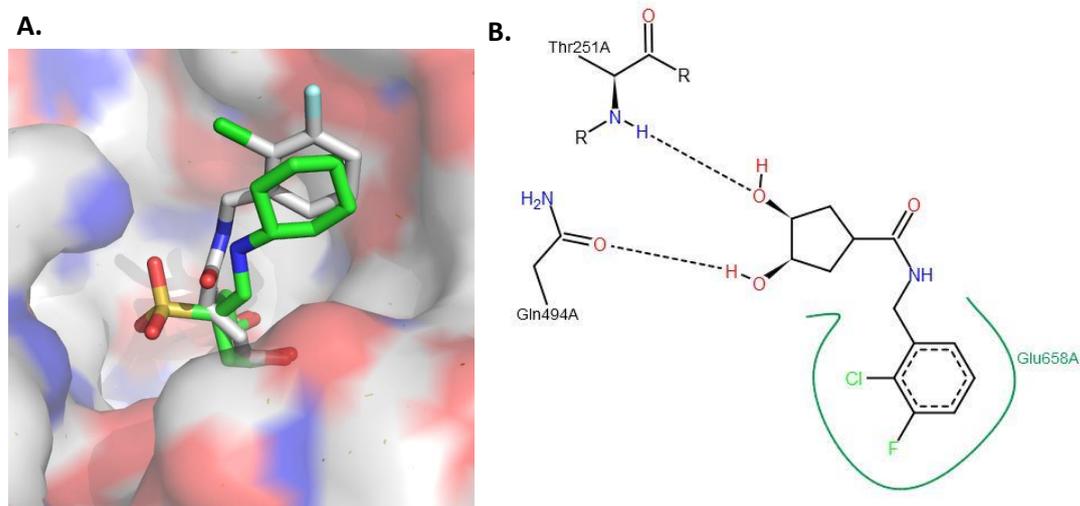
**Scheme 12.** (a) CDI, cat. DMAP, DCM, r.t. 16hrs, 65%. (b) cat. OsO<sub>4</sub>, NMO, THF/H<sub>2</sub>O (9:1) r.t. 16hrs (77: 56%, 78: 18%). (c) 1M NaOH, reflux, 2hrs. (d) CDI, DCM, r.t. 6hrs.

Cyclo-pentene carboxylic acid was esterified with an alcohol using a Steglich esterification (Neises and Steglich, 1978). The ester was then subjected to a dihydroxylation using osmium tetroxide in the presence of NMO. The two meso isomers generated were separated using flash chromatography. The ester was then subjected to hydrolysis using NaOH. After a work-up the crude product was subjected to coupling with an amine in the presence of carbonyl-diimidazole to yield the final amide. This route was used to re-synthesise amides **56** and **57**. NMR comparison of the <sup>1</sup>H spectra of amide **56** synthesised in this route with that of amide **56** synthesised earlier through Scheme 10 was used to assign ester stereochemistry. This route was used to develop some of the candidates for biological testing.

The generated analogues tested in this further optimisation cycle identify that an electron withdrawing group in the R<sup>1</sup> position is key for potency. Amide **75**/Optactamide was identified as the most potent inhibitor. Distortion of the electron density of the ring with two electron withdrawing groups in the R<sup>1</sup> and R<sup>3</sup> position results in improved potency. Three possible interactions occur that improve the potency observed: 1) the halogen atoms contribute to more favourable hydrophobic and steric interactions, 2) the halogen atoms contribute to hydrogen bonding or halogen bonding at these positions and/or 3) a favourable stacking interaction occurs between the aromatic ring and the allosteric site, which is promoted by the electronic withdrawing effects of the two halogens. To determine which is true a NanB-amide **75** crystal structure complex was solved (Figure 51).

### 3.4 Optactamide binding position

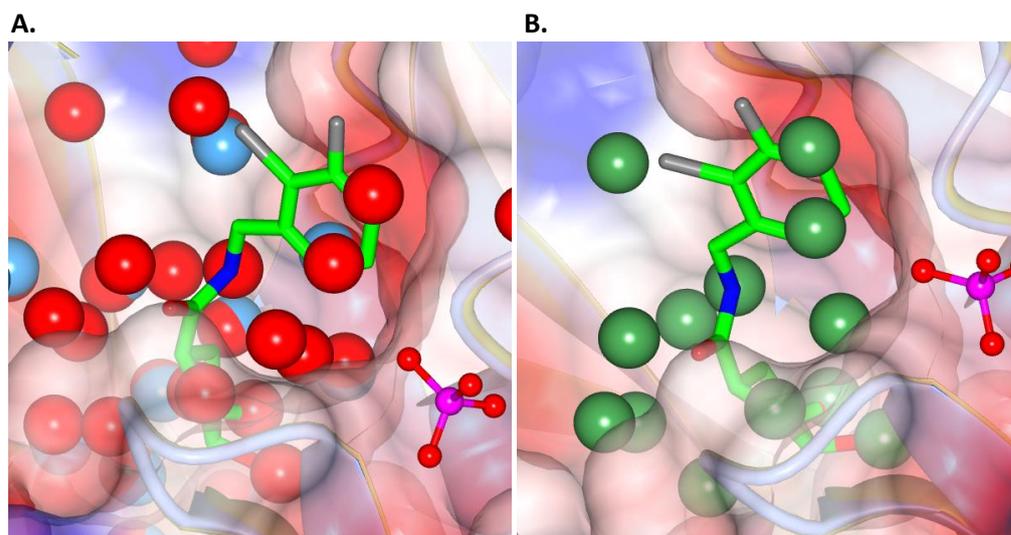
The Optactamide-NanB crystal structure complex was solved. Within this crystal structure Optactamide is observed to bind in the same position as Optactin (**Figure 51**).



**Figure 51.** **A.** Optactamide (grey) bound within the allosteric site of NanB with Optactin superimposed (green) (PDB: 4XYX and PDB: 4XHX). **B.** PoseView generated 2D binding interaction between the LEAD candidate/Optactamide. Similar interactions are made between Optactamide with the allosteric site of NanB as Optactin. PoseView identifies the aromatic region as positioned within a hydrophobic pocket.

PoseView (Stierand et al., 2006) identifies a hydrophobic interaction of the aromatic ring and the halogen substituents with the allosteric site. Given the position of the halogen atoms in the crystal structure, it is possible that the F atom on the R<sup>1</sup> position has a direct hydrogen bonding interaction with the side chain of S599, which is 3.7 Å away. Hydrogen bonding with water is likely given the orientation of the halogens into the water channel exposes them to the solvent. It is possible both hydrogen bonding and solvation contribute to the increased potency of this molecule. The software (PoseView) did not predict an electrostatic/ $\pi$ -stacking interaction to occur between the aromatic group of Optactamide and NanB. However, it is possible that such an interaction does exist as a Glu658 side chain is located 4.0 Å away and is orientated in a face-to-face interaction with the aromatic ring. No electron density for Optactamide was observed within the active site of NanB.

The allosteric site within NanB is located within the water channel of the catalytic domain (as observed within the Optactin-NanB crystal structures). Comparing the water channel of the Optactamide-NanB crystal structure (PDB: 4XYX) with an “apo” crystal structure (PDB: 4XIO) indicates that a number of waters are displaced by the binding of Optactamide (Figure 49).



**Figure 52. A and B.** Overlay of Optactamide-NanB crystal structure (PDB: 4XYX, gold) and “apo” crystal structure (PDB: 4XIO, blue). RMSD of both structures is 0.41 Å (as calculated by CCP4MG). Water molecules are shown as spheres. A. Waters that appear within the Optactamide-NanB crystal structure are in light blue. Waters that appear within the “apo” crystal structure are in red. B. Identified displaced waters from the overlay are in dark green.

It is unclear what role the central pore/water channel has in the sialidase activity of NanB. This structural feature, however is conserved within many sialidases and structurally related glycosidases and must be significant to its evolutionary function (Luo et al., 1998, Crennell et al., 1996). Disruption of ordered waters within this channel could explain the reason reduced sialidase activity is observed in the presence of Optactamide.

### **3.5 Activity assays and synergistic inhibition**

To determine if Optactamide would be useful as a chemical tool in chemical biology studies, control experiments (3.6. Control experiments) and an activity analysis were conducted. A feature of allosteric inhibition is the ability to inhibit selectively proteins within the same protein family. To determine if Optactamide is selective between various sialidases, it was tested against two other sialidases present in *S. pneumoniae*. The two other sialidases present in *S. pneumoniae* are NanA and NanC (Chapter 1.56.1 Streptococcus pneumoniae sialidases). To determine if Optactamide had selectivity for NanB it was tested against NanA and NanC using the 4-Munana assay (Table 9).

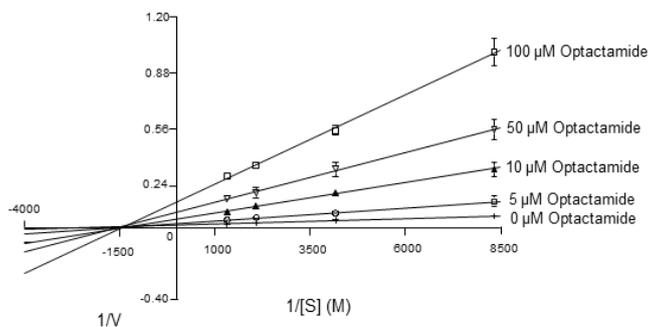
**Table 9.** Table of the percentage inhibition observed for NanA and NanC against Optactamide at 500  $\mu\text{M}$  and 100  $\mu\text{M}$ .

\*NanA catalytic domain expressed and purified by Dr L. Yang. \*NanC expressed and purified by Dr C. D. Owen.

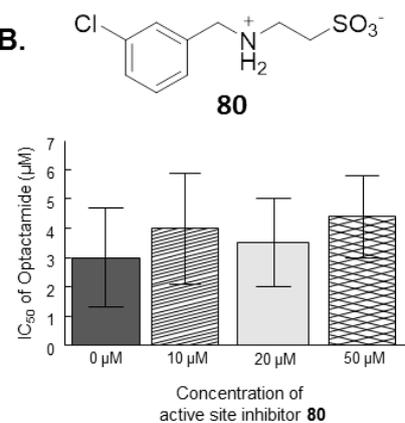
Sialidase	% Inhibition	
	500 $\mu\text{M}$ Optactamide	100 $\mu\text{M}$ Optactamide
NanA*	81.1 $\pm$ 1.1	51.2 $\pm$ 6.4
NanC*	14.1 $\pm$ 7.0	5.9 $\pm$ 3.3

Optactamide was determined to bind in a non-competitive manner against NanB using a Line-Weaver Burke plot against increasing substrate concentrations. In the presence of increasing concentrations of active site inhibitor of NanB (compound **80**), the potency of Optactamide did not change identifying Optactamide as a synergistic inhibitor (**Figure 53**).

**A.**



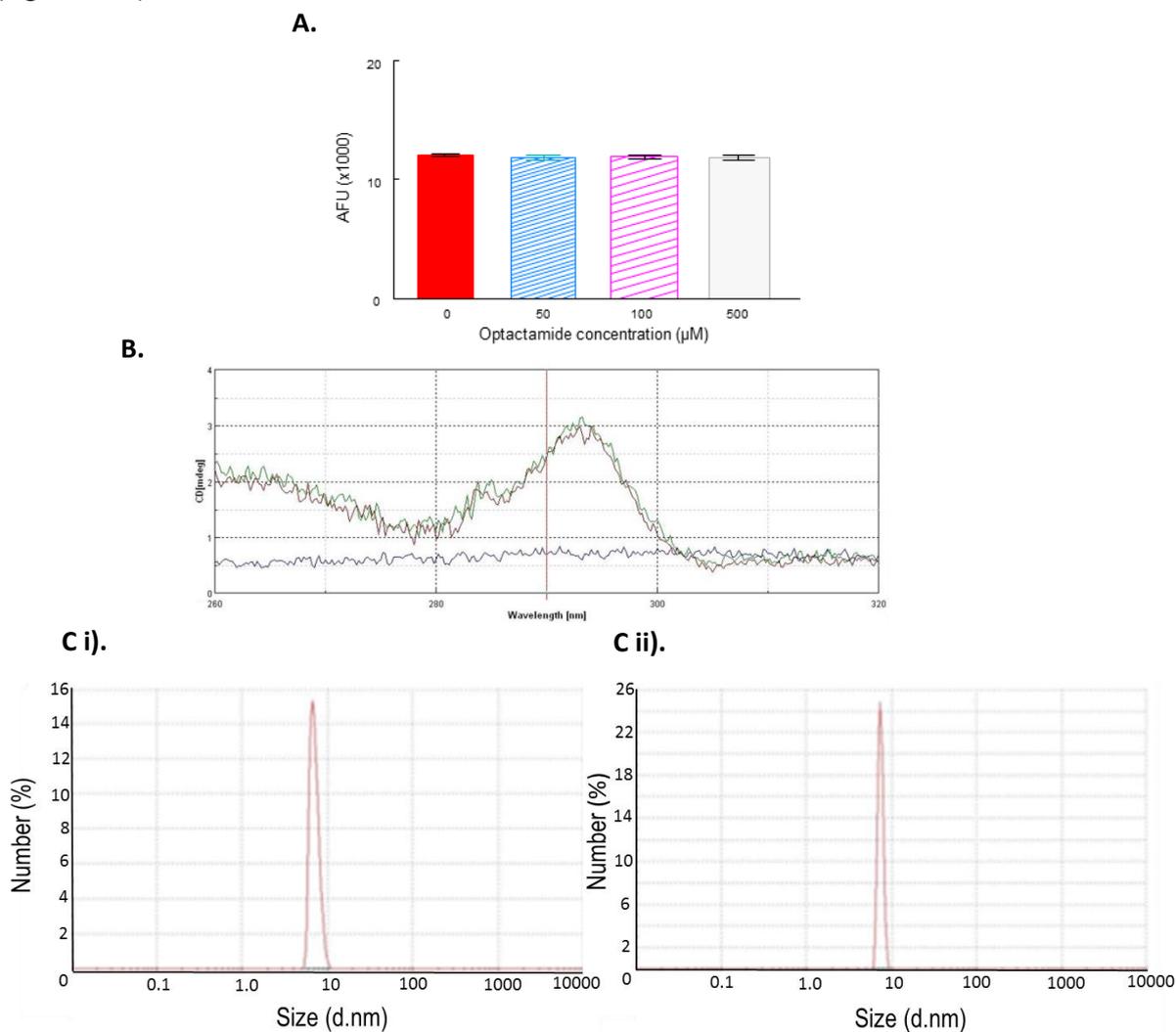
**B.**



**Figure 53. A.** Line weaver-Burke plot of Optactamide with increasing concentrations of substrate/4-Munana. **B.** IC<sub>50</sub> values of Optactamide when in the presence and absence of active site inhibitor compound **80**/2-((3-chlorobenzyl)ammonio)ethane-1-sulfonate (IC<sub>50</sub> of 38.9 $\pm$ 0.8  $\mu\text{M}$ , see Chapter 1.61 Current Inhibitors of NanB).

### 3.6 Control experiments

To determine if the inhibition observed for Optactamide is attributed to binding at the allosteric site alone and not due to promiscuous effects, a number of control experiments were run. In the 4-Munana assay quenching of the fluorophore can occur, which can lead to ambiguous/promiscuous results. Various concentrations of Optactamide were tested against 4-Mu to determine if quenching could occur. At all of the concentrations tested quenching of fluorescence was not observed for 4-Mu (Figure 51.A).



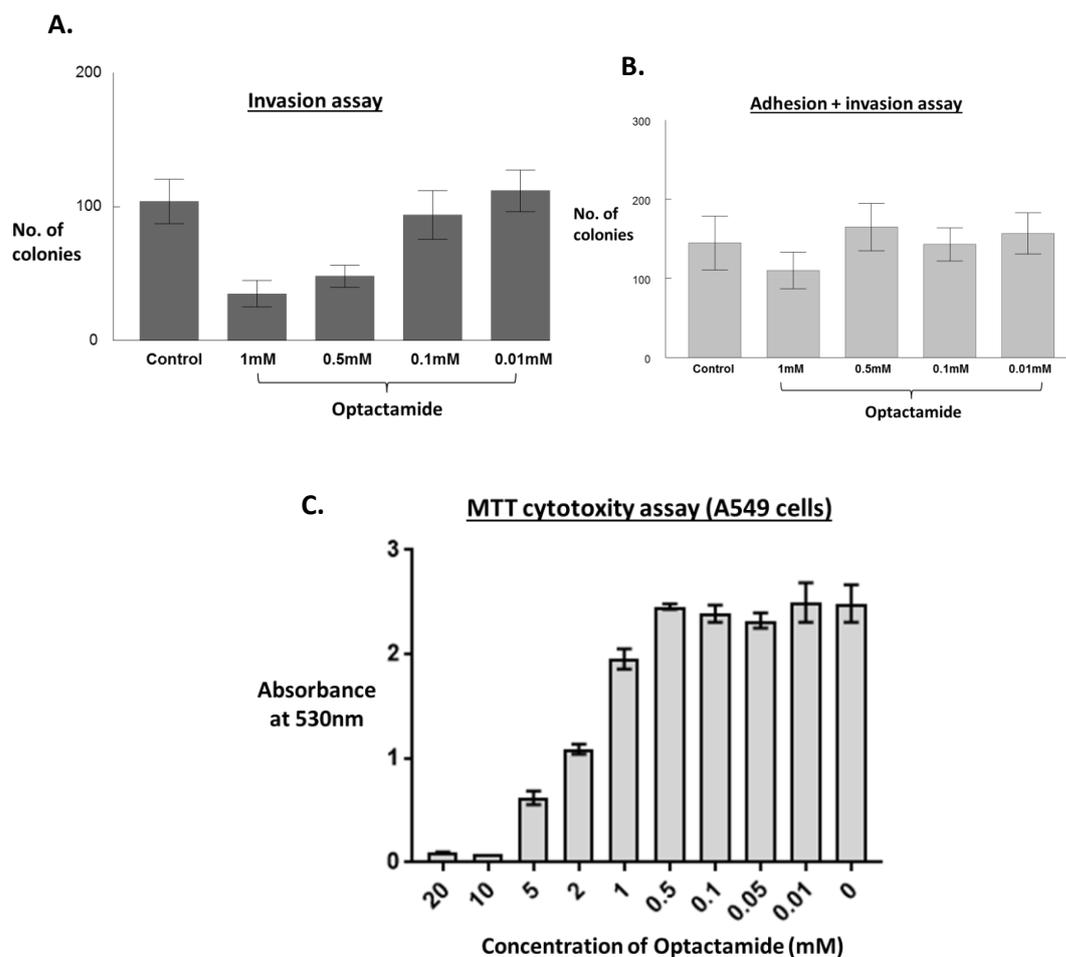
**Figure 54.** **A.** The fluorescence of 4-Mu (100 µM) at various concentrations of Optactamide. **B.** CD near UV spectra of NanB in the absence (red line) and presence of 500 µM Optactamide (green line). CD near UV spectrum was also measured on a solution containing no protein/NanB (blue line). **C.** DLS showing size distribution by diameter of (i) NanB control and (ii) NanB in the presence of Optactamide at 2.5 mM.

To determine that the cause of decreased sialidase activity in the presence of Optactamide was not due to protein folding or aggregation two experiments were performed. Circular-dichroism near

UV spectra of NanB was recorded in the presence and absence of 500  $\mu$ M Optactamide. No protein unfolding, change in tertiary structure or protein-protein interactions were observed in the presence of Optactamide. Small molecules can cause the aggregation of protein leading to promiscuity. Direct light scattering (DLS) is a method useful for the determination of particle size in solution. DLS was performed on NanB in the presence and absence of of Optactamide at 2.5 mM. No change in particle size was observed in solution and verifies that the protein is not aggregated within solution. Attempts to determine the binding affinity of Optactamide against NanB were unsuccessful. Isothermal titration calorimetry (ITC) and surface plasmon resonance (SPR) experiments did not yield a measurable binding affinity for Optactamide. Despite this, these control tests identify Optactamide as a valid allosteric inhibitor of NanB and a useful chemical tool.

### **3.7 Cell and bacterial assay**

Optactamide's use as a chemical tool was tested in preliminary experiments using a cell and bacterial assay. Lung epithelial cells (A549 cells) were used to determine if inhibition of NanB could protect against *S. pneumoniae* adherence and invasion. *S. pneumoniae* was grown overnight in BHI broth. The OD was measured the next day (OD: 1.29) and the bacteria were centrifuged and re-suspended in 10% FBS and DMEM high glucose. This suspension was then diluted 1 in 10 and 100  $\mu$ L added to each well of a 24 well plate containing A549 cells in the presence or absence of Optactamide. Cells were then incubated for six hours. Cell medium was removed and wells washed with PBS. The cells were then treated with trypsin/EDTA and triton and dilutions spread onto blood agar plates and left to incubate overnight. The number of colonies was counted the next day. In the presence of 1 mM and 500  $\mu$ M Optactamide some protection against *S. pneumoniae* was observed. This is more evident when looking at *S. pneumoniae* invasion only. The addition of non-cell permeable antibiotics (gentamycin and penicillin, according to the Winram *et al.* protocol) to the wells prior to cell lysis removes adhered *S. pneumoniae* that has not invaded into the mammalian cells (Figure 52).



**Figure 55.** **A.** The number of colonies determined from the invasion assay (with the antibiotic addition step). Cell suspensions diluted to 1 in 3. **B.** Adhesion and invasion assay (without the addition of the antibiotics). Cell suspensions diluted to 1 in 20. **C.** A549 cell toxicity assay using MTT.

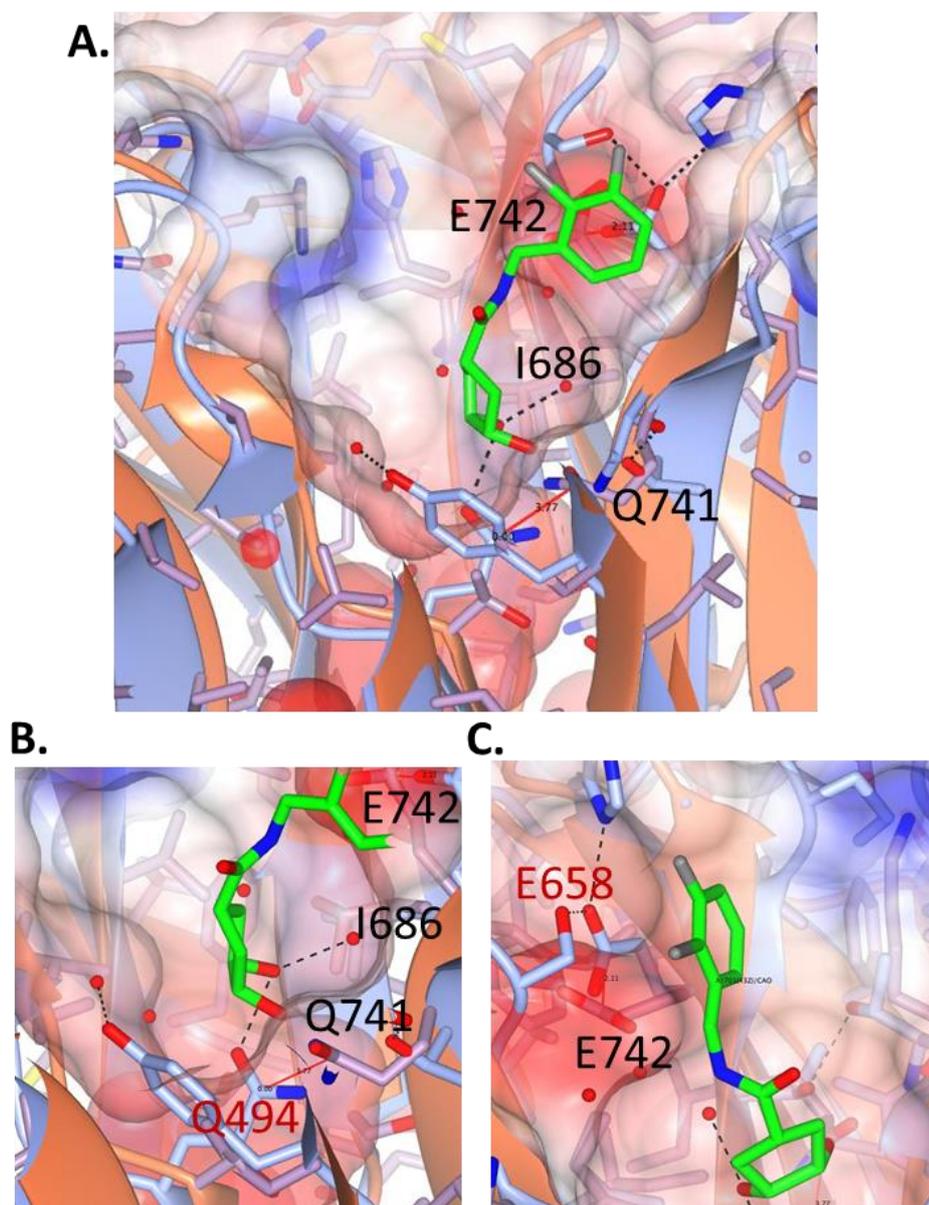
Protection of A549 cells against *S. pneumoniae* adherence and/or invasion was not noticed for lower concentrations of Optactamide. At the higher concentrations protection was detected but at these concentrations almost full inhibition of NanA would be observed (Table 9). It is therefore evident that for protection against *S. pneumoniae* infection, NanB cannot be considered as an independent target, rather NanA would also need to be targeted. However full protection was not observed at 1mM and Optactamide at concentrations of 1 mM and higher cause cellular death of A549 cells using an MTT assay (Figure 52.C.). Optactamide cannot be considered a clinical candidate however its preliminary use as a chemical tool has shed some light on the validity of NanB as a therapeutic target.

### 3.8 Summary

The focus of this chapter was to develop a selective and potent chemical tool for NanB by the optimisation of the NanB allosteric inhibitor Optactin. Optimisation of Optactin proceeded by the removal of the  $\text{SO}_3^-$  (due to mutagenesis studies identifying this functional moiety as redundant), replacement of the amine with an amide and substitution of the cyclohexyl moiety with a benzyl ring improved potency 5-fold. Further optimisation through rational design using the Topliss scheme generated 10-fold further improvement in potency ( $\text{IC}_{50}$ :  $55.4 \pm 2.5 \mu\text{M}$ ) with compound **64**. Different substitution patterns around the aromatic ring led to the discovery of compound **75**/Optactamide with a low micromolar potency ( $\text{IC}_{50}$ :  $3.0 \pm 1.7 \mu\text{M}$ ). Kinetic analysis and structural information show Optactamide to be an allosteric inhibitor of NanB. Optactamide was tested against other sialidases present in *S. pneumoniae*. Against NanA, Optactamide was observed to be 17-fold less potent than NanB ( $51.2 \pm 6.4\%$  inhibition of NanA at  $100 \mu\text{M}$ ). Against NanC, Optactamide was observed to have no inhibition at  $500 \mu\text{M}$  ( $14.1 \pm 7\%$  inhibition of NanC). Optactamide was used as a chemical tool to decipher its role in pneumococcal invasion in an *in vitro* *S. pneumoniae* invasion experiment. NanB and NanA were identified to play a role in the invasion of *S. pneumoniae* into A549 cells, however a high dose was required to cause inhibition ( $500 \mu\text{M}$ ). Analysis against a cell death assay identified that a dose of higher than  $1\text{mM}$  of Optactamide caused cell toxicity and death of A549 cells and so the inhibition observed at  $500 \mu\text{M}$  can be directly attributed to sialidase inhibition. Optactamide could not fully protect A549 cells against *S. pneumoniae* invasion. This preliminary work with a NanB allosteric tool suggests that NanB is not an independent therapeutic target. Targeting NanA, NanB and possibly NanC would provide a better outcome. Optactamide has been useful as a chemical tool in preliminary chemical biology experiments.

## ***4.1 Discussion and Future work***

The development of a novel chemical tool was accomplished against NanB. This “relaxed” chemical tool works by negative allosteric modulation. Orthosteric sites are conserved within the sialidase family and allosteric site modulation provides an alternative route to the development of specific modulators. Optactamide was determined to be specific to NanB when tested against NanC, a sialidase with the highest sequence homology to NanB at 46%. However, Optactamide did inhibit NanA with approximately 50% inhibition at 100 $\mu$ M despite a lower sequence homology of 24%. The allosteric site of NanA contains a very similar binding site to NanB. Further optimisation of Optactamide against NanB should focus on achieving specificity as well as potency. Firstly, an Optactamide-NanA crystal complex should be solved or alternatively, in the absence of an Optactamide-NanA crystal structure, a CADD approach could be used to determine the binding mode of Optactamide within NanA. This will provide valuable information for SBDD assumptions in achieving NanB selectivity in future Optactamide analogues as well as providing an opportunity for the development of a specific NanA inhibitor. Overlay of the structure of NanA (PDB: 2YA4) with Optactamide-NanB crystal structure (PDB: 4XYX) suggests that the binding site maybe similar with only minor positional changes of the key protein side groups involved in van der Waal interactions with Optactamide (structural RMSD: 1.8 Å). In NanA the backbone carbonyl of threonine 251 is replaced with a backbone carbonyl of a lysine (residue 338). Glutamine 494 in NanB (a residue that contributes to direct hydrogen bonding with one of the diols of Optactamide) is replaced with a glycine (residue 556) in NanA. However, glutamine 741 within NanA is positioned 3.7 Å away from the position of glutamine 494 within NanB. This results in a slight steric clash with an OH from the diol of Optactamide (Figure 56). It is unlikely this would render Optactamide inactive, but would change the binding position of Optactamide slightly. Glutamic acid 658 in NanB is thought to contribute to electrostatic interactions with the arene of Optactamide. This glutamic acid is also in a slightly different position in NanA (2.1 Å further away, residue 742). It is possible the flexibility within Optactamide might still accommodate the positional change of the glutamic acid, but both positional changes of the glutamic acid and glutamine would result in sub-optimal binding of Optactamide within NanA (Figure 56). This could explain the reduced potency of Optactamide observed against NanA in the 4-Munana assay (see Chapter 3.5, Activity assays and synergistic inhibition).

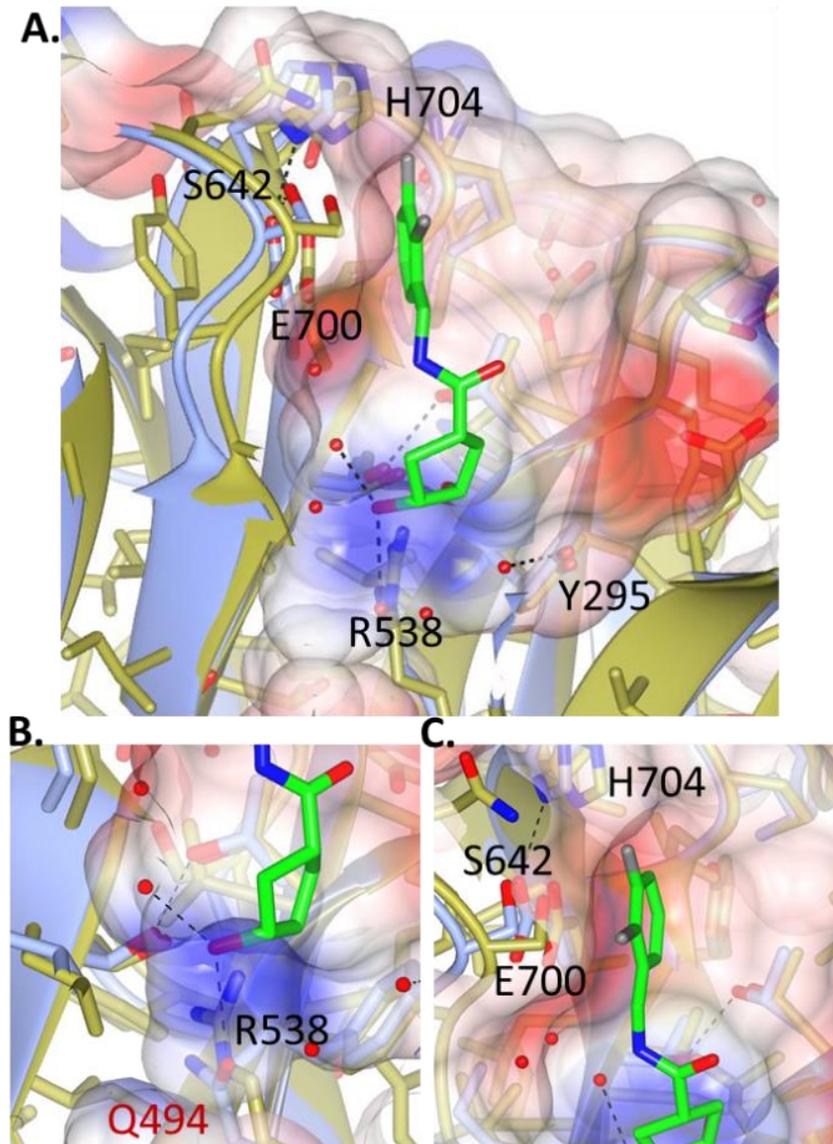


**Figure 56.** Superimposed structure of NanA (PDB: 2YA4) against Optactamide-NanB (PDB: 4XYX) crystal structure (RMSD: 1.8 Å). Electrostatic surface representation shown in each image. **A.** Optactamide (green) binding position within NanB (blue) overlaid in the same position within NanA (pink). **B.** Close up image of the cyclopentane diol binding position of Optactamide. Changes in the position of the glutamine results in a slight clash with Q741 (NanB residue at this position is Q494). **C.** Close up image of the arene binding position of Optactamide. Changes in the position of glutamic acid (NanA: E742 and NanB: E658) results in slightly suboptimal binding for an electrostatic interaction with the arene.

Adaption of Optactamide to a NanA chemical tool could be possible. Improvements in potency and selectivity would include the change of the amide to an amine within Optactamide as this would likely generate a favourable electrostatic interaction with glutamine 742. Addition of a CH<sub>2</sub> to move the position of the amine closer to glutamic acid 742 might result in a more favourable electrostatic

interaction. Further improvements in potency against NanA could be generated by: 1) modification of the benzyl ring using the Topliss series or 2) complete replacement of the arene with another functionality guided by CADD. The analogues synthesised would need to be screened in parallel against NanB using the 4-Munana assay to exclude functionality that cause reduced selectivity.

Comparison of Optactamide's binding position within NanB (PDB: 4XYX) with a structural overlay of NanC (PDB: 4YZ1) confirms that binding within NanC would be unfavourable (structural RMSD is 1.7 Å). Firstly, glutamine 494 within NanB (a residue that contributes to direct hydrogen bonding with one of the diols of Optactamide) is not present in NanC. Instead arginine 538 within NanC occupies this position resulting in a change in the electrostatic and accessible surface of the binding site. The electrostatic surface of this arginine directly clashes with an OH from the diol of Optactamide rendering Optactamide unlikely to bind at this position (Figure 57). Mutation of arginine 538 into a glutamine to develop a NanC<sub>R538Q</sub> mutant might result in Optactamide efficacy. The generation of bacterial resistance to antibiotics is a major health concern (see Chapter 1.42, Antibiotics). The generation of a NanB<sub>Q494R</sub> would likely result in reduced Optactamide activity and could be a mechanism of bacterial resistance against Optactamide. The antibiotic clock (see Chapter 1.42, Antibiotics) dictates how useful an antibiotic will be. If a NAM of NanB or NanA was to be developed as an antibiotic the potential routes of resistance would need to be investigated and evaluated. Other residues that likely contribute to a reduced binding ability include: serine 642, glutamine 700 and histidine 704. These residues are positioned further into the pocket resulting in a steric clash with the arene of Optactamide.



**Figure 57.** Superimposed structure of NanC (PDB: 4YZ1) against Optactamide-NanB (PDB: 4XYX) crystal structure (RMSD: 1.7 Å). Electrostatic surface representation shown in each image. **A.** Optactamide (green) binding position within NanB (blue) overlaid in the same position within NanC (gold). **B.** Close up image of the cyclopentane diol binding position of Optactamide. A direct clash with R538 is observed (NanB residue at this position is Q494). **C.** Close up image of the arene binding position of Optactamide. Changes in the position of the residues that exist in the arene binding pocket (serine, histidine and glutamic acid) are observed.

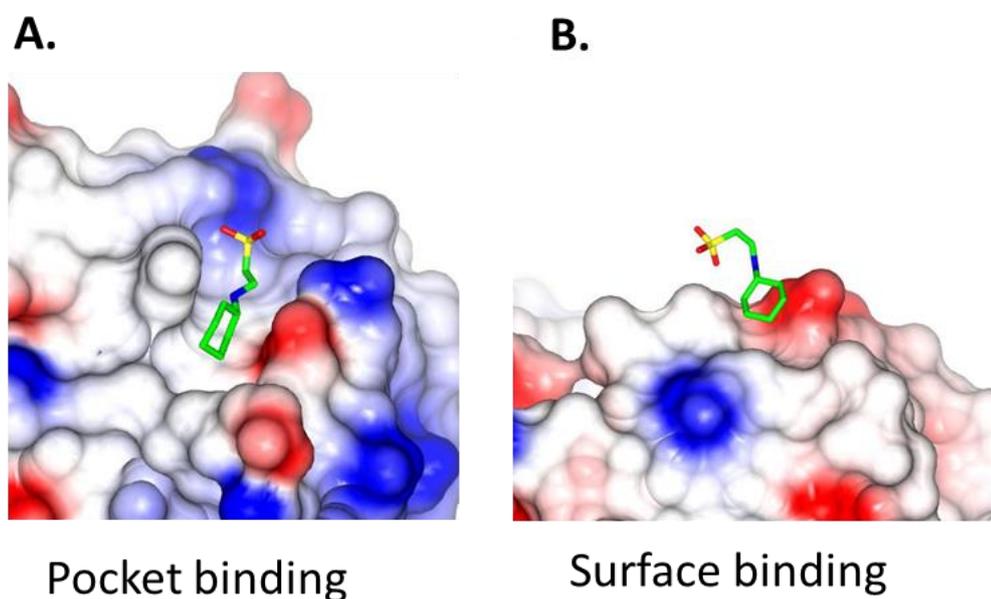
The results obtained from the *S. pneumoniae* infection/invasion assay suggest that NanB has only a minor role in the invasion of *S. pneumoniae* into lung epithelial cells. At high concentrations of Optactamide that would cause full inhibition of NanA, lung epithelial cell protection is observed. Using a specific chemical tool that is selective for NanA would confirm these results. Despite protection being observed in an *in vitro* set up, these results may not correlate with *in vivo*

experiments. Optactamide would not be useful for *in vivo* experimentation as its solubility is low and its toxicity is high (as observed in a MTT toxicity assay) (see Chapter 3.7, Cell and bacterial assay). The limitations of a “relaxed” chemical tool design are evident in this situation. Further optimization of Optactamide will need to follow the “constrained” chemical tool design with the future aim of *in vivo* experimentation. Furthermore, full protection of the lung epithelial cell line against *S. pneumoniae* was not observed with NanA and NanB inhibition suggesting that NanC may also need to be inhibited. Alternatively, these results could indicate that these sialidases are multifaceted and catalytic activity is not the only important function these proteins have in *S. pneumoniae* virulence. The carbohydrate binding domains of these proteins may also have an important role in *S. pneumoniae* virulence. The carbohydrate binding domain in particular, is present in NanA, NanB and NanC and binds to terminal sialic acid of glyconjugates (Owen et al., 2015, Yang et al., 2015, Xu et al., 2008). Development of chemical tools that inhibit sialic acid binding within the carbohydrate binding domains could prove challenging, but inhibition of both the catalytic activity and carbohydrate recognition/binding from the CBM domains could result in full protection from *S. pneumoniae* invasion and adhesion. Alternatively, use of an engineered multi-valent CBM with improved affinity as a competitive inhibitor of *S. pneumoniae* sialidase CBM could provide a novel approach to limiting *S. pneumoniae* invasion and adhesion. Similarly the influenza virus binds to sialic acid present on host cell surfaces (Connaris et al., 2014). A multi-valent CBM has been used in *in vivo* studies and shown to provide mice with complete protection from a lethal challenge of a 2009 pandemic H1N1 influenza virus (Connaris et al., 2014).

Other GH-33 sialidases that allosteric sites could be located in include TcTS (PDB: 1MR5), NanI (PDB: 2VKs), NedA (PDB: 1EUR), VCS (PDB: 1KIT), Neu2, TrSA, STNA, PaNa, Btsa, BDI-2946, Baccac\_01090 and RgNanH (PDB: 4X47) (see Chapter 1.59, Conservation). Further glycosidic hydrolases that allosteric sites could exist in include the GH-1 sialidases as the water channel is conserved with seven internal water molecules conserved across 90% of the known published structures (see Chapter 1.59, Conservation). Direct disruption of one of these conserved waters could result in reduced enzymatic activity. It is also likely allosteric sites exist within water channels of other protein types including: serine proteases, kinases, cytochrome P450 and ATP-synthase (Teze et al., 2013, Knight et al., 2009, Meyer, 1992, Gohlke et al., 2012, Oprea et al., 1997).

It is possible that the PDB provides an as of yet unmined wealth of information of novel unexplored small molecule binding sites in other protein families. Hardy and Wells were the first to identify that small molecules present within crystallization conditions could “serendipitously” bind to

the protein within the crystal structure and identify pockets amenable to small molecule binding and modulation (see Chapter 1.35, Fragment screening and serendipitous binders).



**Figure 58.** Examples of CHES serendipitously bound within crystal structures of: **A.** the NK1 fragment of HGF/SF complexed with CHES (PDB: 5CT1) and **B.** phosphoribosylglycinamide formyltransferase (purN) from *Coxiella burnetii* (PDB: 3TQR). CHES is found bound within a pocket in A. CHES is found bound on the surface of B.

However, care will need to be taken with this approach as it is unlikely that all pockets containing serendipitous small molecule binders in the PDB would be novel allosteric sites. It is likely small molecules bind at the surface and contribute to crystal-crystal contacts (Figure 58). The presence of these small molecules in high millimolar concentrations and errors within crystal structures due to low resolution coordinate errors or density misinterpretation could lead to incorrect assumptions. Additionally, the active sites of enzymes are rich with polar residues and would be more amenable to small molecule binding. A large number of ligands are found bound within structures in the PDB. Evaluating each small molecule bound within each structure is no small task. Computational tools that can screen and identify serendipitous binders within pockets, evaluate their binding and determine if this site is allosteric would be useful in reducing this workload.

The design of a computational tool that screens for allosteric pockets using a machine learning approach was developed by Chen et al, 2016. This computational tool uses structural information from previously determined orthosteric sites (159 protein-ligand complexes from PDBbind), allosteric sites (59 protein-ligand complexes from literature and the AlloSteric Database) and

miscellaneous (99 protein-ligand complexes, where the ligand is neither defined as an orthosteric or allosteric binder and deemed to have no modulatory activity) to rank protein cavities into categories (allosteric, regular, orthosteric)(Chen et al., 2016). A validation test was carried out on CHES, in collaboration with our work. In the PDB 158 CHES-protein binding sites were identified. Only 14 of these sites were defined as buried in pockets and from this only one was identified as an allosteric site (NanB) (Chen et al., 2016).

To further improve the success of a computational screen a direct comparison of multiple structures of the same protein (to evaluate structural conformational shifts in the presence of small molecule binding) should be including in the evaluation. Additionally, the importance of the pocket (identified to serendipitously bind a small molecule) to the protein's function should be examined and included in the evaluation. Water channels within proteins are thought to have important structural and functional roles (see Chapter 1.59, Conservation). Small molecules that bind to important structural features of proteins would be more likely to have a modulatory impact. Research groups depositing protein structures to the PDB should be aware of serendipitous small molecule binders and its potential for allosteric site discovery.

## 4.2. Experimental

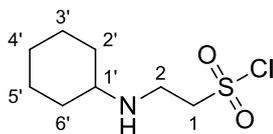
### 4.2.1 Chemistry

All reagents and chemicals were purchased from chemical suppliers including: Sigma Aldrich, Alfa Aesar, Apollo scientific, Carbosynth and Fluorochem. These chemicals were used without any further purification unless otherwise stated. Thin layer chromatography (TLC) was performed on silica coated glass plates purchased from Sigma Aldrich and Fluorochem with fluorescent indicator UV<sub>254</sub>. Developed TLC plates were air dried and analysed under a UV lamp (254/365 nm) or by staining with potassium permanganate. All moisture sensitive reactions were carried out in flame dried glassware and under a positive pressure of nitrogen. Dry THF and toluene were obtained from a dry solvent purification system (MBraun, SPS-800). Ethyl acetate was dried by distillation over K<sub>2</sub>CO<sub>3</sub>. Microwave assisted reactions were carried out using a CEM Discover microwave. Column chromatography was performed on silica gel (40-60 μm, Fluorochem).

NMR spectra were acquired using the Bruker AVANCE 300 (<sup>1</sup>H, 400 MHz; <sup>13</sup>C, 75 MHz), 400 (<sup>1</sup>H, 400 MHz; <sup>13</sup>C, 100 MHz) or 500 (<sup>1</sup>H, 500 MHz; <sup>13</sup>C, 125 MHz) spectrometers. Deuterated solvents were used as the lock for acquiring NMR spectra. All chemical shifts are reported as δ in units of ppm. <sup>13</sup>C spectra was acquired using the PENDANT or DEPTQ pulse sequence. Coupling constants (J) are reported in Hz and the following abbreviations are used for multiplicity: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad. All signals were assigned as far as possible by 2D NMR techniques including COSY, HMBC and HSQC.

