UNIVERSITY OF BIRMINGHAM

Design and synthesis of multivalent receptors towards the detection of biologically relevant carbohydrates

by

Alice Di Pasquale

A thesis submitted to The University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

School of Chemical Engineering

College of Physical Sciences and Engineering

The University of Birmingham

August 2021

Abstract

Changes in the glycan phenotypes of cells is one of the main hallmarks of cancer. Glycans are carbohydrate chains linked to protein carriers, forming glycoproteins. Under or overexpression of glycans has been observed in cancer cells, alongside with expression of entirely new glycans. Many biomarkers currently employed in cancer diagnostics are proteins which are also expressed in benign conditions, resulting in high rates of false positive diagnoses and the associated negative consequences on patients. Conversely to currently employed protein biomarkers, there are particular glycoforms, expressed solely by cancer cells, which are tumour specific and therefore present great potential in cancer diagnostics. In particular, a higher degree of sialylation has been closely associated to cancer. This results in a higher expression of sialosides, glycans terminating with a sialic acid unit. The sialic acid unit is linked to the carbohydrate chain in α2-3 or α2-6 fashion with these glycoforms being expressed by altered and healthy cells, respectively. Thus, the selective detection of α2-3 sialosides, over the α2-6 glycoforms, can be exploited in cancer diagnostics. Furthermore, sialic acid, in its unbound free form, is also a relevant biomarker for certain neurodegenerative diseases and alcoholism. As a result, sialic acid is an important saccharide, for which detection is key in a variety of pathological conditions.

The detection of carbohydrates for diagnostics purposes can be achieved with natural or synthetic lectins. Among these, boron-based receptors, such as phenylboronic acids and benzoboroxoles, present the key advantage of forming covalent complexes with carbohydrates. Boron-based receptors have been known for decades for their sugar-binding properties and have been extensively employed as receptors for neutral monosaccharides. Nonetheless, there is no consensus in the literature with regards to the way these receptors bind to sialic acid. Therefore, it was pivotal to accurately identify the site via which boron-based receptors bind to sialic acid.

Different analogues of sialic acid were employed in calorimetric and spectroscopic studies allowing the unequivocal identification of the α-hydroxyacid group as the only site of binding. This key finding enabled the design and synthesis of small functionalised benzoboroxole receptors to selectively target this monosaccharide. In particular, the functionalisation with charged groups afforded benzoboroxole receptors which display up to a 4.5-fold activity compared to the non-functionalised benzoboroxole. This study resulted in a binding model being postulated highlighting the value of multiple cooperative interactions in sugar recognition.

Following on from this, the focus was moved on the detection of small sialosides representing biologically relevant sialylated glycans. Sialylated trisaccharide epitopes, SAα2-3LacNAc and SAα2-6LacNAc, were enzymatically synthesised to be employed as target molecules. Furthermore, dynamic combinatorial chemistry (DCC) was reviewed as a technique for the generation of carbohydrate receptors. In order to target the aforementioned oligosaccharides, a scaffold-based approach was adopted to provide interactions with a high degree of multivalency. A peptidomimetic scaffold was designed and synthesised and a series of sugar-binding small molecules were selected creating the foundations for the future application of the DCC approach in glycosensing.

Acknowledgements

Firstly, I would like to thank my supervisor, Professor Paula M. Mendes. Her guidance and support during the past four years has been crucial for the success of this work and my development as a researcher. I also would like to thank my second supervisor Professor James Tucker. Thanks also to Prof. Dr. Geert-Jan Boons and his team for hosting me for 2 months in their lab at Utrecht University. I would also like to thank The University of Birmingham, the ERC and the Erasmus Program for funding my research and my placement abroad.

A big thank you to the two Postdoctoral Researchers that taught me the most with regards to synthetic chemistry, Dr. Stefano Tommasone and Dr. Marcos Fernandez-Villamarin. Their support and help within the lab has been indispensable.

A thank you to all the members of the Mendes Group that have been with me during this journey; Barbara Santos-Gomes, Giuseppe Di Palma, Francia Allabush, Kamlesh Patel, Joshua Norman, Barbara Simoes, Yazmin Tagger, Dr. Eduardo Anaya-Plaza, Phillipa Mitchell, Dr. Setareh Vafaei, Dr. Seyed Tabaei, Dr. Craig Ward, Dr. Nasim Mahmoodi, Dr. Joshua Simon Gibson, Laura Brooks, Laura Buccoli, Pushpa Patel, Dario Bazzoli, Miguel Alena-Rodriguez, Charlie Keene. Thank you for your support and indelible moments.

A special thank you to my family in Italy, my mom, Mariella, and my sister, Greta. And a big thank you to my partner Josh, for the support, help and encouragement throughout this time.

Table of Contents

List of definitions and abbreviations

AGA	Anti-Glycan Antibody
Arg	Arginine
ARS	Alizarin red S
Asn	Asparagine
Asp	Aspartic acid
B_0	Magnetic field
B4GalT1	β1-4Galactosyltransferase 1
BA	Boronic acid
BB	Building Block
BOB	Benzoboroxole
Boc	*tert*-Butoxycarbonyl protecting group
Boc_2O	Di-tert-butyl dicarbonate
BPH	Benign Prostate Hyperplasia
BSA	Bovine serum albumin
Cbz	Benzyl carbamate
CD_3COOD	Deuterated acetic acid
CIAP	Calf-intestinal alkaline phosphatase
CMP	Cytidine 5′-monophosphate
COSY	Correlation Spectroscopy
δ	Chemical shift
D_2O	Deuterated water
DCC	Dynamic Combinatorial Chemistry

DCL	Dynamic Combinatorial Library
DEAE	Diethylaminoethanol
DIBAL	Diisobutylaluminium hydride
DIEA	N,N-Diisopropylethylamine
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DP	Differential Power
dt	Double triplet
E	Energy
ECM	Extracellular matrix
EDCI	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EEC	Enthalpy-Entropy Compensation
ESI	Electrospray ionisation
FF	Fast flow
FID	Free Induction Decay
Fmoc	Fluorenylmethyloxycarbonyl protecting group
Fuc	Fucose
γ	Gyromagnetic ratio
G	Free energy of Gibbs
Gal	Galactose
GalNAc	N-Acetylgalactosamine

GalβMe	Methyl-β-D-galactoside
Glc	Glucose
GlcA	Glucuronic acid
GlcNAc	N-Acetylglucosamine
Gln	Glutamine
GLS	Glycosphingolipid
Gly	Glycine
H	Enthalpy
h	Planck constant
H_2SO_4	Sulfuric acid
HCl	Hydrochloric acid
His	Histidine
HOBt	Hydroxybenzotriazole
HPLC	High Performance Liquid Chromatography
HSQC	Heteronuclear Single Quantum Coherence
I	Nuclear spin quantum number
IC_{50}	Half-maximal inhibitory concentration
IdoA	Iduronic acid
IgG	Immunoglobulin G
IgM	Immunoglobulin M
ISSD	Infantile Sialic acid Storage Disease
ITC	Isothermal Titration Calorimetry
J	Coupling constant

K_2HPO_4	Potassium phosphate dibasic
KH_2PO_4	Potassium phosphate monobasic
K_a	Binding constant
K_d	Dissociation constant
KDN	2-Keto-3-deoxy-D-glycerol-D-galacto-nononic acid
L	Ligand
LacNAc	N-Acetyllactosamine
LB	Lysogen broth
LC	Liquid Chromatography
LFA	*Limax Flavus* Agglutinin
LG	Leaving group
Lys	Lysine
μ	Magnetic moment
m/z	Mass to charge ratio
MAH	*Maackia Amurensis* Hemagglutinin
MAL	*Maackia Amurensis* Leukoagglutinin
Man	Mannose
$MgCl_2$	Magnesium chloride
$MgSO_4$	Magnesium sulfate
$MnCl_2$	Manganese chloride
MRI	Magnetic Resonance Imaging
MS	Mass Spectrometry
v	Frequency

n	stoichiometry of binding, number of binding sites
$NaHCO_3$	Sodium bicarbonate
NaOH	Sodium hydroxide
Neu5,9Ac2	N-Acetyl-9-O-acetylneuraminic acid
Neu5Ac	N-Acetylneuraminic acid
Neu5Gc	N-Glycolyl-neuraminic acid
Ni^{2+}-NTA	Nickel-nitrilotriacetic acid
NMR	Nuclear Magnetic Resonance
OD	Optical density
PAPBA	3-(Propionamido)phenylboronic acid
PB	Phosphate buffer
PBA	Phenylboronic acid
PCa	Prostate cancer
Pd/C	Palladium on carbon
Phe	Phenylalanine
PmST1	*Pasteurella multocida* sialyltransferase
PR	Promoter
PSA	Prostate Specific Antigen
Q	Heat of binding
R	Receptor, molar gas constant
RNA	Ribonucleic acid
RNase	Ribonuclease
RPM	Revolutions per minute

rt	Room temperature
S	Entropy, orientation of spin
SA	Sialic Acid
SDS-PAGE	Sodium dodecyl sulphate–polyacrylamide gel electrophoresis
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
Ser	Serine
Sialyl Tn	Sialylated Thomsen-nouvelle antigen
SLe^a	Sialyl Lewis a
SLe^x	Sialyl Lewis x
SLe^y	Sialyl Lewis y
S_N	Nucleophilic substitution
SNA	*Sambucus Nigra* lectin
ST	Sialyltrasferase
T	Temperature
T	Thomsen-Friedenreich antigen
td	Triplet of doublet
Thr	Threonine
TLC	Thin layer chromatography
Tn	Thomsen-nouvelle antigen
TOCSY	Total Correlation Spectroscopy
TOF	Time of flight
TRIS	Tris(hydroxymethyl)aminomethane
Trp	Tryptophan

Tyr	Tyrosine
UDP	Uridine diphosphate
UV	Ultraviolet
V_0	Volume of the sample cell
WGA	Wheat Germ Agglutinin
Xyl	Xylose

Chapter 1 - Introduction

This Chapter focuses on introducing glycans and the receptors employed for their detection. Glycans are polysaccharide chains, which present a great variety of structures, attached to protein carriers forming glycoproteins. Changes in the glycan phenotype have been observed in all types of tumours and have been identified as a hallmark of cancer. Certain glycan modifications and particular sugar motifs have been closely associated to cancer and can be exploited as biomarkers in cancer diagnostics. Increased sialylation has been found in many different types of tumours providing a higher degree of glycans terminating with a sialic acid unit. The detection of glycans for diagnostic purposes can be achieved with natural or synthetic receptors. Lectins and antibodies are natural receptors which can afford reasonable selectivity, however they also present many disadvantages. Synthetic tools are generally preferred and have been extensively reviewed in this Chapter, including covalent and non-covalent receptors. Particular focus has been given to sialic acid detection, given its biological relevance as a free monosaccharide and as the terminating unit in glycans. Further to this, a brief overview on dynamic combinatorial chemistry, as a technique with potential to aid in receptor development, is provided.

1.1 Glycans

1.1.1 Composition and structure of glycans

Glycans are oligo and polysaccharides, which can be found conjugated to protein and lipid scaffolds or as free chains, with key roles in many biological functions. Glycans are exposed on the surface of cells or secreted into the extracellular matrix (ECM)[1]. They mediate a series

of extracellular processes, such as cell to cell and cell to matrix recognition, cell activation, growth, adhesion, signal transduction and motility[2]. The carbohydrate motifs exposed on the cells surfaces also act as receptors for pathogens, such as viruses and bacteria[3]. The aforementioned cellular functions are a result of the interactions between glycans and lectins, which are proteins able to bind glycans present on the surface of cells or in the extracellular matrix. Furthermore, glycans are also involved in intracellular processes such as correct folding, activation, secretion and degradation of glycoproteins[4].

Glycans are carbohydrate chains formed by different combinations of 10 monosaccharides (Figure 1.1)[5]; glucose (Glc), N-acetyl glucosamine (GlcNAc), glucuronic acid (GlcA), galactose (Gal), N acetylgalactosamine (GalNAc), mannose (Man), xylose (Xyl), iduronic acid (IdoA), fucose (Fuc) and N-acetylneuraminic acid, also known as sialic acid (SA)[6].

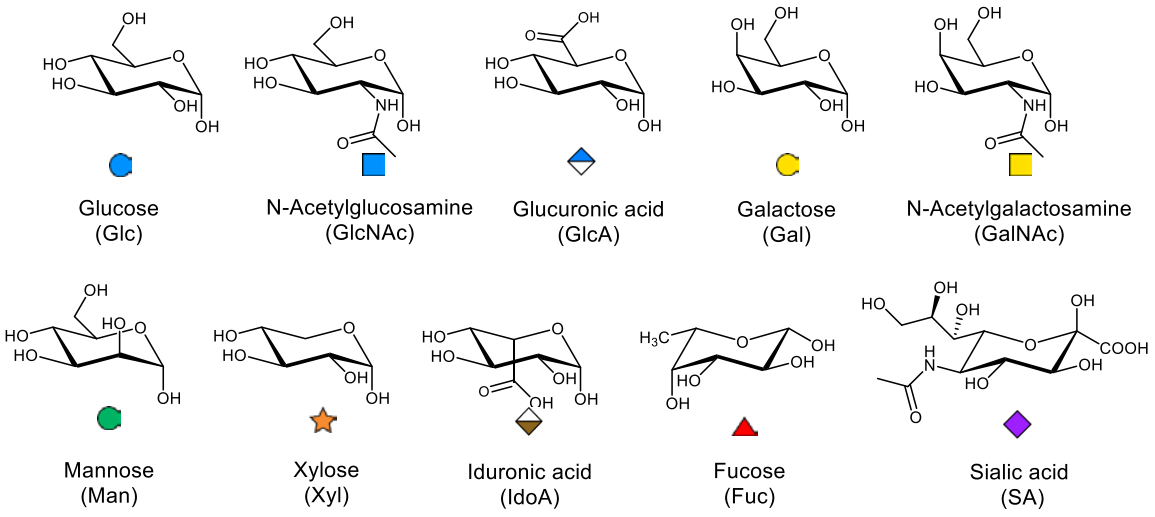


Figure 1.1. Glycan-related monosaccharide structures and associated symbols.

The monosaccharides are linked one to the other in a post-translational process called glycosylation to form di-, oligo- or polysaccharides. The glycosylation is mediated by two main classes of enzymes: glycosyltransferases and glycosidases. The former enzymes catalyse the formation of the glycosidic linkage between two saccharide units, whilst the latter mediates the hydrolysation of the linkage. Each of the over 200 different types of glycosyltransferases catalyse the formation of a certain glycosidic linkage between specific sugar residues[7]. The glycosidic bonds are distinguished in α or β based on the stereochemistry of the anomeric carbon involved in the linkage[8]. For instance, GalNAc and GlcNAc are bound to the protein scaffold through a α and β linkage, respectively[6]. In addition, the glycosidic linkage can be formed with different hydroxyl groups of a given sugar, hence the formation of different isomers. For example, galactose can be linked through position 3 or position 6 to sialic acid, giving two different sialylated epitopes, the α2-3 and the α2-6[6]. Furthermore, a certain sugar residue can be involved simultaneously in more than one glycosidic linkage, which results in branched glycans. For instance, a galactose residue can be linked simultaneously to a GalNAc and to a sialic acid unit through position 1 and 3, giving a branched epitope[6]. Following the glycosylation, the saccharide units can undergo modifications and groups such as phosphate, acetyl and sulfate are introduced. All of these processes are responsible for the great variety of glycans expressed by cells and thus for the immense information content stored, which is referred to as "sugar code"[9].

Furthermore, glycans bound to carrier molecules, proteins or lipids, form glycoconjugates consisting of glycoproteins and glycolipids, respectively. In glycoproteins the sugar chain is anchored to an asparagine residue, to form N-glycans, or alternatively to a serine or threonine residue to give O-glycans. Both N- and O-glycans can be linear or highly branched. The N-glycans are divided into high-mannose type, hybrid type and complex type (Figure 1.2). All

three types present the same core sequence formed by Man and GlcNAc residues[10]. The high-mannose type presents solely mannose units linked to the core, whilst other N-glycans are more heterogenic. The hybrid type N-glycans are mono- or bi-antennaries[11], whilst the complex type are highly branched with up to six branches[10]. N-glycans have key roles in the biosynthesis of glycoproteins, in particular they aid the correct folding[12]. Conversely, mucins and proteoglycans are O-glycans. Mucins are highly branched and involved in cell adhesion and protection against pathogens[13], whilst proteoglycans are linear and participate in cell-matrix interaction, growth and inflammation[1].

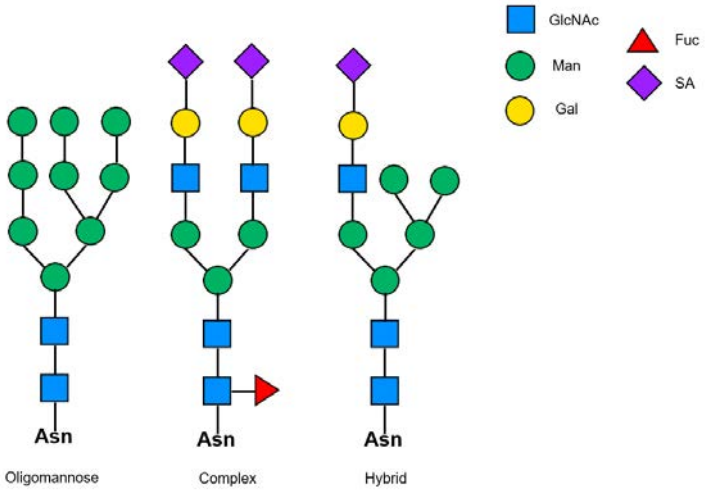


Figure 1.2. Structures of oligomannose, complex and hybrid N-glycans.

Glycolipids present the glycan epitopes linked to lipid scaffolds. Glycosphingolipids (GLS) are glycolipids with a sphingoid or a ceramide scaffold and are divided into neutral and acidic[14]. Neutral GLS include globosides and cerebrosides, whilst gangliosides and sulfatides are acidic GLS[15]. Gangliosides containing sialic acid residues, such as the GM1, are particularly relevant in the nervous system, where they participate in the upkeep of the system and in the neural transmission[14].

1.1.2 Glycans in cancer cells

The change in the glycan phenotype of cancer cells is a hallmark alteration common to all types of tumours. The altered expression of glycans in cancer cells was first observed by Meezan *et al*.[16]. It was then established that aberrant glycosylation occurs in all types of cancer, although the type of alteration differs from tumour to tumour[17]. The alteration of glycans in cancerous cells includes under- or over-expression of glycans normally present in cells or, alternatively, expression of new glycan structures[8]. These modifications are caused by genetic and epigenetic dysregulation of glycogenes which control the expression of glycosyltransferases and glycosidases[3]. The alteration of glycans in glycoproteins or glycolipids results in conformational changes of the whole macromolecule causing variation in the biological functions[6], including variation of cell to cell and cell to matrix interactions[3].

The above-described phenomena are the basis of all stages of the tumour progression: growth, proliferation, invasion, angiogenesis and metastasis[1]. The proliferation, or tumour growth, is caused by anomalous N- and O-glycans. For instance, modified gangliosides are responsible for the growth promotion in melanoma[18], whilst heparan-sulfate proteoglycans modifications have been found to promote proliferation and invasion in pancreatic and ovarian cancers[19]. Following the proliferation phase, the malignant cells detach from the tumour mass and the extracellular matrix in order to migrate to different tissues[1]. In this second phase, the tumour invasion phase, the cell to cell and cell to matrix adhesion are disrupted by alteration of the glycans on the cell surface leading to higher cell mobility[1]. The main changes include increased multiantennary N-glycosylation and sialylation[4]. The increased sialylation reduces the cellular adhesion due to the repulsion present between negatively charged SA residues[1]. Hence, the detachment, release and invasion of the cancer cells[20]. Furthermore, the later stages of the tumour progression are also promoted by altered glycans. For instance, the formation of new

blood vessels, known as angiogenesis, is promoted by proteoglycans[21]. The metastatic process is mediated by O-glycans and specific lectins (*e.g.* siglecs, galectins and selectins) allowing the interaction of tumour cells with blood and epithelial cells[22-23]. The tumour progression is the result of the events described above and is therefore closely related to the expression of altered glycans.

One of the main modifications correlated with cancer growth is the expression of highly β1-6 branched motifs[3],[24] determining an increase in the number of glycan terminal ends. Another hallmark of cancer progression is an over-expression of sialyltransferases and fucosyltransferases[8]. The former catalyses the addition of sialic acid to the terminal end of glycans, whilst the latter mediate the addition of fucose on the glycan core. In particular, certain sialylated epitopes are intimately associated with tumours, these include N-glycan motifs such as sialyl Lewis a (SLe^a), sialyl Lewis x (SLe^x) and polysialic acid[8]. The sialyl Lewis a epitope is formed by 4 sugar units. To the terminal unit, a sialic acid residue, a galactose unit is linked in α2-3 fashion. The galactose residue is β1-4 linked to N-acetylglucosamine which bears a fucose residue in α1-4[25]. The sialyl Lewis x antigen consists of the same carbohydrate sequence as SLe^a but with the fucose unit linked in α1-3 fashion[26]. Polysialic acids are polymers formed exclusively of α2-8-linked SA units[27].

Conversely, O-glycan modifications often result in truncated sequences, such as the Thomsen-Friedenreich (T) antigen, the Thomsen-nouvelle antigen (Tn) and its sialylated counterpart (sialyl Tn). The Tn antigen consists of one single residue of GalNAc liked to the protein scaffold through a Ser or Thr residue. The β1-3 addition of galactose or the α2-6 sialylation of N-GalNAc gives the T and sialyl Tn epitopes, respectively[28]. Over 80% of tumours are reported to express either the Tn antigen or its sialylated derivative[29].

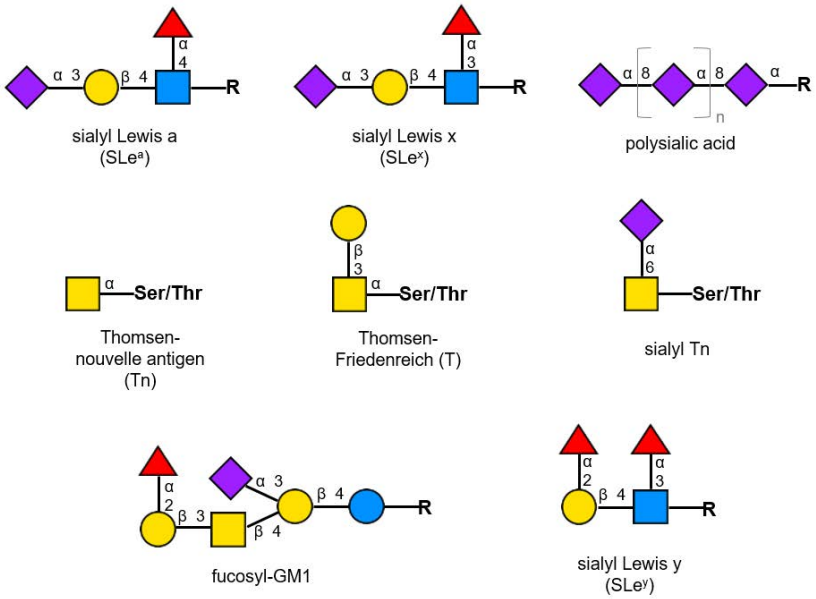


Figure 1.3. Glycan epitopes associated with cancer.

Furthermore, an increase in the fucosylation has been also observed in cancerous cells. The main fucosylated epitopes include Fucosyl-GM1 and Lewis y (Figure 1.3)[8]. The fucosyl-GM1 is biantennary with one end terminating with a fucose and the other with a SA unit. The Lewis y antigen is a difucosylated epitope structurally similar to the Lewis x motif. The aforementioned sialylated, fucosylated and T epitopes are associated to a variety of cancers and are indicative of the growth and metastatic capability of tumours[3]. In particular, the Lewis motifs are common to many epithelial cancers[17], such as pancreatic[30], lung[31], colon[32], gastric[33], breast, prostate, ovarian[34]. High degree of β1-6 N-branching or high levels of Lewis antigens indicate a poor prognosis[35], as in pancreatic, lung and gastric cancer[34].

1.1.3 Glycans as biomarkers

Tumour biomarkers are molecules produced by cancer cells and secreted into biological fluids which can be detected and analysed[36]. The early detection of tumours is pivotal for the survival of the patients. When the detection occurs in latter stages of the tumour progression the survival

rate is significantly decreased[36]. The early detection of a tumour is often hampered by the lack of sensitivity and specificity of the selected biomarkers resulting in misleading results, such as false negative or false positive results[36]. Indeed, many of the biomarkers currently in use are not tumour specific and are instead expressed also by other types of tumours or in benign conditions. For instance, the human chorionic gonadotropin-β biomarker for germ cells tumour is also expressed by the placenta during pregnancy. The human pancreatic ribonuclease 1 (RNase 1) is present in high levels in pancreatic cancer, but also in liver and renal disfunctions [37-39]. High levels of the Prostate Specific Antigen (PSA) are not only found in prostate cancer (PCa)[40] but also in benign conditions such as benign prostate hyperplasia (BPH) and prostatitis which present enlargement and inflammation of the prostate, respectively[41]. Therefore, the current diagnostic test for prostate cancer based on the PSA levels produces high percentages of false positive and false negatives results with serious consequences for the patients, such as biopsy of healthy patients and under detection of ill patients[42].

In order to avoid the aforementioned issues, the chosen biomarker should possess high sensitivity and specificity. Therefore, it should not be based on the levels or activity of a specific protein but on the detection of individual glycoforms as each type of tumour expresses a unique array of N- or O-glycans[34]. Hence, the carbohydrate phenotype of a specific tumour can be employed as a more specific biomarker. Among all the alterations, sialylated motifs are particularly relevant in cancer diagnostic, as they have been found in a wide range of cancers including brain, breast, colorectal, gastric, lung, ovarian, pancreatic, prostatic[17].

1.1.4 Sialylated glycans

As described above, sialylated glycans play an important role in tumour progression. An increase in the degree of sialylation at terminal end of glycoproteins has been observed in many tumours. The sialylation levels of glycans are governed by two main factors; the availability of its active precursor, CMP-sialic acid, and the activity of sialyltransferases and sialidases enzymes[43]. The sialidases are responsible for the cleavage of sialic acid residues, whilst sialyltranferases catalyse the addition of SA at the glycan terminal end. There are four main sialidases (NEU1-4)[44] and approximatively 20 different types of sialyltrasferases which are divided in four classes based on the glycosidic linkage type[45]. The sialyltrasferases ST3Gal1-6 mediate the α2-3 addition of sialic acid to galactose-terminating glycans. The sialyltranferases ST6Gal1 and ST6GalNAc1-6 catalyse the α2-6 addition of sialic acid to galactose or N-acetylgalactosamine, respectively. The ST8Sia1-5 enzymes allow the formation of polysialylated chains via α2-8 addition of sialic acid to another SA residue. Approximatively 10 of the 20 sialyltranferases have been associated with tumours as they altered activity leads to the expression of modified glycoforms[46]. For instance, the upregulation of ST6Gal1 or ST6GalNAc enzymes in cancer cells causes an increase in α2-6 sialylated epitopes[43]. The α2-6 addition of SA to Tn antigen gives the sialyl Tn epitope, thus preventing further chain elongation and resulting in an increase of truncated O-glycoproteins in tumours[47]. High levels of sialyl Tn are associated with cancers that typically have a poor prognosis, such as gastric and ovarian cancer[48]. Conversely, the upregulation of the ST3Gal family of enzymes causes the overexpression of sialyl Lewis a and x as observed in colon, stomach, pancreatic, lung and liver cancers[49].

High levels of sialyl Lewis x and sialyl Tn, which present sialic acid linked α2-3 and α2-6, respectively, are associated with breast cancer[50]. The former is present at high levels in

advanced and metastatic conditions, whilst the latter is associated with decreased survival. On the other hand, the upregulation of the enzyme ST3Gal 3, which generates α2-3 sialylated epitopes, plays a role in pancreatic adenocarcinoma metastasis[51]. An increase of α2-3 sialylated glycans was also observed in PSA glycoforms expressed in patients with prostate cancer[52-53]. Conversely, patients with benign hyperplasia present predominately SA linked in the α2-6 fashion. Therefore, the ability to distinguish between different sialylated glycans could be exploited to provide superior and more accurate diagnostic tests, which are lacking in the field of cancer diagnostics.

1.1.5 Sialic acid family

Herein the term sialic acid has been used to describe one specific monosaccharide, the N-acetylneuraminic acid (SA, Neu5Ac)[46]. However, the term sialic acid is sometimes used to refer to a family of about 50 different negatively charged neuraminic acid derivatives. The main component of this family is the N-acetylneuraminic acid (*e.g.* sialic acid) which consist of a six-membered ring with a carboxylic acid in the anomeric position, an N-acetyl group in position 5 and a glycerol chain in position 6. The other main neuraminic acid derivatives include; N-glycolyl-neuraminic acid (Neu5Gc), N-acetyl-9-O-acetylneuraminic acid (Neu5,9Ac2) and 2-keto-3-deoxy-D-glycerol-D-galacto-nononic acid (KDN)[27, 46] (Figure 1.4).

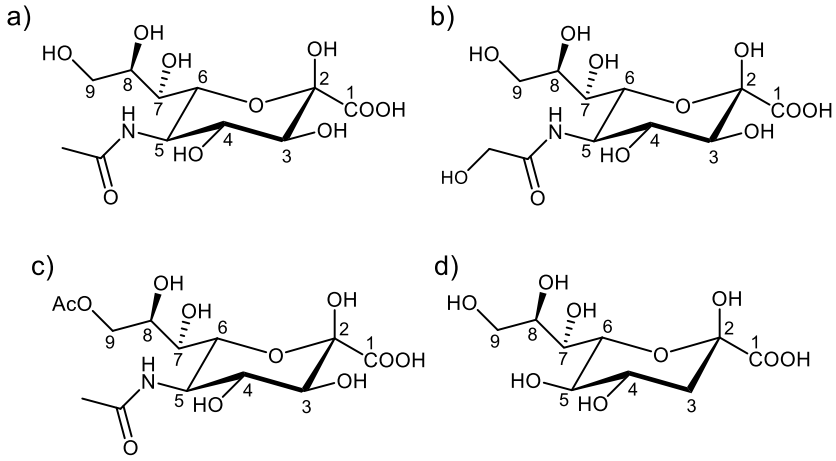


Figure 1.4. a) Sialic acid (SA) or N-acetylneuraminic acid; b) N-glycolyl-neuraminic acid; c) N-acetyl-9-O-acetylneuraminic acid; d) 2-keto-3-deoxy-D-glycerol-D-galacto-nononic acid.

Sialic acid can be found at the terminal end of glycans but also exist as free monosaccharide. Sialic acid presents two different configurations, the α (Figure 1.5 b) and the β (Figure 1.5 a). When sialic acid is unbound and free in solution it is predominantly present in the β anomeric configuration (βSA)[54]. This is characterised by the carboxylic acid and the anomeric hydroxyl group in equatorial and axial positions, respectively. The β configuration composes about 95% of the SA in solution, with the remaining being in the α and acyclic forms (Figure 1.5 c)[55].

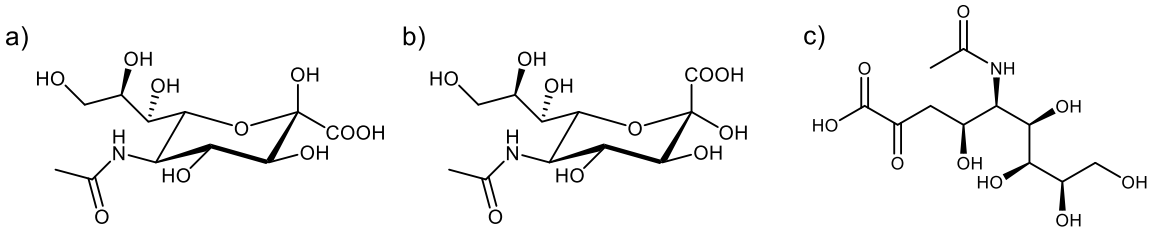


Figure 1.5. Different sialic acid forms; a) β configuration; b) α configuration; c) acyclic form.

On the other hand, when bound in carbohydrate chains the sialic acid adopts the α configuration (αSA), with the carboxylic acid in the axial position and the anomeric hydroxyl group, engaging in the glycoside linkage, in the equatorial position. High levels of free sialic acid indicated

underlying pathological conditions. These include, the neurodegenerative disorders Salla disease[56] and infantile sialic acid storage disease (ISSD)[57] which present high excretion of free sialic acid in urine. Whilst, high levels of free SA in the serum have been correlated to liver dysfunctions caused by alcohol abuse[58] and in patients with prostate and urinary bladder cancer and renal cell carcinoma[59]. Consequently, free sialic acid present great potential as a biomarker for the aforementioned conditions.

1.2 Receptors

1.2.1 Natural receptors

1.2.1.1 Lectins

Lectins are carbohydrate-binding proteins, most of which are glycoproteins, expressed by many living organisms including plants, animals or pathogens[60]. Lectins mediate the invasion of the host cell by pathogens, play a defence role in plants and are involved in cell communication and in the immune system of animals[61]. Lectins, especially ones from plants, have long been employed for the analysis and detection of glycans[62]. Lectins bind carbohydrates by accommodating the ligand in the binding pocket where the amino acid residues form non-covalent interactions, especially hydrogen bonds and CH-π interactions, with the sugar. Hydrogen bonds are pivotal for the specificity of lectins as they are highly directional guaranteeing the correct fit of the ligand in the binding pocket[63]. The hydroxyl groups of the carbohydrate act both as hydrogen bond donor and acceptor, at the same time. The oxygen of the hydroxyl group presents two lone pairs of electrons, thus can accept two hydrogen bonds, whilst the oxygen proton acts as a donor of one hydrogen bond[64]. This cooperative effect provides hydrogen bonds stronger than average[63]. This phenomenon is true for all hydroxyl groups, except for the anomeric one which is exclusively a donor. The carbohydrates may also

contain other polar groups which engage in hydrogen bonding. For instance, the NH group and carbonyl oxygen of the acetamide group (*e.g.* GlcNAc, GalNAc, SA) act as hydrogen bond donors and acceptors, respectively[64]. Furthermore, the lectin counterpart establishes hydrogen bonds with carbohydrates principally through the amide groups of the backbone chain and asparagine and glutamine side chains[64]. In addition, charged (*e.g.* His, Arg, Lys) or bidentate amino acid residues (*e.g.* Arg, Asn, Gln) often decorate the binding pocket[63]. The former residues form charge-reinforced hydrogen bonds which are stronger than their neutral counterparts, whilst the latter create a wider network of hydrogen bonds resulting in stronger interactions[63, 65]. Conversely, amino acids with side chains bearing hydroxyl groups (*e.g.* Tyr, Ser, Thr) are not frequently present in lectin binding pockets[64].

On the other hand, hydrophobic interactions, mainly CH-π interactions, occur between the C-H groups of the sugar and aromatic amino acid residues (*e.g.* Trp, Tyr, Phe and His)[66]. The CH-π interactions are formed between carbohydrate aliphatic protons, which are partially positively charged, and the π electron system of the aromatic moiety[64]. They have been observed with the electropositive C-H faces of the pyranose rings and other non-polar functionalities of sugars, such as the acetamido and the glycerol groups[64]. Furthermore, the aromatic amino acids differ in their tendency of forming CH-π interactions. Tryptophan, which has an electron-rich aromatic system, has the greatest tendency to form CH-π interactions[66]. Tyrosine is preferred to phenylalanine, in forming CH-π interactions, due to the more electron-rich π system, whilst histidine is rarely involved in these interactions. Conversely, amino acids with aliphatic side chains are not commonly present in binding pockets as they do not provide significant interactions with sugars. Moreover, when the carbohydrate ligand is charged electrostatic interactions can occur[64]. For instance, the interactions with anionic groups, such the carboxylate group of sialic acid, occurs via hydrogen bonds or electrostatic interactions. When hydrogen

bonds are employed to bind the carboxylate, they are often observed with Ser residues, whilst electrostatic interactions occur with Arg or Lys side chains[67-68].

The binding affinity of lectins to their ligand is expressed in terms of binding constant (K_a) or dissociation constant (K_d), with the relation between the two constants described in Equation 1.1.

$$K_a = \frac{1}{K_d} \qquad \text{Equation 1.1}$$

The affinity can also be described indirectly as half the maximal concentration of ligand required to inhibit the agglutination or precipitation of the lectin[69], also referred to as IC_{50}. Herein, the IC_{50} parameter is reported to address the affinity of lectins when the binding or dissociation constants are not given, with lower IC_{50} values being representative of higher affinities.

Lectins have been extensively studied and employed for glycan recognition[62]. For instance, lectin affinity chromatography is used for the isolation and purification of glycans. Lectins are also employed in electron microscopy or fluorescence in histological analysis and in blotting or microarrays studies for glycan profiling[62]. Furthermore, lectins are also utilised in biosensors for diagnostic purposes. The biosensors utilise both labelled and label-free lectins, with the analysis being conducted by different means such as fluorescence, surface plasmon resonance, quartz crystal microbalance and electrochemical impedance spectroscopy[62].

Of particular interest are sialic acid-specific lectins, due to the relevance of this monosaccharide in many pathological conditions. These lectins originate from plant or animal sources and are employed as analytical and diagnostic tools. Herein, a few examples of SA-specific lectins are discussed with their binding models being detailed. Different lectins targeting sialic acid can display selectivity towards a particular neuraminic acid derivative (*e.g.* SA, Neu5Gc, KDN), a

certain glycosidic linkage (*i.e.* α2-3, α2-6) or towards a specific carbohydrate sequence containing sialic acid[60]. Lectins with specificity towards particular neuraminic acid derivatives include the Limulin lectin and the *Limax Flavus* Agglutinin (LFA). The former lectin, which is found in the American horseshoe crab (*Limulus polyphemus*), shows specificity towards two neuraminic acid derivatives; SA and N-glycolylneuraminic acid. The Limulin lectin has been employed in detection of sialylated glycans by light and electron microscopy[60]. The LFA, originating from the slug *Limax Flavus*, is specific for SA and binds any sialylated glycan regardless of the glycosidic linkage type[70]. The binding affinity of LFA to SA is reported to be $K_a = 3.8\times10^4\ M^{-1}$, with the glycerol chain being a key point of interaction[70].

Furthermore, some plant lectins can differentiate between different glycosidic linkages including the Wheat Germ Agglutinin (WGA), the *Maackia amurensis* Leukoagglutinin (MAL) and Hemagglutinin (MAH) and the *Sambucus Nigra* lectin (SNA). The Wheat Germ Agglutinin (WGA) binds to SA residues when linked α2-3 to lactose sequences[69]. However, WGA main targets are N-acetylglucosamine and β1-4 GlcNAc-containing sequences. The broad specificity of WGA is due to the structural similarities between GlcNAc and SA with IC_{50} values of 12.5 M and 25 M, respectively. Larger GlcNAc structures (IC_{50} < 2.0 M) and sialylated lactose (IC_{50} = 6.3 M) are more powerful inhibitors. WGA was also reported to bind highly sialylated mucins at μM concentrations, due to the cluster effect (see Section 1.2.2.2). The binding of WGA to SA and to GlcNAc exhibit similar non-covalent interactions due to the structural similarities of the two sugars[64, 71]. In particular, the N-acetyl and 4-OH groups of both ligands are involved in hydrogen bonds with the WGA binding pocket[71]. In addition the SA carboxylate group forms hydrogen bonds with a Ser residue, whilst the glycerol chain engages in apolar interactions with a Tyr[71].

The MAL and MAH lectins, from the *Maackia amurensis* seeds, are selective for glycans containing sialic acid residues linked α2-3 to galactose (K_a in the μM range), with no binding to α2-6 structures[72]. In particular, MAL binds the sialylactosamine structures in N-glycans, whilst MAH interacts with disialylated epitopes in O-glycans[73]. The complex formed between MAL and α2-3-sialylactosamine (SA-α2-3-Gal-β1-4-GlcNAc) presents non-covalent interactions of all three sugars units with the binding pocket amino acid residues[68]. The carbon backbones of the sugars form hydrophobic interactions with various Tyr residues, whilst galactose hydroxyl groups are engaged in hydrogen bonds. The carboxylate group of the SA unit interacts through salt bridges with a Lys residue. Conversely the glycerol chain, a key point of interaction for other lectins, sits outside the binding pocket. The *Maackia amurensis* lectins have been employed to study the glycan phenotype in patients with prostate cancer by lectin affinity chromatography and surface plasmon resonance[53]. More recently MAL has been employed in sandwich-type biosensors[74] and in carbon nanotube-based electrochemical sensors[75] for the detection of α2-3 sialylated glycans in biological samples. These biosensors, despite their low limit of detection (fg ml^{-1}), have no clinical relevance due to their time consuming preparation.

Conversely, the *Sambucus nigra* lectin shows a marked preference towards sialylated glycans with α2-6 linkages to Gal or GalNAc units, over the α2-3 isomers[76-77]. The binding affinity of this elderberry-bark originating lectin to the simple SA-α2-6-lactitol structure is 3.9×10^5 M^{-1}, with the glycerol chain of sialic acid being pivotal for the interaction[78]. However, SNA does not only interact with sialylated glycans but also with Gal or GalNAc terminating glycans[79], hence the low specificity displayed by this lectin. SNA has been employed in lectin affinity chromatography to separate α2-6 sialylated glycans from the α2-3 isomers[77] and in light and electron microscopy to study the glycan phenotype of cells[79]. Furthermore, it has been

employed in biosensors for glycan detection with low limit of detection (fg ml^{-1})[80-81]. However, their convoluted fabrication and low stability mean they are not viable for clinical application.

In addition to plant lectins, human lectins targeting sialic acid have been uncovered. Siglecs are human lectins which mediate the signalling and communication between cells[82]. They are divided into 11 classes based on the preferred type of sialic acid derivative and glycosidic linkage. Siglec-1, also known as sialoadhesin, was shown to bind the α-anomeric configuration of SA, with K_d for the α model molecule, 2-O-methyl-α-sialic acid, of 3 mM[67]. The binding affinity of Siglec-1 exhibits a 2-fold increase for α2-3 sialylated epitopes but is only marginally increased for α2-6 isomers, indicating the preference of Siglec-1 for α2-3 epitopes. Conversely, Siglec-7 binds preferentially α2-6 and α2-8 sialylated epitopes[83]. The interactions between the Siglec-1 binding site and SA-α2-3-lactose occurs through van der Waals and hydrophobic interactions of the N-acetyl group and the glycerol chain with Trp residues[67]. In addition, hydrogen bonds are present between hydroxyl groups and polar amino acids, whilst electrostatic interactions occur between the carboxylate group and an Arg residue.

Many lectins have been uncovered and employed as probes for sialic acid detection, however they present a series of disadvantages limiting their applicability. Lectins are unstable and therefore require storage at 4 °C in order to be conserved for 6 to 12 months[60]. In addition, they have a limited pH range (pH 7 - 9) over which they are stable and active[60]. Lectins are only found in natural sources from which they need to be isolated and purified with yields that can be extremely low, hence the low availability[69]. Moreover, lectins interact with carbohydrates exclusively through non-covalent interactions, resulting in weak binding. When employed in biosensor their affinity is often improved, however the biosensors fabrication can be convoluted and time consuming resulting in lectin-based biosensors not applicable for clinical

implementation[75, 80]. Furthermore, lectins have low specificity as they often target multiple glycan epitopes limiting their use in the analysis of biological samples[81].

1.2.1.2 Antibodies

Anti-glycan antibodies (AGAs) are monoclonal antibodies, mainly IgM or IgG, employed as therapeutic and diagnostic agents to target carbohydrates[84]. They act by recognising glycan sequences or the entire glycosylated macromolecule[84]. Due to relevance in malignant cancer of Tn, sialyl-Tn, TF and Lewis epitopes AGAs towards these antigens have been extensively studied with some being employed for diagnostic purposes[85]. For instance, a monoclonal antibody for sialylated glycans has been employed to detect antigens in the serum of patients with lung carcinomas[86].

The mechanism of binding of the antibodies to the glycans is analogous to the one employed by lectins and thus based solely on non-covalent interactions. Hydrogen bonding can be found between the hydroxyl groups of the carbohydrate chain and polar amino acids of the binding pocket. Whilst, hydrophobic interactions are present between the apolar faces of the sugar units and aromatic residues, such as Trp and Tyr[87]. In addition, AGAs targeting anionic glycans present positively charged functionalities in their binding site allowing ionic interactions to occur[88].

However, AGAs present a series of issues, such as availability, specificity and affinity due to the low immunogenicity of glycans[84]. The low immunogenicity of the glycans is caused by the presence of the antigens not merely in aberrant macromolecules but similarly in self-molecules. This prevents the recruiting of T-cells to produce immunoglobulins[84, 87]. For this reason, the availability of antiglycan antibodies is very low. The number of glycan epitopes has been

estimated to be approximatively 7000, however the number of AGAs is limited to about 1000[84]. Moreover, many of these antibodies share the same target, thus the number of different antigens for which an antibody exists is below 250. In particular, the least represented group of AGAs is the one targeting N-glycans, with the existing AGAs having very low specificity and no commercial availablility[84]. Low specificity is a common feature of many AGAs as many of these do not recognise solely a specific glycan but also other structurally related motifs[89]. Consequently, AGAs capable of discriminating between specific sialic acid derivatives (*e.g.* SA, Neu5Gc, KDN and O-acetylated variants) are still lacking.

1.2.2 Synthetic receptors

1.2.2.1 Aptamers

Aptamers are single stranded deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) oligonucleotides formed from fewer than 100 bases, which can selectively recognise a specific target[85]. They are selected from a pool of strands by a technique called Systematic Evolution of Ligands by Exponential Enrichment (SELEX)[90]. When compared to antibodies, the synthesis of aptamers is more consistent, reproducible and they can be more easily modified by chemical transformation[91]. In spite of these advantages, the main shortcomings of aptamers are the low affinity and low specificity, causing cross reactivity with other glycan epitopes. This is partially due to the limited contact they can provide to carbohydrates. Indeed, aptamers do not contain aromatic moieties or charged functionalities and thus can only interact through hydrogen bonding with carbohydrates. There are not many examples of aptamer targeting sialic acid. For instance RNA aptamers bind SA-terminating glycans with dissociation constants in the low nM range, however they are not able to discriminate between α2-3 and α2-6 isomers or among different Lewis antigens[92-93]. Consequently, chemically modified aptamers have been

developed to introduce additional functional groups (*e.g.* cationic moieties and boronic acids) that provide extra contact points with the target[94-95]. For instance, a single strand DNA aptamer bearing amino groups was developed for sialylactose, with improved binding constant (μM) when compared to the uncharged analogue[94].