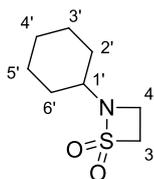
Low resolution mass was obtained from the University of St Andrews using electron spray ionisation (ESI) mass spectrometry. High resolution and low resolution mass was obtained from the ESPRC facility in Swansea. Melting points are uncorrected and measured using Electrothermal 9100 capillary melting point apparatus. FT infrared spectra were obtained using the Perkin Elmer Paragon 1000 FT spectrometer and absorption maxima reported in wavenumbers (cm<sup>-1</sup>).

## 2-(cyclohexylamino)ethanesulfonic acid (**23**)



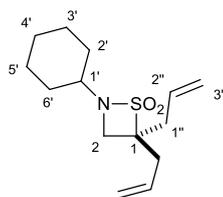
2-(cyclohexylamino)ethanesulfonic acid (4.0 g, 19 mmol) was added to phosphorous pentachloride (4.8 g, 23 mmol, 1.2 eq) and suspended in dichloromethane (40 mL). The reaction was heated under reflux for 3 hrs and then left to cool to room temperature. The desired product **23**, a white powder was then isolated by filtration and used without further purification (4.3 g, 18 mmol, 95%). M.p. 159-161°C (lit 158-160°C).  $^1\text{H}$  NMR (300 MHz,  $\text{CDCl}_3$ ):  $\delta$  9.76 (br. s., 2H,  $\text{NH}_2^+$ ), 4.52 (t,  $J = 5.7\text{Hz}$ , 2H, H1), 3.52-3.48 (m, 2H, H2), 3.06-2.96 (m, 1H, H1'), 2.17-2.15 (m, 2H, 1 x H2', 1 x H6'), 1.88-2.00 (m, 2H, 1 x H2', 1 x H6') 1.28-1.18 (m, 6H, H4', H3', H5');  $^{13}\text{C}$  NMR (75 MHz,  $\text{CDCl}_3$ ):  $\delta$  59.6 (C1), 58.6 (C1'), 39.14 (C2), 29.3 (C2', C6'), 24.6 (C4'), 24.3 (C3', C5'); LRMS ( $\text{EI}^+$ ):  $m/z$  249.10 [M+H].

## 2-Cyclohexylamino-1, 2-thiazetidene 1, 1-dioxide (**24**)



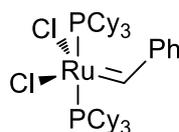
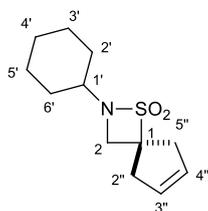
Finely ground **23** (11 g, 42 mmol) and sodium carbonate (8.9 g, 84 mmol, 2 eq) were suspended in dry ethyl acetate (300 mL), under a nitrogen atmosphere. The suspension was stirred vigorously for 48 hrs. The reaction mixture was filtered and the filtrate concentrated to give **24** as white crystals (7.5 g, 35 mmol, 85%); m.p. 78-80 °C (lit 65°C); IR (KBr)  $\nu_{\text{max}}/\text{cm}^{-1}$ : 2941 2789 2733 (CH,  $\text{CH}_2$ ), 1376 1162 ( $\text{SO}_2$ ), 1032 (C-N);  $^1\text{H}$  NMR (400MHz,  $\text{CDCl}_3$ ):  $\delta$  4.01 (t,  $J = 6.5\text{ Hz}$ , 2H, H-3), 3.17 (t,  $J = 6.5\text{ Hz}$ , 2H, H-4) 3.16 (m, 1H, H-1'), 1.86-1.78 (m, 2H, 1 x H-2', 1 x H-6'), 1.70-1.66 (m, 2H, 1 x H-3', 1 x H-5') 1.55-1.51 (m, 1H, 1 x H-4'), 1.39-1.28 (m, 5H, 1 x H-2', 1 x H-3', 1 x H-4', 1 x H-5', 1 x H-6');  $^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ ):  $\delta$  56.9 (C4), 56.7 (C1'), 34.0 (C3), 31.0 (C6', C2'), 25.1 (C4'), 24.0 (C5', C3'); LRMS ( $\text{EI}^+$ ):  $m/z$  212.00 [M+Na].

#### 4,4-Diallyl-2-cyclohexyl-1,2-thiazetidone 1,1-dioxide (25)



A solution of **24** (1.0 g, 5.3 mmol) in dry THF (10 mL) and allyl bromide (1.2 mL, 21 mmol, 4 eq) was cooled to -78 °C and freshly prepared lithium diisopropylamide (13 mmol, 2.5 eq) was added. The reaction was maintained at -78 °C for 1.5 hours. The reaction was quenched by the addition of a saturated solution of NH<sub>4</sub>Cl (50 mL) and allowed to warm to room temperature. The product was extracted with ethyl acetate (3 x 35 mL) and the organic extracts were dried over sodium sulfate and the solvent removed in *vacuo*. This gave the crude product as off white crystals which were purified by column chromatography (ethyl acetate:hexanes 15:85) using Al<sub>2</sub>O<sub>3</sub> as the solid phase to give the product **25** as white crystals (1.1 g, 4.1 mmol, 80%); m.p. 48-49 °C; IR (KBr)  $\nu_{\max}/\text{cm}^{-1}$ : 3070 (H-C=C), 2929 2857 (CH, CH<sub>2</sub>), 1636 (C=C), 1301 1144 (SO<sub>2</sub>); <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  5.80-5.70 (m, 2H, H-2''), 5.18-5.13 (m, 4H, H-3''), 3.08-3.01 (m, 1H, H-1'), 2.82 (s, 2H, H-2), 2.75-2.70 (dd,  $J = 14.6$  Hz,  $J = 7.8$  Hz, 2H, H-1<sub>a</sub>''), 2.63-2.58 (dd,  $J = 14.6, 7.9$  Hz, 2H, H-1<sub>b</sub>''), 1.69-1.65 (m, 2H, 1 x H-3', 1 x H-5'), 1.56-1.49 (m, 2H, 1 x H-2', 1 x H-6'), 1.37-1.17 (m, 6H, 1 x H-2', 1 x H-3', 2 x H-4', 1 x H-5', 1 x H-6'); <sup>13</sup>C NMR (75MHz, CDCl<sub>3</sub>):  $\delta$  132.5 (C2''), 119.7 (C3''), 59.6 (C1'), 59.3 (C1), 48.9 (C2), 36.4 (C1''), 31.2 (C2', C6'), 26.0 (C4'), 24.1 (C3', C5'); LRMS (EI<sup>+</sup>):  $m/z$  292.07 [M+Na]; HRMS (EI<sup>+</sup>): calculated for C<sub>14</sub>H<sub>23</sub>NO<sub>2</sub>SNa 292.1347, found 292.1345.

#### 2-Cyclohexylamino-4-(3-cyclopentene)-1,2-thiazetidone 1,1-dioxide (26)

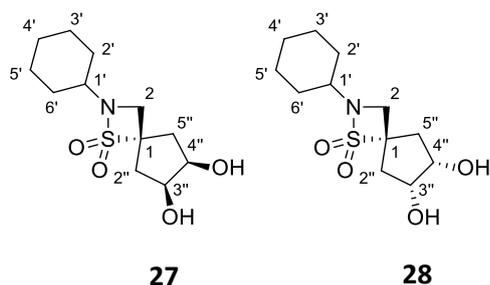


Grubbs 1<sup>st</sup> generation catalyst

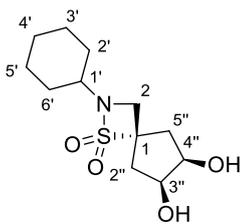
$\beta$ -Sultam **26** (1.4 g, 6.1 mmol) and benzylidene bis(tricyclohexylphosphine)dichlororuthenium (Grubbs 1<sup>st</sup> generation catalyst, 0.25 g, 0.30 mmol) in dry DCM (200 mL) was stirred at room temperature for 1.5 hours. The reaction mixture was filtered through celite and the solvent removed in *vacuo* to give the crude product as a black oil. The crude product was purified by column chromatography (ethyl acetate:hexane 20:80) using Al<sub>2</sub>O<sub>3</sub> as the solid phase to give the product **27** as a white solid (0.80 g, 3.3 mmol, 61%); m.p. 75-76 °C; IR (KBr)  $\nu_{\max}/\text{cm}^{-1}$ : 2935, 2857 (CH, CH<sub>2</sub>),

1614 (C=C), 1304, 1155 (SO<sub>2</sub>); <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ 5.61 (m, 2H, H-3'', H-4''), 3.40 (d, *J* = 16.8 Hz, 2H, H-2<sub>a</sub>'', H-5<sub>a</sub>''), 3.07 (s, 2H, H-2), 3.03 (m, 1H, H-1'), 2.65 (d, *J* = 16.8 Hz, 2H, H-2<sub>b</sub>'', H-5<sub>b</sub>''), 1.84-1.80 (m, 2H, 1 x H-2', 1 x H-6'), 1.70-1.66 (m, 2H, 1 x H-3', 1 x H-5'), 1.56-1.50 (m, 1H, 1 x H-4'), 1.38-1.14 (m, 5H, 1 x H-2', 1 x H-3', 1 x H-4', 1 x H-5', 1 x H-6'); <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>): δ 128.2 (C3'', C4''), 79.2 (C1), 56.4 (C1'), 50.1 (C2), 40.3 (C2'', C5''), 30.9 (C2', C6'), 25.6 (C4'), 23.8 (C3', C5'); LRMS (EI<sup>+</sup>): *m/z* 264.01 [M+Na]; HRMS (EI<sup>+</sup>): calculated for C<sub>12</sub>H<sub>19</sub>NO<sub>2</sub>SNa 264.1034, found 264.1033.

**2-Cyclohexylamino-4-(*cis*-3,4-cyclopentanediol)-1,2-thiazetidine 1,1-dioxide (27) and 2-Cyclohexylamino-4-(*trans*-3,4-cyclopentanediol)-1,2-thiazetidine 1,1-dioxide (28)**

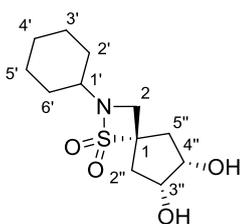


Bicycle **26** (0.80 g, 3.7 mmol) and *N*-methylmorpholine-*N*-oxide (0.65 g, 5.6 mmol) was dissolved in THF/H<sub>2</sub>O (30 mL, 9:1). A solution of osmium tetroxide (2.5% wt in <sup>t</sup>butanol) (0.015 mmol, 0.12 mL) was added to the reaction, which was allowed to stir for 16 hours at room temperature before being quenched by the addition of a saturated solution of sodium sulphite (100 mL). The reaction was extracted with ethyl acetate (3 x 40 mL). The combined organic extracts were dried over magnesium sulphate and the solvent removed in *vacuo* to give the crude products as a white solid (0.78 g). The crude reaction mixture was purified by column chromatography (ethyl acetate:hexane 80:20 to 100:0) to give the products (**27** and **28**) as 2 separate fractions **27** (0.53 g, 2.1 mmol, 57%) and **28** (0.080 g, 0.30 mmol, 8%).



**27**

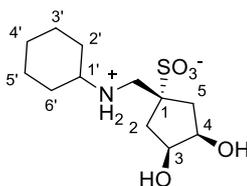
**27:** m.p. 108-109 °C; IR (KBr)  $\nu_{\max}/\text{cm}^{-1}$ : 3446 (OH), 2932, 2852 (CH, CH<sub>2</sub>), 1636 (C=C), 1309, 1145 (SO<sub>2</sub>); <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  4.24-4.19 (m, 2H, H-3'', H-4''), 3.18 (s, 2H, H-2), 3.14-3.08 (m, 1H, H-1'), 2.87 (dd,  $J = 14.8$  Hz,  $J = 5.9$  Hz, 2H, H-2<sub>a</sub>'', H-5<sub>a</sub>''), 2.43 (br. s., 2H, OH), 2.16 (dd,  $J = 14.8$  5.9 Hz, 2H, H-2<sub>b</sub>'', H-5<sub>b</sub>''), 1.90-1.87 (m, 2H, 1 x H-2', 1 x H-6'), 1.78-1.74 (m, 2H, 1 x H-3', 1 x H-5'), 1.66-1.65 (m, 1H, 1 x H-4'), 1.45-1.25 (m, 5H, 1 x H-2', 1 x H-3', 1 x H-4', 1 x H-5', 1 x H-6'); <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>):  $\delta$  76.7 (C1), 72.9 (C3'', C4''), 56.2 (C1'), 51.0 (C2), 38.1 (C2'', C5''), 30.8 (C2', C6'), 25.5 (C4'), 23.7 (C3', C5'); LRMS (EI<sup>+</sup>):  $m/z$  298.04 (M+Na), (EI<sup>-</sup>):  $m/z$  274.09 [M-H]; HRMS (EI<sup>+</sup>): calculated for C<sub>12</sub>H<sub>21</sub>NO<sub>4</sub>Na 298.1089, found 298.1092.



**28**

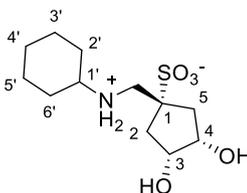
**28:** m.p. 169-170 °C; IR (KBr)  $\nu_{\max}/\text{cm}^{-1}$ : 3431 (OH), 2932, 2857 (CH, CH<sub>2</sub>), 1636 (C=C), 1294, 1150 (SO<sub>2</sub>); <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  4.24-4.19 (m, 2H, H-3'' H-4''), 3.14-3.08 (m, 1H, H-1'), 2.97 (s, 2H, H-2), 2.72 (dd,  $J = 14.8$  Hz,  $J = 3.5$  Hz, 2H, H-2<sub>a</sub>'', H-5<sub>a</sub>''), 2.43 (br. s., 2H, OH), 2.10 (dd,  $J = 14.8$  5.5 Hz, 2H, H-2<sub>b</sub>'', H-5<sub>b</sub>''), 1.90-1.87 (m, 2H, 1 x H-2', 1 x H-6'), 1.78-1.74 (m, 2H, 1 x H-3', 1 x H-5'), 1.66-1.65 (m, 1H, 1 x H-4'), 1.45-1.25 (m, 5H, 1 x H-2', 1 x H-3', 1 x H-4', 1 x H-5', 1 x H-6'); <sup>13</sup>C NMR (75MHz, CDCl<sub>3</sub>):  $\delta$  76.7 (C1), 72.9 (C3'', C4''), 56.2 (C1'), 51.0 (C2), 38.1 (C2'', C5''), 30.8 (C2', C6'), 25.5 (C4'), 23.7 (C3', C5'); LRMS (EI<sup>+</sup>):  $m/z$  298.04 [M+Na]; HRMS (EI<sup>+</sup>): calculated for C<sub>12</sub>H<sub>21</sub>NO<sub>4</sub>Na 298.1089, found 298.1095.

### 1-[(Cyclohexylamino)methyl]-*cis*-3,4-dihydroxycyclopentanesulfonate (Optactin)



**27** (10 mg, 0.035 mmol) was dissolved in H<sub>2</sub>O (5 mL) and irradiated in a sealed glass microwave tube at 140 °C (maximum 300 W) for 8 minutes. The solvent was removed in *vacuo* to give the crude product as a white solid that was purified by recrystallization from H<sub>2</sub>O/EtOH to give Optactin as white crystals (3 mg, 0.010 mmol, 28% yield); m.p. 266-267 (dec) °C; IR (KBr)  $\nu_{\max}/\text{cm}^{-1}$ : 3494 (NH<sub>2</sub><sup>+</sup>), 2935, 2863 (CH<sub>3</sub>, CH<sub>2</sub>, CH), 1220, 1183 (SO<sub>3</sub><sup>-</sup>); <sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O):  $\delta$  4.15 (t,  $J$  = 4.4 Hz, 2H, H-3, H-4), 3.38 (s, 2H, CH<sub>2</sub>), 3.13-3.05 (m, 1H, H-1'), 2.35 (dd,  $J$  = 14.9 Hz, 5.2 Hz, 2H, H-2<sub>a</sub>, H-5<sub>a</sub>), 1.98-1.94 (m, 2H, 1 x H-2', 1 x H-6'), 1.77-1.70 (m, 2H, 1 x H-3', 1 x H-5'), 1.74 (dd,  $J$  = 14.9 5.2 Hz, 2H, H-2<sub>b</sub>, H-5<sub>b</sub>), 1.57-1.52 (m, 1H, 1 x H-4'), 1.36-1.03 (m, 5H, 1 x H-2', 1 x H-3', 1 x H-4', 1 x H-5', 1 x H-6'); <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>):  $\delta$  72.8 (C3, C4), 62.0, (C1'), 58.9 (C1), 50.6 (CH<sub>2</sub>), 36.5 (C2, C5), 28.6 (C2', C6'), 24.4 (C4'), 23.8 (C3', C5'); LRMS (EI<sup>+</sup>):  $m/z$  316.11 [M+Na]; HRMS (EI<sup>+</sup>): calculated for C<sub>12</sub>H<sub>23</sub>NO<sub>5</sub>Na 316.1195, found 316.1205.

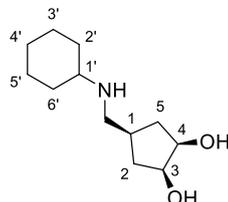
### 1-[(Cyclohexylamino)methyl]-*cis*-3,4-dihydroxycyclopentanesulfonate (29)



**28** (10 mg, 0.035 mmol) was dissolved in H<sub>2</sub>O (5 mL) and irradiated in a sealed glass microwave tube at 140 °C (maximum 300 W) for 8 minutes. The solvent was removed in *vacuo* to give the crude product as a white solid that was purified by recrystallization from H<sub>2</sub>O/EtOH to give **7** as white crystals (5 mg, 0.017 mmol, 48% yield); m.p. 229-231 (dec) °C; IR (KBr)  $\nu_{\max}/\text{cm}^{-1}$ : 3494 (NH<sub>2</sub><sup>+</sup>), 2935, 2863 (CH<sub>3</sub>, CH<sub>2</sub>, CH), 1220, 1183 (SO<sub>3</sub><sup>-</sup>); <sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O):  $\delta$  4.06 (q,  $J$  = 7.0 Hz, 1H, H-3), 3.93 (q,  $J$  = 7.3 Hz, 1H, H-4), 3.38 (s, 2H, CH<sub>2</sub>), 3.15-3.03 (m, 1H, H-1'), 2.48 (dd,  $J$  = 14.5 Hz,  $J^3$  = 7.0 Hz, 1H, H-2<sub>a</sub>), 2.10 (dd,  $J$  = 14.5 7.3 Hz, 1H, H-5<sub>a</sub>), 2.05 (dd,  $J$  = 14.5 Hz, 7.7 Hz, 1H, H-5<sub>b</sub>), 2.01-1.92 (m, 2H, 1 x H-2', 1 x H-6'), 1.77-1.68 (m, 2H, 1 x H-3', 1 x H-5'), 1.61 (dd,  $J$  = 14.5 Hz, 7.0 Hz, 1H, H-2<sub>b</sub>), 1.57-1.51 (m, 1H, 1 x H-4'), 1.39-1.15 (m, 4H, 1 x H-2', 1 x H-3', 1 x H-5', 1 x H-6'), 1.15-1.00 (m, 1H, 1 x H-4'); <sup>13</sup>C NMR (75 MHz, D<sub>2</sub>O):  $\delta$  76.1 (C3), 75.6 (C4), 60.1 (C1), 59.0 (C1'), 50.4 (CH<sub>2</sub>), 36.6 (C2), 36.2 (C5), 30.2

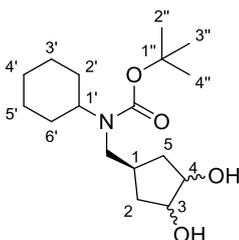
(C2'), 28.6 (C6'), 28.5 (C4'), 24.4 (C3'), 23.9 (C5'); LRMS (EI<sup>+</sup>): *m/z* 291.72 [M-H]; HRMS (EI<sup>-</sup>): calculated for C<sub>12</sub>H<sub>22</sub>NO<sub>5</sub>S 292.1219, found 292.1211.

**(1R,2S, 4r)-4-((Cyclohexylamino)methyl)cyclopentane-1,2-diol (31)**



To (1r, 3R, 4S)-*N*-Cyclohexyl-3,4-dihydroxycyclopentanecarboxamide (5.4 mg, 0.024 mmol) in dry THF (5 mL), LiAlH<sub>4</sub> (0.19 mL, 0.19 mmol, 8.0 eq) was added carefully. The reaction was stirred for 16 hrs and quenched by the addition of H<sub>2</sub>O (0.19 ml), NaOH (0.19 ml, 1 M) and H<sub>2</sub>O (0.58 ml) and then filtered through celite. The crude product was purified by silica column chromatography (treated with triethylamine) and eluted with DCM/MeOH (9:1) to give **31** as an oil (3 mg, 0.014 mmol, 30%). IR (thin film)  $\nu_{\max}/\text{cm}^{-1}$ : 3426 (OH); <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>)  $\delta$  3.88 (ddd, *J* = 5.6, 4.4, 1.5 Hz, 2H, H3, H4), 2.58 (d, *J* = 3.0 Hz, 2H, CH<sub>2</sub>), 2.38 (tt, *J* = 10.5, 3.7 Hz, 1H, H1'), 2.26 (dtd, *J* = 8.8, 5.4, 4.3, 2.0 Hz, 1H, H1), 2.14 – 2.00 (m, 2H, 1 x C2, 1 x C5), 1.97 – 1.84 (m, 2H, 1 x C2, 1 x C5), 1.66 – 1.56 (m, 1H), 1.66 – 1.56 (m, 1H, NH), 1.39 – 0.98 (m, 10H, C2', C3', C4', C5', C6'). LRMS (EI<sup>+</sup>): *m/z* 214.13 (M+H); HRMS (EI<sup>+</sup>): calculated for C<sub>12</sub>H<sub>23</sub>NO<sub>2</sub>H 214.18, found 214.18.

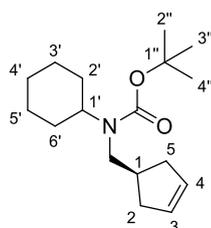
***tert*-butyl cyclohexyl((3, 4-dihydroxycyclopentyl)methyl)carbamate (32)**



To *tert*-butyl cyclohexyl(cyclopent-3-en-1-ylmethyl)carbamate (10 mg, 0.036 mmol) and *N*-methylmorpholine-*N*-oxide (72 mg, 0.60 mmol) dissolved in THF/H<sub>2</sub>O (9:1, 1 mL : 0.1 mL) was added a catalytic amount of osmium tetroxide (1.2  $\mu\text{mol}$ , 0.01 mL). After stirring at room temperature for 16 hrs, the reaction was quenched with sodium sulfite (5 mL). The product was extracted with ethyl acetate (10 mL), dried over magnesium sulfate and concentrated. The crude was then subject to column chromatography DCM/MeOH (9:1) to give **32** a mixture of diastereomers (5.9 mg, 0.019 mmol, 52%). IR (ATR)  $\nu_{\max}/\text{cm}^{-1}$ : 3428 (OH), 2976 - 2854 (CH) and 1635 (C=O) <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  4.14 (s, 2H, H3,H4), 3.10 (s, 2H, CH<sub>2</sub>), 2.52 (q, *J* = 7.9 Hz, 1H, H1), 2.40-2.35 (br, s, 2H, OH)

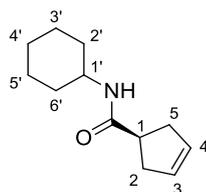
1.83 – 1.74 (m, 3H, 1 x H2, 1x H5, 1 x H1' ), 1.73 – 1.67 (m, 2H, 1 x H2', 1 x H6'), 1.65 – 1.56 (m, 2H, 1 x H2, 1 H5), 1.45 (s, 9H, H1'', H2'', H3''), 1.32 – 1.20 (m, 2H, 1 x H2', 1x H6'), 1.16 – 1.03 (m, 6H, H3', H4', H5'). LRMS (EI<sup>+</sup>): m/z 314.23 (100) (M +H); HRMS (EI<sup>+</sup>): calculated for C<sub>17</sub>H<sub>31</sub>NO<sub>4</sub>H 314.2326, found 314.2330.

***tert*-Butyl cyclohexyl(cyclopent-3-en-1-ylmethyl)carbamate (**33**)**



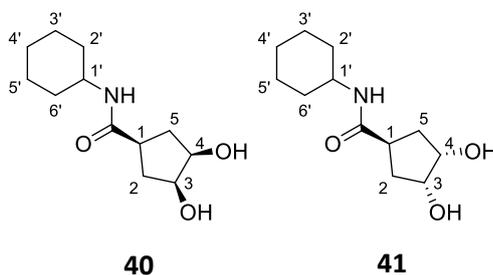
3-Cyclopentene carboxylic acid (0.064 ml, 0.54 mmol) was added to dry THF (5 mL) and LiAlH<sub>4</sub> (2.16 mmol, 4.0 eq) and left stirring at room temperature for 16 hrs. The reaction was quenched by the sequential addition of H<sub>2</sub>O (0.19 mL), NaOH (0.19 mL, 1 M) and H<sub>2</sub>O (0.58 mL) and then filtered through Celite<sup>®</sup>. The filtrate was washed with DCM. Dess-Martin Periodinane (0.46 g, 1.08 mmol, 2.0 eq) was then added to the DCM solution containing the crude product. The reaction was left for 16 hrs. The reaction was then quenched with sodium thiosulfate (5 mL) and sodium bicarbonate (5 mL). The crude product was then subject to extraction with DCM. Activated molecular sieves were added to the crude product and then cyclohexylamine (0.54 mmol, 1.0 eq) was added. The reaction was left stirring for 1 hr at room temperature and then NaBH<sub>4</sub> (30 mg, 0.81 mmol, 1.5 eq) added. 4 mL of NH<sub>4</sub>Cl was then added and aqueous separated from DCM layer. DIPEA (28 mg, 0.24 mmol) and di-*tert*-butyl dicarbonate (0.14 g, 0.65 mmol, 1.1 eq) were added to the mixture and left at room temperature for 30 mins. The crude product was concentrated and subject to column chromatography (DCM/MeOH 19:1) to afford **33** as an oil (24 mg, 0.076 mmol, 14%). IR (ATR)  $\nu_{\text{max}}/\text{cm}^{-1}$ : 2976 - 2854 (CH), 1685 (C=O) and 1543 (C=C); <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  5.54 (s, 2H, H3, H4), 3.01 – 2.96 (s, 2H, CH<sub>2</sub>), 2.53 – 2.37 (m, 1H, H1), 2.37 – 2.16 (m, 2H, 1x H2, 1 x H5), 2.06 – 1.84 (m, 3H, H1', 1 x H2, 1 x H5), 1.74 – 1.03 (br. m, 19H, H1'', H3'', H4'', H2', H6', H5', H3' H4' ). LRMS (EI<sup>+</sup>): m/z 279.84 (100).

### ***N*-Cyclohexylcyclopent-3-enecarboxamide (39)**



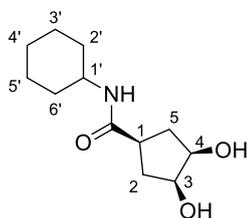
To a solution of 3-cyclopentene carboxylic acid **37** (0.46 mL, 4.5 mmol, 1 eq) in DCM (mL), 1, 1'-carbonyl diimidazole (0.56 g, 4.91 mmol, 1.1 eq) was added in portions. After 45 mins, cyclohexylamine (0.56 g, 4.91 mmol, 1.1 eq.) was added and left to stir for 4-5 hrs. The reaction was quenched with an aqueous solution of HCl (60 mL, 1 M). The product was extracted using DCM (50 mL). The product was then washed with an aqueous solution of HCl (30 mL, 1M), H<sub>2</sub>O (30 mL), a 50% saturated aqueous Na<sub>2</sub>CO<sub>3</sub>/NaCl solution (30 mL), dried and filtered through a cotton plug. The filtrate was diluted with 10% EtOAc/CH<sub>2</sub>Cl<sub>2</sub> and filtered through silica plug eluting with 10% EtOAc/CH<sub>2</sub>Cl<sub>2</sub> to remove the baseline color. Subsequent concentration in *vacuo* affords amide alkene **39** as a white solid (0.80 g, 3.6 mmol, 80% yield); m.p. 151-153 °C. IR (ATR)  $\nu_{\max}/\text{cm}^{-1}$ : 3283(NH), 2924 - 2846 (CH), 1635 (C=O) and 1543 (C=C); <sup>1</sup>H NMR (500 MHz, Methanol-*d*<sub>4</sub>): 5.65 – 5.60 (m, 2H, H<sub>3</sub>, H<sub>4</sub>), 3.64 - 3.58 (m, 1H, H<sub>1'</sub>), 3.05 - 2.94 (m, 1H, H<sub>1</sub>), 2.58 - 2.49 (m, 4H, H<sub>2</sub>, H<sub>5</sub>), 1.91 - 1.15 (m, 10H, H<sub>2'</sub>, H<sub>3'</sub>, H<sub>4'</sub>, H<sub>5'</sub>, H<sub>6'</sub>); <sup>13</sup>C NMR (126 MHz, Methanol-*d*<sub>4</sub>)  $\delta$  176.4, (C=O), 128.6 (C<sub>3</sub>, C<sub>4</sub>), 48.3 (C<sub>1'</sub>), 42.8 (C<sub>1</sub>), 36.6 (C<sub>2</sub>, C<sub>5</sub>), 32.4, 25.2, 24.8 (C<sub>2'</sub>, C<sub>3'</sub>, C<sub>4'</sub>, C<sub>5'</sub>, C<sub>6'</sub>). LR-MS (ES<sup>+</sup>): *m/z* 194.02 [M+H]. HRMS (EI<sup>+</sup>): calculated for C<sub>12</sub>H<sub>20</sub>NO 194.1539, found 194.1538.

**(3*S*,4*R*)-*N*-Cyclohexyl-3,4-dihydroxycyclopentanecarboxamide (40) and (3*R*,4*S*)-*N*-Cyclohexyl-3,4-dihydroxycyclopentanecarboxamide (41).**



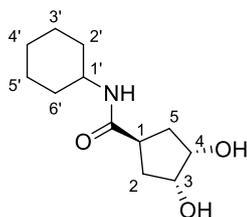
To a solution of **39** (100 mg, 0.52 mmol, 1 eq.) and *N*-methylmorpholine-*N*-oxide (1.2 eq.) dissolved in THF/H<sub>2</sub>O (9:1) was added a catalytic amount of osmium tetroxide (0.012 mmol, 0.1 mL) and stirred at room temperature overnight. The reaction was then quenched by the addition of sodium sulphite (30 mL). The product was extracted with ethyl acetate (100 mL), dried over

magnesium sulphate and concentrated in *vacuo*. The crude product was then purified using flash column chromatography (10% MeOH/DCM) to give products **40** (40% yield) and **41** (17% yield) as white solids.



**40**

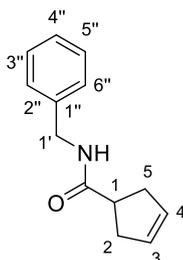
**40:** m.p. 134-135 °C; IR (thin film)  $\nu_{\max}/\text{cm}^{-1}$ : 3299 (OH), 1639 (C=O) and 740;  $^1\text{H}$  NMR (500 MHz, Methanol-*d*4)  $\delta$  3.98 – 3.91 (m, 2H, H3, H4), 3.64 – 3.57 (m, 1H, H1'), 2.76 – 2.68 (m, 1H, H1), 2.17 – 2.08 (m, 2H, 1 x H2, 1 x H5), 1.86 (m, 2H, 1 x H2', 1 x H6'), 1.83 – 1.73 (m, 4H, 1 x H2, 1x H5, 1x H2', 1 x H6'), 1.67 – 1.62 (m, 1H, 1 x H4'), 1.43 – 1.33 (m, 2H, 1 x H3', 1 x H5'), 1.23 (m, 3H, 1 x H3', 1 x H4', 1x H5');  $^{13}\text{C}$  NMR (126 MHz, Methanol-*d*4)  $\delta$  177.2 (C=O), 73.6 (C3, C4), 48.4 (C1'), 40.3 (C1), 34.3 (C2, C5), 32.2 (C2', C6'), 25.2 (C4'), 24.6 (C3', C5'); LR-MS (ES+):  $m/z$  228.15 [M+H]; HRMS (EI<sup>+</sup>): calculated for  $\text{C}_{12}\text{H}_{22}\text{NO}_3$  228.1594, found 228.1594.



**41**

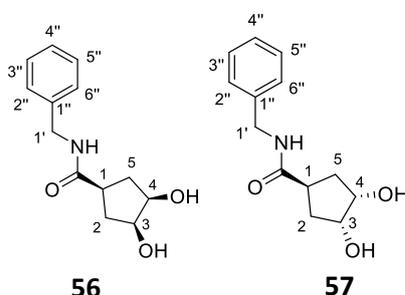
**41:** m.p. 179-181 °C; IR (thin film)  $\nu_{\max}/\text{cm}^{-1}$ : 3299 (OH), 1635 (C=O) 1540, 1264 and 746; (C=C)  $^1\text{H}$  NMR (500 MHz, Methanol-*d*4)  $\delta$  4.15 – 4.09 (m, 2H, H3, H4), 3.65 – 3.55 (m, 1H, H1'), 2.98 (m, 1H, H1), 1.97 – 1.81 (m, 6H, H2, H5, 1 x H6', 1 x H2'), 1.81 – 1.73 (m, 1H, 1 x H4'), 1.66 (m, 2H, 1 x H5', 1 x H3'), 1.41 – 1.30 (m, 2H, 1 x H5', 1x H3'), 1.26 – 1.15 (m, 3H, 1x H4', 1 x H6', 1 x H2');  $^{13}\text{C}$  NMR (126 MHz, Methanol-*d*4)  $\delta$  175.2 (C=O), 73.6 (C3, C4), 48.3 (C1'), 40.2 (C1), 34.5 (C2, C5), 32.4 (C2', C6'), 25.2 (C4'), 24.8 (C3', C5'); LR-MS (ES+):  $m/z$  228.15 [M+H]. HRMS (EI<sup>+</sup>): calculated for  $\text{C}_{12}\text{H}_{22}\text{NO}_3$  228.1594, found 228.1594.

### N-benzylcyclopent-3-ene-1-carboxamide (**58**)



To a solution of 3-cyclopentene carboxylic acid **37** was reacted with phenylmethanamine (1 eq) using carbonyl diimidazole (1.2 eq). The reaction was quenched with an aqueous solution of HCl (60 mL, 1 M). The product was extracted using DCM (50 mL). The product was then washed with an aqueous solution of HCl (30 mL, 1M), H<sub>2</sub>O (30 mL), a 50% saturated aqueous Na<sub>2</sub>CO<sub>3</sub>/NaCl solution (30 mL), dried and filtered through a cotton plug. The filtrate was diluted with 10% EtOAc/CH<sub>2</sub>Cl<sub>2</sub> and filtered through silica plug eluting with 10% EtOAc/CH<sub>2</sub>Cl<sub>2</sub> to remove the baseline color. Subsequent concentration in *vacuo* affords **58** as a white solid (81% yield); m.p. 92-93 °C; ; IR (ATR)  $\nu_{\text{max}}/\text{cm}^{-1}$ : 3271 (NH), 3055-2846 (CH), 1635 (C=O), 1543 (C=C), 817 (C-Cl), 678 (C-H); <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>)  $\delta$  7.31 (d, J = 8.2 Hz, 1H, H3'), 7.27 (d, J = 2.0 Hz, 1H, H4'), 7.03 (dd, J = 8.2, 2.1 Hz, 1H, H2'), 5.96 (m, 1H, NH), 5.68 – 5.54 (m, 2H, H3, H4), 4.31 (d, J = 6.1 Hz, 2H, H1''), 2.93 (quin, J = 7.9 Hz, 1H, H1), 2.64 -2.50 (m, 4H, H2, H5). <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>)  $\delta$  176.4 (C=O), 139.3 (C1'), 131.0 (C3'), 129.9 (C4'), 129.6 (C3, C4), 127.4 (C2'), 77.5 (2 x C-Cl), 43.8 (C1), 42.8 (C1''), 37.4 (C2, C5). LR-MS (ES+): m/z 271.98 (M+H).

**(3S,4R)-N-benzyl-3,4-dihydroxycyclopentanecarboxamide (56)** and **(3R,4S)-N-benzyl-3,4-dihydroxycyclopentanecarboxamide (57)**.



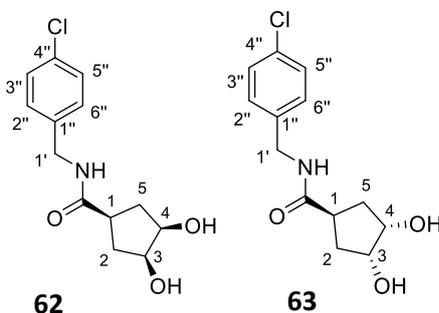
To a solution of **58** and *N*-methylmorpholine-*N*-oxide (1.2 eq.) dissolved in THF/H<sub>2</sub>O (9:1) was added a catalytic amount of osmium tetroxide (1 mol %) and stirred at room temperature overnight.

The reaction was then quenched by the addition of sodium sulphite (30 mL). The product was extracted with ethyl acetate (100 mL), dried over magnesium sulphate and concentrated in *vacuo*. The crude product was then purified using flash column chromatography (10% MeOH/DCM) to give products **56** (43% yield) and **57** (40% yield) as white solids.

**56**: m.p. 125-127 °C; IR (ATR)  $\nu_{\text{max}}/\text{cm}^{-1}$ : 3295 (OH), 1635 (C=O), 1539 (Ar C=C), 692 (C-H);  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  7.40–7.28 (m, 5H, H2', H3', H4', H5', H6'), 6.11 (s, 1H, NH), 4.46 (d,  $J = 5.2$  Hz, 2H, H1'), 3.99-3.91 (m, 2H, H3, H4), 3.90 (d,  $J = 8.4$  Hz, 2H, 2 x OH), 2.80-2.66 (m, 1H, H1), 2.25-2.17 (m, 2H, 1 x H2, 1 x H5), 1.94-1.88 (m, 2H, 1 x H2, 1 x H5);  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  178.6 (C=O), 138.4 (C1'), 128.9 and 127.9 (C2', C3', C4, C5', C6'), 74.3 (C3, C4), 44.1 (C1''), 42 (C1), 35.3 (C2, C5); LR-MS (ES+):  $m/z$  236.11 [M+H]; HRMS (EI+): calculated for  $\text{C}_{13}\text{H}_{17}\text{NO}_3$  235.1200, found 236.1200 [M+H].

**57**: m.p. 154-155 °C; IR (ATR)  $\nu_{\text{max}}/\text{cm}^{-1}$ : 3275 (OH), 1635 (Amide C=O), 1541 (Ar C=C), 692 (C-H);  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  7.38-7.25 (m, 5H, H2', H3', H4', H5', H6'), 5.70 (br s, 1H, NH), 4.43 (d,  $J = 6.26$  Hz, 2H, H1'), 4.31-4.25 (s, 2H, H3, H4), 2.99-2.92 (m, 1H, H1), 2.18-2.10 (m, 2H, 1 x H2, 1 x H5), 2.00-1.95 (m, 2H, 1 x H2, 1 x H5);  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  177.5 (C=O), 137.9 (C1'), 128.7 and 127.6 (C2', C3', C4, C5', C6'), 73.6 (C3, C4), 43.8 (C1''), 41.3 (C1), 35.4 (C2, C5); LR-MS (ES+):  $m/z$  236.11 [M+H]; HRMS (EI+): calculated for  $\text{C}_{13}\text{H}_{17}\text{NO}_3$  235.1200, found 236.1200 [M+H].

**(3S,4R)-N-(4-chlorobenzyl)-3,4-dihydroxycyclopentanecarboxamide (62)** and **(3R,4S)-N-(4-chlorobenzyl)-3,4-dihydroxycyclopentanecarboxamide (63)**



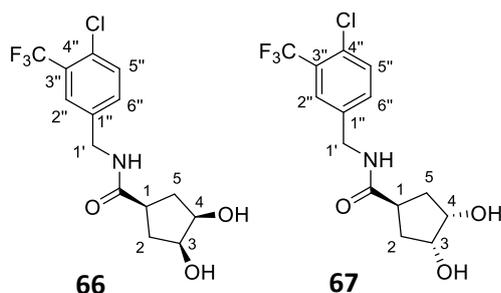
To a solution of **85** and *N*-methylmorpholine-*N*-oxide (1.2 eq.) dissolved in THF/ $\text{H}_2\text{O}$  (9:1) was added a catalytic amount of osmium tetroxide (1 mol %) and stirred at room temperature overnight. The reaction was then quenched by the addition of sodium sulphite (30 mL). The product was

extracted with ethyl acetate (100 mL), dried over magnesium sulphate and concentrated in *vacuo*. The crude product was then purified using flash column chromatography (10% MeOH/DCM) to give products **62** (50% yield) and **63** (48% yield) as white solids.

**62**: m.p. 127-128 °C; IR (ATR)  $\nu_{\text{max}}/\text{cm}^{-1}$ : 3296.35 (OH), 1637.56 (C=O), 1527.92 (Ar C=C), 1089.78 (CH);  $^1\text{H NMR}$  (300 MHz,  $\text{CDCl}_3$ )  $\delta$  7.32 (d,  $J = 8.04$  Hz, 2H, H3', H4'), 7.20 (d,  $J = 8.04$  Hz, 2H, H2', H5'), 6.15 (br s, 1H, NH), 4.40-4.38 (m, 2H, H1''), 4.08-3.97 (m, 2H, H3, H4), 2.77-2.67 (m, 1H, H1), 2.26-2.11 (m, 2H, 1 x H2, 1 x H5), 1.93-1.80 (m, 2H, 1 x H2, 1 x H5);  $^{13}\text{C NMR}$  (75 MHz,  $\text{CDCl}_3$ )  $\delta$  179.2 (C=O), 136.8 (C1'), 129.5 (C2', C5'), 74.7 (C3, C4), 64.9 (C-Cl), 43.7 (C1''), 42.4 (C1), 35.7 (C2, C5), LR-MS (ES+):  $m/z$  292.05 [M+Na]; HRMS (EI+): calculated for  $\text{C}_{13}\text{H}_{16}\text{ClNO}_3$  269.0819, found 270.0894 [M+H].

**63**: m.p. 166-168 °C; IR (ATR)  $\nu_{\text{max}}/\text{cm}^{-1}$ : 3271.27 (OH), 1633.71 (Amide RCONHR), 1539.20 (Ar C=C), 1062 (C-H);  $^1\text{H NMR}$  (300 MHz,  $\text{CDCl}_3$ )  $\delta$  7.23 (d,  $J = 8.25$  Hz, 2H, H3', H4'), 7.12 (d,  $J = 8.25$  Hz, 2H, H2', H5'), 5.67 (br s, 1H, NH), 4.34-4.32 (m, 2H, H1''), 4.25-4.17 (m, 2H, H3, H4), 2.96-2.83 (m, 1H, H1), 2.10 – 1.84 (m, 4H, H2, H5);  $^{13}\text{C NMR}$  (75 MHz,  $\text{CDCl}_3$ )  $\delta$  179.2 (C=O), 136.8 (C1'), 129.5 (C2', C5'), 129.4 (C3', C4'), 74.7 (C3, C4), 64.9 (C-Cl), 43.7 (C1''), 42.4 (C1), 35.7 (C2, C5); LR-MS (ES+):  $m/z$  291.99 [M+Na];

**(3S,4R)-N-(4-chloro-3-(trifluoromethyl)benzyl)-3,4-dihydroxycyclopentanecarboxamide (66)**  
**and (3R,4S)-N-(4-chloro-3-(trifluoromethyl)benzyl)-3,4-dihydroxycyclopentanecarboxamide (67)**

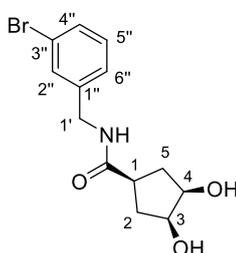


To a solution of **85** and *N*-methylmorpholine-*N*-oxide (1.2 eq.) dissolved in THF/ $\text{H}_2\text{O}$  (9:1) was added a catalytic amount of osmium tetroxide (1 mol %) and stirred at room temperature overnight. The reaction was then quenched by the addition of sodium sulphite (30 mL). The product was extracted with ethyl acetate (100 mL), dried over magnesium sulphate and concentrated in *vacuo*. The crude product was then purified using flash column chromatography (10% MeOH/DCM) to give products **66** (39 % yield) and **67** (34 % yield) as white solids.

**66**: m. p. 145-146 °C; IR (ATR)  $\nu_{\text{max}}/\text{cm}^{-1}$ : 3302 (OH), 2939 (CH), 1635 (C=O), 1134 (C-F), 663 (C-Cl);  $^1\text{H}$  NMR (300 MHz, Methanol- $d_4$ )  $\delta$  7.60-7.56 (m, 1H, H4'), 7.48-7.42 (m, 1H, H3'), 7.42-7.38 (m, 1H, H2'), 4.30 (s, 2H, H1''), 3.90-3.79 (m, 2H, H3, H4), 2.68 (tt,  $J = 9.2, 6.8$  Hz, 1H, H1), 2.14-1.96 (m, 2H, 1 x H2, 1 x H5), 1.83 - 1.65 (m, 2H, 1 x H2, 1 x H5);  $^{13}\text{C}$  NMR (75 MHz, Methanol- $d_4$ )  $\delta$  179.7 (C=O), 140.6 (C1'), 134.0 (C2'), 133.2 (C3'), 128.1 (C4'), 126.5 (C-CF3), 75.2 (C3, C4), 48.9 (CF3), 43.6 (C1''), 41.8 (C1), 36.1 (C2, C5); LR-MS (ES+):  $m/z$  337.99 [M+H]; HRMS (EI+): calculated for  $\text{C}_{14}\text{H}_{15}\text{ClF}_3\text{NO}_3\text{Na}$  360.0590, found 360.0580 [M+Na].

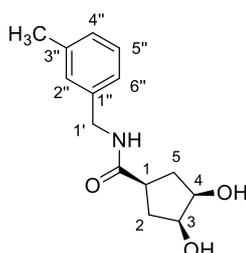
**67**: m. p. 147-149 °C; IR (ATR)  $\nu_{\text{max}}/\text{cm}^{-1}$ : 3282 (OH), 1635 (C=O), 1539 (C=C), 1114 (C-F), 659 (C-Cl);  $^1\text{H}$  NMR (300 MHz, Methanol- $d_4$ )  $\delta$  7.59-7.54 (m, 1H, H4'), 7.49-7.41 (m, 1H, H3'), 7.41-7.32 (m, 1H, H2'), 4.28 (s, 2H, H1''), 4.04-3.96 (m, 2H, H3, H4), 2.97 (quin,  $J = 8.2$  Hz, 1H, H1), 1.87-1.78 (m, 4H, H2, H5);  $^{13}\text{C}$  NMR (75 MHz, Methanol- $d_4$ )  $\delta$  178.6 (C=O), 140.7 (C1'), 133.9 (C2'), 133.2 (C3'), 128.0 (C4'), 127.8 (C-CF3), 75.2 (C3, C4), 47.9 (CF3), 43.5 (C1''), 42.0 (C1), 36.2 (C2, C5). LR-MS (ES+):  $m/z$  338.07 [M+H]; HRMS (EI+): calculated for  $\text{C}_{14}\text{H}_{15}\text{ClF}_3\text{NO}_3\text{Na}$  360.0590, found 360.0591 [M+Na].

**(1*r*,3*R*,4*S*)-N-(3-bromobenzyl)-3,4-dihydroxycyclopentane-1-carboxamide (68)**



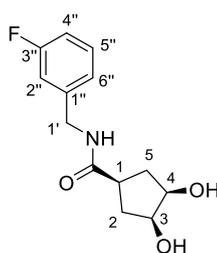
To a solution of **87** and *N*-methylmorpholine-*N*-oxide (1.2 eq.) dissolved in THF/ $\text{H}_2\text{O}$  (9:1) was added a catalytic amount of osmium tetroxide (1 mol %) and stirred at room temperature overnight. The reaction was then quenched by the addition of sodium sulphite (30 mL). The product was extracted with ethyl acetate (100 mL), dried over magnesium sulphate and concentrated in *vacuo*. The crude product was then purified using flash column chromatography (10% MeOH/DCM). This led to the isolation of only one of the two diastereomers to give **68** (33% yield) as a white solid; m. p. 103-105 °C; IR (ATR)  $\nu_{\text{max}}/\text{cm}^{-1}$ : 3336-3244 (OH), 2920 (CH), 1631 (C=O), 1543 (C=C), 690 (C-Br);  $^1\text{H}$  NMR (500 MHz, Methanol- $d_4$ )  $\delta$  7.48-7.45 (m, 1H, H5'), 7.43-7.41 (m, 1H, H4'), 7.27-7.24 (m, 1H, H2', H3'), 4.36 (s, 2H, H1''), 4.00-3.92 (m, 2H, H3, H4), 2.84-2.74 (m, 1H, H1), 2.22-2.11 (m, 2H, 1 x H2, 1 x H5), 1.91 - 1.80 (m, 2H, 1 x H2, 1 x H5);  $^{13}\text{C}$  NMR (126 MHz, Methanol- $d_4$ )  $\delta$  177.8 (C=O), 130.1 (C5'), 129.9 (C4'), 129.8 (C2'), 125.9 (C3'), 73.4 (C3, C4), 42.2 (C1''), 40.1 (C1), 34.8 (C2, C5); LR-MS (ES+):  $m/z$  313.98 (M+H).

**(1*r*,3*R*,4*S*)-3,4-dihydroxy-N-(3-methylbenzyl)cyclopentane-1-carboxamide (69)**



Ester **77** was subjected to a hydrolysis using NaOH. After a work up the crude product was amide coupled with *m*-tolylmethanamine (1eq) using carbonyl-diimidazole (1.2eq). After an acid base work up, the product **70** was extracted with DCM (100 mL), dried over magnesium sulphate and concentrated in *vacuo* to afford a white powder (23% yield). m.p. 102-104°C. IR (ATR)  $\nu_{\text{max}}/\text{cm}^{-1}$ : 3402-3296 (OH), 2968-2883 (CH), 1656 (C=O), 1529-1440 (C=C), 1348 (C-H).  $^1\text{H}$  NMR (500 MHz, Methanol- $d_4$ )  $\delta$  7.22-7.17 (t,  $J = 7.5$  Hz, 1H, H5''), 7.11-7.04 (m, 3H, H2'', H4'', H6''), 4.32 (s, 2H, H1'), 3.98-3.92 (m, 2H, H3, H4), 2.78 (tt,  $J = 9.3, 6.6$  Hz, 1H, H1), 2.19-2.10 (m, 2H, 1 x H2, 1 x H5), 1.89-1.81 (m, 2H, 1 x H2, 1 x H5).  $^{13}\text{C}$  NMR (75 MHz, Methanol- $d_4$ )  $\delta$  179.2 (C=O), 139.8 (C3''), 139.3 (C1''), 129.5 (C2''), 129.2 (C4''), 128.9 (C6'') 125.6 (C5'') 74.9 (C3, C4), 44.3 (CH<sub>3</sub>), 43.6 (C1'), 41.6 (C1), 35.7 (C2, C5). LR-MS (ES<sup>+</sup>):  $m/z$  251.1 [M+H]; HRMS (EI<sup>+</sup>): calculated for C<sub>14</sub>H<sub>20</sub>O<sub>3</sub>N 250.1438, found 250.1441.

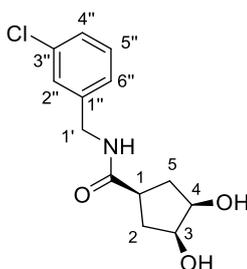
**(1*r*,3*R*,4*S*)-N-(3-fluorobenzyl)-3,4-dihydroxycyclopentane-1-carboxamide (70)**



Ester **77** was subjected to a hydrolysis using NaOH. After a work up the crude product was amide coupled with (3-fluorophenyl)methanamine (1 eq) using carbonyl-diimidazole (1.2eq). After an acid base work up, the product **70** was extracted with DCM (100 mL), dried over magnesium sulphate and concentrated in *vacuo* to afford a white powder (23% yield). m.p. 106-108°C. IR (ATR)  $\nu_{\text{max}}/\text{cm}^{-1}$ : 3267 (OH), 3059-2839 (CH), 1616 (C=O), 1539-1435 (C=C), 1251-1114 (C-F).  $^1\text{H}$  NMR (500 MHz, Methanol- $d_4$ )  $\delta$  7.36-7.29 (m, 1H, H5''), 7.11-7.07 (m, 1H, H2''), 7.03-6.94 (m, 2H, H4'', H6''), 4.36 (s, 2H, H1'), 3.98-3.92 (m, 2H, H3, H4), 2.78 (tt,  $J = 9.2, 6.7$  Hz, 1H, H1), 2.19-2.10 (m, 2H, 1 x H2, 1 x H5),

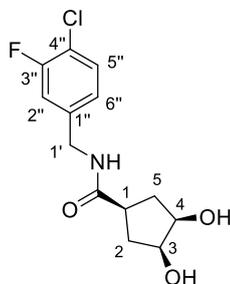
1.89–1.80 (m, 2H, 1 x H2, 1 x H5).  $^{13}\text{C}$  NMR (75 MHz, Methanol-d4)  $\delta\text{C}$  179.5 (C=O), 163.4 (C3''), 143.2 (C1''), 131.5 (d, J= 8.4 Hz, C2''), 124.4 (C6''), 115.4 (d, J= 8.4 Hz, C4''), 115.2 (d, J= 8.4 Hz, C5''), 75.1 (C3, C4), 43.9 (C1'), 41.7 (C1), 35.9 (C2, C5). LR-MS (ES+): m/z 255.1 [M+H]; HRMS (EI+): calculated for  $\text{C}_{13}\text{H}_{17}\text{O}_3\text{NF}$  254.1190, found 254.1194.

**(1r,3R,4S)-N-(3-chlorobenzyl)-3,4-dihydroxycyclopentane-1-carboxamide (71)**



Ester **77** was subjected to a hydrolysis using NaOH. After a work up the crude product was amide coupled with (3-chlorophenyl)methanamine (1eq) using carbonyl-diimidazole (1.2eq). After an acid base work up, the product **71** was extracted with DCM (100 mL), dried over magnesium sulphate and concentrated in *vacuo* to afford a white powder (34% yield). m.p. 105-107°C. IR (ATR)  $\nu_{\text{max}}/\text{cm}^{-1}$ : 3250 (OH), 2922 (CH), 1633 (C=O), 1552-1396 (C=C), 788 (C-Cl).  $^1\text{H}$  NMR (500 MHz, Methanol-d4)  $\delta$  7.33-7.25 (m, 2H, H4'', H6''), 7.27-7.23 (m, 1H, H4''), 7.22-7.18 (m, 1H, H2''), 4.35 (s, 2H, H1'), 3.97-3.92 (m, 2H, H3, H4), 2.78 (tt, J = 9.2 Hz, 6.7 Hz, 1H, H1), 2.19-2.11 (m, 2H, 1 x H2, 1 x H5), 1.88–1.81 (m, 2H, 1 x H2, 1 x H5).  $^{13}\text{C}$  NMR (75 MHz, Methanol-d4)  $\delta\text{C}$  179.3 (C=O), 142.5 (C1''), 135.3 (C3''), 131.1 (C2''), 128.5 (C6''), 128.3 (C4''), 126.9 (C5''), 74.8 (C3, C4), 43.6 (C1'), 41.5 (C1), 35.7 (C2, C5). LR-MS (ES+): m/z 271.1 [M+H]; HRMS (EI+): calculated for  $\text{C}_{13}\text{H}_{17}\text{ClNO}_3$  270.0896 found 270.0896

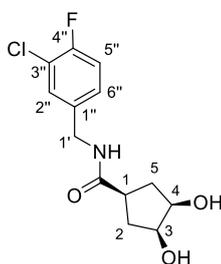
**(1r,3R,4S)-N-(4-chloro-3-fluorobenzyl)-3,4-dihydroxycyclopentane-1-carboxamide (72)**



Ester **77** was subjected to a hydrolysis using NaOH. After a work up the crude product was amide coupled with (4-chloro-3-fluorophenyl)methanamine (1eq) using carbonyl-diimidazole (1.2eq). After

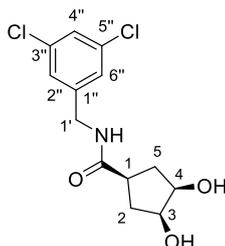
an acid base work up, the product **72** was extracted with DCM (100 mL), dried over magnesium sulphate and concentrated in *vacuo* to afford a white powder (22% yield). m.p. 127-129°C. IR (ATR)  $\nu_{\text{max/cm-1}}$ : 3329 (OH), 2939-2872 (CH), 1628 (C=O), 1585-1489 (C=C), 1099-1014 (C-F) 817 (C-Cl).  $^1\text{H}$  NMR (500 MHz, Methanol- $d_4$ )  $\delta$  7.43 (t,  $J = 9.2, 6.8$  Hz, 1H, H5''), 7.17 (dd,  $J = 9.2, 6.8$  Hz, 1H, H6''), 7.13-7.09 (m, 1H, H2''), 4.36 (s, 2H, H1'), 3.95-3.94 (m, 2H, H3, H4), 2.84-2.75 (tt,  $J = 9.2, 6.8$  Hz, 1H, H1), 2.20-2.12 (m, 2H, 1 x H2, 1 x H5), 1.90-1.80 (m, 2H, 1 x H2, 1 x H5).  $^{13}\text{C}$  NMR (75 MHz, Methanol- $d_4$ )  $\delta_{\text{C}}$  179.7 (C=O), 164.9 (C3''), 142.3 (C1''), 132.1 (C5''), 125.7 (C6''), 117.1 (C4''), 116.9 (C2''), 75.2 (C3, C4), 43.7 (C1'), 41.9 (C1), 36.1 (C2, C5).

**(1r,3R,4S)-N-(3-chloro-4-fluorobenzyl)-3,4-dihydroxycyclopentane-1-carboxamide (73)**



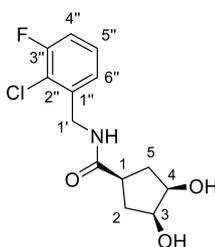
Ester **77** was subjected to a hydrolysis using NaOH. After a work up the crude product was amide coupled with (3-chloro-4-fluorophenyl)methanamine (1eq) using carbonyl-diimidazole (1.2eq) After an acid base work up, the product **73** was extracted with DCM (100 mL), dried over magnesium sulphate and concentrated in *vacuo* to afford a white powder (18% yield). IR (ATR)  $\nu_{\text{max/cm-1}}$ : 3280 (OH), 2918 (CH), 1640 (C=O), 1543-1500 (C=C), 1114-1026 (C-F) 819 (C-Cl). m.p. 112-113°C.  $^1\text{H}$  NMR (500 MHz, Methanol- $d_4$ )  $\delta$  7.38 (dd,  $J = 7.1, 2.1$  Hz, 1H, H2''), 7.24-7.15 (m, 2H, H5'', H6''), 4.30 (s, 2H, H1'), 4.13-4.08 (m, 2H, H3, H4), 3.06 (tt,  $J = 7.1, 2.1$  Hz, 1H, H1), 2.01-1.83 (m, 4H, H2, H5).  $^{13}\text{C}$  NMR (75 MHz, Methanol- $d_4$ )  $\delta_{\text{C}}$  177.5 (C=O), 158.3 (C4''), 148.2 (C1''), 136.7 (C2''), 129.3 (C6''), 129.8 (C3''), 116.9 (C5''), 75.2 (C3, C4), 41.6 (C1'), 40.2 (C1), 34.4 (C2, C5). LR-MS (ES+):  $m/z$  289.1 [M+H]; HRMS (EI+): calculated for  $\text{C}_{13}\text{H}_{15}\text{ClFNO}_3$  288.0800 found 288.0801

**(1r,3R,4S)-N-(3,5-dichlorobenzyl)-3,4-dihydroxycyclopentane-1-carboxamide (74)**



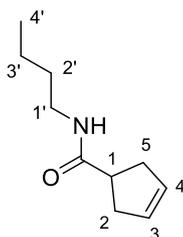
Ester **77** was subjected to a hydrolysis using NaOH. After a work up the crude product was amide coupled with (3,5-dichlorophenyl)methanamine (1eq) using carbonyl-diimidazole (1.2eq). After an acid base work up, the product **74** was extracted with DCM (100 mL), dried over magnesium sulphate and concentrated in *vacuo* to afford a white powder (22% yield). m.p. 121-123°C. <sup>1</sup>H NMR (500 MHz, Methanol-d<sub>4</sub>) δ 7.32 (t, J = 9.2, 6.9 Hz, 1H, H<sub>4''</sub>), 7.25-7.24 (m, 2H, H<sub>2''</sub>, H<sub>6''</sub>), 4.33 (s, 2H, H<sub>1'</sub>), 3.96-3.93 (m, 2H, H<sub>3</sub>, H<sub>4</sub>), 2.78 (tt, J = 9.2, 6.9 Hz, 1H, H<sub>1</sub>), 2.20-2.10 (m, 2H, 1 x H<sub>2</sub>, 1 x H<sub>5</sub>), 1.90-1.80 (m, 2H, 1 x H<sub>2</sub>, 1 x H<sub>5</sub>). <sup>13</sup>C NMR (75 MHz, Methanol-d<sub>4</sub>) δC 178.9 (C=O), 143.9 (C<sub>1''</sub>), 135.7 (C<sub>3''</sub>, C<sub>5''</sub>), 127.6 (C<sub>2''</sub>, C<sub>6''</sub>), 126.7 (C<sub>4''</sub>), 74.4 (C<sub>3</sub>, C<sub>4</sub>), 42.8 (C<sub>1'</sub>), 41.0 (C<sub>1</sub>), 35.3 (C<sub>2</sub>).

**(1r,3R,4S)-N-(2-chloro-3-fluorobenzyl)-3,4-dihydroxycyclopentane-1-carboxamide (75/Optactamide)**



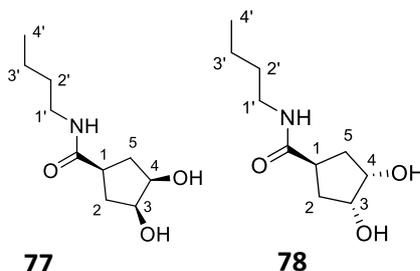
Ester **77** was subjected to a hydrolysis using NaOH. After a work up the crude product was amide coupled with (2-chloro-3-fluorophenyl)methanamine (1 eq) using carbonyl-diimidazole (1.2eq). After an acid base work up, the product **75** was extracted with DCM (100 mL), dried over magnesium sulphate and concentrated in *vacuo* to afford a white powder (22% yield). m.p. 118-120°C. IR (ATR)  $\nu_{\text{max/cm-1}}$ : 3304 (OH), 2937-2906 (CH), 1583 (C=O), 1568 (C=C), 1103 (C-F), 794 (C-Cl). <sup>1</sup>H NMR (500 MHz, Methanol-d<sub>4</sub>) δ 7.32 (t, J = 9.2, 6.9 Hz, 1H, H<sub>5''</sub>), 7.25-7.24 (m, 2H, H<sub>4''</sub>, H<sub>6''</sub>), 4.33 (s, 2H, H<sub>1'</sub>), 3.98-3.92 (m, 2H, H<sub>3</sub>, H<sub>4</sub>), 2.78 (tt, J = 9.2, 6.9 Hz, 1H, H<sub>1</sub>), 2.20-2.10 (m, 2H, 1 x H<sub>2</sub>, 1 x H<sub>5</sub>), 1.89-1.80 (m, 2H, 1 x H<sub>2</sub>, 1 x H<sub>5</sub>). <sup>13</sup>C NMR (75 MHz, Methanol-d<sub>4</sub>) δC 177.9 (C=O), 159.2 (C<sub>3''</sub>), 138.3 (C<sub>2''</sub>), 127.7 (C<sub>6''</sub>), 127.5 (C<sub>5''</sub>) 124.1 (d, J = 7.8 Hz, C<sub>4''</sub>), 114.8 (C<sub>4''</sub>), 73.4 (C<sub>3</sub>, C<sub>4</sub>), 40.5 (C<sub>1'</sub>), 40.0 (C<sub>1</sub>), 34.2 (C<sub>2</sub>). LR-MS (ES<sup>+</sup>): m/z 289.1 [M+H]; HRMS (EI<sup>+</sup>): calculated for C<sub>13</sub>H<sub>15</sub>ClFNO<sub>3</sub> 288.0800 found 288.0801

### N-butylcyclopent-3-ene-1-carboxamide (**76**)



To a solution of 3-cyclopentene carboxylic acid **37** was reacted with butan-1-amine (1eq) using carbonyl diimidazole (1.2 eq). The reaction was quenched with an aqueous solution of HCl (60 mL, 1 M). The product was extracted using DCM (50 mL). The product was then washed with an aqueous solution of HCl (30 mL, 1M), H<sub>2</sub>O (30 mL), a 50% saturated aqueous Na<sub>2</sub>CO<sub>3</sub>/NaCl solution (30 mL), dried and filtered through a cotton plug. The filtrate was diluted with 10% EtOAc/CH<sub>2</sub>Cl<sub>2</sub> and filtered through silica plug eluting with 10% EtOAc/CH<sub>2</sub>Cl<sub>2</sub> to remove the baseline color. Subsequent concentration in *vacuo* affords **76** as a yellow oil (70% yield). IR (ATR)  $\nu_{\text{max}}/\text{cm}^{-1}$ : 2468-2100 (CH), 1745 (C=O), 1630 (C=C). <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>)  $\delta$  5.63-5.6 (m, 2H, H<sub>3</sub>, H<sub>4</sub>), 4.05 (m, 5.614 2H, H<sub>1'</sub>), 3.10-3.03 (m, 1H, H<sub>1</sub>), 2.64-2.58 (m, 4H, H<sub>2</sub>, H<sub>5</sub>), 1.61-1.54(m, 2H, H<sub>2'</sub>), 1.40-1.30 (m, 2H, H<sub>3'</sub>), 0.89 (t, J=7.4 Hz, 3H, H<sub>4'</sub>). <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>)  $\delta$  177.8 (C=O), 128.6 (C<sub>3</sub>, C<sub>4</sub>), 67.1 (C<sub>1'</sub>), 39.2 (C<sub>1</sub>), 34.8 (C<sub>2</sub>,C<sub>5</sub>), 18.5 (C<sub>3'</sub>) 10.2 (C<sub>4'</sub>).

### (1*r*,3*R*,4*S*)-N-butyl-3,4-dihydroxycyclopentane-1-carboxamide (**77**) and (1*s*,3*R*,4*S*)-N-butyl-3,4-dihydroxycyclopentane-1-carboxamide (**78**)

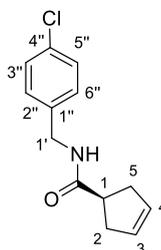


To a solution of **76** and *N*-methylmorpholine-*N*-oxide (1.2 eq.) dissolved in THF/H<sub>2</sub>O (9:1) was added a catalytic amount of osmium tetroxide (1 mol %) and stirred at room temperature overnight. The reaction was then quenched by the addition of sodium sulphite (30 mL). The product was extracted with ethyl acetate (100 mL), dried over magnesium sulphate and concentrated in *vacuo*. The crude product was then purified using flash column chromatography (10% MeOH/DCM) to give products: **77** as yellow oil (56% yield) and **78** as yellow oil (18% yield).

**77:** IR (ATR)  $\nu_{\text{max}}/\text{cm}^{-1}$ : 3329-3204 (OH), 2461-2069 (CH), 1712 (C=O).  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  4.30-4.20 (m, 2H, H3, H4), 4.15-4.06 (m, 5.614 2H, H1'), 3.20-3.13 (m, 1H, H1), 2.64-2.58 (m, 4H, H2, H5), 1.61-1.54(m, 2H, H2'), 1.40-1.30 (m, 2H, H3'), 0.89 (t,  $J=7.4$  Hz, 3H, H4').  $^{13}\text{C}$  NMR (75 MHz,  $\text{CDCl}_3$ )  $\delta$  176.6 (C=O), 73.46 (C3, C4), 64.6 (C1'), 39.25 (C1), 34.57 (C2,C5), 19.1 (C3') 13.7 (C4'). HRMS (EI+): calculated for  $\text{C}_{10}\text{H}_{18}\text{O}_4$  203.1277 found 203.1278

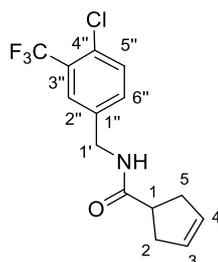
**78:** IR (ATR)  $\nu_{\text{max}}/\text{cm}^{-1}$ : 3315-3196 (OH), 2459-2034 (CH), 1705 (C=O).  $\delta$  4.30-4.20 (m, 2H, H3, H4), 4.15-4.06 (m, 5.614 2H, H1'), 2.26-2.19 (m, 1H, H1), 2.16-2.10 (m, 2H, 1 x H2, 1 x H5), 2.06-1.99 (m, 2H, 1 x H2, 1 x H5), 1.65-1.60(m, 2H, H2'), 1.42-1.36 (m, 2H, H3'), 0.98-0.93 (m, 3H, H4').

### N-(4-chlorobenzyl)cyclopent-3-ene-1-carboxamide **85**



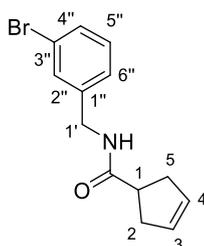
To a solution of 3-cyclopentene carboxylic acid **37** was reacted with (4-chlorophenyl)methanamine (1eq) using carbonyl diimidazole (1.2 eq). The reaction was quenched with an aqueous solution of HCl (60 mL, 1 M). The product was extracted using DCM (50 mL). The product was then washed with an aqueous solution of HCl (30 mL, 1M),  $\text{H}_2\text{O}$  (30 mL), a 50% saturated aqueous  $\text{Na}_2\text{CO}_3/\text{NaCl}$  solution (30 mL), dried and filtered through a cotton plug. The filtrate was diluted with 10% EtOAc/ $\text{CH}_2\text{Cl}_2$  and filtered through silica plug eluting with 10% EtOAc/ $\text{CH}_2\text{Cl}_2$  to remove the baseline color. Subsequent concentration in *vacuo* affords a white solid **85** (64% yield); m.p. 106-107 °C; IR (ATR)  $\nu_{\text{max}}/\text{cm}^{-1}$ : 3271 (NH), 3055-2843 (CH), 1535 (C=C), 802 (C-Cl), 682 (C-H);  $^1\text{H}$  NMR (500 MHz, Methanol- $d_4$ )  $\delta$  7.33 – 7.31 (m, 2H, H3', H4'), 7.28 – 7.26 (m, 2H, H2', H5'), 5.72 – 5.62 (m, 2H, H3, H4), 4.36 (s, 2H, H1''), 3.14 – 3.03 (m, 1H, H1), 2.68 – 2.53 (m, 4H, H2, H5).  $^{13}\text{C}$  NMR (126 MHz, Methanol- $d_4$ )  $\delta$  137.7 (C=O), 132.5 (C1'), 128.7 (C2', C5'), 128.6 (C3, C4), 128.18 (C3', C4'), 42.82 (C1), 42.05 (C1''), 36.58 (C2, C5). LR-MS (ES+):  $m/z$  235.95 (M+H).

### N-(4-chloro-3-(trifluoromethyl)benzyl)cyclopent-3-ene-1-carboxamide **86**



To a solution of 3-cyclopentene carboxylic acid **37** was reacted with (4-chloro-3-(trifluoromethyl)phenyl)methanamine (1 eq) using carbonyl diimidazole (1.2 eq). The reaction was quenched with an aqueous solution of HCl (60 mL, 1 M). The product was extracted using DCM (50 mL). The product was then washed with an aqueous solution of HCl (30 mL, 1M), H<sub>2</sub>O (30 mL), a 50% saturated aqueous Na<sub>2</sub>CO<sub>3</sub>/NaCl solution (30 mL), dried and filtered through a cotton plug. The filtrate was diluted with 10% EtOAc/CH<sub>2</sub>Cl<sub>2</sub> and filtered through silica plug eluting with 10% EtOAc/CH<sub>2</sub>Cl<sub>2</sub> to remove the baseline color. Subsequent concentration in *vacuo* affords a yellow solid (80% yield); m.p. 63-65 °C; IR (ATR)  $\nu_{\text{max}}/\text{cm}^{-1}$ : 3255 (NH), 3074-2846 (CH), 1643 (C=O), 1543 (C=C), 1111 (C-F), 659; <sup>1</sup>H NMR (300 MHz, Methanol-d<sub>4</sub>)  $\delta$  7.70 – 7.65 (m, 1H, H3'), 7.54 (m, 1H, H4'), 7.53 – 7.47 (m, 1H, H2'), 5.76 – 5.56 (m, 2H, H3, H4), 4.61 (m, 1H, NH), 4.41 (s, 2H, H1''), 3.18 – 3.00 (m, 1H, H1), 2.63 – 2.56 (m, 4H, H2, H5). <sup>13</sup>C NMR (75 MHz, Methanol-d<sub>4</sub>)  $\delta$  178.94 (C=O), 140.3 (C1'), 133.5 (C3'), 132.7 (C4'), 129.9 (C3, C4), 127.6 (C3'), 126.2 (C-CF<sub>3</sub>), 64.11 (C-Cl), 44.18 (C1), 43.8 (CF<sub>3</sub>), 43.07 (C1''), 37.90 (C2, C5). LR-MS (ES<sup>+</sup>): m/z 303.93 (M+H).

### N-(3-bromobenzyl)cyclopent-3-ene-1-carboxamide **87**



To a solution of 3-cyclopentene carboxylic acid **37** was reacted with (3-bromophenyl)methanamine (1eq) using carbonyl diimidazole (1.2 eq). The reaction was quenched with an aqueous solution of HCl (60 mL, 1 M). The product was extracted using DCM (50 mL). The product was then washed with an aqueous solution of HCl (30 mL, 1M), H<sub>2</sub>O (30 mL), a 50% saturated aqueous Na<sub>2</sub>CO<sub>3</sub>/NaCl solution (30 mL), dried and filtered through a cotton plug. The filtrate was diluted with 10% EtOAc/CH<sub>2</sub>Cl<sub>2</sub> and filtered through silica plug eluting with 10%

EtOAc/CH<sub>2</sub>Cl<sub>2</sub> to remove the baseline color. Subsequent concentration in *vacuo* affords to produce a dark yellow solid (91% yield); m.p. 85.0-86.0 °C; IR (ATR)  $\nu_{\text{max}}/\text{cm}^{-1}$ : 3267 (NH), 2904 (CH), 1631 (C=O), 1527 (C=C), 663 (C-Br); <sup>1</sup>H NMR (500 MHz, Methanol-d<sub>4</sub>)  $\delta$  7.42 -7.38 (s, 1H, H5'), 7.40 (m, 1H, H4'), 7.28 -7.21 (m, 2H, H2', H3'), 5.79 - 5.57 (m, 2H, H3, H4), 4.36 (s, 2H, H1''), 3.15 - 3.04 (m, 1H, H1), 2.69 - 2.54 (m, 4H, H2, H5). <sup>13</sup>C NMR (126 MHz, Methanol-d<sub>4</sub>)  $\delta$  177.4 (C=O), 141.6 (C1'), 130.1 (C5'), 129.9 (C4'), 129.8 (C3'), 128.6 (C3, C4), 125.90 (C2'), 63.4 (C-Br), 42.8 (C1), 42.1 (C1''), 36.6 (C2, C5). LR-MS (ES+): m/z 279.88 (M+H).

## 4.2.2 Biology

### Computational modelling:

CORINA was used to create 3D coordinates of low energy states of Optactin analogues. These were modelled into the allosteric site of NanB (protein-ligand complex crystal structure of Optactin (Brear, 2012)) using GOLD and ranked based on a fitness algorithm called CHEMSCORE. CORINA was also used to create a small library of structural analogues of amide **40**. GOLD was again the docking programme used to dock this ligands against the allosteric pocket of NanB. The Optatin-NanB<sub>K499G</sub> crystal structure was used as the basis for receptor file generation. Optactin bound within the allosteric site of NanB was used as a basis to generate a receptor file. A circumference cut-off of between a 10 Å around the bound ligand was used to create the receptor for docking. GOLD docking was performed by using the software's GUI. The scores were compiled and compared.

### Protein expression:

The NanB plasmid 5  $\mu$ l (pet23b vector) supplied by the Taylor group (University of St. Andrews) was transformed into 500  $\mu$ L of *E.Coli* (BL21 GOLD) by incubation on ice for 45 mins, then heat shocked for 30 seconds and then put back on ice for 2 minutes. The cells were then added to 1 mL of L-B Broth and incubated at 37 °C overnight. After incubation for 1 hr 200  $\mu$ l of cells were spread onto L-B agar plates containing 100 mg/mL L-ampicillin. These cells were incubated overnight at 37 °C. A single colony was then inoculated into 10 mL of L-B Broth containing 0.1 mg/mL. From this overnight inoculation 1 mL was transferred into 1000 mL x 4 of L-B Broth containing 0.1 mg/mL L-ampicillin. The cultures were shaken at 37 °C until reaching optical density of 0.6. At this point IPTG was added at a final concentration of 1 mM to initiate protein expression. The cultures were incubated overnight at 37 °C before harvesting the cells by centrifugation at 8000 rpm for 25 minutes.

For the autoinduction method; 1 mL of overnight inoculation was transferred into 1000 mL of autoinduction media (Terrific broth base with trace elements, ForMedium) containing 0.1 mg/mL L-

ampicillin and incubated for 2 hrs at 37 °C. The temperature was then turned down to 22 °C and incubated for 48 hrs stopping only to add a further 50 µg/mL L-ampicillin. Cells were then harvested by centrifugation at 8000 rpm for 25 minutes.

#### **Protein purification:**

The cell pellet was resuspended in 200 mL of PBS with 10mM imidazole, 2 protease cocktail tablets (Roche Diagnostics) (or without protease cocktail inhibitor in later preps), DNase I (Sigma, 20 µg/ml final concentration) and 5 mM NaCl. The cells were lysed on ice by sonication (5 x 30 second bursts). Cell debris was then removed by centrifugation at 20,000 rpm for 25 minutes at 4 °C. The supernatant was then filtered using a syringe 0.2 µm filter. The filtered supernatant was loaded onto a nickel HisTrap HP column (GE Healthcare) and protein eluted with a gradient of 10 mM-500 mM imidazole. The fractions containing NanB were collected and dialysed overnight at 4 °C into 10 mM of Tris HCl, pH 7.5. The protein was then applied to a 5 ml Q-FF DEAE anion exchange column and eluted with a NaCl gradient from 0-500 mM. The NanB containing fractions were concentrated and applied to a 120 ml Sephacryl S-100 gel filtration column equilibrated with 10 mM Tris-HCl, 50 mM NaCl, pH 7.5. Protein was concentrated to 1 mg/mL for biological assays. For crystallisation experiments protein was concentrated to 6.73 mg/mL.

#### **Protein crystallisation and X-ray crystallography:**

Crystal trials were set up using sitting drop vapour diffusion 96 well plates (Douglas instruments). 2 µL of 6.73 mg/mL protein and 2 µL of buffer was added into the sitting drop well. In the reservoir 50 µL of buffer was added. The buffer consisted of PEG 8K and imidazole pH 8.0. Different concentrations of PEG 8K were used (4, 5, 6, 7, 8, 10 % PEG 8K). The crystals generated were used for seeding. For seeding a single crystal was aspirated into 50 µL of buffer in a seed bead tube and vortexed for 60 seconds. Crystallisation conditions were set up as before with addition of crystal seeds by use of a cat whisker dipped into the seed solution and then into the sitting drop. The plates were then sealed and stored at room temperature.

Crystal soaking was performed with the generated NanB<sub>D643G</sub> and NanB<sub>WT</sub> crystals. Compounds were made up in stock concentrations of 1M. This was performed with DMSO. The compounds were kept at a stock concentration of 1M dissolved in DMSO. The buffer contained compounds at a final concentration of 5 mM and at a final DMSO concentration of 5%. Crystals were soaked for periods of time ranging from 10 mins to 3 days. It was noticed that NanB<sub>D643G</sub> crystals survived for longer under these conditions. The crystals were cryo-protected with DMSO (30%) or MPD (20%) and sodium acetate buffer pH 5.0 for 30 seconds prior to placement in liquid nitrogen.

Data collection	NanB <sub>WT</sub> -CHES	NanB <sub>WT</sub> <sub>T</sub>	NanB <sub>WT</sub> -Optactin	NanB-amide 40	NanB <sub>D643G</sub> -Optactamide (PDB: 4XYX)
Beamline	In-house	In-house	In-house	In-house	In-house
Wavelength (Å)	1.54178	1.54178	1.54178	1.54178	1.54178
Unit-cell parameters					
a (Å)	75.873	76.672	75.70	75.7	75.6
b (Å)	82.047	82.834	82.83	82.83	82.65
c (Å)	113.85	117.627	116.47	116.36	116.32
α (°)	90.00	90.00	90.00	90.00	90.00
β (°)	90.00	97.00	97.00	90.00	90.00
γ (°)	90.00	90.00	90.00	90.00	90.00
Molecules per symmetric unit	1	1	1	1	1
Resolution (Å)	50-3.32	50-1.93	44.55-1.84	50-2.1	50-2.1
Unique reflections	11450	1381		1458	1953
Redundancy <sup>a</sup>	3.6 (3.3)	7.5(5.1)	6.5(5.0)	6.2(2.3)	5.5 (4.2)
Completeness (%) <sup>a</sup>	70.1(76.1)	99.8(97.4)	99.6(95.5)	99.6(92.7)	98.6(92.6)
Rmerge (%) <sup>a,b</sup>	11.2(23.4)	4.8(12.8)	6.5(5.0)	5.6(19)	7.1(22.1)
$I/\sigma I^a$	13.5(8.18)	55.6(16.6)	34.9(11.1)	36.5(10.2)	38.0 (9.2)

Refinement	NanB <sub>WT</sub>	NanB <sub>WT</sub> -Optactin	NanB-amide <b>40</b>	NanB <sub>D643G</sub> -Optactamide
Resolution range (Å)	50- 3.32	44.5 2-1.84	44.55-1.84	50-2.1 50-2.1
Mean B-factor (Å <sup>2</sup> )				
Protein	-	20.3	18.1	22.5 24.9
Ligands	-	N/A	21.3	24.2 -
Water	-	27.6	25.5	19.5 -
<i>R</i> -factor (%) <sup>c</sup>	-	16.9	15.1	17.1 17.6
<i>R</i> <sub>free</sub> (%) <sup>d</sup>	-	21.2	19.9	21.4 24.7
r.m.s.d. bond lengths (Å) <sup>e</sup>	-	0.02	0.02	0.02 0.02
r.m.s.d. bond angles (°) <sup>e</sup>	-	1.92	2.35	2.14 2.37
Number of reflections	-	5392 0	55944	56450 40395
No. of atoms				
Protein atoms	-	5254	5218	5193 5602
Ligands	-	-	Optactin 27	Amide <b>40</b> 34 -
Water molecules	-	156	148	158 -

Table 4.3. Data collection, refinement and validation statistics of NanB<sub>D643G</sub> and NanB<sub>WT</sub> in complex with Optactin, Amide 40 and Optactamide.

- a. Values in parentheses correspond to the highest resolution shell.
- b.  $R_{\text{merge}} = \frac{\sum hkl \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum hkl \sum_i I_i(hkl)}$ , where  $I_i(hkl)$  is the intensity for all observations  $i$  of reflection  $hkl$ , and  $\langle I(hkl) \rangle$  is the weighted average intensity for all observations  $i$  of reflection  $hkl$ .
- c.  $R_{\text{factor}}$  and  $R_{\text{free}} = \frac{(\sum |F_o| - |F_c|)}{(\sum |F_o|)}$ .
- d.  $R_{\text{free}}$  was calculated for a 5% set of reflections excluded from the refinement.
- e. r.m.s.d. is the root-mean-square deviation from ideal geometry.

Protein crystallography data was recorded in house on the Rigaku-MSX Micromax-007 X-ray generator and R-Axis detector. This was recorded at 100 K and data was scaled and integrated using MOSFLM. Data was integrated by MOSFLM and run on pointless and scaled by scalar using CCP4. Alternatively the data was processed with HKL2000. SCALEPACKtoMTZ was used within the CCP4 suite to generate an MTZ file. Phaser was used with molecular replacement to solve the initial phases. The data was then subject to Refmac for refinement. Coot was then used to fit the ligand to the observed density.

#### **Kinetic assay:**

The MUNANA assay was used to measure the activity of inhibitors on NanB. Stock solutions were made of each compound in DMSO (0.1 M). Each compound was initially assayed at 500  $\mu$ M. The MUNANA was assayed in 96 well plates with each well containing a 50  $\mu$ l total volume containing sodium acetate (50 mM) buffer pH 5.0 a final concentration of 60 ng/ml NanB, 120  $\mu$ M MUNANA and compound (2 mM, 1mM, 500  $\mu$ M, 250  $\mu$ M, 100  $\mu$ M, 20  $\mu$ M). The assay was performed using the Stratagene MX3005P PCR system and fluorescence measured at 320 nm excitation and 430 nm emission. Recordings were taken at 37 °C every 30 seconds for 20 minutes. The initial rate was calculated and compared to a control giving the percentage inhibition.

#### **Fluorescence quenching:**

The fluorescence quenching assay was used to determine if compounds quench the fluorescence of **4-Mu**. The assay was performed at 37 °C using 96 well plates with a total volume of 50  $\mu$ l on the Stratagene MX3005P PCR system. The fluorescence was measured at 365 nm excitation and 420 nm emission and recordings were taken every 30 seconds for 10 minutes. Each well had **4-Mu** at a final concentration of 20  $\mu$ M, the compound at 500  $\mu$ M or 1 mM, 2% DMSO in 50 mM sodium acetate buffer pH 5.0. The amount of fluorescence was compared at each condition against the control.

### **Dynamic light scattering (DLS) assay:**

DLS assay was performed in a Quartz Suprasil 1.25 mm precision cell (Hellma Analytics). A 125 mM stock of Optactamide was made up in DMSO. A 50mM stock of Optactin was made up in DMSO. The reaction cell contained a total volume of 200  $\mu$ L, with a final concentration of 2mg/mL NanB and 2% DMSO (final concentration of 2.5mM Optactamide or 0 mM Optactamide for Control)(final concentration of 1mM Optactin or 0 mM Optactin for Control) in 50 mM sodium acetate buffer pH 5.0. DLS of each sample was measured using a Zetasizer  $\mu$ V (Malvern) spectrometer and analysed using Zetasizer software (Malvern).

### **Cell culture:**

The A549 cell line was obtained from Professor Randall's group (the original location these cells were obtained from was the ECACC). This cell line was maintained in a T75 flask in 10% FBS, DMEM high glucose and 1% penicillin/streptomycin media. Cells were passaged when reaching 80-95% confluence and maintained for 20 passages.

*S.pneumonia* (50 $\mu$ L) was cultured in 5mL BHI from D39 *S.pneumonia* frozen stock vials stored by the Professor Taylor group. This inoculation was left to incubate overnight at 37°C in anaerobic conditions. The OD was measured the next day (OD: 1.29) and the bacteria was centrifuged and re-suspended in 10% FBS, DMEM high glucose and 1% penicillin/streptomycin media ready for use in the adhesion and invasions assay. Contamination was checked by the streaking of cells on blood (horse blood) agar plates with sensitivity to optochin checked with optochin disks.

### **Cytotoxicity/MTT assay:**

The MTT assay was performed using a 96 well plate with 10,000 A549 cells/well seeded 24hrs before the start of the experiment in 10% FBS, DMEM high glucose and 1% penicillin media. The cells were then treated with or without Optactamide at various concentrations (20, 10, 5, 2, 1, 0.5, 0.1, 0.05, 0.01 mM) and left to incubate overnight. MTT was added (20  $\mu$ L) from a stock solution of 12 mM to each well and left to incubate for 4hrs. The media was then removed and formazan product was re-suspended in 200  $\mu$ L of DMSO. Absorbance at 530nm was read using a SpectraMax plate reader.