1.2.2.2 Non-covalent synthetic receptors

In addition to macromolecules, a great variety of synthetic receptors have been developed as recognition tools for glycans. The advantage of synthetic receptors over lectins, antibodies and aptamers, is the great variability in their structures allowing interaction with carbohydrates through multiple and different functionalities. The formation of multiple interactions with the target, known as a multivalent approach, is always desirable in order to enhance the binding affinity. In the multivalent approach the formation of multiple weak interactions with the ligand results in a stronger affinity than the sum of the single interactions[96]. This phenomenon, called the "cluster effect" indeed at the basis of lectin-carbohydrate interactions[96]. For instance, the binding affinity of lectins to monosaccharides is in the mM to μM range, whilst for oligosaccharides, which offer more sites for interactions, is up to nM[97]. The multivalent approach has been highly exploited for the design of synthetic receptors. Synthetic receptors, also known as synthetic lectins, are divided into receptors acting solely through non-covalent interactions and receptors able to form covalent bonds with the target.

In this Section non-covalent synthetic receptors are discussed[98]. These supramolecular receptors interact with carbohydrates mainly through combinations of hydrogen bonds, CH-π and electrostatic interactions. They are frequently formed by aromatic cage-like or macrocyclic structures in order to encapsulate the sugar and maximise the binding through multiple

non-covalent interactions. The aromatic scaffold employed in these receptors have included porphyrin[99], calixarene[100], biphenyl[101] and phenyl rings[102]. In addition, the scaffold is frequently functionalised with polar groups such as, hydroxyl, amide, amine (Figure 1.6 a) in order to provide hydrogen bonds[103-104]. Particularly, amine and guanidine groups are often incorporated in the receptors as they can provide more than one hydrogen bond allowing binding of vicinal *cis*-diol units.

Other bidentate ligands such as naphthyridine, phenantroline and oxime moieties are also present in sugar receptors[105-107]. Functional groups mimicking amino acid residues of lectin binding pockets can also be introduced. For instance, aminopyridine and carboxylate groups resemble Arg and Gln residues[108], respectively, whilst imidazole and indole rings imitate Trp and His residues[109]. In addition, receptors with charged moieties provide charge-reinforced hydrogen bonds or electrostatic interactions with neutral and charged sugar groups, respectively[110]. Charge-reinforced hydrogen bonds are stronger than neutral hydrogen bonds[65], hence charged functionalities are preferred.

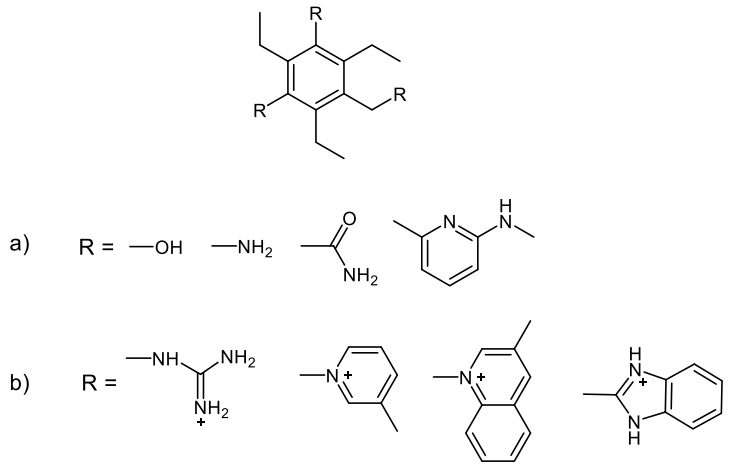


Figure 1.6. Phenyl-based scaffold receptors functionalised with; a) Polar groups; b) Positively charged groups.

Many of the synthetic lectins developed to target sialic acid are acyclic receptors. These consist of an aromatic core and functionalised branches. The phenyl ring provides CH-π interactions with the CH groups of the carbohydrate apolar faces, whilst the polar or charged branches create hydrogen bonds or electrostatic interactions[103, 108, 110] . Acyclic receptors for the detection of SA present binding constants in the order of 10^5 M^{-1}. They are based on branches bearing charged groups, such as guanidinium, benzimidazolium, pyridinium and quinolinium groups (Figure 1.6 b)[104, 110]. These groups can provide electrostatic interactions and charge-reinforced hydrogen bonds with the sialic acid carboxylate and hydroxyl groups, respectively.

The main issue limiting the applicability of the abovementioned receptors is the very low or negligible solubility in water. Therefore, acyclic receptors are employed in organic solvents or in water/DMSO mixtures, with the complexation decreasing as the polarity of the solvent increases. Indeed, when synthetic lectins are used in organic solvents, they interact with the carbohydrates mainly through hydrogen bonds which are particularly strong in non-polar media as there is no competition between the carbohydrate and water molecules. On the other hand, in the aqueous media CH-π interactions are stronger due to the hydrophobic effect. Nevertheless, the recognition of carbohydrates in aqueous media can be challenging, due to the desolvation required prior interaction between the two species[98]. Indeed, the examples of synthetic lectins acting in water are very limited[111-112]. For instance, molecular-imprinted polymers for the detection of sialic acid can either be functionalised with 4-vinylpyridine or N,N,N-trimethylaminoethyl methacrylate to bind through the sialic acid carboxylate group[113]. However, only the N,N,N-trimethylaminoethyl methacrylate-based polymers can recognise SA in aqueous conditions, whilst the 4-vinylpyridine-based one acts solely in organic solvents. Furthermore, Ferrand *et al.* developed a water-soluble supramolecular receptor[112], which binds the disaccharide cellobiose with $K_a = 650$ M^{-1}. This receptor consists of an aromatic floor and

roof linked by isophthalamide units with tricarboxylate branches to increase the water solubility. Further studies from the same research group led to water-soluble receptors (Figure 1.7) for the recognition of charged monosaccharides, such as α-sialic acid[111]. These receptors present pyrene platforms, to provide hydrophobic and CH-π interactions with the apolar face of α-SA, and guanidino branches to form hydrogen bonds and electrostatic interactions.

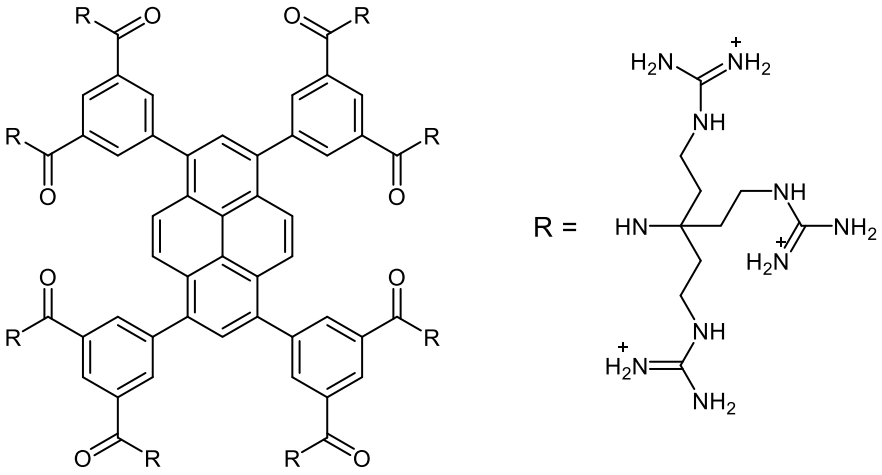


Figure 1.7. Water-soluble pyrene-based receptor for sugar recognition.

The difficulties in developing water compatible synthetic receptors for sugars are due to the competition with water encountered by the receptors when forming hydrogen bonds with carbohydrates. Hence, when working in aqueous solutions boron-based receptors are preferred as they covalently bind saccharides[98].

1.2.2.3 Covalent synthetic receptors – boronic acids

Boronic acids (BAs) are Lewis acids employed in carbohydrate recognition due to their ability to covalently bind diols through the formation of boronate esters[96]. The formation constant of

the boronate ester with the ligand is pH dependent. The optimal binding pH is based on the pK_a of the boronic acid and the pK_a of the ligand, as shown in Equation 1.2[114].

$$pH = \frac{(pK_aBA + pK_aligand)}{2} \qquad \text{Equation 1.2}$$

Phenylboronic acid (PBA, Figure 1.9), an aromatic boronic acid incorporated in many carbohydrate receptors, is known to bind sugars preferentially at basic pH[96]. The pK_a of PBA is 8.86[115], thus when binding, for instance, fructose (pK_a = 12.1) the theoretical optimal pH is 10.5[116]. Furthermore, the boron centre electronic structure changes upon the pH. At pH values below the pK_a the boron centre is in the neutral and trigonal form whilst, at pH values above the pK_a the boron is negatively charged in the tetrahedral form. Therefore, the binding of boronic acid to saccharides at basic pH occurs through the anionic tetrahedral boronate, to give a negatively charged boronate ester (Figure 1.8 a)[117].

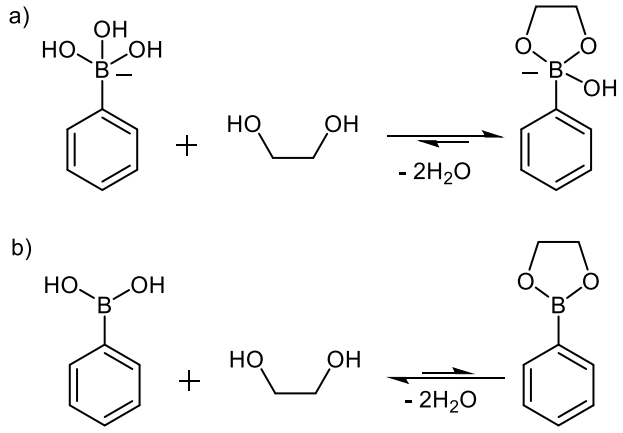


Figure 1.8. The boronate ester formation between a diol and; a) Anionic tetrahedral boronic acid; b) Neutral trigonal boronic acid.

Conversely, the trigonal boronate ester which would be formed with the diol at more acidic pH is not stable and is easily hydrolysed (Figure 1.8 b). Therefore, as a general rule for all neutral carbohydrates, the binding to boronic acid occurs at pH values above the BA pK_a with the tetrahedral boron centre[96]. On the other hand, ligands with lower pK_a are bound preferentially

at pH values below the BA pK_a, hence through the trigonal boron. For instance, the dye Alizarin Red S (ARS) (pK_a = 5.5) and sialic acid (pK_a = 2.6) are bound by boronic acids at neutral and acidic pH, respectively[116, 118].

BA receptors have limited applicability in the analysis of biological fluid as the binding affinity to neutral monosaccharides is greatly reduced at physiological pH (7.4)[116]. At pH 7.4 the binding affinity of PBA is the highest for ARS and catechol (K_a > 800 M^{-1}) and decreases significantly for carbohydrates with the following trend; sorbitol, fructose, mannitol (K_a > 100 M^{-1}) > arabinose, sialic acid, glucuronic acid, galactose, mannose (> 10 M^{-1}) > glucose, lactose (< 10 M^{-1})[116]. For instance, the binding affinity of PBA to fructose[116] at 8.5 pH is 560 M^{-1}, whilst at pH 7.4 is reduced to 160 M^{-1}.

The binding affinity to carbohydrates at physiological pH can be increased by employing more acidic boronic acids. The acidity of the boron can be increased by introducing electro-withdrawing substituents on the aromatic ring, which makes the boron more electron deficient[119]. For instance, the addition of a nitro group in *meta* position of PBA increases the acidity of the boron, resulting in a pK_a of 6.9[119]. The *m*-nitrophenylboronic acid (Figure 1.9 b) at physiological pH presents an 8-fold enhancement of the affinity to fructose, K_a = 1350 M^{-1}, when compared to PBA.

Moreover, the acidity of the boron centre can also be increased by introducing a tertiary amine in the *ortho* position, to give Wulff-type boronic acids[120]. The nitrogen atom of the *ortho*-amine group is believed to stabilises the boronate ester through a B-N intramolecular bond or through ion pairing in aprotic and protic solvents, respectively[121]. The Wulff-type BA 2-(dimethylaminomethyl)phenylboronic acid (Figure 1.9 c), with pK_a of 5.3, can be employed at acidic pH for the detection of neutral monosaccharides[122].

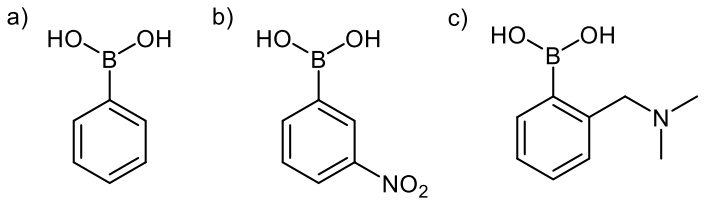


Figure 1.9. Structures of; a) Phenylboronic acid (PBA); b) *m*-Nitrophenylboronic acid; c) 2-(Dimethylaminomethyl)phenylboronic acid.

Boronic acid receptors have been extensively employed for the detection of monosaccharides, with their binding affinity varying with the chosen ligand[116]. For instance, at pH 8.5 PBA presents the highest affinity to fructose (K_a = 560 M^{-1}). The affinity is significantly lower for galactose (K_a = 80 M^{-1}) and further reduced for glucose (K_a = 11 M^{-1})[116]. The same trend in affinities is retained at physiological pH.

In solution, the aforementioned monosaccharides are in an equilibrium between the furanose and the pyranose forms. BA receptors preferentially bind to the furanose form[123-124]. In fact, the stability of the boronate ester is greatest for endocylic *vic*-diol furanose, whilst is up to 30 times lower for exocyclic-1,2-diol pyranose[114]. An even lower stability is shown for other carbohydrate diols, in the following order exocyclic-1,2-diol furanose > *cis*-*vic*-diol pyranose > exocyclic 4,6-diol pyranose (Figure 1.10)[114].

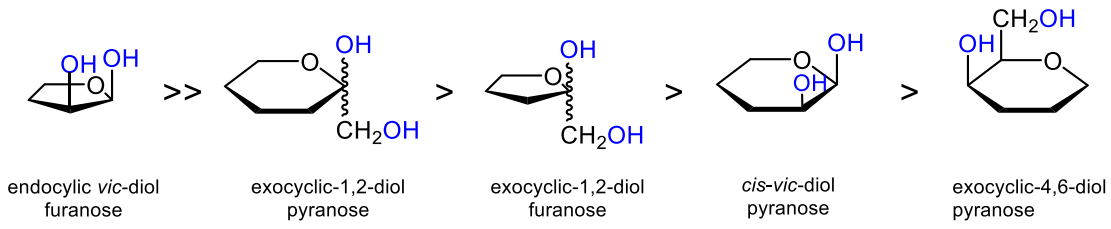


Figure 1.10. Sugar diols in order of affinity, from highest to lowest, for boronic acid receptors.

The preferential binding to the furanose form limits the application of boronic acids in the detection of glycans presenting solely pyranose sugar residues. Nevertheless, BA receptors can

be employed for the recognition of monosaccharides or di- and oligosaccharides formed by furanose units.

The first boronic acid receptor for carbohydrate recognition was developed by Yoon *et al.*, which consisted of an anthracene-based boronic acid (Figure 1.11 a) employed for fluorescence studies of monosaccharides[125]. Since then many functionalised boronic acids have been developed containing one or more boronic acid units and polar or apolar functional groups. The presence of additional functionalities improves the affinity and selectivity of the probe by allowing multipoint interactions with the ligand. For instance, Tsukagoshi *et al.* developed the first receptor selective for glucose[126] by inserting two appropriately spaced boronic acid units in the receptor design (Figure 1.11 b). Mono phenylboronic acid displays a greater affinity to fructose than glucose. However, contrary to fructose, glucose presents two binding sites which can be simultaneously accessed by the diboronic acid receptor resulting in preferential binding, or selectivity, for glucose[96]. Diboronic acid receptors have also been proposed for oligosaccharides. For instance, a fluorescent diboronic acid receptor was developed for the selective recognition of the sialyl Lewis x epitope[127]. This is reported to bind the tetrasaccharide despite the sugar unit being the pyranose forms but, unfortunately, no binding model was described.

Many PBA-based receptors have been functionalised by a variety of groups, including polar and apolar groups, neutral and charged ones in order to achieve multivalent interactions with the monosaccharides. For instance, urea and thiourea (Figure 1.11 c) functionalised boronic acids cooperatively bind sialic acid through boronate esterfication and hydrogen bonding[128]. Charged moieties are often introduced in PBA receptors for the detection of negatively charged carbohydrates. Guanidino-functionalised PBA receptors have been developed for the recognition of D-glucarate and SA[129-130]. The guanidino group provides electrostatic

interactions with the carboxylate group of glucarate which, in addition to the boronate ester, results in a two-point interaction and a 5-fold increase of the affinity. Triboronic acid receptors[131] functionalised with amino groups are reported to bind heparin, an anionic polysaccharide, with the contribution of electrostatic interactions leading to binding constants up to 10^8 M^{-1}. Furthermore, 5-indolylboronic acid (Figure 1.11 d), which mimics the tryptophan residues in lectins, interacts with oligosaccharides through hydrophobic and CH-π interactions[132]. The 5-indolylboronic acid, by intercalating between sugar residues, results in selective binding for glucose oligosaccharides.

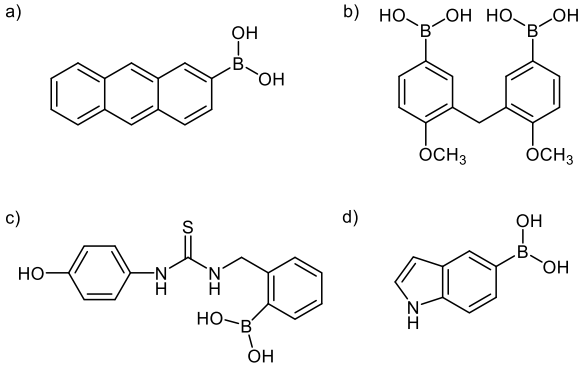


Figure 1.11. Boronic acid derivatives developed for carbohydrate recognition; a) Anthrylboronic acid; b) 2,2'-Dimethoxydiphenylmethane-5,5'-diboronic acid; c) 2-{[3-(4-Nitrophenyl)thioureido]methyl}phenylboronic acid; d) 5-Indolylboronic acid.

Boronic acid can also be introduced in peptide scaffolds in order to mimic the lectin-saccharide interaction. The peptidomimetic molecules often include charged amino acid residues, such as Arg and Lys, polar residues as Tyr, and aromatic ones, such as Phe and Trp[133-135]. Peptides with different amino acid sequences results in different affinities and selectivities for carbohydrates[135]. In addition, boronic acid can also be included in nucleic acids and polymers to increase the affinity through formation of covalent complexes[136]. For instance, DNA-aptamers bearing boronic acid groups showed an improved affinity towards glycosylated proteins when compared to strands lacking boron units[95]. PBA-functionalised hydrogels can

detect glucose at mM levels[137], whilst PBA-functionalised micelles target cells with high sialylation levels and thus can be employed for the delivery of anticancer drugs[138]. Boronic acids have also been incorporated in surfaces or nanoparticles for carbohydrate detection. BA-functionalised gold surfaces have been developed for the selective recognition of glucose[139], sialic acid[140] and glycoproteins[141,142]. Furthermore, boronic acids have been employed in molecular imprinted surfaces in which a binding pocket is formed on the surface providing high specificity and selectivity for the ligand species. Molecular imprinted BA-functionalised gold surfaces proved to be highly selective for glycosylated PSA over other glycosylated and non-glycosylated proteins, with K_d = 1.8 μM[143]. Whilst, nanoparticles with boronic acid binding sites can be imprinted with mono- or oligosaccharides, with the reported binding affinities in the low mM range[144].

Two of the main issues limiting the applicability of boronic acids; the preferential binding to furanose rings and the pH at which they operate. The former issue limits the applicability of PBA receptors for the recognition of cancer related glycans, as these present sugar residues in the pyranose form. Whilst the latter limits their suitability in the analysis of biological fluids. This issue is partially overcome by Wulff-type and electro-withdrawing boronic acids. However, these, similar to many other PBA receptors, are poorly soluble in water and thus often require the presence of a co-solvents.

1.2.2.4 Covalent synthetic receptors – benzoboroxoles

As previously described, phenylboronic acid receptors exhibit a series of disadvantages which limit their applicability. These include the pH range in which they operate, their preferential binding to furanose forms and the poor solubility in water. A cyclic dehydrated from boronic

acid derivative, the *ortho*-hydroxymethyl phenylboronic acid (Figure 1.12), first presented by Torssell in 1957[145], overcomes many of the issues of BAs.

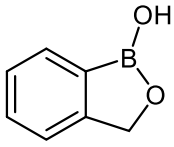


Figure 1.12. *ortho*-Hydroxymethyl phenylboronic acid or benzoboroxole (BOB).

This species is also known as benzoboroxole (BOB) or benzoxaborole and consist of an oxaborole heterocyclic ring fused to the benzene ring[120]. Despite their discovery in 1957 it was not until much later, in 2006, when Dowult and Hall showed the ability of benzoboroxole to bind sugars in the pyranose form at physiological pH[146]. In addition, benzoboroxole presents better solubility in water and lower pK_a (pK_a = 7.2) when compared to PBA. The increased acidity of BOB is caused by the ring strain of the trigonal and hybridised sp^2 boron which is relieved upon hybridization change to sp^3 giving the anionic tetrahedral boron centre[114]. In addition, contrary to PBA, when BOB forms a boronate ester with a diol the complex is further stabilised by intramolecular hydrogen bonds[147], contributing to higher affinity of BOB to carbohydrates. For instance, at neutral pH PBA exhibits negligible binding to glucose and moderate binding for fructose (K_a = 160 M^{-1})[116], whilst BOB shows enhanced affinities with K_a of 17 M^{-1} and 606 M^{-1}, respectively[147].

Despite the superiority of benzoboroxoles not many receptors containing this unit have been developed for carbohydrate recognition. For instance, fluorescent dyes containing one or three BOB units were developed for the detection of galactose[148],[149]. The monomeric receptor presents low binding affinity to galactose (K_a = 30 M^{-1}), whilst the trimeric probe has an affinity in the nM range. Furthermore, the BOB unit was included in Ugi-scaffold molecules and peptides. Ugi-scaffold are small molecules mimicking peptides and when functionalised with a

BOB unit can be employed for recognition of glycoproteins at physiological pH[150]. Benzoboroxole as well as phenylboronic acid have been incorporated in linear and cyclic peptide for the recognition of monosaccharides[151]. A bis-BOB functionalised peptide (Figure 1.13) has been developed for the selective recognition of the TF antigen, with K_d = 0.9 mM[152].

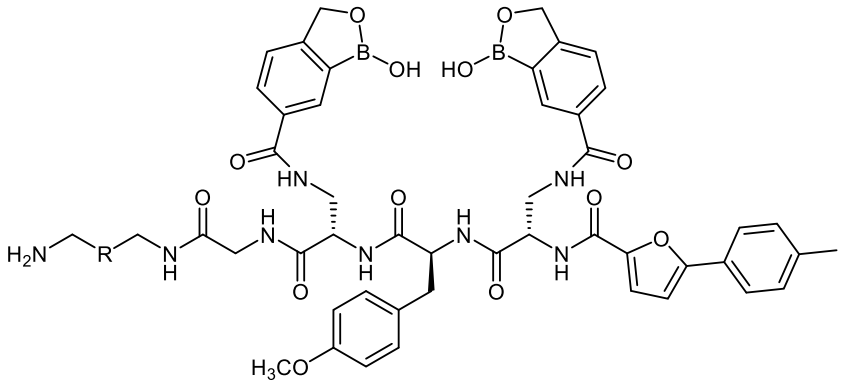


Figure 1.13. Bis-benzoboroxole functionalised peptide for the recognition of the TF antigen.

This Gal-β1-3-GalNAc antigen contains two diol binding sites to which the BOB units bind with formation of two esters. In addition, non-covalent interactions (*e.g.* H bonds, and CH-π interactions) are formed between the amino acid residues of the peptide backbone and the ligand. A key aspect of this receptor is the spacing between the two BOB units which needs to be appropriate in order to allow access to both diols. The correct spacing also determines the selectivity towards the Gal-β1-3-GalNAc-like structure over different mono- di- or oligosaccharide. Benzoboroxole units have also been utilised to functionalise nanoparticles and surfaces. At physiological pH BOB-decorated nanoparticles present a 2-fold increase in the affinity to fructose when compared to the binding of one BOB unit[153]. BOB-functionalised surfaces have been imprinted with different oligosaccharides and glycoproteins to create highly selective binding scaffolds with K_d in the low mM range[154].

1.2.2.5 Sialic acid recognition by covalent receptors

Sialic acid and sialylated glycans are biomarkers for a series of pathological conditions, such as cancer, sialic acid storage diseases and alcoholism[50-52, 56-59]. Synthetic receptors based solely on non-covalent interactions to target SA were described in Section 1.2.2.2. However, these present a series of disadvantages, mainly low or negligible solubility in water and weak interactions with the target. On the other hand, boronic acid and benzoboroxole present greater solubility and through the formation of boronate ester can covalently bind sugars, including sialic acid. Thus, BA and BOB receptors can be employed as probes in sialic acid recognition.

Boronic acids and its analogues present an anomalous binding profile to sialic acid. The binding occurs preferentially at acidic pH, contrary to all other monosaccharides[118]. Thus, at acidic pH, boron-based receptors are selective for SA with the binding occurring through the neutral and trigonal boron. Moreover, despite many BA receptors have been developed for SA, the site at which binding occurs is still under debate. There are two possible binding sites, the glycerol chain and the α-hydroxyacid group, with three different binding models described in literature. Otsuka *et al.* studied the binding of 3-(propionamido)phenylboronic acid (PAPBA) to sialic acid at acidic pH and proposed a binding mechanism involving the glycerol chain (Figure 1.14 a)[118]. The binding to the glycerol chain was proposed to occur through the *threo* geminal 1,2-diol C8/C9 or the 1,3-diol C7/C9, whilst the binding through the *cis*-diol C8/C9 is prohibited as a result of the *erythro* configuration[114]. The binding to the 1,2-diol is preferred as it forms a five-membered ring which is more stable than the six-membered ring resulting from binding the 1,3-diol[114]. However, these findings contradict with the binding affinities of BA to other diols in linear chains. Indeed, the binding affinity of boronic acids to molecules structurally similar to the glycerol chain, such as glycerine and other alkyl diols[155] has been assessed to be negligible at pH 7.4. Conversely, the binding affinity reported for SA under the

physiological conditions is 37.6 M^{-1}, with this value increasing at more acidic pH[118]. At acidic pH the binding occurs with the trigonal boron centre which would normally give an unstable boronate ester complex prone to hydrolysis. Otsuka *et al.* suggested a role of the SA acetamide group in stabilising the otherwise unstable boronate ester complex through intramolecular interaction[118]. However, in later studies this explanation was rejected[155-156]. Despite the conflicting information regarding the glycerol chain binding site this model is still utilised in many publications to describe the binding of boron-based receptors to sialic acid and sialylated glycans[128, 133-134, 156-163].

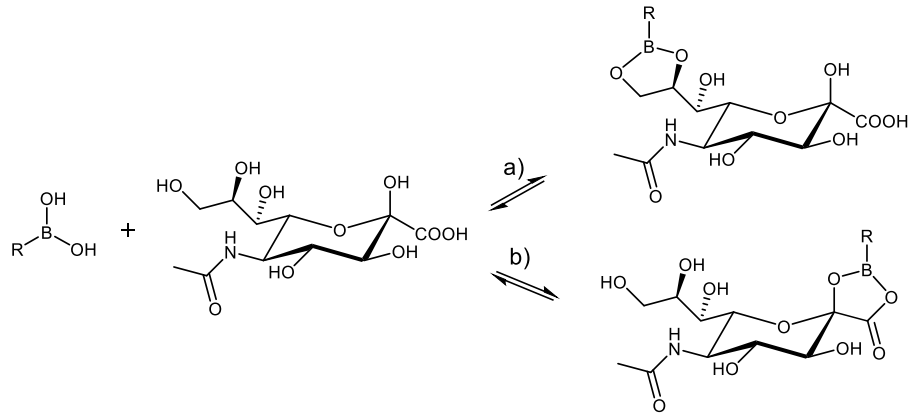


Figure 1.14. Boronate esterification of sialic acid through the; a) The glycerol chain binding site; b) The α-hydroxyacid binding site.

Alternative models to the one proposed by Otsuka *et al*. have been described in literature involving binding to the α-hydroxyacid moiety (Figure 1.14 b). Djanashvili *et al*. described the binding site as being pH dependent[156]. According to this model, at acidic and neutral pH values (pH 2 – 8) the complexation occurs with the negatively charged α-hydroxyacid moiety. Whilst, at pH > 8 the binding site switches to be the glycerol chain diols[156]. Moreover, at pH < 2 no binding is present due to the lower nucleophilicity of the protonated α-hydroxyacid group[164]. A third binding model was more recently detailed by Nishitani *et al*.[155]. In this model, the binding occurs exclusively with the α-hydroxyacid moiety and does not vary with the pH.

Furthermore, this model is consistent with the binding profile of boron-based receptors to other α-hydroxyacid containing molecules. Indeed, lactic, tartaric, gluconic, glucaric and glycolyc acids are all bound by BA receptors preferentially at acidic pH[165-167]. In addition, it was also shown that the boronic acid exhibits a stronger binding to α-hydroxyacid groups over diols[130].

A summary of the sialic acid receptors bearing boronic acid or benzoboroxole units is presented herein. The majority of these receptors are believed to covalently bind the glycerol chain diols. It can be assumed that the binding models discussed herein refer to the glycerol chain binding model, unless explicitly stated. Phenylboronic acid receptors are often functionalised with different groups to provide multipoint interactions with sialic acid. Urea and thiourea functionalities were included in PBA receptors to provide hydrogen bonds to SA, however these receptors are not water soluble and thus operate in DMSO[128]. On the contrary, heterocyclic boronic acid, such as 5-boronopicolinic acid (Figure 1.15 a), operates in aqueous media[158]. These receptors bind sialic acid at pH 6.5 with $K_a > 1000$ M^{-1}. In these receptors the nitrogen atom of the heterocyclic ring is believed to form hydrogen bonds with SA carboxylate group. A two-point interaction model, is also provided by metal chelate boronic acid receptors, with the metal chelate (*i.e.* lanthanide(III) and zinc (II)) interacting with the carboxylate ion[157, 159]. For instance, the lanthanide (III) based receptors at physiological pH present a binding constant of 151 M^{-1} for SA and 79 M^{-1} for the α-SA derivative, 2-O-methyl-α-sialic acid. A PBA-based MRI agent, containing an amine group and a lanthanide chelate, has been employed to detect SA on cells[168]. The amine group is reported to undergo electrostatic interactions with the carboxylate group and thus aid in the sialic acid recognition.

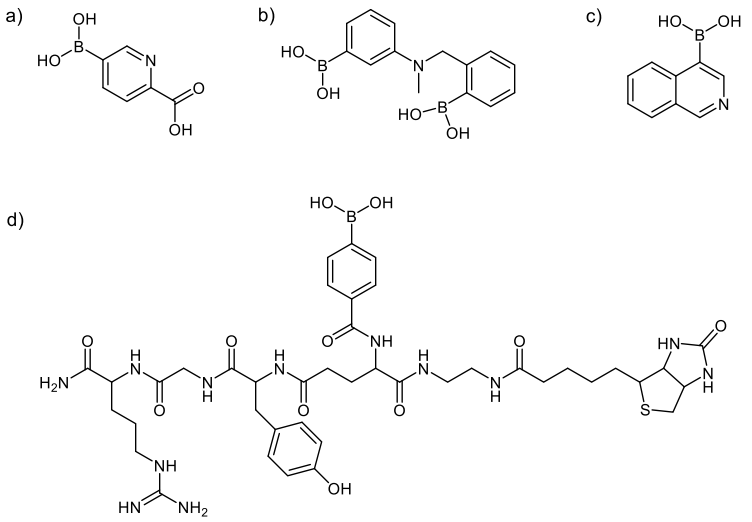


Figure 1.15. Boronic acids for sialic acid recognition; a) 5-Boronopicolinic acid; b) Bis-boronic acid; c) 4-Isoquinolineboronic acid; b) Phenylboronic acid-functionalised peptidomimetic.

The PBA or BOB units have also been included in peptide scaffolds. A peptidomimetic molecule (Figure 1.15 d) functionalised with phenylboronic acid and arginine provides multipoint interactions and display binding affinity to SA comparable to the *Sambucus nigra* lectin[133]. Moreover, the binding of a library of peptidomimetics bearing a BOB unit to various monosaccharides was investigated[134]. The highest affinity of these peptidomimetics was achieved with fructose (K_a = 700 – 1000 M^{-1}) followed by SA with K_a up to 167 M^{-1}. Moreover, BA-functionalised surfaces and nanoparticles have been developed as biosensors for sialic acid. PBA-functionalised nanoparticles and carbon dots have been employed for the detection of SA in biological sample with a limit of detection in the µM range[160-161, 163]. Fluorescent molecular imprinted nanoparticles present mM affinity to SA and are able to detect sialylated glycans via multivalent interactions[162].

All the aforementioned receptors are believed to bind SA through the glycerol chain diols. Only a few examples which consider alternative binding site exist in literature. A bis-boronic acid (Figure 1.15 b), is reported to simultaneously bind the glycerol chain and the α-hydroxyacid

group[169], whilst 4-isoquinolineboronic acid (Figure 1.15 c), is the only boron-based receptor to be reported to bind sialic acid exclusively through the α-hydroxyacid unit[164].

1.3 Dynamic combinatorial chemistry

Carbohydrate receptors, as detailed previously, are frequently centred on functional groups which mimic the key interactions between lectins and carbohydrates[108-109, 133-135, 150-151]. For instance, phenyl rings are introduced to mimic Phe and Trp residues, whilst charged or polar moieties mimic Arg, Lys, His amino acids[63, 66]. In addition, receptors can contain a boronic acid unit, or analogues, in order to form a covalently bound complex[96]. Thus, a great variety of receptors can be developed comprising of different functional groups linked in different fashions. The design of a receptor is often based on a core structure (*i.e.* aromatic, macrocyclic, peptido, boronic acid) which is then functionalised in different manners in order to generate distinct analogues[96, 98, 135]. Since the candidate receptors requiring independent synthesis and testing can be many, this approach can result extremely time consuming. This is especially marked when larger receptor molecules are required, thus limiting the number of analogues that can be evaluated. When targeting larger carbohydrates, such as oligosaccharides, the receptors must provide a high degree of multivalency with the target[170]. In order to achieve this, a large chemical space must be investigated, requiring many different analogues and isomers to be developed. However, this is often limited by the cumbersome and long synthesis resulting in only very few analogues being developed and tested.

A parallel can be drawn between this approach and the combinatorial chemistry approach widely adopted in medicinal chemistry. In medicinal chemistry large libraries of hit analogues are synthesised and tested to identify improved binders. This approach is particularly time

consuming and can be affected by a lack of positive results. In order to overcome the issues posed by the classic combinatorial approach Huc and Lehn applied the concept of dynamic combinatorial chemistry (DCC), introduced first by Brady *et al.*[171], in the medicinal chemistry field (Figure 1.16)[172]. In their seminal publication they detailed target-driven generation of carbonic anhydrase inhibitors. Two different sets of molecules, known as building blocks, with complementary reactivities were employed for the *in situ* generation of a series of adducts, forming a dynamic combinatorial library (DCL). In particular, they employed aldehyde- and amine-bearing building blocks (BBs) which form imine adducts via a reversible reaction. The carbonic anhydrase acts as a template driving the formation of its best binder(s) among all possible imine adducts.

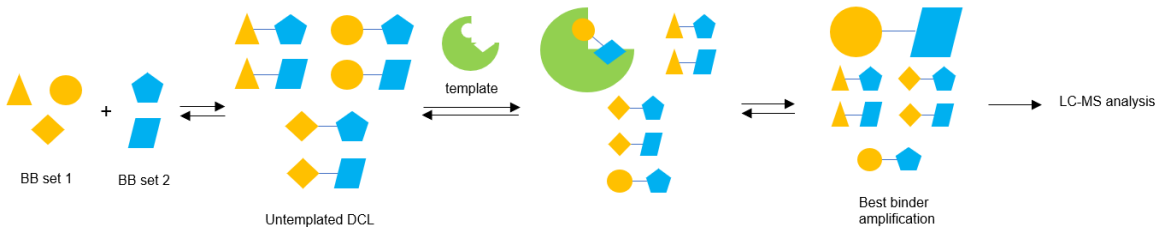


Figure 1.16. Schematic representation of the dynamic combinatorial chemistry (DCC) approach.

In general terms, the DCC approach consist of generating a library of adducts (DCL) by reversibly connecting a series of building blocks, or by connecting building blocks to a scaffold or framework[173]. The DCC is therefore based on a reversible reaction between the BBs, such as imine exchange, hydrazone and acylhydrazone exchange, oxime exchange, disulfide exchange, thioester exchange, Michael addition, reversible Dies-Alder, alkene metathesis, boronate esterification [174-175]. The building blocks and the adducts are under constant interconversion, with this being driven by the thermodynamic equilibrium of the system[176-177]. When a template, or target molecule, is added to the DCL the equilibrium is perturbated and

shifted towards the components which best bind the template[172]. This phenomenon is defined as thermodynamic templating; by shifting the equilibrium towards the amplified member, which is stabilised by the interactions with the template, the library can achieve the lowest Gibbs free energy[176]. In other words, the molecular recognition between the template and one or more library members stabilises the fittest binder at the expenses of the other library members[178]. Furthermore, in order to identify the amplified members for a templated DCL a comparative approach must be adopted. This consists in the generation, under the same experimental conditions, of one templated library and one control library without the target[179]. The two libraries are then analysed by chromatography providing two different chromatograms which are compared to identify the amplified peaks[179]. Often the analysis is performed by HPLC-MS or LC-MS with the mass spectroscopy allowing identification of the adducts by their mass[174].

Since the pioneering paper by Huc and Lehn, the DCC approach has been widely applied in medicinal chemistry for the discovery of new small molecules targeting biologically relevant proteins[180]. Moreover, DCC has also found applications in molecular recognition in the generation of receptors for biologically relevant small molecules or ions. These include sensing of nicotine[181], nucleosides[182], acetylcholine[183], carboxylic acids[184], dipeptides[185], anions[186-187] and metal ions[188-191]. Despite the potential of DCC for molecular recognition, its application in the glycosensing field has been very limited. Indeed, only a couple of publications, concerning the DCC approach for sugar detection, can be found in the literature and these are limited to mono- and disaccharide templates. Rauschenberg *et al.* presented the DCC generation of cyclic peptido receptors for the recognition of different mono- and disaccharides[192], whilst Yang *et al*. developed metal-organic cages for monosaccharides[193]. The application of DCC to oligosaccharides comprising three or more sugar units is, to this date, still unexplored.

1.4 Conclusions

Many of the proteins currently employed as biomarkers in cancer diagnostics lack tumour specificity and can lead to incorrect diagnosis[36]. Conversely, glycan modifications induced by cancer cells are closely associated with specific tumours rendering these glycan motifs more attractive biomarkers. In particular, many cancer-related glycans present a higher degree of sialylation[48-49]. Specifically, glycans terminating with a sialic acid unit bound α2-3 to the carbohydrate chain have been related to several tumour conditions[49]. Furthermore, sialic acid is also a relevant biomarker in its unbound form, especially for the detection of certain neurodegenerative diseases and alcoholism[56-58]. Therefore, the recognition of sialic acid in its free and bound forms can offer significant advantages in the diagnosis of related diseases. A series of natural and synthetic sugar receptors have been described herein[60, 84, 96, 98], demonstrating the superiority of boron-based receptors as sugar-binding tools[146], although only few of these have been developed for sialic acid recognition. New receptors, developed via rational or combinatorial means, are therefore in need for the detection of sialic acid and sialylated glycans and for future diagnostics applications.

1.5 Aims of this work

This work focuses on developing receptors for sialic acid and sialylated glycans for future applications as diagnostic tools.

1) Pursue an understanding of the molecular recognition of sialic acid as a free, unbound monosaccharide. The interaction of boron-based receptors to the sugar is investigated to elucidate the nature of the binding site, given the contradicting theories presented in

the literature. Following this, functionalised benzoboroxole receptors are designed and synthetised to enhance affinity and selectivity to sialic acid. The binding to sialic acid is measured by isothermal titration calorimetry (ITC) and a binding model was postulated, granting an understanding of the significance of different functional groups on the binding.

2) Dynamic combinatorial chemistry is proposed and detailed as a technique to generate highly selective receptors for oligosaccharides, including sialylated glycans. Preliminary works to enable the future application of this technique are performed, including the enzymatic synthesis of the sialylated epitopes of interest.

List of references

(1) Fuster, M. M.; Esko, J. D., The sweet and sour of cancer: Glycans as novel therapeutic targets. *Nat. Rev. Cancer* **2005,** *5* (7), 526-542.

(2) Sell, S., Cancer-associated carbohydrates identified by monoclonal antibodies. *Hum. Pathol.* **1990,** *21* (10), 1003-1019.

(3) Ho, W. L.; Hsu, W. M.; Huang, M. C.; Kadomatsu, K.; Nakagawara, A., Protein glycosylation in cancers and its potential therapeutic applications in neuroblastoma. *J. Hematol. Oncol.* **2016,** *9*, 15.

(4) Josic, D.; Martinovic, T.; Pavelic, K., Glycosylation and metastases. *Electrophoresis* **2019,** *40* (1), 140-150.

(5) Varki, A.; Cummings, R. D.; Aebi, M.; Packer, N. H.; Seeberger, P. H.; Esko, J. D.; Stanley, P.; Hart, G.; Darvill, A.; Kinoshita, T.; Prestegard, J. J.; Schnaar, R. L.; Freeze, H. H.; Marth, J. D.; Bertozzi, C. R.; Etzler, M. E.; Frank, M.; Vliegenthart, J. F. G.; Lütteke, T.; Perez, S.; Bolton, E.; Rudd, P.; Paulson, J.; Kanehisa, M.; Toukach, P.; Aoki-Kinoshita, K. F.; Dell, A.; Narimatsu, H.; York, W.; Taniguchi, N.; Kornfeld, S., Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **2015,** *25* (12), 1323-1324.

(6) Stowell, S. R.; Ju, T. Z.; Cummings, R. D., Protein Glycosylation in Cancer. *Annu. Rev. Pathol. Vol 10* **2015,** *10*, 473-510.

(7) Cummings, R. D.; Pierce, J. M., The Challenge and Promise of Glycomics. *Chem. Biol.* **2014,** *21* (1), 1-15.

(8) Dube, D. H.; Bertozzi, C. R., Glycans in cancer and inflammation. Potential for therapeutics and diagnostics. *Nat. Rev. Drug Discov.* **2005,** *4* (6), 477-488.

(9) Gabius, H. J.; Roth, J., An introduction to the sugar code. *Histochem. Cell Biol.* **2017,** *147* (2), 111-117.

(10) Stanley, P.; Taniguchi, N.; Aebi, M., N-Glycans. *Essentials of Glycobiology* **2017**.

(11) N-Glycans. In *Essentials of Glycobiology*, Varki, A.; Cummings, R. D.; Esko, J. D., Eds. Cold Spring Harbor Laboratory Press: New York, 1999.

(12) Trombetta, E. S., The contribution of N-glycans and their processing in the endoplasmic reticulum to glycoprotein biosynthesis. *Glycobiology* **2003,** *13* (9), 77R-91R.

(13) Dhanisha, S. S.; Guruvayoorappan, C.; Drishya, S.; Abeesh, P., Mucins: Structural diversity, biosynthesis, its role in pathogenesis and as possible therapeutic targets. *Crit. Rev. Oncol./Hematol.* **2018,** *122*, 98-122.

(14) Yu, R. K.; Tsai, Y. T.; Ariga, T.; Yanagisawa, M., Structures, Biosynthesis, and Functions of Gangliosides-an Overview. *J. Oleo Sci.* **2011,** *60* (10), 537-544.

(15) Rabu, C.; McIntosh, R.; Jurasova, Z.; Durrant, L., Glycans as targets for therapeutic antitumor antibodies. *Future Oncol.* **2012,** *8* (8), 943-960.

(16) Meezan, E.; Wu, H. C.; Black, P. H.; Robbins, P. W., Comparative Studies on the Carbohydrate-containing membrane components of normal and virus-transformed mouse fibroblasts. II. Separation of glycoproteins and glycopeptides by Sephadex chromatography. *Biochemistry* **1969,** *8* (6), 2518-2524.

(17) Orntoft, T. F.; Vestergaard, E. M., Clinical aspects of altered glycosylation of glycoproteins in cancer. *Electrophoresis* **1999,** *20* (2), 362-371.

(18) Kanter-Lewensohn, L.; Dricu, A.; Wang, M.; Wejde, J.; Kiessling, R.; Larsson, O., Expression of the insulin-like growth factor-1 receptor and its anti-apoptotic effect in malignant melanoma: a potential therapeutic target. *Melanoma Res.* **1998,** *8* (5), 389-397.

(19) Lai, J. P.; Chien, J.; Staub, J.; Avula, R.; Greene, E. L.; Matthews, T. A.; Smith, D. I.; Kaufmann, S. H.; Roberts, L. R.; Shridhar, V., Loss of HSulf-1 up-regulates heparin-binding growth factor signaling in cancer. *J. Biol. Chem.* **2003,** *278* (25), 23107-23117.

(20) Hakomori, S., Tumor malignancy defined by aberrant glycosylation and sphingo(glyco)lipid metabolism. *Cancer Res.* **1996,** *56* (23), 5309-5318.

(21) Vlodavsky, I.; Friedmann, Y., Molecular properties and involvement of heparanase in cancer metastasis and angiogenesis. *J. Clin. Invest.* **2001,** *108* (3), 341-347.

(22) Krause, T.; Turner, G. A., Are selectins involved in metastasis? *Clin. Exp. Metastasis* **1999,** *17* (3), 183-192.

(23) Borsig, L.; Wong, R.; Hynes, R. O.; Varki, N. M.; Varki, A., Synergistic effects of L- and P-selectin in facilitating tumor metastasis can involve non-mucin ligands and implicate leukocytes as enhancers of metastasis. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (4), 2193-2198.

(24) Ogata, S.-I.; Muramatsu, T.; Kobata, A., New structural characteristic of the large glycopeptides from transformed cells. *Nature* **1976,** *259* (5544), 580-582.

(25) Trinchera, M.; Aronica, A.; Dall'Olio, F., Selectin Ligands Sialyl-Lewis a and Sialyl-Lewis x in Gastrointestinal Cancers. *Biology-Basel* **2017,** *6* (1), 18.

(26) Lan, Y.; Hao, C.; Zeng, X.; He, Y. L.; Zeng, P. J.; Guo, Z. H.; Zhang, L. J., Serum glycoprotein-derived N- and O-linked glycans as cancer biomarkers. *Am. J. Cancer Res.* **2016,** *6* (11), 2390-2415.

(27) Schauer, R., Sialic acids as regulators of molecular and cellular interactions. *Curr. Opin. Struct. Biol.* **2009,** *19* (5), 507-514.

(28) Loureiro, L. R.; Sousa, D. P.; Ferreira, D.; Chai, W. G.; Lima, L.; Pereira, C.; Lopes, C. B.; Correia, V. G.; Silva, L. M.; Li, C. X.; Santos, L. L.; Ferreira, J. A.; Barbas, A.; Palma, A. S.; Novo, C.; Videira, P. A., Novel monoclonal antibody L2A5 specifically targeting sialyl-Tn and short glycans terminated by alpha-2-6 sialic acids. *Sci. Rep.* **2018,** *8*, 16.

(29) Ju, T. Z.; Lanneau, G. S.; Gautam, T.; Wang, Y. C.; Xia, B. Y.; Stowell, S. R.; Willard, M. T.; Wang, W. Y.; Xia, J. Y.; Zuna, R. E.; Laszik, Z.; Benbrook, D. M.; Hanigan, M. H.; Cummings, R. D., Human tumor antigens to and sialyl Tn arise from mutations in Cosmc. *Cancer Res.* **2008,** *68* (6), 1636-1646.

(30) Kaur, S.; Kumar, S.; Momi, N.; Sasson, A. R.; Batra, S. K., Mucins in pancreatic cancer and its microenvironment. *Nat. Rev. Gastroenterol. Hepatol.* **2013,** *10* (10), 607-620.

(31) Abutaily, A. S.; Addis, B. J.; Roche, W. R., Immunohistochemistry in the distinction between malignant mesothelioma and pulmonary adenocarcinoma: a critical evaluation of new antibodies. *J. Clin. Pathol.* **2002,** *55* (9), 662-668.

(32) Itzkowitz, S. H.; Yuan, M.; Montgomery, C. K.; Kjeldsen, T.; Takahashi, H. K.; Bigbee, W. L.; Kim, Y. S., Expression of Tn, sialosyl-Tn, and T antigens in human colon cancer. *Cancer Res.* **1989,** *49* (1), 197-204.

(33) David, L.; Nesland, J. M.; Clausen, H.; Carneiro, F.; Sobrinhosimoes, M., Simple mucin-type carbohydrate antigens (Tn, sialosyl-Tn and T) in gastric mucosa, carcinomas and metastases. *Apmis* **1992,** *100*, 162-172.

(34) Lan, Y.; Hao, C.; Zeng, X.; He, Y.; Zeng, P.; Guo, Z.; Zhang, L., Serum glycoprotein-derived N- and O-linked glycans as cancer biomarkers. *Am. J. Cancer Res.* **2016,** *6* (11), 2390-2415.

(35) Dennis, J. W.; Granovsky, M.; Warren, C. E., Glycoprotein glycosylation and cancer progression. *Biochim. Biophys. Acta-Gen. Subj.* **1999,** *1473* (1), 21-34.

(36) Peracaula, R.; Barrabes, S.; Sarrats, A.; Rudd, P. M.; de Llorens, R., Altered glycosylation in tumours focused to cancer diagnosis. *Dis. Markers* **2008,** *25* (4-5), 207-218.

(37) Perkins, G. L.; Slater, E. D.; Sanders, G. K.; Prichard, J. G., Serum tumor markers. *Am. Fam. Physician* **2003,** *68* (6), 1075-1082.

(38) Kurihara, M.; Ogawa, M.; Ohta, T.; Kurokawa, E.; Kitahara, T.; Murata, A.; Matsuda, K.; Kosaki, G.; Watanabe, T.; Wada, H., Radioimmunoassay for human pancreatic ribonuclease and measurement of serum immunoreactive pancreatic ribonuclease in patients with malignant tumors. *Cancer Res.* **1984,** *44* (5), 2240-2243.

(39) Peterson, L. M., Serum RNase in the diagnosis of pancreatic carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* **1979,** *76* (6), 2630-2634.

(40) Catalona, W. J.; Smith, D. S.; Ratliff, T. L.; Dodds, K. M.; Coplen, D. E.; Yuan, J. J. J.; Petros, J. A.; Andriole, G. L., Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. *N. Engl. J. Med.* **1991,** *324* (17), 1156-1161.

(41) Catalona, W. J.; Richie, J. P.; Ahmann, F. R.; Hudson, M. A.; Scardino, P. T.; Flanigan, R. C.; Dekernion, J. B.; Ratliff, T. L.; Kavoussi, L. R.; Dalkin, B. L.; Waters, W. B.; Macfarlane, M. T.; Southwick, P. C., Comparison of Digital Rectal Examination and Serum Prostate Specific Antigen in the Early Detection of Prostate Cancer: Results of a Multicenter Clinical Trial of 6,630 Men. *J. Urol.* **1994,** *151* (5), 1283-1290.

(42) Thompson, I. M.; Pauler, D. K.; Goodman, P. J.; Tangen, C. M.; Lucia, M. S.; Parnes, H. L.; Minasian, L. M.; Ford, L. G.; Lippman, S. M.; Crawford, E. D.; Crowley, J. J.; Coltman, C. A., Prevalence of prostate cancer among men with a prostate-specific antigen level <= 4.0 ng per milliliter. *N. Engl. J. Med.* **2004,** *350* (22), 2239-2246.

(43) Schultz, M. J.; Swindall, A. F.; Bellis, S. L., Regulation of the metastatic cell phenotype by sialylated glycans. *Cancer Metastasis Rev.* **2012,** *31* (3-4), 501-518.

(44) Miyagi, T.; Yamaguchi, K., Mammalian sialidases: Physiological and pathological roles in cellular functions. *Glycobiology* **2012,** *22* (7), 880-896.

(45) Harduin-Lepers, A.; Vallejo-Ruiz, V.; Krzewinski-Recchi, M. A.; Samyn-Petit, B.; Julien, S.; Delannoy, P., The human sialyltransferase family. *Biochimie* **2001,** *83* (8), 727-737.

(46) Pearce, O. M. T.; Laubli, H., Sialic acids in cancer biology and immunity. *Glycobiology* **2016,** *26* (2), 111-128.

(47) Brockhausen, I., Pathways of O-glycan biosynthesis in cancer cells. *Biochim. Biophys. Acta-Gen. Subj.* **1999,** *1473* (1), 67-95.

(48) Munkley, J., The Role of Sialyl-Tn in Cancer. *Int. J. Mol. Sci.* **2016,** *17* (3), 275-275.

(49) Ugorski, M.; Laskowska, A., Sialyl Lewis(a): a tumor-associated carbohydrate antigen involved in adhesion and metastatic potential of cancer cells. *Acta Biochim. Pol.* **2002,** *49* (2), 303-311.

(50) Cazet, A.; Julien, S.; Bobowski, M.; Krzewinski-Recchi, M.-A.; Harduin-Lepers, A.; Groux-Degroote, S.; Delannoy, P., Consequences of the expression of sialylated antigens in breast cancer. *Carbohydr. Res.* **2010,** *345* (10), 1377-1383.

(51) Perez-Garay, M.; Arteta, B.; Pages, L.; de Llorens, R.; de Bolos, C.; Vidal-Vanaclocha, F.; Peracaula, R., alpha 2,3-Sialyltransferase ST3Gal III Modulates Pancreatic Cancer Cell Motility and Adhesion In Vitro and Enhances Its Metastatic Potential In Vivo. *PLoS One* **2010,** *5* (9), 11.

(52) Tajiri, M.; Ohyama, C.; Wada, Y., Oligosaccharide profiles of the prostate specific antigen in free and complexed forms from the prostate cancer patient serum and in seminal plasma: A glycopeptide approach. *Glycobiology* **2008,** *18* (1), 2-8.

(53) Ohyama, C.; Hosono, M.; Nitta, K.; Oh-eda, M.; Yoshikawa, K.; Habuchi, T.; Arai, Y.; Fukuda, M., Carbohydrate structure and differential binding of prostate specific antigen to Maackia amurensis lectin between prostate cancer and benign prostate hypertrophy. *Glycobiology* **2004,** *14* (8), 671-679.

(54) Schauer, R., Characterization of sialic acids. *Methods Enzymol.* **1978,** *50*, 64-89.

(55) Klepach, T.; Carmichael, I.; Serianni, A. S., C-13-labeled N-acetyl-neuraminic acid in aqueous solution: Detection and quantification of acyclic keto, keto hydrate, and enol forms by C-13 NMR spectroscopy. *J. Am. Chem. Soc.* **2008,** *130* (36), 11892-11900.

(56) Renlund, M.; Aula, P.; Raivio, K. O.; Autio, S.; Sainio, K.; Rapola, J.; Koskela, S. L., Salla disease: a new lysosomal storage disorder with disturbed sialic acid metabolism. *Neurology* **1983,** *33* (1), 57-66.

(57) Paschke, E.; Trinkl, G.; Erwa, W.; Pavelka, M.; Mutz, I.; Roscher, A., Infantile type of sialic acid storage disease with sialuria. *Clin. Genet.* **1986,** *29* (5), 417-424.

(58) Chrostek, L.; Cylwik, B.; Korcz, W.; Krawiec, A.; Koput, A.; Supronowicz, Z.; Szmitkowski, M., Serum free sialic acid as a marker of alcohol abuse. *Alcoholism (NY)* **2007,** *31* (6), 996-1001.

(59) Lagana, A.; PardoMartinez, B.; Marino, A.; Fago, G.; Bizzarri, M., Determination of serum total lipid and free N-acetylneuraminic acid in genitourinary malignancies by fluorimetric high performance liquid chromatography. Relevance of free N-acetylneuraminic acid as tumour marker. *Clin. Chim. Acta* **1995,** *243* (2), 165-179.

(60) Mandal, C.; Mandal, C., Sialic-acid binding lectins. *Experientia* **1990,** *46* (5), 433-441.

(61) Lehmann, F.; Tiralongo, E.; Tiralongo, J., Sialic acid-specific lectins: occurrence, specificity and function. *Cell. Mol. Life Sci.* **2006,** *63* (12), 1331-1354.

(62) Hendrickson, O. D.; Zherdev, A. V., Analytical Application of Lectins. *Crit. Rev. Anal. Chem.* **2018,** *48* (4), 279-292.

(63) Quiocho, F. A., Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions. *Annu. Rev. Biochem.* **1986,** *55*, 287-315.

(64) Weis, W. I.; Drickamer, K., Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* **1996,** *65*, 441-473.

(65) Fersht, A. R.; Shi, J. P.; Knilljones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M. Y.; Winter, G., Hydrogen-Bonding and Biological Specificity Analyzed by Protein Engineering. *Nature* **1985,** *314* (6008), 235-238.

(66) Hudson, K. L.; Bartlett, G. J.; Diehl, R. C.; Agirre, J.; Gallagher, T.; Kiessling, L. L.; Woolfson, D. N., Carbohydrate-Aromatic Interactions in Proteins. *J. Am. Chem. Soc.* **2015,** *137* (48), 15152-15160.

(67) Crocker, P. R.; Vinson, M.; Kelm, S.; Drickamer, K., Molecular analysis of sialoside binding to sialoadhesin by NMR and site-directed mutagenesis. *Biochem. J.* **1999,** *341*, 355-361.

(68) Imberty, A.; Gautier, C.; Lescar, J.; Perez, S.; Wyns, L., An unusual carbohydrate binding site revealed by the structures of two Maackia amurensis lectins complexed with sialic acid-containing oligosaccharides. *J. Biol. Chem.* **2000,** *275* (23), 17541-17548.

(69) Bhavanandan, V. P.; Katlic, A. W., The interaction of wheat germ agglutinin with sialoglycoproteins. The role of sialic acid. *J. Biol. Chem.* **1979,** *254* (10), 4000-4008.

(70) Knibbs, R. N.; Osborne, S. E.; Glick, G. D.; Goldstein, I. J., Binding determinants of the sialic acid-specific lectin from the slug Limax flavus. *J. Biol. Chem.* **1993,** *268* (25), 18524-18531.

(71) Wright, C. S., 2.2 A-Resolution Structure-Analysis of 2 Refined N-Acetylneuraminyl-Lactose - Wheat-Germ-Agglutinin Isolectin Complexes. *J. Mol. Biol.* **1990,** *215* (4), 635-651.

(72) Yamamoto, K.; Konami, Y.; Irimura, T., Sialic Acid-Binding Motif of Maackia amurensis Lectins. *J. Biochem.* **1997,** *121* (4), 756-761.

(73) Konami, Y.; Yamamoto, K.; Osawa, T.; Irimura, T., Strong affinity of Maackia amurensis hemagglutinin (MAH) for sialic acid-containing Ser/Thr-linked carbohydrate chains of N-terminal octapeptides from human glycophorin A. *FEBS Lett.* **1994,** *342* (3), 334-338.

(74) Yuan, Q. Q.; He, J. L.; Niu, Y. Z.; Chen, J.; Zhao, Y. L.; Zhang, Y. C.; Yu, C., Sandwich-type biosensor for the detection of alpha 2,3-sialylated glycans based on fullerene-palladium-platinum alloy and 4-mercaptophenylboronic acid nanoparticle hybrids coupled with Au-methylene blue-MAL signal amplification. *Biosens. Bioelectron.* **2018,** *102*, 321-327.

(75) Niu, Y. Z.; He, J. L.; Li, Y. L.; Zhao, Y. L.; Xia, C. Y.; Yuan, G. L.; Zhang, L.; Zhang, Y. C.; Yu, C., Determination of alpha 2,3-sialylated glycans in human serum using a glassy carbon electrode modified with carboxylated multiwalled carbon nanotubes, a polyamidoamine dendrimer, and a glycan-recognizing lectin from Maackia Amurensis. *Microchim. Acta* **2016,** *183* (7), 2337-2344.

(76) Shibuya, N.; Goldstein, I. J.; Broekaert, W. F.; Nsimba-Lubaki, M.; Peeters, B.; Peumans, W. J., The elderberry (Sambucus nigra L.) bark lectin recognizes the Neu5Ac(alpha 2-6)Gal/GalNAc sequence. *J. Biol. Chem.* **1987,** *262* (4), 1596-601.

(77) Shibuya, N.; Goldstein, I. J.; Broekaert, W. F.; Nsimbalubaki, M.; Peeters, B.; Peumans, W. J., Fractionation of sialylated oligosaccharides, glycopeptides, and glycoproteins on immobilized elderberry (Sambucus nigra L.) bark lectin. *Arch. Biochem. Biophys.* **1987,** *254* (1), 1-8.

(78) Suttajit, M.; Winzler, R. J., Effect of Modification of N-Acetylneuraminic Acid on the Binding of Glycoproteins to Influenza Virus and on Susceptibility to Cleavage by Neuraminidase. *J. Biol. Chem.* **1971,** *246* (10), 3398-3404.

(79) Taatjes, D. J.; Roth, J.; Peumans, W.; Goldstein, I. J., Elderberry bark lectin--gold techniques for the detection of Neu5Ac (alpha 2,6) Gal/GalNAc sequences: applications and limitations. *Histochem.J.* **1988,** *20* (9), 478-490.

(80) Li, Y. L.; He, J. L.; Niu, Y. Z.; Yu, C., Ultrasensitive electrochemical biosensor based on reduced graphene oxide-tetraethylene pentamine-BMIMPF6 hybrids for the detection of alpha 2,6-sialylated glycans in human serum. *Biosens. Bioelectron.* **2015,** *74*, 953-959.

(81) Bertok, T.; Gemeiner, P.; Mikula, M.; Gemeiner, P.; Tkac, J., Ultrasensitive impedimetric lectin based biosensor for glycoproteins containing sialic acid. *Microchim. Acta* **2013,** *180* (1-2), 151-159.

(82) Crocker, P. R., Siglecs: sialic-acid-binding immunoglobulin-like lectins in cell–cell interactions and signalling. *Curr. Opin. Struct. Biol.* **2002,** *12* (5), 609-615.

(83) Attrill, H.; Takazawa, H.; Witt, S.; Kelm, S.; Isecke, R.; Brossmer, R.; Ando, T.; Ishida, H.; Kiso, M.; Crocker, P. R.; van Aalten, D. M. F., The structure of siglec-7 in complex with sialosides: leads for rational structure-based inhibitor design. *Biochem. J.* **2006,** *397*, 271-278.

(84) Sterner, E.; Flanagan, N.; Gildersleeve, J. C., Perspectives on Anti-Glycan Antibodies Gleaned from Development of a Community Resource Database. *ACS Chem. Biol.* **2016,** *11* (7), 1773-1783.

(85) Tommasone, S.; Allabush, F.; Tagger, Y. K.; Norman, J.; Kopf, M.; Tucker, J. H. R.; Mendes, P. M., The challenges of glycan recognition with natural and artificial receptors. *Chem. Soc. Rev.* **2019,** *48* (22), 5488-5505.

(86) Shitara, K.; Hanai, N.; Yoshida, H., Distribution of lung adenocarcinoma-associated antigens in human tissues and sera defined by monoclonal antibodies KM-52 and KM-93. *Cancer Res.* **1987,** *47* (5), 1267-1272.

(87) Haji-Ghassemi, O.; Blackler, R. J.; Young, N. M.; Evans, S. V., Antibody recognition of carbohydrate epitopes. *Glycobiology* **2015,** *25* (9), 920-952.

(88) Muller-Loennies, S.; MacKenzie, C. R.; Patenaude, S. I.; Evans, S. V.; Kosma, P.; Brade, H.; Brade, L.; Narang, S., Characterization of high affinity monoclonal antibodies specific for chlamydial lipopolysaccharide. *Glycobiology* **2000,** *10* (2), 121-130.

(89) Manimala, J. C.; Roach, T. A.; Li, Z. T.; Gildersleeve, J. C., High-throughput carbohydrate microarray profiling of 27 antibodies demonstrates widespread specificity problems. *Glycobiology* **2007,** *17* (8), 17C-23C.

(90) Tuerk, C.; Gold, L., Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **1990,** *249* (4968), 505-510.

(91) Gong, S.; Ren, H. L.; Tian, R. Y.; Lin, C.; Hu, P.; Li, Y. S.; Liu, Z. S.; Song, J.; Tang, F.; Zhou, Y.; Li, Z. H.; Zhang, Y. Y.; Lu, S. Y., A novel analytical probe binding to a potential carcinogenic factor of N-glycolylneuraminic acid by SELEX. *Biosens. Bioelectron.* **2013,** *49*, 547-554.

(92) Cho, S.; Lee, B. R.; Cho, B. K.; Kim, J. H.; Kim, B. G., In vitro selection of sialic acid specific RNA aptamer and its application to the rapid sensing of sialic acid modified sugars. *Biotechnol. Bioeng.* **2013,** *110* (3), 905-913.

(93) Jeong, S.; Eom, T. Y.; Kim, S. J.; Lee, S. W.; Yu, J., In vitro selection of the RNA aptamer against the Sialyl Lewis X and its inhibition of the cell adhesion. *Biochem. Biophys. Res. Commun.* **2001,** *281* (1), 237-243.

(94) Masud, M. M.; Kuwahara, M.; Ozaki, H.; Sawai, H., Sialyllactose-binding modified DNA aptamer bearing additional functionality by SELEX. *Bioorg. Med. Chem.* **2004,** *12* (5), 1111-1120.

(95) Li, M.; Lin, N.; Huang, Z.; Du, L.; Altier, C.; Fang, H.; Wang, B., Selecting Aptamers for a Glycoprotein through the Incorporation of the Boronic Acid Moiety. *J. Am. Chem. Soc.* **2008,** *130* (38), 12636-12638.

(96) Wu, X.; Li, Z.; Chen, X. X.; Fossey, J. S.; James, T. D.; Jiang, Y. B., Selective sensing of saccharides using simple boronic acids and their aggregates. *Chem. Soc. Rev.* **2013,** *42* (20), 8032-8048.

(97) Badjic, J. D.; Nelson, A.; Cantrill, S. J.; Turnbull, W. B.; Stoddart, J. F., Multivalency and cooperativity in supramolecular chemistry. *Accounts Chem. Res.* **2005,** *38* (9), 723-732.

(98) Davis, A. P.; Wareham, R. S., Carbohydrate recognition through noncovalent interactions: A challenge for biomimetic and supramolecular chemistry. *Angew. Chem. Int. Ed.* **1999,** *38* (20), 2978-2996.

(99) Kral, V.; Rusin, O.; Schmidtchen, F. P., Novel porphyrin-cryptand cyclic systems: Receptors for saccharide recognition in water. *Org. Lett.* **2001,** *3* (6), 873-876.

(100) Kikuchi, Y.; Tanaka, Y.; Sutarto, S.; Kobayashi, K.; Toi, H.; Aoyama, Y., Highly cooperative binding of alkyl glucopyranosides to the resorcinol cyclic tetramer due to intracomplex guest-guest hydrogen-bonding: solvophobicity/solvophilicity control by an alkyl group of the geometry, stoichiometry, stereoselectivity, and cooperativity. *J. Am. Chem. Soc.* **1992,** *114* (26), 10302-10306.

(101) Mooibroek, T. J.; Crump, M. P.; Davis, A. P., Synthesis and evaluation of a desymmetrised synthetic lectin: an approach to carbohydrate receptors with improved versatility. *Org. Biomol. Chem.* **2016,** *14* (6), 1930-1933.

(102) Schmuck, C.; Schwegmann, M., Recognition of anionic carbohydrates by an artificial receptor in water. *Org. Lett.* **2005,** *7* (16), 3517-3520.

(103) Mazik, M.; Kuschel, M., Amide, amino, hydroxy and aminopyridine groups as building blocks for carbohydrate receptors. *Eur. J. Org. Chem.* **2008,** *2008* (9), 1517-1526.

(104) Mazik, M.; Cavga, H., Molecular recognition of N-acetylneuraminic acid with acyclic benzimidazolium- and aminopyridine/guanidinium-based receptors. *J. Org. Chem.* **2007,** *72* (3), 831-838.

(105) Mazik, M.; Sicking, W., Molecular recognition of carbohydrates by artificial receptors: Systematic studies towards recognition motifs for carbohydrates. *Chem. Eur. J.* **2001,** *7* (3), 664-670.

(106) Mazik, M.; Hartmann, A.; Jones, P. G., Highly Effective Recognition of Carbohydrates by Phenanthroline-Based Receptors: alpha- versus beta-Anomer Binding Preference. *Chem.-Eur. J.* **2009,** *15* (36), 9147-9159.

(107) Mazik, M.; Buthe, A. C., Oxime-based receptors for mono- and Disaccharides. *J. Org. Chem.* **2007,** *72* (22), 8319-8326.

(108) Mazik, M.; Cavga, H., Carboxylate-based receptors for the recognition of carbohydrates in organic and aqueous media. *J. Org. Chem.* **2006,** *71* (8), 2957-2963.

(109) Mazik, M.; Kuschel, M., Highly effective acyclic carbohydrate receptors consisting of aminopyridine, imidazole, and indole recognition units. *Chem. Eur. J.* **2008,** *14* (8), 2405-2419.

(110) Geffert, C.; Kuschel, M.; Mazik, M., Molecular Recognition of N-Acetylneuraminic Acid by Acyclic Pyridinium- and Quinolinium-Based Receptors in Aqueous Media: Recognition through Combination of Cationic and Neutral Recognition Sites. *J. Org. Chem.* **2013,** *78* (2), 292-300.

(111) Carter, T. S.; Mooibroek, T. J.; Stewart, P. F. N.; Crump, M. P.; Galan, M. C.; Davis, A. P., Platform Synthetic Lectins for Divalent Carbohydrate Recognition in Water. *Angew. Chem. Int. Ed.* **2016,** *55* (32), 9311-9315.

(112) Ferrand, Y.; Crump, M. P.; Davis, A. P., A synthetic lectin analog for biomimetic disaccharide recognition. *Science* **2007,** *318* (5850), 619-622.

(113) Kugimiya, A.; Matsui, J.; Takeuchi, T., Sialic acid-imprinted polymers using noncovalent interactions. *Mater. Sci. Eng. C-Biomimetic Mater. Sens. Syst.* **1997,** *4* (4), 263-266.

(114) Peters, J. A., Interactions between boric acid derivatives and saccharides in aqueous media: Structures and stabilities of resulting esters. *Coord. Chem. Rev.* **2014,** *268*, 1-22.

(115) Nakatani, H.; Hiromi, K., Binding of m-nitrobenzeneboronic acid to the active site of subtilisin BPN′. *Biochim. Biophys. Acta* **1978,** *524* (2), 413-417.

(116) Springsteen, G.; Wang, B. H., A detailed examination of boronic acid-diol complexation. *Tetrahedron* **2002,** *58* (26), 5291-5300.

(117) Lorand, J. P.; Edwards, J. O., Polyol Complexes and Structure of the Benzeneboronate Ion. *J. Org. Chem.* **1959,** *24* (6), 769-774.

(118) Otsuka, H.; Uchimura, E.; Koshino, H.; Okano, T.; Kataoka, K., Anomalous binding profile of phenylboronic acid with N-acetylneuraminic acid (Neu5Ac) in aqueous solution with varying pH. *J. Am. Chem. Soc.* **2003,** *125* (12), 3493-3502.

(119) Mulla, H. R.; Agard, N. J.; Basu, A., 3-methoxycarbonyl-5-nitrophenyl boronic acid: high affinity diol recognition at neutral pH. *Bioorg. Med. Chem. Lett.* **2004,** *14* (1), 25-27.

(120) Wulff, G., Selective Binding to Polymers via Covalent Bonds - The Construction of Chiral Cavities as Specific Receptor-Site. *Pure Appl. Chem.* **1982,** *54* (11), 2093-2102.

(121) Zhu, L.; Shabbir, S. H.; Gray, M.; Lynch, V. M.; Sorey, S.; Anslyn, E. V., A structural investigation of the N-B interaction in an o-(N,N-dialkylaminomethyl)arylboronate system. *J. Am. Chem. Soc.* **2006,** *128* (4), 1222-1232.

(122) Brooks, W. L. A.; Deng, C. C.; Sumerlin, B. S., Structure–Reactivity Relationships in Boronic Acid–Diol Complexation. *ACS Omega* **2018,** *3* (12), 17863-17870.

(123) Norrild, J. C.; Eggert, H., Evidence for Mono- and Bisdentate Boronate Complexes of Glucose in the Furanose Form. Application of 1JC-C Coupling Constants as a Structural Probe. *J. Am. Chem. Soc.* **1995,** *117* (5), 1479-1484.

(124) Stones, D.; Manku, S.; Lu, X. S.; Hall, D. G., Modular solid-phase synthetic approach to optimize structural and electronic properties of oligoboronic acid receptors and sensors for the aqueous recognition of oligosaccharides. *Chem. Eur. J.* **2004,** *10* (1), 92-100.

(125) Yoon, J.; Czarnik, A. W., Fluorescent chemosensors of carbohydrates. A means of chemically communicating the binding of polyols in water based on chelation-enhanced quenching. *J. Am. Chem. Soc.* **1992,** *114* (14), 5874-5875.

(126) Tsukagoshi, K.; Shinkai, S., Specific complexation with mono- and disaccharides that can be detected by circular dichroism. *J. Org. Chem.* **1991,** *56* (13), 4089-4091.

(127) Yang, W. Q.; Fan, H. Y.; Gao, X. M.; Gao, S. H.; Karnati, V. V. R.; Ni, W. J.; Hooks, W. B.; Carson, J.; Weston, B.; Wang, B. H., The first fluorescent diboronic acid sensor specific for hepatocellular carcinoma cells expressing sialyl Lewis X. *Chem. Biol.* **2004,** *11* (4), 439-448.

(128) Regueiro-Figueroa, M.; Djanashvili, K.; Esteban-Gomez, D.; de Blas, A.; Platas-Iglesias, C.; Rodriguez-Blas, T., Towards Selective Recognition of Sialic Acid Through Simultaneous Binding to Its cis-Diol and Carboxylate Functions. *Eur. J. Org. Chem.* **2010,** (17), 3237-3248.

(129) Yang, W. Q.; Yan, J.; Fang, H.; Wang, B. H., The first fluorescent sensor for D-glucarate based on the cooperative action of boronic acid and guanidinium groups. *Chem. Commun.* **2003,** (6), 792-793.

(130) Wiskur, S. L.; Lavigne, J. L.; Metzger, A.; Tobey, S. L.; Lynch, V.; Anslyn, E. V., Thermodynamic analysis of receptors based on guanidinium/boronic acid groups for the complexation of carboxylates, alpha-hydroxycarboxylates, and diols: Driving force for binding and cooperativity. *Chem. Eur. J.* **2004,** *10* (15), 3792-3804.

(131) Wright, A. T.; Zhong, Z. L.; Anslyn, E. V., A functional assay for heparin in serum using a designed synthetic receptor. *Angew. Chem. Int. Ed.* **2005,** *44* (35), 5679-5682.

(132) Nagai, Y.; Kobayashi, K.; Toi, H.; Aoyama, Y., Stabilization of Sugar-Boronic Esters of Indolylboronic Acid in Water via Sugar–Indole Interaction: A Notable Selectivity in Oligosaccharides. *Bull. Chem. Soc. Jpn.* **1993,** *66* (10), 2965-2971.

(133) Chaudhary, P. M.; Murthy, R. V.; Yadav, R.; Kikkeri, R., A rationally designed peptidomimetic biosensor for sialic acid on cell surfaces. *Chem. Commun.* **2015,** *51* (38), 8112-8115.

(134) Kowalczyk, W.; Sanchez, J.; Kraaz, P.; Hutt, O. E.; Haylock, D. N.; Duggan, P. J., The binding of boronated peptides to low affinity mammalian saccharides. *Pept. Sci.* **2018,** *110* (3), 12.

(135) Duggan, P. J.; Offermann, D. A., Remarkably selective saccharide recognition by solid-supported peptide boronic acids. *Tetrahedron* **2009,** *65* (1), 109-114.

(136) Dai, C. F.; Sagwal, A.; Cheng, Y. F.; Peng, H. J.; Chen, W. X.; Wang, B. H., Carbohydrate biomarker recognition using synthetic lectin mimics. *Pure Appl. Chem.* **2012,** *84* (11), 2479-2498.

(137) Bruen, D.; Delaney, C.; Diamond, D.; Florea, L., Fluorescent Probes for Sugar Detection. *ACS Appl. Mater. Interfaces* **2018,** *10* (44), 38431-38437.

(138) Deshayes, S.; Cabral, H.; Ishii, T.; Miura, Y.; Kobayashi, S.; Yamashita, T.; Matsumoto, A.; Miyahara, Y.; Nishiyama, N.; Kataoka, K., Phenylboronic Acid-Installed Polymeric Micelles for Targeting Sialylated Epitopes in Solid Tumors. *J. Am. Chem. Soc.* **2013,** *135* (41), 15501-15507.

(139) Stephenson-Brown, A.; Wang, H. C.; Iqbal, P.; Preece, J. A.; Long, Y. T.; Fossey, J. S.; James, T. D.; Mendes, P. M., Glucose selective Surface Plasmon Resonance-based bis-boronic acid sensor. *Analyst* **2013,** *138* (23), 7140-7145.

(140) Matsumoto, A.; Sato, N.; Kataoka, K.; Miyahara, Y., Noninvasive Sialic Acid Detection at Cell Membrane by Using Phenylboronic Acid Modified Self-Assembled Monolayer Gold Electrode. *J. Am. Chem. Soc.* **2009,** *131* (34), 12022-+.

(141) Calisir, M.; Bakhshpour, M.; Yavuz, H.; Denizli, A., HbA1c detection via high-sensitive boronate based surface plasmon resonance sensor. *Sens. Actuator B-Chem.* **2020,** *306*, 8.

(142) Mitchell, P.; Tommasone, S.; Angioletti-Uberti, S.; Bowen, J.; Mendes, P. M., Precise Generation of Selective Surface-Confined Glycoprotein Recognition Sites. *ACS Appl. Bio Mater.* **2019,** *2* (6), 2617-2623.

(143) Stephenson-Brown, A.; Acton, A. L.; Preece, J. A.; Fossey, J. S.; Mendes, P. M., Selective glycoprotein detection through covalent templating and allosteric click-imprinting. *Chem. Sci.* **2015,** *6* (9), 5114-5119.

(144) Gunasekara, R. W.; Zhao, Y., A General Method for Selective Recognition of Monosaccharides and Oligosaccharides in Water. *J. Am. Chem. Soc.* **2017,** *139* (2), 829-835.

(145) Torssell, K., Zur kenntnis der arylborsauren 0. 3. Bromierung der tolylborsauren nach wohl-ziegler. *Ark. Kemi* **1957,** *10*, 507-511.

(146) Dowlut, M.; Hall, D. G., An improved class of sugar-binding boronic acids, soluble and capable of complexing glycosides in neutral water. *J. Am. Chem. Soc.* **2006,** *128* (13), 4226-4227.

(147) Berube, M.; Dowlut, M.; Hall, D. G., Benzoboroxoles as efficient glycopyranoside-binding agents in physiological conditions: Structure and selectivity of complex formation. *J. Org. Chem.* **2008,** *73* (17), 6471-6479.

(148) Sorensen, M. D.; Martins, R.; Hindsgaul, O., Assessing the terminal glycosylation of a glycoprotein by the naked eye. *Angew. Chem. Int. Ed.* **2007,** *46* (14), 2403-2407.

(149) Claes, D.; Memmel, E.; Holzapfel, M.; Seibel, J.; Maison, W., High-Affinity Carbohydrate Binding by Trimeric Benzoboroxoles Measured on Carbohydrate Arrays. *ChemBioChem* **2014,** *15* (16), 2450-2457.

(150) Rowe, L.; El Khoury, G.; Lowe, C. R., A benzoboroxole-based affinity ligand for glycoprotein purification at physiological pH. *J. Mol. Recognit.* **2016,** *29* (5), 232-238.

(151) Otremba, T.; Ravoo, B. J., Linear and Cyclic Carbohydrate Receptors Based on Peptides Modified with Boronic Acids. *ChemistrySelect* **2016,** *1* (9), 2079-2084.

(152) Pal, A.; Berube, M.; Hall, D. G., Design, Synthesis, and Screening of a Library of Peptidyl Bis(Boroxoles) as Oligosaccharide Receptors in Water: Identification of a Receptor for the Tumor Marker TF-Antigen Disaccharide. *Angew. Chem. Int. Ed.* **2010,** *49* (8), 1492-1495.

(153) Schumacher, S.; Katterle, M.; Hettrich, C.; Paulke, B. R.; Hall, D. G.; Scheller, F. W.; Gajovic-Eichelmann, N., Label-free detection of enhanced saccharide binding at pH 7.4 to nanoparticulate benzoboroxole based receptor units. *J. Mol. Recognit.* **2011,** *24* (6), 953-959.

(154) Tommasone, S.; Tagger, Y. K.; Mendes, P. M., Targeting Oligosaccharides and Glycoconjugates Using Superselective Binding Scaffolds. *Adv. Funct. Mater.* **2020**, 2002298.

(155) Nishitani, S.; Maekawa, Y.; Sakata, T., Understanding the Molecular Structure of the Sialic Acid-Phenylboronic Acid Complex by using a Combined NMR Spectroscopy and DFT Study: Toward Sialic Acid Detection at Cell Membranes. *ChemistryOpen* **2018,** *7* (7), 513-519.

(156) Djanashvili, K.; Frullano, L.; Peters, J. A., Molecular recognition of sialic acid end groups by phenylboronates. *Chem. Eur. J.* **2005,** *11* (13), 4010-4018.

(157) Regueiro-Figueroa, M.; Djanashvili, K.; Esteban-Gomez, D.; Chauvin, T.; Toth, E.; de Blas, A.; Rodriguez-Blas, T.; Platas-Igleslas, C., Molecular Recognition of Sialic Acid by Lanthanide(III) Complexes through Cooperative Two-Site Binding. *Inorg. Chem.* **2010,** *49* (9), 4212-4223.

(158) Matsumoto, A.; Stephenson-Brown, A. J.; Khan, T.; Miyazawa, T.; Cabral, H.; Kataoka, K.; Miyahara, Y., Heterocyclic boronic acids display sialic acid selective binding in a hypoxic tumor relevant acidic environment. *Chem. Sci.* **2017,** *8* (9), 6165-6170.

(159) Yamamoto, M.; Takeuchi, M.; Shinkai, S., Molecular design of a PET-based chemosensor for uronic acids and sialic acids utilizing a cooperative action of boronic acid and metal chelate. *Tetrahedron* **1998,** *54* (13), 3125-3140.

(160) Xu, S. M.; Che, S. T.; Ma, P. Y.; Zhang, F. M.; Xu, L. B.; Liu, X.; Wang, X. H.; Song, D. Q.; Sun, Y., One-step fabrication of boronic-acid-functionalized carbon dots for the detection of sialic acid. *Talanta* **2019,** *197*, 548-552.

(161) Sankoh, S.; Thammakhet, C.; Numnuam, A.; Limbut, W.; Kanatharana, P.; Thavarungkul, P., 4-mercaptophenylboronic acid functionalized gold nanoparticles for colorimetric sialic acid detection. *Biosens. Bioelectron.* **2016,** *85*, 743-750.

(162) Shinde, S.; El-Schich, Z.; Malakpour, A.; Wan, W.; Dizeyi, N.; Mohammadi, R.; Rurack, K.; Wingren, A. G.; Sellergren, B., Sialic Acid-Imprinted Fluorescent Core-Shell Particles for Selective Labeling of Cell Surface Glycans. *J. Am. Chem. Soc.* **2015,** *137* (43), 13908-13912.

(163) Jayeoye, T. J.; Cheewasedtham, W.; Putson, C.; Rujiralai, T., Colorimetric determination of sialic acid based on boronic acid-mediated aggregation of gold nanoparticles. *Microchim. Acta* **2018,** *185* (9), 8.

(164) Wellington, N.; Macklai, S.; Britz-McKibbin, P., Elucidating the Anomalous Binding Enhancement of Isoquinoline Boronic Acid for Sialic Acid Under Acidic Conditions: Expanding Biorecognition Beyond Vicinal Diols. *Chem. Eur. J.* **2019,** *25* (67), 15277-15280.

(165) Pizer, R., Boron acid complexation reactions with polyols and alpha-hydroxy carboxylic acids: Equilibria, reaction mechanisms, saccharide recognition. *Inorg. Chim. Acta* **2017,** *467*, 194-197.

(166) Kustin, K.; Pizer, R., Temperature-Jump Study of Rate and Mechanism of Boric Acid-Tartaric Acid Complexation. *J. Am. Chem. Soc.* **1969,** *91* (2), 317-&.

(167) Zhao, J. Z.; Fyles, T. M.; James, T. D., Chiral binol-bisboronic acid as fluorescence sensor for sugar acids. *Angew. Chem. Int. Ed.* **2004,** *43* (26), 3461-3464.

(168) Djanashvili, K.; Koning, G. A.; van der Meer, A.; Wolterbeek, H. T.; Peters, J. A., Phenylboronate Tb-160 complexes for molecular recognition of glycoproteins expressed on tumor cells. *Contrast Media Mol. Imaging* **2007,** *2* (1), 35-41.

(169) Levonis, S. M.; Kiefel, M. J.; Houston, T. A., Boronolectin with divergent fluorescent response specific for free sialic acid. *Chem. Commun.* **2009,** (17), 2278-2280.

(170) Ulrich, S.; Dumy, P., Probing secondary interactions in biomolecular recognition by dynamic combinatorial chemistry. *Chem. Commun.* **2014,** *50* (44), 5810-5825.

(171) Brady, P. A.; BonarLaw, R. P.; Rowan, S. J.; Suckling, C. J.; Sanders, J. K. M., 'Living' macrolactonisation: Thermodynamically-controlled cyclisation and interconversion of oligocholates. *Chem. Commun.* **1996,** (3), 319-320.

(172) Huc, I.; Lehn, J. M., Virtual combinatorial libraries: Dynamic generation of molecular and supramolecular diversity by self-assembly. *Proc. Natl. Acad. Sci. U. S. A.* **1997,** *94* (6), 2106-2110.

(173) Lehn, J. M., Dynamic combinatorial chemistry and virtual combinatorial libraries. *Chem. Eur. J.* **1999,** *5* (9), 2455-2463.

(174) Frei, P.; Hevey, R.; Ernst, B., Dynamic Combinatorial Chemistry: A New Methodology Comes of Age. *Chem. Eur. J.* **2019,** *25* (1), 60-73.

(175) Mondal, M.; Hirsch, A. K. H., Dynamic combinatorial chemistry: a tool to facilitate the identification of inhibitors for protein targets. *Chem. Soc. Rev.* **2015,** *44* (8), 2455-2488.

(176) Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J.-L.; Sanders, J. K. M.; Otto, S., Dynamic Combinatorial Chemistry. *Chem. Rev.* **2006,** *106* (9), 3652-3711.

(177) Otto, S.; Furlan, R. L. E.; Sanders, J. K. M., Selection and amplification of hosts from dynamic combinatorial libraries of macrocyclic disulfides. *Science* **2002,** *297* (5581), 590-593.

(178) Otto, S.; Furlan, R. L. E.; Sanders, J. K. M., Recent developments in dynamic combinatorial chemistry. *Curr. Opin. Chem. Biol.* **2002,** *6* (3), 321-327.

(179) Hartman, A. M.; Gierse, R. M.; Hirsch, A. K. H., Protein-Templated Dynamic Combinatorial Chemistry: Brief Overview and Experimental Protocol. *Eur. J. Org. Chem.* **2019,** *2019* (22), 3581-3590.

(180) Huang, R.; Leung, I. K. H., Protein-Directed Dynamic Combinatorial Chemistry: A Guide to Protein Ligand and Inhibitor Discovery. *Molecules* **2016,** *21* (7), 910.

(181) Hamieh, S.; Ludlow, R. F.; Perraud, O.; West, K. R.; Mattia, E.; Otto, S., A Synthetic Receptor for Nicotine from a Dynamic Combinatorial Library. *Org. Lett.* **2012,** *14* (21), 5404-5407.

(182) Chung, M.-K.; Severin, K.; Lee, S. J.; Waters, M. L.; Gagné, M. R., Constitutionally selective amplification of multicomponent 84-membered macrocyclic hosts for (−)-cytidine•H+. *Chem. Sci.* **2011,** *2* (4), 744-747.

(183) Lam, R. T. S.; Belenguer, A.; Roberts, S. L.; Naumann, C.; Jarrosson, T.; Otto, S.; Sanders, J. K. M., Amplification of Acetylcholine-Binding Catenanes from Dynamic Combinatorial Libraries. *Science* **2005,** *308* (5722), 667.

(184) Ulatowski, F.; Jurczak, J., Oligocarboxylates as useful templates in dynamic combinatorial chemistry. *Pure Appl. Chem.* **2017,** *89* (6), 801-807.

(185) Klepel, F.; Ravoo, B. J., A dynamic combinatorial library for biomimetic recognition of dipeptides in water. *Beilstein J. Org. Chem.* **2020,** *16*, 1588-1595.

(186) Bru, M.; Alfonso, I.; Bolte, M.; Burguete, M. I.; Luis, S. V., Structurally disfavoured pseudopeptidic macrocycles through anion templation. *Chem. Commun.* **2011,** *47* (1), 283-285.

(187) Beeren, S. R.; Sanders, J. K. M., Ferrocene-amino acid macrocycles as hydrazone-based receptors for anions. *Chem. Sci.* **2011,** *2* (8), 1560-1567.

(188) Ceborska, M.; Tarnowska, A.; Ziach, K.; Jurczak, J., Dynamic combinatorial libraries of macrocycles derived from phthalic aldehydes and α,ω-diamines. *Tetrahedron* **2010,** *66* (49), 9532-9537.

(189) Klein, J. M.; Saggiomo, V.; Reck, L.; McPartlin, M.; Pantoş, G. D.; Lüning, U.; Sanders, J. K. M., A remarkably flexible and selective receptor for Ba2+ amplified from a hydrazone dynamic combinatorial library. *Chem. Commun.* **2011,** *47* (12), 3371-3373.

(190) Klein, J. M.; Saggiomo, V.; Reck, L.; Lüning, U.; Sanders, J. K. M., Dynamic combinatorial libraries for the recognition of heavy metal ions. *Org. Biomol. Chem.* **2012,** *10* (1), 60-66.

(191) González-Álvarez, A.; Alfonso, I.; López-Ortiz, F.; Aguirre, Á.; García-Granda, S.; Gotor, V., Selective Host Amplification from a Dynamic Combinatorial Library of Oligoimines for the Syntheses of Different Optically Active Polyazamacrocycles. *Eur. J. Org. Chem.* **2004,** *2004* (5), 1117-1127.

(192) Rauschenberg, M.; Bomke, S.; Karst, U.; Ravoo, B. J., Dynamic Peptides as Biomimetic Carbohydrate Receptors. *Angew. Chem. Int. Ed.* **2010,** *49* (40), 7340-7345.

(193) Yang, D.; von Krbek, L. K. S.; Yu, L.; Ronson, T. K.; Thoburn, J. D.; Carpenter, J. P.; Greenfield, J. L.; Howe, D. J.; Wu, B.; Nitschke, J. R., Glucose Binding Drives Reconfiguration of a Dynamic Library of Urea-Containing Metal–Organic Assemblies. *Angew. Chem. Int. Ed.* **2021,** *60* (9), 4485-4490.

Chapter 2 - Techniques

2.1 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a technique employed for the direct measurement of the heat and thermodynamic parameters associated with a binding event[1]. The binding event is generally represented as the equilibrium between a receptor (R) and a ligand (L) to form the corresponding complex (RL) (Equation 2.1).

$$R + L \rightleftharpoons RL \qquad \text{Equation 2.1}$$

During an ITC experiment the ligand is titrated into a sample cell containing the receptor. At each injection, the interaction between the two species causes a change in the heat. The binding event causes a release or absorption of heat for exothermic and endothermic reactions, respectively. The heat associated with the binding event Q (cal) is defined by Equation 2.2, where V_0 is the volume of the sample cell, ΔH is the enthalpy of binding (cal mol^{-1}) and $[RL]$ is the concentration of complex[2].

$$Q = [RL]V_0\Delta H \qquad \text{Equation 2.2}$$

The heat for each injection is proportional to the amount of ligand that interacts with the receptor. As the titration progresses a greater proportion of R is already engaged in the complex, therefore a smaller fraction of R is available to bind the injected ligand. This results in a decrease of the heat change as the titration progress towards saturation. As can be seen in Figure 2.1.