### ***S.pneumonia* adhesion and invasion assay:**

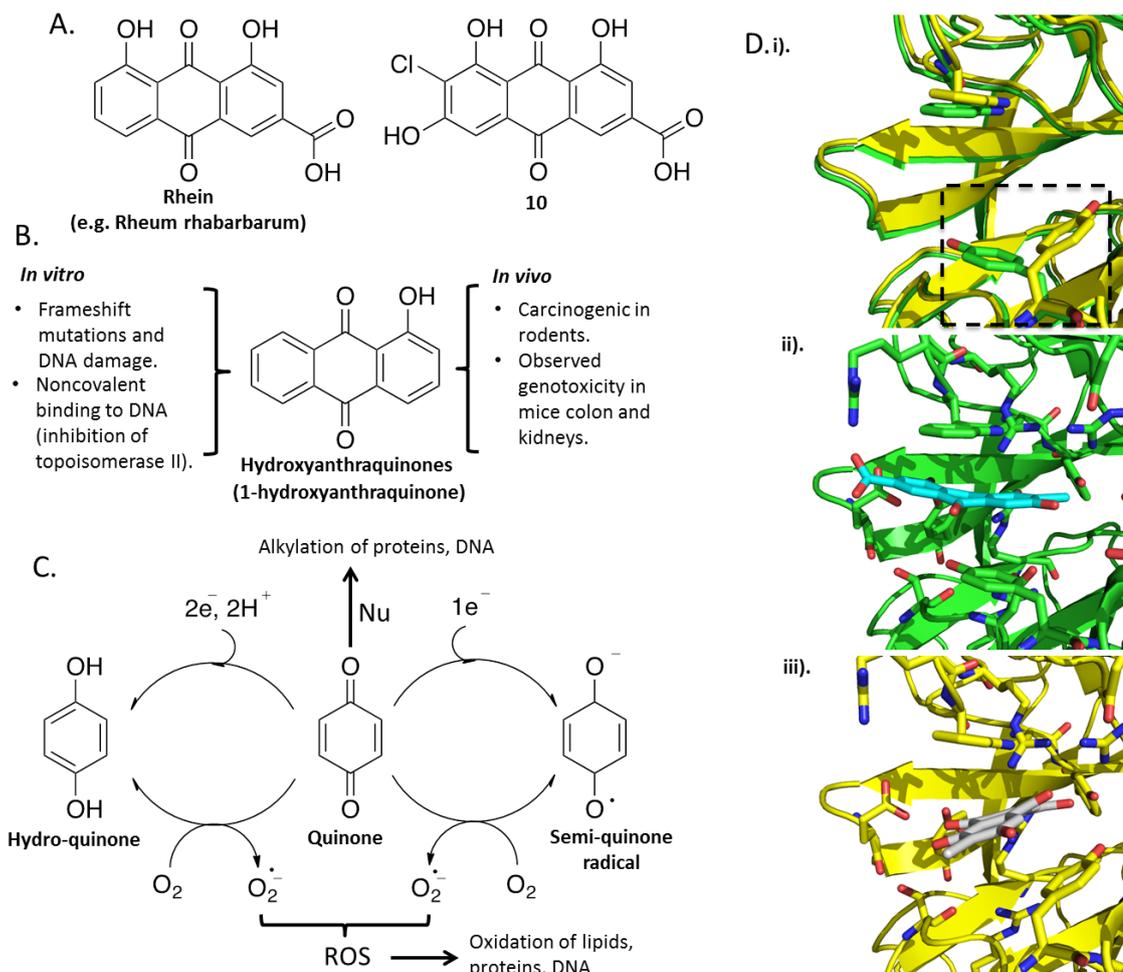
The *S.pneumonia* adhesion and invasion assay was performed using 24 well plates seeded with  $2.5 \times 10^5$  cells per well 24hrs prior to the start of the experiment. To the monolayers of A549 cells 100  $\mu$ L of a 1 in 10 dilution of the *S.pneumonia* cell suspension was added to each well. Optactamide (made up in PBS and DMSO (2% final DMSO concentration in each well) at various concentrations (1, 05, 0.1 and 0.01 mM) was immediately added to each well (for the control PBS + DMSO (2% final DMSO concentration in each well) was added), centrifuged (800 x g for 10 mins) and incubated at 37°C in anaerobic conditions for six hours. After this incubation period each well was washed carefully with 3 x 500  $\mu$ L PBS (for the invasion assay the addition of 500  $\mu$ L of DMEM high glucose containing gentamycin (100  $\mu$ g/mL) and penicillin (10  $\mu$ g/mL) was added and incubated for a 2hr prior to a further wash with 3 x 500  $\mu$ L PBS) and 0.2 mLs of trypsin/EDTA plus 0.025% Triton-X-100 was added to cause cell detachment and lysis. Dilutions (1 in 3, 1 in 10, 1 in 20 and 1 in 100 were trialed) of each cell lysis sample were spread onto blood agar plates and left to incubate at 37°C in anaerobic conditions overnight. The number of colonies was then counted and only the control dilutions that produced colonies of 30 to 300 were analysed and reported.

## ***5.0 The discovery of constrained tools for TcTS.***

In this chapter an attempt was made to discover an allosteric pocket within TcTS. The sequence identity between TcTS and NanB is low (19%), however the structural features of the catalytic domains are similar with both containing a 6-bladed  $\beta$  topology with a water channel core that runs through the centre of the catalytic domain. The aim of the work in this chapter was to identify a similar allosteric site within TcTS as in NanB. Using the position of the allosteric site in NanB, a potential allosteric site within TcTS was mapped and used for fragment screening. A CADD approach was used for “hit” identification. A 704,041 fragment library obtained from the ZINC database was filtering down to a smaller “focused” library of 1,015 fragments using ligand similarity searching. This smaller “focused” fragment library was screened against the mapped site using FlexX, a faster docking programme than GOLD. Three “hits” were chosen from the evaluation of binding scores and removal of potential active site binders. Only weak inhibition was observed and no ligand binding was observed through crystal soaking.

### ***5.1 The TcTS story so far.***

The structure of TcTS was solved by Buschiazzi et al., 2002 (see Chapter 1.57.1). Despite an available crystal structure and interest in TcTS as a potential drug target candidate for the NTD, Chagas disease (see Chapter 1.40, 1.41 and 1.57.1), no protein-inhibitor complex of TcTS has been published apart from DANA-TcTS, covalent intermediate (2,3-difluoro-N-acetylneuraminic acid) and benzoylated NANA (5-(acetylamino)-9-(benzoylamino)-3,5,9-trideoxy- 3-fluoro-D-erythro-alpha-L-manno-non-2-ulopyranosonic acid). All of these inhibitors are substrate mimetics and weak inhibitors of TcTS. In the pursuit of more potent TcTS inhibitors, SBDD has been used by a variety of groups to use structural observations within the active site as a basis for improved binding and potency (see Chapter 1.62). An example of a TcTS inhibitor discovered using a HTS approach is in the discovery of compound **10** (Figure 59).



**Figure 59** **A.** Chemical structures of two similar anthraquinones, a natural laxative Rhein and compound **10**. **B.** The chemical structure of a hydroxyanthraquinone and its *in vitro* and *in vivo* toxicity. **C.** The chemical structure of a quinone with potential routes of toxicity including direct alkylation and generation of reactive oxygen species (ROS). **D.** i) The structure of the TcTS active site within two separate crystal structures (unpublished TcTS<sub>F58N</sub> crystal structures, see Chapter 4.7). The conformation of Tyr119 differs between the two structures (closed (yellow) and open (green)) highlighted by a dashed square. ii). The open conformation of the TcTS active site with the proposed binding conformation of compound **10**. iii). Proposed binding conformation of compound **10** in the closed conformation of the TcTS active site. Figures created using CHEMdraw, GOLD docking programme and PYMOL.

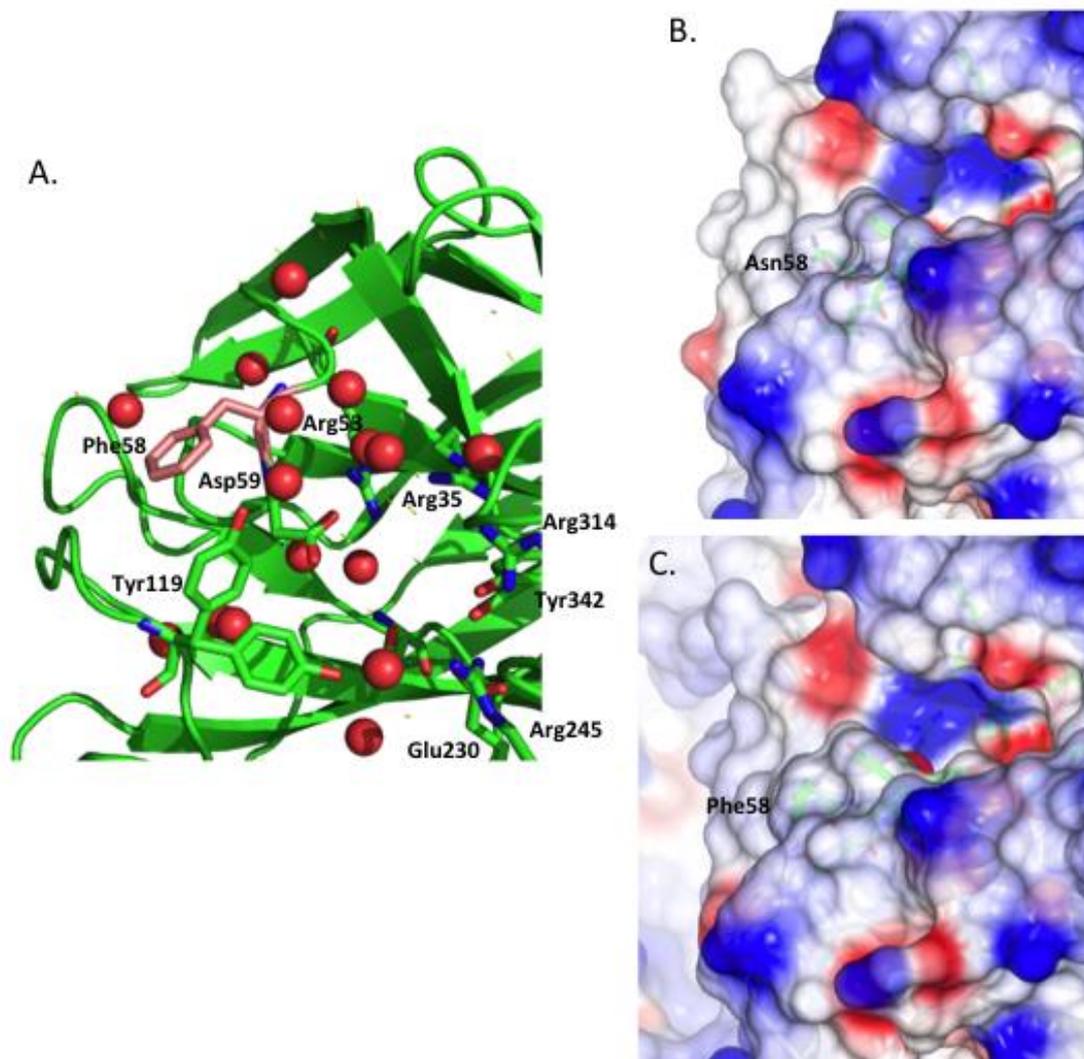
This TcTS inhibitor was discovered using a natural product library of 2283 moieties. This led to the initial discovery of 103 inhibitors of TcTS. Filtering out promiscuous binders, following Lipinski's constraints (Chapter 1.71) and limiting based on structural availability narrowed this subset to two "hits". SAR was performed on these two hits, which led to the most potent inhibitor to date (compound **10**) an anthraquinone derivative with a potency of 580 nM (Chapter 1.62). Tested against human Neu2, compound **10** was observed to have a selectivity of 170 times greater for TcTS than for the human sialidase. Tyr119 exists in two separate conformations observed within crystal structures solved within the group and in the paper published by Buschiazzo et al., 2002 (Figure 56).

Compound **10** was run against these two conformations of the active site (termed open and closed based on the positioning of Tyr119) using the docking programme GOLD. From the docking run, the anthraquinone core of compound **10** was observed to form aromatic stacking interactions with the side chains of residues, Tyr119 and Trp312. The position of the carboxylic acid from compound **10** was not observed to form van der Waals contacts with the arginine triad in any of the ten poses generated against the open conformation of TcTS. Within the closed conformation, the carboxylic acid from compound **10** was observed to form van der Waals contacts with the arginine triad in 7 of the 10 binding poses generated. Surprisingly, the CHEMPLP docking score generated by GOLD for the highest ranked binding pose for the open conformation (155.05) is greater than that of the highest ranked binding pose (with arginine triad van der Waals contacts as a filter) in the closed conformation (146.62). The carboxylic acid of compound **10** should form strong van der Waals contacts with the arginine triad. These data suggest that compound **10** either binds in an unusual manner or it may bind to or stabilise the closed conformation of the TcTS active site. Without a solved TcTS-ligand complex of compound **10** the exact binding pose is unclear.

Despite the successful identification of a novel potent TcTS inhibitor problems exist with this class of chemical as a constrained chemical tool. A number of drugs that have made it through to the clinic contain an anthraquinone moiety (Malik and Müller, 2016), however a concern with anthraquinones as drug molecules is the presence of a quinone, which is a well-known toxicophore (Figure 56). Quinones can act as electrophilic Michael acceptors (Malik and Müller, 2016). However, the quinone in anthraquinone is unable to act as a Michael acceptor due to the position of the aromatic rings (Malik and Müller, 2016). The toxicity of anthraquinone derivatives is attributed to the generation of free radicals and for their potential to non-covalently bind to DNA causing topoisomerase II inhibition (Genov et al., 2016). Anthraquinone, without any substitution, is determined as a B2 possible carcinogen to humans by the International Agency for Research on Cancer (IARC) (Malik and Müller, 2016). Rhein is a hydroxyanthraquinone with an almost identical substitution pattern to compound **10**. Rhein is the major toxic chemical in *Cassia occidentalis* seeds (Panigrahi et al., 2015). The consumption of these seeds has been linked to the deaths of children in northern parts of India (Panigrahi et al., 2014). Rhein is toxic to rat primary hepatocytes and has been demonstrated to interfere with the cellular cytoskeleton, mitochondrial function and depletion of reduced glutathione in a number of other cell lines (Yuan et al., 2016, Panigrahi et al., 2015). In addition to the toxicity risk, further problems exist with compound solubility of anthraquinones, thereby limiting the versatility of the chemical class as a chemical tool. Despite the problems that exist with this class of molecule, the discovery of compound **10** and other active site inhibitors by other research groups (see Chapter 1.62) have highlighted key compound functionalities required for

potency. It is clear that solely targeting the sialic acid binding pocket results in low potency. All the potent inhibitors of TcTS contain aromatic groups and have predicted affinity for the lactose binding pocket.

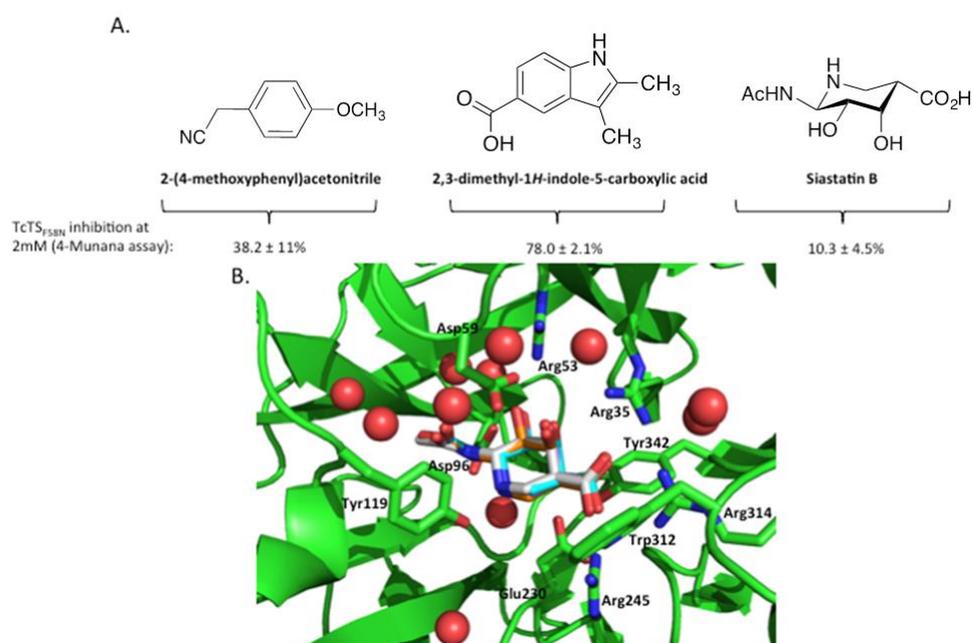
The crystal structure of TcTS solved by Buschiazzo et al., 2002, contains 7 mutations at the molecular surface. These mutations were generated using a “systematic surface-mutagenesis approach” and proposed to aid crystallisation (amino acid positions involved in crystal contacts from the TrSA (a protein with 70% sequence identity to TcTS) crystal structure were used as the basis for TcTS mutagenesis). The majority of these mutations fall within the lectin-like binding domain of TcTS at proposed crystal-crystal contact positions. However, a single mutation (Phe58) exists within the catalytic domain and within close proximity to the catalytic site residue Asp59.



**Figure 60.** A. TcTS active site with residue Phe58 highlighted in pink (1MS9) B. Surface potential (electrostatic) of the TcTS active site (TcTS<sub>F58N</sub>) with residue Asn58 highlighted (unpublished). C. Surface potential (electrostatic) of the TcTS active site with residue Phe58 highlighted (1MS3). The active site of 1MS3 was aligned against the active site of Dr Telford's TcTS<sub>F58N</sub> structure (unpublished) and calculated to have an RMSD of 0.12 Å over all atoms. Figures created using PyMOL and CCP4MG (surface potentials). RMSD calculated using the alignment function in PyMOL. Active sites for 1MS9 and TcTS<sub>F58N</sub> were generated using residue 58 as the epicenter to a 12 Å diameter cut-off.

Although determined by Buschiazzo et al., 2002 to have no impact on TcTS activity using the 4-Munana assay, it is possible a mutation this close to the active site might bias the structural integrity of the active site and mislead SBDD approaches. To generate a structure with an increased similarity to wild-type TcTS, Dr Telford within the Taylor group performed a point mutation on this construct, expressed and purified this TcTS<sub>F58N</sub> enzyme. This TcTS<sub>F58N</sub> enzyme was crystallised and used for structural determination. Despite very little structural difference between the TcTS<sub>F58N</sub> construct and

the version used by Buschiazzo et al., 2002 (Figure 57.), the TcTS<sub>F58N</sub> construct was observed to form crystals with excellent diffraction in house and so TcTS<sub>F58N</sub> was used for further experiments. Using this new TcTS<sub>F58N</sub> construct, Dr Telford began an initial search for small molecule inhibitors against TcTS<sub>F58N</sub>. This involved screening 672 fragments from the Maybridge RO3 library. All fragments were tested against TcTS<sub>F58N</sub> using two screening assays. The first assay used was the standard sialidase 4-Munana assay. This 4-Munana assay identified 66 possible inhibitors of TcTS<sub>F58N</sub> from the Maybridge library. The second assay, a <sup>1</sup>H water-LOGSY NMR technique was used to screen 472 fragments from the Maybridge library and identified only 24 fragments as inhibitors. Only one fragment (2-(4-methoxyphenyl)acetonitrile) was identified by both assays to inhibit TcTS<sub>F58N</sub>. Potency of this inhibitor was deemed to be weak by the 4-Munana assay ( $38.2 \pm 11\%$  at 2 mM). The most potent inhibitor identified from the Maybridge library was 2,3-dimethyl-1*H*-indole-5-carboxylic acid ( $78.0 \pm 2.1\%$  at 2mM) (Figure 61).



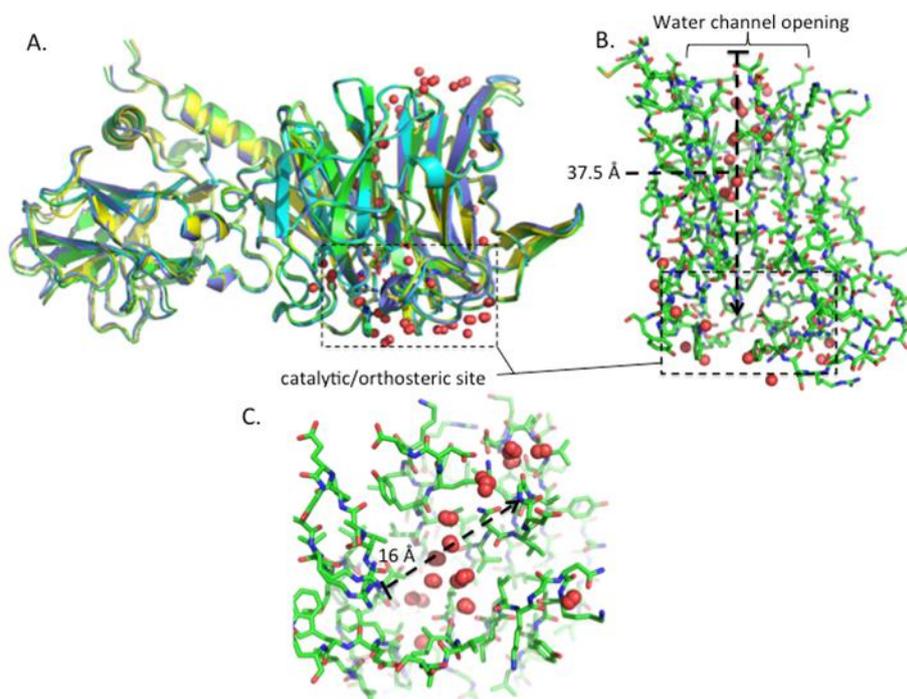
**Figure 61. A.** The chemical structure of the TcTS inhibitors identified by Dr Telford and the percentage inhibition observed for each inhibitor at 2mM using the 4-Munana assay. **B.** The active site of TcTS<sub>F58N</sub>-siastatinB complex (siastatinB in grey, Dr Telford, unpublished). The docking of siastatinB performed by Dr Telford using GOLD was replicated here. Three poses (teal, orange and purple) were generated (all within an RMSD of 1.5) with all poses observed to bind in a similar manner to the observed binding mechanism in the crystal structure.

Due to the disappointing results obtained from these assays, an alternative approach was sought that could use the structural information provided by the TcTS<sub>F58N</sub> crystal structure to generate a more potent inhibitor. A CADD approach was investigated to evaluate if ligand binding poses could

be accurately predicted. GOLD was tested and binding modes compared against the solved crystal-ligand complex (Siasatin B-TcTS<sub>F58N</sub>). GOLD accurately predicted the binding pose of this ligand and indicated that a CADD approach to TcTS inhibitor discovery and design could be successful. A CADD approach is used within this thesis to attempt to discover a novel TcTS inhibitor.

## ***5.2 The water channel within TcTS***

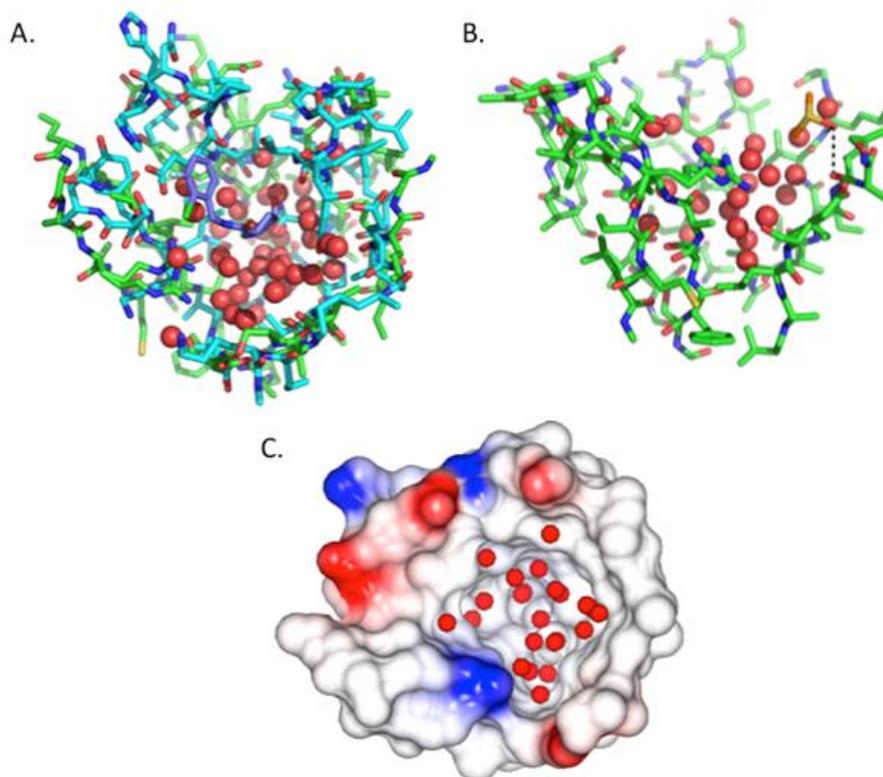
TcTS contains 19% sequence similarity to NanB (see Chapter 1.59.3). This enzyme is a trans-sialidase and works *via* a different mechanism to NanB (see Chapter 1.59.2). The allosteric site on NanB exists within a conserved water channel. The water channel and conserved water molecules are found across the glycoside hydrolase family and proposed to have important structural and functional roles for enzymatic activity (see Chapter 1.59.3). The allosteric site found within the water channel of NanB (see Chapter 2 and Chapter 3) could be present in other enzymes within the glycoside hydrolase family. Within TcTS, the conserved water channel runs as a central pore through the canonical propeller of the domain ending at the catalytic site (Figure 59.). The TcTS water channel is approximately 37.5 Å long with a width of 16 Å (measurements estimated within PyMOL using solved crystal structure of TcTS<sub>F58N</sub>) at the channel opening (located on the opposite side to the active site of TcTS). TcTS crystal structures were solved in three different space groups by Buschiazzo et al., 2002. Alignment and comparison of the three space groups (1MS3: monoclinic (P12<sub>1</sub>1), 1MS4: triclinic (P1) and 1MS9: orthorhombic (P2<sub>1</sub>2<sub>1</sub>2)) showed no significant difference between the solved structures (see Appendix Table 1.). Additionally the solved structure within this work (using Dr Telford's methods, see Chapters 4.6 and 4.7) is a monoclinic (P2<sub>1</sub>) space group with a RMSD (calculated with the alignment function in PyMOL) of 0.38, 0.50 and 0.42 Å compared to that of 1MS3, 1MS4 and 1MS9 respectively (Figure 61).



**Figure 62.** A. Aligned structures of TcTS<sub>F58N</sub> and solved structures by Buschiazzo et al., 2002 in monoclinic (1MS3, yellow), triclinic (1MS4, purple) and orthorhombic (1MS9, teal) space groups. The catalytic site of TcTS is highlighted with a dashed box. B. A stick representation of the water channel running through the catalytic domain of TcTS<sub>F58N</sub>. The length of the water channel from Tyr342 to the opening is estimated to be 37.5 Å. C. An image of the water channel opening (approximate width is 16 Å). A number of conserved waters exist at the channel opening and within the channel. Conserved water molecules from TcTS<sub>F58N</sub>, 1MS3, 1MS4 and 1MS9 are shown as red spheres. Conserved water molecules are water molecules that appeared in a similar position in all of the structures analysed. Waters that did not appear in a similar position in all of the structures were not shown in this image. Figures were created using PyMOL and Microsoft PowerPoint.

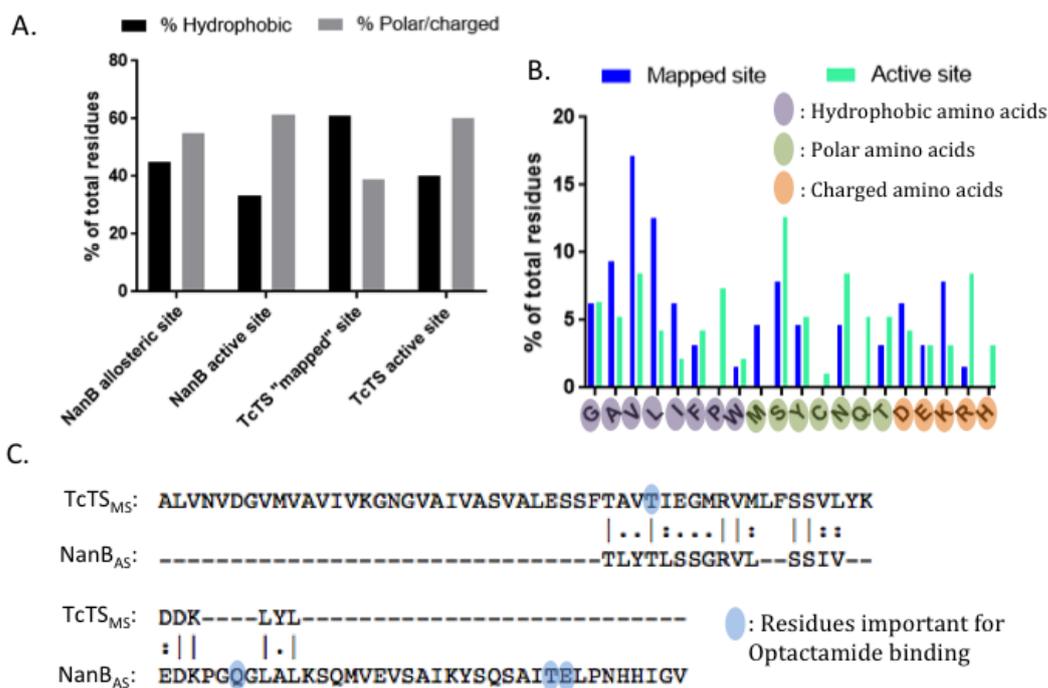
A number of conserved water molecules are found within the water channel of the TcTS<sub>F58N</sub> structure and across three published TcTS structures analysed (Figure 62). A large number of the conserved water molecules are located at the channel opening and continue to an approximate depth of 18 Å. Disruption of these conserved water molecules through small molecule binding might result in decreased activity of TcTS leading to the identification of an allosteric site. To generate a model of a potential allosteric site within TcTS, the allosteric site of NanB was used as a basis to map the “secondary” site coordinates. An RMSD of the secondary site compared to NanB was calculated to be 2.9 Å. Whereas the RMSD of the whole NanB catalytic domain compared with TcTS is calculated to have an RMSD of 3.7 Å. Due to the low sequence similarity between both proteins a high RMSD was expected. The coordinates from NanB did map out the opening of the water channel on TcTS, which provides a starting point for CADD (Figure 60.). A molecule of DMSO was found

within this mapped secondary site on TcTS. No other ligand was found to serendipitously bind within this mapped site in all 14 published TcTS crystal structures on the PDB.



**Figure 63 A.** Overlaid structures of TcTS (green) and NanB (4XYX, cyan) secondary site using Optactamide (purple) as the centre coordinate (12 Å distance filter). B. An image of the resulting mapped site of TcTS showing the opening of the water channel. A molecule of DMSO is observed to bind within this site (forming a hydrogen bond contact with the backbone carbonyl of residue Val184). C. Electrostatic surface representation of the mapped TcTS site. Waters are shown as red spheres. Figures created using PyMOL and Microsoft PowerPoint. Electrostatic surface representation was created using CCP4 QTMG.

The two sites between TcTS and NanB were compared and both contain a higher ratio of hydrophobic amino acids than polar/charged groups. The electrostatic surface of the mapped “secondary” site on TcTS is mainly hydrophobic (Figure 63 C.).



**Figure 64.** **A.** Bar chart representing the hydrophobic and polar/charged residues as a percentage of the total for NanB (allosteric site and active site) and TcTS (mapped site and active site). **B.** Bar chart representing the frequency of each amino acid within the mapped and active site of TcTS as a percentage of the total. **C.** Sequence alignment of the TcTS mapped site (TcTS<sub>MS</sub>) and NanB allosteric site (NanB<sub>AS</sub>). Residues within NanB<sub>AS</sub> that form van der Waals contacts with Optactamide are highlighted in blue. All TcTS and NanB sites created within PyMOL using a distance filter from a centre coordinate. The frequency of amino acids in each site was tallied within Microsoft® Excel® 14.6. Bar charts were created using GraphPad Prism 7. Sequence of TcTS<sub>MS</sub> and NanB<sub>AS</sub> were aligned using EMBOSS Needle. Images manipulated using Microsoft® PowerPoint® 14.6.

Allosteric sites have a different composition to that of active sites (see Chapter 1.31) (Li et al., 2013). The active site of NanB and TcTS contain a higher proportion of polar/charged groups than hydrophobic groups (Figure 61.). The frequency of glycine in the NanB active site is 7.5% and 6.1% in the allosteric site fitting with the description by Li et al., 2013. The ratio of hydrophobic groups within the allosteric site using a 12 Å distance filter from Optactamide (PDB:4XYX) is 0.8:1 (hydrophobic to polar/charged groups). The frequency of glycine is identical within the TcTS secondary site and active site at 6.3%. The number of polar/charged groups within this mapped secondary site is lower than that of the hydrophobic groups (1.56:1 of hydrophobic to polar/charged groups), consistent with the description by Li et al., 2013. In contrast, the active site of TcTS contains a high proportion of polar/charged sites and a lower frequency of hydrophobic sites with a ratio of 1.5:1 of polar/charged to hydrophobic groups. TcTS contains a higher frequency of hydrophobic

amino acids than the NanB allosteric site. The make-up of the mapped site in TcTS is very different to that of the allosteric site in NanB with a sequence match of 11.5% (calculated using EMBL-EBI pairwise sequence alignment tool EMBOSS Needle (Figure 61.)). Only one of the residues important for Optactamide binding (Figure 61.) within the allosteric site of NanB is found within the TcTS mapped site. Unsurprisingly, Optactamide did not inhibit TcTS at 1mM using the 4-Munana assay (see Appendix Figure 5). With a possible allosteric site mapped within TcTS the next step was to use this in CADD to identify if any small molecules would bind to this site.

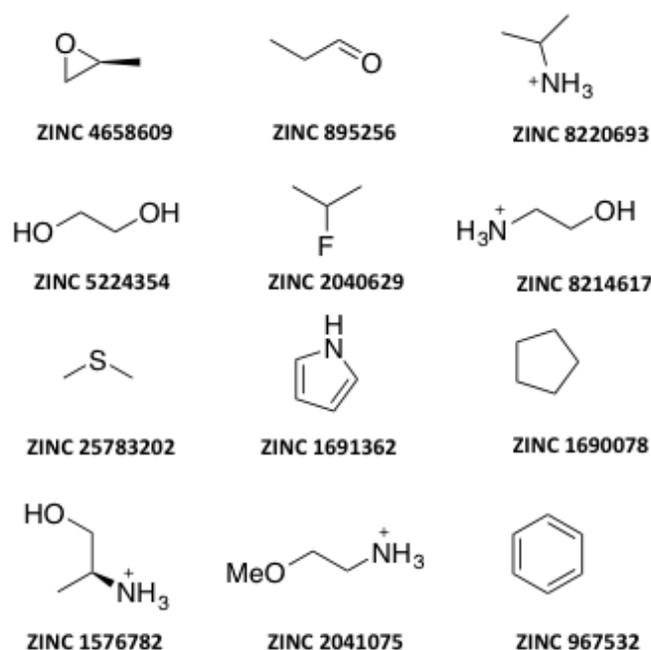
### ***5.3 Fragment library generation***

The generation of a good screening library requires a diverse set of synthetically viable chemicals. Fragment screening has been in use for more than 20 years (Siegal et al., 2007) and has contributed to the discovery and development of two approved drugs and over thirty clinical candidates (Erlanson et al., 2016). The first drug approval from a fragment based discovery approach was in 2011 for an inhibitor (vemurafenib) of a mutant BRAF kinase (Bollag et al., 2012, Erlanson et al., 2016). Fragments are typically small molecules of less than 300 Da (Ruda et al., 2010). Chemical space can be exploited much more efficiently with fragments than by traditional high-throughput screening (HTS), as there are fewer possible fragments than larger lead sized drug molecules (Davis and Erlanson, 2013). Consequently, fragments are usually screened in the thousands in the hope of identifying different pharmacophore moieties required for high affinity binding. Fragments make better starting points for drugs as they will contain fewer interfering moieties than HTS hits (Davis and Erlanson, 2013).

The availability of fragments through chemical suppliers was the main criteria for the generation of a “virtual” fragment-screening library against TcTS<sub>M5</sub>. ZINC is a free public resource of commercially available compounds from chemical supplier catalogues (Figure 65.) (Irwin et al., 2012) and was used to generate a virtual library. ZINC provides 3D structural information for each compound in a ready-to-dock format (sdf or smile formats) with each compound in different protonation states (pH 5 - 9.5) and tautomeric forms (Irwin et al., 2012). The ZINC<sup>12</sup> database contains 22,724,825 purchasable compounds that include make-on-demand and in-stock compounds.

The ZINC<sup>12</sup> library contains 3,459,596 fragments (make-on-demand and in-stock compounds). This set is filtered using the ZINC filtering capacity and reduced to 704,041 fragments that meet certain logP, molecular weight and rotatable bond criteria (logp equal to or less than 3.5, molecular weight equal to or less than 250 Da and rotatable bonds equal to or less than 5). This filtering

capacity was used to follow the constrained approach to chemical tool design. Fragments that adhere to these guidelines are more likely to lead to the development of chemical tools that conform to Lipinski's "Rule of Five" to achieve drug-like properties and valuable properties for use in chemical biology (see Chapter 1.23 and 1.71).

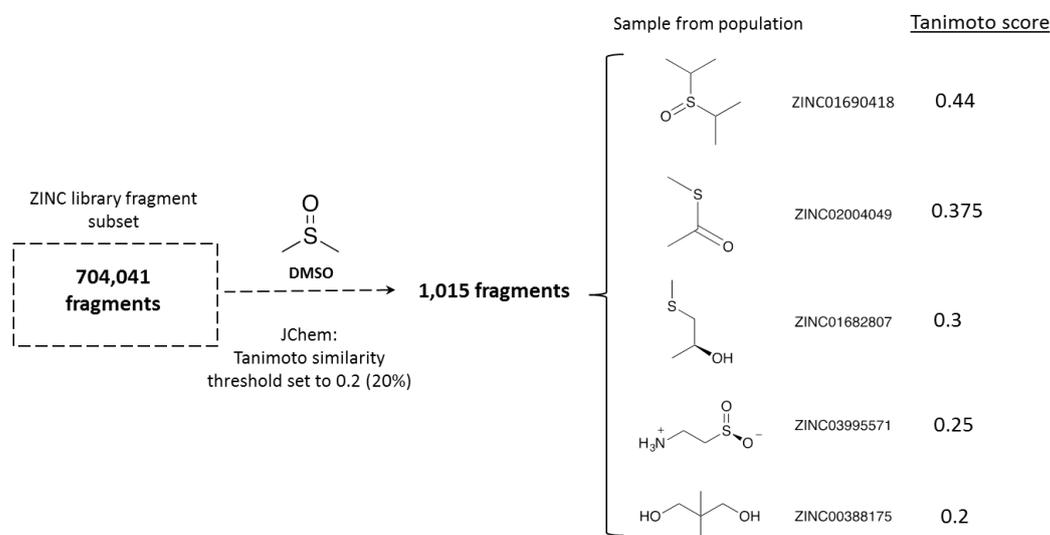


**Figure 65.** Examples of fragments from the ZINC<sup>12</sup> database. Each compound contains a unique ZINC ID (fragments and ZINC ID from in stock/Frags Now subset on the ZINC<sup>12</sup> database available at <http://zinc.docking.org/browse/subsets/>). Figure created using ChemDraw<sup>®</sup> 15.0 and Microsoft<sup>®</sup> PowerPoint<sup>®</sup>.

## 5.4 In Silico docking and "HIT" identification

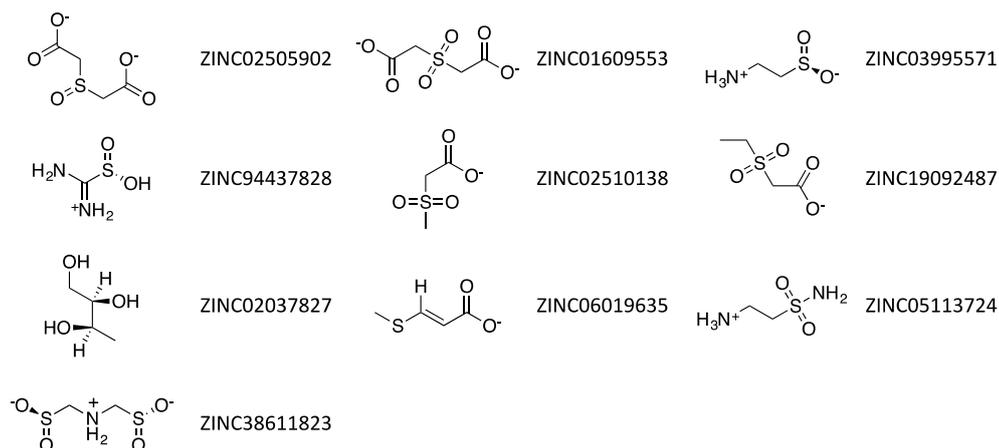
Generation of an enriched library can be performed by ligand similarity searching. Ligands observed to bind to the site was used as a similarity search criteria. This criterion should filter out fragments without the appropriate functionality for binding and should generate a library of ligands with a higher probability of binding. DMSO is a small ligand and used in high millimolar concentrations as a cryoprotectant in crystallography. As this was the only observed molecule to bind to this site and its likely weak affinity (forming a single hydrogen bond with Val184), it provides only limited functionality information. The Tanimoto similarity threshold was set to 0.2 (20%) to enable filtering of potential binding functionalities that might prove useful and decrease the library set to reduce the computational run time. Library filtering and reduction was performed using JChem and reduced the library set from 704,041 fragments to 1,015 fragments (a more appropriate load for docking) (Figure 66). The docking software FlexX from BioSolveIT LeadIT was used to run the screen as this software has a faster runtime than GOLD (see Chapter 1.73). These hits were screened

independently against the active site. The hits were then profiled based upon FlexX scores for both the mapped secondary site and the active site.



**Figure 66.** Schematic of the ligand similarity results generated from JChem. DMSO was used as the initial ligand similarity criteria. The threshold was set low at 0.2 and this generated 1,015 fragments (five examples are shown in the image with their corresponding Tanimoto score). Figure created using ChemDraw 14.0 and Microsoft PowerPoint.

Hits that scored high for binding against the mapped secondary site and low for the active site were ranked (Figure 67 and Table 10). The top three were bought in from commercial sources (hypotaurine, 1,2,3-butanetriol and (2-methylsulfonyl)acetic acid).



**Figure 67.** Top ten hits and ZINC ID's from the FlexX screen of 1,015 fragments.

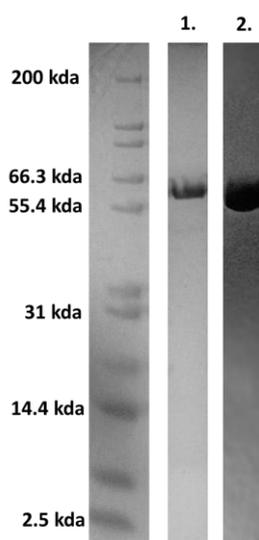
**Table 10.** Table of the scoring information generated from FlexX for the top ten highest scoring fragments. Score: Total score of the docking solution generated by FlexX. Match Score: the calculated contribution of the matched interacting groups. Lipo: the calculated contribution of the lipophilic area. Ambig score: the contribution of the lipophilic-hydrophilic contact area. Clash score: the contribution of the clash penalty to the score. Rot: the ligand conformational entropy score (Merzoug et al., 2013).

ZINC ID	Score	Match	Lipo	Ambig	Clash	Rot
ZINC02505902	-19.96	-24.85	-3.18	-5.69	2.76	5.6
ZINC01609553	-19.88	-24.38	-2.52	-5.15	1.17	5.6
ZINC03995571	-19.70	-24.50	-2.34	-3.46	1.00	4.2
ZINC94437828	-19.32	-20.49	-1.64	-4.90	0.91	1.4
ZINC02510138	-19.18	-23.95	-1.96	-4.09	2.62	2.8
ZINC19092487	-18.24	-23.23	-2.61	-4.35	2.35	4.2
ZINC06019635	-17.62	-19.52	-3.42	-3.07	1.59	1.4
ZINC02037827	-17.45	-27.31	-1.78	-3.58	2.82	7.0
ZINC05113724	-17.38	-24.57	-1.10	-4.08	1.36	5.6
ZINC38611823	-17.23	-22.22	-2.13	-5.45	1.58	5.6
DMSO	-5.32	-6.10	-3.71	-2.28	1.37	0.0

Fragments typically have low (mM) affinity for their targets and so sensitive and robust methods are required for the experimental analysis of the fragment “hits” generated from the CADD screen (Davis and Erlanson, 2013). The two methods run in parallel were the standard sialidase 4-Munana assay and X-ray crystallography (ligand soaking) to evaluate activity and binding. To do this TcTS<sub>F58N</sub> protein was expressed, purified and crystallised.

## 5.5 TcTS construct, expression and purification

The expression of TcTS followed Dr Telford's protocol. Unfortunately, a copy of the TcTS<sub>F58N</sub> plasmid could not be located, so mutagenesis was performed using the original template provided by Alejandro Buschiazzi. Primers were designed (Appendix Figure 84.) and site-directed mutagenesis was performed on the Buschiazzi construct. Colony PCR was performed and full sequencing by GATC confirmed the sequence of TcTS<sub>F58N</sub> (see Appendix for full sequence). This TcTS<sub>F58N</sub> construct was transformed into BL21 DE3 GOLD *E.coli* and expressed using the autoinduction method (Figure 68.).

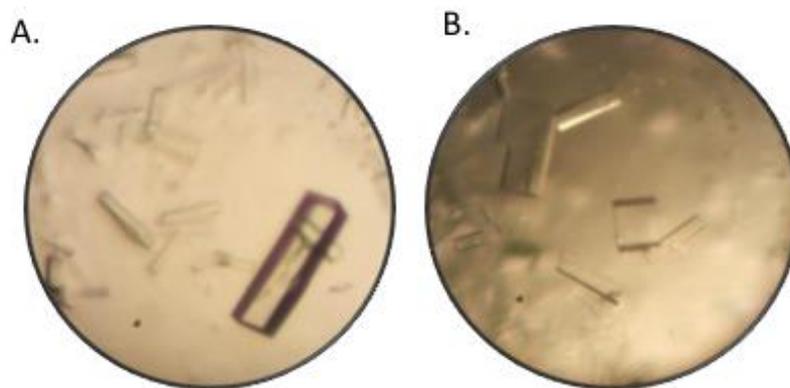


**Figure 68.** SDS Page of TcTS<sub>F58N</sub> after nickel-column chromatography (1) and size-exclusion chromatography (2).

The activity of TcTS<sub>F58N</sub> was assessed after each stage of protein purification. The activity of TcTS<sub>F58N</sub> was the highest after the size-exclusion chromatography purification step (345 AFU/sec compared to 100 AFU after IMAC).

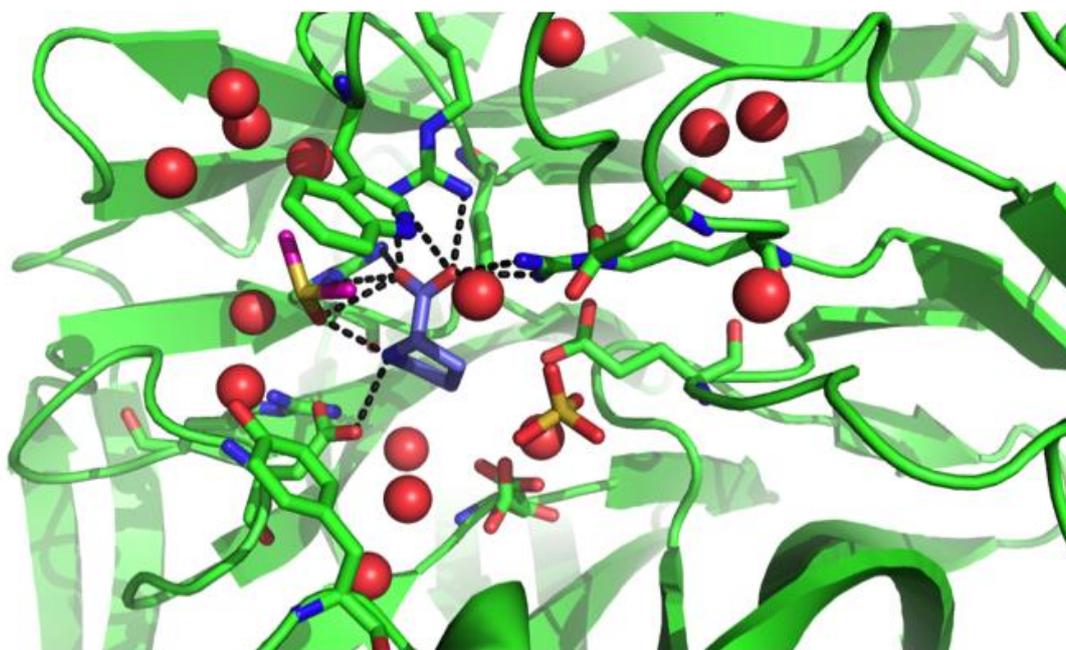
## 5.6 TcTS crystal generation and structures

TcTS<sub>F58N</sub> protein purified using the method detailed in Chapter 4.6 was added to crystal drop vapour diffusion plates for protein crystallisation. The crystallisation conditions discovered by Dr Telford (Telford, 2014) were used. The first condition used was 200mM L-Proline, 100mM Hepes and 10% PEG 3350. Crystals in this condition grew within a week and after two weeks were observed to diffract to the resolution limits of the in house detector (Figure 69.).



**Figure 69.** Images of TcTS crystals grown in conditions, A: 200 mM L-proline, 100 mM Hepes and 10% PEG 3350, B: 100 mM Tris pH 8.5 and 10% PEG 8000.

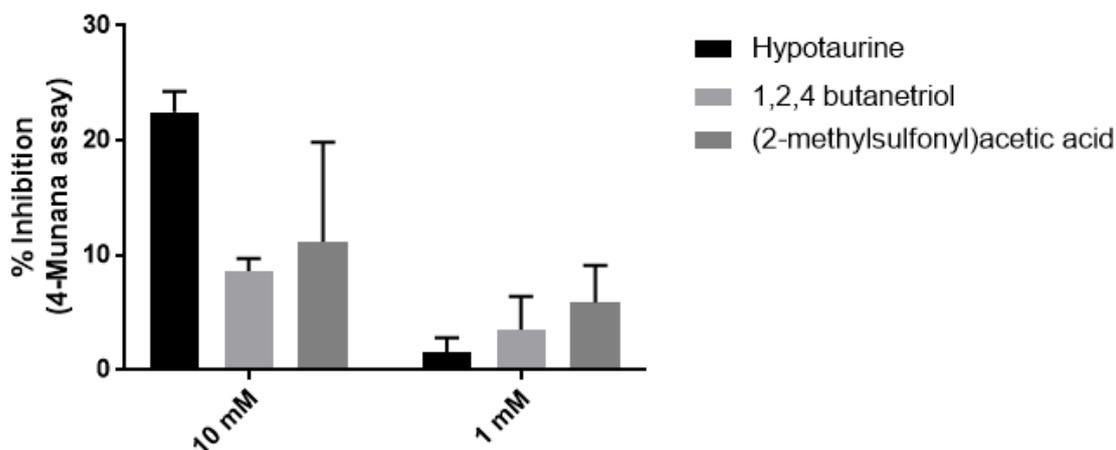
The second condition used was 100 mM Tris pH 8.5 and 10% PEG 8000. Crystals grew within two weeks and after one month were observed to diffract to the resolution limits of the in house detector. This condition produces an apo crystal structure with no small molecules bound within the active site. The time for the crystals to grow could be optimised and streak seeding of previously grown crystals into the crystallisation drop drastically shortened the length of time needed for crystals to grow. The first condition required one week and the second condition only required two weeks after crystal seeding. L-proline was observed to serendipitously bind within the active site of TcTS and form a number of van der Waal interactions with TcTS active site (Figure 70).



**Figure 70.** Proline (purple) found bound within the active site of TcTS. Interactions of Proline with surrounding molecule (DMSO, pink) and protein highlighted with dashed bonds (black). Phosphate also located within the active site. Water molecules are shown as red spheres. Figure created using PyMOL.

### ***5.7 Fragment activity and crystal soaking***

The top three hits (hypotaurine, 1,2,3-butanetriol and (2-methylsulfonyl)acetic acid) were bought in from commercial sources. A single molecule of DMSO is the only “serendipitous” binder within this secondary site. The lack of information on “serendipitous” binders within this site provides a major challenge to identify key functionalities for binding and selectivity. All fragments run by FlexX had low binding scores and are therefore likely to be very weak binders. Unsurprisingly, low inhibition was observed at 1mM or 10mM for hypotaurine, 1,2,3-butanetriol and (2-methylsulfonyl) acetic acid in the 4-Munana assay. Despite low inhibitory activity for hypotaurine and (2-methylsulfonyl) acetic acid these fragments were screened against the mapped site using ligand soaking and X-ray crystallography. Unfortunately, no fragment was identified to bind to the mapped “allosteric” site of TcTS through crystal soaking. No structural information and very low potency (Figure 71) provides a challenging prospect for further development.



**Figure 71.** Graph showing the percentage activity of each fragment against TcTS using the 4-Munana assay. Graph created using Microsoft Excel and GraphPad.

This CADD approach to identifying secondary site binders was unsuccessful. An alternative approach to the identification of TcTS chemical tools was pursued. With the wealth of knowledge generated from groups targeting the active site, a re-task of CADD screening towards this site was likely to be more successful and the following work details our approach to identifying an active site directed inhibitor.

## 5.8 Summary

The water channel within TcTS runs through the catalytic domain forming a central pore. This structural feature also appears within NanB despite only a 19% sequence similarity between the two sialidases. The TcTS water channel is 37.5 Å long and has a width of 16 Å. A number of conserved waters appear within the water channel of TcTS structures published in the PDB. Using a solved TcTS<sub>F58N</sub> crystal structure (previously solved by Dr Telford) a potential allosteric site was mapped using the allosteric site of NanB as a template. Creation of an enriched “virtual” docking library proceeded by obtaining a freely available fragment library database in a ready-to-dock format (only including fragments with a logp equal to or less than 3.5, molecular weight equal to or less than 250 Da and rotatable bonds equal to or less than 5) and filtering it using a ligand similarity search set at 0.2 (20%) against DMSO (the only known molecule to bind within this mapped pocket). This reduced the library size from 3,459,596 to 704,041 fragments. This enriched “virtual” library was then docked against the potential allosteric site within TcTS. The library was also docked against the active site to remove potential promiscuous binders and the top three hits that scored low for active site binding and high for binding within the potential allosteric pocket of TcTS were bought in from commercial

sources. The activity of these three hits (hypotaurine, 1,2,3-butanetriol and (2-methylsulfonyl)acetic acid) against TcTS in the 4-Munana assay was low. Crystal soaking of hypotaurine, 1,2,3-butanetriol and (2-methylsulfonyl)acetic acid performed in parallel to the 4-Munana assay did not generate any ligand-crystal complexes. With very low activity and no further structural information to aid design the SBDD approach to the development of a TcTS allosteric inhibitor is limited. It is possible an allosteric site exists within the water channel of TcTS, but a large compound screen would be required in order to identify functionalities important for binding within this potential allosteric pocket. The CADD approach was unsuccessful at identifying ligands that would bind within this mapped site. A wealth of information exists on inhibitors that bind within the active site of TcTS and a re-task of CADD screening towards the active site would likely be successful.

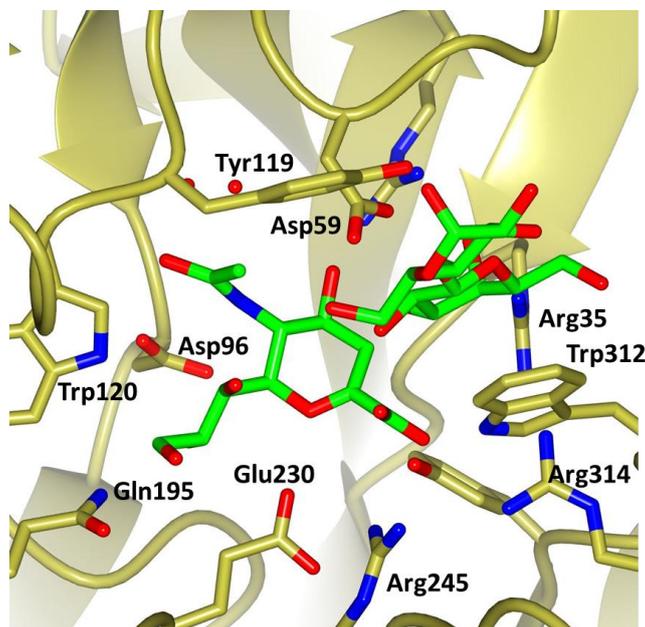
## ***6.0 Development of a TcTS constrained chemical tool that targets the active site***

This chapter is focused on the development of a constrained chemical tool that targets the active site of TcTS. To do this a CADD approach was used to identify novel TcTS active site binders. The crystal structures previously generated within the group were utilised to generate receptor files for docking. Within these structures and the TcTS structures published in the PDB, the active site of TcTS has been observed to form two conformations (Tyr119 flipped up (open conformation) and Tyr119 flipped down (closed conformation)) (see Chapter 1.62, Current Inhibitors of TcTS). Therefore, two receptor files for CADD were created using both the open and closed conformations. A large “lead-like” library was obtained from the ZINC<sup>12</sup> online database. This “lead-like” library was filtered using a subset of known active site binders of TcTS to create an enriched library (increasing the likelihood of finding an active site binder). The “lead-like” library was filtered from 13,760,200 to 60,256 compounds using a Tanimoto chemical similarity search criteria. This enriched library was docked against the two receptors files using FlexX. A KNIME workflow was used to speed up the docking process. Two “Hits” from this CADD screen were chosen for activity analysis. Activity analysis using the 4-Munana assay observed that the first “hit” inhibits TcTS activity. Crystal soaking however did not result in an inhibitor-TcTS complex. The second “hit” was evaluated, but due to cost of material the compound was only investigated as fragments. The core unit of this inhibitor was evaluated for activity and found to have moderate potency. A ligand-TcTS complex was solved indicating that the core fragment binds to the active site of TcTS. Analysis of analogues of this five-membered-ring fragment suggests that this core is suboptimal for binding and could be improved. The activity and structural analysis of these inhibitors provides a starting point for the development of a constrained chemical tool targeting the active site of TcTS.

### ***6.1 TcTS active site***

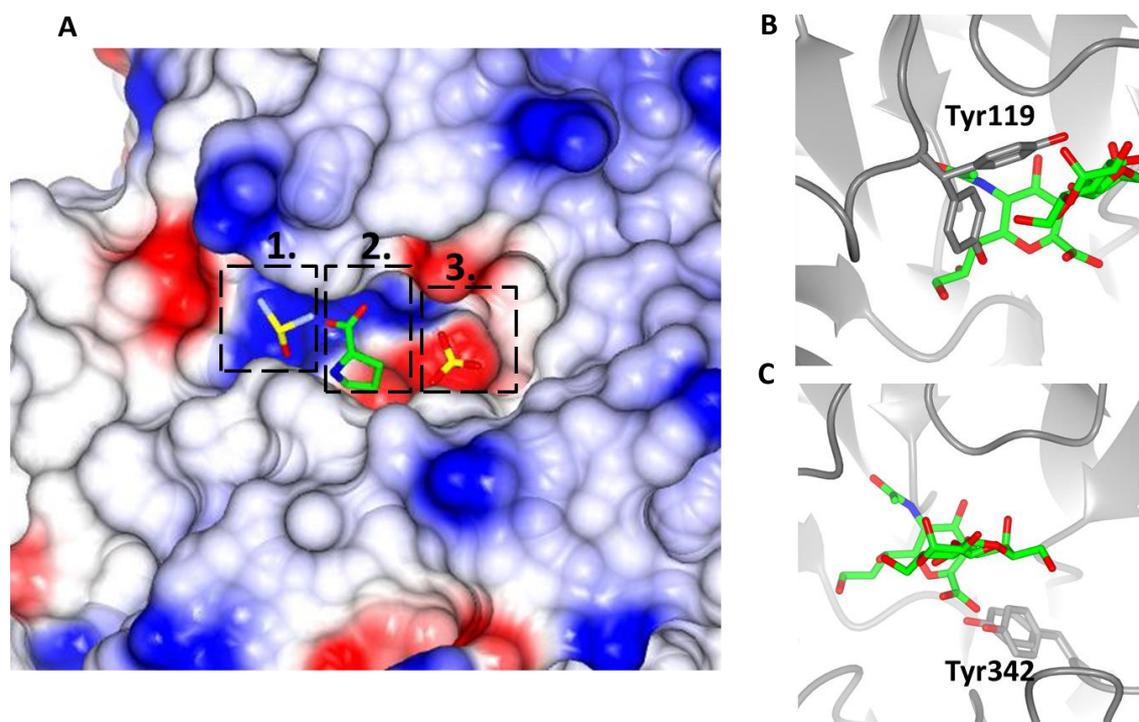
The TcTS active site is located within a canonical  $\beta$ -propeller fold in the N-terminal catalytic domain (see Chapter 1.57.1, TcTS Sialidase). The active site accommodates the binding of  $\alpha$ -2,3 linked sialic acid from host glycoconjugates. The key residues within the active site are Trp312 and Tyr119 as the side chains of these two aromatic amino acids form stacking interactions with aglycon moieties. Lactose has very few protein hydrogen bonds but abundant water hydrogen bonds within the TcTS active site. The TcTS active site has a greater affinity for lactose than Sia suggesting that binding is driven by hydrophobic affinity from these two amino acids (Trp312 and Tyr119). Other key

residues that form important ligand interactions include the arginine triad (Arg35, Arg245 and Arg314) and two aspartic acid residues (Asp59 and Asp96) (Figure 72).



**Figure 72.** Binding position of lactose (green) and DANA (green) within the active site of TcTS (PDB: 1MS0). The key residues important for substrate binding are highlighted.

Three pockets have been identified within the TcTS active site. Within the proline crystal structure solved by Dr Telford, L-proline, a phosphate ion and DMSO molecules occupy these pockets. Developing a chemical tool that can bind in all three pockets might prove challenging as it is unlikely a flat molecule will occupy all pockets forming all the key interactions required for the best affinity. A molecule with stereochemistry would be needed complicating ligand design and synthetic tractability.



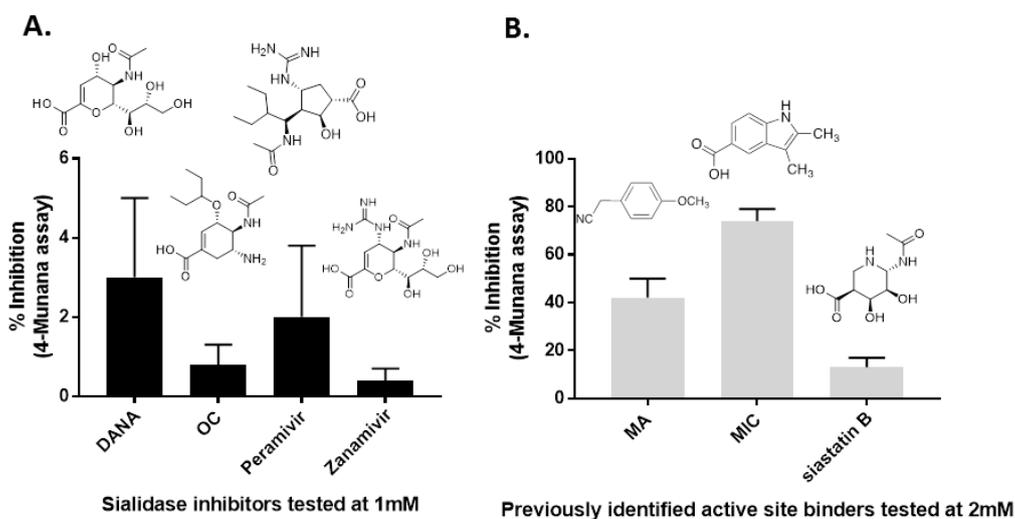
**Figure 73.** A. Three pockets identified within the TcTS binding site identified by Dr Telford. 1. Lactose binding site (DMSO bound). 2. Sia binding site with L-proline bound (green). 3. Phosphate binding pocket. B. Tyr119 exists in two conformations (PDB: 1MS3). The two positions of Tyr119 are highlighted in this image. C. Tyr342 also exists in two conformations these positions are highlighted in this image.

Further difficulty in inhibitor design is presented by the presence of two conformational changes in the Tyr119 (Figure 73). In the *holo* structure, Tyr119 flips up and into an open conformation for lactose binding. In the *apo* structure, Tyr119 flips down filling the pocket and forming a closed conformation. Two receptor files were created containing the TcTS active site (one with the open conformation and one with the closed conformation of Tyr119). The residue Tyr342 also exists between two states, the inactive state and the active state (Figure 73). In the active state, a hydrogen bond exists between Tyr342 and the carboxylate of Glu230 orientating the tyrosine underneath the scissile glycosidic linkage. Tyr342 in this position is thought to stabilise the substrate and/or the intermediate formed in the reaction. In the inactive state/unliganded structure, Tyr342 does not form a hydrogen bond with Glu230. Both the receptor files created contain the active state of Tyr342 to provide the most appropriate TcTS active site models for ligand docking.

### **6.1.1 Known binders of the TcTS active site**

A number of research groups have developed inhibitors of TcTS. These focus mainly on substrate mimetics of sialic acid (see Chapter 1.62, Current Inhibitors of TcTS). These inhibitors have been reported in the literature to have a weak potency against TcTS. To confirm these results the sialidase mimetics we had available in the lab were tested against TcTS (Figure 74). These sialidase mimetics were observed to have low potency against TcTS confirming previous results from other labs.

The most potent TcTS inhibitor reported to date is an anthraquinone, 6-chloro-9,10-dihydro-4,5,7-trihydroxy-9,10-dioxo-2-anthracenecarboxylic acid developed by Arioka *et al.* 2010. The TcTS inhibitor 6-chloro-9,10-dihydro-4,5,7-trihydroxy-9,10-dioxo-2-anthracenecarboxylic acid was not commercially available to purchase, however a structurally similar analogue anthraquinone-2-carboxylic acid was available from Sigma-Aldrich. This compound was difficult to dissolve in DMSO and so testing of this compound at final concentrations of 50 $\mu$ M and higher proved difficult. At 50 $\mu$ M anthraquinone-2-carboxylic acid was found to have little inhibitory activity against TcTS (7 $\pm$ 3 % inhibition of TcTS in the 4-Munana assay). According to Arioka *et al.* anthraquinone-2-carboxylic acid has an IC<sub>50</sub> of 430  $\mu$ M and at concentrations of 0.1mM or higher, the activity of TcTS is inhibited in the 4-Munana assay. An attempt to generate an anthraquinone-2-carboxylic acid-TcTS crystal structure was unsuccessful. This was probably due to the inability to dissolve anthraquinone-2-carboxylic acid at a concentration suitable for crystal soaking. It is likely the inhibitor 6-chloro-9,10-dihydro-4,5,7-trihydroxy-9, 10-dioxo-2-anthracenecarboxylic acid would be more soluble due to the presence of a number of polar groups that would contribute to hydrogen bonding with solvent.



**Figure 74.** **A.** Graph showing the percentage inhibition of known sialidase inhibitors against TcTS activity using the 4-Munana assay (DANA, Oseltamivir carboxylate (OC), Peramivir and Zanamivir). **B.** Graph showing the percentage inhibition of previously identified active site binders against TcTS activity using the 4-Munana assay (2-(4-methoxyphenyl)acetonitrile (MA), 2-3-dimethyl-1H-indole-5-carboxylic acid (MIC) and siastatin B).

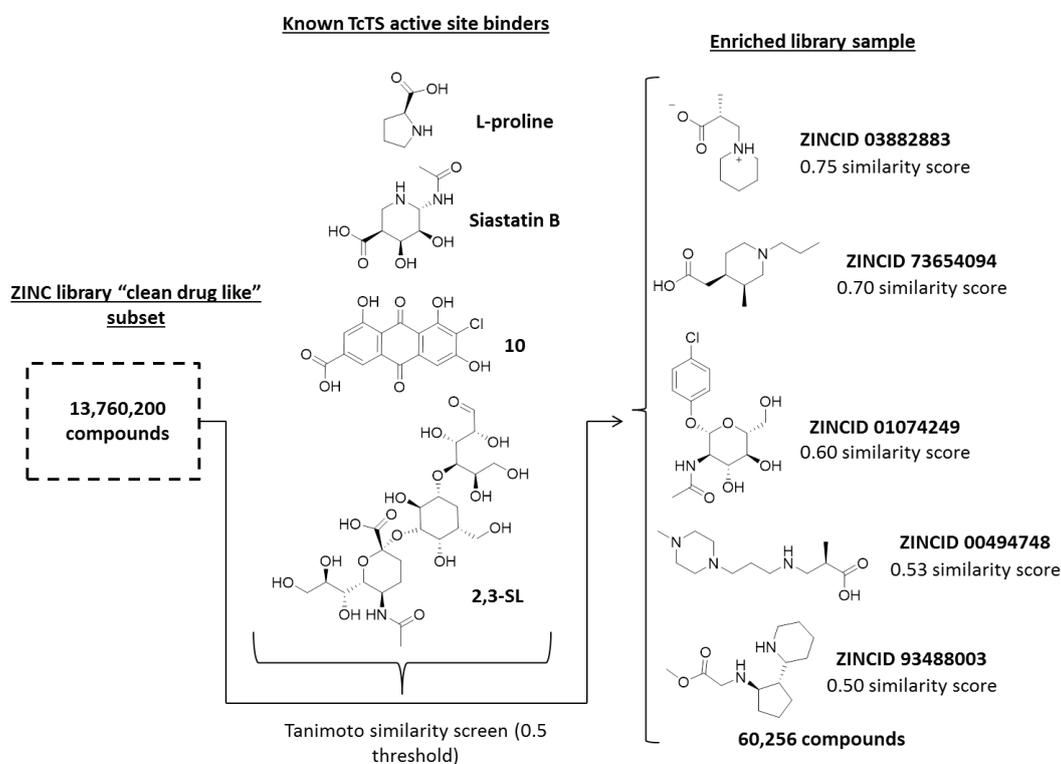
Dr Telford in previous work identified three main active site binders and inhibitors of TcTS (2-(4-methoxyphenyl)acetonitrile, 2-3-dimethyl-1H-indole-5-carboxylic acid and siastatin B). These inhibitors were re-tested against TcTS and observed to have weak potency confirming Dr Telford's results (Figure 71.B.). The percentage inhibition of these compounds at 2mM was determined to be  $42 \pm 8\%$  ( $38.2 \pm 11\%$ ),  $74 \pm 5\%$  ( $78.0 \pm 2.1\%$ ) and  $13 \pm 4\%$  ( $10.3 \pm 4.5\%$ ) respectively (percentage inhibition obtained by Dr Telford shown in parentheses). Within the TcTS crystal structure generated by Dr Telford, L-proline was identified as a serendipitous binder within the active site (Figure 67).

## 6.2 Library generation and docking

To improve the likelihood of finding a 'hit' against the active site a larger 'virtual' library was taken from the ZINC<sup>12</sup> database. To comply with the constrained chemical tool approach the "clean drug-like" library containing 13,760,200 molecules was chosen. The following filters applied to this set of molecules include: molecular weight (molecular weight above 150 Da and below 500 Da), hydrogen bond donor (maximum of 5), hydrogen bond acceptor (maximum of 10), logP (maximum of 5) and polar surface area (PSA) criteria (maximum of 150 kcal/mol). These comply with Lipinski's RO5 guidelines for increasing the likelihood of drug permeability and adsorption. PSA is a useful parameter for the prediction of adsorption (Ertl et al., 2000). PSA is defined by the sum of all polar surfaces within a molecule and correlates well with passive molecular transport data (Ertl et al.,

2000) enabling successful prediction of Caco-2 penetration (Palm et al., 1998) , blood-brain barrier transport (Kelder et al., 1999) and intestinal absorption (Palm et al., 1997).

To create an enriched library, these 13,760,200 compounds were filtered using a Tanimoto similarity ligand screen of known substrates and active site binders of TcTS (0.5 set threshold) (Figure 72). The known binders of TcTS were profiled and tested against TcTS using the 4-Munana assay. These binders had low potency against TcTS, but have been evaluated as TcTS inhibitors. This subset library of TcTS inhibitors will be used for chemical similarity comparison. As most of the inhibitors included in the subset library have been identified as TcTS active site binders through X-ray crystallographic studies, this provides a useful basis for obtaining the required chemical functionality for an active site inhibitor.



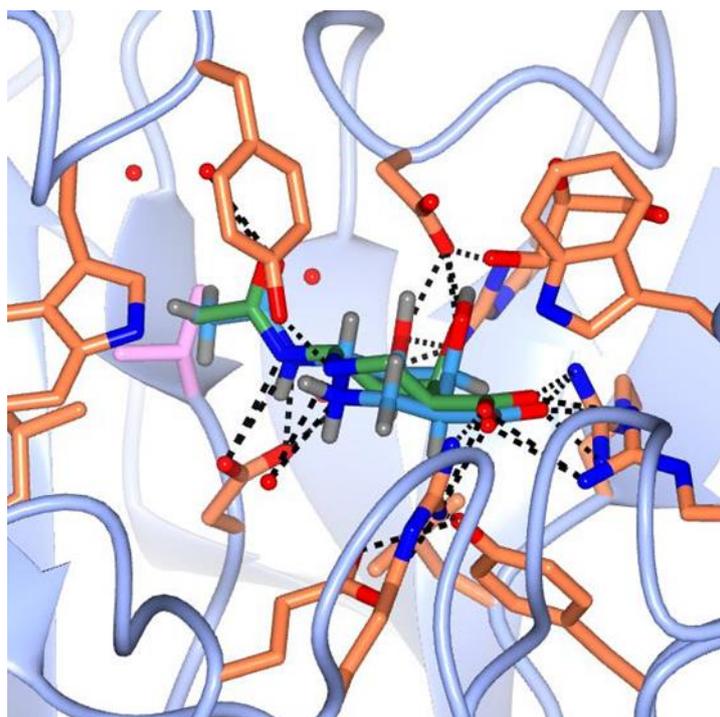
**Figure 75.** A schematic of the ligand similarity selection applied to the large "clean drug like" ZINC<sup>12</sup> library.

This reduced the library to 60,256 compounds. This was split into 100 sdf files containing between 203 to 863 molecules in each docking file. This allows the use of the graphical user interface (GUI) input within LeadIT. This user interface allows the maximum of 1,000 ligands docked at one time. To dock larger libraries of 1,000 ligands or more the commandline interface or a

Konstantz Information Miner (KNIME) workflow can be used. This 60,256 compound library was screened against the active site using FlexX within the docking programme LeadIT.

### 6.2.1 FlexX docking and validation

LeadIT was the docking programme of choice for this large CADD screen due to its faster run time than GOLD. This docking programme was validated through the evaluation of a known binder of TcTS. FlexX could accurately predict the binding of siastatin B (the binding mode of siastatin B was experimentally determined by Dr Telford in previous work).

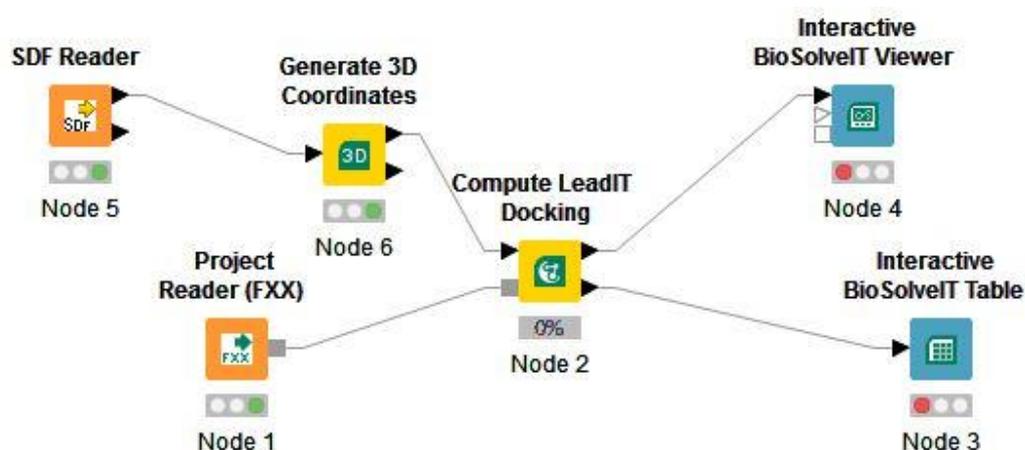


**Figure 76.** Siastatin B binding position within the TcTS active site (Telford, 2014). FlexX prediction of siastatin B within TcTS (blue) and binding position determined by siastatin B-TcTS crystal structure (green).

The overlay of the experimentally determined binding pose of siastatin B matches with the best binding pose generated by FlexX (Figure 76). As FlexX docking programme can accurately predict the binding pose of this inhibitor, docking of the large library was then initiated using the GUI interface within LeadIT.

## 6.2.2 Faster docking using the KNIME workflow

The use of the GUI interface within LeadIT is limited to one input file (containing less than 1000 ligands) per docking run. After each docking run a new library needs to be manually input if docking ligand files containing libraries of 1000 or more. The KNIME workflow was used to decrease the need for the manual input of each ligand batch file. KNIME workflow enables the user to visually assemble and interactively execute data within a pipeline (Berthold et al., 2009). Data is processed or visualised by nodes (Warr, 2012). Data processed and completed within one node is then passed onto the next node within the series (Warr, 2012). A number of advantages exist with the use of KNIME as multiple sdf files can be input into the pipeline and data can be viewed on intermediate results even after execution of the workflow (Warr, 2012). Furthermore, the workflow can be restarted at any intermediate node within the pipeline (Warr, 2012).



**Figure 77.** Schematic representing the KNIME workflow for FlexX docking within LeadIT.

The enriched library created within Instant JChem was loaded from the individual batch files into the SDF Reader Node 5. The 3D coordinates of each ligand are generated using the Generate 3D Coordinates Node 6. The receptor files were created within the GUI of LeadIT and loaded into the Project Reader (FXX) Node 1. The ligands were then docked using the Compute LeadIT Docking Node 2 and the docking output visualised using the BioSolveIT viewer Node 4 and Interactive Table Node 3. The KNIME workflow method was used to aid the docking of this large library (Figure 77).

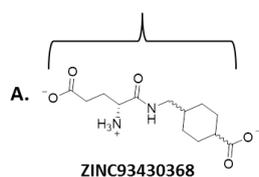
### **6.2.3 Docking scores.**

The large “clean drug like” library was docked against the open conformation using FlexX. The FlexX docking scores generated for each of the 60,256 compounds were filtered to include compounds that only scored higher than -55 (resulting in 1277 identified “hits”). These were compiled into a spreadsheet and then ranked according to FlexX score, the compounds were then docked using GOLD. GOLD was used to confirm and validate the docking scores obtained by FlexX for the open conformation. GOLD was also used to screen these 1277 identified “hits” against the closed conformation of the TcTS active site (Figure 78). As it was unclear as to which conformation would be useful for inhibitor generation, compounds that scored high for both receptor conformations were selected for further assessment. The top ten “hits” from all tables were profiled and the chemical structures were compared. Within the top ten “hits” for the open and the closed conformation (identified using FlexX and GOLD), the most common chemical structure was a five membered ring containing a carboxylic acid and two amines. Within the top ten “hits” for the open conformation identified within GOLD, the most common chemical structure is a heterocyclic ring. This is different to the results from FlexX. However, one structure within the top ten “hits” (for the open conformation) identified using GOLD contains a five membered ring containing a carboxylic acid and two amines.

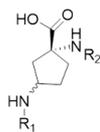
**Table 11.** Top ten “hits” identified using FlexX for the open conformation of TcTS. **Table 12.** Top ten “hits” identified using GOLD for the open conformation of TcTS. **Table 3.** Top ten “hits” identified using GOLD for the closed conformation of TcTS.

**Table 6.**

ZINCID	Score (FlexX, open)
ZINC93430368	-72.3463
ZINC93823870	-67.7992
ZINC93825333	-67.3708
ZINC93827840	-66.6174
ZINC93828541	-66.5453
ZINC93825344	-66.5293
ZINC93823888	-66.4955
ZINC93829071	-66.3571
ZINC93832105	-66.267
ZINC93822772	-66.2573



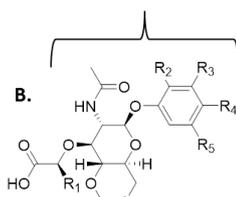
ZINC93430368



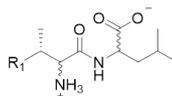
ZINC93823870, ZINC93825333,  
ZINC93827840, ZINC93828541,  
ZINC93825344, ZINC93823888,  
ZINC93829071, ZINC93832105 &  
ZINC93822772

**Table 7.**

ZINCID	Score (GOLD, open)
ZINC35057649	178.68
ZINC35057016	177.82
ZINC35057694	177.04
ZINC35057637	174.79
ZINC35057625	174.69
ZINC40471238	172.72
ZINC25624992	172.12
ZINC94560181	171.92
ZINC01532221	170.63
ZINC93830107	170.5



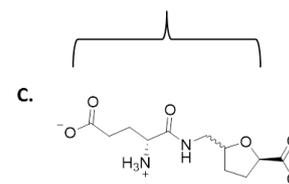
ZINC35057649, ZINC35057016, ZINC35057694,  
ZINC35057637 & ZINC35057625



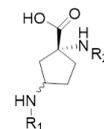
ZINC40471238 & ZINC01532221

**Table 8.**

ZINCID	Score (GOLD, closed)
ZINC93824779	193.82
ZINC94465165	189.16
ZINC93825175	189.08
ZINC93824790	188.99
ZINC94560180	188.77
ZINC93830470	187.96
ZINC93825180	187.36
ZINC93824779	187.34
ZINC93828898	187.3
ZINC93825180	187.12



ZINC94560180



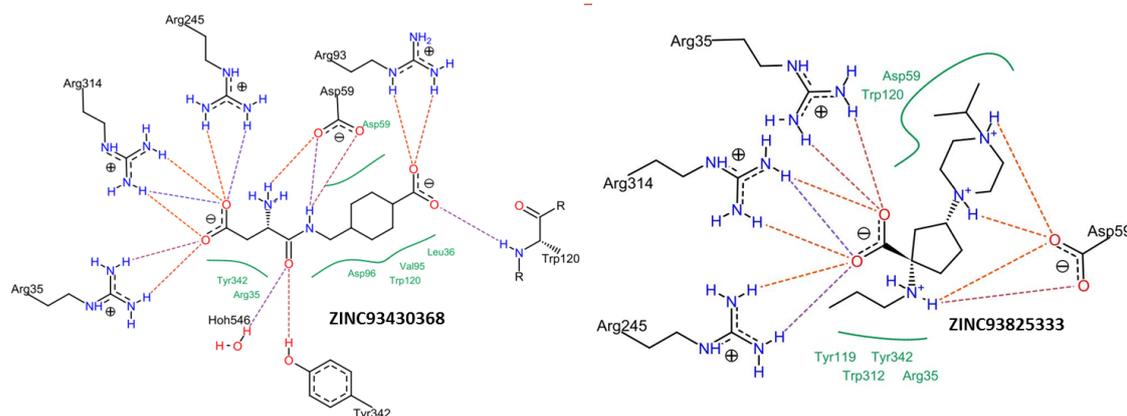
ZINC93824779, ZINC94465165,  
ZINC93825175, ZINC93824790,  
ZINC93830470, ZINC93825180 &  
ZINC93828898

**Figure 78.** **A.** Common chemical structure found within the top ten “hits” identified for the open TcTS active site structure by FlexX. **B.** Common chemical structure found within the top ten “hits” identified for the open TcTS active site structure by GOLD. **C.** Common chemical structure found within the top ten “hits” identified for the closed TcTS active site structure by GOLD.

All of the top ten hits contain a carboxylic acid moiety. This suggests that for a good docking score/affinity, the presence of carboxylic acid is important (presumably this is required for an interaction with the arginine triad). The presence of aromatic moieties within the top “hits” suggests that this is also required for a good docking score. It is likely that the presence of an aromatic functionality results in electrostatic stacking with Tyr119 and Trp312. Additionally, the presence of an amine within most of the top “hits” suggests this functionality is also important. It is likely this functionality forms van der Waal interactions with the aspartic acids (Asp 59 and Asp96). To evaluate if these assumptions hold true, manual evaluation of the binding pose generated for each of the chemical structures in Figure 78 were conducted.

### 6.2.4 Active site binding assessment

An example was chosen from each common chemical structure within the top “hits” identified from the CADD screen (Figure 79). The main residues involved in receptor-substrate interactions include: Trp312, Tyr119, the arginine triad (Arg53, Arg245 and Arg314) and two aspartic acid residues (Asp59 and Asp96).



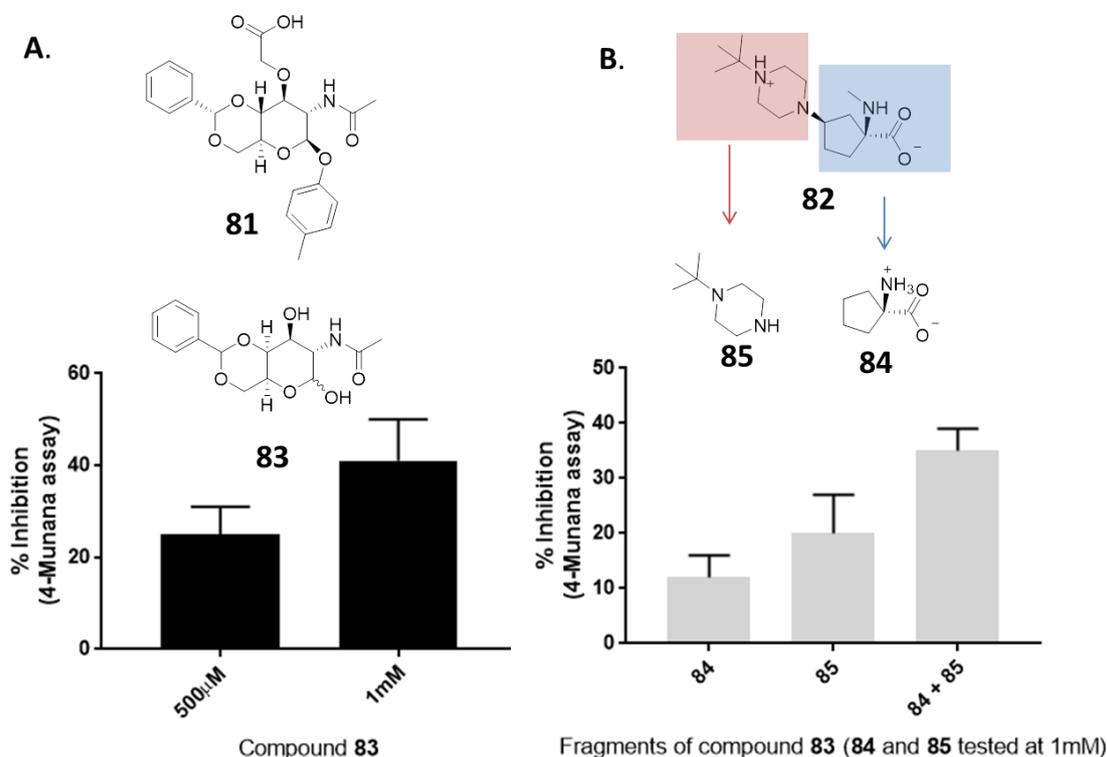
**Figure 79.** The binding pose generated from the FlexX screen for two examples of “hits” in Figure 75. Image created by PoseView within LeadIT.

For both the FlexX identified inhibitors a number of key interactions are observed between the arginine triad and Asp59. ZINC93430368 has a hydrogen bond interaction with Tyr342. ZINC93825333 has the potential for binding with Tyr119 and Trp312 with position of the alkane chain. It is likely this group is suboptimal and could be modified for a key interaction with Tyr119 and Trp312. The binding of “hits” from the GOLD screen were also evaluated. Based on the interactions with the key residues, two inhibitors were selected for further development and analysis. These two inhibitors were compounds ZINC35057649/**81** and ZINC93827840/**82** (Figure 80).

### 6.3 Biological testing of compound **81** and **82**.

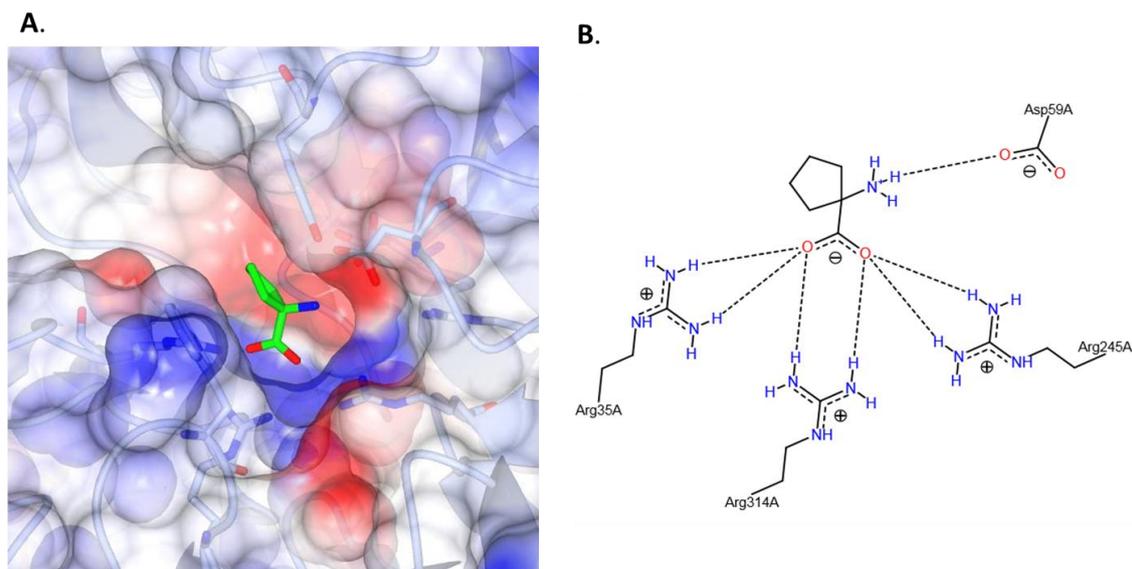
A close analogue of **81** was readily available to purchase through Carbosynth. This compound is 2-acetamido-4,6-O-benzylidene-2-deoxy-D-galactopyranose/compound **83**. Activity analysis using the 4-Munana screen evaluated compound **83** to have moderate potency in the 4-Munana assay (25±6% inhibition of TcTS activity at 500µM and 41±9% inhibition at 1mM). This is the best potency of all evaluated inhibitors so far. However, various attempts at crystal soaking with this inhibitor did not generate a co-crystal complex. Compound **82** was available commercially, but this compound is a make-on-demand compound requiring a 6 week development time. It is also expensive to buy. Evaluation of compound ZINC93827840 instead proceeded through fragment analysis. Compound **82** was instead broken down into functional units and investigated as individual fragments. As the

fragments were readily available and cheap to purchase, this was advantageous. Additionally, this core fragment featured commonly in many of the top hits and so activity and structural evaluation of this fragment alone would be very useful in identifying if it inhibits TcTS and binds in the active site.



**Figure 80.** **A.** Chemical structure of ZINC35057649/**81** and structurally similar analogue **83** and a graph representing the % inhibition of TcTS activity in the presence of **81** at 500 μM and 1 mM using the 4-Munana assay. **B.** Chemical structure of ZINC93827840/**82** and fragments **84**, **85** and a graph representing the % inhibition of TcTS activity in the presence of **81** at 500 μM and 1 mM using the 4-Munana assay. Evaluation of the fragments of this 'hit' was conducted using the 4-Munana assay.