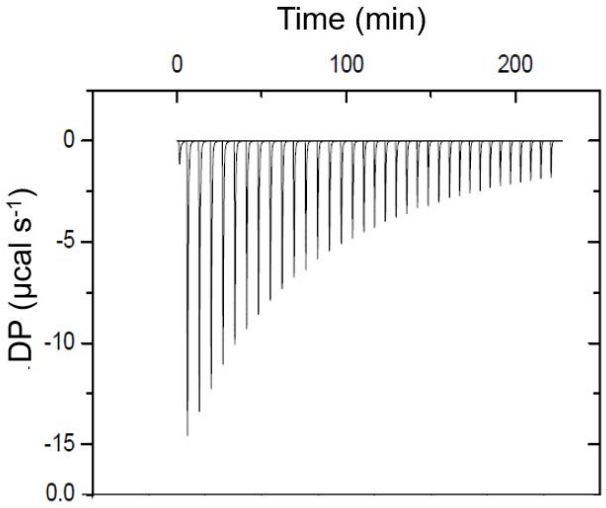


Figure 2.1. Titration graph for an exothermic binding event.

The data collected during the titration are integrated and the fitted to provide the thermodynamic parameters associate with the binding event. In particular, ITC is largely used to determine the binding constant K_a (M^{-1}) between receptor and ligand (Equation 2.3).

$$K_a = \frac{[RL]}{[R][L]} \qquad \text{Equation 2.3}$$

From the K_a the dissociation constant K_d can be determined from Equation 2.4.

$$K_d = \frac{1}{K_a} \qquad \text{Equation 2.4}$$

In addition, ITC provides the stoichiometry of binding (n), the enthalpy of binding (ΔH, cal mol^{-1}) and entropy of binding (ΔS, cal mol^{-1} deg^{-1}). The free energy of binding (ΔG, cal mol^{-1}) is determined from (Equation 2.5), with the molar gas constant R being 1.987 cal mol^{-1} K^{-1} and T, the absolute temperature, in K.

$$\Delta G = -RT \ln K_a = \Delta H - T\Delta S \qquad \text{Equation 2.5}$$

ITC can be employed to study the thermodynamics of any binding event as virtually all interactions are characterised by heat changes. ITC is widely employed to assess the binding of

proteins with other macromolecules, carbohydrates or small molecules[3]. ITC can also be applied to other systems, such as binding between small molecules[4-5] or with nanoparticle systems[6]. One of the significant benefits of ITC versus other techniques is that it does not require chemical modifications (*e.g.* labelling) or surface immobilisation of the receptor[3, 7].

A typical ITC instrument is composed of a sample cell and an injection syringe (Figure 2.2). The sample cell sits in an adiabatic shield together with the reference cell[1]. The sample cell contains the receptor solution and in this the ligand is injected during the titration. The volume of the sample cell depends upon the specific instrument but typically ranges from 300 µl to 3 ml. The reference cell, the volume of which is identical to the sample cell, contains only the solvent and no receptor. For experiments performed in aqueous media the reference cell usually contains pure water, whilst for experiments in organic media the reference cell is filled with the corresponding solvent. Before the start of the experiment both cells are heated to a selected temperature which can range between 2 and 80 °C[8], but it is normally set at 25 °C.

A thermoelectric device is present between the two cells in order to measure the difference between their temperatures (ΔT)[1]. A constant power ($\leq$ 1 mW) is provided to the reference cell during the experiment, which activates a feedback circuit to maintain the ΔT = 0 °C between the cells by suppling power to the sample cell[1]. When no binding event is occurring the differential power (DP), also known as feedback power, provided to the sample cell is constant and it represents the baseline of the experiment[1]. On the other hand, the heat change associated with the binding event causes a variation of the sample cell temperature. For a exothermic reaction the released heat provides, at least partially, the heat otherwise supplied by the DP[8]. Thus, less power, compare to the baseline, is required in order to maintain the ΔT = 0 °C and the DP decreases[1]. Conversely, for exothermic reactions the absorption of heat triggers an increase of the DP as more power is needed to keep the ΔT = 0 °C[1]. The DP supplied to the

sample cell is the experimental data recorded during the titration (DP, µcal s^{-1}), which can be plotted against the time[2]. For an exothermic reaction this corresponds to negative peaks as the DP is reduced compared to the baseline (Figure 2.1). For endothermic events, the positive DP results in positive peaks.

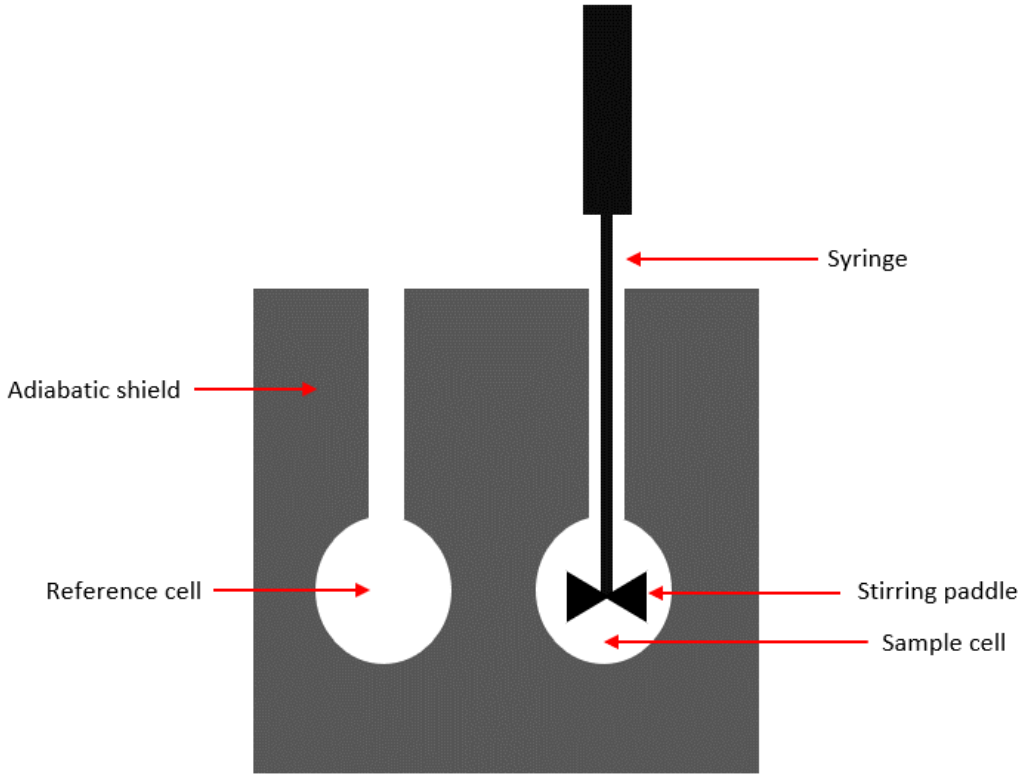


Figure 2.2. Schematic representation of the ITC instrument.

The heat associated to the injection (ΔH, cal mol^{-1}) is the time integral of the deflection from the baseline which corresponds to the DP applied to the sample cell from the beginning of the injection until the system returns to the baseline[9]. In other words, the integral of the peak area provides the ΔH associated with the injection.

Furthermore, the ΔH is plotted against the molar ratio to give the calorimetric binding curve. The shape of the binding curve influences the fitting and the accuracy of the calculated thermodynamic parameters[10]. The shape of the binding curves is determined by a parameter

called *c*-value. The *c*-value is given by Equation 2.6 and is dependent on the binding stoichiometry *n*, binding constant K_a and the concentration of receptor in the sample cell $[R]_{cell}$[7].

$$c = nK_a[R]_{cell} \quad \text{Equation 2.6}$$

For a binding stoichiometry of 1 ($n = 1$) this can be simplified to (Equation 2.7)

$$c = K_a[R]_{cell} \quad \text{Equation 2.7}$$

The ideal *c*-value lies between 10 and 500, as this affords a well-defined sigmoidal binding curve (Figure 2.3)[10]. Systems with very high affinities ($K_a > 10^7\ M^{-1}$) cannot be easily studied by ITC as they present high *c*-values. For *c*-values above 1000 the titration peaks are all of the same intensity and the integration gives a flat curve hampering the determination of the thermodynamic parameters[9],[11]. In order to reduce the *c*-value and obtain a sigmoidal curve the concentration of *R* must be lowered. However, exceedingly low concentrations (10^{-7} M) could generate signals which are below the limit of detection[9, 12]. To overcome this limitation very high affinity systems can be studied by ITC through competition studies, with more details given in reference[13].

On the other hand, for low affinity systems ($K_a < 10^4\ M^{-1}$) the concentrations required for a *c*-value ≥ 10 are generally too high to be achieved due to either solubility or availability issues[7]. Thus, for low affinity systems the *c*-value is generally < 10. At low *c*-values ($c \sim 1.0$) high concentrations of ligand are required to reach saturation and the data of the first part of the sigmoidal curve cannot be collected[9]. Moreover, for extremely low *c*-values ($c < 0.1$) is not possible to determine with accuracy all the thermodynamic parameters. This can be overcome by freezing the number of binding site (*n*) as this is often known from prior studies or literature data[14]. By fixing the binding stoichiometry, experiments with extremely low *c*-values will still provide a reliable K_a value[7]. However, if concentrations of the solutions are not accurately

known for experiments where n is fixed, the fitting will provide erroneous ΔH values[7, 14]. Thus, is pivotal to precisely know the concentration of both ligand and receptor prior analysis.

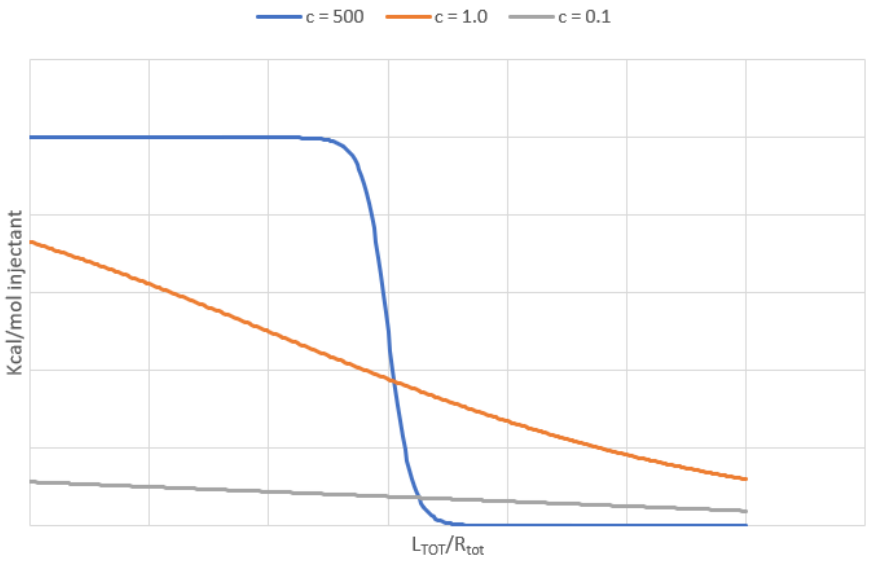


Figure 2.3. The effect of the c-value on the shape of the binding curve.

Furthermore, the c-value can be used to determine the required experimental receptor concentration when K_a is known, or extrapolated from a similar system, as shown by Equation 2.8.[9]

$$[R]_{cell} = \frac{c}{K_a} \qquad \text{Equation 2.8}$$

Once the concentration of receptor has been established, the ligand concentration can consequently be determined. The ligand concentration must be high enough to guarantee the saturation of the system. For high affinity systems, with $n = 1$ and $c > 10$, the saturation can be achieved for a molar ratio receptor to ligand of 1 to 2 or 1 to 3[15]. The molar ratio is the ratio of the receptor concentration to ligand concentration in the sample cell at the end of the titration[14], as per Equation 2.9.

$$molar\ ratio = \frac{[R]}{[L]} \qquad \text{Equation 2.9}$$

For systems with low *c*-values (c < 10) several equivalents are required (molar ratio up to 1 to 10)[7] to push the weak affinity systems towards saturation. Based on the required molar ratio and the total injected volume (40 – 500 µl)[16] the ligand concentration can be calculated[9].

The ligand solution is placed in the injection syringe which is then located inside the sample cell. The ligand solution is injected in the sample cell sequentially, usually over 10 to 60 injections. Each injection consists of a small volume (1 - 10 µl)[17] of titrant, with the volume selected in order to provide a heat change large enough to be detected[9]. The magnitude of the heat change required is determined by the instrument sensitivity which generally 0.1 µcal for the VP-ITC instrument used in this work[1, 8]. The spacing, or time interval, between each injection must be long enough to guarantee the return to the baseline of the system prior the following injection and is normally set between 180 s and 480 s [9]. The injection syringe is also responsible for mixing the cell content through the stirring paddle attached to its end[1]. The stirring is typically set at a rate between 250 and 300 RPM[15]. The experimental baseline is determined prior to the injection of the ligand but with the stirring activated in order to include the heat associated with the stirring in the baseline[1]. In addition to the above-mentioned parameters the reference power is also selected prior the experiment. The reference power is the value near to which the differential power will equilibrate[8]. When large exothermic reactions are expected the selected reference power must be larger (~ 30 µCal sec^{-1}), whilst for endothermic reaction a smaller reference power is required (~ 2 µCal sec^{-1}).

For each ITC experiment a control titration must be conducted in order to establish the heat of dilution[9]. This consists in the injection of the ligand solution into buffer in order to determine the heat associated to the dilution of a highly concentrated solution into a big volume[9]. In addition, mismatch of temperature or buffer can cause large heat of dilution. In particular, buffer mismatch must be avoided as they result in big heat changes and anomalous titration curves[8].

Hence, the heat of dilution is routinely measured and subtracted from the ligand-receptor titration data.

The collected data are then analysed with the software provided by the manufacturer. For the analysis of the data an appropriate binding model must be selected. The binding model can be one-binding site, two or more independent binding sites or a cooperative sites models[12]. Once the binding model is selected the software initially estimates the thermodynamic parameters K_a ΔH and n, unless n is fixed. These preliminary values are then used to calculate the heat associated with each injection ΔQ and compared to the experimental values. This allows a better estimation of the thermodynamic parameters by least-squares fitting[18]. This procedure is iterated until the final values for n, K_a and ΔH are provided.

2.2 Nuclear Magnetic Resonance Spectroscopy

Nuclear Magnetic Resonance (NMR) Spectroscopy, first described by Bloch and Purcell in 1946[19], is an analytical technique employed for defining the structure of organic molecules[20]. This technique is based on atoms possessing a quantum property called spin, which is frequently visualised as the spinning of the nucleus around its axis[21]. Atoms presenting spin include, but are not limited to, ^{1}H, ^{11}B, ^{13}C, ^{15}N, ^{17}O, ^{19}F and ^{31}P[22]. Atoms with spin also possess a non-zero magnetic moment ($\mu \neq 0$). Upon application of a magnetic field (B_0) the magnetic dipole of these nuclei aligns along the axis of the applied field[21]. The number of possible orientations (S) is defined by Equation 2.10.

$$S = 2I + 1 \qquad \text{Equation 2.10}$$

The variable *I* represents the nuclear spin quantum number which is a fixed value for each atom[19]. For instance, protons have *I* = ½ and therefore can assume two different orientations, parallel and antiparallel to the applied magnetic field[21]. When experiencing an external magnetic field the two orientations exhibit two different energy levels, thus a difference in energy (ΔE) exist between them with this being dependent on the applied magnetic field B_0 and the Planck constant *h*, as per Equation 2.11[22].

$$\Delta E = \gamma \frac{h}{2\pi} B_o \qquad \text{Equation 2.11}$$

The variable γ is the gyromagnetic ratio, a property defined by the nucleus type. When electromagnetic radiation of frequency *v*, corresponding to an energy equal to ΔE, is supplied the nuclei are excited to higher energy levels producing a spin inversion[22]. Given that $\Delta E = hv$ this can be substituted into Equation 2.11 resulting in Equation 2.12 and consequently obtaining Equation 2.13.

$$hv = \Delta E = \gamma \frac{\mathrm{h}}{2\pi} \mathrm{B}_0 \qquad \text{Equation 2.12}$$

$$v = \frac{\gamma}{2\pi} B_0 \qquad \text{Equation 2.13}$$

After excitation, the nuclei relax each emitting at their resonant frequency *v* resulting in an interferogram of superimposed waves known as FID (free induction decay)[22]. The signal is then processed by Fourier transformation to produce the NMR spectrum. In the NMR spectrum each of the resonant frequencies for each nucleus is displayed as a signal, after transformation into a chemical shift (δ) and quoted in ppm[22].

$$\delta = \frac{(v_i - v_R)}{v_0} \qquad \text{Equation 2.14}$$

The chemical shift is defined by the ν_i, the frequency in Hz of the nucleus, ν_R, a reference frequency, and ν_0 the frequency of the instrument in MHz, as per Equation 2.14. Moreover, the chemical shift is a measure of the influence of the chemical environment on the resonant frequency of a specific nucleus[19]. The chemical environment in which the nucleus is present defines the shielding to which the nucleus is subject to. The shielding is determined by the electrons near a nucleus providing an opposing magnetic field to B_0, which reduces the effective magnetic field seen by the nucleus, resulting in a lower resonance frequency following Equation 2.13. Therefore, for higher electron densities the shielding is greater and the signal is shifted to higher fields, corresponding to a smaller chemical shift. Conversely, when the nucleus is de-shielded this will result in a shift to lower fields, providing larger δ values. For instance, the OH proton of ethanol (**1**) is downfield (δ = 2.56 ppm) compared to the SH proton of ethanethiol (**2**) (δ = 1.46 ppm) (Figure 2.4)[22]. This is due to the higher electronegativity of the O atom providing a lower shielding to the proton linked to it, when compared to the S atom. The effect of the heteroatom electronegativity also influences the chemical shift of the α proton (methylene protons), whilst the effect on β protons is minimal.

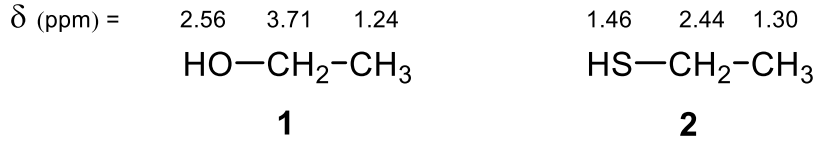


Figure 2.4. Ethanol (**1**) and ethanethiol (**2**) and their respective chemical shift values in ppm.

The chemical shift is pivotal for the interpretation ^{1}H NMR spectra. Other key parameters include signal multiplicity, coupling constants (J) and signal integrals[19]. The signals can be displayed as singlets or as a multiplets, such as doubles, triplets, quartets, doublet of doublets, etc. The multiplicity provides information on the number of NMR-active nuclei that are within

a 3-bond distance from a given nucleus[19]. This occurs because the local field of a given nucleus is influenced by the magnetic moment of the nearby nuclei.

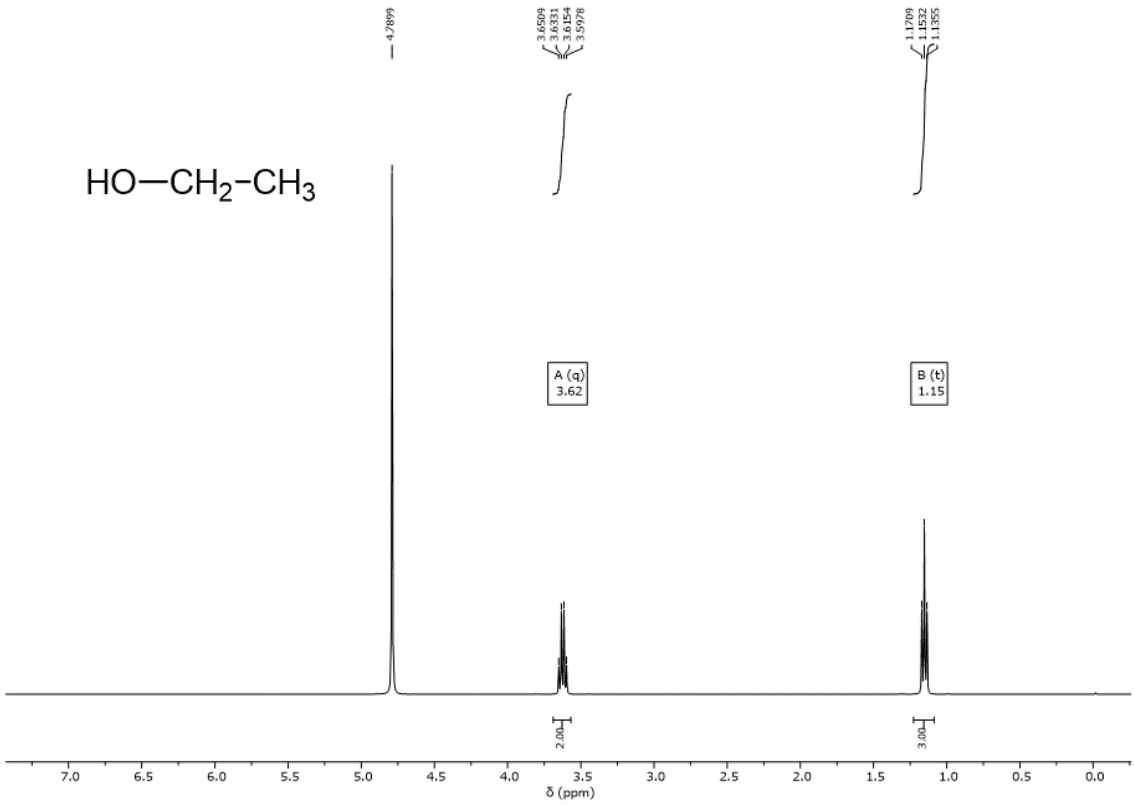


Figure 2.5. ^{1}H NMR spectrum of ethanol in D_2O obtained from the BMRB database. The CH_2 group (peak A) is a quartet at 3.62 ppm of integrated area 2. The CH_3 group is the triplet at 1.15 ppm (peak B) integral of 3.

For instance, when considering the methyl group of ethanol this appears in the ^{1}H NMR spectra (Figure 2.5)[23] as triplet, due to the influence of the methylene protons. Each of the CH_2 protons have two different orientations relative to the field (I = ½, therefore S = + ½ or -½). When both methylene protons are antiparallel to the field this reduces the local field of the CH_3 protons causing a shift to higher fields. When one proton is parallel to the field and the other is antiparallel this has no effect on the methyl chemical shift, whilst when they are both parallel there is a shift to lower fields. Consequently, the methyl signal will results as split in 3, a triplet.

The difference in chemical shift between the signal of the methyl triplet is referred to as coupling constant, J (Hz) which is independent by the external applied field $B_{0.}$ When interpreting a ^{1}H NMR the signal multiplicity is equivalent to the number of neighbouring protons, plus one. Therefore, if a signal appears as a triplet, it means that the nucleus is "next to" two equivalent protons. As a further confirmation of this the J constants of the methyl and methylene signals can be calculated, and they will result to be of the same value. Furthermore, the integrated area of the peak is directly proportional to the number of protons producing the signal. For instance, in the ethanol example the CH_3 and the CH_2 signal will have an area ratio of 3:2 (Figure 2.5). For easier interpretation if a peak is known to be attributable to a specific signal, for instance CH_2, the integrated area can be normalised to 2 simplifying the interpretation of the remaining signals.

For the identification of organic molecules, the ^{13}C NMR is also routinely recorded. Conversely to ^{1}H NMR the sensitivity of the ^{13}C NMR is much lower as the isotope ^{13}C represents only 1.1% of the total carbon, whilst the most abundant ^{12}C is not NMR-active[22]. For this reason, the ^{13}C NMR analysis requires more concentrated samples. In the ^{13}C spectrum the peaks appear as singlets, as the ^{13}C NMR is generally recorded decoupled from the ^{1}H (Figure 2.6)[23].

In addition to the 1-D experiments detailed above, 2-D experiments can also be performed to provide additional information on the molecular structure. Many different 2D experiments can be performed, with details able to be found in literature, although only techniques that are employed in this work are presented here.

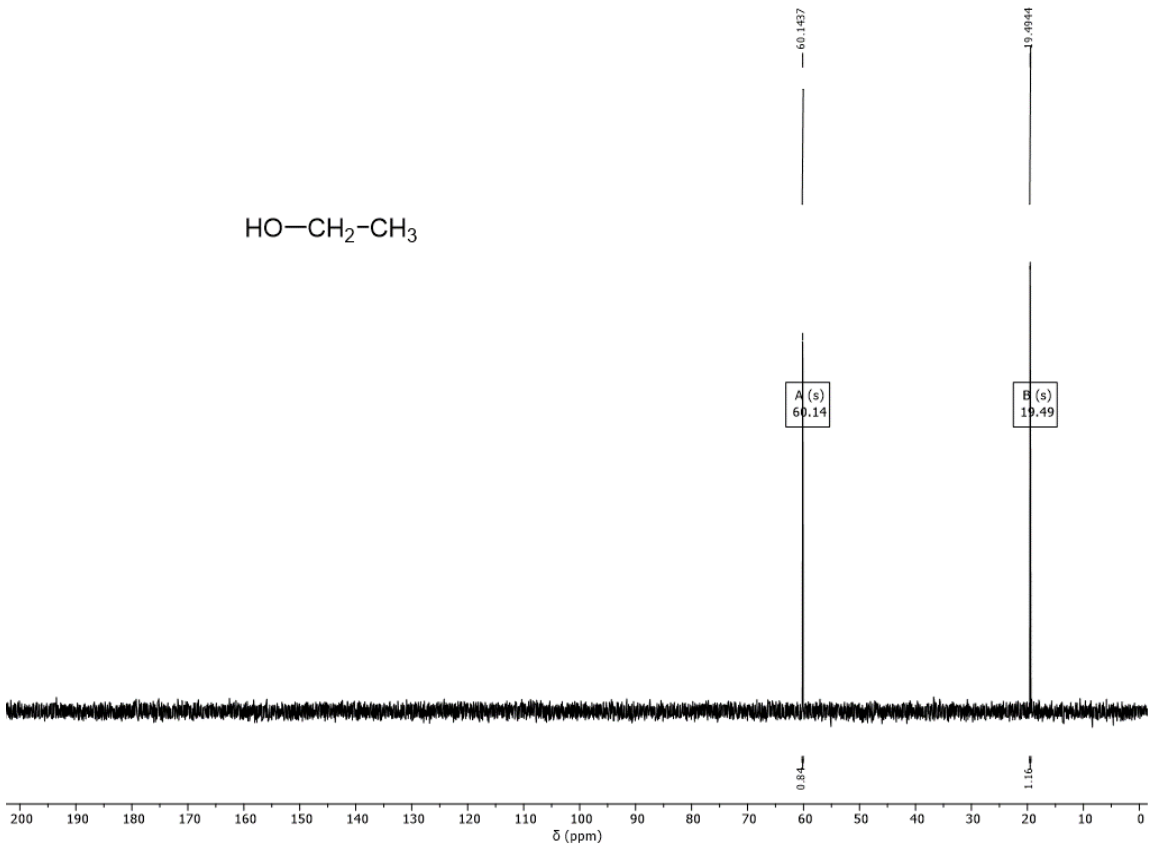


Figure 2.6. ^{13}C NMR spectrum of ethanol in D_2O obtained from the BMRB database. The C atom of the methylene group (peak A) has chemical shift of 60.14 ppm, whilst the C atom of the methyl group (peak B) has chemical shift of 19.49 ppm.

The ^{1}H-^{1}H Correlation Spectroscopy, also known as COSY, display the correlation between nuclei separated by a maximum of 3 bonds[20]. The correlation between protons is displayed as cross peaks in the 2-D spectra. The COSY NMR provide the same information regarding proton coupling as the ^{1}H NMR, however there are situations in which the 2D plot can be of additional value. The COSY NMR can be employed to better assess coupling of protons, via analysis of cross peaks, for cases where the *J* constants cannot be determined in the ^{1}H NMR due to overlapping signals. Furthermore, correlation of nuclei further apart can also be observed, for instance the ^{1}H-^{1}H Total Correlation Spectroscopy (TOCSY) displays the correlation, as cross peaks, between protons further apart as long as they are part of the same spin system

(Figure 2.6.)[24]. This is particularly useful for structural elucidation of carbohydrate chains, as each monosaccharide unit is an independent spin system[24].

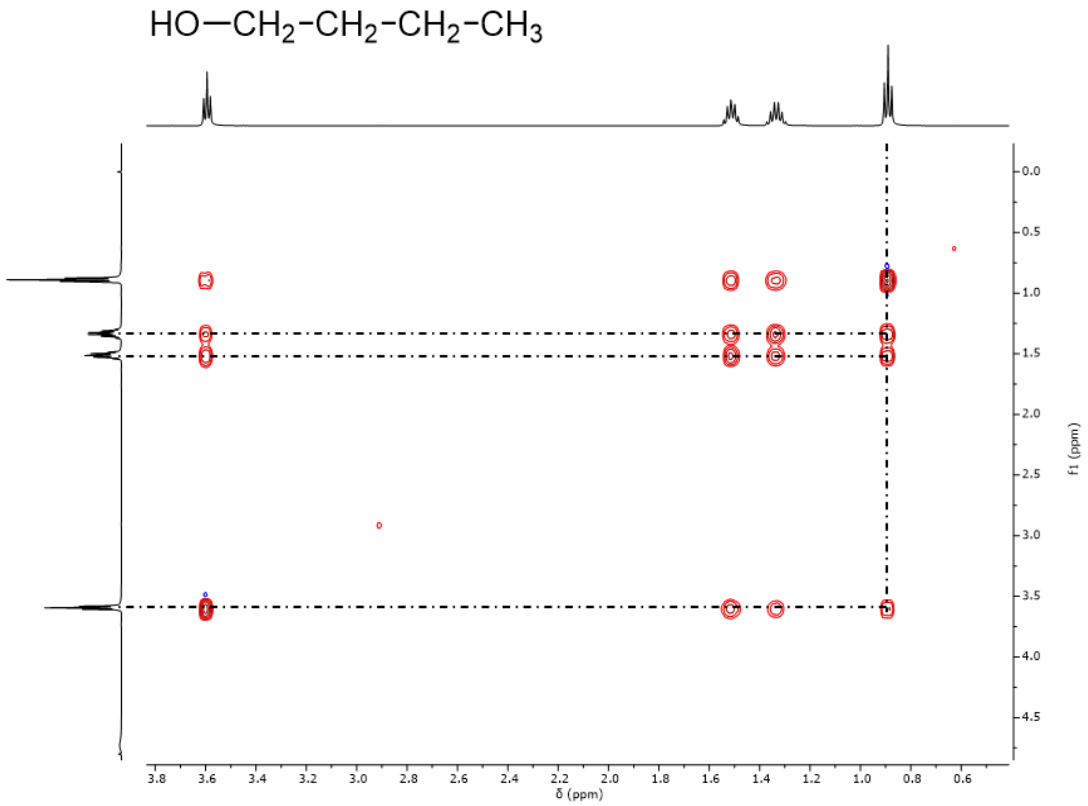


Figure 2.6. ^{1}H-^{1}H TOCSY NMR spectrum of butanol in D_2O showing the long-distance correlation between the methyl group and the methylene protons.

Furthermore, it is also possible to observe correlation of protons attached to a specific carbon atom via ^{1}H-^{13}C Heteronuclear Single Quantum Coherence (HSQC), this is especially useful in the assignment of the peaks in ^{13}C NMR spectra (Figure 2.7).

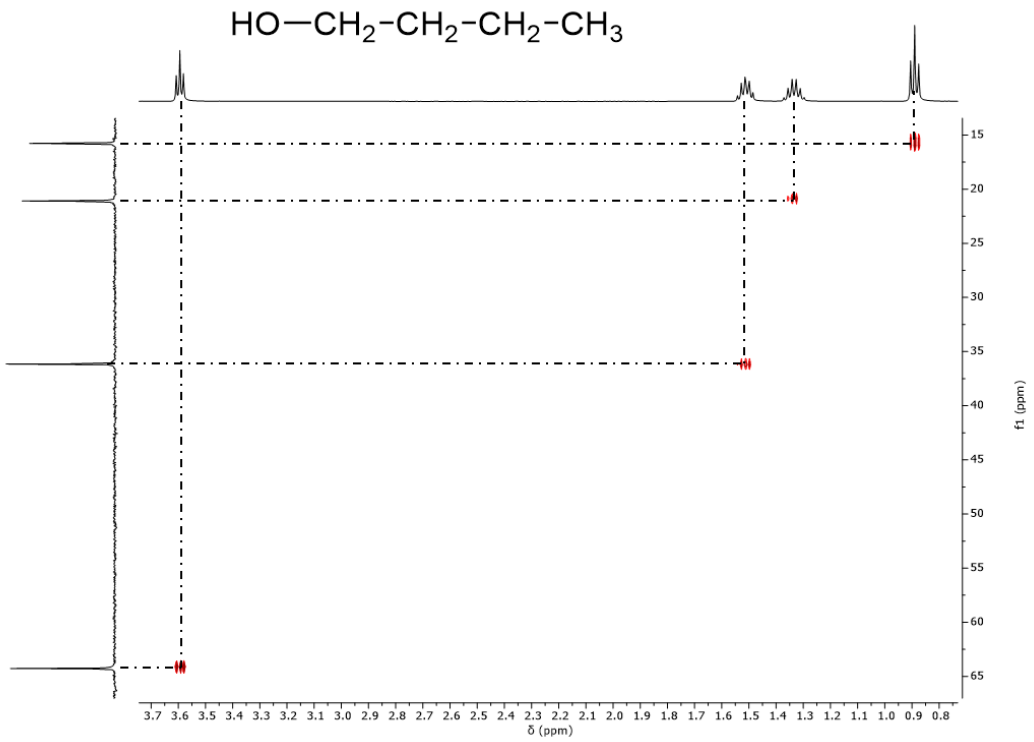


Figure 2.7. ^{1}H-^{13}C HSQC NMR spectrum of butanol in D_2O displaying the long-distance correlation between the methyl group and the methylene protons.

2.3 Liquid chromatography – mass spectrometry

Liquid chromatography – mass spectrometry (LC-MS) is an analytical technique for the analysis of liquid mixtures[25]. LC-MS finds clinical application in the analysis of biological samples and is routinely employed in organic chemistry for monitoring reaction progression and purification of complex mixtures. This technique combines the ability of liquid chromatography (LC) to separate complex mixtures and analyte identification via mass spectrometry.

Liquid chromatography (LC) and high-performance liquid chromatography (HPLC) are chromatography techniques employed for the separation of complex mixtures (Figure 2.8)[26]. Chromatography techniques are based on having two phases, a stationary and a mobile phase,

for which different analytes display different affinities. Herein, only reverse phase chromatography, which is commonly employed in organic chemistry, is described. In reverse phase chromatography, the stationary phase, which is packed inside a column, is apolar (as standard a C18 column). The mobile phase, or eluent, is passed through the column. The mobile phase is a polar system usually consisting of a mixture of water and acetonitrile or water and methanol. The composition of the eluent can be changed during the chromatographic run to aid in the separation of similar species, referred to as gradient elution. A mixture is loaded onto the column and the various components of the mixture will be separated based on their different affinities for the stationary phase versus the mobile phase. Thus, each component will be eluted out of the column at a distinct time referred to as retention time. Polar components, which do not display a strong affinity to the stationary phase, will be eluted more quickly and therefore present shorter retention times. Conversely, apolar components will present longer retention times as they present a higher affinity to the stationary phase and are retained in the column. Each component will be represented in the chromatogram as a peak with a corresponding retention time. The area beneath the peak can be calculated and utilised for quantification purposes[27]. The mobile phase is then fed to the mass spectrometer for analysis[25]. The analytes present in the mobile phase are firstly ionised. Many are the possible ionisation methods, with the electrospray ionisation (ESI) method being one of the most widespread in LC-MS and HPLC-MS systems. Following ionisation, the ions are separated by an electromagnetic field, on the basis of their mass to charge ratio (*m/z*), in the mass analyser[28-29]. Ionisation of small molecules by electrospray generally affords singly charged ions with little to no fragmentation. Therefore, the *m/z* ratio obtained is representative of the molecular weight of the analyte as an ion. Indeed, in the ionisation process a proton or a cation is added to the analyte, for ionisation in positive ion mode. Therefore the *m/z* values correspond to the mass of the analyte plus a

proton $[M + H]^+$, or plus a cation, usually sodium, $[M + Na]^+$. Furthermore, when the ionisation is performed in negative ion mode a proton is lost from the analyte and the *m/z* value corresponds to the mass of the analyte minus a proton $[M - H]^-$. The MS analysis can therefore provide a *m/z* value for each peak of the chromatogram allowing identification of the chemical compound, based on the molecular weight.

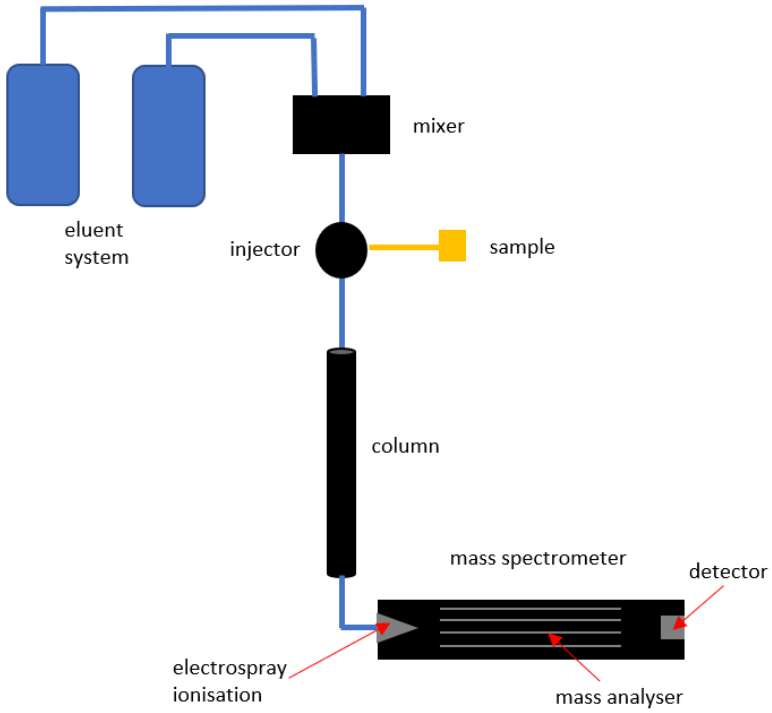


Figure 2.8. Schematic representation of a LC-MS or HPLC-MS instrument.

List of references

1. Wiseman, T.; Williston, S.; Brandts, J. F.; Lin, L. N., Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **1989,** *179* (1), 131-137.

2. Freire, E.; Mayorga, O. L.; Straume, M., Isothermal titration calorimetry. *Anal. Chem.* **1990,** *62* (18), 950A-959A.

3. Doyle, M. L., Characterization of binding interactions by isothermal titration calorimetry. *Curr. Opin. Biotechnol.* **1997,** *8* (1), 31-35.

4. Wiskur, S. L.; Lavigne, J. L.; Metzger, A.; Tobey, S. L.; Lynch, V.; Anslyn, E. V., Thermodynamic analysis of receptors based on guanidinium/boronic acid groups for the complexation of carboxylates, alpha-hydroxycarboxylates, and diols: Driving force for binding and cooperativity. *Chem. Eur. J.* **2004,** *10* (15), 3792-3804.

5. Linton, B.; Hamilton, A. D., Calorimetric investigation of guanidinium-carboxylate interactions. *Tetrahedron* **1999,** *55* (19), 6027-6038.

6. Prozeller, D.; Morsbach, S.; Landfester, K., Isothermal titration calorimetry as a complementary method for investigating nanoparticle–protein interactions. *Nanoscale* **2019,** *11* (41), 19265-19273.

7. Turnbull, W. B.; Daranas, A. H., On the value of c: Can low affinity systems be studied by isothermal titration calorimetry? *J. Am. Chem. Soc.* **2003,** *125* (48), 14859-14866.

8. MicroCal, L., *VP-ITC MicroCalorimeter User's Manual, MAU130030 REV. E*. MicroCal, LLC: Northampton, MA.

9. Bhatnagar, R. S.; Gordon, J. I., Thermodynamic studies of myristoyl-CoA: protein N-myristoyltransferase using isothermal titration calorimetry. *Methods Enzymol.* **1995,** *250*, 467-486.

10. Bundle, D. R.; Sigurskjold, B. W., Determination of accurate thermodynamics of binding by titration microcalorimetry. *Neoglycoconjugates, Pt B: Biomedical Applications* **1994,** *247*, 288-305.

11. Saboury, A. A., Application of a new method for data analysis of isothermal titration calorimetry in the interaction between human serum albumin and Ni2+. *J. Chem. Thermodyn.* **2003,** *35* (12), 1975-1981.

12. Fisher, H. F.; Singh, N., Calorimetric methods for interpreting protein—Ligand interactions. In *Methods Enzymol.*, Academic Press: 1995; Vol. 259, pp 194-221.

13. Sigurskjold, B. W., Exact Analysis of Competition Ligand Binding by Displacement Isothermal Titration Calorimetry. *Anal. Biochem.* **2000,** *277* (2), 260-266.

14. Tellinghuisen, J., Isothermal titration calorimetry at very low c. *Anal. Biochem.* **2008,** *373* (2), 395-397.

15. Holdgate, G., Isothermal Titration Calorimetry and Differential Scanning Calorimetry. In *Ligand-Macromolecular Interactions in Drug Discovery: Methods and Protocols*, Roque, A. C. A., Ed. Humana Press Inc: Totowa, 2010; Vol. 572, pp 101-133.

16. Werber, L.; Mastai, Y., Isothermal titration calorimetry for chiral chemistry. *Chirality* **2018,** *30* (5), 619-631.

17. Wadso, I., Isothermal microcalorimetry for the characterization of interactions between drugs and biological materials. *Thermochim. Acta* **1995,** *267*, 45-59.

18. MicroCal, L., *ITC Data Analysis in Origin - Tutorial Guide, MAU130010 Rev D*. MicroCal, LLC: Northampton, MA.

19. Darbeau, R. W., Nuclear Magnetic Resonance (NMR) Spectroscopy: A Review and a Look at Its Use as a Probative Tool in Deamination Chemistry. *Appl. Spectrosc. Rev.* **2006,** *41* (4), 401-425.

20. Elyashberg, M., Identification and structure elucidation by NMR spectroscopy. *Trends Anal. Chem.* **2015,** *69*, 88-97.

21. Tognarelli, J. M.; Dawood, M.; Shariff, M. I. F.; Grover, V. P. B.; Crossey, M. M. E.; Cox, I. J.; Taylor-Robinson, S. D.; McPhail, M. J. W., Magnetic Resonance Spectroscopy: Principles and Techniques: Lessons for Clinicians. *J Clin Exp Hepatol* **2015,** *5* (4), 320-328.

22. Manfred Hesse, H. M., Bernd Zeeh, *Spectroscopic Methods in Organic Chemistry*. 2008.

23. Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Wenger, R. K.; Yao, H. Y.; Markley, J. L., BioMagResBank. *Nucleic Acids Res.* **2008,** *36*, D402-D408.

24. Bross-Walch, N.; Kühn, T.; Moskau, D.; Zerbe, O., Strategies and Tools for Structure Determination of Natural Products Using Modern Methods of NMR Spectroscopy. *Chem. Biodivers.* **2005,** *2* (2), 147-177.

25. Pitt, J. J., Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *Clin Biochem Rev* **2009,** *30* (1), 19-34.

26. Malviya, R.; Bansal, V.; Pal, O.; Sharma, P., High performance liquid chromatography: A short review. *J. Glob. Pharma Technol.* **2010,** *2*, 22-26.

27. Hartman, A. M.; Gierse, R. M.; Hirsch, A. K. H., Protein-Templated Dynamic Combinatorial Chemistry: Brief Overview and Experimental Protocol. *Eur. J. Org. Chem.* **2019,** *2019* (22), 3581-3590.

28. Ho, C. S.; Lam, C. W. K.; Chan, M. H. M.; Cheung, R. C. K.; Law, L. K.; Lit, L. C. W.; Ng, K. F.; Suen, M. W. M.; Tai, H. L., Electrospray ionisation mass spectrometry: principles and clinical applications. *Clin Biochem Rev* **2003,** *24* (1), 3-12.

29. Karas, M., Time-of-flight mass spectrometer with improved resolution. *J. Mass Spectrom.* **1997,** *32* (1), 1-3.

Chapter 3 - Sialic acid sensing by positively charged benzoboroxole receptors

As discussed in the Introduction Chapter glycan recognition is a significant area of research in the attempt to detecting diseases more efficiently and more effectively. Glycans, which are polysaccharides chains attached to protein carriers, can be altered by pathological conditions in their composition and structure. One of the main modifications involves the alteration of the sialic acid content. Alteration of the expression of sialic acid can indicate underlying pathological conditions, such as cancer, neurodegenerative diseases and alcoholism[1-5]. Sialylated glycans have been linked to different tumours, whilst free sialic acid was found in high levels in certain neurodegenerative diseases, termed sialic acid storage diseases. In alcoholism both bound and unbound sialic acid levels are present at elevated levels. Thus, this monosaccharide is extremely relevant as a biomarker for diagnostic purposes and demand a highly effective and selective probe for its detection. Small synthetic molecules are deemed to be more suitable, when compared to natural lectins, to act as probes due to their higher availability[6]. Synthetic receptors should be obtained from inexpensive starting materials and straightforward synthetic routes. Moreover, the receptor should bind the sugar via multiple interactions including both covalent and non-covalent bonds. The multivalent nature of the binding interaction between receptor and sugar determines affinity and selectivity enhancement, within the so-called "cluster effect"[7]. Additionally, in order for the probe to be employed in diagnostics it should be compatible with biological fluids. Hence, the probe should be soluble in aqueous media and show activity at physiological pH.