The first fragment/**84** was observed to have 12±4% inhibition at 1 mM. Ligand soaking with **84** resulted in a **84**-TcTS crystal structure. The binding of **84** to the active site matched with that predicted from LeadIT. Further evaluation of an additional fragment **85** (tert-butyl piperazine) resulted in inhibition against TcTS 20±7% at 1 mM. Addition of both fragments together resulted in slightly improved inhibition of 35±4% inhibition at 1 mM suggesting that synergistic inhibition has occurred. However, ligand soaking with tert-butyl piperazine did not result in a ligand-TcTS complex. A crystal structure has been obtained with the core structural unit bound within the Sia pocket of TcTS (Figure 81).

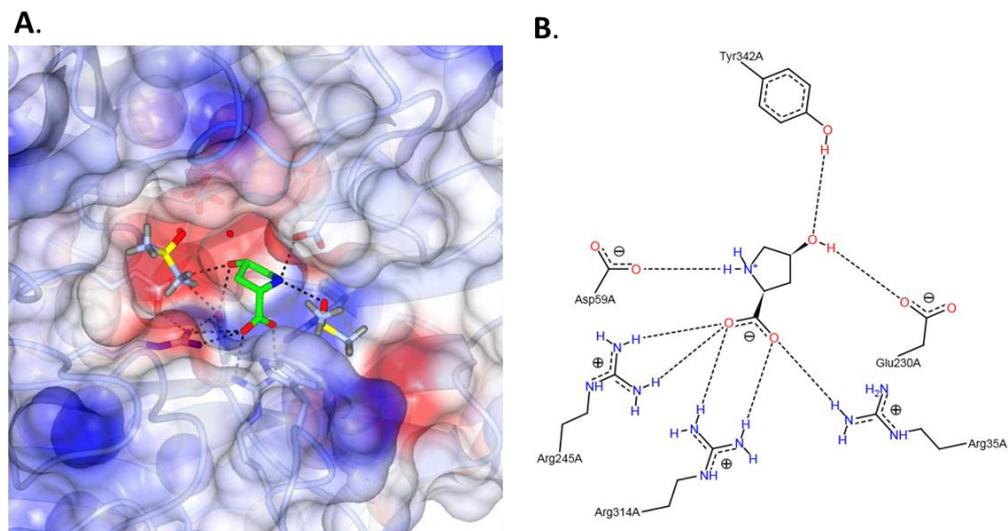


**Figure 81.** **A.** Binding position of **84** within the active site of TcTS. **B.** A 2D schematic of the interactions of **84** with the active site of TcTS as predicted by PoseView.

To gain further information on the optimal binding within the Sia binding site a number of other five membered analogues were tested.

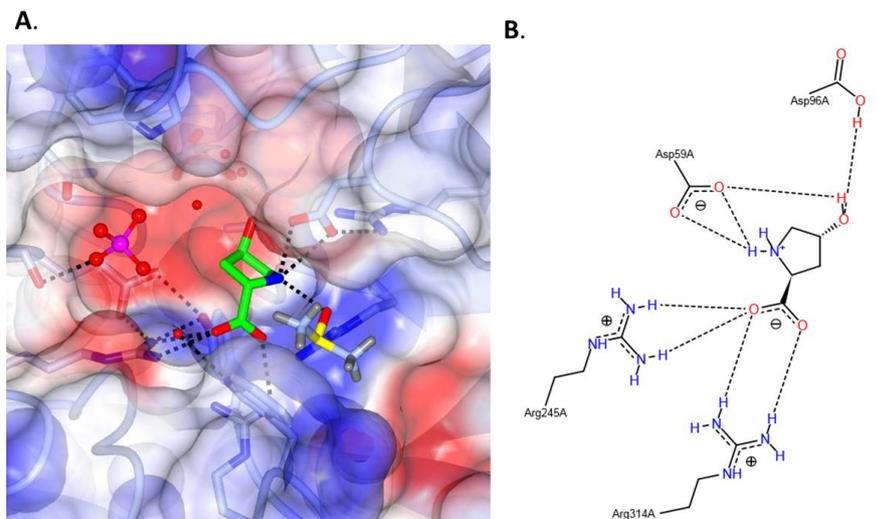
#### **6.4 SAR with analogues of 84**

Two analogues of **84** were trailed (*cis*-4-hydroxy proline and *trans*-4-hydroxy proline) and evaluated to identify the optimum position of the five membered ring within the TcTS active site. These analogues were chosen to also evaluate the position and required stereochemistry for the addition of functional groups to the five membered ring. A crystal structure of *cis*-4-hydroxy proline-TcTS was obtained (Figure 79). *Cis*-4-hydroxy proline was bound within the same position as L-proline and in a slightly different position to **84** (Figure 82).



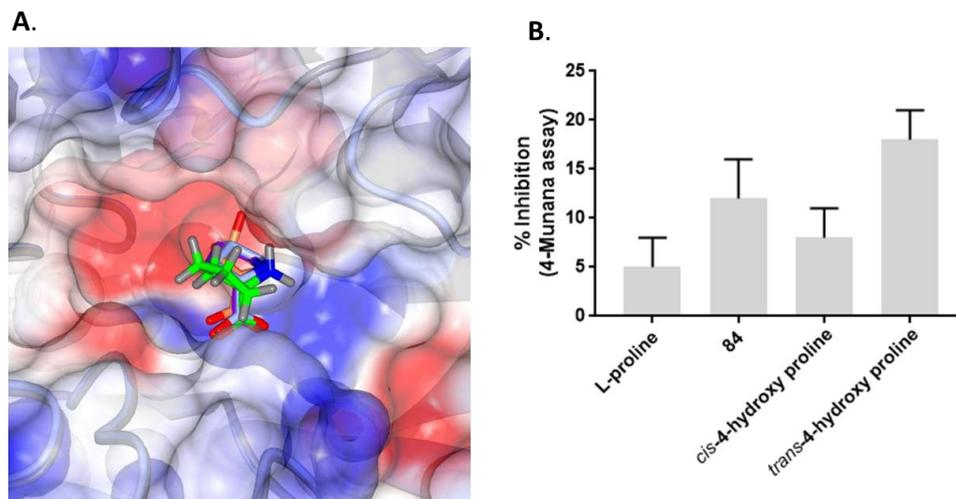
**Figure 82.** **A.** Binding position of *cis*-4-hydroxy proline within the active site of TcTS. **B.** A 2D schematic of the interactions of *cis*-4-hydroxy proline with the active site of TcTS as predicted by PoseView.

The van der Waals interaction between the amine and the aspartic acid residue (Asp59) of TcTS orientates the *cis*-4-hydroxy proline into a different position than that of **84** (Figure 82). Interactions exist between the arginine triad and the carboxylic acid of *cis*-4-hydroxy proline. The hydroxyl group of *cis*-4-hydroxy proline has additional hydrogen bonds with Tyr342 and Glu230, potentially increasing binding affinity. The isomer of *cis*-4-hydroxy proline, *trans*-4-hydroxy proline was also investigated as an active site binder and analogue of **84**. A crystal structure of *trans*-4-hydroxy proline-TcTS was obtained (Figure 80). *Trans*-4-hydroxy proline was observed to bind in a similar position to L-proline and *cis*-4-hydroxy proline (Figure 81). *Trans*-4-hydroxy proline was observed to have slightly better potency against TcTS than *cis*-4-hydroxy proline and **84**. This could be due to the presence of a hydrogen bond between aspartic acid (Asp96) and the hydroxyl group of *trans*-4-hydroxy proline ( ).



**Figure 83.** **A.** Binding position of *trans*-4-hydroxy proline within the active site of TcTS. **B.** A 2D schematic of the interactions of *trans*-4-hydroxy proline with the active site of TcTS as predicted by PoseView.

The slightly increased potency of *trans*-4-hydroxy proline suggests that the further optimisation of an active site inhibitor of TcTS should proceed with the use of *trans*-4-hydroxy proline or the addition of functionality to **84** that would result in a direct hydrogen bond interaction with Asp96.



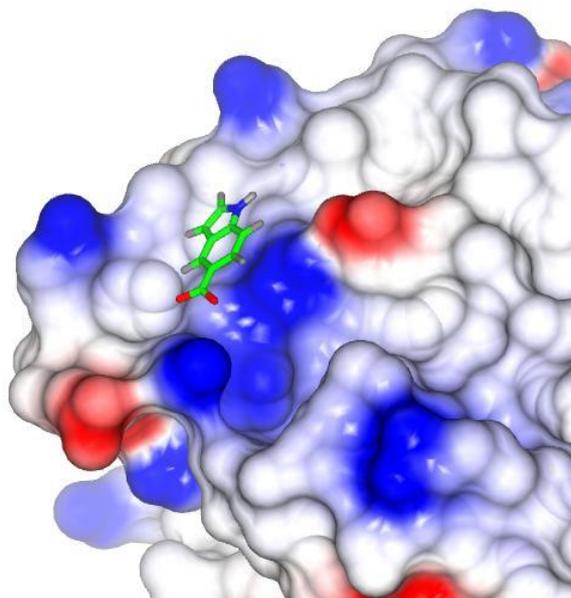
**Figure 84.** **A.** Overlaid binding positions of **84** (green), *trans*-4-hydroxy proline (orange), *cis*-4-hydroxy proline (blue) and L-proline (purple) within the TcTS active site. **B.** Graph representing the % inhibition of TcTS activity observed for **84**, *trans*-4-hydroxy proline, *cis*-4-hydroxy proline and L-proline (all tested at 1mM) using the 4-Munana assay.

## 6.5 Summary

Within this chapter a fragment was identified as a suitable starting point for chemical tool development. A CADD approach was used to identify an active site inhibitor. A large “clean drug like” library was used as a platform to generate an enriched library of compounds likely to bind to the TcTS active site. The information gained from the literature and previously identified active site binders within the group (Telford, 2014) were used to generate a table of active site binders for a chemical search filter. Similarly to Chapter 5, a Tanimoto similarity metric was used with a threshold of 50%. This filtered the initial 13,760,200 down to a smaller more manageable library of 60,256 compounds. This enriched library of 60,256 compounds was then screened against the active site of TcTS using FlexX. This was performed initially using the GUI interface within LeadIT and later completed with a KNIME workflow. The results from this screen were further filtered to include “hits” with a total FlexX score of more than -55. The docking was performed for a second cycle on this library of “hits” using GOLD. Within this second docking cycle, the two conformations of Tyr119 were used and so the library of “hits” was docked against two receptor files. The top ten hits from each docking cycle were evaluated. Common chemical structures appeared within the top ten “hits”. Two “hits” **81** and **82** were chosen for further evaluation based on the binding analysis of each of these “hits” from PoseView. The purchase and test of **83** a close analogue to **81** identified a moderate inhibitor of TcTS activity within the 4-Munana assay. However, no inhibitor-TcTS structure was solved for this heterocyclic compound **83** raises questions over the validity of this inhibitor as an active site TcTS inhibitor and would require further investigation. To improve the potency of **83** against TcTS a ligand-protein structure would be needed for SBDD. Testing of **83** proceeded through fragment analysis. Fragments of **82** also had moderate potency against TcTS using the 4-Munana assay. A crystal structure of the core fragment (**84**) of **83** bound to the active site of TcTS was generated. Further evaluation of analogues of **84** offers valuable information for the generation of a novel active site inhibitor of TcTS. This provides a good basis for future SBDD and sets the groundwork for the generation of a TcTS active site targeted constrained chemical tool.

## ***7.0 Discussion and Future work***

An allosteric site was not located within TcTS with a CADD fragment screen. It is unlikely that a CADD approach would be successful in identifying a small molecule binder of an allosteric site within TcTS without the identification of a key functionality required for binding to this site. Additionally, a higher ratio of hydrophobic residues within this mapped allosteric site would also lead to difficulties in generating/finding a small molecule that would bind to this site. DMSO was the only small molecule/fragment found to bind within this mapped region on TcTS. This is used at a high concentration (20-30% DMSO in crystallisation buffer) as a cryoprotectant for crystal structure generation. Even though DMSO is observed within the crystal structure of TcTS, it is likely to have a low binding affinity. Therefore, DMSO is not a good molecule on which to base a chemical similarity search and so the Tanimoto threshold was set low at 20%. Despite the low threshold the library was filtered from 704,041 to 1,015 (filtering the library by 99.9%). The reduction of this library is too high, greater chemical variety is needed in the future to identify functional groups important for binding at this site. The full fragment library should be screened against the mapped secondary site using the CADD approach. Without the availability of SBDD information provided by serendipitous binders, conventional methods should be used to determine if an allosteric site exists on TcTS. A large HTS should be utilised to identify fragments that bind and inhibit the TcTS. Crystal soaking and structural determination using X-ray crystallography should be used to identify if the fragment/“hits” determined to inhibit TcTS activity bind to an allosteric site. Using crystallisations conditions discovered by Dr Telford, it is possible to generate crystals of consistent quality and resolution for ligand binding determination.



**Figure 85.** Binding of indole-5-carboxylic acid (green) found bound on the surface of TcTS. Electrostatic surface of the protein shown.

An analogue of 2-3-dimethyl-1H-indole-5-carboxylic acid (MIC) was also tested in the 4-Munana assay and resulted in similar potency ( $70\pm 4\%$  inhibition of activity in the 4-Munana assay at 2mM). Although indole-5-carboxylic acid had weak potency against TcTS, it was used in crystal soaking to determine mode of binding. Within the crystal structure, indole-5-carboxylic acid was not observed to bind to the active site or the predicted allosteric site. It was found bound to a position outside of the water channel at the surface of the catalytic domain (Figure 85). Low potency of the indole-5-carboxylic acid suggests this inhibitor lacks affinity for TcTS. The position of the indole-5-carboxylic acid is positioned between crystal-crystal contacts. This binding could be an artefact of crystallisation and would need to be further investigated as this ligand may have serendipitously identified an allosteric site. Due to the disappointing results obtained from screening ligands against the mapped secondary site, the CADD approach was then used to determine if a new active site inhibitor could be discovered. The benefit of this approach is that examples exist of TcTS active site inhibitors and this information can be used to determine the functionality required for binding within the TcTS active site. However, the development of a potent active site inhibitor has proven to be difficult. Inhibitors that are sialic acid mimetics typically have low potency against TcTS. It is evident from past work that the lactose pocket, specifically binding to Trp312 and Tyr119 needs to occur in order for increased potency. "Hits" generated from the CADD screen were analysed with this in mind. Two

candidates (**81** and **82**) were selected based on these criteria for further analysis. The two compounds selected were difficult to obtain commercially and so close an analogue of **81**, **83** was obtained instead. The compound **82** was investigated as fragments. The difficulty in generating a **83**-TcTS crystal structure suggests that **83** may be promiscuous and not a true active site inhibitor of TcTS. This should be further investigated to determine if **83** binds to the active site of TcTS. Compound **84** is the core unit for many of the top “hits” from the CADD screen and a fragment of **82**. This compound has reasonable inhibitory activity (for a fragment) against TcTS despite only binding within the Sia binding pocket of the active site. The development of a more potent chemical tool should proceed with the addition of functionality to **84** that will occupy the lactose pocket of TcTS. Analysis of analogues of **84** suggest that this core fragment could be further improved. The positioning of a group that can hydrogen bond to Asp96 would slightly improve potency. However, addition of an aromatic functionality forming aromatic stacking interactions with Tyr119 and Trp312 would probably improve potency significantly. Further increases in potency could be generated by the addition of functionality into the third pocket (phosphate binding pocket) identified by Dr Telford. The work here has identified a valid starting point for further SAR development. It is highly likely a potent TcTS active site directed constrained chemical tool could be developed using fragment **84** as the core unit. Compound **83** would need to be further investigated. This analogue of **81** lacks the important carboxylic acid functional group for binding to the arginine triad. Addition of this group might significantly increase its potency.

## ***7.1 Experimental***

All reagents were purchased from chemical suppliers: Sigma, Carbosynth and Fluorochem. These chemicals were used without further purification.

### ***7.1.1 Computational modelling:***

Computational experiments were conducted using a Dell Inspiron 5520 Intel® Core i5-3210M 2.50 GHz CPU. A 64-bit operating system running Windows 10 was used. BioSolve IT LeadIT software containing the docking programme FlexX was operated for computational docking. The Cambridge Crystallographic Data Centre (CCDC) software GOLD was also employed for computational docking. Ligand libraries were obtained from the freely available chemical database ZINC<sup>12</sup> (obtained from: <http://zinc.docking.org/>). Two libraries were downloaded from the ZINC<sup>12</sup> database. These were the ZINC<sup>12</sup> “Frag Now” library containing 704,041 fragments and the ZINC<sup>12</sup> “Clean Drug-like” library containing 13,760,200 compounds. Chemical similarity profiling was performed by ChemAxon software Instant JChem. A Tanimoto similarity metric was used to search for ligands within these libraries that had similar chemical characteristics as the target group. The target group chosen was

based on using known binders/inhibitors of the sites prepared for docking. Docking receptor files were created using the GUI within LeadIT. Receptor files for docking were generated from structures obtained within the group ((Telford, 2014)). Ligands bound within these sites were used as a basis to generate a receptor file. A circumference cut-off of between a 6 and 15 Å around the bound ligand was used to create the receptor for docking. Docking was performed using the GUI within LeadIT or KNIME workflow provided by BioSolveIT and KNIME. GOLD docking was performed by using the software's GUI. Docking scores obtained from LeadIT were manually compiled into excel files. GOLD automatically creates log files containing docking scores.

### **7.1.2 Protein expression:**

The TcTS plasmid 5 µl (pTrcHis) supplied by the Busciazzo group was subjected to site-directed mutagenesis using designed primers (Appendix Figure 2) and the QuickChange Lightning Site-Directed Mutagenesis Kit (Stratagene). The resulting construct was then transformed into DH5α *E.Coli* and plated onto 100µg/mL L-ampicillin plates. These plates were then incubated at 37°C overnight. Colonies were selected and amplified in 5mL of LB with a final concentration of 100µg/mL of L-ampicillin. The DNA plasmid was isolated by mini-prep (Qiagen) and sent to GATC Biotech for sequencing (in accordance with previous work (Telford, 2014)). Once the sequence was confirmed the new TcTS construct (TcTS<sub>F58N</sub>) was transformed into 100 µL of *E.Coli* (BL21 GOLD) by incubation on ice for 45 mins. The *E.Coli* was then heat shocked for 30 seconds and then put back on ice for 2 minutes. The cells were then added to 1 mL of L-B Broth and incubated at 37 °C. After incubation for 1 hr 200 µl of cells were spread onto L-B agar plates containing 100 mg/mL L-ampicillin. These cells were incubated overnight at 37 °C. A single colony was then inoculated into 10 mL of L-B Broth containing 0.1 mg/mL. From this overnight inoculation 1 mL was transferred into autoinduction media (Terrific broth base with trace elements, ForMedium) containing 0.1 mg/mL L-ampicillin and incubated for 2 hrs at 37 °C at 200 rpm. The temperature was then turned down to 22 °C and incubated for 48 hrs briefly stopping after 30hrs to add a further 50 µg/mL L-ampicillin. Cells were then harvested by centrifugation at 9000 rpm for 25 minutes and stored at -80°C overnight.

### **7.1.3. Protein purification:**

The cell pellet was thawed and resuspended in 200 mL of Tris buffer (50 mM Tris. pH 7.5) with 2 protease cocktail tablets (Roche Diagnostics) and DNase I (Sigma, 20 µg/ml final concentration). The cells were lysed by a cell disruptor (Constant Systems). Cell debris was then removed by centrifugation at 20,000 rpm for 25 minutes at 4 °C. The supernatant was then filtered using a syringe 0.2 µm filter. The filtered supernatant was loaded onto a nickel HisTrap HP column (GE Healthcare) equilibrated with 50mM Tris, pH 7.5). The protein was eluted with a gradient of 10 mM-

500 mM imidazole. The fractions containing TcTS were collected and to a 120 ml Sephacryl S-200 gel filtration column. Fractions containing TcTS were collected and analysed using an SDS-PAGE gel. Protein was concentrated to 12mg/mL for crystallisation experiments.

#### **7.1.4. Kinetic analysis:**

The MUNANA assay was used to measure the activity of “hits” identified from computational screening. Stock solutions were made of each compound in DMSO (0.1 M) or in water depending on solubility. Compound were initially assayed at concentrations ranging between 2-1mM. The MUNANA was assayed in 96 well plates with each well containing a 100  $\mu$ l total volume containing Hepes, 100mM NaCl pH 7.5 and a final concentration of 100 ng/mL TcTS, 120  $\mu$ M MUNANA and compound (concentrations ranging between 2 mM, 1mM, 500  $\mu$ M). The assay was performed using the Stratagene MX3005P PCR system and fluorescence measured at 365 nm excitation and 420 nm emission. Recordings were taken at 37 °C every 10 seconds for 10 minutes. The initial rate was calculated and compared to a control giving the percentage inhibition.

#### **7.1.5. Protein crystallisation and X-ray crystallography:**

Crystal trials were set up using sitting drop vapour diffusion 96 well plates (Douglas instruments). 2  $\mu$ L of 12 mg/mL protein and 2  $\mu$ L of buffer was added into the sitting drop well. In the reservoir 70  $\mu$ L of buffer was added. The buffer consisted of 200mM L-proline, 100mM Hepes and 10% PEG 3350. Crystals in this condition grew within a week and after two weeks had grown to suitable size. A second crystallisation condition was also used. The buffer within this condition was 100 mM Tris pH 8.5 and 10% PEG 8000. Crystal were slower to grow in this condition. Crystal seeding was employed to speed crystal growth. For seeding a single crystal (obtained from the L-proline condition) was lifted using a crystal loop and was suspended in 50  $\mu$ L of 100 mM Tris pH 8.5 and 10% PEG 8000 buffer. A seed bead was added and the solution vortexed for 60 seconds. Crystallisation conditions were set up as before with addition of crystal seeds by use of a cat whisker dipped into the seed solution and then into the sitting drop. The plates were then sealed and stored at room temperature.

Crystal soaking was performed with the generated TcTS crystals. Compounds were made up in stock concentrations of 1M. This was performed with DMSO for insoluble compounds or water for soluble compounds. The stock solution was added into the wells (to generate a 100mM final concentration of compound) containing the crystal for periods of time ranging from 2 mins to overnight. Additional soaking experiments included the addition of compound (water soluble compounds) at 200mM with 100mM Hepes and 10% PEG 3350.

Crystals were cryoprotected with 30% DMSO in crystallisation buffer for a few seconds prior to placing into a liquid nitrogen stream. All X-ray structural data were collected in house at 100 K using a Rigaku MSC Micromax-007 X-ray generator and Saturn 944+ CCD detector. HKL2000 was used to integrate and scale the diffraction images. SCALEPACKtoMTZ was used within the CCP4 suite to generate an MTZ file. Phaser was used with molecular replacement to solve the initial phases. The data was then subject to Refmac for refinement. Coot was then used to fit the ligand to the observed density.

Data collection	TcTS proline- indole-5- carboxylic acid	TcTS- cycloleucine	TcTS- <i>cis</i> hydroxyproline	TcTS- <i>trans</i> hydroxyproline
Beamline	In-house	In- house	In-house	In-house
Wavelength (Å)	1.54178	1.54178	1.54178	1.54178
Unit-cell parameters				
a (Å)	54.069	54.39	54.39	53.923
b (Å)	86.932	129.5	129.54	129.1
c (Å)	73.815	54.38	54.38	54.24
$\alpha$ (°)	90.00	90.00	90.00	90.0
$\beta$ (°)	99.557	107.79	107.794	107.89
$\gamma$ (°)	90.00	90.00	90.00	90.0
Molecules per symmetric unit	1	1	1	1
Resolution (Å)	60-1.87	50-1.95	50-1.97	50-1.68
Redundancy <sup>a</sup>	2.5(2.2)	2.8 (1.8)	3.0(4.8)	3.2(1.8)
Completeness (%) <sup>a</sup>	84.0(83.8)	91.6(52.5)	98.9(92.4)	60.2(2.6)
Rmerge (%) <sup>a,b</sup>	7.8(10.5)	7.0(21.3)	7.8(16.8)	3.0(13.0)
$I/\sigma I$ <sup>c</sup>	37(10.8)	24.3(3.96)	56.1(30.2)	51.5(8.96)

## 8.0 Bibliography

- ABED, Y., PIZZORNO, A. & BOIVIN, G. 2012. Therapeutic activity of intramuscular peramivir in mice infected with a recombinant influenza A/WSN/33 (H1N1) virus containing the H275Y neuraminidase mutation. *Antimicrobial agents and chemotherapy*, 56, 4375-4380.
- ADAM, M. 2005. Integrating research and development: the emergence of rational drug design in the pharmaceutical industry. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 36, 513-537.
- AGUSTÍ, R., PARÍS, G., RATIER, L., FRASCH, A. C. & DE LEDERKREMER, R. M. 2004. Lactose derivatives are inhibitors of *Trypanosoma cruzi* trans-sialidase activity toward conventional substrates in vitro and in vivo. *Glycobiology*, 14, 659-670.
- AHMAD, O. B., LOPEZ, A. D. & INOUE, M. 2000. The decline in child mortality: a reappraisal. *Bulletin of the World Health Organization*, 78, 1175-1191.
- AIR, G. M. 2012. Influenza neuraminidase. *Influenza and Other Respiratory Viruses*, 6, 245-256.
- AL-ALI, H. 2016. The evolution of drug discovery: from phenotypes to targets, and back. *MedChemComm*, 7, 788-798.
- AL-MAJHDI, F. N. 2009. Structure of the sialic acid binding site in influenza A virus: Hemagglutinin.
- AMAYA, M. A. F., WATTS, A. G., DAMAGER, I., WEHENKEL, A., NGUYEN, T., BUSCHIAZZO, A., PARIS, G., FRASCH, A. C., WITHERS, S. G. & ALZARI, P. M. 2004. Structural insights into the catalytic mechanism of *Trypanosoma cruzi* trans-sialidase. *Structure*, 12, 775-784.
- AMINOV, R. I. 2009. The role of antibiotics and antibiotic resistance in nature. *Environmental microbiology*, 11, 2970-2988.
- ANDERSON, A. C. 2003. The Process of Structure-Based Drug Design. *Chemistry & Biology*, 10, 787-797.
- ANGHEBEN, A., BOIX, L., BUONFRATE, D., GOBBI, F., BISOFFI, Z., PUPPELLA, S., GANDINI, G. & APRILI, G. 2015. Chagas disease and transfusion medicine: a perspective from non-endemic countries. *Blood Transfusion*, 13, 540.
- ANGULO, F. J., COLLIGNON, P., WEGENER, H. C., BRAAM, P. & BUTLER, C. D. 2005. The routine use of antibiotics to promote animal growth does little to benefit protein undernutrition in the developing world. *Clinical infectious diseases*, 41, 1007-1013.
- ARIAS, D. G., MARQUEZ, V. E., CHIRIBAO, M. L., GADELHA, F. R., ROBELLO, C., IGLESIAS, A. A. & GUERRERO, S. A. 2013. Redox metabolism in *Trypanosoma cruzi*: functional characterization of tryparedoxins revisited. *Free Radical Biology and Medicine*, 63, 65-77.
- ARIOKA, S., SAKAGAMI, M., UEMATSU, R., YAMAGUCHI, H., TOGAME, H., TAKEMOTO, H., HINO, H. & NISHIMURA, S.-I. 2010. Potent inhibitor scaffold against *Trypanosoma cruzi* trans-sialidase. *Bioorganic & medicinal chemistry*, 18, 1633-1640.
- ARMAH, F. A., QUANSAH, R., LUGINA, I., CHUENPAGDEE, R., HAMBATI, H. & CAMPBELL, G. 2015. Historical perspective and risk of multiple neglected tropical diseases in Coastal Tanzania: compositional and contextual determinants of disease risk. *PLoS Negl Trop Dis*, 9, e0003939.
- ARMOUR, D., DE GROOT, M. J., EDWARDS, M., PERROS, M., PRICE, D. A., STAMMEN, B. L. & WOOD, A. 2006. The Discovery of CCR5 Receptor Antagonists for the Treatment of HIV Infection: Hit-to-Lead Studies. *ChemMedChem*, 1, 706-709.
- ARONOV, A. M. 2005. Predictive in silico modeling for hERG channel blockers. *Drug Discovery Today*, 10, 149-155.
- ARROWSMITH, C. H., AUDIA, J. E., AUSTIN, C., BAELL, J., BENNETT, J., BLAGG, J., BOUNTRA, C., BRENNAN, P. E., BROWN, P. J. & BUNNAGE, M. E. 2015. The promise and peril of chemical probes. *Nature chemical biology*, 11, 536-541.

- ASHWOOD, V. A., FIELD, M. J., HORWELL, D. C., JULIEN-LAROSE, C., LEWTHWAITE, R. A., MCCLEARY, S., PRITCHARD, M. C., RAPHY, J. & SINGH, L. 2001. Utilization of an Intramolecular Hydrogen Bond To Increase the CNS Penetration of an NK1 Receptor Antagonist. *Journal of Medicinal Chemistry*, 44, 2276-2285.
- AUSTEL, V. 1983. Features and problems of practical drug design. *In: AUSTEL, V., BALABAN, A. T., BONCHEV, D., CHARTON, M., FUJITA, T., IWAMURA, H., MEKENYAN, O. & MOTOC, I. (eds.) Steric Effects in Drug Design.* Berlin, Heidelberg: Springer Berlin Heidelberg.
- BAHRAMI, G., MOHAMMADI, B. & KIANI, A. 2008. Determination of oseltamivir carboxylic acid in human serum by solid phase extraction and high performance liquid chromatography with UV detection. *Journal of Chromatography B*, 864, 38-42.
- BAILEY, D., ZANDERS, E. & DEAN, P. 2001. The end of the beginning for genomic medicine. *nature biotechnology*, 19, 207-208.
- BALDWIN, J. J., PONTICELLO, G. S., ANDERSON, P. S., CHRISTY, M. E., MURCKO, M. A., RANDALL, W. C., SCHWAM, H., SUGRUE, M. F., SPRINGER, J. & GAUTHERON, P. 1989. Thienothiopyran-2-sulfonamides: novel topically active carbonic anhydrase inhibitors for the treatment of glaucoma. *Journal of medicinal chemistry*, 32, 2510-2513.
- BANTIA, S., ARNOLD, C. S., PARKER, C. D., UPSHAW, R. & CHAND, P. 2006. Anti-influenza virus activity of peramivir in mice with single intramuscular injection. *Antiviral Research*, 69, 39-45.
- BARAGANA, B., HALLYBURTON, I., LEE, M. C., NORCROSS, N. R., GRIMALDI, R., OTTO, T. D., PROTO, W. R., BLAGBOROUGH, A. M., MEISTER, S., WIRJANATA, G., RUECKER, A., UPTON, L. M., ABRAHAM, T. S., ALMEIDA, M. J., PRADHAN, A., PORZELLE, A., MARTINEZ, M. S., BOLSCHER, J. M., WOODLAND, A., NORVAL, S., ZUCCOTTO, F., THOMAS, J., SIMEONS, F., STOJANOVSKI, L., OSUNA-CABELLO, M., BROCK, P. M., CHURCHER, T. S., SALA, K. A., ZAKUTANSKY, S. E., JIMENEZ-DIAZ, M. B., SANZ, L. M., RILEY, J., BASAK, R., CAMPBELL, M., AVERY, V. M., SAUERWEIN, R. W., DECHERING, K. J., NOVIYANTI, R., CAMPO, B., FREARSON, J. A., ANGULO-BARTUREN, I., FERRER-BAZAGA, S., GAMO, F. J., WYATT, P. G., LEROY, D., SIEGL, P., DELVES, M. J., KYLE, D. E., WITTLIN, S., MARFURT, J., PRICE, R. N., SINDEN, R. E., WINZELER, E. A., CHARMAN, S. A., BEBREVSKA, L., GRAY, D. W., CAMPBELL, S., FAIRLAMB, A. H., WILLIS, P. A., RAYNER, J. C., FIDOCK, D. A., READ, K. D. & GILBERT, I. H. 2015. A novel multiple-stage antimalarial agent that inhibits protein synthesis. *Nature*, 522, 315-20.
- BEAUMIER, C. M., GILLESPIE, P. M., STRYCH, U., HAYWARD, T., HOTEZ, P. J. & BOTTAZZI, M. E. 2016. Status of vaccine research and development of vaccines for Chagas disease. *Vaccine*, 34, 2996-3000.
- BEHERA, S. K., PRAHARAJ, A. B., DEHURY, B. & NEGI, S. 2015. Exploring the role and diversity of mucins in health and disease with special insight into non-communicable diseases. *Glycoconjugate journal*, 32, 575-613.
- BENNETT, E. P., MANDEL, U., CLAUSEN, H., GERKEN, T. A., FRITZ, T. A. & TABAK, L. A. 2012. Control of mucin-type O-glycosylation: A classification of the polypeptide GalNAc-transferase gene family. *Glycobiology*, 22, 736-756.
- BENVENUTI, M. & MANGANI, S. 2007. Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. *Nature protocols*, 2, 1633-1651.
- BERGFORS, T. 2003. Seeds to crystals. *Journal of structural biology*, 142, 66-76.
- BERMAN, H. M., WESTBROOK, J., FENG, Z., GILLILAND, G., BHAT, T. N., WEISSIG, H., SHINDYALOV, I. N. & BOURNE, P. E. 2000. The protein data bank. *Nucleic acids research*, 28, 235-242.
- BERMUDEZ, J., DAVIES, C., SIMONAZZI, A., REAL, J. P. & PALMA, S. 2016. Current drug therapy and pharmaceutical challenges for Chagas disease. *Acta tropica*, 156, 1-16.
- BERRY, A. M., LOCK, R. A. & PATON, J. C. 1996. Cloning and characterization of nanB, a second *Streptococcus pneumoniae* neuraminidase gene, and purification of the NanB enzyme from recombinant *Escherichia coli*. *Journal of bacteriology*, 178, 4854-4860.
- BIERE, B., BAUER, B. & SCHWEIGER, B. 2010. Differentiation of influenza B virus lineages Yamagata and Victoria by real-time PCR. *Journal of clinical microbiology*, 48, 1425-1427.

- BISSANTZ, C., KUHN, B. & STAHL, M. 2010. A Medicinal Chemist's Guide to Molecular Interactions. *Journal of Medicinal Chemistry*, 53, 5061-5084.
- BLAIR, J. M., WEBBER, M. A., BAYLAY, A. J., OGBOLU, D. O. & PIDDOCK, L. J. 2015. Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology*, 13, 42-51.
- BOSSART-WHITAKER, P., CARSON, M., BABU, Y., SMITH, C., LAVER, W. & AIR, G. M. 1993. Three-dimensional structure of influenza A N9 neuraminidase and its complex with the inhibitor 2-deoxy 2, 3-dehydro-N-acetyl neuraminic acid. *Journal of molecular biology*, 232, 1069-1083.
- BREAR, P. 2012. *Thesis*. University of St Andrews.
- BREAR, P., TELFORD, J., TAYLOR, G. L. & WESTWOOD, N. J. 2012. Synthesis and Structural Characterisation of Selective Non-Carbohydrate-Based Inhibitors of Bacterial Sialidases. *Chembiochem*, 13, 2374-2383.
- BREITLING, J. & AEBI, M. 2013. N-Linked Protein Glycosylation in the Endoplasmic Reticulum. *Cold Spring Harbor Perspectives in Biology*, 5, a013359.
- BRITTAN, J., BUCKERIDGE, T., FINN, A., KADIOGLU, A. & JENKINSON, H. 2012. Pneumococcal neuraminidase A: an essential upper airway colonization factor for *Streptococcus pneumoniae*. *Molecular oral microbiology*, 27, 270-283.
- BROCKHAUSEN, I., SCHACHTER, H. & STANLEY, P. 2009. O-GalNAc glycans.
- BRODER, S. & VENTER, J. C. 2000. Sequencing the entire genomes of free-living organisms: the foundation of pharmacology in the new millennium. *Annual Review of Pharmacology and Toxicology*, 40, 97-132.
- BROOIJMANS, N. & KUNTZ, I. D. 2003. Molecular recognition and docking algorithms. *Annual review of biophysics and biomolecular structure*, 32, 335-373.
- BUCCI, M., GOODMAN, C. & SHEPPARD, T. L. 2010. A decade of chemical biology. *Nature chemical biology*, 6, 847-854.
- BUCHINI, S., BUSCHIAZZO, A. & WITHERS, S. G. 2008. A New Generation of Specific *Trypanosoma cruzi* trans-Sialidase Inhibitors. *Angewandte Chemie International Edition*, 47, 2700-2703.
- BUNNAGE, M. E., CHEKLER, E. L. P. & JONES, L. H. 2013. Target validation using chemical probes. *Nature chemical biology*, 9, 195-199.
- BUNNAGE, M. E., GILBERT, A. M., JONES, L. H. & HETT, E. C. 2015. Know your target, know your molecule. *Nature chemical biology*, 11, 368-372.
- BUSCHIAZZO, A. & ALZARI, P. M. 2008a. Structural insights into sialic acid enzymology. *Current opinion in chemical biology*, 12, 565-572.
- BUSCHIAZZO, A. & ALZARI, P. M. 2008b. Structural insights into sialic acid enzymology. *Curr Opin Chem Biol*, 12, 565-72.
- BUSCHIAZZO, A., AMAYA, M. A. F., CREMONA, M. A. L., FRASCH, A. C. & ALZARI, P. M. 2002. The crystal structure and mode of action of trans-sialidase, a key enzyme in *Trypanosoma cruzi* pathogenesis. *Molecular cell*, 10, 757-768.
- BUSCHIAZZO, A., TAVARES, G. A., CAMPETELLA, O., SPINELLI, S., CREMONA, M. L., PARÍS, G., AMAYA, M. F., FRASCH, A. C. & ALZARI, P. M. 2000. Structural basis of sialyltransferase activity in trypanosomal sialidases. *The EMBO Journal*, 19, 16-24.
- BUSSE, H., HAKODA, M., STANLEY, M. & STREICHER, H. 2007. Galactose-Phosphonates as Mimetics of the Sialyltransfer by Trypanosomal Sialidases. *Journal of Carbohydrate Chemistry*, 26, 159-194.
- CAMPBELL, S. F. 2000. Science, art and drug discovery: a personal perspective. *Clinical Science*, 99, 255.
- CANALS, M., SEXTON, P. M. & CHRISTOPOULOS, A. 2011. Allosterism in GPCRs: 'MWC' revisited. *Trends in biochemical sciences*, 36, 663-672.
- CAPDEVILLE, R., BUCHDUNGER, E., ZIMMERMANN, J. & MATTER, A. 2002. Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nature reviews Drug discovery*, 1, 493-502.
- CARR, R. A., CONGREVE, M., MURRAY, C. W. & REES, D. C. 2005. Fragment-based lead discovery: leads by design. *Drug discovery today*, 10, 987-992.

- CARREA, A. & DIAMBRA, L. 2016. Systems Biology Approach to Model the Life Cycle of Trypanosoma cruzi. *PLoS one*, 11, e0146947.
- CASTILLO, R., HOLLAND, L. & BOLTZ, D. 2010. Peramivir and its use in H1N1 influenza. *Drugs of today (Barcelona, Spain: 1998)*, 46, 399-408.
- CAZY. CAZY Classification [Online]. Available: <http://www.cazy.org/> [Accessed].
- CDC. 2014. Influenza (Flu) [Online]. Available: <http://www.cdc.gov/flu/about/viruses/types.htm> [Accessed].
- CDC. 2016. Chapter 3 Infectious Diseases Related to Travel: Pneumococcal Disease [Online]. Available: <http://wwwnc.cdc.gov/travel/yellowbook/2016/infectious-diseases-related-to-travel/pneumococcal-disease> [Accessed].
- CHAN, J. & BENNET, A. J. 2012. Enzymology of influenza virus sialidase. *Influenza Virus Sialidase-A Drug Discovery Target*. Springer.
- CHAN, J., WATSON, J. N., LU, A., CERDA, V. C., BORGFORD, T. J. & BENNET, A. J. 2011. Bacterial and viral sialidases: contribution of the conserved active site glutamate to catalysis. *Biochemistry*, 51, 433-441.
- CHAND, P., BANTIA, S., KOTIAN, P. L., EL-KATTAN, Y., LIN, T.-H. & BABU, Y. S. 2005. Comparison of the anti-influenza virus activity of cyclopentane derivatives with oseltamivir and zanamivir in vivo. *Bioorganic & Medicinal Chemistry*, 13, 4071-4077.
- CHANDA, S. K. & CALDWELL, J. S. 2003. Fulfilling the promise: drug discovery in the post-genomic era. *Drug discovery today*, 8, 168-174.
- CHANG, G., GUIDA, W. C. & STILL, W. C. 1989. An internal-coordinate Monte Carlo method for searching conformational space. *Journal of the American Chemical Society*, 111, 4379-4386.
- CHANG, Y.-C., UCHIYAMA, S., VARKI, A. & NIZET, V. 2012. Leukocyte inflammatory responses provoked by pneumococcal sialidase. *MBio*, 3, e00220-11.
- CHANGEUX, J.-P. 2012. Allostery and the Monod-Wyman-Changeux model after 50 years. *Annual review of biophysics*, 41, 103-133.
- CHANGEUX, J.-P. 2013. 50 years of allosteric interactions: the twists and turns of the models. *Nature reviews Molecular cell biology*, 14, 819-829.
- CHARIFSON, P. S., CORKERY, J. J., MURCKO, M. A. & WALTERS, W. P. 1999. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. *Journal of Medicinal Chemistry*, 42, 5100-5109.
- CHARIFSON, P. S. & KUNTZ, I. D. 1997. *Recent successes and continuing limitations in computer-aided drug design*, Dekker: New York.
- CHAVAS, L. M., TRINGALI, C., FUSI, P., VENERANDO, B., TETTAMANTI, G., KATO, R., MONTI, E. & WAKATSUKI, S. 2005. Crystal Structure of the Human Cytosolic Sialidase Neu2 EVIDENCE FOR THE DYNAMIC NATURE OF SUBSTRATE RECOGNITION. *Journal of Biological Chemistry*, 280, 469-475.
- CHAYAMA, K., IMAMURA, M. & HAYES, C. N. 2016. Hepatitis C virus treatment update—A new era of all-oral HCV treatment. *Advances in Digestive Medicine*.
- CHEN, A. S. Y., WESTWOOD, N. J., BREAR, P., ROGERS, G. W., MAVRIDIS, L. & MITCHELL, J. B. 2016. A random forest model for predicting allosteric and functional sites on proteins. *Molecular Informatics*.
- CHEN, Z. & WEBER, S. G. 2007. A High-Throughput Method for Lipophilicity Measurement. *Analytical chemistry*, 79, 1043-1049.
- CHEKASOV, A., MURATOV, E. N., FOURCHES, D., VARNEK, A., BASKIN, I. I., CRONIN, M., DEARDEN, J., GRAMATICA, P., MARTIN, Y. C., TODESCHINI, R., CONSONNI, V., KUZ'MIN, V. E., CRAMER, R., BENIGNI, R., YANG, C., RATHMAN, J., TERFLOTH, L., GASTEIGER, J., RICHARD, A. & TROPSHA, A. 2014. QSAR Modeling: Where have you been? Where are you going to? *Journal of medicinal chemistry*, 57, 4977-5010.

- CHLANDA, P., SCHRAIDT, O., KUMMER, S., RICHES, J., OBERWINKLER, H., PRINZ, S., KRÄUSSLICH, H.-G. & BRIGGS, J. A. 2015. Structural analysis of the roles of influenza A virus membrane-associated proteins in assembly and morphology. *Journal of virology*, 89, 8957-8966.
- CHOFFNES, E. R. & RELMAN, D. A. 2011. *The Causes and Impacts of Neglected Tropical and Zoonotic Diseases: Opportunities for Integrated Intervention Strategies: Workshop Summary*, National Academies Press.
- CHOI, H., KIM, J.-Y., CHANG, Y. T. & NAM, H. G. 2014. Forward Chemical Genetic Screening. In: SANCHEZ-SERRANO, J. J. & SALINAS, J. (eds.) *Arabidopsis Protocols*. Totowa, NJ: Humana Press.
- CHOKSHI, D. A. 2006. Improving access to medicines in poor countries: The role of universities. *PLoS Med*, 3, e136.
- CHRISTOPOULOS, A. 2002. Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. *Nature Reviews Drug Discovery*, 1, 198-210.
- CHUENKOVA, M. & PEREIRA, M. 1995. Trypanosoma cruzi trans-sialidase: enhancement of virulence in a murine model of Chagas' disease. *The Journal of experimental medicine*, 181, 1693-1703.
- CLADER, J. W. 2004. The discovery of ezetimibe: a view from outside the receptor. *Journal of medicinal chemistry*, 47, 1-9.
- CLARDY, J., FISCHBACH, M. A. & WALSH, C. T. 2006. New antibiotics from bacterial natural products. *Nature biotechnology*, 24, 1541-1550.
- CLARKE, A. J. & COOPER, D. N. 2010. GWAS: heritability missing in action? *European Journal of Human Genetics*, 18, 859.
- COATS, M. T. 2016. Combating Malaria: Where do We Stand? *Journal of infectious disease and therapy*, 4.
- COHEN, M. & VARKI, A. 2010. The sialome—far more than the sum of its parts. *Omic: a journal of integrative biology*, 14, 455-464.
- COHEN, M. & VARKI, A. 2014. Modulation of glycan recognition by clustered saccharide patches. *Int. Rev. Cell Mol. Biol*, 308, 75-125.
- COHEN, M. L. 2000. Changing patterns of infectious disease. *Nature*, 406, 762-767.
- COLLI, W. 1993. Trans-sialidase: a unique enzyme activity discovered in the protozoan Trypanosoma cruzi. *The FASEB Journal*, 7, 1257-1264.
- COMINI, M. A. & FLOHÉ, L. 2013. Trypanothione-Based Redox Metabolism of Trypanosomatids. *Trypanosomatid Diseases*. Wiley-VCH Verlag GmbH & Co. KGaA.
- CONN, P. J., CHRISTOPOULOS, A. & LINDSLEY, C. W. 2009. Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders. *Nature reviews Drug discovery*, 8, 41-54.
- CONNARIS, H., GOVORKOVA, E. A., LIGERTWOOD, Y., DUTIA, B. M., YANG, L., TAUBER, S., TAYLOR, M. A., ALIAS, N., HAGAN, R. & NASH, A. A. 2014. Prevention of influenza by targeting host receptors using engineered proteins. *Proceedings of the National Academy of Sciences*, 111, 6401-6406.
- COPLEY, R. R., RUSSELL, R. B. & PONTING, C. P. 2001. Sialidase-like Asp-boxes: sequence-similar structures within different protein folds. *Protein Science*, 10, 285-292.
- COUCH, R. B. 1996. Orthomyxoviruses.
- COURA, J. R. & BORGES-PEREIRA, J. 2010. Chagas disease: 100 years after its discovery. A systemic review. *Acta tropica*, 115, 5-13.
- CRAIN, M., WALTMAN, W., TURNER, J., YOTHER, J., TALKINGTON, D., MCDANIEL, L., GRAY, B. M. & BRILES, D. 1990. Pneumococcal surface protein A (PspA) is serologically highly variable and is expressed by all clinically important capsular serotypes of Streptococcus pneumoniae. *Infection and immunity*, 58, 3293-3299.
- CRENNELL, S. J., GARMAN, E. F., PHILIPPON, C., VASELLA, A., LAVER, G. W., VIMR, E. R. & TAYLOR, G. L. 1996. The Structures of Salmonella typhimuriumLT2 Neuraminidase and its Complexes with Three Inhibitors at High Resolution. *Journal of molecular biology*, 259, 264-280.
- CUI, Q. & KARPLUS, M. 2008. Allostery and cooperativity revisited. *Protein science*, 17, 1295-1307.