Boron-based receptors have been extensively employed as detection tools in carbohydrate recognition[7]. Boronic acids (BAs) are small molecules which bind to saccharide diols in a covalent and reversible fashion. However, BAs find limited applicability in the glyco sensing field. Biologically relevant saccharides are primarily present in their pyranose conformation, whilst BA receptors interact predominantly with furanose rings[8]. BAs can therefore mediate the recognition of monosaccharides, which often are present in the furanose form, but not the recognition of pyranose residues in oligosaccharides or glycans. In addition, BAs generally form boronate ester complexes with the sugar diols at pH values above the pK_a of the boronic acid[9], which for phenylboronic acid (pK_a = 8.86)[10] derivatives is basic. Thus, BA receptors bind saccharides primarily at basic pH. This is due to the different stability of the boronate ester formed with the sugar at different pH values[11]. For pH values below the pK_a, the complex generated by the trigonal boron is unstable and prone to hydrolysis, whilst for pH > pK_a the boronate ester display greater stability. This, in addition to their low solubility, limits their applicability in the analysis of biological samples. The issues of BAs are overcome by a class of analogues; benzoboroxoles[12]. Benzoboroxole (BOB) is a cyclic analogue of phenylboronic acid with the key advantage of interacting with pyranose rings[13]. In addition, the receptor BOB can interact with sugars at neutral pH due to the lower pK_a value, of 7.2. Benzoboroxoles also generally present higher binding affinities towards saccharides, compared to the corresponding phenylboronic acids[13]. Despite the above-mentioned advantages not many examples of benzoboroxole receptors for sugar recognition are present in the literature[14-18]. Even fewer are the BOB-based receptors developed to target sialic acid despite its relevance as a biomarker in various diseases[19].

In the work presented herein, and published in The Journal of Organic Chemistry with the title of *Cooperative Multipoint Recognition of Sialic Acid by Benzoboroxole-Based Receptors*

Bearing Cationic Hydrogen-Bond Donors[20], benzoboroxole receptors were developed for sialic acid recognition. Initially the binding site of boron-based receptors to sialic acid was investigated, as different and opposing theories are reported in literature[21-23]. This was performed by isothermal titration calorimetry (ITC) binding studies of non-functionalised benzoboroxole **B1** and sialic acid derivatives (*e.g.* sialic acid, 2-O-methyl-α-sialic acid and sialic acid methyl ester). Isothermal titration calorimetry studies, alongside with nuclear magnetic resonance (NMR) and mass spectrometry (MS) studies, allowed unequivocal identification of the binding site. On this basis, a series of functionalised benzoboroxole receptors **B2** – **B7** were designed and synthesised. Their binding affinities to sialic acid were measured by calorimetry and a binding model, comprised of multivalent and cooperative interactions, was postulated. The selectivity of the receptors was assessed with different neutral and anionic monosaccharides. A functionalised benzoboroxole receptor was found to bind sialic acid with a 4.7-fold increase in the binding affinity and good selectivity, showing potential to be employed as detection probe.

3.1 Optimisation of isothermal titration calorimetry experimental conditions

Isothermal titration calorimetry measures the heat involved (released or absorbed) in a binding event allowing the assessment of the following thermodynamic parameters; n (number of binding sites), K_a (binding constant, M^{-1}), ΔH (change in enthalpy, kcal mol^{-1}) and ΔS (change in entropy, kcal mol^{-1} K^{-1})[24]. ITC, unlike other techniques, does not require labelling or immobilisation of the receptor[25]. Therefore, binding studies conducted in solution eliminate any possible artifacts caused by the receptor derivatisation. ITC was employed to assess the formation of the boronate ester complex in solution obtained by the reaction between a benzoboroxole receptor and sialic acid, usually this is an exothermic binding event

characterised by heat development[19, 26-27]. The ITC experiment consist of the titration of a ligand solution, the monosaccharide, into a solution of receptor, the benzoboroxole. The ligand is added in small aliquots (µl) to the cell containing the receptor. When the ligand binds the receptor, forming the boronate ester complex, heat is released in the system and a negative heat change is recorded. In the first few injections a large amount of heat is released as a larger fraction of receptor is available. The heat released reduces progressively, as a smaller fraction of receptor is free, until saturation is reached.

Prior to performing the binding studies, the experimental conditions had to be optimised. The experimental conditions (*e.g.* concentrations, heat change and shape of the isotherm) are linked to a parameter called *c*-value[28]. The *c*-values is given by Equation 3.1 and should ideally be ≥ 10. For *c*-values ≥ 10 the titration data can be fitted with a sigmoidal shape which allow the all the thermodynamic parameters (n, K_a, ΔH and ΔS) to be calculated with accuracy.

$$c = n[R]_{cell}K_a \qquad \text{Equation 3.1}$$

The *c*-value can be arbitrarily selected in order to guarantee a sigmoidal binding curve. Once the *c*-value is selected the required receptor concentration can be calculated from Equation 3.2, provided K_a and n are known values.

$$[R]_{cell} = \frac{c}{nK_a} \qquad \text{Equation 3.2}$$

The binding constant of benzoboroxole (BOB, **B1**) to sialic acid was not known prior to this work. Thus, for a preliminary estimation of the $[R]_{cell}$ the binding constant of PBA to sialic acid at pH 7.4 was considered (K_a = 21 M^{-1})[29]. Furthermore, the number of binding sites, n, was known to be 1, as one molecule of SA is bound exclusively by one boron-based receptor. Weak affinity systems ($K_a < 10^4$) such as the boronate ester formation, tend to present low *c*-values[30].

Therefore, an initial *c*-value of 10, the lowest value to provide a sigmoidal curve, was selected. For c = 10 the required $[R]_{cell}$ was approximatively 470 mM. In addition, for c = 10 the molar ratio at the end of the titration should at least be 1:2. The molar ratio is the receptor concentration to the ligand concentration present in the sample cell at the end of the titration[31], as per Equation 3.3.

$$molar\ ratio = \frac{[R]}{[L]} \qquad \text{Equation 3.3}$$

The required ligand concentration for the aforementioned conditions was 5.4 M. Understandably, these conditions cannot be met due to the low water solubility of the receptor and the large amount of material required. Thus, a lower *c*-value was selected. For a *c*-value of 1 the concentration of receptor required, despite being 10 times lower ($[R]_{cell}$ = 47 mM), was still too high given the low solubility of **B1** in aqueous media. Consequently, a receptor concentration of 2 mM was arbitrarily selected. This concentration also matches the concentration employed in published ITC binding studies for the interaction of benzoboroxole to fructose[26]. For a 2 mM cell solution the resulting *c*-value becomes extremely low ($c < 0.1$). For systems with very low *c*-values two provisions must be adopted in order to obtain reliable results. The first one requires fixing the number of binding sites during the fitting[31]. Thus, the stoichiometry (n) was fixed to 1 for all experiments presented in this Chapter. The second provision is with regards to the molar ratio[30]. For weak binding systems with very low *c*-values, including these boron-based receptors, saturation is hardly ever achieved. Nevertheless, the system must be pushed towards the saturation state, thus a high molar ratio is required[30]. Additionally, the heat developed during an injection must be large enough to be detected by the instrument and present a high signal to noise ratio. The smallest quantity of heat that can be detected accurately by the VP-ITC instrument is between 3 and 5 μcal[32].

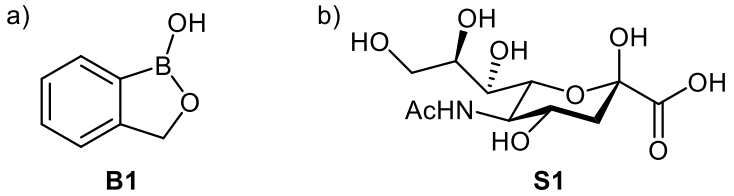


Figure 3.1. Structure of a) benzoboroxole, BOB (**B1**); b) sialic acid, SA (**S1**).

A series of optimisation experiments for the binding of non-functionalised **B1** to sialic acid (**S1**) (Figure 3.1) were conducted in 0.1 M PBS buffer at pH 7.4. Initially, a molar ratio of 1 : 7.5 was selected resulting in a 2 mM solution of **B1** and an 80 mM solution of sialic acid. The volume of injection was 5 µL for a total of 40 injections. In these conditions the heat released upon binding was below 2.5 µcal s^{-1}. Thus, the concentration of titrant was increased to 100 mM to give a higher molar ratio of approximatively 1 : 10. The titration was conducted with the same volume and number of injections of the previous experiment. With this higher molar ratio the heat released was approximatively 4 µcal s^{-1} which allowed accurate measurements with the VP-ITC instrument. Although a 100 mM solution of monosaccharide might be adequate when studying the inexpensive sialic acid, it is certainly not suitable for more expensive sugars (*e.g.* 2-O-methyl-α-sialic acid). Therefore, lower concentrations of ligand are required whilst maintaining a heat change above 3 µcal s^{-1}, this was achieved by increasing the injection volume from 5 µl to 7 µl. Thus, a 2 mM solution of receptor **B1** was titrated with an 80 mM solution of **S1** with 35 injections of a volume of 7 µl resulting in a heat change of approximatively 5 µcal s^{-1}. These conditions were deemed to be optimal for the system under investigation and were employed for all subsequent ITC experiments presented in this Chapter.

Table 3.1. Isothermal titration calorimetry experimental parameters.

Ligand concentration	80 mM
Receptor concentration	2 mM
Number of injections	35
Cell temperature	25°C
Reference Power	10 – 25 μcal s^{-1}
Initial delay	60 s
Stirring	307 RPM
Volume of each injection	7.0 μl
Duration of each injection	4.0 s
Spacing	280 – 480 s
Filter	2.0 s

In addition to the aforementioned conditions, the ITC experiments require other parameters to be selected (Table 3.1). Given an injection volume, the instrument automatically selects the appropriate duration of the single injection, which is 4.0 s for a 7 μl injection. The stirring speed was selected to be 307 RPM and the filter period to 2 s, as recommended in the VP-ITC manual[32]. The filter period is used to average the differential power (DP) signal[24]. The DP is the experimental data recorded during the titration and it represents the power provided to the sample cell[24]. The DP decreases when the heat is released during an exothermic event. The decrease in DP is represented on the titration graph as a negative peak (Figure 3.2), whilst in the absence of heat change the DP value rests on the baseline. The baseline value at the start of experiment is determined by the selected reference power. If a large exothermic reaction is expected, a large reference power (≈ 30μcal s^{-1}) must be selected, whilst for endothermic reactions the reference power should be small (≈ 2 μcal s^{-1})[32]. All experiments presented herein have a reference power between 10 and 25 μcal s^{-1}. A reference power of 10 μcal s^{-1} and 25 μcal s^{-1} were selected for small (≈ 5 μcal s^{-1}) and large (≈ 25 μcal s^{-1}) heat releases, respectively.

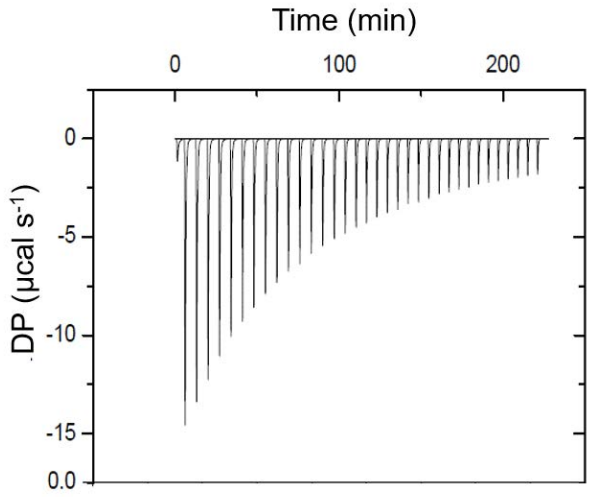


Figure 3.2. Titration graph with negative peaks corresponding to a decrease in the DP value determined by an exothermic binding event.

The magnitude of the heat change also influences the spacing, or time interval, required between injections. The spacing must allow the DP to return to the baseline value prior the following injection. For titrations with small and large heat change the spacing was set at 280 s and 480 s, respectively. Furthermore, the temperature, at which the titrations were performed, was set at 25 °C for all experiments presented here. The titration experiment does not begin until sample cell, reference cell and adiabatic jacket are all at 25 °C. In addition, for the experiment to start the baseline must be flat, with the DP value being stable. The system equilibrates the temperature and the DP and the experiment starts only when the differential temperature between the cells and the adiabatic jacket is 0 ± 0.05 °C and the DP is stable at a value equal to the selected reference power (± 1 µcal sec^{-1}). Once the experiment starts there is an initial delay, set at 60 s, in which the instrument collects enough data points to establish a baseline. Only after this the titration of the ligand into the sample cell begins.

In order to ensure that the data is reliable each experiment is repeated three times, with the binding constant calculated as the average value obtained in the three experiments, whilst the

error is expressed as twice the standard deviation. This method applies to all the experiments presented in here. The binding of **B1** to sialic acid in 0.1 M PBS buffer at pH 7.4 was measured three times (Table 3.2 Entry 1) with the average binding constant found to be 27.5 ± 0.2 M^{-1}.

Table 3.2. Binding constants (K_a, M^{-1}) measure by ITC by titrating a 80 mM solution of sialic acid **S1** into a 2 mM solution of benzoboroxole **B1** under the following conditions; Entry 1: in 0.1 M PBS buffer without subtraction of the dilution of heat; Entry 2: in 0.1 M PBS buffer with subtraction of the heat of dilution; Entry 3: in 0.1 M PB buffer with subtraction of the heat of dilution. Each experiment consisted of three titrations with final K_a value being the average of the three titrations.

	1st titration	2nd titration	3rd titration	average
Entry 1	27.6 ± 0.2	27.6 ± 0.2	27.4 ± 0.2	27.5 ± 0.2
Entry 2	29.4 ± 0.4	30.6 ± 0.3	30.3 ± 0.3	30.1 ± 1.2
Entry 3	39.4 ± 0.4	39.1 ± 0.2	39.7 ± 0.2	39.4 ± 0.6

Further to the collection of binding data, the heat of dilution was also recorded. When titrating a highly concentrate solution into the sample cell the dilution of the ligand into a bigger volume generally causes a heat change. This heat, known as heat of dilution, must be subtracted from the titration data of the binding experiment[32]. The heat of dilution was measured by titrating the ligand solution, 80 mM solution of ligand, into the buffer employing the same conditions of the binding experiment. The heat of dilution was then subtracted (Table 3.2 Entry 2) from each titration dataset and the thermodynamic parameters extracted from the fitting. The parameters of the three titrations are then averaged out, resulting in a modest increase of binding constant when titrating **S1** into **B1** (K_a = 30.1 ± 1.2 M^{-1}) as shown in Entry 2 of Table 3.2.

The effect of the ionic strength on the boronate ester formation between sialic acid and **B1** was also assessed. The binding experiments were initially conducted in commercially available 0.1 M PBS buffer, which contains 1.37 M of NaCl. When 0.1 M phosphate buffer (PB) was used instead a 1.3-times increase in the binding affinity was measured as in Table 3.2 Entry 3 with a K_a value of 39.4 ± 0.6 M^{-1}. In addition, different solutions in phosphate buffer containing

different concentrations of NaCl (0.34 M, 0.69 M and 1.03 M), corresponding to 25%, 50% and 75% of the NaCl in PBS, were investigated. As shown in Figure 3.3, the binding affinity decreased proportionally with the increase in the NaCl concentration. The data collected here appears to suggest that the ionic strength of the solution has a negative effect on the boronate esterification. There does not appear to be any literature on the reasons for this, warranting further investigation into the cause. This, however, was not the focus of this particular project, therefore was not pursued further. Consequently, all experiments at pH 7.4 and 6.5 were conducted in 0.1 M PB buffer. Whilst, for experiments conducted at other pH values different buffers were utilised to allow pH control (see Section 3.2).

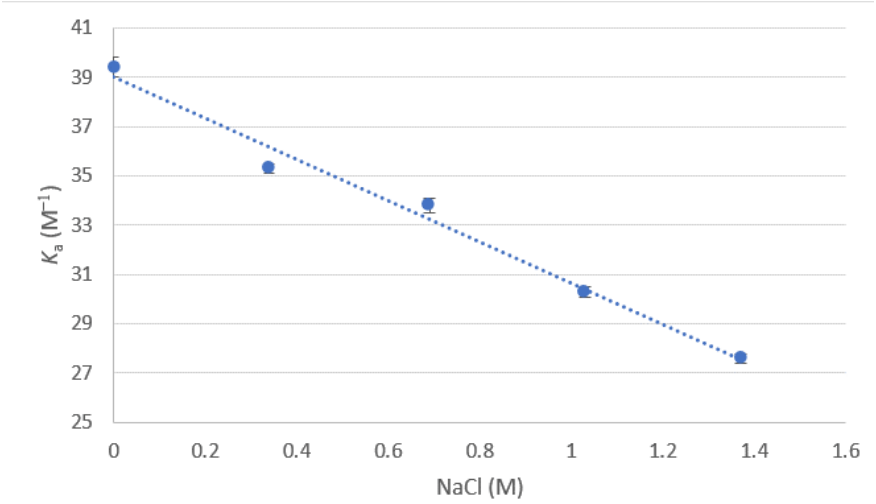


Figure 3.3. Graph representing the influence of the ionic strength (NaCl concentration, M) on the binding affinity (K_a, M^{-1}) for **B1** to **S1**.

In summary, all ITC experiments presented in this Chapter are conducted with a 2 mM solution of benzoboroxole-based receptor, which is placed in the sample cell. An 80 mM solution of sialic acid, or SA derivative, is placed in the syringe and titrated into the cell in 35 injections of volume 7 µl at 25 °C. Each ITC experiment consists of three reproducible titrations from which the heat of dilution is measured and subtracted. When fitting the data, the binding stoichiometry

n is fixed to 1. The thermodynamic parameters K_a, ΔH and ΔS calculated for each experiment are reported as the average of the three titrations with the error being twice the standard deviation.

3.2 Binding site

3.2.1 The pH dependency of the binding event

Boronic acids and benzoboroxoles are reported to bind monosaccharides at basic pH through the negative tetrahedral boron[11]. This applies to all monosaccharides, excluding sialic acid, with the binding of boron-based receptors to sialic acid occurring preferentially at acidic pH[21]. To confirm the optimal pH for the boronate ester formation between **B1** and **S1**, 5 different pH values were examined; 5.5 (0.1 M acetate buffer), 6.5 (0.1 M phosphate buffer), 7.4 (0.1 M phosphate buffer), 8.5 (0.1 M ammonium acetate) and 10.0 (0.1 M ammonium acetate). As seen in Table 3.3 the highest K_a was reached at pH 5.5 and 6.5, which presented, within error, the same binding constant of 51.2 M^{-1}. Conversely, the binding affinity was progressively reduced at the increasing of the pH, with the lowest binding constant at pH 10 (K_a =12.5 ± 2.5 M^{-1}).

Table 3.3. Binding constants (K_a, M^{-1}) measure by ITC by titrating an 80 mM solution of **S1** into a 2 mM solution of **B1** at 5 different pH values; 5.5 in 0.1 M acetate buffer, 6.5 in 0.1 M PB buffer, 7.4 in 0.1 M PB buffer, 8.5 in 0.1 M ammonium acetate and 10.0 in 0.1 M ammonium acetate. The experiments consist in three titrations and were conducted at 25 °C. The heat of dilution was measured and subtracted.

pH	K_a (M^{-1})
5.5	51.2 ± 1.2
6.5	50.7 ± 1.3
7.4	39.3 ± 0.6
8.5	23.5 ± 2.1
10.0	12.5 ± 2.5

It was confirmed that the boronate ester formation with sialic acid occurs more favourably at acidic pH. With a pK_a of 7.2 [13] the boron centre of benzoboroxole at acidic pH is neutral and trigonal (Figure 3.4). Thus, the interaction between benzoboroxole and sialic acid occurs via the neutral and trigonal boron centre. Conversely, the binding to sialic acid via the tetrahedral boron, which is the predominant species at pH ≥ 7.4, is less favourable.

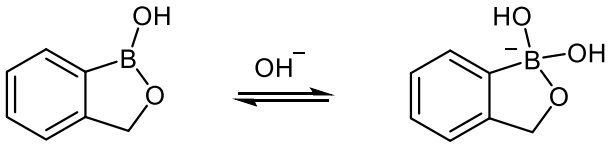


Figure 3.4. Equilibrium between the trigonal and tetrahedral form of the boron centre of benzoboroxole.

The atypical binding for sialic acid was described by Otsuka *et al.*[21]. They depicted the binding of a phenylboronic acid derivative to sialic acid occurring via the *cis*-diols of the glycerol chain. At acidic pH the interaction with the trigonal boron would normally generate an unstable complex prone to hydrolysis. However, they proposed the presence of an intramolecular bond, between the sialic acid acetamide group and the boron centre, responsible for the stabilisation of the otherwise unstable complex. This hypothesis was later discredited [22-23], nevertheless this model is incorrectly referred to in many publications, including recent ones [19, 33-37]. Indeed, most of boron-based receptors are believed to bind sialic acid through the diols of the glycerol chain. In particular, the binding is reported to occur either through the 1,2-*cis* diol in position 8 and 9 or the 1,3-*cis* diol of position 7 and 9. On the other hand, the binding via position 8 and 7 is not deemed possible due to the *erythro* configuration of this diol[33]. The binding to the hydroxyl groups in position 8 and 9 would be preferred as 5-membered boronate ester rings are more stable than 6-membered ones[8]. This model, despite being accepted by many, is in contrast with studies presenting low to no affinity for diols in linear chains. For instance, a Wulff-type boronic acid receptor showed very low, non-detectable, affinity towards 1,2-alkanediols[27],

whilst showing high affinity to catechol and α-hydroxyacid groups. The lack of binding to linear chains was also confirmed by NMR studies[22]. In addition to the glycerol chain binding site model, two other models are described in the literature. In the model proposed by Nishitani *et al.* the binding is understood to occur exclusively with the α-hydroxyacid moiety of sialic acid [22]. Conversely, Djanashvili *et al.* depict the binding site as being pH dependent [23]. In this model the glycerol chain is the main binding site at pH > 8, whilst at lower pH, between pH 2 and 8, the binding is via the α-hydroxyacid group. Hence, three different theories are debated in the literature with two possible binding sites being described; the glycerol chain and the α-hydroxyacid.

3.2.2 The binding of non-functionalised benzoboroxole to sialic acid derivatives

The binding site was clarified by assessing the binding affinity of **B1** to different sialic acid derivatives by calorimetric and analytical techniques. Three different sialic acid derivatives were employed; sialic acid (**S1**), 2-O-methyl-α-sialic acid (**S2**) and sialic acid methyl ester (**S3**) (Figure 3.5). Sialic acid (**S1**) when free and unbound in solution is present predominately in the β anomer configuration (> 90%) with both proposed binding sites, the glycerol chain and the α-hydroxyacid, available. Conversely, when it is linked to another sugar residue, as in glycans, the hydroxyl group in position 2 is engaged in a glycosidic linkage and the α-hydroxyacid is unavailable.

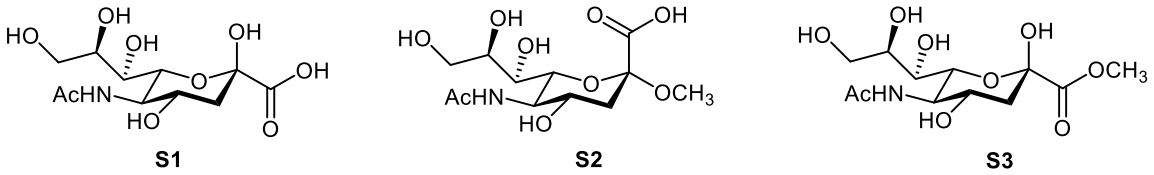


Figure 3.5. Structures of sialic acid (**S1**), 2-O-methyl-α-sialic acid (**S2**) and sialic acid methyl ester (**S3**).

To mimic the structure of sialic acid in chains the methyl glycoside of sialic acid, the 2-O-methyl-α-sialic acid (**S2**), was selected. This derivative presents the sugar in the α configuration and the α-hydroxyacid group unavailable due to the methylation of position 2. The α-hydroxyacid is unavailable also in the esterified derivative of sialic acid, the sialic acid methyl ester (**S3**). Therefore, in both derivatives of sialic acid, **S2** and **S3**, the only possible site of interaction for boron-based receptors is the glycerol chain. The 2-O-methyl-α-sialic acid was purchased, whilst the sialic acid methyl ester was synthesised following a published procedure in anhydrous methanol with 1.2 equivalent of TFA[38], with a yield of 97% (Figure 3.6).

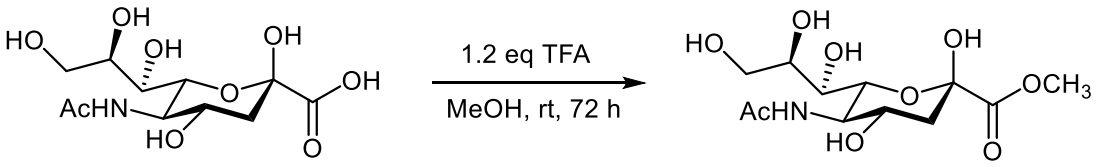


Figure 3.6. Esterification of sialic acid **S1** with 1.2 equivalents of trifluoroacetic acid in methanol at room temperature for 72 hours to give **S3**.

For the methyl ester sialic acid, it was essential to verify the stability of the ester group in aqueous conditions, prior ITC studies, as the ester group could face hydrolysis. If the hydrolysis were to occur during the ITC experiment the released free **S1**, presenting both binding sites available, would affect the study conclusions. The stability of the ester **S3** in aqueous conditions was assessed by NMR at two different pD 5.1 (~ pH 5.5) and 7.0 (~ pH 7.4)[39]. In Figure 3.7 the region between 2.50 ppm and 1.60 ppm was analysed showing the protons of position 3 giving two distinct signals for the equatorial (H_{3eq}) and axial (H_{3ax}) protons. The ^{1}H NMR spectrum of sialic acid (Figure 3.7a) shows the equatorial (H_{3eq}) and axial proton (H_{3ax}) at chemical shift of 2.22 ppm and 1.83 ppm, respectively. When the acid is esterified in **S3** the H_{3eq} and H_{3ax} signals are shifted to 2.30 ppm and 1.90 ppm, respectively (Figure 3.7b). If the methyl ester is

hydrolysed to the corresponding acid new signals will appear at higher fields, as **S1** is being formed. The stability of **S3** was firstly assessed at pD 5.1 (approx. pH 5.5.) in 0.1 M CD_3COOD buffer in D_2O. Multiple NMR spectra were recorded over time for a total of 36 hours with no hydrolysis being detected, as no shifting of H_{3eq} and H_{3ax} signals was observed (Figure 3.7c). Therefore, **S3** is stable in aqueous conditions at pH 5.5.

Consequently, **S3** was studied in 0.1 M PB buffer in D_2O at pD 7.0. The ^1H NMR spectra (Figure 3.7d) recorded 6 hours after sample preparation showed the presence of two peaks at 2.30 ppm and 1.90 ppm corresponding to **S3** and two other signals at 2.18 ppm and 1.81 ppm corresponding to the hydrolysed **S1**. Thus, after 6 hours **S3** was partially hydrolysed. The full hydrolysis of the methyl ester was achieved in 36 hours, as shown by the loss of the signals at 2.30 and 1.90 ppm (Figure 3.7e). A single ITC titration, from sample preparation to the last injection, can require up to 5 hours, therefore sialic acid methyl ester cannot be employed for ITC studies at neutral or basic pH, due to the observed hydrolysis.

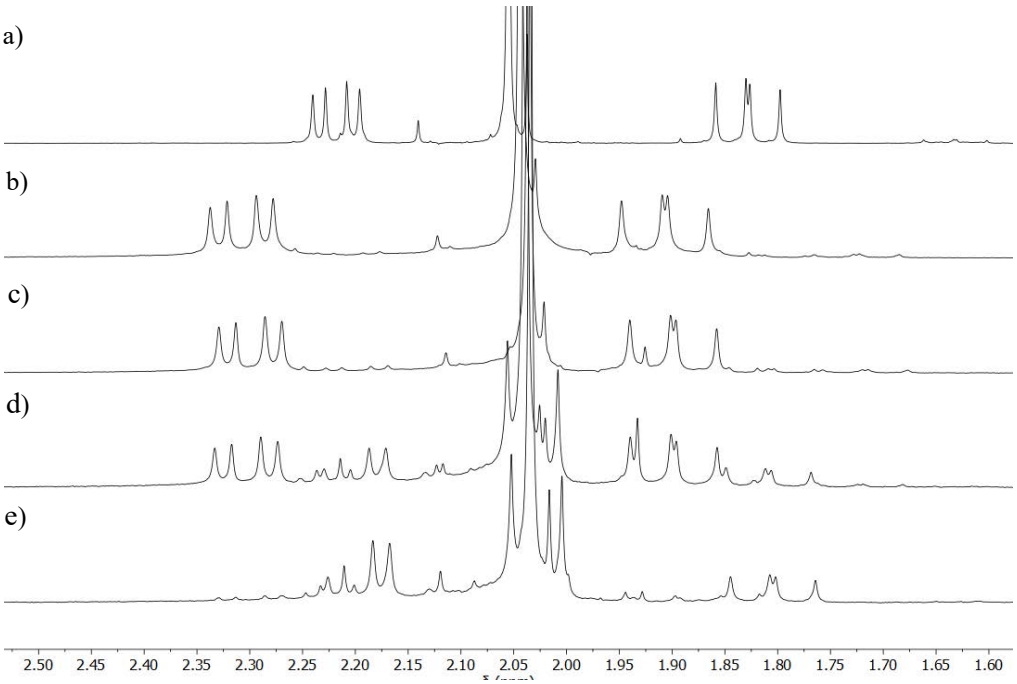


Figure 3.7. ^{1}H NMR spectra of the region between 2.50 ppm and 1.60 ppm of: a) sialic acid (**S1**) in 0.1 M CD_3COOD buffer in D_2O, pD = 5.1; b) sialic acid methyl ester (**S3**) in 0.1 M CD_3COOD buffer in D_2O, pD = 5.1, after 10 mins from sample preparation; c) sialic acid methyl ester (**S3**) in 0.1 M CD_3COOD buffer in D_2O, pD = 5.1, after 36 hours; d) sialic acid methyl ester (**S3**) in 0.1 M PB buffer in D_2O, pD = 7.0, after 6 hours; e) sialic acid methyl ester (**S3**) in 0.1 M PB buffer in D_2O, pD = 7.0, after 36 hours.

3.2.3 ITC studies of the binding site

The binding affinity of **B1** to different sialic acid derivatives was assessed by ITC across a wide pH range (5.5 – 10). In particular, the binding affinity to **S1** and **S2** was measured at pH 5.5 (0.1 M acetate buffer), 7.4 (0.1 M phosphate buffer) and 10.0 (0.1 M ammonium acetate buffer). On the other hand, the affinity to sialic acid methyl ester was only measured at pH 5.5.

Table 3.4. Binding constants (K_a, M^{-1}) measure by ITC by titrating a 80 mM solution of **S1**, **S2** or **S3** into a 2 mM solution of **B1** at three different pH values; 5.5 in 0.1 M acetate buffer, 7.4 in 0.1 M phosphate buffer and 10.0 in 0.1 M ammonium acetate. The experiments consist in three titrations and were conducted at 25 °C. The heat of dilution was measured and subtracted.

	K_a (M^{-1})		
pH	**S1**	**S2**	**S3**
5.5	51.2 ± 1.2	1.6 ± 2.4	< 1.0
7.4	39.3 ± 0.6	< 1.0	-
10.0	12.5 ± 2.5	< 1.0	-

If the glycerol chain were the main binding site, as reported in many publications [19, 21, 33-37], all three sugar derivatives should interact with **B1** providing similar binding constants, as the glycerol chain is unaltered by the modification at the C1 or C2. On the other hand, if the binding site were pH dependent no binding would occur with **S2** and **S3**, at pH ≤ 7.4, due the α-hydroxyacid group being not accessible. At basic pH, the binding site would switch to be the glycerol chain, hence the boronate ester formation with **S2** would be observed. Conversely, if the α-hydroxyacid were the only site of interaction binding would only occur with **S1**. No significant binding was detected for **S2** and **S3** at any pH value (Table 3.4). The binding was observed solely for the **S1** which is the only ligand presenting the α-hydroxyacid moiety available. Therefore, the calorimetric studies indicate the α-hydroxyacid as the only site of interaction for boron-based receptors. Once this was established, it was possible to interpret the anomalous binding profile of this class of receptors to sialic acid. The lower affinity observed at pH ≥ 7.4 can be ascribed to the charged state of the receptor and ligand. At pH above the **B1** pK_a (pK_a = 7.2)[12] the boron centre is tetrahedral and negatively charged, whilst the α-hydroxyacid (pK_a = 2.6)[21] is negatively charged across the entire pH range. Therefore, at neutral and basic pH the repulsion between the two species hinders the binding[40]. On the other hand, at acidic pH the boron is trigonal and neutral, and no repulsion occurs with the binding

site, hence the higher affinity. This binding profile is consistent with the binding of phenylboronic acid receptors to other α-hydroxyacid-containing molecules[41].

3.2.4 NMR and MS studies of the binding site

As further confirmation to the ITC studies, NMR and MS analysis were performed with the benzoboroxole receptor and the **S1** and **S2** ligands. The NMR studies required higher concentrations of benzoboroxole compared to the ITC experiments. Due to the low solubility in aqueous media an organic co-solvent was needed to fully dissolve **B1**. The receptor **B1** has a reasonable solubility in methanol, therefore methanol-d_4 was considered as a co-solvent. It was important to assess the effect of the methanol on the boronate equilibrium, thus it was necessary to evaluate the binding of a benzoboroxole receptor at different % of methanol-d_4 (0 – 30%). The receptor **B1** cannot be employed for this particular study, as this is not soluble in samples containing low percentages of methanol (0 – 20%). Therefore, a benzoboroxole receptor with higher solubility in water was needed in order to evaluate if methanol negatively alters the binding compared to a sample in 100% D_2O. For this reason, the receptor **B2**, which is highly soluble in water, was utilised (for its synthesis see Section 3.3.2). Different samples of a 1:1 mixture of **B2** and **S1** containing different percentages of methanol-d_4 were analysed by NMR to assess the effect of the co-solvent in the boronate ester formation.

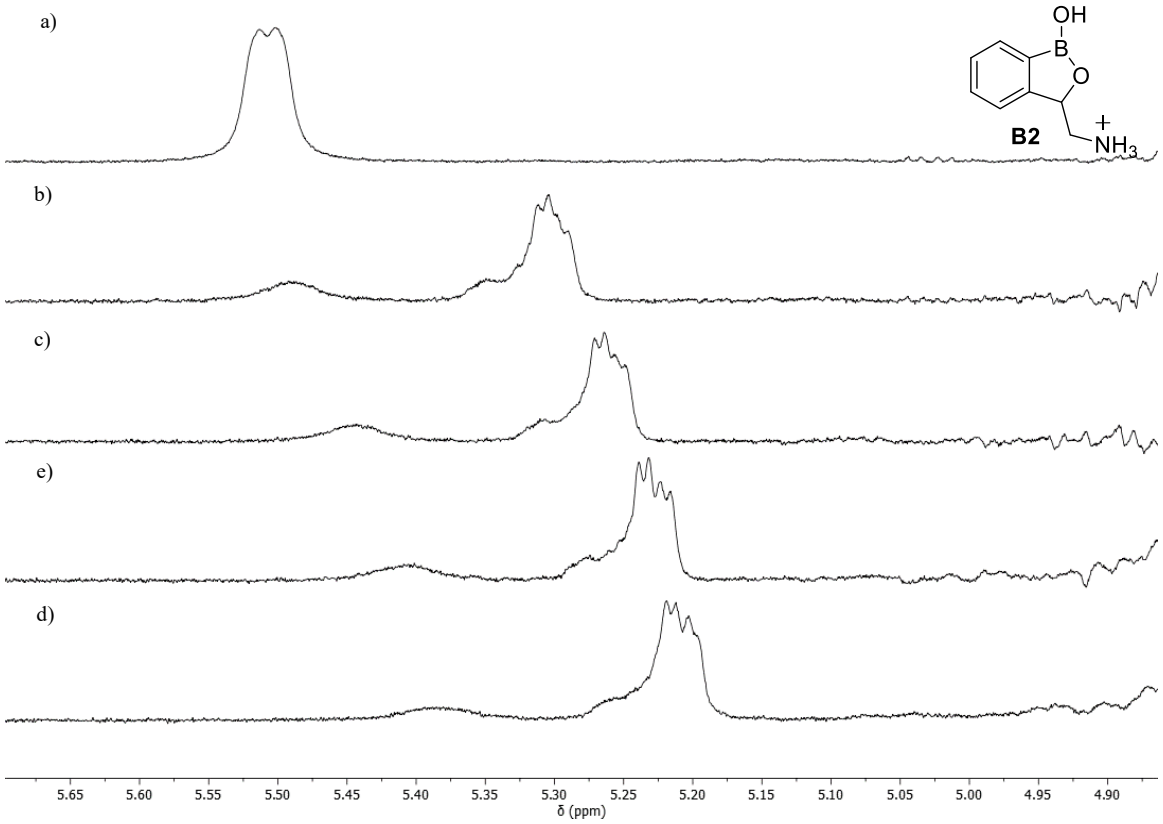


Figure 3.8. ^{1}H NMR spectra of; a) **B2** in D_2O; b) 1:1 mixture of **B2** and **S1** in D_2O with 0% methanol-d_4; c) 1:1 mixture of **B2** and **S1** in D_2O with 10% methanol-d_4; d) 1:1 mixture of **B2** and **S1** in D_2O with 20% methanol-d_4; e) 1:1 mixture of **B2** and **S1** in D_2O with 30% methanol-d_4.

In the NMR spectra (Figure 3.8a) the unbound receptor **B2** can be identified by the oxaborole ring CH_2 signal being a double doublet at 5.45 ppm. When the receptor is bound to the sugar in the **B2−S1** complex, the oxaborole CH_2 signal is shifted at 5.25 ppm. With the boronate ester formation being an equilibrium both species, **B2** and **B2−S1**, will be present in the sample and the spectra will present both CH_2 signals at 5.45 ppm and 5.25 ppm (Figure 3.8b,c,d,e). The integral area of the free **B2** peak at 5.45 ppm was set to 1, as an internal reference in all spectra. The peak of complex was then integrated, and the percentage of complex was measure with Equation 3.4 giving results in Table 3.5.

$$\%\ complex = \frac{area_{complex}}{area_{unbound}} \qquad \text{Equation 3.4}$$

Table 3.5. Peaks integrated area of unbound **B2** and complex **B2-S1** and corresponding % of complex at different concentrations of methanol-d_4.

% CD_3OD	$Area_{unbound}$	$Area_{complex}$	% complex
0	1.00	3.35	77
10	1.00	4.36	81
20	1.00	4.40	81
30	1.00	4.68	82

When no methanol-d_4 was present in the sample the percentage of complex was 77%. The addition of 10% methanol-d_4 marginally promoted the boronate ester formation with the complex percentage increasing from 77% to 81%. The addition of methanol in higher concentration, 20% and 30%, did not significantly change the degree of complexation. These results are consistent with the routine use of methanol, and other organic solvents, to enhance boron-based receptor solubility[33, 42]. Consequently, it was concluded that the presence of methanol-d_4 in the buffer does not negatively affect the boronate esterification. Thus, methanol-d_4 can be employed as co-solvent for NMR studies of the interaction between benzoboroxole and sialic acids.

The NMR studies of different 1:1 mixtures of **B1** with **S1** and **S2** were conducted of at pD = 5.1 (~ pH 5.5)[39] in a 0.1 M deuterated acetate buffer with 30% methanol-d_4. The ^{1}H NMR spectra for the two separate species, receptor and ligand, were recorded to allow comparison with the mixtures[12]. Figure 3.9 shows the aromatic region of the spectrum between 7.75 ppm and 7.15 ppm. This region of the spectrum presents signals solely from the aromatic protons of **B1**, as the sugar does not contain any aromatic protons. The sugar presents signals at higher fields, below 4.00 ppm. However, this region is not suited for assessing the complexation as is too crowded of overlapping signals rendering the identification and integrations of peaks difficult. Figure 3.9c show the spectra of **B1**, with three peaks corresponding to the four aromatic protons; a double triplet (dt) at 7.68 ppm, a triplet of doublet (td) at 7.48 ppm and a multiplet at 7.36 ppm.

The ^{1}H NMR spectra of sialic acid is not shown in Figure 3.9 as sialic acid has no aromatic protons. Figure 3.9b shows the aromatic region of a 1:1 mixture of **B1** and **S1**. Receptor **B1** reacts with **S1** to form the boronate ester. The complexation is an equilibrium and three species are present in solution: free **B1**, free **S1** and the boronate ester complex.

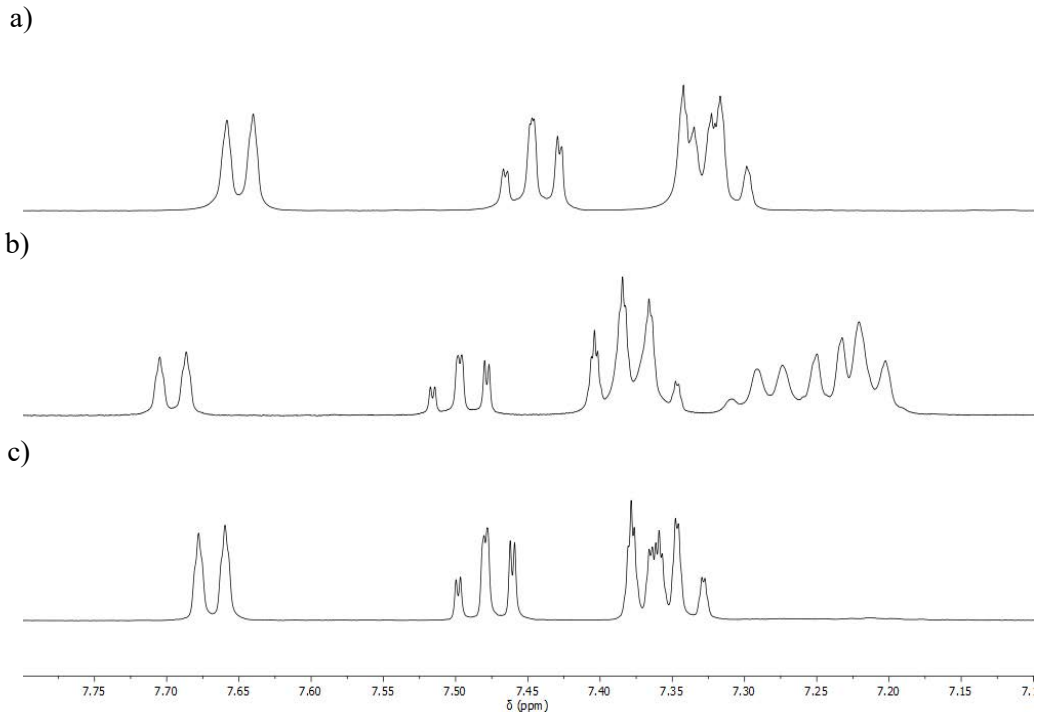


Figure 3.9. ^{1}H NMR spectra recorded in 0.1 M CD_3COOD (pD = 5.1) with 30% methanol-d_4 a of; a) 1:1 mixture of **B1** and **S2**; b) 1:1 mixture of **B1** and **S1**; c) **B1**.

In Figure 3.9b the dt and the td peaks are shifted from 7.68 ppm and 7.48 ppm to 7.70 ppm and 7.50 ppm, respectively. These peaks are slightly shifted (< 0.05 ppm) due to changes in the environment caused by the presence of other species in solution. The multiplet signal, only shifted by 0.02 ppm, is also broader. The new broad peak between 7.32 ppm and 7.18 ppm and the broader multiplet between 7.41 ppm and 7.34 ppm are ascribable to the boronate ester species. The broadness is caused by the overlapping of the peaks for unbound and bound **B1** species[12]. A 1:1 mixture of **B1** and **S2** was analysed by ^{1}H NMR using the same conditions employed for the **S1** mixture (Figure 3.9a). The mixture of **B1** and **S2** presents only three

signals, all of which are only slightly shifted (< 0.05 ppm) when compared to **B1**. No extra signals are present; hence no complexation was seen between **B1** and **S2**.

Furthermore, to confirm that the signals at 7.36 ppm and 7.24 ppm can be assigned to the complex different receptor : ligand ratios were explored. The integral area of these signals is expected to change upon changing the ratio. Mixtures with 1:1, 1:2 and 1:3 ratios of **B1** and **S1** were analysed by NMR in 0.1 M CD_3COOD buffer with 30% MeOD-d_4 at pD 5.0 (Figure 3.10).

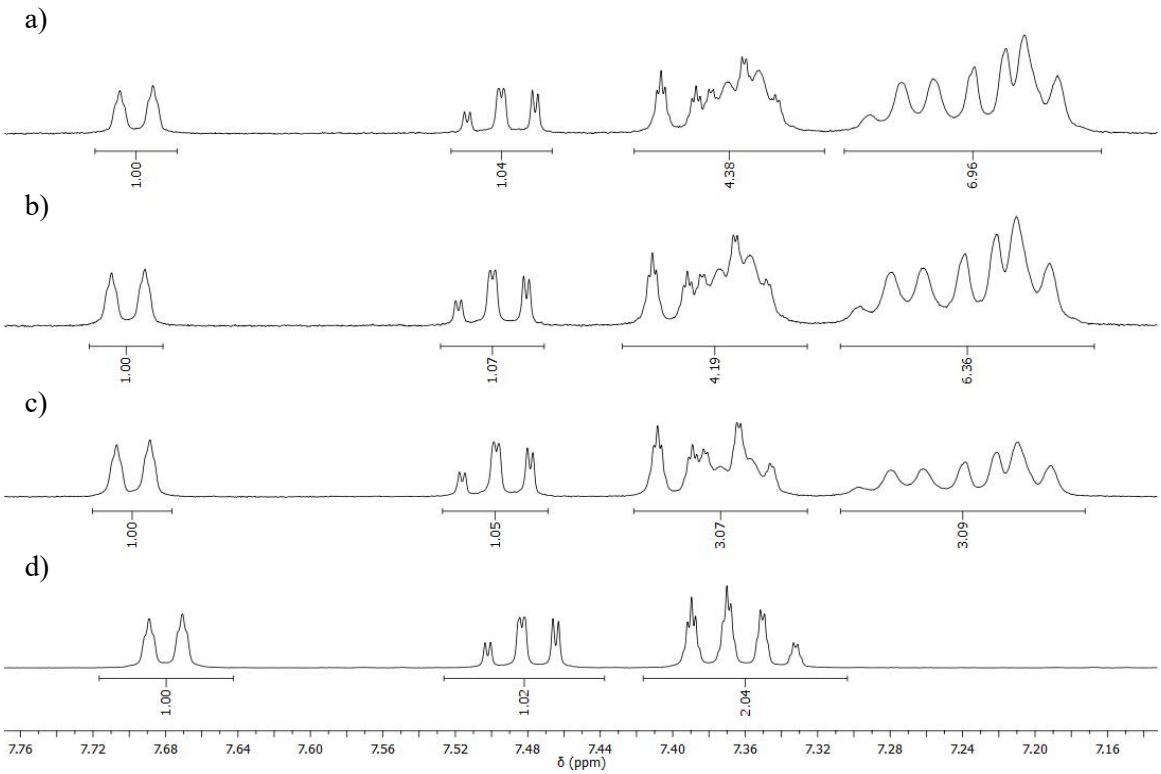


Figure 3.10. ^{1}H NMR spectra in 0.1 M CD_3COOD with 30% methanol-d_4 at pD 5.1 of mixtures of **B1** and **S1** with ratio of; a) 1:3; b) 1:2; c) 1:1; d) 1:0.

The double triplet at 7.68 ppm of the unbound **B1** (Figure 3.10d), was integrated to 1 as reference. The triplet of doublets at 7.48 ppm belongs to the unbound **B1** as its integral value does not vary with the ratios (Figure 3.10a,b,c,d). On the other hand, the integrated areas of the broad multiplets at 7.36 ppm and 7.24 ppm increases with the change in the **B1**: **S1** ratio. For

instance, the peak at 7.36 ppm has an integral area of 2 in Figure 3.10d corresponding to two aromatic protons of **B1**, which upon boronate esterification increases with the ratio, up to 4.38 for a 1:3 ratio (Figure 3.10a,b,c). This confirms that the signal is a result of the unbound **B1** and complex peaks overlapping. Furthermore, the integral area of the broad peak at 7.24 ppm also increases with the ratio, with an area of 6.96 for the 1:3 ratio (Figure 3.10a,b,c). This peak is present solely in mixtures of **B1** and **S1**, hence is ascribable entirely to the complex. It was therefore ascertained that the multiplets at 7.36 ppm and 7.24 ppm are signals of the boronate ester complex.

In addition to the NMR studies, ESI(-) MS analysis was performed on both mixtures in 0.1 M acetate buffer. The MS spectra (Figure 3.11a) of a 1:1 mixture of **B1** and **S1** showed the unbound **S1** peak at m/z = 308.10 $[M]^-$ and the boronate ester species at m/z = 424.14 $[M]^-$. The **B1** (m/z = 134.05) is not shown as its below the lower mass limit of the instrument. On the other hand, a mixture of **B1** and **S2** (Figure 3.11b) presented solely the peak of unbound **S2** at m/z = 322.12 whilst the boronate ester peak at m/z = 437.15 was not observed.

The NMR and MS analysis support the ITC findings by confirming the formation of the boronate ester for **B1** occurs solely with free sialic acid (**S1**). It was therefore established that the α-hydroxyacid group is the only binding site. Thus, the binding of boron-based receptors can arise only with the free form of sialic acid, and no complexation can be seen with sialic acid in glycoforms. These outcomes are believed to apply not only for benzoboroxoles but also to all boron-based receptors, such as phenylboronic acids and Wulff-type boronic acids, due to the analogy in the structures and binding properties. Furthermore, it is reasonable to assume that boron-based derivatives reported to bind sialylated glycans are instead interacting with other sugar residues of the carbohydrate chains[34, 43]. Nevertheless, boron-based receptors play an important role in the detection of sialic acid in its free form, which is a biomarker for sialic acid

related neurodegenerative diseases, certain kinds of tumours (*e.g.* prostate, urinary bladder and renal cell cancers)[1] and alcoholism-related liver dysfunctions[5].

a)

b)

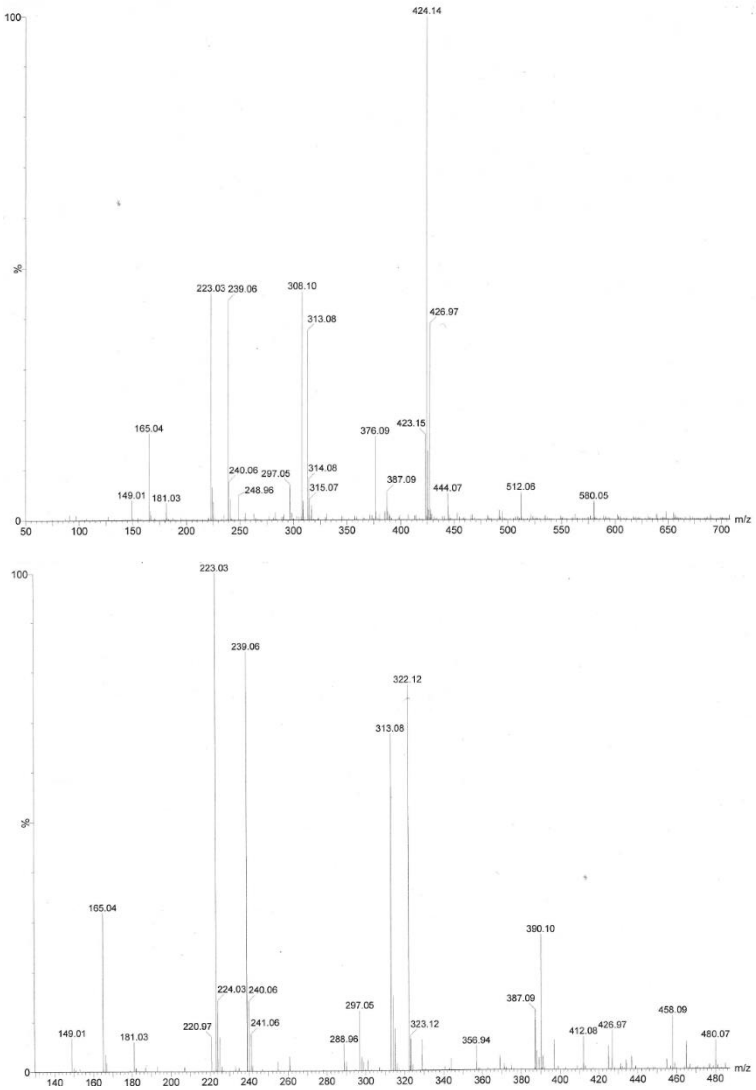


Figure 3.11. Electrospray ionisation (ESI) mass spectrometry (MS) in negative mode analysis in 0.1 M acetate buffer at pH 5.5 of a 1:1 mixture of; a) **B1** and **S1**; b) **B1** and **S2**.

3.3 Functionalised benzoboroxole receptors

3.3.1 Design of functionalised benzoboroxole receptors

Following the clarification of the binding site, a series of functionalised benzoboroxole receptors for sialic acid were designed and synthesised. The receptors were designed to contain multiple functional groups, which provide multipoint interactions with sialic acid and determine an enhancement in the affinity and selectivity. The benzoboroxole unit was selected in order to form a covalent and reversible boronate ester complex with the sugar[13]. Additional functional groups were designed to create supplementary non-covalent interactions with the ligand. To direct the rational design of the functionalised receptors, firstly the main points of interaction with sialic acid needed to be identified. These included hydroxyl groups, the carboxylic acid, the acetamide moiety and the carbon backbone. A variety of non-covalent interactions can be exploited to interact with the aforementioned functional groups. The hydroxyl and the acetamide groups can be bound via hydrogen bonds. Hydrogen bonds can be neutral or charge-reinforced, with the latter known to be stronger and therefore preferred[44]. Indeed, the hydrogen bond between two uncharged groups (*e.g.* hydroxyl groups) provides a small contribution to the binding energy of solely 0.5 - 1.5 kcal mol^{-1}. Conversely, the charge-reinforced hydrogen bond between a charged group and a neutral group is a stronger interaction with energy up to 4.5 kcal mol^{-1}. For instance, charge-reinforced hydrogen bonds have been shown to play an important role in, both natural and synthetic, sugar-binding lectins[45-47]. Furthermore, hydrophobic and CH-π interactions can be exploited to interact with the carbon backbone. On the other hand, the carboxylic acid cannot be a non-covalent point of interaction as it is engaged in the boronate ester. Thus, charged groups present in the receptor will only form charged-reinforced hydrogen bonds with the sugar polar group, and will not form any electrostatic interactions with the carboxylate.

Many boron-based receptors presented in the literature present functional groups linked to the phenyl ring of a boronic acid or benzoboroxole unit[7, 16, 33]. Functionalisation of benzoboroxoles at the oxaborole ring are less common. Nevertheless, functionalisation of the oxaborole, via position 7, can offer a greater flexibility of the side chain compared to aromatic functional groups. All the receptors presented herein were functionalised at the oxaborole ring. In position 7, a side chain bearing a positively charged group was introduced in order to exploit charge-reinforced hydrogen bonds. Two different cationic moieties were selected; the amino and the guanidino group. The guanidino group was expected to have a great effect in enhancing the affinity as it can generate a wider network of hydrogen bonds with the ligand. The guanidino group, and other bidentate groups, charged or uncharged, have been employed in synthetic lectins to create networks of non-covalent interactions[47-51]. In addition to the cationic group a phenyl ring was introduced in the side chain of two of the six receptors. Aromatic rings can be found in the binding pocket lectins and as core structure in synthetic receptors[38, 45-46, 52]. The phenyl group provides hydrophobic and CH-π interactions with the carbon backbone of sialic acid.

Furthermore, it was important to establish the optimal distance between the two main binding units of the receptors, the benzoboroxole and the charged group. The distance between functional groups of a receptor can define the selectivity and affinity towards a specific ligand. The prime example of this is a diboronic acid receptor developed by Tsukagoshi *et al*.[53] The binding affinity of boronic acids to glucose is low, with BA preferred ligand being fructose. However, when two boronic acids were linked together and appropriately distanced the selectivity was switched from fructose to glucose and this was associated also to affinity enhancement. Hence, the relative position of the functional groups in a receptor can significantly alter the binding properties. In order to maximise affinity and selectivity it was

pivotal to investigate the optimal distance between the receptor functional groups. For this reason, a variety of receptors with different distances between the benzoboroxole core and the charged group were synthesised (Figure 3.12). In receptors **B2** and **B3** the two groups are in close proximity with the amino and guanidino groups, respectively, linked to the oxaborole ring via a methylene group, a 1 carbon chain. All other receptors were obtained from the corresponding **B2** and **B3** via amide coupling and therefore present an amide group. In receptors **B4** and **B5** the charged group is separated from the benzoboroxole ring by a chain containing 2 extra carbon atoms. Whilst, in receptors **B6** and **B7** the side chain consists in a benzyl group linked to the amide. The presence of an amide in the side chain is a potential point of constraint for the receptor, but it is also provides an additional point of interaction for the sugar.

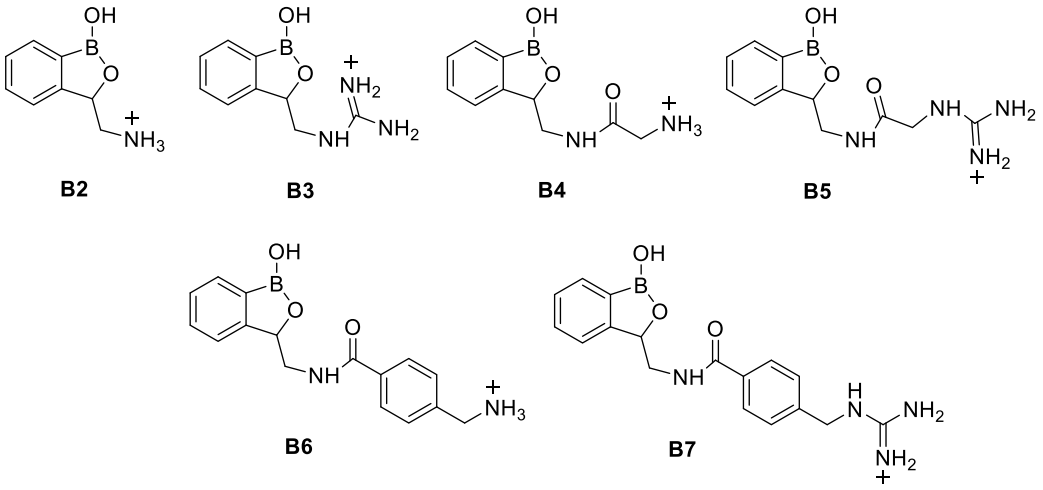


Figure 3.12. Positively charged benzoboroxole receptors **B2** – **B7**.

3.3.2 Synthesis of amino-benzoboroxole receptors

In order to synthesise the abovementioned functionalised benzoboroxole receptors a precursor with a functional group on the oxaborole ring, which can be further derivatised, is needed.

Benzoboroxole derivatives can be obtained through three main different synthetic strategies starting from *ortho*-substituted aromatic bromide (Figure 3.13a), 2-formylphenylboronic acid (Figure 3.13b), or boronic pinacol esters (Figure 3.13c)[54].

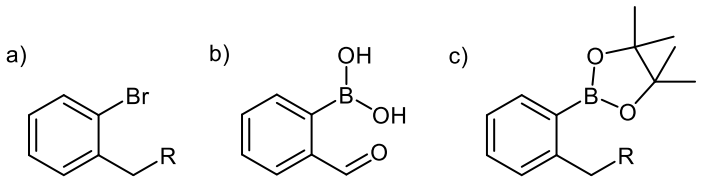


Figure 3.13. The different starting material for the synthesis of the oxaborole ring; a) Aromatic bromide; b) 2-Formylphenylboronic acid; c) Boronic pinacol ester.

The synthesis of benzoboroxole derivatives from boronic pinacol esters is the most common route to obtain a benzoboroxole ring. However, this strategy does not allow functionalisation of the newly generated oxaborole ring. Conversely, the synthesis of benzoboroxole derivatives from *ortho*-substituted aryl bromide and 2-formylphenylboronic acid allows functionalisation of the oxaborole ring. The synthesis of benzoboroxole derivatives from an appropriate *ortho*-substituted aryl bromide requires a bromo-lithium exchange which is followed by *trans*-metalation with a source of boron. This is then followed by dehydration and cyclisation to give the substituted benzoboroxole ring. This method requires a strong base, such as butyllithium, which was not considered due to safety concerns relating to its extreme reactivity[55]. On the other hand, the 2-formylphenylboronic acid synthetic route affords a variety of benzoboroxole substituted in position 7 employing common and safer reagents. The functional groups which can be introduced on the oxaborole ring include phenyl and heterocyclic rings, amides, alkenes, tertiary amines, esters.

Whilst the introduction of a primary amine, to give **B2**, is not directly possible. A primary amine can be obtained by reduction of a nitro group, which can be introduced through a nitroaldol reaction, also known as Henry reaction[56]. The nitro functionalised benzoboroxole (**B8**) was

obtained from 2-formylphenylboronic acid and nitromethane in basic aqueous solution. The reaction proceeds with the deprotonation of the nitromethane followed by the nucleophilic attack of the nitromethane anion at the carbonyl group of 2-formylphenylboronic acid, the resulting intermediates are then protonated to give the nitro alcohol[56]. The nitro alcohol undergoes spontaneous dehydration leading to the cyclisation of the oxaborole ring (Figure 3.14). Thus, the resulting benzoboroxole is functionalised in position 7 with a nitromethyl group.

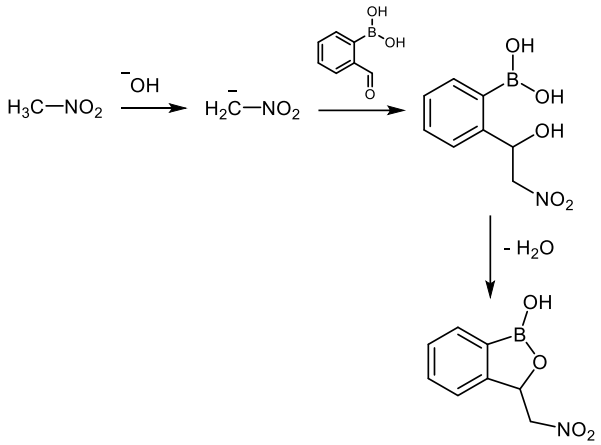


Figure 3.14. Henry reaction between nitromethane and 2-formylphenylboronic acid followed by dehydration leading to the formation of the oxaborole ring.

Furthermore, the functionalisation of position 7 renders this carbon a chiral centre with the reaction providing a racemic mixture. The two enantiomers were not separated at this or at any later stage, thus all benzoboroxole derivatives synthesised herein were employed as racemic mixtures.

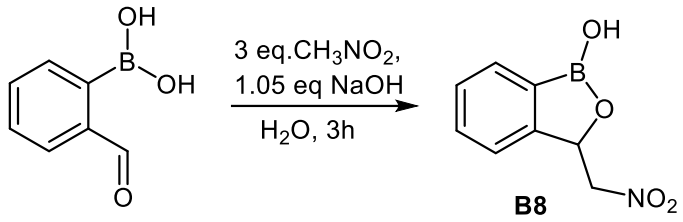


Figure 3.15. Synthesis of **B8** from the 2-formylphenylboronic acid with nitromethane under basic conditions.

The synthesis of **B8** (Figure 3.15) was performed following a literature procedure which was modified affording improved yield[57]. The reaction was performed in the presence of 3 equivalents of nitromethane in order to promote completion. The reaction was stopped by acidification with 2 M hydrochloric acid and the precipitate filtered immediately after in order to avoid its partial dissolution in water. It was observed that the direct acidification of the reaction mixture generates by-products. In order to avoid this the solution must be diluted with water prior to acidification. This procedure afforded the pure **B8** in high yields (92%). The ^{1}H NMR of **B8** (Figure 3.16) presents the hydroxyl group of boron as a singlet at 9.51 ppm. The region between 7.80 ppm and 7.35 ppm shows the four aromatic protons of the benzoboroxole aromatic ring. The proton H_7 of the CH of oxaborole ring gives a double doublet at 5.78 ppm. Furthermore, the CH_2 present two stereochemically different protons, H_{8a} and H_{8b}, which gives two different signals, each of integral of 1, at 5.34 ppm and 4.58 ppm.

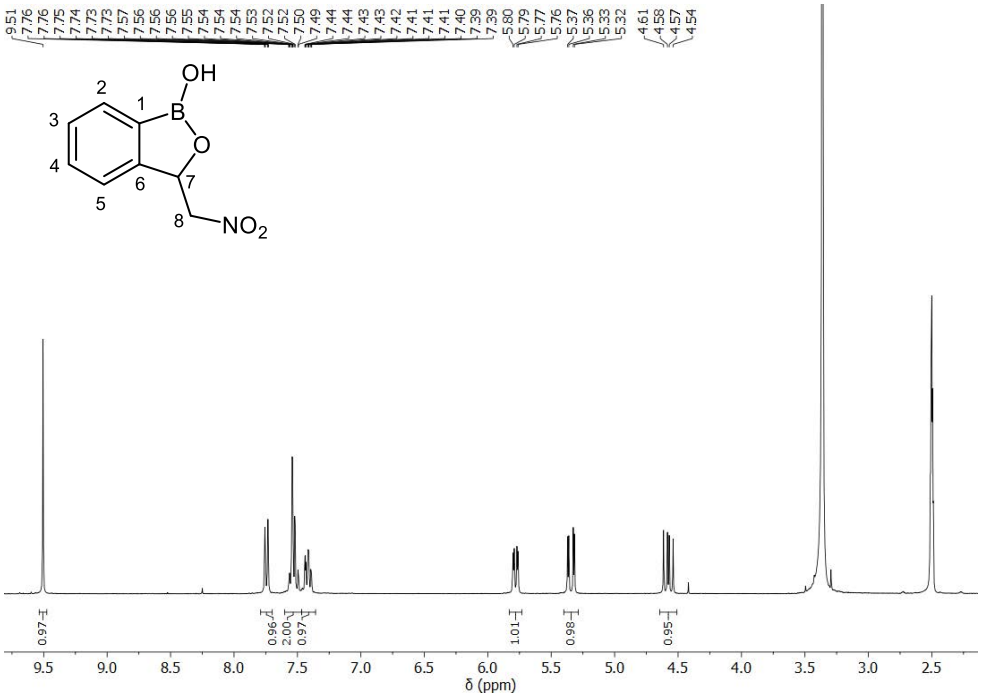


Figure 3.16. ^{1}H NMR spectrum of **B8** (300 MHz, DMSO-d_6, 298 K).

The nitro group of **B8** was then reduced to give the corresponding amino benzoboroxole derivative, **B2**. Initially, the reaction was attempted without success with common reductive strategies, such as H_2 at atmospheric pressure with different palladium catalysts (*i.e.* $Pd(OH)_2$ and Pd/C). The reaction was also attempted with hydrazine and palladium catalysis. However, no conversion to the corresponding amino group was observed by TLC or NMR. This is due to the lower reactivity to hydrogenation of aliphatic nitro groups when compared to aromatic ones. For instance, the hydrogenation of **B8**, which proceed at atmospheric pressure for an aromatic nitro group, requires a high pressure of hydrogen for an aliphatic group. Transfer hydrogenation with ammonium formate was also attempted without any conversion being observed. The reduction of **B8** was achieved in the presence of $NiCl_2$ hexahydrate and $NaBH_4$ in methanol (Figure 3.17)[58]. $NaBH_4$ reduces $NiCl_2$ to Ni_2B which then catalyses the *in situ* release of H_2 from $NaBH_4$[59].

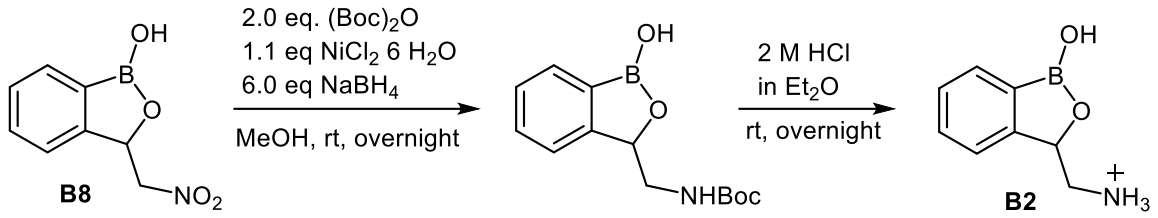


Figure 3.17. Two-step synthesis of **B2** from **B8**.

Upon formation the amine is protected *in situ* with Boc_2O according to the literature procedure[60], this is necessary to avoid the dimerisation of the amine with the imine intermediate[58] (Figure 3.18).

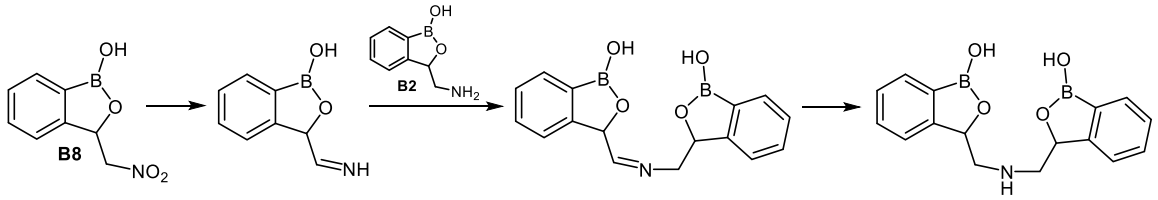


Figure 3.18. Side reaction of dimerisation between the amine **B2** and imine intermediates in the reduction of the **B8**.

Following the reduction, the catalyst was removed by filtration and the crude was directly deprotected by acidic treatment with 2 M hydrogen chloride in diethyl ether to afford the complete cleavage of the Boc group. The precipitate was then triturated with diethyl ether giving the pure derivative **B2** as hydrochloride salt with a yield, over the two steps, of 78%.

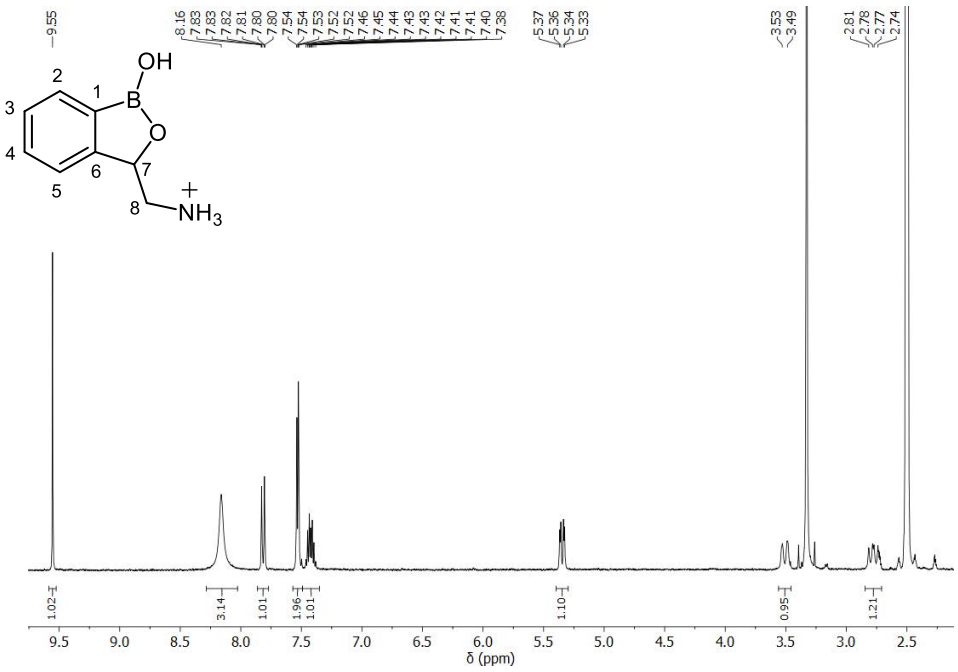


Figure 3.19. ^{1}H NMR spectrum of **B2** (300 MHz, DMSO-d_6, 298 K).

The reduction of the nitro group into the corresponding amino hydrochloride salt was confirmed by ^{1}H NMR (Figure 3.19). The signals of the protons of the CH_2 group, H_{8a} and H_{8b}, are shifted to a higher field from 5.34 ppm and 4.58 ppm to 3.51 ppm and 2.79 ppm. This is due to the increased shielding provided by the NH_3^+ group compared to the nitro group. The H_7 proton of the oxaborole ring is also shifted upfield, from 5.78 ppm to 5.35 ppm. Furthermore, the NH_3^+ group presents a broad singlet peak of integral 3 at 8.16 ppm. The broadness of this peak is due to the exchange of protons with water, thus confirming the presence of the NH_3^+ group.

The derivative **B2**, in addition to being employed in the binding studies as receptor, was also the starting material for the synthesis of all other receptors via amide coupling or guanidinylation. Receptors **B4** and **B6** were synthesised via amide coupling of **B2** with Boc-glycine-OH and 4-(Boc-aminomethyl)benzoic acid, respectively. The coupling reaction was performed with 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDCI) and 4-dimethylaminopyridine (DMAP). DMAP mediated the deprotonation of charged amino

group rendering **B2** available for the coupling. DMAP and EDCI both contribute towards the activation of the carboxylic acid. The solvent in which the amide coupling was performed was selected based on the high polarity of **B2**, which is insoluble in most organic solvents except for methanol and dimethylformamide (DMF). Methanol cannot be employed under these reaction conditions as it would react with the carboxylic acid to give the correspondent methyl ester via Steglich esterification[61]. Thus, DMF was selected as solvent for the coupling.

Receptor **B4** was synthesised by a two-step procedure. In the first step **B2** reacted with an excess of Boc-glycine-OH (1.2 equivalent) to give the Boc-protected intermediate (Figure 3.20).

DMAP and EDCI were easily removed by extraction with 1 M HCl, whilst the unreacted Boc-glycine-OH was removed by washing with a saturated solution of sodium bicarbonate. Unfortunately, this work up procedure was responsible for the low overall yield of this synthetic procedure, as the Boc-protected intermediate was partially lost during the acidic wash. This could be due to the compound being partially soluble in water. Alternatively, the hydrochloric acid could have mediated the deprotection of the Boc group giving the deprotected product **B4**, which has high water solubility. Following the extraction, the crude was treated with 2 M hydrogen chloride in diethyl ether to cleave the Boc group. The deprotection was quantitative and no further purification was required. The **B4** derivative was obtained with a yield of 31% across the two-step synthetic procedure.

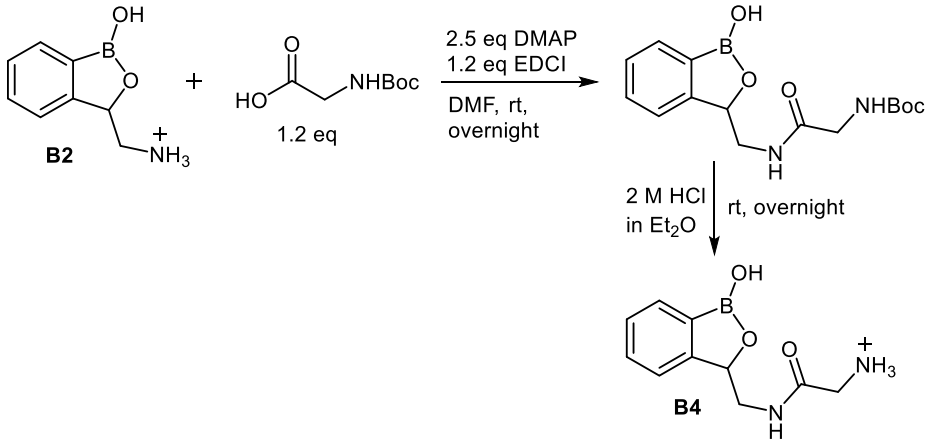


Figure 3.20. Amide coupling of **B2** with Boc-Gly-OH followed by Boc deprotection under acidic conditions to give **B4**.

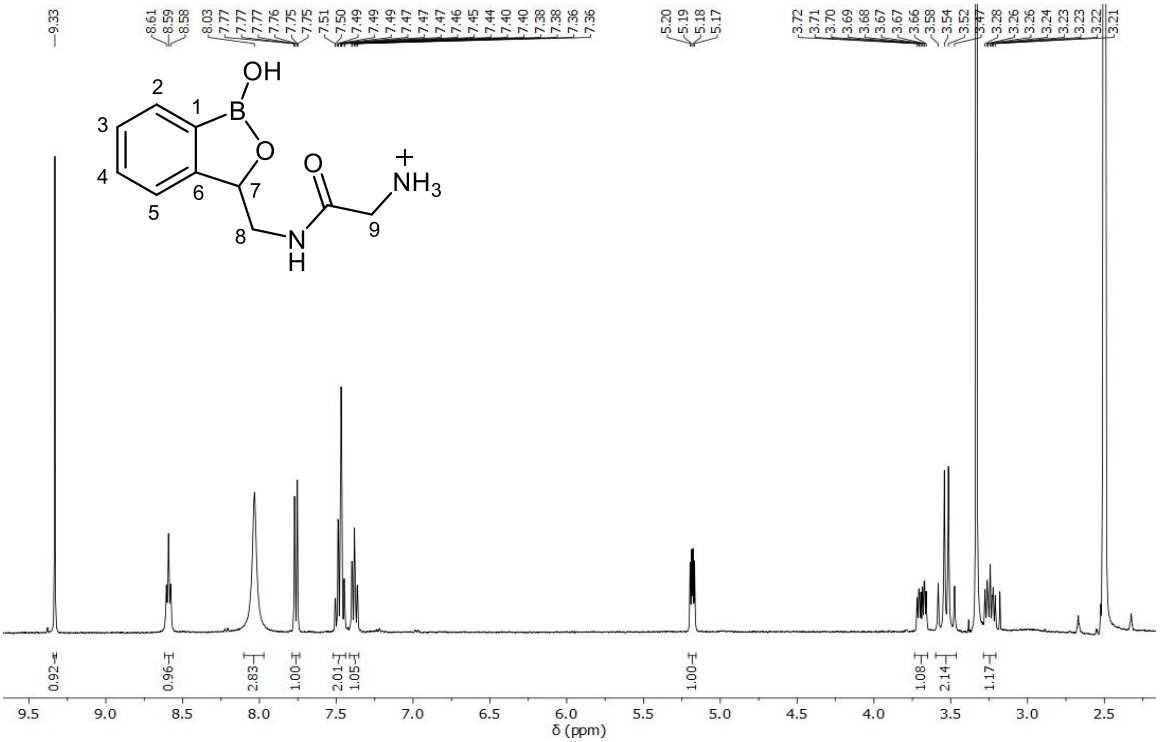


Figure 3.21. ^{1}H NMR spectrum of **B4** (400 MHz, DMSO-d_6, 298 K).

The ^{1}H NMR spectrum of **B4** (Figure 3.21) shows the two protons, H_{8a} and H_{8b} as two doublet of doublet of doublets signals at 3.69 ppm and 3.24 ppm. These signals are shifted downfield as the amide group is more electron withdrawing than the amino group. The NH proton of the

amide group is shown as a triplet at 8.59 ppm. The protons H_9 present a quartet at 3.55 ppm, with an integral area of 2. The NH_3^+ signal is a broad peak at 8.03 ppm.

Benzoboroxole derivative **B6** (Figure 3.22) was synthesised from **B2** and 4-(Boc-aminomethyl)benzoic acid with the same reaction conditions as described for the synthesis of **B4**.

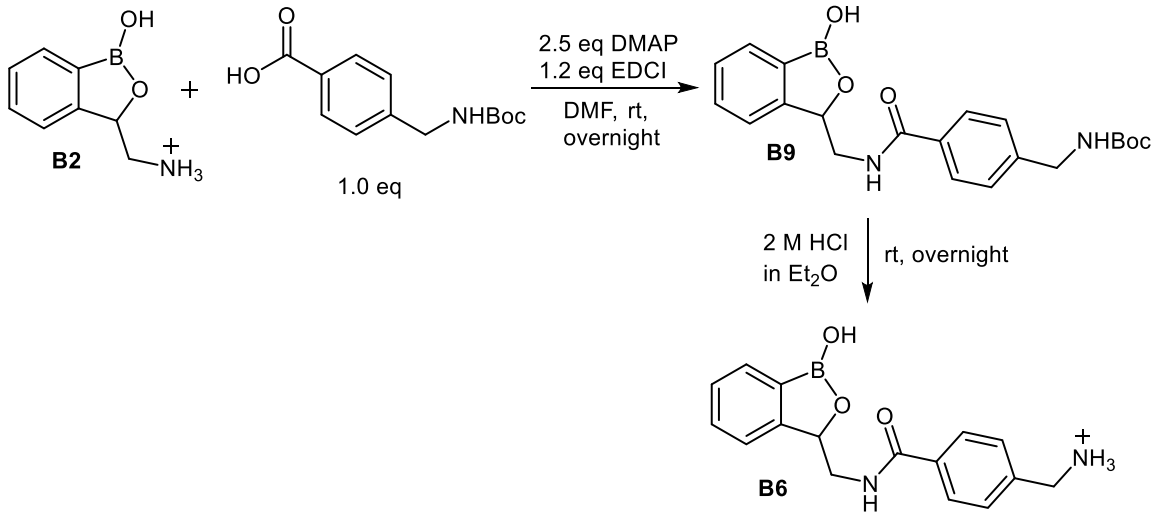


Figure 3.22. Amide coupling between **B2** and 4-(Boc-aminomethyl)benzoic acid to give Boc protected intermediate **B9** which is then treated under acidic conditions to afford **B6**.

In spite of this, the work up of this reaction differed from the one previously described. Initially the organic phase was washed with 1 M HCl and then with saturated $NaHCO_3$. However, the low water solubility of 4-(Boc-aminomethyl)benzoic acid only allowed partial removal of the unreacted reagent. Following extraction, the crude was purified by recrystallisation with a mixture of hexane and ethyl acetate to afford the Boc-protected intermediate **B9** with a yield of 31%. It was important for the purification to be performed prior to cleavage of the Boc protection. Had the acidic treatment been performed on the crude this would have produced a mixture of 4-(aminomethyl)benzoic acid and **B6** which, due to their high polarity, could not be easily purified by recrystallisation or column chromatography.

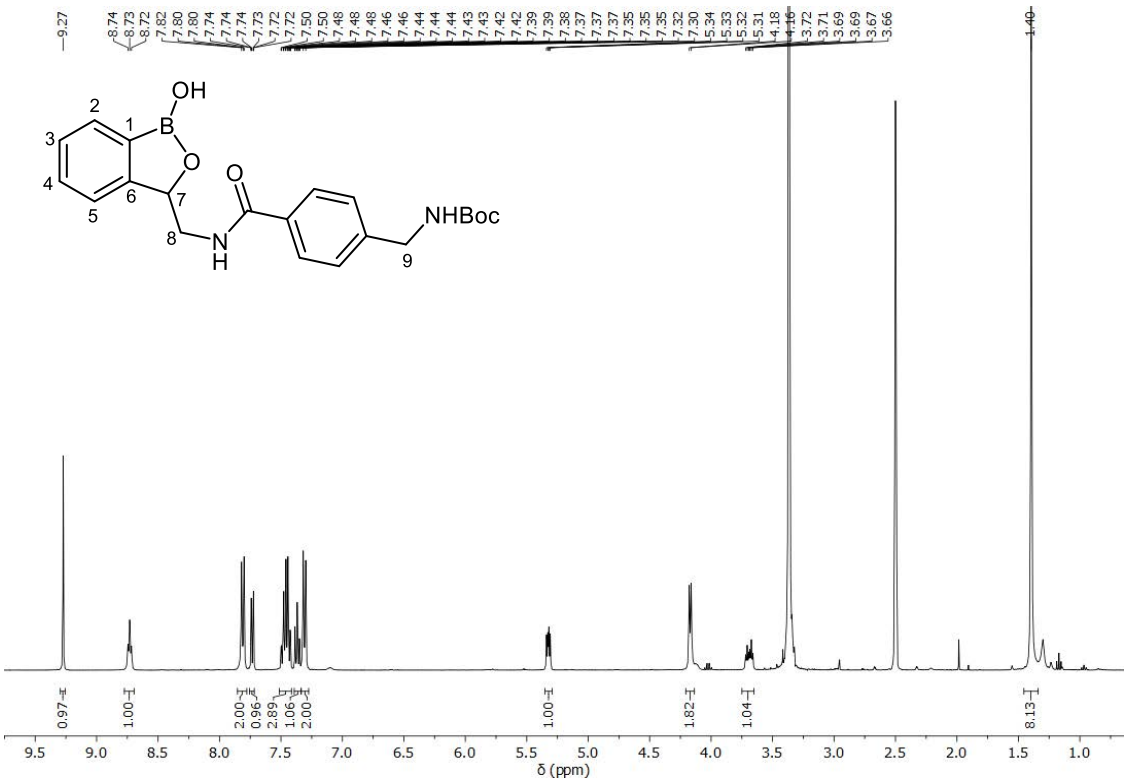


Figure 3.23. ^{1}H NMR spectrum of **B9** (400 MHz, DMSO-d_6, 298 K).

The ^{1}H NMR spectra of **B9** (Figure 3.23) presents five aromatic multiplets, between 7.90 ppm and 7.25 ppm, corresponding to the protons of both phenyl groups. The proton of the amide NH shows a triplet signal at 8.73 ppm. Of the two protons H_{8a} and H_{8b} of the oxaborole ring only one signal at 3.69 ppm can be seen in the ^{1}H NMR. The second signal of the oxaborole CH_2 group overlaps with the water peak. The protons H_9 present a doublet of integral area of 2 at 4.17 ppm. For the protected amine group the N*H* signal overlaps with the aromatic protons, whilst the Boc methyl groups present a singlet at 1.40 ppm. The **B9** was then deprotected with a 2 M hydrogen chloride in diethyl ether to give the **B6** as an off-white solid with quantitative yield. In the ^{1}H NMR spectrum of **B6** (Figure 3.24) the Boc signal at 1.40 ppm is not present, whilst a broad singlet corresponding to the NH_3^+ is observed at 8.52 ppm. In addition, H_9 is

shifted from 4.17 ppm to 4.07 ppm, hence the NMR analysis confirms the full deprotection of the Boc group.

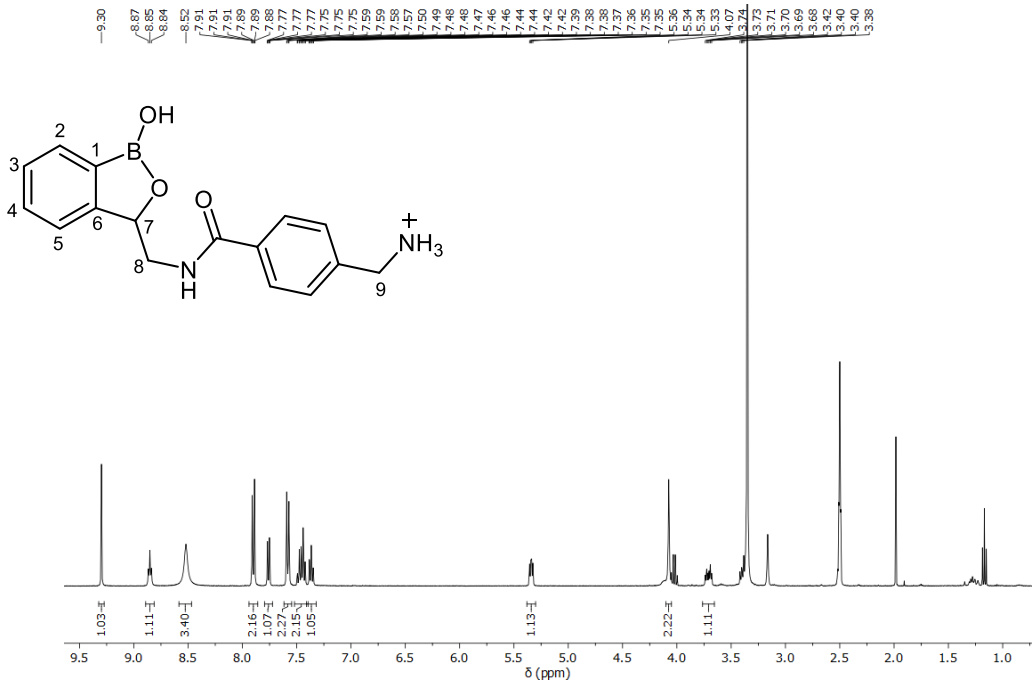


Figure 3.24. ^{1}H NMR spectrum of **B6** (400 MHz, DMSO-d_6, 298 K).

3.3.3 Synthesis of guanidino-benzoboroxole receptors

The guanidino benzoboroxole receptors, **B3**, **B5** and **B7**, were obtained by guanidinylation of the corresponding amino-benzoboroxole derivatives **B2**, **B4** and **B6**. Their synthesis consists of two steps (Figure 3.25). The first step was the guanidinilation of the amino derivative with 1.1 equivalent of N,N′-di-Boc-1H-pyrazole-1-carboxamidine in the presence of triethylamine. This was followed by the cleavage of the Boc groups in acidic conditions to give **B3**, **B5** and **B7**.

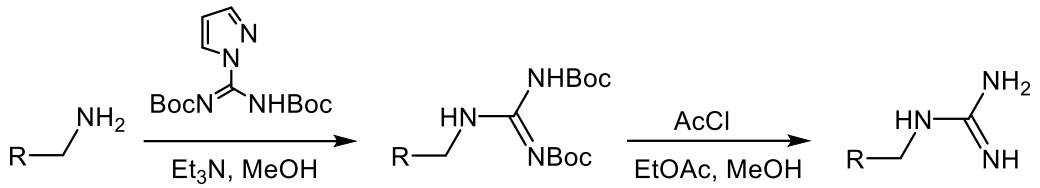


Figure 3.25. Schematic representation of the two-step guanidinylation reaction.

The reagent 1H-pyrazole-1-carboxamidine is well known for guanidinylation of primary amines[62]. In this work the protected analogue, N,N′-di-Boc-1H-pyrazole-1-carboxamidine, is used instead. In this fashion the guanidinylated receptors are initially obtained in their Boc-protected form. The Boc-protected intermediates are less polar and therefore can be purified by column chromatography, whilst the unprotected derivative would require reverse phase chromatography due to their high polarity. Following purification, the protected intermediates **B10 - B12** were treated in acidic condition to cleave the Boc groups. The cleavage of the Boc groups was initially attempted with common reagents such as TFA in DCM[63], 2 M HCl in diethyl ether[60] and in dioxane[64], however none of these methods were able to fully remove both protecting groups. According to the literature, this could be a result of low solubility of the reaction intermediates in the aforementioned solvents[65]. The full deprotection was achieved generating hydrochloric acid *in situ* by addition of acetyl chloride to a mixture of methanol and ethyl acetate. Acetyl chloride reacts with methanol to form hydrochloric acid gas and methyl acetate. The addition of acetyl chloride in the methanol and ethyl acetate mixture was performed at 0 °C to reduce evaporation of the HCl gas. The mixture of methanol and ethyl acetate ensures that starting material and intermediates are fully dissolved. The reaction was stopped by evaporation of the solvents under reduced pressure. The main limitation of this procedure is the requirement of multiple additions of acetyl chloride across 48 to 72 hours to achieve completion.

Receptors **B3**, **B5** and **B7** were all synthesised following the procedures described above (Figure 3.25) but differ in the purification strategies due to the different polarities of the intermediates **B10**, **B11** and **B12**.

Intermediate **B10** was synthesised from **B2** following the procedure described above (Figure 3.26). The purification was performed by column chromatography in 85% hexane and 15% ethyl acetate, with a yield of 25%. The low yield is caused by the high polarity of **B10** which is retained by the column.

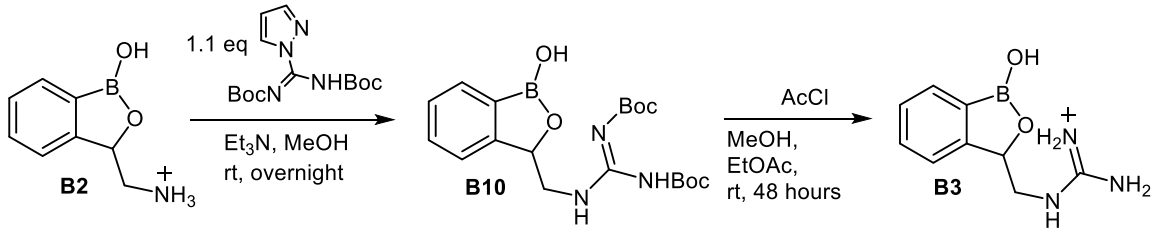


Figure 3.26. Two-step guanidinylation of **B2** to afford receptor **B3**.

The ^{1}H NMR spectrum of **B10** (Figure 3.27) confirms the successful guanidinylation. In particular, the CH_2NH signal is a triplet at 8.46 ppm, whilst the N*H*Boc signal is a singlet at 11.48 ppm. The methyl protons of two Boc groups give two singlets at 1.47 ppm and 1.39 ppm. In addition, due to the conversion of the NH_3^+ group into a Boc-guanidine, the protons H_{8a} and H_{8b} are shifted to lower fields, from 3.51 ppm and 2.79 ppm to 3.96 ppm and 3.25 ppm, respectively.