- CUMMING, J. G., DAVIS, A. M., MURESAN, S., HAEBERLEIN, M. & CHEN, H. 2013. Chemical predictive modelling to improve compound quality. *Nat Rev Drug Discov*, 12, 948-962.
- CUSHMAN, D. W., CHEUNG, H., SABO, E. & ONDETTI, M. 1977. Design of potent competitive inhibitors of angiotensin-converting enzyme. Carboxyalkanoyl and mercaptoalkanoyl amino acids. *Biochemistry*, 16, 5484-5491.
- DALKAS, G. A., VLACHAKIS, D., TSAGKRASOULIS, D., KASTANIA, A. & KOSSIDA, S. 2012. State-of-the-art technology in modern computer-aided drug design. *Briefings in bioinformatics*, bbs063.
- DALL'OLIO, F. & CHIRICOLO, M. 2001. Sialyltransferases in cancer. *Glycoconjugate journal*, 18, 841-850.
- DAVIES, G. & HENRISSAT, B. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure*, 3, 853-859.
- DAVIES, J. & DAVIES, D. 2010. Origins and evolution of antibiotic resistance. *Microbiology and Molecular Biology Reviews*, 74, 417-433.
- DAVIS, A. & WARD, S. E. 2014. *The handbook of medicinal chemistry: principles and practice*, Royal Society of Chemistry.
- DAVIS, A. M., TEAGUE, S. J. & KLEYWEGT, G. J. 2003. Application and Limitations of X-ray Crystallographic Data in Structure-Based Ligand and Drug Design. *Angewandte Chemie International Edition*, 42, 2718-2736.
- DE BEER, S., VERMEULEN, N. P. & OOSTENBRINK, C. 2010. The role of water molecules in computational drug design. *Current topics in medicinal chemistry*, 10, 55-66.
- DE CLERCQ, E. 2013. Antivirals: Past, present and future. *Biochemical Pharmacology*, 85, 727-744.
- DEMIR, Ö. & ROITBERG, A. E. 2009. Modulation of Catalytic Function by Differential Plasticity of the Active Site: Case Study of Trypanosoma cruzi trans-Sialidase and Trypanosoma rangeli Sialidase. *Biochemistry*, 48, 3398-3406.
- DENNIS, M. S., EIGENBROT, C., SKELTON, N. J., ULTSCH, M. H., SANTELL, L., DWYER, M. A., O'CONNELL, M. P. & LAZARUS, R. A. 2000. Peptide exosite inhibitors of factor VIIa as anticoagulants. *Nature*, 404, 465-470.
- DESSAU, M. A. & MODIS, Y. 2011. Protein crystallization for X-ray crystallography. *JoVE (Journal of Visualized Experiments)*, e2285-e2285.
- DEVITA, V. T. & CHU, E. 2008. A history of cancer chemotherapy. *Cancer research*, 68, 8643-8653.
- DONG, G., PENG, C., LUO, J., WANG, C., HAN, L., WU, B., JI, G. & HE, H. 2015. Adamantane-Resistant Influenza A Viruses in the World (1902–2013): Frequency and Distribution of M2 Gene Mutations. *PLoS ONE*, 10, e0119115.
- DREWS, J. 1996. Genomic sciences and the medicine of tomorrow. *Nature biotechnology*, 14, 1516-1518.
- DREWS, J. & RYSER, S. 1997. Classic drug targets. *Nature Biotechnology*, 15, 1350-1350.
- DU, X., LI, Y., XIA, Y.-L., AI, S.-M., LIANG, J., SANG, P., JI, X.-L. & LIU, S.-Q. 2016. Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. *International Journal of Molecular Sciences*, 17, 144.
- EDMOND, J., JOHNSTON, R., KIDD, D., RYLANCE, H. & SOMMERVILLE, R. 1966. The inhibition of neuraminidase and antiviral action. *British journal of pharmacology and chemotherapy*, 27, 415-426.
- ENGLAND, P. H. 2016. Dichloromethane: incident management.
- ENGLISH, A. C., GROOM, C. R. & HUBBARD, R. E. 2001. Experimental and computational mapping of the binding surface of a crystalline protein. *Protein engineering*, 14, 47-59.
- ENGSTLER, M., REUTER, G. & SCHAUER, R. 1992. Purification and characterization of a novel sialidase found in procyclic culture forms of Trypanosoma brucei. *Molecular and biochemical parasitology*, 54, 21-30.
- ESPRIPT. *ESPrIPT* [Online]. Available: <http://espript.ibcp.fr> [Accessed].
- FAN, T.-J., HAN, L.-H., CONG, R.-S. & LIANG, J. 2005. Caspase family proteases and apoptosis. *Acta biochimica et biophysica Sinica*, 37, 719-727.

- FANZANI, A., ZANOLA, A., FAGGI, F., PAPINI, N., VENERANDO, B., TETTAMANTI, G., SAMPAOLESI, M. & MONTI, E. 2012. Implications for the mammalian sialidases in the physiopathology of skeletal muscle. *Skeletal Muscle*, 2, 23-23.
- FAUBER, B. P., RENÉ, O., DE LEON BOENIG, G., BURTON, B., DENG, Y., EIDENSCHENK, C., EVERETT, C., GOBBI, A., HYMOWITZ, S. G. & JOHNSON, A. R. 2014. Reduction in lipophilicity improved the solubility, plasma–protein binding, and permeability of tertiary sulfonamide RORc inverse agonists. *Bioorganic & medicinal chemistry letters*, 24, 3891-3897.
- FDA. 2016a. *Developing Products for Rare Diseases and Conditions* [Online]. Available: <http://www.fda.gov/ForIndustry/DevelopingProductsforRareDiseasesConditions/ucm2005525.htm> [Accessed].
- FDA 2016b. Novel Drugs 2015 Summary.
- FDA. 2016c. *Orphan Drug Act - Relevant Excerpts* [Online]. Available: <http://www.fda.gov/ForIndustry/DevelopingProductsforRareDiseasesConditions/HowtoapplyforOrphanProductDesignation/ucm364750.htm> [Accessed].
- FEASEY, N., WANSBROUGH-JONES, M., MABEY, D. C. & SOLOMON, A. W. 2010. Neglected tropical diseases. *British medical bulletin*, 93, 179-200.
- FELDMANN, H., JONES, S., KLENK, H.-D. & SCHNITTLER, H.-J. 2003. Ebola virus: from discovery to vaccine. *Nature Reviews Immunology*, 3, 677-685.
- FENG, E., YE, D., LI, J., ZHANG, D., WANG, J., ZHAO, F., HILGENFELD, R., ZHENG, M., JIANG, H. & LIU, H. 2012. Recent Advances in Neuraminidase Inhibitor Development as Anti-influenza Drugs. *ChemMedChem*, 7, 1527-1536.
- FENLEY, A. T., MUDDANA, H. S. & GILSON, M. K. 2012. Entropy–enthalpy transduction caused by conformational shifts can obscure the forces driving protein–ligand binding. *Proceedings of the National Academy of Sciences*, 109, 20006-20011.
- FERMINI, B. & FOSSA, A. A. 2003. The impact of drug-induced QT interval prolongation on drug discovery and development. *Nat Rev Drug Discov*, 2, 439-447.
- FERREIRA, L. G., DOS SANTOS, R. N., OLIVA, G. & ANDRICOPULO, A. D. 2015. Molecular docking and structure-based drug design strategies. *Molecules*, 20, 13384-13421.
- FILIPPAKOPOULOS, P., QI, J., PICAUD, S., SHEN, Y., SMITH, W. B., FEDOROV, O., MORSE, E. M., KEATES, T., HICKMAN, T. T. & FELLETAR, I. 2010. Selective inhibition of BET bromodomains. *Nature*, 468, 1067-1073.
- FLOWER, D. R. 1998. On the properties of bit string-based measures of chemical similarity. *Journal of Chemical Information and Computer Sciences*, 38, 379-386.
- FORUM, W. E. 2016. Insight Report: The Global Risks Report 2016 11th Edition.
- FRALISH, B. H. & TARLETON, R. L. 2003. Genetic immunization with LYT1 or a pool of trans-sialidase genes protects mice from lethal *Trypanosoma cruzi* infection. *Vaccine*, 21, 3070-3080.
- FREARSON, J. A. & COLLIE, I. T. 2009. HTS and hit finding in academia—from chemical genomics to drug discovery. *Drug discovery today*, 14, 1150-1158.
- FREIRE-DE-LIMA, L., FONSECA, L., OELTMANN, T., MENDONÇA-PREVIATO, L. & PREVIATO, J. 2015. The trans-sialidase, the major *Trypanosoma cruzi* virulence factor: three decades of studies. *Glycobiology*, cww057.
- FREIRE, E. 2008. Do enthalpy and entropy distinguish first in class from best in class? *Drug discovery today*, 13, 869-874.
- FRIED, M. W., SHIFFMAN, M. L., REDDY, K. R., SMITH, C., MARINOS, G., GONÇALES, F. L. J., HÄUSSINGER, D., DIAGO, M., CAROSI, G., DHUMEAUX, D., CRAXI, A., LIN, A., HOFFMAN, J. & YU, J. 2002. Peginterferon Alfa-2a plus Ribavirin for Chronic Hepatitis C Virus Infection. *New England Journal of Medicine*, 347, 975-982.
- FUGUET, E., RETA, M., GIBERT, C., ROSÉS, M., BOSCH, E. & RÀFOLS, C. 2008. Critical evaluation of buffering solutions for pKa determination by capillary electrophoresis. *Electrophoresis*, 29, 2841-2851.

- FUTAMURA, Y., MUROI, M. & OSADA, H. 2013. Target identification of small molecules based on chemical biology approaches. *Mol Biosyst*, 9, 897-914.
- GAO, M., NETTLES, R. E., BELEMA, M., SNYDER, L. B., NGUYEN, V. N., FRIDELL, R. A., SERRANO-WU, M. H., LANGLEY, D. R., SUN, J. H., O'BOYLE, D. R., 2ND, LEMM, J. A., WANG, C., KNIPE, J. O., CHIEN, C., COLONNO, R. J., GRASELA, D. M., MEANWELL, N. A. & HAMANN, L. G. 2010. Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect. *Nature*, 465, 96-100.
- GASKELL, A., CRENNELL, S. & TAYLOR, G. 1995. The three domains of a bacterial sialidase: a  $\beta$ -propeller, an immunoglobulin module and a galactose-binding jelly-roll. *Structure*, 3, 1197-1205.
- GERSHELL, L. J. & ATKINS, J. H. 2003. A brief history of novel drug discovery technologies. *Nature Reviews Drug Discovery*, 2, 321-327.
- GIACOPUZZI, E., BRESCIANI, R., SCHAUER, R., MONTI, E. & BORSANI, G. 2012. New insights on the sialidase protein family revealed by a phylogenetic analysis in metazoa. *PLoS one*, 7, e44193.
- GILBERT, I. H. 2013. Drug discovery for neglected diseases: molecular target-based and phenotypic approaches: miniperspectives series on phenotypic screening for anti-infective targets. *Journal of medicinal chemistry*, 56, 7719-7726.
- GLEESON, M. P., LEESON, P. D. & VAN DE WATERBEEMD, H. 2015. Physicochemical properties and compound quality. *The handbook of medicinal chemistry: principles and practice*, The Royal Society of Chemistry, Cambridge, UK, 1-31.
- GOETZ, T., ARSLAN, A., WIDEN, W. & WULFF, P. 2007. GABA A receptors: structure and function in the basal ganglia. *Progress in brain research*, 160, 21-41.
- GOHLKE, H., HENDLICH, M. & KLEBE, G. 2000. Knowledge-based scoring function to predict protein-ligand interactions. *Journal of molecular biology*, 295, 337-356.
- GOHLKE, H., SCHLIEPER, D. & GROTH, G. 2012. Resolving the negative potential side (n-side) water-accessible proton pathway of F-type ATP synthase by molecular dynamics simulations. *Journal of Biological Chemistry*, 287, 36536-36543.
- GOOD, A. C., KRISTEK, S. R. & MASON, J. S. 2000. High-throughput and virtual screening: core lead discovery technologies move towards integration. *Drug Discovery Today*, 5, 61-69.
- GOODFORD, P. J. 1985. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *Journal of medicinal chemistry*, 28, 849-857.
- GREENFIELD, N. J. 2004. Circular dichroism analysis for protein-protein interactions. *Protein-Protein Interactions: Methods and Applications*, 55-77.
- GRUBBS, R. H. 2006. Olefin-metathesis catalysts for the preparation of molecules and materials (Nobel lecture). *Angewandte Chemie International Edition*, 45, 3760-3765.
- GRUBBS, R. H., MILLER, S. J. & FU, G. C. 1995. Ring-closing metathesis and related processes in organic synthesis. *Accounts of chemical research*, 28, 446-452.
- GUNASEKARAN, K., MA, B. & NUSSINOV, R. 2004. Is allostery an intrinsic property of all dynamic proteins? *Proteins: Structure, Function, and Bioinformatics*, 57, 433-443.
- GUT, H., KING, S. J. & WALSH, M. A. 2008. Structural and functional studies of Streptococcus pneumoniae neuraminidase B: An intramolecular trans-sialidase. *FEBS Letters*, 582, 3348-3352.
- GUT, H., XU, G., TAYLOR, G. L. & WALSH, M. A. 2011. Structural basis for Streptococcus pneumoniae NanA inhibition by influenza antivirals zanamivir and oseltamivir carboxylate. *Journal of molecular biology*, 409, 496-503.
- HAKOMORI, S. 2004. Glycosynapses: microdomains controlling carbohydrate-dependent cell adhesion and signaling. *Anais da Academia Brasileira de Ciências*, 76, 553-572.
- HANAHAH, D. & WEINBERG, R. A. 2011. Hallmarks of cancer: the next generation. *cell*, 144, 646-674.
- HANN, M. M. & KESERÜ, G. M. 2012. Finding the sweet spot: the role of nature and nurture in medicinal chemistry. *Nat Rev Drug Discov*, 11, 355-365.

- HARDY, J. A., LAM, J., NGUYEN, J. T., O'BRIEN, T. & WELLS, J. A. 2004. Discovery of an allosteric site in the caspases. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 12461-12466.
- HARDY, J. A. & WELLS, J. A. 2004. Searching for new allosteric sites in enzymes. *Current opinion in structural biology*, 14, 706-715.
- HASHEMI, S., NASROLLAH, A. & RAJABI, M. 2013. Irrational antibiotic prescribing: a local issue or global concern? *EXCLI journal*, 12, 384.
- HAYDEN, F. G., OSTERHAUS, A. D., TREANOR, J. J., FLEMING, D. M., AOKI, F. Y., NICHOLSON, K. G., BOHNEN, A. M., HIRST, H. M., KEENE, O. & WIGHTMAN, K. 1997. Efficacy and safety of the neuraminidase inhibitor zanamivir in the treatment of influenza virus infections. *New England Journal of Medicine*, 337, 874-880.
- HEATH, R. J., RUBIN, J. R., HOLLAND, D. R., ZHANG, E., SNOW, M. E. & ROCK, C. O. 1999. Mechanism of triclosan inhibition of bacterial fatty acid synthesis. *Journal of Biological Chemistry*, 274, 11110-11114.
- HEFTI, F. F. 2008. Requirements for a lead compound to become a clinical candidate. *BMC Neuroscience*, 9, S7.
- HEIFETZ, A., BARKER, O., VERQUIN, G., WIMMER, N., MEUTERMANS, W., PAL, S., LAW, R. J. & WHITTAKER, M. 2013. Fighting Obesity with a Sugar-Based Library: Discovery of Novel MCH-1R Antagonists by a New Computational-VAST Approach for Exploration of GPCR Binding Sites. *Journal of Chemical Information and Modeling*, 53, 1084-1099.
- HENTRICH, K., LÖFLING, J., PATHAK, A., NIZET, V., VARKI, A. & HENRIQUES-NORMARK, B. Streptococcus pneumoniae Senses a Human-like Sialic Acid Profile via the Response Regulator CiaR. *Cell Host & Microbe*.
- HERNANDEZ, J. E., ADIGA, R., ARMSTRONG, R., BAZAN, J., BONILLA, H., BRADLEY, J., DRETHER, R., ISON, M. G., MANGINO, J. E., MAROUSHEK, S., SHETTY, A. K., WALD, A., ZIEBOLD, C., ELDER, J., HOLLISTER, A. S., SHERIDAN, W. & ON BEHALF OF THE E, I. N. D. P. I. 2011. Clinical Experience in Adults and Children Treated with Intravenous Peramivir for 2009 Influenza A (H1N1) Under an Emergency IND Program in the United States. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 52, 695-706.
- HIOUAL, K., CHIKHI, A., BENSEGUENI, A., MERZOUG, A. & BOUCHERIT, H. 2012. Comparative data on docking algorithms: Keeping the update in the field knowledge. *Int J Applied Information Systems*, 2, 2249-0868.
- HIRST, R. A., GOSAI, B., RUTMAN, A., GUERIN, C. J., NICOTERA, P., ANDREW, P. W. & O'CALLAGHAN, C. 2008. Streptococcus pneumoniae deficient in pneumolysin or autolysin has reduced virulence in meningitis. *Journal of Infectious Diseases*, 197, 744-751.
- HOPKINS, A. L. & GROOM, C. R. 2002. The druggable genome. *Nature reviews Drug discovery*, 1, 727-730.
- HOPKINS, A. L., KESERU, G. M., LEESON, P. D., REES, D. C. & REYNOLDS, C. H. 2014. The role of ligand efficiency metrics in drug discovery. *Nat Rev Drug Discov*, 13, 105-121.
- HORN, J. R. & SHOICHET, B. K. 2004. Allosteric inhibition through core disruption. *Journal of molecular biology*, 336, 1283-1291.
- HORTON JR, A. M., NOGGLE, C. A. & DEAN, R. S. 2011. *The encyclopedia of neuropsychological disorders*, Springer Publishing Company.
- HUANG, S.-Y. & ZOU, X. 2010. Advances and Challenges in Protein-Ligand Docking. *International Journal of Molecular Sciences*, 11, 3016-3034.
- HUANG, S. S., JOHNSON, K. M., RAY, G. T., WROE, P., LIEU, T. A., MOORE, M. R., ZELL, E. R., LINDER, J. A., GRIJALVA, C. G. & METLAY, J. P. 2011. Healthcare utilization and cost of pneumococcal disease in the United States. *Vaccine*, 29, 3398-3412.
- HUGHES, J., REES, S., KALINDJIAN, S. & PHILPOTT, K. 2011. Principles of early drug discovery. *British journal of pharmacology*, 162, 1239-1249.

- IMMING, P., SINNING, C. & MEYER, A. 2006. Drugs, their targets and the nature and number of drug targets. *Nature reviews Drug discovery*, 5, 821-834.
- ISON, M. G. 2011. Antivirals and resistance: influenza virus. *Current opinion in virology*, 1, 563-573.
- JAIN, K. K. 2000. An assessment of rufinamide as an anti-epileptic in comparison with other drugs in clinical development. *Expert opinion on investigational drugs*, 9, 829-840.
- JANEWAY JR, C. A., TRAVERS, P., WALPORT, M. & SHLOMCHIK, M. J. 2001. The front line of host defense.
- JEDRZEJAS, M. J. 2001. Pneumococcal virulence factors: structure and function. *Microbiology and Molecular Biology Reviews*, 65, 187-207.
- JONES, G., WILLETT, P., GLEN, R. C., LEACH, A. R. & TAYLOR, R. 1997. Development and validation of a genetic algorithm for flexible docking. *Journal of molecular biology*, 267, 727-748.
- JONES, K. E., PATEL, N. G., LEVY, M. A., STOREYGARD, A., BALK, D., GITTLEMAN, J. L. & DASZAK, P. 2008. Global trends in emerging infectious diseases. *Nature*, 451, 990-993.
- JORGE, S. D., PALACE-BERL, F., MASUNARI, A., CECHINEL, C. A., ISHII, M., PASQUALOTO, K. F. M. & TAVARES, L. C. 2011. Novel benzofuroxan derivatives against multidrug-resistant *Staphylococcus aureus* strains: Design using Topliss' decision tree, synthesis and biological assay. *Bioorganic & Medicinal Chemistry*, 19, 5031-5038.
- JUGE, N., TAILFORD, L. & OWEN, C. D. 2016. Sialidases from gut bacteria: a mini-review. *Biochemical Society Transactions*, 44, 166-175.
- JUMAT, M. R., SUGRUE, R. J. & TAN, B.-H. 2014. Genetic characterisation of influenza B viruses detected in Singapore, 2004 to 2009. *BMC research notes*, 7, 1.
- KADIOGLU, A., WEISER, J. N., PATON, J. C. & ANDREW, P. W. 2008. The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nature Reviews Microbiology*, 6, 288-301.
- KAIDA, K., KANZAKI, M., MORITA, D., KAMAKURA, K., MOTOYOSHI, K., HIRAKAWA, M. & KUSUNOKI, S. 2006. Anti-ganglioside complex antibodies in Miller Fisher syndrome. *Journal of Neurology, Neurosurgery, and Psychiatry*, 77, 1043-1046.
- KAPETANOVIC, I. 2008. Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach. *Chemico-biological interactions*, 171, 165-176.
- KAWASUMI, M. & NGHIEM, P. 2007. Chemical genetics: elucidating biological systems with small-molecule compounds. *Journal of Investigative Dermatology*, 127, 1577-1584.
- KELLENBERGER, E., RODRIGO, J., MULLER, P. & ROGNAN, D. 2004. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins: Structure, Function, and Bioinformatics*, 57, 225-242.
- KELLY, S. M., JESS, T. J. & PRICE, N. C. 2005. How to study proteins by circular dichroism. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1751, 119-139.
- KERNS, E. H. & DI, L. 2003. Pharmaceutical profiling in drug discovery. *Drug Discovery Today*, 8, 316-323.
- KESERÚ, G. R. M., ERLANSON, D. A., FERENCZY, G. R. G., HANN, M. M., MURRAY, C. W. & PICKETT, S. D. 2016. Design Principles for Fragment Libraries: Maximizing the Value of Learnings from Pharma Fragment-Based Drug Discovery (FBDD) Programs for Use in Academia. *Journal of medicinal chemistry*.
- KIM, S., OH, D.-B., KANG, H. A. & KWON, O. 2011. Features and applications of bacterial sialidases. *Applied microbiology and biotechnology*, 91, 1-15.
- KIRCHHOFF, L. V. & PEARSON, R. D. 2007. The emergence of Chagas disease in the United States and Canada. *Current infectious disease reports*, 9, 347-350.
- KIRCHMAIR, J., GOLLER, A. H., LANG, D., KUNZE, J., TESTA, B., WILSON, I. D., GLEN, R. C. & SCHNEIDER, G. 2015. Predicting drug metabolism: experiment and/or computation? *Nat Rev Drug Discov*, 14, 387-404.

- KIRCHMAIR, J., WILLIAMSON, M. J., TYZACK, J. D., TAN, L., BOND, P. J., BENDER, A. & GLEN, R. C. 2012. Computational prediction of metabolism: sites, products, SAR, P450 enzyme dynamics, and mechanisms. *Journal of chemical information and modeling*, 52, 617-648.
- KITCHEN, D. B., DECORNEZ, H., FURR, J. R. & BAJORATH, J. 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature reviews Drug discovery*, 3, 935-949.
- KLIBANOV, A. M. 1990. Asymmetric transformations catalyzed by enzymes in organic solvents. *Accounts of Chemical Research*, 23, 114-120.
- KLUGER, G., KURLEMANN, G., HABERLANDT, E., ERNST, J.-P., RUNGE, U., SCHNEIDER, F., MAKOWSKI, C., BOOR, R. & BAST, T. 2009. Effectiveness and tolerability of rufinamide in children and adults with refractory epilepsy: first European experience. *Epilepsy & Behavior*, 14, 491-495.
- KNIGHT, J. D., HAMELBERG, D., MCCAMMON, J. A. & KOTHARY, R. 2009. The role of conserved water molecules in the catalytic domain of protein kinases. *Proteins: Structure, Function, and Bioinformatics*, 76, 527-535.
- KOGEJ, T., BLOMBERG, N., GREASLEY, P. J., MUNDT, S., VAINIO, M. J., SCHAMBERGER, J., SCHMIDT, G. & HÜSER, J. 2013. Big pharma screening collections: more of the same or unique libraries? The AstraZeneca–Bayer Pharma AG case. *Drug discovery today*, 18, 1014-1024.
- KOHANSKI, M. A., DWYER, D. J., HAYETE, B., LAWRENCE, C. A. & COLLINS, J. J. 2007. A common mechanism of cellular death induced by bactericidal antibiotics. *Cell*, 130, 797-810.
- KONGKAMNERD, J., MILANI, A., CATTOLI, G., TERREGINO, C., CAPUA, I., BENEDUCE, L., GALLOTTA, A., PENGO, P., FASSINA, G. & MONTHAKANTIRAT, O. 2011. The quenching effect of flavonoids on 4-methylumbelliferone, a potential pitfall in fluorimetric neuraminidase inhibition assays. *Journal of Biomolecular screening*, 16, 755-764.
- KOOPMANS, R. 2009. *Advances in Chemical Engineering: Engineering Aspects of Self-Organising Materials*, Academic Press.
- KOSHLAND, D. E. 1953. Stereochemistry and the mechanism of enzymatic reactions. *Biological Reviews*, 28, 416-436.
- KRAMER, B., RAREY, M. & LENGAUER, T. 1999. Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. *Proteins: Structure, Function, and Bioinformatics*, 37, 228-241.
- KROEMER, R. T. 2007. Structure-based drug design: docking and scoring. *Current Protein and Peptide Science*, 8, 312-328.
- KRÓL, E., RYCHŁOWSKA, M. & SZEWCZYK, B. 2014. Antivirals-current trends in fighting influenza. *Acta Biochimica Polonica*, 61, 495-504.
- KUBINYI, H. 2006. Success stories of computer-aided design. *Computer applications in pharmaceutical research and development*, 2, 377.
- KUHN, B., MOHR, P. & STAHL, M. 2010. Intramolecular hydrogen bonding in medicinal chemistry. *Journal of medicinal chemistry*, 53, 2601-2611.
- KUMAR, S. 2012. *Textbook of microbiology*, JP Medical Ltd.
- KUMAR, S. & NUSSINOV, R. 2002. Close-range electrostatic interactions in proteins. *ChemBioChem*, 3, 604-617.
- KUNTZ, I. D., BLANEY, J. M., OATLEY, S. J., LANGRIDGE, R. & FERRIN, T. E. 1982. A geometric approach to macromolecule–ligand interactions. *Journal of molecular biology*, 161, 269-288.
- LADBURY, J. E. 1996. Just add water! The effect of water on the specificity of protein–ligand binding sites and its potential application to drug design. *Chemistry & biology*, 3, 973-980.
- LEE, N., CHOI, K., CHAN, P., HUI, D., LUI, G., WONG, B., WONG, R., SIN, W., HUI, W. & NGAI, K. 2010. Outcomes of adults hospitalised with severe influenza. *Thorax*, 65, 510-515.
- LEESON, P. D. & SPRINGTHORPE, B. 2007. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nature Reviews Drug Discovery*, 6, 881-890.

- LEMM, J. A., O'BOYLE, D., LIU, M., NOWER, P. T., COLONNO, R., DESHPANDE, M. S., SNYDER, L. B., MARTIN, S. W., LAURENT, D. R. S. & SERRANO-WU, M. H. 2010. Identification of hepatitis C virus NS5A inhibitors. *Journal of virology*, 84, 482-491.
- LENGAUER, T., LEMMEN, C., RAREY, M. & ZIMMERMANN, M. 2004. Novel technologies for virtual screening. *Drug Discovery Today*, 9, 27-34.
- LEO, A., JOW, P., SILIPO, C. & HANSCH, C. 1975. Calculation of hydrophobic constant (log P) from  $\pi$  and  $f$  constants. *Journal of medicinal chemistry*, 18, 865-868.
- LEVITZKI, A. & KOSHLAND, D. 1969. Negative cooperativity in regulatory enzymes. *Proceedings of the National Academy of Sciences*, 62, 1121-1128.
- LI, W., ESCARPE, P. A., EISENBERG, E. J., CUNDY, K. C., SWEET, C., JAKEMAN, K. J., MERSON, J., LEW, W., WILLIAMS, M., ZHANG, L., KIM, C. U., BISCHOFBERGER, N., CHEN, M. S. & MENDEL, D. B. 1998. Identification of GS 4104 as an Orally Bioavailable Prodrug of the Influenza Virus Neuraminidase Inhibitor GS 4071. *Antimicrobial Agents and Chemotherapy*, 42, 647-653.
- LI, X., CHEN, Y., LU, S., HUANG, Z., LIU, X., WANG, Q., SHI, T. & ZHANG, J. 2013. Toward an understanding of the sequence and structural basis of allosteric proteins. *Journal of Molecular Graphics and Modelling*, 40, 30-39.
- LI, Y., LUBCHENKO, V. & VEKILOV, P. G. 2011. The use of dynamic light scattering and Brownian microscopy to characterize protein aggregation. *Review of Scientific Instruments*, 82, 053106.
- LIAO, C., SITZMANN, M., PUGLIESE, A. & NICKLAUS, M. C. 2011. Software and resources for computational medicinal chemistry. *Future medicinal chemistry*, 3, 1057-1085.
- LIMA, M. D. S., FIRMO, A. A. M. & MARTINS-MELO, F. R. 2016. Trends in AIDS-related mortality among people aged 60 years and older in Brazil: a nationwide population-based study. *AIDS care*, 1-8.
- LINDBERG, G., EKLUND, G. A., GULLBERG, B. & RÅSTAM, L. 1991. Serum sialic acid concentration and cardiovascular mortality. *Bmj*, 302, 143-146.
- LINDEGARDH, N., HANPITHAKPONG, W., KAMANIKOM, B., FARRAR, J., HIEN, T. T., SINGHASIVANON, P., WHITE, N. J. & DAY, N. P. J. 2011. Quantification of the anti-influenza drug zanamivir in plasma using high-throughput HILIC-MS/MS. *Bioanalysis*, 3, 157-165.
- LINDEMANN, L., JACOBSEN, H., SCHUHBAUER, D., KNOFLACH, F., GATTI, S., WETTSTEIN, J. G., LOETSCHER, H., CHU, T., EBELING, M. & PAULSON, J. C. 2010. In vitro pharmacological selectivity profile of oseltamivir prodrug (Tamiflu®) and active metabolite. *European journal of pharmacology*, 628, 6-10.
- LINDSAY, M. A. 2003. Target discovery. *Nature Reviews Drug Discovery*, 2, 831-838.
- LIONTA, E., SPYROU, G., VASSILATIS, D. K. & COURNIA, Z. 2014. Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. *Current Topics in Medicinal Chemistry*, 14, 1923-1938.
- LIPINSKI, C. & HOPKINS, A. 2004. Navigating chemical space for biology and medicine. *Nature*, 432, 855-861.
- LIPINSKI, C. A., LOMBARDO, F., DOMINY, B. W. & FEENEY, P. J. 1997. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 23, 3-25.
- LIPINSKI, C. A., LOMBARDO, F., DOMINY, B. W. & FEENEY, P. J. 2012. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, 64, 4-17.
- LIVINGSTONE, D. J. & DAVIS, A. M. 2011. *Drug design strategies: Quantitative approaches*, Royal Society of Chemistry.
- LOMBARD, V., RAMULU, H. G., DRULA, E., COUTINHO, P. M. & HENRISSAT, B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research*, 42, D490-D495.

- LOMBARDINO, J. G. & LOWE, J. A. 2004. The role of the medicinal chemist in drug discovery—then and now. *Nature Reviews Drug Discovery*, 3, 853-862.
- LONG, R. D., HILLIARD JR, N. P., CHHATRE, S. A., TIMOFEEVA, T. V., YAKOVENKO, A. A., DEI, D. K. & MENSAH, E. A. 2010. Comparison of zwitterionic N-alkylaminomethanesulfonic acids to related compounds in the Good buffer series. *Beilstein journal of organic chemistry*, 6, 31.
- LOUNNAS, V., RITSCHER, T., KELDER, J., MCGUIRE, R., BYWATER, R. P. & FOLOPPE, N. 2013. CURRENT PROGRESS IN STRUCTURE-BASED RATIONAL DRUG DESIGN MARKS A NEW MINDSET IN DRUG DISCOVERY. *Computational and Structural Biotechnology Journal*, 5, 1-14.
- LOWDEN, J. A. & O'BRIEN, J. S. 1979. Sialidosis: a review of human neuraminidase deficiency. *American Journal of Human Genetics*, 31, 1-18.
- LUNDBLAD, A. 2015. Gunnar Blix and his discovery of sialic acids. Fascinating molecules in glycobiology. *Uppsala Journal of Medical Sciences*, 120, 104-112.
- LUO, Y., LI, S.-C., CHOU, M.-Y., LI, Y.-T. & LUO, M. 1998. The crystal structure of an intramolecular trans-sialidase with a NeuA $\alpha$ 2 $\rightarrow$ 3Gal specificity. *Structure*, 6, 521-530.
- MACARRON, R. 2006. Critical review of the role of HTS in drug discovery. *Drug discovery today*, 11, 277-279.
- MACARRON, R., BANKS, M. N., BOJANIC, D., BURNS, D. J., CIROVIC, D. A., GARYANTES, T., GREEN, D. V., HERTZBERG, R. P., JANZEN, W. P. & PASLAY, J. W. 2011. Impact of high-throughput screening in biomedical research. *Nature reviews Drug discovery*, 10, 188-195.
- MAESTRO, B. & SANZ, J. M. 2016. Choline Binding Proteins from *Streptococcus pneumoniae*: A Dual Role as Enzybiotics and Targets for the Design of New Antimicrobials. *Antibiotics*, 5, 21.
- MAKLEY, L. N. & GESTWICKI, J. E. 2013. Expanding the Number of 'Druggable' Targets: Non-Enzymes and Protein-Protein Interactions. *Chemical biology & drug design*, 81, 22-32.
- MANCO, S., HERNON, F., YESILKAYA, H., PATON, J. C., ANDREW, P. W. & KADIOGLU, A. 2006. Pneumococcal Neuraminidases A and B Both Have Essential Roles during Infection of the Respiratory Tract and Sepsis. *Infection and Immunity*, 74, 4014-4020.
- MANIS, J. P. 2007. Knock out, knock in, knock down—genetically manipulated mice and the Nobel Prize. *New England Journal of Medicine*, 357, 2426-2429.
- MANTA, B., COMINI, M., MEDEIROS, A., HUGO, M., TRUJILLO, M. & RADI, R. 2013. Trypanothione: A unique bis-glutathionyl derivative in trypanosomatids. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1830, 3199-3216.
- MARION, C., BURNAUGH, A. M., WOODIGA, S. A. & KING, S. J. 2011. Sialic Acid Transport Contributes to Pneumococcal Colonization. *Infection and Immunity*, 79, 1262-1269.
- MARTIN, W. H., HOOVER, D. J., ARMENTO, S. J., STOCK, I. A., MCPHERSON, R. K., DANLEY, D. E., STEVENSON, R. W., BARRETT, E. J. & TREADWAY, J. L. 1998. Discovery of a human liver glycogen phosphorylase inhibitor that lowers blood glucose in vivo. *Proceedings of the National Academy of Sciences*, 95, 1776-1781.
- MARTÍNEZ, J. L. 2008. Antibiotics and antibiotic resistance genes in natural environments. *Science*, 321, 365-367.
- MAUN, H. R., EIGENBROT, C. & LAZARUS, R. A. 2003. Engineering exosite peptides for complete inhibition of factor VIIa using a protease switch with substrate phage. *Journal of Biological Chemistry*, 278, 21823-21830.
- MAY, L. T., LEACH, K., SEXTON, P. M. & CHRISTOPOULOS, A. 2007. Allosteric modulation of G protein-coupled receptors. *Annu. Rev. Pharmacol. Toxicol.*, 47, 1-51.
- MCGOVERN, S. L., CASELLI, E., GRIGORIEFF, N. & SHOICHET, B. K. 2002. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *Journal of medicinal chemistry*, 45, 1712-1722.
- MCGOVERN, S. L., HELFAND, B. T., FENG, B. & SHOICHET, B. K. 2003. A specific mechanism of nonspecific inhibition. *Journal of medicinal chemistry*, 46, 4265-4272.
- MCLAUGHLIN, M. M., SKOGLUND, E. W. & ISON, M. G. 2015. Peramivir: an intravenous neuraminidase inhibitor. *Expert opinion on pharmacotherapy*, 16, 1889-1900.

- MEINDL, P., BODO, G., PALESE, P., SCHULMAN, J. & TUPPY, H. 1974a. Inhibition of neuraminidase activity by derivatives of 2-deoxy-2,3-dehydro-N-acetylneuraminic acid. *Virology*, 58, 457-463.
- MEINDL, P., BODO, G., PALESE, P., SCHULMAN, J. & TUPPY, H. 1974b. Inhibition of neuraminidase activity by derivatives of 2-deoxy-2, 3-dehydro-N-acetylneuraminic acid. *Virology*, 58, 457-463.
- MENDONÇA-PREVIATO, L., REGINA TODESCHINI, A., FREIRE DE LIMA, L. & OSVALDO PREVIATO, J. 2010. The trans-sialidase from *Trypanosoma cruzi* a putative target for trypanocidal agents. *The Open Parasitology Journal*, 4.
- MENG, X.-Y., ZHANG, H.-X., MEZEI, M. & CUI, M. 2011. Molecular Docking: A powerful approach for structure-based drug discovery. *Current computer-aided drug design*, 7, 146-157.
- MEYER, E. 1992. Internal water molecules and H-bonding in biological macromolecules: A review of structural features with functional implications. *Protein science*, 1, 1543-1562.
- MICHEL, J., TIRADO-RIVES, J. & JORGENSEN, W. L. 2009. Energetics of Displacing Water Molecules from Protein Binding Sites: Consequences for Ligand Optimization. *Journal of the American Chemical Society*, 131, 15403-15411.
- MILO, R., HOU, J. H., SPRINGER, M., BRENNER, M. P. & KIRSCHNER, M. W. 2007. The relationship between evolutionary and physiological variation in hemoglobin. *Proceedings of the National Academy of Sciences*, 104, 16998-17003.
- MITCHELL, T. J. 2003. The pathogenesis of streptococcal infections: from tooth decay to meningitis. *Nature Reviews Microbiology*, 1, 219-230.
- MIYAGI, T. & YAMAGUCHI, K. 2012. Mammalian sialidases: physiological and pathological roles in cellular functions. *Glycobiology*, 22, 880-896.
- MOFFAT, J. G., RUDOLPH, J. & BAILEY, D. 2014. Phenotypic screening in cancer drug discovery—past, present and future. *Nature Reviews Drug Discovery*, 13, 588-602.
- MOHR, J. T., KROUT, M. R. & STOLTZ, B. M. 2008. Natural products as inspiration for the development of asymmetric catalysis. *Nature*, 455, 323-332.
- MONOD, J., CHANGEUX, J.-P. & JACOB, F. 1963. Allosteric proteins and cellular control systems. *Journal of molecular biology*, 6, 306-329.
- MONOD, J., WYMAN, J. & CHANGEUX, J.-P. 1965. On the nature of allosteric transitions: a plausible model. *Journal of molecular biology*, 12, 88-118.
- MONTAGNA, G., CREMONA, M. L., PARIS, G., AMAYA, M. F., BUSCHIAZZO, A., ALZARI, P. M. & FRASCH, A. C. 2002. The trans-sialidase from the african trypanosome *Trypanosoma brucei*. *European Journal of Biochemistry*, 269, 2941-2950.
- MONTI, E., BONTEN, E., D'AZZO, A., BRESCIANI, R., VENERANDO, B., BORSANI, G., SCHAUER, R. & TETTAMANTI, G. 2010. Sialidases in vertebrates: a family of enzymes tailored for several cell functions. *Advances in carbohydrate chemistry and biochemistry*, 64, 403-479.
- MORAN, M., GUZMAN, J., ROPARS, A.-L., MCDONALD, A., JAMESON, N., OMUNE, B., RYAN, S. & WU, L. 2009. Neglected disease research and development: how much are we really spending? *PLoS Med*, 6, e1000030.
- MORENS, D. M., FOLKERS, G. K. & FAUCI, A. S. 2004. The challenge of emerging and re-emerging infectious diseases. *Nature*, 430, 242-249.
- MORGAN, P., VAN DER GRAAF, P. H., ARROWSMITH, J., FELTNER, D. E., DRUMMOND, K. S., WEGNER, C. D. & STREET, S. D. 2012. Can the flow of medicines be improved? Fundamental pharmacokinetic and pharmacological principles toward improving Phase II survival. *Drug discovery today*, 17, 419-424.
- MORIUCHI, H., KATSUSHIMA, N., NISHIMURA, H., NAKAMURA, K. & NUMAZAKI, Y. 1991. Community-acquired influenza C virus infection in children. *The Journal of pediatrics*, 118, 235-238.
- MORLEY, T. J., WILLIS, L. M., WHITFIELD, C., WAKARCHUK, W. W. & WITHERS, S. G. 2009. A New Sialidase Mechanism: BACTERIOPHAGE K1F ENDO-SIALIDASE IS AN INVERTING GLYCOSIDASE. *The Journal of Biological Chemistry*, 284, 17404-17410.