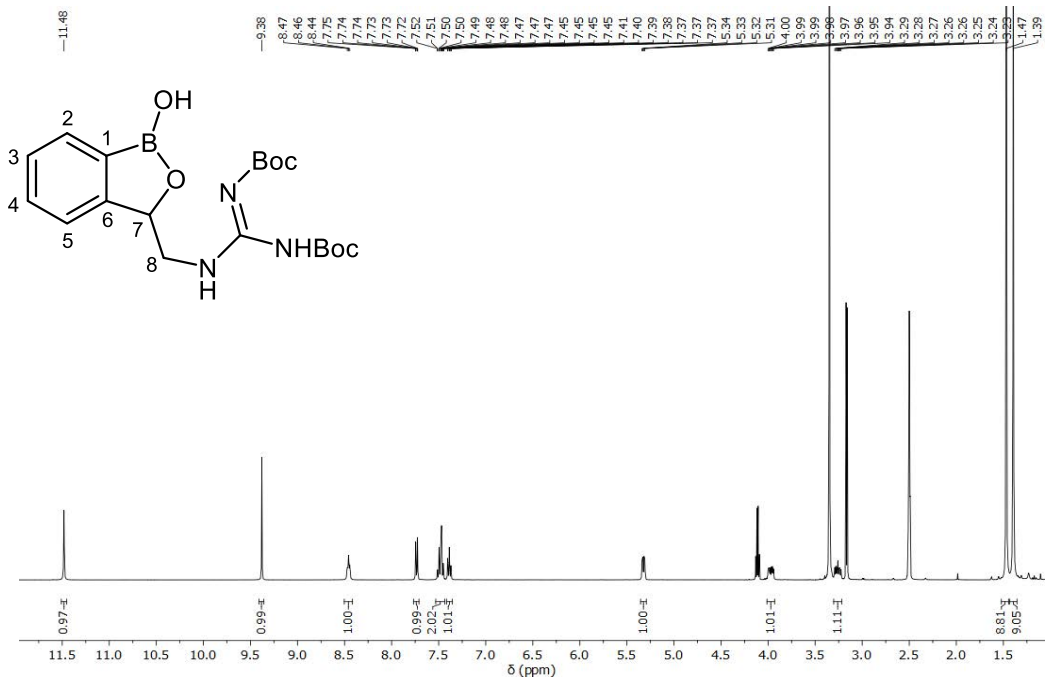


Figure 3.27. ^{1}H NMR spectrum of **B10** (400 MHz, DMSO-d_6, 298 K).

Derivative **B10** was deprotected in acidic conditions described previously to give **B3**. The complete cleavage of the protecting group was verified by ^{1}H NMR (Figure 3.28). The two singlet peaks at 1.47 ppm and 1.39 ppm of the Boc groups are no longer present after the deprotection. Moreover, the N*H*Boc signal at 11.48 ppm is not present either, and the CH_2N*H* is shifted at 7.61 ppm. The guanidino group presents only one broad peak at 6.96 ppm corresponding to the NH_2^+, due to the resonance of the guanidino group. In addition, the broad water peak at 3.37 ppm indicates exchanging of protons with the guanidino moiety.

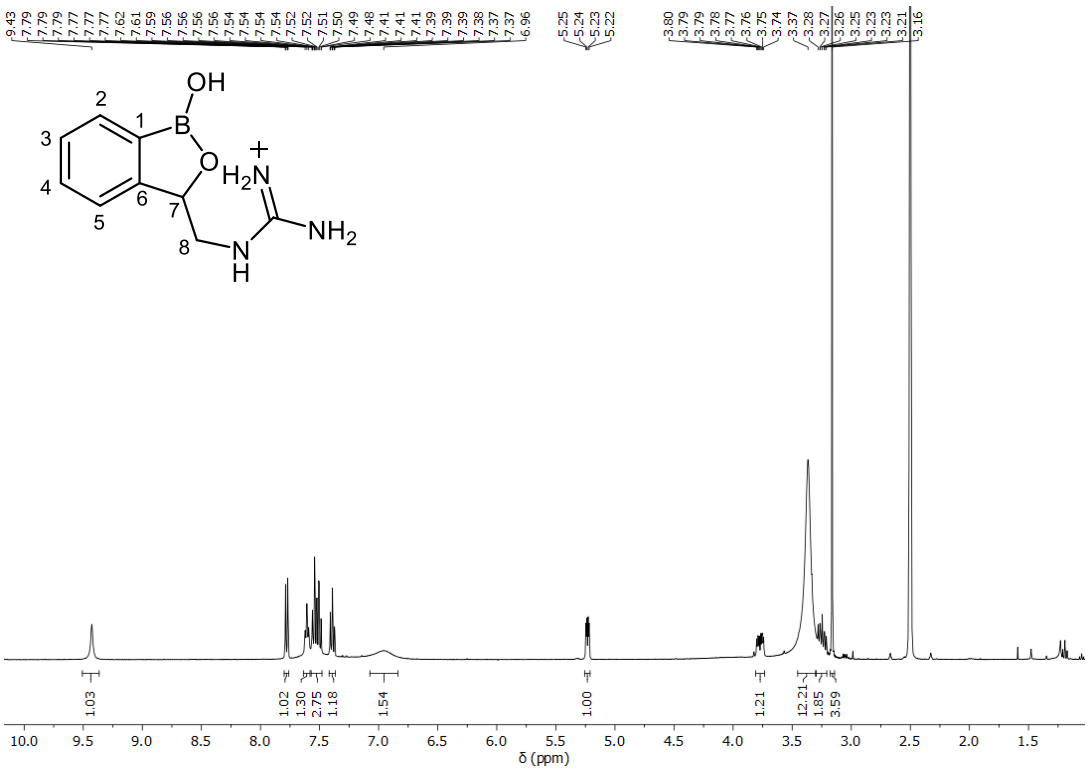


Figure 3.28. ^{1}H NMR spectrum of **B3** (400 MHz, DMSO-d_6, 298 K).

In the ^{1}H NMR spectra of **B3** small traces of impurities are present. Purification by recrystallisation was attempted with different mixtures of methanol and diethyl ether, ethyl acetate or chloroform but without success, probably due to the hygroscopic nature of the molecule. It was observed that even in the presence of small traces of moisture from air the compound turned into a sticky oil, preventing recrystallisation. Unfortunately, no other solvent was suitable for the recrystallisation of **B3** and the compound was too polar to be purified by column chromatography. Since the impurities were present only in trace quantities (< 5%) the compound was not purified further.

Derivative **B4** was treated with N,N′-di-Boc-1H-pyrazole-1-carboxamidine to give the Boc-protected derivative **B11** (Figure 3.29). This intermediate was purified by column chromatography with a gradient solvent system, 80% hexane and 20% ethyl acetate to 95% ethyl acetate and 5% methanol, to afford **B11** in 61% yield.

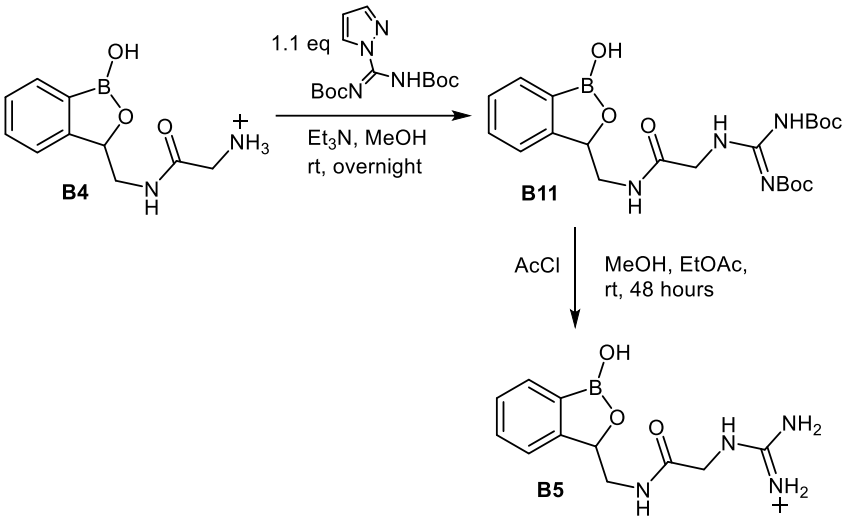


Figure 3.29. Two-step guanidinylation of **B4** to afford receptor **B5**.

The ^{1}H NMR of **B11** confirmed the guanidinylation of **B4** (Figure 3.30). The amino group of **B4** was converted into an amide group and the N*H* of the amide is a triplet at 8.69 ppm. The N*H*Boc shows a singlet at 11.43 ppm and the two Boc groups as two singlets at 1.48 and 1.38 ppm. The guanidinylation caused a shift of the CH_2 protons (H_9) from 3.55 ppm to 3.93 ppm.

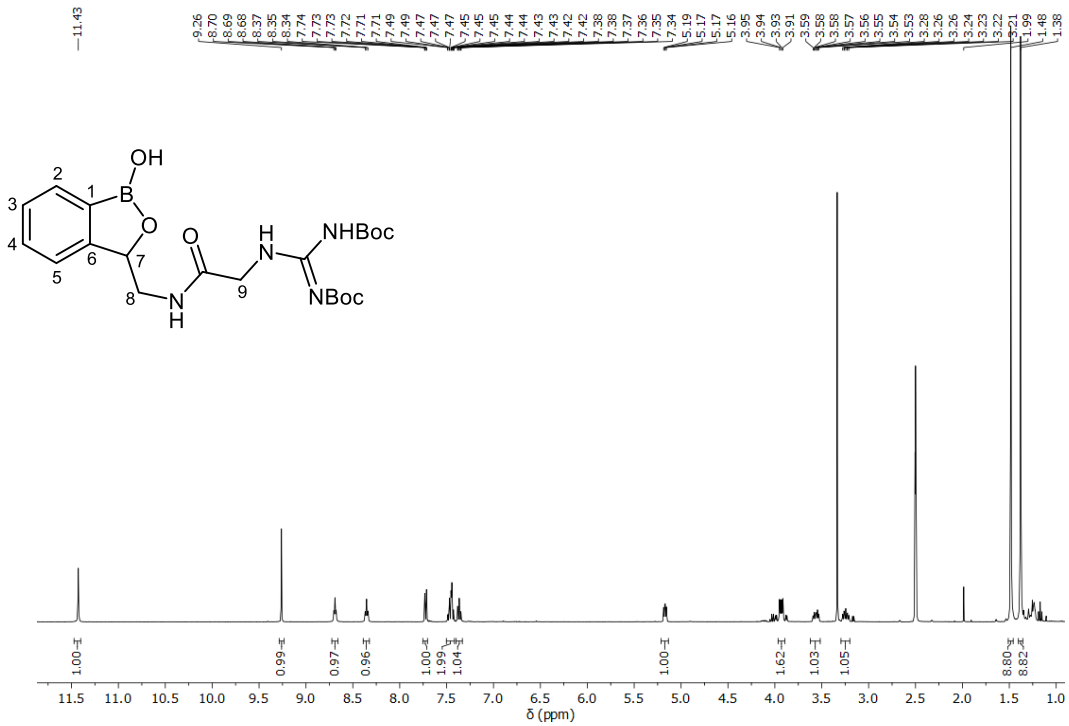


Figure 3.30. ^{1}H NMR spectrum of **B11** (400 MHz, DMSO-d_6, 298 K).

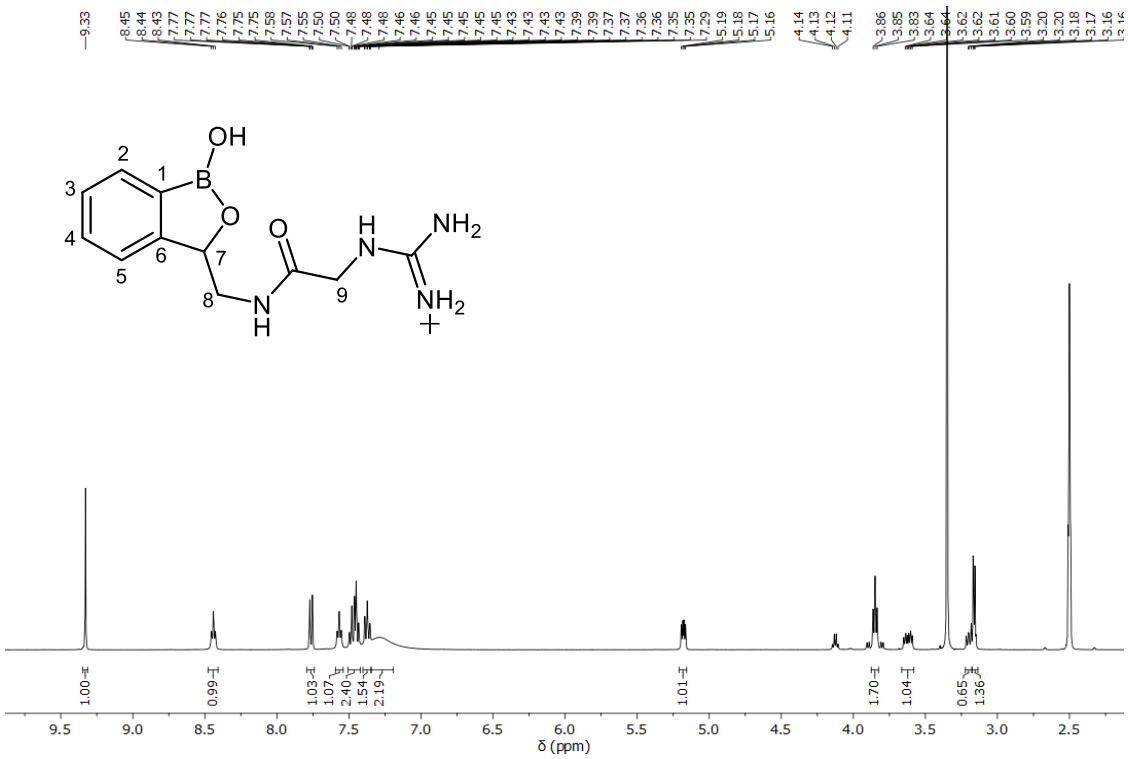


Figure 3.31. ^{1}H NMR spectrum of **B5** (400 MHz, DMSO-d_6, 298 K).

The intermediate **B11** was deprotected in acidic condition. The crude was recrystallised in a mixture of methanol and diethyl ether to afford the pure **B5** with a yield of 30%. The full cleavage of both Boc groups was confirmed by ^{1}H NMR (Figure 3.31) with no peaks being present in the region around 1.45 ppm. The N*H*Boc peak at 11.43 ppm is also not present, whilst the CH_2NH is shifted to 7.57 ppm. The NH_2^+ group presents a broad peak at 7.29 ppm, which partially overlaps with the aromatic signals.

The synthesis of **B7** consisted of the guanidinylation of **B6** to give the Boc-protected derivative **B12** (Figure 3.32). The crude was then purified by a gradient solvent system with increasing polarity, 80% hexane and 20% ethyl acetate to 100% ethyl acetate, to give the pure **B12** with a yield of 55%.

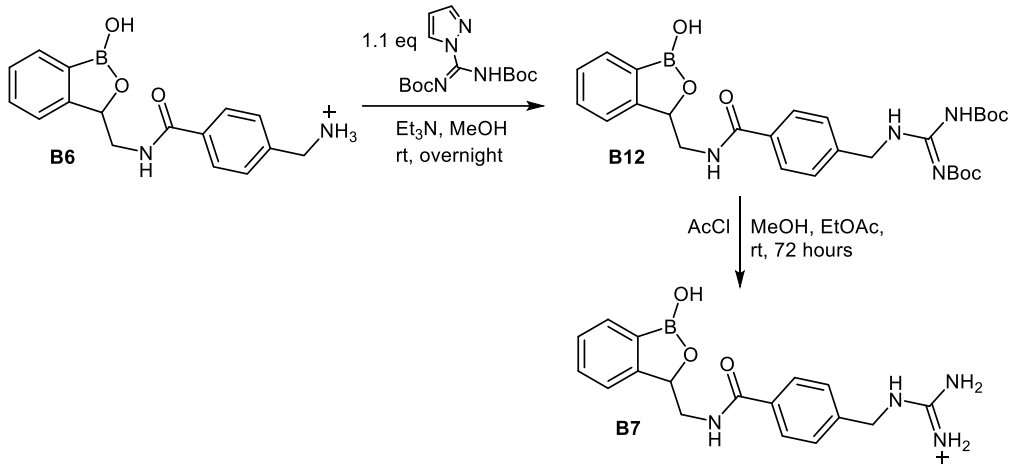


Figure 3.32. Two-step guanidinylation of **B6** to afford receptor **B7**.

The guanidinylation of **B6** was confirmed by NMR. The ^{1}H NMR of **B12** (Figure 3.33) presents the two singlets of Boc methyl groups at 1.48 ppm and 1.37 ppm. The N*H*Boc signal is a singlet at 11.54 ppm whilst the CH_2NH signal overlaps with the amide signal at 8.76 ppm. The protons H_9 are shifted to lower fields from 4.07 ppm to 4.58 ppm, due to the guanidinylation of the amino group.

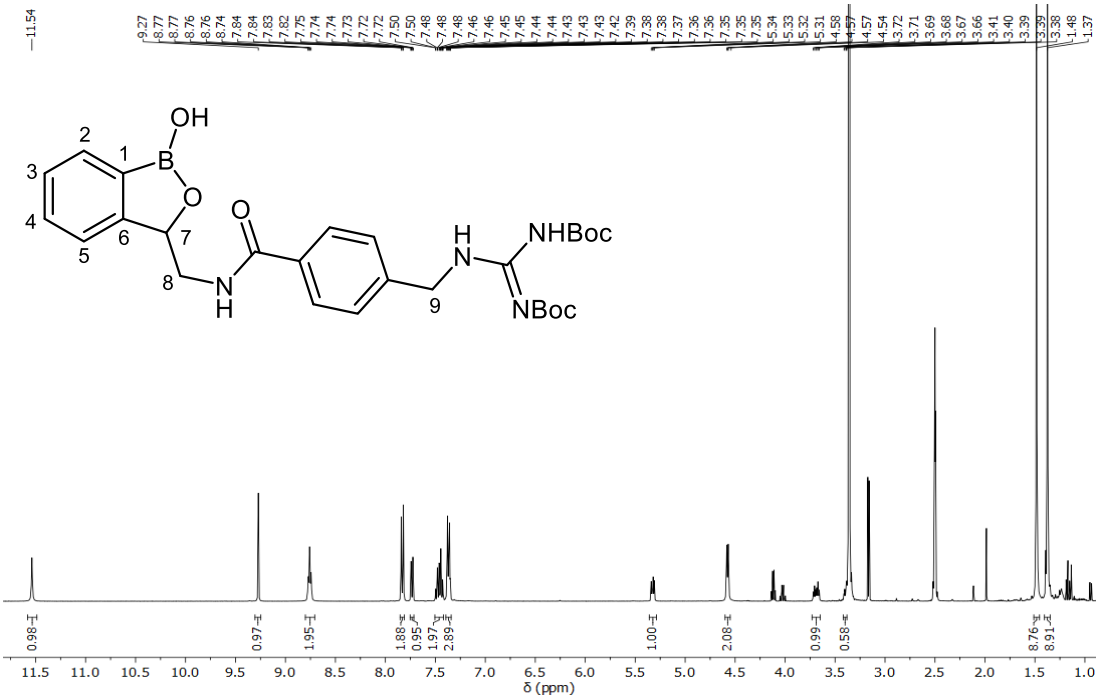


Figure 3.33. ^{1}H NMR spectrum of **B12** (400 MHz, DMSO-d_6, 298 K).

Derivative **B12** was deprotected in acidic conditions to give the pure **B7** as an off-white solid with quantitative yield. The full deprotection is confirmed by ^{1}H NMR analysis of **B7** (Figure 3.34). The two singlets of the Boc group at 1.48 ppm and 1.37 ppm are not present as well as the N*H*Boc signal at 11.54 ppm. The CH_2NH signal is shifted at higher fields, with a chemical shift of 8.23 ppm. The NH_2^+ signal of **B7** is a broad peak overlapping with other signals in the aromatic region.

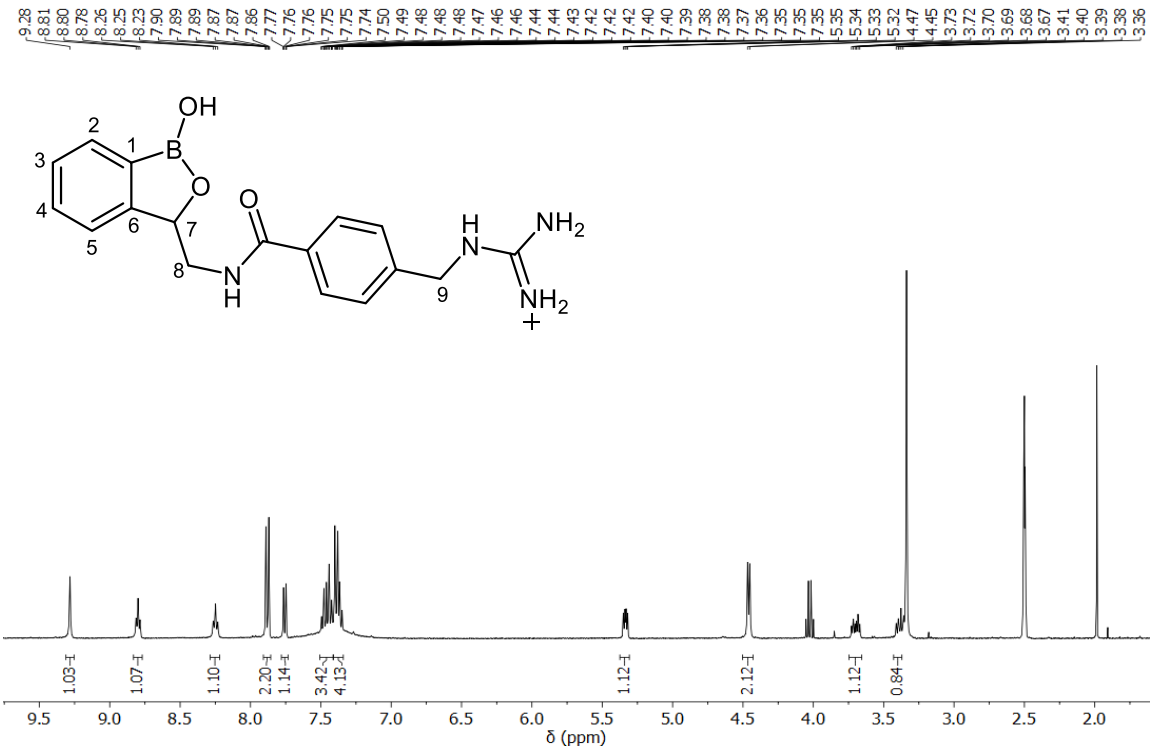


Figure 3.34. ^{1}H NMR spectrum of **B7** (400 MHz, DMSO-d_6, 298 K).

In addition, to the benzoboroxole receptors a control molecule was synthesised. The N-benzylguanidine (**B13**), as hydrochloride salt, was employed to assess the role of the charged and phenyl groups in the binding of sialic acid.

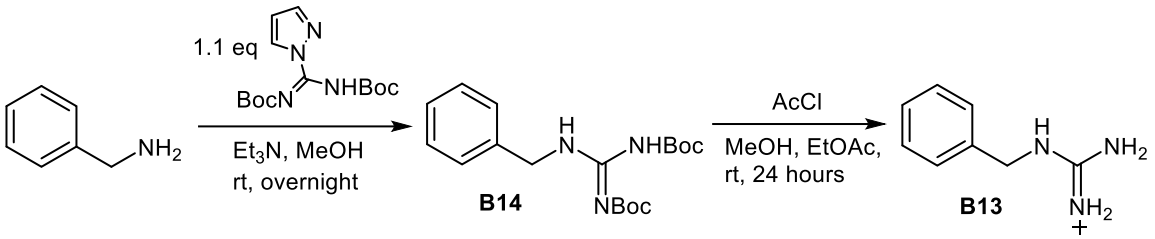


Figure 3.35. Two-step synthesis of control molecule **B13**.

Compound **B13** was synthesised from benzylamine with the same experimental conditions previously detailed for the guanidino receptors (Figure 3.35). Upon formation the Boc-protected intermediate a white precipitate was formed and filtered. This was deprotected with acetyl chloride to afford the molecule **B13**, with 18.0% yield. The ^{1}H NMR spectrum of **B13** (Figure 3.36) showed the NH_2^+ peak as a broad peak overlapping with the aromatic signal and the the CH_2NH as a triplet at 8.23 ppm.

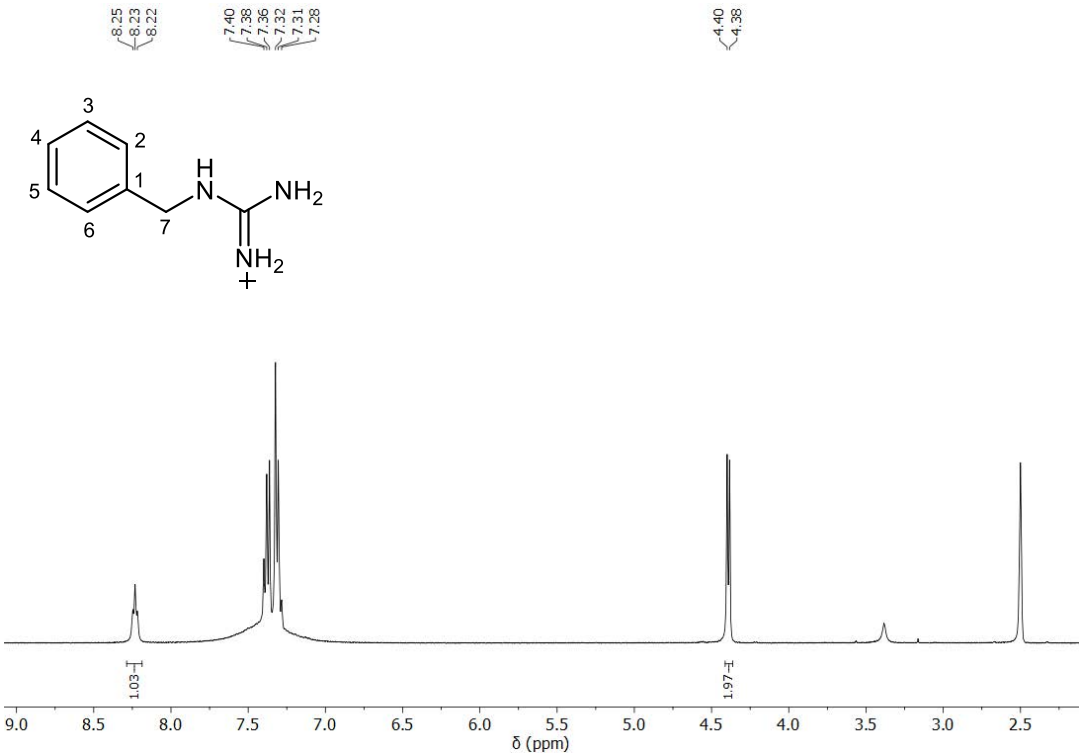


Figure 3.36. ^{1}H NMR spectrum of **B13** (400 MHz, DMSO-d_6, 298 K).

3.4 ITC binding studies of positively charged benzoboroxole receptors

3.4.1 Binding affinities

The binding affinity of synthesised benzoboroxole receptors **B2** – **B7** to sialic acid was evaluated by ITC at four different pH values of 5.5 (0.1 M acetate buffer), 6.5 (0.1 M phosphate buffer), 7.4 (0.1 M phosphate buffer) and 8.5 (0.1 M ammonium acetate buffer). All experiments reported in Table 3.6 were repeated three times with the heat of dilution subtracted. The values reported in Table 3.6 were calculated as the mean of the three different measurements with the error corresponding to twice the standard deviation.

Table 3.6. Binding constants (K_a, M^{-1}) measure by ITC by titrating a 80 mM solution of **S1** into a 2 mM solution of receptors **B2** – **B7** at four different pH values; 5.5 in 0.1 M acetate buffer, 6.5 in 0.1 M PB buffer, 7.4 in 0.1 M PB buffer and, 8.5 in 0.1 M ammonium acetate. The experiments consist in three titrations and were conducted at 25 °C. The heat of dilution was measured and subtracted.

Receptor	pH 5.5	pH 6.5	pH 7.4	pH 8.5
B1	51.2 ± 1.2	50.7 ± 0.9	39.3 ± 0.6	23.5 ± 2.1
B2	150.4 ± 7.9	64.7 ± 3.7	26.8 ± 2.1	21.9 ± 1.4
B3	234.3 ± 8.0	105.0 ± 7.3	43.1 ± 4.8	37.5 ± 1.0
B4	129.6 ± 0.8	84.7 ± 3.8	43.7 ± 0.9	28.1 ± 1.9
B5	141.0 ± 6.8	112.0 ± 5.5	55.5 ± 1.0	33.7 ± 1.6
B6	104.4 ± 5.2	85.9 ± 4.0	50.5 ± 2.1	36.3 ± 1.4
B7	110.2 ± 3.4	82.8 ± 3.9	60.9 ± 4.3	41.4 ± 0.6

All the functionalised receptors **B2** – **B7** displayed their highest affinity at pH 5.5 with the lowest binding occurring at pH 8.5 (Figure 3.37). This behaviour is consistent with the binding profile of non-functionalised **B1**, as the main binding event is the formation of the boronate ester. The formation of the boronate ester is enhanced at acidic pH, as no repulsion is present between the trigonal boron centre and the α-hydroxyacid binding site. Furthermore, when compared to **B1**, all functionalised receptors **B2** – **B7** presented increased binding affinities across the studied pH range (Figure 3.37), with the exception of **B2** at pH ≥ 7.4. The affinity

enhancement was ascribed to the additional interactions that **B2** – **B7** created with sialic acid. Indeed, the receptors **B2** – **B7** bind to sialic acid via multipoint interactions comprising of covalent boronate ester formation and non-covalent interactions. The non-covalent interactions provided by **B2** – **B7** consists primarily of hydrogen bonding and CH-π interactions with the charged and aromatic groups, respectively. The formation of non-covalent interactions within the complex provide stabilisation of the boronate ester resulting in an increased affinity.

The highest binding affinity of receptors **B2** – **B7** was observed at acidic pH, with the non-covalent interactions taking place at this pH providing the greatest stabilisation to the complex. On the other hand, at basic pH the affinity is reduced partially due to the lower proportion of boronate ester formed at pH ≥ 7.4, but also to weaker non-covalent interactions being formed. This effect can be distinctly observed when considering receptors **B2** and **B3**, which present only one additional point of interaction, the charged moiety. The receptors **B2** and **B3** showed the strongest binding to sialic acid at pH 5.5. Lacking any other point of interaction, such as amide or aromatic moieties, it is reasonable to ascribe the large increase in the affinity exclusively to the charged moiety. The electrostatically charged head creates charge-reinforced hydrogen bonds with the hydrogen bond acceptors of sialic acid[20]. Charge-reinforced hydrogen bonds are known to provide a higher contribution to the binding energy than their neutral counterparts[66]. For this reason, the strongest binding was observed at pH 5.5 where the contribution of charge-reinforced hydrogen bonds is especially significant. At basic pH the contribution of charge-reinforced hydrogen bonds is reduced, with neutral hydrogen bonds prevailing, however these are weaker bonds leading to lower binding affinities. In addition to hydrogen bonding, the cationic head can also promote ion pairing with the negatively charged boron centre of the boronate ester complex providing further stabilisation.

Ion pairing was shown to contribute to the stability of the complex in other positively charged boron-based receptors[67].

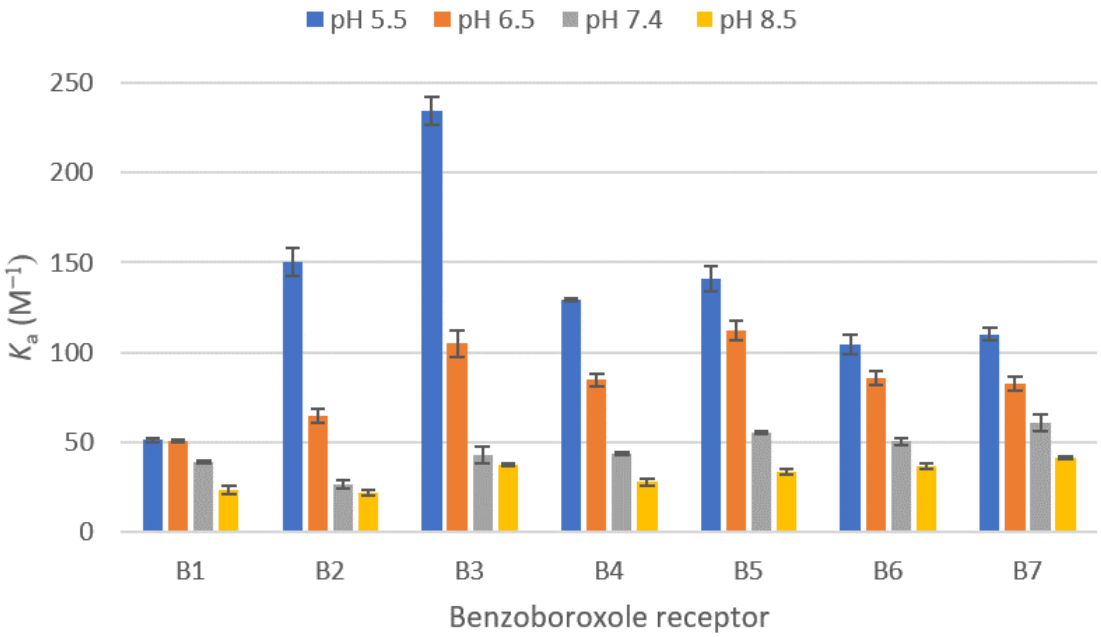


Figure 3.37. Isothermal titration calorimetry binding studies of receptors **B1** and **B2** – **B7** with sialic acid (**S1**) at pH 5.5 (0.1 M acetate buffer), 6.5 (0.1 M phosphate buffer), 7.4 (0.1 M phosphate buffer), and 8.5 (0.1 M ammonium acetate buffer) and 25 °C. Each experiment consists of three titrations. The heat of dilution was measured and subtracted.

Furthermore, the nature of the charged head significantly influenced the affinity. Receptors bearing a guanidino head presented higher K_a than their corresponding amino derivatives across the entire pH range, excluding aromatic derivatives **B6** and **B7** at pH 6.5. It is proposed that the guanidino moiety creates wider networks of charge-reinforced hydrogen bonds with sialic acid acceptors, thus resulting in more stable boronate ester complexes. This is particularly clear at pH 5.5 with the receptor **B3** ($K_a = 234.3 \pm 8.0\ M^{-1}$) presenting a 1.6-fold increase in the affinity compared to the corresponding amino receptor **B2** ($K_a = 150.4 \pm 7.9\ M^{-1}$). In addition, the guanidino group can create stronger ion pairs with anionic groups, when compared to the amino moiety[68].

Another parameter to consider is the distance between the boron-unit and the charged group and how this influences the binding affinity. This is particularly important at acidic pH where the charge head has an influential role in the binding. Conversely, at basic pH other types of non-covalent interaction (*i.e.* CH-π interactions and neutral hydrogen bonds) provide a greater contribution to the complex stabilisation. Thus, at basic pH the relative position of the charged head in relation to the boron centre is less relevant. At pH 5.5 all synthesised receptors **B2** – **B7** showed an increased affinity compared to **B1**. In particular, **B2** and **B3** presented a 3- and 4.5-fold affinity increase due to charge-reinforced hydrogen bonds. The increase is more marked for **B3** as this can create a wider network of interactions. Given the large contribution provided by the charged head at this pH, it can be noticed that when this is further away from the benzoboroxole ring the binding affinity is reduced. Considering amino-functionalised benzoboroxoles **B2**, **B4** and **B6** it can be seen that the larger derivatives present decreased binding affinities. Receptor **B2** presented a binding constant of 150.4 ± 7.9 M^{-1}, whilst **B4** and **B6** have lower K_a of 129.6 ± 0.8 M^{-1} and 104.4 ± 5.2 M^{-1}, respectively. This same trend was also observed for guanidino derivatives. Receptor **B3** presented the highest affinity to sialic acid amongst all receptors, with K_a = 234.3 ± 8.0 M^{-1}. Larger guanidino derivatives, **B5** and **B7**, have lower binding affinities with K_a = 141.0 ± 6.8 M^{-1} and K_a = 110.2 ± 3.4 M^{-1}, respectively. Despite the lower affinity of larger derivatives, the contribution to the binding offered by the guanidino moiety can still be observed for **B5** and **B7** with these presenting higher affinity to sialic acid than their corresponding amino derivatives **B4** and **B6**. Nonetheless, for these derivatives the enhancement caused by the substitution of the amino with a guanidino group is modest, when compared to **B2** and **B3**. Given the lower affinity of larger derivatives, it is reasonable to assume that the receptors **B2** and **B3** created charge-reinforced hydrogen bonds with groups in proximity to the boronate ester or with the boronate ester itself.

In the case of **B4 – B7** the cationic group is too far away from the boronate ester to provide comparable non-covalent interactions to the ones formed by **B2** and **B3**, hence a lower degree of stabilisation is provided. In addition to the distance, also the constraint to the flexibility of the side chain, provided by the amide and phenyl groups, could be responsible for the lower affinities of these derivatives. This could be particularly relevant for **B6** and **B7** which contain a phenyl ring in the side chain, reducing its flexibility. Indeed, these derivatives present the lowest binding constants at acidic pH. Nevertheless, **B6** and **B7** receptors still display a 2-fold increase in the binding affinity when compared to **B1**, suggesting that other non-covalent interactions (*e.g.* CH-π interactions and neutral hydrogen bonds), although weaker, are formed. Moreover, in derivatives synthesised via amide coupling, **B4 - B7**, the amide group can form hydrogen bonds with sialic acid providing and additional point of interaction. However, the effect of this group in the affinity is believed to be small due to the low binding energy provided by neutral hydrogen bonds[66].

At pH > 5.5 the binding affinities of the synthesised receptors decreased steadily, with increasing pH. The greatest reduction was observed for **B2** and **B3**. This suggests that at a more basic pH the contribution of charge-reinforced hydrogen bonds and ion pairing on the binding is reduced. On the other hand, derivatives with additional points of interaction present higher binding affinities as they offer a more diverse range of interactions. Indeed, the contribution of non-covalent interactions (*e.g.* CH-π and neutral H bonds) in the binding is more relevant at pH > 5.5.

At pH 6.5 receptor **B2** presents the lowest binding constant (K_a = 64.7 ± 3.7 M^{-1}), among the synthesised receptors, with a binding affinity 2.3 times lower compared to pH 5.5. This indicates that the contribution of the charge-reinforced hydrogen bonds is reduced, as these are partially substituted by weaker neutral hydrogen bonds. The stabilisation provided by the ion

pairing with the boron centre could also be reduced at this pH. At pH 6.5 the receptor **B3** still displayed higher affinity than **B2**, with K_a = 105.0 ± 7.3 M^{-1}, confirming that importance of the wider network of hydrogen bonds in the stabilisation of the boronate ester. Moreover, at pH 6.5 the highest binding affinity for sialic acid is achieved by receptor **B5**, with K_a = 112.0 ± 5.5 M^{-1}. This suggests that there is an increased contribution of additional points of interaction to the boronate ester stabilisation. Indeed, receptors **B4** and **B5** can provide additional neutral hydrogen bonds through the amide group. Moreover, receptors **B6** and **B7** can also form CH-π interactions with the apolar face of sialic acid. On the other hand, **B2** and **B3** do not contain other functional groups, aside from the charged moiety, resulting in fewer points of interaction with sialic acid. In addition to these considerations, the differences in the affinities between pH 5.5 and 6.5 cannot neglect the influence of the different buffer solutions (*e.g.* sodium acetate, phosphate buffer)[9]. Acetate buffer, used for pH 5.5, has been reported to accelerate the reaction between boron-based receptors and a ligand via addition of the acetate ion to the boron[69]. Thus, the lower affinities at pH 6.5, compared to pH 5.5, could also be partially attributed to the different buffer employed as the boronate ester formation in phosphate buffer (pH 6.5) cannot take advantage of the support provided by the acetate ions. As different buffers are used for different pHs, the direct comparison of data from different pH values must be interpreted cautiously. Nevertheless, at a given pH, the comparison of K_a values for the different functionalised receptors is valid.

At physiological pH, the binding affinity of all receptors **B2** – **B7** is greatly reduced. This is mainly determined by the lower affinity of the tetrahedral boron to the α-hydroxyacid caused by electrostatic repulsion (see Section 3.2.3). Furthermore, at neutral pH the contribution of charge-reinforced hydrogen bonds and ion pairing is further reduced. For instance, **B2** with K_a = 26.8 ± 2.1 M^{-1} presents lower affinity to sialic acid than non-functionalised **B1**

(K_a = 39.3 ± 0.6 M^{-1}), suggesting that the amino group plays no role in the binding at pH 7.4. The affinity of receptor **B3** is greatly reduced, when compared to acidic pH, however it still presents a modest increase compared to **B1**, with K_a = 43.1 ± 4.8 M^{-1}. This finding is in contradiction to that of the amino functionalised receptors, suggesting that there is a small contribution of the guanidino group in the binding, confirming the positive effect provided by hydrogen bonding network, even when in absence of charge-reinforced hydrogen bonds. Furthermore, receptors **B4** and **B5** display binding constants of 43.7 ± 0.9 M^{-1} and 55.5 ± 1.0 M^{-1}, respectively. Whilst the aromatic receptors **B6** (K_a = 50.5 ± 2.1 M^{-1}) and **B7** (K_a = 60.9 ± 4.3 M^{-1}) have the highest binding affinity amongst functionalised receptors, respectively. The binding of these receptors to sialic acid is reinforced by CH-π interactions between the side chain phenyl ring and the sugar carbon backbone. The observations reported for pH 7.4 are also valid for pH 8.5 as the trend is the similar, though the binding constants are further reduced. At pH 8.5 the highest binding constant is shown by **B7** (K_a = 41.4 ± 0.6 M^{-1}), confirming the relevance of CH-π interactions at neutral and basic pH.

Furthermore, the chirality of the receptors was examined. The functionalised benzoboroxole receptors **B2** – **B7** presented a chiral centre in position 7 of the oxaborole ring. The benzoboroxole compounds were synthesised and studied as racemic mixtures. However, the two enantiomers of a given receptor could theoretically display significantly different binding affinities to sialic acid. Furthermore, due to the chirality of the boron atom in the complexed species the receptor racemic mixture generates a mixture of four diastereomeric complexes which could present different stabilities and binding energies. The ITC data can be utilised to gauge whether the racemic mixture is formed by two enantiomers with significantly different binding affinities to sialic acid[70]. For the case in which two enantiomers exist with significantly different affinities the shape of the isotherm would be composed by two curves, which

manifests itself as a step in the ITC isotherm. On the other hand, when the two enantiomers have similar binding affinities the ITC graph is composed by one single curve, not presenting any step. Herein, all ITC curves comprised of a single curve, thus both enantiomers of receptors **B2** – **B7** present comparable binding constants to the racemic mixture.

3.4.2 Enthalpy-entropy compensation

The ITC experimental data can be used to extract other valuable thermodynamic data in relation to the binding event, in addition to the binding constant K_a. Statistical analysis of fits to the titration curve allow extraction of two other important thermodynamic parameters; the change in enthalpy (ΔH, kcal mol^{-1}) and the change in entropy (ΔS, cal mol^{-1} deg^{-1}). The Gibbs free energy (ΔG, kcal mol^{-1}) of the system was then calculated through application of Equation 3.5 with temperature T being 298.15 K and R being 1.98720×10−3 kcal K^{-1} mol^{-1}.

$$\Delta G = -RT \ln K_a = \Delta H - T\Delta S \qquad \text{Equation 3.5}$$

Table 3.7. Thermodynamic parameters ΔH (kcal mol^{-1}), ΔS (cal mol^{-1} K^{-1}), ΔG (kcal mol^{-1}) for the binding of receptors **B1**, **B2** – **B7** to **S1** at pH 5.5, 6.5, 7.4 and 8.5 at 25°C (298.15 K).

	pH	**B1**	**B2**	**B3**	**B4**	**B5**	**B6**	**B7**
ΔH	5.5	−5.6 ± 0.1	−3.3 ± 0.1	−3.5 ± 0.1	−4.9 ± 0.0	−4.6 ± 0.2	−3.9 ± 0.2	−3.3 ± 0.1
	6.5	−5.1 ± 0.1	−2.7 ± 0.1	−3.8 ± 0.3	−4.0 ± 0.0	−3.9 ± 0.1	−3.6 ± 0.1	−2.9 ± 0.1
	7.4	−4.7 ± 0.1	−2.9 ± 0.0	−2.8 ± 0.3	−3.7 ± 0.1	−2.4 ± 0.0	−3.4 ± 0.1	−2.9 ± 0.1
	8.5	−3.5 ± 0.2	−2.6 ± 0.1	−2.3 ± 0.2	−3.2 ± 0.1	−3.4 ± 0.1	−3.3 ± 0.1	−2.8 ± 0.1
ΔS	5.5	−10.9 ± 0.4	−1.2 ± 0.3	−1.0 ± 0.2	−6.6 ± 0.2	−5.8 ± 0.8	−3.7 ± 0.4	−1.8 ± 0.3
	6.5	−9.3 ± 0.4	−0.7 ± 0.4	−3.4 ± 1.0	−4.7 ± 0.2	−3.9 ± 0.2	−3.2 ± 0.3	−1.1 ± 0.5
	7.4	−8.3 ± 0.5	−3.3 ± 0.2	−2.0 ± 1.1	−4.8 ± 0.4	−2.6 ± 0.1	−3.6 ± 0.5	−1.7 ± 0.5
	8.5	−5.3 ± 1.0	−2.7 ± 0.5	−0.5 ± 0.9	−4.1 ± 0.6	−4.3 ± 0.4	−4.0 ± 0.4	−2.0 ± 0.2
ΔG	5.5	−2.3 ± 0.2	−3.0 ± 0.2	−3.2 ± 0.1	−2.9 ± 0.1	−2.9 ± 0.4	−2.8 ± 0.2	−2.8 ± 0.2
	6.5	−2.3 ± 0.3	−2.5 ± 0.2	−2.8 ± 0.6	−2.6 ± 0.1	−2.8 ± 0.2	−2.6 ± 0.1	−2.6 ± 0.3
	7.4	−2.2 ± 0.3	−1.9 ± 0.1	−2.2 ± 0.6	−2.2 ± 0.2	−2.4 ± 0.8	−2.3 ± 0.3	−2.4 ± 0.3
	8.5	−1.9 ± 0.5	−1.8 ± 0.3	−2.1 ± 0.5	−2.0 ± 0.3	−2.1 ± 0.2	−2.1 ± 0.2	−2.2 ± 0.1

The binding event (*i.e.* the formation of the boronate ester) promotes the release of heat and, thus a negative change in the enthalpy of the system, as observed in Table 3.7. This was observed for all the ITC experiments presented within this study. The negative change in the enthalpy associated to the binding event indicates that the formation of the boronate ester is favourable and enthalpic driven. It must be noted that the change in the enthalpy of the system consist of a combination of various phenomena occurring during the binding event[71], and it is not only ascribable to the boronate ester formation. The overall change in enthalpy is given by a combination of interactions that are being broken, between the unbound species in solution and the solvent, and others that are being formed, between receptor and ligand. In addition, different buffers, with different ionisation enthalpies, provide different reaction enthalpy changes. The obtained ΔH value is therefore a sum of the enthalpy of binding and the enthalpy of ionisation of the buffer[72]. Therefore, the enthalpy of system evaluated at different pH cannot be directly compared. Moreover, the interpretation of the enthalpy of systems with very low *c*-values, such as the ones studied herein (c = 0.04 - 0.46), must be performed with extra care. As described in Section 3.1, when the *c*-value is < 1 the number of binding site *n* must be fixed prior fitting of the data. However, in such cases the fitted ΔH values are reliable only if the following conditions are met; high molar ratio, appropriate levels of signal to noise and the concentration of ligand and receptors must be known with accuracy[30]. The experiments were designed to meet all these conditions (see Section 3.1). Thus, the experimental ΔH values are believed to be accurate and can be used for the thermodynamic interpretation of the binding event.