- MRAZEK, M. F. & MOSSIALOS, E. 2003. Stimulating pharmaceutical research and development for neglected diseases. *Health policy*, 64, 75-88.
- MUELLER-LANGER, F. 2013. Neglected infectious diseases: are push and pull incentive mechanisms suitable for promoting drug development research? *Health Economics, Policy and Law*, 8, 185-208.
- MULLARD, A. 2016. 2015 FDA drug approvals. *Nature Reviews Drug Discovery*, 15, 73-76.
- MURPHEY, R. D., STERN, H. M., STRAUB, C. T. & ZON, L. I. 2006. A chemical genetic screen for cell cycle inhibitors in zebrafish embryos. *Chemical biology & drug design*, 68, 213-219.
- MURRAY, C. W., CARR, M. G., CALLAGHAN, O., CHESSARI, G., CONGREVE, M., COWAN, S., COYLE, J. E., DOWNHAM, R., FIGUEROA, E. & FREDERICKSON, M. 2010. Fragment-based drug discovery applied to Hsp90. Discovery of two lead series with high ligand efficiency. *Journal of medicinal chemistry*, 53, 5942-5955.
- NEISES, B. & STEGLICH, W. 1978. Simple method for the esterification of carboxylic acids. *Angewandte Chemie International Edition in English*, 17, 522-524.
- NEMETH, E. 2006. Misconceptions about calcimimetics. *Annals of the New York Academy of Sciences*, 1068, 471-476.
- NERES, J., BONNET, P., EDWARDS, P. N., KOTIAN, P. L., BUSCHIAZZO, A., ALZARI, P. M., BRYCE, R. A. & DOUGLAS, K. T. 2007. Benzoic acid and pyridine derivatives as inhibitors of Trypanosoma cruzi trans-sialidase. *Bioorganic & medicinal chemistry*, 15, 2106-2119.
- NERES, J., BREWER, M. L., RATIER, L., BOTTI, H., BUSCHIAZZO, A., EDWARDS, P. N., MORTENSON, P. N., CHARLTON, M. H., ALZARI, P. M. & FRASCH, A. C. 2009. Discovery of novel inhibitors of Trypanosoma cruzi trans-sialidase from in silico screening. *Bioorganic & medicinal chemistry letters*, 19, 589-596.
- NERES, J., BRYCE, R. A. & DOUGLAS, K. T. 2008. Rational drug design in parasitology: trans-sialidase as a case study for Chagas disease. *Drug discovery today*, 13, 110-117.
- NEU, H. C. 1992. The crisis in antibiotic resistance. *Science*, 257, 1064-1073.
- NEWSTEAD, S., WATSON, J. N., KNOLL, T. L., BENNET, A. J. & TAYLOR, G. 2005. Structure and mechanism of action of an inverting mutant sialidase. *Biochemistry*, 44, 9117-9122.
- NICHOLLS, J. M., MOSS, R. B. & HASLAM, S. M. 2013. The use of sialidase therapy for respiratory viral infections. *Antiviral research*, 98, 401-409.
- NOIREAU, F., DIOSQUE, P. & JANSEN, A. M. 2009. Trypanosoma cruzi: adaptation to its vectors and its hosts. *Veterinary research*, 40, 1-23.
- NUSSINOV, R. & TSAI, C.-J. 2012. The different ways through which specificity works in orthosteric and allosteric drugs. *Current pharmaceutical design*, 18, 1311-1316.
- NUSSINOV, R. & TSAI, C.-J. 2013. Allostery in disease and in drug discovery. *Cell*, 153, 293-305.
- NWAKA, S. & RIDLEY, R. G. 2003. Virtual drug discovery and development for neglected diseases through public-private partnerships. *Nature Reviews Drug Discovery*, 2, 919-928.
- OIKONOMAKOS, N. G., SKAMNAKI, V. T., TSITSANOOU, K. E., GAVALAS, N. G. & JOHNSON, L. N. 2000. A new allosteric site in glycogen phosphorylase b as a target for drug interactions. *Structure*, 8, 575-584.
- OO, C., SNELL, P., BARRETT, J., DORR, A., LIU, B. & WILDING, I. 2003. Pharmacokinetics and delivery of the anti-influenza prodrug oseltamivir to the small intestine and colon using site-specific delivery capsules. *International journal of pharmaceuticals*, 257, 297-299.
- OPREA, T. I., HUMMER, G. & GARCÍA, A. E. 1997. Identification of a functional water channel in cytochrome P450 enzymes. *Proceedings of the National Academy of Sciences*, 94, 2133-2138.
- OSTLER, E. L. 2007. Chemical Biology is. *Chemistry Central Journal*, 1, 5-5.
- OVERINGTON, J. P., AL-LAZIKANI, B. & HOPKINS, A. L. 2006. How many drug targets are there? *Nature reviews Drug discovery*, 5, 993-996.
- OWEN, C. D., LUKACIK, P., POTTER, J. A., SLEATOR, O., TAYLOR, G. L. & WALSH, M. A. 2015. Streptococcus pneumoniae NanC STRUCTURAL INSIGHTS INTO THE SPECIFICITY AND

- MECHANISM OF A SIALIDASE THAT PRODUCES A SIALIDASE INHIBITOR. *Journal of Biological Chemistry*, 290, 27736-27748.
- OWENS, G. M. 2015. New FDA Drug Approvals Hit an 18-Year High in 2014. *American Health & Drug Benefits*, 8, 15-17.
- OXFORD, J. 2000. Zanamivir (Glaxo Wellcome). *IDrugs: the investigational drugs journal*, 3, 447-459.
- PALESE, P., TOBITA, K., UEDA, M. & COMPANS, R. W. 1974. Characterization of temperature sensitive influenza virus mutants defective in neuraminidase. *Virology*, 61, 397-410.
- PALUMBO, E. 2011. Pegylated Interferon and Ribavirin Treatment for Hepatitis C Virus Infection. *Therapeutic Advances in Chronic Disease*, 2, 39-45.
- PARIS, G., RATIER, L., AMAYA, M. F., NGUYEN, T., ALZARI, P. M. & FRASCH, A. C. C. 2005. A Sialidase Mutant Displaying trans-Sialidase Activity. *Journal of Molecular Biology*, 345, 923-934.
- PARKER, D., SOONG, G., PLANET, P., BROWER, J., RATNER, A. J. & PRINCE, A. 2009. The NanA neuraminidase of *Streptococcus pneumoniae* is involved in biofilm formation. *Infection and immunity*, 77, 3722-3730.
- PATERSON, I. & ANDERSON, E. A. 2005. The renaissance of natural products as drug candidates. *Science*, 310, 451-453.
- PATON, J. C. 1998. Novel pneumococcal surface proteins: role in virulence and vaccine potential. *Trends in microbiology*, 6, 85-87.
- PATTERSON, M. J. 1996. *Streptococcus*. In: BARON, S. (ed.) *Medical Microbiology*. 4th ed. Glaveston (TX): University of Texas Medical Branch.
- PAUL GLEZEN, W., SCHMIER, J. K., KUEHN, C. M., RYAN, K. J. & OXFORD, J. 2013. The burden of influenza B: a structured literature review. *American journal of public health*, 103, e43-e51.
- PEREIRA, M. 1983. A developmentally regulated neuraminidase activity in *Trypanosoma cruzi*. *Science*, 219, 1444-1446.
- PETTIGREW, M. M., FENNIE, K. P., YORK, M. P., DANIELS, J. & GHAFAR, F. 2006. Variation in the Presence of Neuraminidase Genes among *Streptococcus pneumoniae* Isolates with Identical Sequence Types. *Infection and Immunity*, 74, 3360-3365.
- PHILLIPS, I., CASEWELL, M., COX, T., DE GROOT, B., FRIIS, C., JONES, R., NIGHTINGALE, C., PRESTON, R. & WADDELL, J. 2004. Does the use of antibiotics in food animals pose a risk to human health? A critical review of published data. *Journal of Antimicrobial Chemotherapy*, 53, 28-52.
- PHARMA. *Members* [Online]. Available: <http://www.phrma.org/about/members> [Accessed].
- PHARMA 2015. Profile Biopharmaceutical Research Industry.
- PLENGE, R. M., SCOLNICK, E. M. & ALTSHULER, D. 2013. Validating therapeutic targets through human genetics. *Nature reviews Drug discovery*, 12, 581-594.
- POLICYCURES 2014. G-FINDER Neglected Disease Research and Development: Emerging Trends.
- POLICYCURES 2015. G-FINDER Neglected Disease Research and Development: The Ebola Effect.
- POLICYCURES. 2016. *G-Finder* [Online]. Available: <http://policycures.org/gfinder.html> [Accessed].
- POLITZ, M. C., COPELAND, M. F. & PFLEGER, B. F. 2013. Artificial repressors for controlling gene expression in bacteria. *Chemical communications*, 49, 4325-4327.
- POLLASTRI, M. P. 2014. Finding new collaboration models for enabling neglected tropical disease drug discovery. *PLoS Negl Trop Dis*, 8, e2866.
- PREVIATO, J., ANDRADE, A. F., PESSOLANI, M. C. V. & MENDONÇA-PREVIATO, L. 1985. Incorporation of sialic acid into *Trypanosoma cruzi* macromolecules. A proposal for a new metabolic route. *Molecular and biochemical parasitology*, 16, 85-96.
- PROLO, L. M., VOGEL, H. & REIMER, R. J. 2009. The lysosomal sialic acid transporter sialin is required for normal CNS myelination. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29, 15355.
- PSHEZHETSKY, A. V. & HINEK, A. 2011. Where catabolism meets signalling: neuraminidase 1 as a modulator of cell receptors. *Glycoconjugate journal*, 28, 441-452.

- PUBCHEM. *1,2-dichloroethane* [Online]. Available: <https://pubchem.ncbi.nlm.nih.gov/compound/11#section=Top> [Accessed].
- PUBCHEM. *DICHLOROMETHANE* [Online]. Available: <https://pubchem.ncbi.nlm.nih.gov/compound/6344> [Accessed].
- PUJADAS, G., VAQUE, M., ARDEVOL, A., BLADE, C., SALVADO, M., BLAY, M., FERNANDEZ-LARREA, J. & AROLA, L. 2008. Protein-ligand docking: A review of recent advances and future perspectives. *Current Pharmaceutical Analysis*, 4, 1-19.
- QUAGLIARELLO, V. J. & SCHELD, W. M. 1997. Treatment of bacterial meningitis. *New England Journal of Medicine*, 336, 708-716.
- QUISTGAARD, E. M. & THIRUP, S. S. 2009. Sequence and structural analysis of the Asp-box motif and Asp-box beta-propellers; a widespread propeller-type characteristic of the Vps10 domain family and several glycoside hydrolase families. *BMC structural biology*, 9, 1.
- RAFELSON JR, M. E. 1962. THE NEURAMINIDASES AND THEIR ACTION ON GLYCOPROTEINS AND OTHER SIALIC ACID-CONTAINING COMPOUNDS. *Exposés annuels de biochimie médicale*, 24, 121-132.
- RAICU, I. 2009. *Many-task computing: bridging the gap between high-throughput computing and high-performance computing*, ProQuest.
- RANG, H. P., RITTER, J. M., FLOWER, R. J. & HENDERSON, G. 2014. *Rang & Dale's Pharmacology: With student consult online access*, Elsevier Health Sciences.
- RAREY, M., KRAMER, B., LENGAUER, T. & KLEBE, G. 1996. A fast flexible docking method using an incremental construction algorithm. *Journal of molecular biology*, 261, 470-489.
- RASSI, A. & MARIN-NETO, J. A. 2010. Chagas disease. *The Lancet*, 375, 1388-1402.
- RATH, V. L., AMMIRATI, M., DANLEY, D. E., EKSTROM, J. L., GIBBS, E. M., HYNES, T. R., MATHIOWETZ, A. M., MCPHERSON, R. K., OLSON, T. V. & TREADWAY, J. L. 2000. Human liver glycogen phosphorylase inhibitors bind at a new allosteric site. *Chemistry & biology*, 7, 677-682.
- RATIER, L., URRUTIA, M., PARIS, G., ZAREBSKI, L., FRASCH, A. C. & GOLDBAUM, F. A. 2008. Relevance of the diversity among members of the Trypanosoma cruzi trans-sialidase family analyzed with camelids single-domain antibodies. *PLoS One*, 3, e3524.
- RAUTIO, J., KUMPULAINEN, H., HEIMBACH, T., OLIYAI, R., OH, D., JARVINEN, T. & SAVOLAINEN, J. 2008. Prodrugs: design and clinical applications. *Nat Rev Drug Discov*, 7, 255-270.
- RCSB. *PDB* [Online]. Available: [www.rcsb.org](http://www.rcsb.org) [Accessed].
- REBUFFET, E., GROISILLIER, A., THOMPSON, A., JEUDY, A., BARBEYRON, T., CZIZEK, M. & MICHEL, G. 2011. Discovery and structural characterization of a novel glycosidase family of marine origin. *Environmental microbiology*, 13, 1253-1270.
- REDDY, A. S., PATI, S. P., KUMAR, P. P., PRADEEP, H. & SASTRY, G. N. 2007. Virtual screening in drug discovery-a computational perspective. *Current Protein and Peptide Science*, 8, 329-351.
- REDFERN, W. S., CARLSSON, L., DAVIS, A. S., LYNCH, W. G., MACKENZIE, I., PALETHORPE, S., SIEGL, P. K. S., STRANG, I., SULLIVAN, A. T., WALLIS, R., CAMM, A. J. & HAMMOND, T. G. 2003. Relationships between preclinical cardiac electrophysiology, clinical QT interval prolongation and torsade de pointes for a broad range of drugs: evidence for a provisional safety margin in drug development. *Cardiovascular Research*, 58, 32-45.
- REES, D. C., CONGREVE, M., MURRAY, C. W. & CARR, R. 2004. Fragment-based lead discovery. *Nature Reviews Drug Discovery*, 3, 660-672.
- RESISTANCE, R. O. A. 2015. Antimicrobials in Agriculture and the Environment: Reducing Unnecessary Use and Waste. In: O'NEILL, J. (ed.).
- RISHTON, G. M. 2003. Nonleadlikeness and leadlikeness in biochemical screening. *Drug discovery today*, 8, 86-96.
- ROBERGE, M., SANTELL, L., DENNIS, M. S., EIGENBROT, C., DWYER, M. A. & LAZARUS, R. A. 2001. A novel exosite on coagulation factor VIIa and its molecular interactions with a new class of peptide inhibitors. *Biochemistry*, 40, 9522-9531.

- ROBERT, X. & GOUET, P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research*, 42, W320-W324.
- ROSANO, G. L. & CECCARELLI, E. A. 2014. Recombinant protein expression in Escherichia coli: advances and challenges. *Recombinant protein expression in microbial systems*, 7.
- ROSSENU, S., DEWITTE, D., VANDEKERCKHOVE, J. & AMPE, C. 1997. A phage display technique for a fast, sensitive, and systematic investigation of protein–protein interactions. *Journal of protein chemistry*, 16, 499-503.
- ROTA, P. A., WALLIS, T. R., HARMON, M. W., ROTA, J. S., KENDAL, A. P. & NEROME, K. 1990. Cocirculation of two distinct evolutionary lineages of influenza type B virus since 1983. *Virology*, 175, 59-68.
- ROY, A., R MCDONALD, P., SITTAMPALAM, S. & CHAGUTURU, R. 2010. Open access high throughput drug discovery in the public domain: a Mount Everest in the making. *Current pharmaceutical biotechnology*, 11, 764-778.
- RU, X., SONG, C. & LIN, Z. 2016. A genetic algorithm encoded with the structural information of amino acids and dipeptides for efficient conformational searches of oligopeptides. *Journal of computational chemistry*.
- RUSSELL, R. J., HAIRE, L. F., STEVENS, D. J., COLLINS, P. J., LIN, Y. P., BLACKBURN, G. M., HAY, A. J., GAMBLIN, S. J. & SKEHEL, J. J. 2006. The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature*, 443, 45-49.
- SABINO, E. C., LEE, T. H., MONTALVO, L., NGUYEN, M. L., LEIBY, D. A., CARRICK, D. M., OTANI, M. M., VINELLI, E., WRIGHT, D. & STRAMER, S. L. 2013. Antibody levels correlate with detection of Trypanosoma cruzi DNA by sensitive polymerase chain reaction assays in seropositive blood donors and possible resolution of infection over time. *Transfusion*, 53, 1257-1265.
- SADOWSKY, J. D., BURLINGAME, M. A., WOLAN, D. W., MCCLENDON, C. L., JACOBSON, M. P. & WELLS, J. A. 2011. Turning a protein kinase on or off from a single allosteric site via disulfide trapping. *Proceedings of the National Academy of Sciences*, 108, 6056-6061.
- SANCHEZ, A., KSIAZEK, T. G., ROLLIN, P. E., PETERS, C. J., NICHOL, S. T., KHAN, A. S. & MAHY, B. 1995. Reemergence of Ebola virus in Africa. *Emerging Infectious Diseases*, 1, 96.
- SANGUINETTI, M. C. & TRISTANI-FIROUZI, M. 2006. hERG potassium channels and cardiac arrhythmia. *Nature*, 440, 463-469.
- SASAKI, S., CHO, N., NARA, Y., HARADA, M., ENDO, S., SUZUKI, N., FURUYA, S. & FUJINO, M. 2003. Discovery of a Thieno[2,3-d]pyrimidine-2,4-dione Bearing a p-Methoxyureidophenyl Moiety at the 6-Position: A Highly Potent and Orally Bioavailable Non-Peptide Antagonist for the Human Luteinizing Hormone-Releasing Hormone Receptor. *Journal of Medicinal Chemistry*, 46, 113-124.
- SATYANARAYANAJOIS, S. D. & HILL, R. A. 2011. Medicinal chemistry for 2020. *Future medicinal chemistry*, 3, 1765-1786.
- SCHAUER, R. 2000. Achievements and challenges of sialic acid research. *Glycoconjugate journal*, 17, 485-499.
- SCHAUER, R. 2004. Sialic acids: fascinating sugars in higher animals and man. *Zoology*, 107, 49-64.
- SCHENKMAN, S. & EICHINGER, D. 1993. Trypanosoma cruzi trans-sialidase and cell invasion. *Parasitology Today*, 9, 218-222.
- SCHENONE, M., DANČÍK, V., WAGNER, B. K. & CLEMONS, P. A. 2013. Target identification and mechanism of action in chemical biology and drug discovery. *Nature chemical biology*, 9, 232-240.
- SCHNAAR, R., SUZUKI, A. & STANLEY, P. 2009. Chapter 10: Glycosphingolipids. *Essentials of Glycobiology 2nd edit.* (Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME, eds.). Cold Spring Harbor, New York.
- SCHNAAR, R. L., GERARDY-SCHAHN, R. & HILDEBRANDT, H. 2014. Sialic Acids in the Brain: Gangliosides and Polysialic Acid in Nervous System Development, Stability, Disease, and Regeneration. *Physiological Reviews*, 94, 461-518.

- SCHNEIDER, G. & BÖHM, H.-J. 2002. Virtual screening and fast automated docking methods. *Drug Discovery Today*, 7, 64-70.
- SCHNELL, J. R. & CHOU, J. J. 2008. Structure and mechanism of the M2 proton channel of influenza A virus. *Nature*, 451, 591-595.
- SCHOFIELD, C. J. & DIAS, J. C. P. 1991. A cost-benefit analysis of chagas disease control. *Memórias do Instituto Oswaldo Cruz*, 86, 285-295.
- SCHOLL, M., DING, S., LEE, C. W. & GRUBBS, R. H. 1999. Synthesis and Activity of a New Generation of Ruthenium-Based Olefin Metathesis Catalysts Coordinated with 1, 3-Dimesityl-4, 5-dihydroimidazol-2-ylidene Ligands. *Organic Letters*, 1, 953-956.
- SCHREIBER, S. L. 1998. Chemical genetics resulting from a passion for synthetic organic chemistry. *Bioorganic & medicinal chemistry*, 6, 1127-1152.
- SCHRODI, Y. & PEDERSON, R. L. 2007. Evolution and applications of second-generation ruthenium olefin metathesis catalysts. *Aldrichimica Acta*, 40, 45-52.
- SCHWAB, C. H. 2011. Conformations and 3D pharmacophore searching. *Drug Discovery Today: Technologies*, 7, e245-e253.
- SCHWAB, P., FRANCE, M. B., ZILLER, J. W. & GRUBBS, R. H. 1995. A Series of Well-Defined Metathesis Catalysts—Synthesis of [RuCl<sub>2</sub>(CHR')(PR<sub>3</sub>)<sub>2</sub>] and Its Reactions. *Angewandte Chemie International Edition in English*, 34, 2039-2041.
- SCIENCE, A. A. F. T. A. O. 1884. THE IMPORTANCE OF CHEMISTRY IN BIOLOGY AND MEDICINE. *Science*, 454-456.
- SCOTT, A. R. 2016. Chemical probes: A shared toolbox. *Nature*, 533, S60-S61.
- SEBOLT-LEOPOLD, J. S., DUDLEY, D. T., HERRERA, R., VAN BECELAERE, K., WILAND, A., GOWAN, R. C., TECLE, H., BARRETT, S. D., BRIDGES, A. & PRZYBRANOWSKI, S. 1999. Blockade of the MAP kinase pathway suppresses growth of colon tumors in vivo. *Nature medicine*, 5, 810-816.
- SERVICE, U. S. P. H. 2001. Agency for Toxic Substances and Disease Registry (ATSDR). TOXICOLOGICAL PROFILE FOR 1,2-DICHLOROETHANE
- SEVERI, E., HOOD, D. W. & THOMAS, G. H. 2007. Sialic acid utilization by bacterial pathogens. *Microbiology*, 153, 2817-2822.
- SEYRANTEPE, V., LANDRY, K., TRUDEL, S., HASSAN, J. A., MORALES, C. R. & PSHEZHETSKY, A. V. 2004. Neu4, a novel human lysosomal lumen sialidase, confers normal phenotype to sialidosis and galactosialidosis cells. *Journal of Biological Chemistry*, 279, 37021-37029.
- SEYRANTEPE, V., POUPETOVA, H., FROISSART, R., ZABOT, M. T., MAIRE, I. & PSHEZHETSKY, A. V. 2003. Molecular pathology of NEU1 gene in sialidosis. *Human mutation*, 22, 343-352.
- SHAH, N. S., WRIGHT, A., BAI, G.-H., BARRERA, L., BOULAHBAL, F., DROBNIOWSKI, F., GILPIN, C., HAVELKOV, M., LEPE, R. & LUMB, R. 2007. Worldwide emergence of extensively drug-resistant tuberculosis.
- SHAMES, S. L., FAIRLAMB, A. H., CERAMI, A. & WALSH, C. T. 1986. Purification and characterization of trypanothione reductase from *Crithidia fasciculata*, a new member of the family of disulfide-containing flavoprotein reductases. *Biochemistry*, 25, 3519-3526.
- SHANMUGAM, S. 2015. Zanamivir oral delivery: possibilities revisited. *Therapeutic delivery*, 6, 403-405.
- SHARMA, A., JACOB, A., TANDON, M. & KUMAR, D. 2010. Orphan drug: development trends and strategies. *Journal of Pharmacy and Bioallied Sciences*, 2, 290.
- SHIE, J.-J., FANG, J.-M., LAI, P.-T., WEN, W.-H., WANG, S.-Y., CHENG, Y.-S. E., TSAI, K.-C., YANG, A.-S. & WONG, C.-H. 2011. A Practical Synthesis of Zanamivir Phosphonate Congeners with Potent Anti-influenza Activity. *Journal of the American Chemical Society*, 133, 17959-17965.
- SHOICHET, B. K. 2006. Interpreting steep dose-response curves in early inhibitor discovery. *Journal of medicinal chemistry*, 49, 7274-7277.
- SHTYRYA, Y. A., MOCHALOVA, L. V. & BOVIN, N. V. 2009. Influenza Virus Neuraminidase: Structure and Function. *Acta Naturae*, 1, 26-32.
- SILVANI, H. L. 1947. The Use of Antibiotics in Surgical Practice. *California medicine*, 67, 171.

- SLIWOSKI, G., KOTHIWALE, S., MEILER, J. & LOWE, E. W. 2014. Computational Methods in Drug Discovery. *Pharmacological Reviews*, 66, 334-395.
- SMEE, D. F. & SIDWELL, R. W. 2002. Peramivir (BCX-1812, RWJ-270201): potential new therapy for influenza. *Expert opinion on investigational drugs*, 11, 859-869.
- SMITH, A. 2002. Screening for drug discovery: the leading question. *Nature*, 418, 453-459.
- SMITH, C. 2003. Drug target validation: Hitting the target. *Nature*, 422, 341-347.
- SOSA-ESTANI, S. & SEGURA, E. L. 2015. Integrated control of Chagas disease for its elimination as public health problem-A Review. *Memórias do Instituto Oswaldo Cruz*, 110, 289-298.
- SOUSA, A. F., GOMES-ALVES, A. G., BENÍTEZ, D., COMINI, M. A., FLOHÉ, L., JAEGER, T., PASSOS, J., STUHLMANN, F., TOMÁS, A. M. & CASTRO, H. 2014. Genetic and chemical analyses reveal that trypanothione synthetase but not glutathionylspermidine synthetase is essential for *Leishmania infantum*. *Free Radical Biology and Medicine*, 73, 229-238.
- SOUSA, S. F., FERNANDES, P. A. & RAMOS, M. J. 2006. Protein–ligand docking: current status and future challenges. *Proteins: Structure, Function, and Bioinformatics*, 65, 15-26.
- SPRATT, B. G. 1983. Penicillin-binding Proteins and the Future of  $\beta$ -Lactam Antibiotics: The Seventh Fleming Lecture. *Microbiology*, 129, 1247-1260.
- SPREER, A., KERSTAN, H., BÖTTCHER, T., GERBER, J., SIEMER, A., ZYSK, G., MITCHELL, T. J., EIFFERT, H. & NAU, R. 2003. Reduced release of pneumolysin by *Streptococcus pneumoniae* in vitro and in vivo after treatment with nonbacteriolytic antibiotics in comparison to ceftriaxone. *Antimicrobial agents and chemotherapy*, 47, 2649-2654.
- SPRING, D. R. 2005. Chemical genetics to chemical genomics: small molecules offer big insights. *Chemical Society Reviews*, 34, 472-482.
- SRIWILAJAROEN, N., MAGESH, S., IMAMURA, A., ANDO, H., ISHIDA, H., SAKAI, M., ISHITSUBO, E., HORI, T., MORIYA, S. & ISHIKAWA, T. 2016. A Novel Potent and Highly Specific Inhibitor against Influenza Viral N1–N9 Neuraminidases: Insight into Neuraminidase–Inhibitor Interactions. *Journal of medicinal chemistry*, 59, 4563-4577.
- STANLEY, P., SCHACHTER, H. & TANIGUCHI, N. 2009. N-glycans.
- STEIN, R. L. 2003. High-throughput screening in academia: the Harvard experience. *Journal of biomolecular screening*, 8, 615-619.
- STEINMETZER, T., HARDES, K., BÖTTCHER-FRIEBERTSHÄUSER, E. & GARTEN, W. 2015. Strategies for the Development of Influenza Drugs: Basis for New Efficient Combination Therapies. In: DIEDERICH, E. W. & STEUBER, H. (eds.) *Therapy of Viral Infections*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- STENCEL-BAERENWALD, J. E., REISS, K., REITER, D. M., STEHLE, T. & DERMODY, T. S. 2014. The sweet spot: defining virus-sialic acid interactions. *Nat Rev Micro*, 12, 739-749.
- STEVERDING, D. 2014. The history of Chagas disease. *Parasites & vectors*, 7, 1.
- STIERAND, K., MAAß, P. C. & RAREY, M. 2006. Molecular complexes at a glance: automated generation of two-dimensional complex diagrams. *Bioinformatics*, 22, 1710-1716.
- STOCKMAN, B. J., KOTHE, M., KOHLS, D., WEIBLEY, L., CONNOLLY, B. J., SHEILS, A. L., CAO, Q., CHENG, A. C., YANG, L. & KAMATH, A. V. 2009. Identification of Allosteric PIF-Pocket Ligands for PDK1 using NMR-Based Fragment Screening and 1H-15N TROSY Experiments. *Chemical biology & drug design*, 73, 179-188.
- STOCKWELL, B. R. 2000. Chemical genetics: ligand-based discovery of gene function. *Nature Reviews Genetics*, 1, 116-125.
- STOCKWELL, B. R. 2004. Exploring biology with small organic molecules. *Nature*, 432, 846-854.
- STREICHER, H. & BUSSE, H. 2006. Building a successful structural motif into sialylmimetics—cyclohexenephosphonate monoesters as pseudo-sialosides with promising inhibitory properties. *Bioorganic & medicinal chemistry*, 14, 1047-1057.
- STUDIER, F. W. 2005. Protein production by auto-induction in high-density shaking cultures. *Protein expression and purification*, 41, 207-234.

- STURA, E. A. & WILSON, I. A. 1991. Applications of the streak seeding technique in protein crystallization. *Journal of crystal growth*, 110, 270-282.
- SUI, J., HWANG, W. C., PEREZ, S., WEI, G., AIRD, D., CHEN, L.-M., SANTELLI, E., STEC, B., CADWELL, G. & ALI, M. 2009. Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nature structural & molecular biology*, 16, 265-273.
- SUMMA, V., PETROCCHI, A., BONELLI, F., CRESCENZI, B., DONGHI, M., FERRARA, M., FIORE, F., GARDELLI, C., GONZALEZ PAZ, O. & HAZUDA, D. J. 2008. Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. *Journal of medicinal chemistry*, 51, 5843-5855.
- SWINNEY, D. C. 2013. The contribution of mechanistic understanding to phenotypic screening for first-in-class medicines. *Journal of biomolecular screening*, 18, 1186-1192.
- SWINNEY, D. C. 2014. Opportunities for phenotypic screening in drug discovery. *Drug Discov. World (Fall)*, 33-40.
- SWINNEY, D. C. & ANTHONY, J. 2011. How were new medicines discovered? *Nature Reviews Drug Discovery*, 10, 507-519.
- TAILFORD, L. E., OWEN, C. D., WALSHAW, J., CROST, E. H., HARDY-GODDARD, J., LE GALL, G., DE VOS, W. M., TAYLOR, G. L. & JUGE, N. 2015. Discovery of intramolecular trans-sialidases in human gut microbiota suggests novel mechanisms of mucosal adaptation. *Nature communications*, 6.
- TALELE, T. T., KHEDKAR, S. A. & RIGBY, A. C. 2010. Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. *Current topics in medicinal chemistry*, 10, 127-141.
- TAYLOR, G. 1996. Sialidases: structures, biological significance and therapeutic potential. *Current opinion in structural biology*, 6, 830-837.
- TAYLOR, G. 2003. Influenza Virus Neuraminidase Inhibitors. *Handbook of Cell Signaling, Three-Volume Set*, 105.
- TAYLOR, G. & RUSSELL, R. 2010. Chapter 16 - Influenza Virus Neuraminidase Inhibitors A2 - Bradshaw, Ralph A. In: DENNIS, E. A. (ed.) *Handbook of Cell Signaling (Second Edition)*. San Diego: Academic Press.
- TAYLOR, R. D., JEWSEBURY, P. J. & ESSEX, J. W. 2002. A review of protein-small molecule docking methods. *Journal of computer-aided molecular design*, 16, 151-166.
- TERRETT, N. K., BELL, A. S., BROWN, D. & ELLIS, P. 1996. Sildenafil (Viagra™), a potent and selective inhibitor of type 5 cGMP phosphodiesterase with utility for the treatment of male erectile dysfunction. *Bioorganic & Medicinal Chemistry Letters*, 6, 1819-1824.
- TERSTAPPEN, G. C., SCHLÜPEN, C., RAGGIASCHI, R. & GAVIRAGHI, G. 2007. Target deconvolution strategies in drug discovery. *Nature Reviews Drug Discovery*, 6, 891-903.
- TEZE, D., HENDRICKX, J., DION, M., TELLIER, C., WOODS JR, V. L., TRAN, V. & SANEJOUAND, Y.-H. 2013. Conserved water molecules in family 1 glycosidases: a DXMS and molecular dynamics study. *Biochemistry*, 52, 5900-5910.
- THIEL, K. A. 2004. Structure-aided drug design's next generation. *Nat Biotech*, 22, 513-519.
- TODESCHINI, A. R. & HAKOMORI, S.-I. 2008a. Functional role of glycosphingolipids and gangliosides in control of cell adhesion, motility, and growth, through glycosynaptic microdomains. *Biochimica et biophysica acta*, 1780, 421-433.
- TODESCHINI, A. R. & HAKOMORI, S.-I. 2008b. Functional role of glycosphingolipids and gangliosides in control of cell adhesion, motility, and growth, through glycosynaptic microdomains. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1780, 421-433.
- TODESCHINI, A. R., MENDONÇA-PREVIATO, L., PREVIATO, J. O., VARKI, A. & VAN HALBEEK, H. 2000. Trans-sialidase from *Trypanosoma cruzi* catalyzes sialoside hydrolysis with retention of configuration. *Glycobiology*, 10, 213-221.
- TOPLISS, J. G. 1972. Utilization of operational schemes for analog synthesis in drug design. *Journal of Medicinal Chemistry*, 15, 1006-1011.

- TRAVING, C. & SCHAUER, R. 1998. Structure, function and metabolism of sialic acids. *Cellular and Molecular Life Sciences CMLS*, 54, 1330-1349.
- TRNKA, T. M. & GRUBBS, R. H. 2001. The development of L2X2Ru CHR olefin metathesis catalysts: an organometallic success story. *Accounts of Chemical Research*, 34, 18-29.
- TSAI, C.-J., DEL SOL, A. & NUSSINOV, R. 2009. Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. *Molecular Biosystems*, 5, 207-216.
- TU, A.-H. T., FULGHAM, R. L., MCCRORY, M. A., BRILES, D. E. & SZALAI, A. J. 1999. Pneumococcal surface protein A inhibits complement activation by *Streptococcus pneumoniae*. *Infection and immunity*, 67, 4720-4724.
- TUOMANEN, E., LIU, H., HENGSTLER, B., ZAK, O. & TOMASZ, A. 1985. The induction of meningeal inflammation by components of the pneumococcal cell wall. *Journal of Infectious Diseases*, 151, 859-868.
- UNAIDS 2015. Fact sheet - 2014: Global HIV statistics. .
- UNWIN, N. 1989. The structure of ion channels in membranes of excitable cells. *Neuron*, 3, 665-676.
- VAN BOECKEL, T. P., BROWER, C., GILBERT, M., GRENFELL, B. T., LEVIN, S. A., ROBINSON, T. P., TEILLANT, A. & LAXMINARAYAN, R. 2015. Global trends in antimicrobial use in food animals. *Proceedings of the National Academy of Sciences*, 112, 5649-5654.
- VAN EPPS, S. 2016. Carboxylic-Acid-Based Neuraminidase Inhibitors. *Bioactive Carboxylic Compound Classes: Pharmaceuticals and Agrochemicals*.
- VANDEKERCKHOVE, F., SCHENKMAN, S., DE CARVALHO, L. P., TOMLINSON, S., KISO, M., YOSHIDA, M., HASEGAWA, A. & NUSSENZWEIG, V. 1992. Substrate specificity of the *Trypanosoma cruzi* trans-sialidase. *Glycobiology*, 2, 541-548.
- VARGHESE GUPTA, S., GUPTA, D., SUN, J., DAHAN, A., TSUME, Y., HILFINGER, J., LEE, K.-D. & AMIDON, G. L. 2011. Enhancing the intestinal membrane permeability of zanamivir: a carrier mediated prodrug approach. *Molecular pharmaceutics*, 8, 2358-2367.
- VARGHESE, J. N. 1999. Development of neuraminidase inhibitors as anti-influenza virus drugs. *Drug development research*, 46, 176-196.
- VARKI, A. 1992. Diversity in the sialic acids. *Glycobiology*, 2, 25-40.
- VARKI, A. 2008. Sialic acids in human health and disease. *Trends Mol Med*, 14, 351-60.
- VARKI, A. & GAGNEUX, P. 2012. Multifarious roles of sialic acids in immunity. *Annals of the New York Academy of Sciences*, 1253, 16-36.
- VARKI, A. & SCHAUER, R. 2009. Sialic acids.
- VARKI, A. & SHARON, N. 2009. Chapter 1 Historical Background and Overview. *Essentials of Glycobiology*.
- VARKI, N. M. & VARKI, A. 2007. Diversity in cell surface sialic acid presentations: implications for biology and disease. *Laboratory investigation*, 87, 851-857.
- VAVRICKA, C. J., LIU, Y., KIYOTA, H., SRIWILAIJAROEN, N., QI, J., TANAKA, K., WU, Y., LI, Q., LI, Y. & YAN, J. 2013. Influenza neuraminidase operates via a nucleophilic mechanism and can be targeted by covalent inhibitors. *Nature communications*, 4, 1491.
- VELEC, H. F., GOHLKE, H. & KLEBE, G. 2005. DrugScoreCSD knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *Journal of medicinal chemistry*, 48, 6296-6303.
- VENKATESH, S. & LIPPER, R. A. 2000. Role of the development scientist in compound lead selection and optimization. *Journal of Pharmaceutical Sciences*, 89, 145-154.
- VENTOLA, C. L. 2015. The antibiotic resistance crisis: part 1: causes and threats. *Pharmacy and Therapeutics*, 40, 277.
- VERDONK, M. L., COLE, J. C., HARTSHORN, M. J., MURRAY, C. W. & TAYLOR, R. D. 2003. Improved protein-ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics*, 52, 609-623.
- VERMA, R. K. & GARG, S. 2001. Drug delivery technologies and future directions. *Pharmaceut. Technol. On-Line*, 25, 1-14.

- VERWEIJ, J., DE JONGE, M., ESKENS, F. & SLEIJFER, S. 2012. Moving molecular targeted drug therapy towards personalized medicine: issues related to clinical trial design. *Molecular oncology*, 6, 196-203.
- VEZINA, C., KUDELSKI, A. & SEHGAL, S. 1975. Rapamycin (AY-22,989), a new antifungal antibiotic. I. Taxonomy of the producing streptomycete and isolation of the active principle. *The Journal of antibiotics*, 28, 721-726.
- VIMR, E. R. 1994. Microbial sialidases: does bigger always mean better? *Trends in microbiology*, 2, 271-277.
- VIMR, E. R., KALIVODA, K. A., DESZO, E. L. & STEENBERGEN, S. M. 2004. Diversity of Microbial Sialic Acid Metabolism. *Microbiology and Molecular Biology Reviews*, 68, 132-153.
- VON ITZSTEIN, M. 2007. The war against influenza: discovery and development of sialidase inhibitors. *Nat Rev Drug Discov*, 6, 967-974.
- VON ITZSTEIN, M., DYASON, J. C., OLIVER, S. W., WHITE, H. F., WU, W.-Y., KOK, G. B. & PEGG, M. S. 1996. A study of the active site of influenza virus sialidase: an approach to the rational design of novel anti-influenza drugs. *Journal of medicinal chemistry*, 39, 388-391.
- VON ITZSTEIN, M. & THOMSON, R. 2009. Anti-influenza drugs: the development of sialidase inhibitors. *Antiviral strategies*. Springer.
- VON ITZSTEIN, M., WU, W.-Y., KOK, G. B., PEGG, M. S., DYASON, J. C., JIN, B., PHAN, T. V., SMYTHE, M. L., WHITE, H. F., OLIVER, S. W., COLMAN, P. M., VARGHESE, J. N., RYAN, D. M., WOODS, J. M., BETHELL, R. C., HOTHAM, V. J., CAMERON, J. M. & PENN, C. R. 1993. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*, 363, 418-423.
- VOUGIOUKALAKIS, G. C. & GRUBBS, R. H. 2009. Ruthenium-Based Heterocyclic Carbene-Coordinated Olefin Metathesis Catalysts†. *Chemical reviews*, 110, 1746-1787.
- WADE, R. C. 1997. 'Flu'and structure-based drug design. *Structure*, 5, 1139-1145.
- WALSH, C. 1999. Deconstructing vancomycin. *Science*, 284, 442-443.
- WANG, Q., ZHENG, M., HUANG, Z., LIU, X., ZHOU, H., CHEN, Y., SHI, T. & ZHANG, J. 2012. Toward understanding the molecular basis for chemical allosteric modulator design. *Journal of Molecular Graphics and Modelling*, 38, 324-333.
- WARREN, L. 1963. The distribution of sialic acids in nature. *Comparative biochemistry and physiology*, 10, 153-171.
- WATTS, A. G., DAMAGER, I., AMAYA, M. L., BUSCHIAZZO, A., ALZARI, P., FRASCH, A. C. & WITHERS, S. G. 2003. Trypanosoma cruzi Trans-sialidase Operates through a Covalent Sialyl-Enzyme Intermediate: Tyrosine Is the Catalytic Nucleophile. *Journal of the American Chemical Society*, 125, 7532-7533.
- WATTS, A. G., OPPEZZO, P., WITHERS, S. G., ALZARI, P. M. & BUSCHIAZZO, A. 2006. Structural and kinetic analysis of two covalent sialosyl-enzyme intermediates on Trypanosoma rangeli sialidase. *Journal of Biological Chemistry*, 281, 4149-4155.
- WEBSTER, R. G., BEAN, W. J., GORMAN, O. T., CHAMBERS, T. M. & KAWAOKA, Y. 1992. Evolution and ecology of influenza A viruses. *Microbiological reviews*, 56, 152-179.
- WEISS, P., TIETZE, F., GAHL, W., SEPPALA, R. & ASHWELL, G. 1989. Identification of the metabolic defect in sialuria. *Journal of Biological Chemistry*, 264, 17635-17636.
- WELLMAN-LABADIE, O. & ZHOU, Y. 2010. The US Orphan Drug Act: rare disease research stimulator or commercial opportunity? *Health Policy*, 95, 216-228.
- WENTHUR, C. J., GENTRY, P. R., MATHEWS, T. P. & LINDSLEY, C. W. 2014. Drugs for allosteric sites on receptors. *Annual review of pharmacology and toxicology*, 54, 165.
- WHITESIDES, G. M. & WONG, C. H. 1985. Enzymes as catalysts in synthetic organic chemistry [new synthetic methods (53)]. *Angewandte Chemie International Edition in English*, 24, 617-638.
- WHITTAKER, P. A. 2004. The role of bioinformatics in target validation. *Drug Discovery Today: Technologies*, 1, 125-133.
- WHO. Media centre [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs375/en/> [Accessed].

- WHO 2010. First WHO report on neglected tropical diseases: Working to overcome the global impact of neglected tropical diseases.
- WHO 2012. Accelerating work to overcome the global impact of neglected tropical diseases: a roadmap for implementation: executive summary.
- WHO. 2015. *Media centre: Antibiotic resistance* [Online]. Available: <http://www.who.int/mediacentre/factsheets/antibiotic-resistance/en/> [Accessed].
- WHO. 2016a. *Global Health Observatory (GHO) data* [Online]. Available: <http://www.who.int/gho/tb/en/> [Accessed].
- WHO. 2016b. *Media Centre* [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs259/en/> [Accessed].
- WHO. 2016c. *Media Centre* [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs199/en/> [Accessed].
- WILLETT, P. 2006. Similarity-based virtual screening using 2D fingerprints. *Drug Discovery Today*, 11, 1046-1053.
- WILLETT, P., BARNARD, J. M. & DOWNS, G. M. 1998. Chemical Similarity Searching. *Journal of Chemical Information and Computer Sciences*, 38, 983-996.
- WITCHEL, H. J. 2011. Drug-induced hERG Block and Long QT Syndrome. *Cardiovascular Therapeutics*, 29, 251-259.
- WITT, A., MACDONALD, N. & KIRKPATRICK, P. 2004. Memantine hydrochloride. *Nature Reviews Drug Discovery*, 3, 109-110.
- WLODAWER, A., MINOR, W., DAUTER, Z. & JASKOLSKI, M. 2008. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *The FEBS journal*, 275, 1-21.
- WONG, C. 1989. Enzymatic catalysts in organic synthesis. *Science*, 244, 1145-1152.
- WONG, S.-S. & WEBBY, R. J. 2013. Traditional and new influenza vaccines. *Clinical microbiology reviews*, 26, 476-492.
- WORKMAN, P. & COLLINS, I. 2010. Probing the probes: fitness factors for small molecule tools. *Chemistry & biology*, 17, 561-577.
- WU, P., NIELSEN, T. E. & CLAUSEN, M. H. 2016. Small-molecule kinase inhibitors: an analysis of FDA-approved drugs. *Drug Discovery Today*, 21, 5-10.
- XING, L. & GLEN, R. C. 2002. Novel Methods for the Prediction of logP, pKa, and logD. *Journal of Chemical Information and Computer Sciences*, 42, 796-805.
- XU, G., KIEFEL, M. J., WILSON, J. C., ANDREW, P. W., OGGIONI, M. R. & TAYLOR, G. L. 2011. Three *Streptococcus pneumoniae* sialidases: three different products. *Journal of the American Chemical Society*, 133, 1718-1721.
- XU, G., POTTER, J. A., RUSSELL, R. J., OGGIONI, M. R., ANDREW, P. W. & TAYLOR, G. L. 2008. Crystal structure of the NanB sialidase from *Streptococcus pneumoniae*. *Journal of molecular biology*, 384, 436-449.
- YADAV, M. & SINGH, G. 2013. Virtual screening of ligand molecules for target protein CYP26A1 by using AutoDock-Vina. *Int J Innov Res Sci Eng Technol*, 2, 4915-20.
- YAMAGUCHI, K., HATA, K., KOSEKI, K., SHIOZAKI, K., AKITA, H., WADA, T., MORIYA, S. & MIYAGI, T. 2005. Evidence for mitochondrial localization of a novel human sialidase (NEU4). *Biochemical Journal*, 390, 85-93.
- YANG, L., CONNARIS, H., POTTER, J. A. & TAYLOR, G. L. 2015. Structural characterization of the carbohydrate-binding module of NanA sialidase, a pneumococcal virulence factor. *BMC Structural Biology*, 15, 15.
- YANG, X., STEUKERS, L., FORIER, K., XIONG, R., BRAECKMANS, K., VAN REETH, K. & NAUWYNCK, H. 2014. A beneficiary role for neuraminidase in influenza virus penetration through the respiratory mucus. *PLoS one*, 9, e110026.
- YANG, Y., ADELSTEIN, S. J. & KASSIS, A. I. 2009. Target discovery from data mining approaches. *Drug discovery today*, 14, 147-154.

- ZARTLER, E. R. & SHAPIRO, M. J. 2005. Fragonomics: fragment-based drug discovery. *Current opinion in chemical biology*, 9, 366-370.
- ZHANG, Q., YANG, J., LIANG, K., FENG, L., LI, S., WAN, J., XU, X., YANG, G., LIU, D. & YANG, S. 2008. Binding interaction analysis of the active site and its inhibitors for neuraminidase (N1 subtype) of human influenza virus by the integration of molecular docking, FMO calculation and 3D-QSAR CoMFA modeling. *Journal of chemical information and modeling*, 48, 1802-1812.
- ZIMMERMAN, R. K., RUBEN, F. L. & AHWESH, E. R. 1997. Influenza, influenza vaccine, and amantadine/rimantadine. *Journal of family practice*, 45, 107-124.