The entropic component in Table 3.7 was negative for all experiments. The change of entropy (ΔS) represents the disorder of the system and is a combination of solvent release and changes in the conformational, translational and rotational freedom of the species upon binding. In

particular, the formation of the boronate ester caused a loss of degrees of freedom of the species in solution which resulted in an overall negative change in the entropy. The negative entropy indicates that the binding is not favourable from the entropic point of view. The Gibbs free energy (Table 3.7) is negative for all experiments indicating that the formation of the boronate ester, between receptor **B1** – **B7** and sialic acid, is a spontaneous event. Furthermore, in system where covalent bonds are formed the ΔG is dominated by the enthalpic component[73], as observed in these studies.

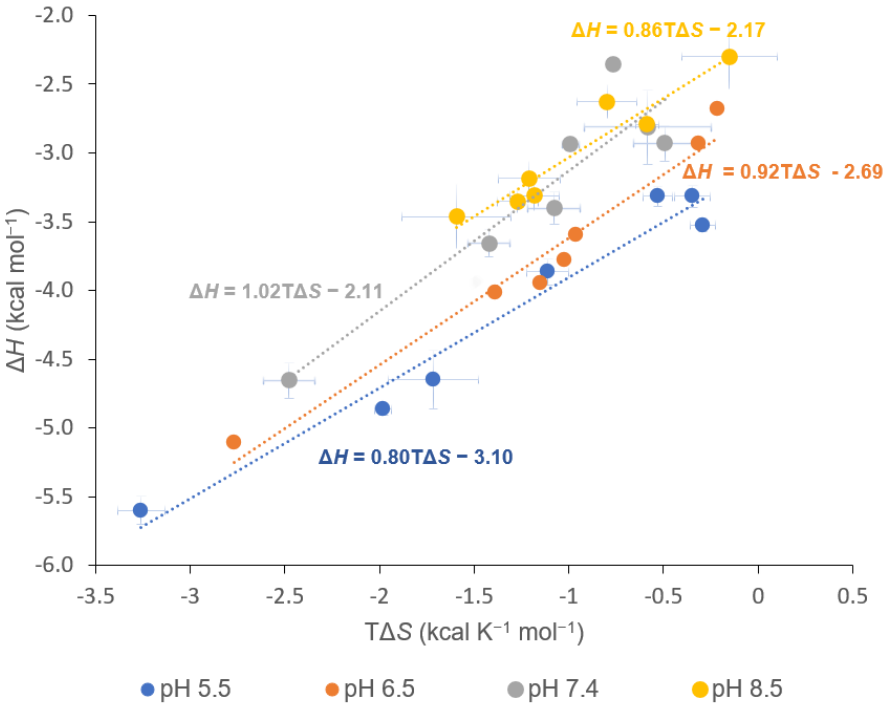


Figure 3.38. Enthalpy (ΔH, kcal mol^{-1}) against entropy ($T\Delta S$, kcal K^{-1} mol^{-1}) plot at pH 5.5 (0.1 M acetate buffer), 6.5 (0.1 M phosphate), 7.4 (0.1 M phosphate buffer), and 8.5 (0.1 M ammonium acetate buffer) and 25 °C.

The complexation causes a favourable negative change in the enthalpy, whilst simultaneously causing an unfavourable negative change of the entropy. Thus, the enthalpic component is compensated by the entropic one with this resulting in a moderate change in the free energy

(from −1.8 to −3.2 to kcal mol^{-1}) [71,74]. Furthermore, the ΔG is linked to the binding constant K_a by Equation 3.5. Greater binding constants correspond to more negative values of ΔG therefore in order to achieve higher affinities the negative change in ΔG should be larger. However, in the boronate ester formation, as in many other systems[73], greater changes in the ΔG are prevented by the compensation between enthalpy and entropy. This effect is called enthalpy-entropy compensation (EEC). The EEC effect is the linear relationship existing between the enthalpic component ΔH and the entropic component $T\Delta S$. The ΔH data were plotted against the $T\Delta S$ values for receptors **B1** – **B7** at a selected pH. A linear dependency was found for all four pH values (Figure 3.38), with the slope of the linear dependency indicating the degree of compensation. For slopes of 1 the EEC is complete, as observed at pH 7.4. Slopes > 1 indicate that the binding is driven by entropy and is more sensitive to the enthalpic component[75], whilst slopes below 1 indicate that the binding is enthalpic driven and more sensitive to changes in the entropy of the system[27, 74]. The latter condition (slopes < 1) was observed at pH 5.5, 6.5 and 8.5 with slopes of 0.80, 0.92 and 0.86, respectively. The EEC effect was initially reported as a property of the solvent, particularly water, and to be independent from the species in solution[73]. Further studies described the EEC effect as a characteristic of all systems based on weak interactions [73]. The EEC has been extensively observed for protein-ligand interactions[76] and was also shown for boronic acid-based receptors[27]. The analysis of the thermodynamic data showed that the boronate ester formation between benzoboroxole receptors **B1** – **B7** and sialic acid is enthalpy driven and subject to enthalpy-entropy compensation.

3.4.3 Cooperativity of the binding event

The binding of functionalised-benzoboroxole receptors to sialic acid consists of a combination of covalent and non-covalent interactions and the cooperative nature of these interactions was investigated. As multiple binding events occur in the binding process these can be distinguish in a primary and a secondary event. Herein, the formation of the boronate ester is defined as the primary binding event, whilst the sum of the non-covalent interactions (*e.g.* H-bonds, CH-π interactions, ion pairing) are defined as the secondary binding event. The two binding events could be independent or dependent with respect to each other. If the events are independent the secondary binding event would occur regardless of the formation of the boronate ester. Alternatively, one event can be dependent on the other, hence the non-covalent interaction will occur only as a consequence of boronate esterification. In particular, the aim was to assess whether the formation of non-covalent interactions is dependent on the boronate ester formation.

A control molecule **B13** was synthesised (see Section 3.3.3) and the binding of this to sialic acid was measured by ITC. The control molecule **B13** contains solely an aromatic ring functionalised with a methyl guanidino group. The guanidino group was chosen, rather than the amino, due to its greater effect on the binding affinity. The compound **B13** can interact with sialic acid solely through non-covalent interactions as it lacks the boron-unit. The cationic and aromatic moieties can offer charge-reinforced hydrogen bonds and CH-π interactions. In addition, the guanidino could form electrostatic interactions with the carboxylate as this will not be engaged in a boronate ester. The non-covalent interactions which provided the highest contribution to the binding are charge-reinforced hydrogen bonds and these are particularly relevant at pH 5.5. Therefore, the binding of **B13** to **S1** was measured at pH 5.5. If the two binding events, boronate esterification and non-covalent interactions, were independent **B13**

would bind **S1** through non-covalent interactions despite the absence of the boronate ester. The formation of non-covalent interactions would be confirmed by significant, albeit weaker, binding being observed by ITC. On the other hand, if the secondary binding event can only occur exclusively following the covalent complexation no significant binding would be present for **B13**. Experimental ITC data showed that there was not a significant change in heat detected when titrating an 80 mM solution of sialic acid into a 2 mM solution of **B13**, with $K_a < 1.0\ M^{-1}$. Hence, the secondary binding event appears dependent on the primary one with the non-covalent interactions occurring solely after the boronate ester formation. Thus, the binding of the positively charged receptors **B2** – **B7** to sialic acid was found to be cooperative.

3.4.4 Selectivity studies

In order to assess the selectivity of positively charged benzoboroxoles towards sialic acid the binding affinity to neutral (*e.g.* fructose, galactose, glucose and mannose) and negatively charged (glucuronic acid) monosaccharides (Figure 3.39) was assessed.

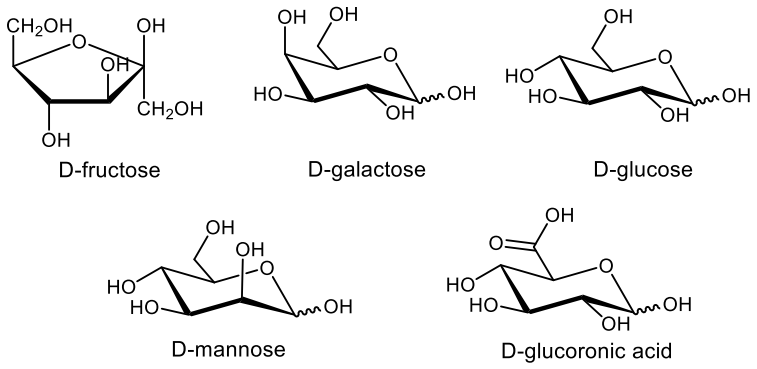


Figure 3.39. Neutral and anionic monosaccharides.

The binding of boron-based receptors to neutral monosaccharides occurs preferentially for pH values above the pK_a (pK_a = 7.2 for **B1**)[13]. Therefore, at physiological and basic pH the binding arises between the tetrahedral boron and the sugar diols[9]. For instance, the binding of benzoboroxole **B1** to fructose at physiological pH is reported to be K_a = 508.0 ± 7.0 M^{-1} [26], via the 2,3 diol[77], whilst with glucose the binding is with the *trans*-4,6 diol (K_a = 31 M^{-1})[13]. The binding affinity to galactose varies between the α and β anomeric configurations, with K_a in the 20-30 M^{-1} range, and occurs with the 3,4 *cis*-diol or with the 4,6 *cis*-diol, with the former being the preferred [13]. Finally, the K_a for mannose is K_a = 24 M^{-1}.[13]

The selectivity studies were conducted by ITC at pH 5.5, as at this pH the functionalised receptor presents the highest binding affinity to sialic acid (Table 3.8).

Table 3.8. Binding constants (K_a, M^{-1}) measure by ITC by titrating a 80 mM solution of D-fructose, D-galactose, D-glucose, D-mannose or D-glucuronic acid into a 2 mM solution of **B1**, **B2** or **B3** at pH 5.5 in 0.1 M acetate buffer, The experiments consist in three titrations and were conducted at 25 °C. The heat of dilution was measured and subtracted.

	B1	**B2**	**B3**
S1	51.2 ± 1.2	150.4 ± 7.3	234.3 ± 8.0
D-Fructose	18.5 ± 0.6	198.6 ± 16.1	162.6 ± 12.6
D-Galactose	7.5 ± 0.4	10.6 ± 4.1	5.6 ± 2.4
D-Glucose	< 1.0	10.1 ± 4.4	10.0 ± 0.9
D-Mannose	< 1.0	9.6 ± 4.2	4.8 ± 1.1
D-Glucuronic acid	< 1.0	11.6 ± 2.2	12.5 ± 0.3

Three receptors were selected for the study: **B1**, **B2** and **B3**. The non-functionalised benzoboroxole **B1** was employed to assess the sole contribution of the boronate ester formation. Whilst cationic benzoboroxole receptors **B2** and **B3** were chosen as they offer the highest affinity to sialic acid. The binding of benzoboroxole receptors **B1**, **B2** and **B3** to the aforementioned neutral monosaccharides at acidic pH was expected to be low to negligible.

Indeed, the binding affinity of **B1** to fructose and galactose was greatly reduced, with $K_a = 18.5 \pm 0.6$ M^{-1} and $K_a = 7.5 \pm 0.4$ M^{-1}, respectively, whilst the binding to glucose and mannose was too low to be detected ($K_a < 1.0$ M^{-1}). The presence of the amino group in **B2** caused an increase in the affinity of the benzoboroxole to the neutral monosaccharides. This increase led to modest binding constants for galactose, glucose and mannose, whilst the affinity to fructose saw a 10-fold increase. The binding affinity of **B2** to fructose ($K_a = 198.6 \pm 16.1$ M^{-1}) is higher than the binding affinity to sialic acid ($K_a = 150.4 \pm 7.3$ M^{-1}). This indicated that positively charged groups, through charge-reinforced hydrogen bonds, can enhance the affinity also for other monosaccharides. Hence, charge-reinforced hydrogen bonds have the potential to be highly valuable tools in the design of sugar probes. Interestingly, when comparing the amino to the guanidino, the latter functionality caused no further enhancement in the binding affinity for glucose ($K_a = 10.0 \pm 0.9$ M^{-1}) whilst it displayed a reduced affinity for fructose ($K_a = 162.6 \pm 12.6$ M^{-1}), galactose and mannose. Consequently, **B3** is observed to selectively bind sialic acid amongst the neutral monosaccharides.

In addition to the aforementioned monosaccharides, the binding affinity to glucuronic acid was also measured. Glucuronic acid is a glucose derivative with the CH_2OH group in position 6 oxidised to a carboxylic acid. Thus, it is an anionic sugar and was employed to verify the cross reactivity with anionic species. It is important to note that the carboxylic acid of glucuronic acid is not a α-hydroxyacid as no OH group is present in the α position; thus, this cannot be bound by the boron unit. In addition, the oxidation of position 6 renders the 4,6 diol, the preferred binding site, unavailable. Hence, only the less favourable *trans*-3,4 diol is accessible[13]. The binding to glucuronic acid is expected to be consistent or lower than the binding to glucose, unless the interaction of the positively charged moiety with the carboxylic acid is significant. As reported in Table 3.8 all three receptors **B1**, **B2** and **B3** present negligible or low binding

constants to glucuronic acid, which are consistent with the K_a values observed for glucose. Hence, it is reasonable to conclude that no significant interaction occurs between the cationic moieties and the carboxylic acid, as already shown by the cooperative studies (see Section 3.4.3).

There was no evidence of cross reactivity of the positively charged receptor **B2** and **B3** with galactose, glucose, mannose and the negatively charged glucuronic acid. Whilst receptor **B2** was found to preferentially bind fructose at pH 5.5, suggesting a lack of selectivity for sialic acid. Conversely, **B3** has been shown to be selective for sialic acid under acidic conditions.

3.5 Summary and outlook

Initially the binding affinity of non-functionalised receptor **B1** to sialic acid was measured by ITC confirming that the highest affinity could be achieved under acidic conditions, due to lack of repulsion between ligand and receptor. The formation of the boronate ester was proven to arise exclusively with the α-hydroxyacid group and no binding with the glycerol chain was detected at any pH value. Following these studies, functionalised benzoboroxole receptors bearing cationic groups were synthesised and their affinity to sialic acid was assessed by ITC. The highest binding was recorded at pH 5.5. At this pH the neutral, trigonal boron reacts with the negatively charged α-hydroxyacid group to form the boronate ester complex with elimination of one equivalent of water. Furthermore, functionalised receptors **B2** – **B7** provide additional interactions with sialic acid resulting in stabilisation of the reversible boronate ester and, thus enhanced binding affinities. Particularly relevant at acidic pH are charge-reinforced hydrogen bonds and the ionic pairing. The charge-reinforced hydrogen bonds are formed between the amino or guanidino groups and the polar groups of sialic acid[20]. Ionic pairing

occurs between the cationic group and the negative boron centre in the complex. Hence, the binding of functionalised benzoboroxole to sialic acid is characterised by multivalent interactions and a cooperative nature. Indeed, the formation of the boronate ester is essential for the non-covalent interactions to occur and in the absence of the benzoboroxole unit no binding was detected. Furthermore, it is reasonable to assume that the charge-reinforced hydrogen bonds occur in close proximity to the boron centre as the smaller derivatives **B2** and **B3** present greater binding affinities. In particular, **B3** presents the highest binding affinity among synthesised benzoboroxoles, with $K_a = 234.3 \pm 8.0\ M^{-1}$. The guanidino moiety enhances the binding by providing a wider network of hydrogen bonds which further stabilise the complex. For receptors **B4 – B7**, which present longer side chains, the affinity is decreased at pH 5.5 as they cannot provide comparable interactions and stabilisation to **B2** and **B3**. Conversely, at more basic pH, receptors **B4 – B7** presented higher affinity than **B2** and **B3**. This is due to the charge-reinforced hydrogen bonds being less relevant at higher pH values with other interactions prevailing. Indeed, receptors **B4 – B7** present additional functional groups (*e.g.* amide and phenyl groups) offering additional points of interaction with sialic acid. Nevertheless, neutral hydrogen bonds and CH-π interaction interactions provide a lower contribution to the binding. This, combined with the repulsion at basic pH, results in an overall progressive decrease of the affinity at more basic pH values.

It was concluded that the optimal pH for the detection of sialic acid by boron-based receptors is 5.5. At this pH the benzoboroxole receptors **B1** and **B3** have shown to be selective towards sialic acid versus other monosaccharides, with the exception of **B2** binding preferentially fructose. Consequently, the best receptor for sialic acid was determined to be **B3** as it presents the highest binding affinity to sialic acid ($K_a = 234.3 \pm 8.0$) and good selectivity.

List of references

(1) Lagana, A.; PardoMartinez, B.; Marino, A.; Fago, G.; Bizzarri, M., Determination of serum total lipid and free N-acetylneuraminic acid in genitourinary malignancies by fluorimetric high performance liquid chromatography. Relevance of free N-acetylneuraminic acid as tumour marker. *Clin. Chim. Acta* **1995,** *243* (2), 165-179.

(2) Haga, Y.; Uemura, M.; Baba, S.; Inamura, K.; Takeuchi, K.; Nonomura, N.; Ueda, K., Identification of Multisialylated LacdiNAc Structures as Highly Prostate Cancer Specific Glycan Signatures on PSA. *Anal. Chem.* **2019,** *91* (3), 2247-2254.

(3) Renlund, M.; Aula, P.; Raivio, K. O.; Autio, S.; Sainio, K.; Rapola, J.; Koskela, S. L., Salla disease: a new lysosomal storage disorder with disturbed sialic acid metabolism. *Neurology* **1983,** *33* (1), 57-66.

(4) Paschke, E.; Trinkl, G.; Erwa, W.; Pavelka, M.; Mutz, I.; Roscher, A., Infantile type of sialic acid storage disease with sialuria. *Clin. Genet.* **1986,** *29* (5), 417-424.

(5) Chrostek, L.; Cylwik, B.; Korcz, W.; Krawiec, A.; Koput, A.; Supronowicz, Z.; Szmitkowski, M., Serum free sialic acid as a marker of alcohol abuse. *Alcoholism (NY)* **2007,** *31* (6), 996-1001.

(6) Bhavanandan, V. P.; Katlic, A. W., The interaction of wheat germ agglutinin with sialoglycoproteins. The role of sialic acid. *J. Biol. Chem.* **1979,** *254* (10), 4000-4008.

(7) Wu, X.; Li, Z.; Chen, X. X.; Fossey, J. S.; James, T. D.; Jiang, Y. B., Selective sensing of saccharides using simple boronic acids and their aggregates. *Chem. Soc. Rev.* **2013,** *42* (20), 8032-8048.

(8) Peters, J. A., Interactions between boric acid derivatives and saccharides in aqueous media: Structures and stabilities of resulting esters. *Coord. Chem. Rev.* **2014,** *268*, 1-22.

(9) Springsteen, G.; Wang, B. H., A detailed examination of boronic acid-diol complexation. *Tetrahedron* **2002,** *58* (26), 5291-5300.

(10) Nakatani, H.; Hiromi, K., Binding of m-nitrobenzeneboronic acid to the active site of subtilisin BPN′. *Biochim. Biophys. Acta* **1978,** *524* (2), 413-417.

(11) Lorand, J. P.; Edwards, J. O., Polyol Complexes and Structure of the Benzeneboronate Ion. *J. Org. Chem.* **1959,** *24* (6), 769-774.

(12) Dowlut, M.; Hall, D. G., An improved class of sugar-binding boronic acids, soluble and capable of complexing glycosides in neutral water. *J. Am. Chem. Soc.* **2006,** *128* (13), 4226-4227.

(13) Berube, M.; Dowlut, M.; Hall, D. G., Benzoboroxoles as efficient glycopyranoside-binding agents in physiological conditions: Structure and selectivity of complex formation. *J. Org. Chem.* **2008,** *73* (17), 6471-6479.

(14) Sorensen, M. D.; Martins, R.; Hindsgaul, O., Assessing the terminal glycosylation of a glycoprotein by the naked eye. *Angew. Chem. Int. Ed.* **2007,** *46* (14), 2403-2407.

(15) Claes, D.; Memmel, E.; Holzapfel, M.; Seibel, J.; Maison, W., High-Affinity Carbohydrate Binding by Trimeric Benzoboroxoles Measured on Carbohydrate Arrays. *ChemBioChem* **2014,** *15* (16), 2450-2457.

(16) Rowe, L.; El Khoury, G.; Lowe, C. R., A benzoboroxole-based affinity ligand for glycoprotein purification at physiological pH. *J. Mol. Recognit.* **2016,** *29* (5), 232-238.

(17) Pal, A.; Berube, M.; Hall, D. G., Design, Synthesis, and Screening of a Library of Peptidyl Bis(Boroxoles) as Oligosaccharide Receptors in Water: Identification of a Receptor for the Tumor Marker TF-Antigen Disaccharide. *Angew. Chem. Int. Ed.* **2010,** *49* (8), 1492-1495.

(18) Tommasone, S.; Tagger, Y. K.; Mendes, P. M., Targeting Oligosaccharides and Glycoconjugates Using Superselective Binding Scaffolds. *Adv. Funct. Mater.* **2020**, 2002298.

(19) Kowalczyk, W.; Sanchez, J.; Kraaz, P.; Hutt, O. E.; Haylock, D. N.; Duggan, P. J., The binding of boronated peptides to low affinity mammalian saccharides. *Pept. Sci.* **2018,** *110* (3), 12.

(20) Di Pasquale, A.; Tommasone, S.; Xu, L.; Ma, J.; Mendes, P. M., Cooperative Multipoint Recognition of Sialic Acid by Benzoboroxole-Based Receptors Bearing Cationic Hydrogen-Bond Donors. *J. Org. Chem.* **2020,** *85* (13), 8330-8338.

(21) Otsuka, H.; Uchimura, E.; Koshino, H.; Okano, T.; Kataoka, K., Anomalous binding profile of phenylboronic acid with N-acetylneuraminic acid (Neu5Ac) in aqueous solution with varying pH. *J. Am. Chem. Soc.* **2003,** *125* (12), 3493-3502.

(22) Nishitani, S.; Maekawa, Y.; Sakata, T., Understanding the Molecular Structure of the Sialic Acid-Phenylboronic Acid Complex by using a Combined NMR Spectroscopy and DFT Study: Toward Sialic Acid Detection at Cell Membranes. *ChemistryOpen* **2018,** *7* (7), 513-519.

(23) Djanashvili, K.; Frullano, L.; Peters, J. A., Molecular recognition of sialic acid end groups by phenylboronates. *Chem. Eur. J.* **2005,** *11* (13), 4010-4018.

(24) Wiseman, T.; Williston, S.; Brandts, J. F.; Lin, L. N., Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **1989,** *179* (1), 131-137.

(25) Doyle, M. L., Characterization of binding interactions by isothermal titration calorimetry. *Curr. Opin. Biotechnol.* **1997,** *8* (1), 31-35.

(26) Schumacher, S.; Katterle, M.; Hettrich, C.; Paulke, B. R.; Hall, D. G.; Scheller, F. W.; Gajovic-Eichelmann, N., Label-free detection of enhanced saccharide binding at pH 7.4 to nanoparticulate benzoboroxole based receptor units. *J. Mol. Recognit.* **2011,** *24* (6), 953-959.

(27) Wiskur, S. L.; Lavigne, J. L.; Metzger, A.; Tobey, S. L.; Lynch, V.; Anslyn, E. V., Thermodynamic analysis of receptors based on guanidinium/boronic acid groups for the

complexation of carboxylates, alpha-hydroxycarboxylates, and diols: Driving force for binding and cooperativity. *Chem. Eur. J.* **2004,** *10* (15), 3792-3804.

(28) Bundle, D. R.; Sigurskjold, B. W., Determination of accurate thermodynamics of binding by titration microcalorimetry. *Neoglycoconjugates, Pt B: Biomedical Applications* **1994,** *247*, 288-305.

(29) Springsteen, G.; Wang, B. H., Alizarin Red S. as a general optical reporter for studying the binding of boronic acids with carbohydrates. *Chem. Commun.* **2001,** (17), 1608-1609.

(30) Turnbull, W. B.; Daranas, A. H., On the value of c: Can low affinity systems be studied by isothermal titration calorimetry? *J. Am. Chem. Soc.* **2003,** *125* (48), 14859-14866.

(31) Tellinghuisen, J., Isothermal titration calorimetry at very low c. *Anal. Biochem.* **2008,** *373* (2), 395-397.

(32) MicroCal, L., *VP-ITC MicroCalorimeter User's Manual, MAU130030 REV. E*. MicroCal, LLC: Northampton, MA.

(33) Regueiro-Figueroa, M.; Djanashvili, K.; Esteban-Gomez, D.; de Blas, A.; Platas-Iglesias, C.; Rodriguez-Blas, T., Towards Selective Recognition of Sialic Acid Through Simultaneous Binding to Its cis-Diol and Carboxylate Functions. *Eur. J. Org. Chem.* **2010,** (17), 3237-3248.

(34) Matsumoto, A.; Stephenson-Brown, A. J.; Khan, T.; Miyazawa, T.; Cabral, H.; Kataoka, K.; Miyahara, Y., Heterocyclic boronic acids display sialic acid selective binding in a hypoxic tumor relevant acidic environment. *Chem. Sci.* **2017,** *8* (9), 6165-6170.

(35) Regueiro-Figueroa, M.; Djanashvili, K.; Esteban-Gomez, D.; Chauvin, T.; Toth, E.; de Blas, A.; Rodriguez-Blas, T.; Platas-Igleslas, C., Molecular Recognition of Sialic Acid by Lanthanide(III) Complexes through Cooperative Two-Site Binding. *Inorg. Chem.* **2010,** *49* (9), 4212-4223.

(36) Yamamoto, M.; Takeuchi, M.; Shinkai, S., Molecular design of a PET-based chemosensor for uronic acids and sialic acids utilizing a cooperative action of boronic acid and metal chelate. *Tetrahedron* **1998,** *54* (13), 3125-3140.

(37) Chaudhary, P. M.; Murthy, R. V.; Yadav, R.; Kikkeri, R., A rationally designed peptidomimetic biosensor for sialic acid on cell surfaces. *Chem. Commun.* **2015,** *51* (38), 8112-8115.

(38) Carter, T. S.; Mooibroek, T. J.; Stewart, P. F. N.; Crump, M. P.; Galan, M. C.; Davis, A. P., Platform Synthetic Lectins for Divalent Carbohydrate Recognition in Water. *Angew. Chem. Int. Ed.* **2016,** *55* (32), 9311-9315.

(39) Godoy-Alcantar, C.; Yatsimirsky, A. K.; Lehn, J. M., Structure-stability correlations for imine formation in aqueous solution. *J. Phys. Org. Chem.* **2005,** *18* (10), 979-985.

(40) Jayeoye, T. J.; Cheewasedtham, W.; Putson, C.; Rujiralai, T., A selective probe based on 3-aminophenyl boronic acid assembly on dithiobis(succinimidylpropionate) functionalized gold nanoparticles for sialic acid detection in human serum. *J. Mol. Liq.* **2019,** *281*, 407-414.

(41) Friedman, S.; Pace, B.; Pizer, R., Complexation of Phenylboronic Acid with Lactic-acid. Stability Constants and Reaction Kinetics. *J. Am. Chem. Soc.* **1974,** *96* (17), 5381-5384.

(42) Duggan, P. J.; Offermann, D. A., Remarkably selective saccharide recognition by solid-supported peptide boronic acids. *Tetrahedron* **2009,** *65* (1), 109-114.

(43) Djanashvili, K.; Koning, G. A.; van der Meer, A.; Wolterbeek, H. T.; Peters, J. A., Phenylboronate Tb-160 complexes for molecular recognition of glycoproteins expressed on tumor cells. *Contrast Media Mol. Imaging* **2007,** *2* (1), 35-41.

(44) Fersht, A. R.; Shi, J. P.; Knilljones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M. Y.; Winter, G., Hydrogen bonding and biological specificity analysed by protein engineering. *Nature* **1985,** *314* (6008), 235-238.

(45) Wright, C. S., 2.2 A-Resolution Structure-Analysis of 2 Refined N-Acetylneuraminyl-Lactose - Wheat-Germ-Agglutinin Isolectin Complexes. *J. Mol. Biol.* **1990,** *215* (4), 635-651.

(46) Geffert, C.; Kuschel, M.; Mazik, M., Molecular Recognition of N-Acetylneuraminic Acid by Acyclic Pyridinium- and Quinolinium-Based Receptors in Aqueous Media: Recognition through Combination of Cationic and Neutral Recognition Sites. *J. Org. Chem.* **2013,** *78* (2), 292-300.

(47) Mazik, M.; Cavga, H., Molecular recognition of N-acetylneuraminic acid with acyclic benzimidazolium- and aminopyridine/guanidinium-based receptors. *J. Org. Chem.* **2007,** *72* (3), 831-838.

(48) Huang, C. Y.; Cabell, L. A.; Lynch, V.; Anslyn, E. V., Intermolecular versus intramolecular hydrogen-bonding competition in the complexation of cyclitols by a twisted polyaza cleft. *J. Am. Chem. Soc.* **1992,** *114* (5), 1900-1901.

(49) Cuntze, J.; Owens, L.; Alcazar, V.; Seiler, P.; Diederich, F., Molecular Clefts Derived from 9,9′-spirobi[9H-fluorene] for enantioselective complexation of pyranosides and dicarboxylic acids. *Helv. Chim. Acta* **1995,** *78* (2), 367-390.

(50) Mazik, M., Molecular recognition of carbohydrates by acyclic receptors employing noncovalent interactions. *Chem. Soc. Rev.* **2009,** *38* (4), 935-956.

(51) Mazik, M.; Sicking, W., Molecular recognition of carbohydrates by artificial receptors: Systematic studies towards recognition motifs for carbohydrates. *Chem. Eur. J.* **2001,** *7* (3), 664-670.

(52) Imberty, A.; Gautier, C.; Lescar, J.; Perez, S.; Wyns, L., An unusual carbohydrate binding site revealed by the structures of two Maackia amurensis lectins complexed with sialic acid-containing oligosaccharides. *J. Biol. Chem.* **2000,** *275* (23), 17541-17548.

(53) Tsukagoshi, K.; Shinkai, S., Specific complexation with mono- and disaccharides that can be detected by circular dichroism. *J. Org. Chem.* **1991,** *56* (13), 4089-4091.

(54) Adamczyk-Wozniak, A.; Borys, K. M.; Sporzynski, A., Recent Developments in the Chemistry and Biological Applications of Benzoxaboroles. *Chem. Rev.* **2015,** *115* (11), 5224-5247.

(55) Schwindeman, J. A.; Woltermann, C. J.; Letchford, R. J., Safe handling of organolithium compounds in the laboratory. *J Chem Health Saf.* **2002,** *9* (3), 6-11.

(56) Rosini, G., The Henry (Nitroaldol) Reaction. *Comprehensive Organic Synthesis* **1991**, 321-340.

(57) Hernandez, V.; Crepin, T.; Palencia, A.; Cusack, S.; Akama, T.; Baker, S. J.; Bu, W.; Feng, L. S.; Freund, Y. R.; Liu, L.; Meewan, M.; Mohan, M.; Mao, W. M.; Rock, F. L.; Sexton, H.; Sheoran, A.; Zhang, Y. C.; Zhang, Y. K.; Zhou, Y.; Nieman, J. A.; Anugula, M. R.; Keramane, E.; Savariraj, K.; Reddy, D. S.; Sharma, R.; Subedi, R.; Singh, R.; O'Leary, A.; Simon, N. L.; De Marsh, P. L.; Mushtaq, S.; Warner, M.; Livermore, D. M.; Alley, M. R. K.; Plattner, J. J., Discovery of a Novel Class of Boron-Based Antibacterials with Activity against Gram-Negative Bacteria. *Antimicrob. Agents Chemother.* **2013,** *57* (3), 1394-1403.

(58) John O. Osby, B. G., Rapid and efficient reduction of aliphatic nitro compounds to amines. *Tetrahedron Lett.* **1985,** *26* (52), 6413-6416.

(59) Paul, R.; Buisson, P.; Joseph, N., Catalytic Activity of Nickel Borides. *Ind. Eng. Chem. Res.* **1952,** *44* (5), 1006-1010.

(60) Hernandez, V. S. Benzoboroxole derivatives for treating bacterial infections. 2012. WO12/033858.

(61) Neises, B.; Steglich, W., Simple Method for the Esterification of Carboxylic Acids. *Angew. Chem. Int. Ed.* **1978,** *17* (7), 522-524.

(62) Bernatowicz, M. S.; Wu, Y. L.; Matsueda, G. R., 1-H-Pyrazole 1-carboxamidine hydrochloride - an attractive reagent for guanylation of amines and its application to peptide synthesis. *J. Org. Chem.* **1992,** *57* (8), 2497-2502.

(63) Zhang, P.; Zhuang, R. Q.; Guo, Z. D.; Su, X. H.; Chen, X. Y.; Zhang, X. Z., A Highly Efficient Copper-Mediated Radioiodination Approach Using Aryl Boronic Acids. *Chem. Eur. J.* **2016,** *22* (47), 16782-16785.

(64) Yang, W. Q.; Yan, J.; Fang, H.; Wang, B. H., The first fluorescent sensor for D-glucarate based on the cooperative action of boronic acid and guanidinium groups. *Chem. Commun.* **2003,** (6), 792-793.

(65) Hugues Miel, S. R., Total deprotection of N,N′-bis(tert-butoxycarbonyl)guanidines using SnCl4. *Tetrahedron Lett.* **1997,** *38* (45), 7865-7866.

(66) Fersht, A. R.; Shi, J. P.; Knilljones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M. Y.; Winter, G., Hydrogen-Bonding and Biological Specificity Analyzed by Protein Engineering. *Nature* **1985,** *314* (6008), 235-238.

(67) Wellington, N.; Macklai, S.; Britz-McKibbin, P., Elucidating the Anomalous Binding Enhancement of Isoquinoline Boronic Acid for Sialic Acid Under Acidic Conditions: Expanding Biorecognition Beyond Vicinal Diols. *Chem. Eur. J.* **2019,** *25* (67), 15277-15280.

(68) Petrauskas, V.; Maximowitsch, E.; Matulis, D., Thermodynamics of Ion Pair Formations Between Charged Poly(Amino Acid)s. *J. Phys. Chem. B* **2015,** *119* (37), 12164-12171.

(69) Suzuki, Y.; Sugaya, T.; Iwatsuki, S.; Inamo, M.; Takagi, H. D.; Odani, A.; Ishihara, K., Detailed Reaction Mechanism of Phenylboronic Acid with Alizarin Red S in Aqueous Solution: Re-Investigation with Spectrophotometry and Fluorometry. *ChemistrySelect* **2017,** *2* (10), 2956-2964.

(70) Fokkens, J.; Klebe, G., A simple protocol to estimate differences in protein binding affinity for enantiomers without prior resolution of racemates. *Angew. Chem. Int. Ed.* **2006,** *45* (6), 985-989.

(71) Du, X.; Li, Y.; Xia, Y. L.; Ai, S. M.; Liang, J.; Sang, P.; Ji, X. L.; Liu, S. Q., Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. *Int. J. Mol. Sci.* **2016,** *17* (2), 34.

(72) Saboury, A. A., Application of a new method for data analysis of isothermal titration calorimetry in the interaction between human serum albumin and Ni2+. *J. Chem. Thermodyn.* **2003,** *35* (12), 1975-1981.

(73) Dunitz, J. D., Win Some, Lose Some - Enthalpy-Entropy Compensation In Weak Intermolecular Interactions. *Chem. Biol.* **1995,** *2* (11), 709-712.

(74) Denamor, A. F. D.; Ritt, M. C.; Schwingweill, M. J.; Arnaudneu, F.; Lewis, D. F. V., Solution thermodynamics of amino acid–18-crown-6 and amino acid–cryptand 222 complexes in methanol and ethanol. Linear enthalpy–entropy compensation effect. *J. Chem. Soc.-Faraday Trans.* **1991,** *87* (19), 3231-3239.

(75) Sun, S. G.; Fazal, M. A.; Roy, B. C.; Chandra, B.; Mallik, S., Thermodynamic studies on the recognition of flexible peptides by transition-metal complexes. *Inorg. Chem.* **2002,** *41* (6), 1584-1590.

(76) Reynolds, C. H.; Holloway, M. K., Thermodynamics of Ligand Binding and Efficiency. *ACS Med. Chem. Lett.* **2011,** *2* (6), 433-437.

(77) Suzuki, Y.; Shimizu, M.; Okamoto, T.; Sugaya, T.; Iwatsuki, S.; Inamo, M.; Takagi, H. D.; Odani, A.; Ishihara, K., Detailed Mechanism of the Reaction of Phenylboronic Acid Derivatives with D-Fructose in Aqueous Solution: A Comprehensive Kinetic Study. *ChemistrySelect* **2016,** *1* (16), 5141-5151.

Chapter 4 - Towards the detection of sialylated glycans with scaffold-based receptors

Glycoproteins are proteins presenting one or more oligosaccharide chains, named glycans, linked to the peptide backbone[1]. Many different types of glycans have been identified, such as hybrid and complex types, mono-, bi- and multiantennary[2]. Among these, sialylated glycans are polysaccharide chains terminating with a sialic acid unit and have been identified as relevant biomarkers for cancerous conditions[3], such as colon, stomach, pancreatic, lung, breast and prostate cancers[4-6]. Cancer-associated sialylated glycans are structurally very similar to glycans expressed by normal cells, with the main difference being the glycosidic linkage configuration anchoring the sialic acid residue to a galactose (Gal) or a N-acetylgalactosamine (GalNAc) unit[6]. Sialic acid can be bound to galactose in α2-3 fashion, in which the glycosidic bond is between the 2-OH group of sialic acid and the 3'-OH group of galactose (Figure 4.1a). The α2-3 sialylated glycans have been found to be expressed by abnormal cancerous cells in, prostate cancer[6] and pancreatic adenocarcinoma metastasis[7]. On the other hand, non-altered cells display α2-6 sialylated glycans, in which the sialic acid residue is linked to the galactose through position 6 (Figure 4.1b).

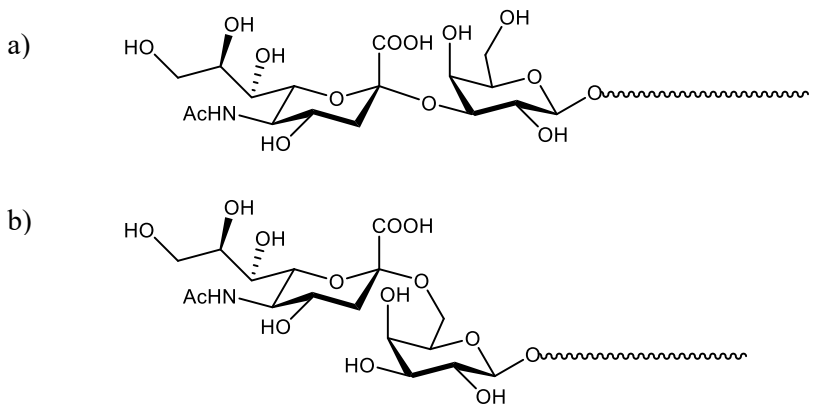


Figure 4.1. Representation of a) α2-3 sialylated glycans; b) α2-6 sialylted glycans.

For instance, the Prostate-Specific Antigen (PSA) of patients with prostate cancer (PCa) is rich in α2-3 sialylated glycans, whilst patients with Benign Prostatic Hyperplasia (BHP) present an abundancy of α2-6 linkages[6]. The current diagnostic test for prostate cancer is based on measuring the PSA levels, however the PSA levels can be enhanced not only in PCa patients but also in BHP patients resulting in high rates (up to 20%)[8] of false positive results[9-11]. On the other hand, the analysis of the PSA glycans can overcome this issue resulting in a more specific diagnosis and therefore reducing misdiagnosis and the associated negative consequences[10]. It must be noted that other cancerous conditions, such as breast cancer[5], are characterised by increased expression of α2-6 sialosides compared to non-altered cells which display α2-3 sialylated glycans. The differences in the sialylation of glycans can be exploited as biomarkers in the diagnosis of cancer. In order to achieved this, a highly selective diagnostic tool able to specifically target a certain glycan structure over the other glycoforms is required.

In order to achieve high affinity and selectivity, the receptor must provide multivalent interactions with the sialoside[12-14]. Indeed, multivalency is at the basis of the interactions between saccharides and their natural binders, lectins. Lectins are able to bind saccharides with μM affinities and display reasonable selectivity by forming a network of non-covalent interactions with the sugars[15]. Indeed, non-covalent interactions are weak interaction, and it is solely in a multivalent context that these become relevant allowing sugar recognition, with this being known as cluster effect[16]. This effect was also exploited in synthetic receptors[17-18], with one of the prime example of how this principle significantly affects the selectivity described by Tommasone *et al.*[19]. Within their work they developed molecular imprinted surfaces presenting high affinity and selectivity for a particular glycan over other oligosaccharides which share high structural analogy with the targeted one. Conversely to lectins, the cluster effect is achieved via multiple covalent interactions between the oligosaccharide and benzoboroxole units of the

surface binding pocket. In particular, each sugar unit is bound to at least one benzoboroxole affording dissociation constants comparable to lectins. The affinity and selectivity enhancement offered by cooperative binding was also demonstrated in Chapter 3 of this work. A receptor combining both covalent and non-covalent interactions displayed good affinity and selectivity for the monosaccharide sialic acid. When targeting oligosaccharides, a higher degree of multivalency is required compared to monosaccharides. In particular, a receptor with at least one functional group, or point of interaction, per sugar unit is needed. Unlike the case for small molecules, scaffold-based receptors are particularly suitable when targeting oligosaccharides, as they can be modified with different functional groups and therefore offer several points of interactions[20]. Many functional groups can be included in the receptor to promote additional binding, as glycans can engage in a variety of interactions, including hydrogen bonds, electrostatic and hydrophobic interactions, and covalent bonds. Given the great variety of potential scaffolds and functional groups there are many possible receptors requiring independent design, synthesis and testing. Such approach can be extremely time consuming, hence in order to overcome this issue dynamic combinatorial chemistry (DCC) was proposed as a promising technique for the discovery of oligosaccharide receptors. Dynamic combinatorial chemistry was first applied by Huc and Lehn in 1997 for the target-driven generation of carbonic anhydrase inhibitors[21]. In this approach the independent synthesis and testing of a large quantity of different potential inhibitors is replaced by the *in situ* formation of different adducts, obtained via a reversible reaction between different small molecules (*i.e* building blocks). The addition of the target protein, also known as template, to the library determines amplification of the best binders amongst the other adducts. DCC is a versatile technique which can be tailored to different substrates. Indeed, DCC has not only found application in medicinal chemistry[22], but also in the field of sensing specially focusing on ions

and small molecules[23]. Nevertheless, its applications in glycosensing are very limited and only apply to mono- and disaccharides[24-25], making this an unexplored technique for the generation of biologically relevant oligosaccharide receptors.

The aim of this work is to lay the foundations for the development of a universal system to generate highly specific receptors for biologically relevant glycans. In order for the development to proceed glycan analogues must first be synthesised. The cancer related α2-3 sialoside, and the associated α2-6 glycoform, were synthesised at Utrecht University under the supervision of Prof. Dr. Boons within a placement funded by the Erasmus program. Furthermore, consideration of the DCC components, including the synthesis of a peptidomimetic scaffold and the selection of sugar-binding building blocks, for the future application of this technique are also detailed within. This work will enable the future application of DCC in glycosensing.

Unfortunately, the COVID-19 pandemic significantly affected the progress of the research planned within this Chapter. In particular, the University closure in March 2020, due to the national lockdown, and the restrictions in place within the laboratories once the University re-opened in mid-July substantially limited my ability to perform further experimental work for this Chapter.

4.1 Enzymatic synthesis of sialylated epitopes

In this section the enzymatic synthesis of sialylated epitopes is detailed, alongside with the enzyme expression and NMR analysis. This work was performed within an Erasmus-funded placement at Utrecht University under the supervision of Prof. Dr. Boons, a world-leading expert in carbohydrate synthesis.

4.1.1 Glycan synthesis overview

The synthesis of di- or oligosaccharides involves the condensation of two sugar units, with formation of a glycosidic bond, to obtain a glycoside which can be achieved by chemical means. The chemical synthesis of a glycoside requires a sugar donor and a sugar acceptor[26]. The sugar donor must present a good leaving group in the anomeric position. In addition, both donor and the acceptor must present protecting groups in all the positions not involved in the glycosidic linkage to guarantee a strict control of the reaction regiochemistry. During the condensation, the donor leaving group is displaced by the nucleophilic free hydroxyl group of the acceptor through a S_N1 mechanism[27]. In particular, the leaving group is firstly activated by a promoter which supports the dissociation of the leaving group. The resulting glycosyl cation is stabilised via oxocarbenium ion intermediate which is then attacked by the acceptor nucleophilic OH group resulting in the formation of the glycosidic bond (Figure 4.2).

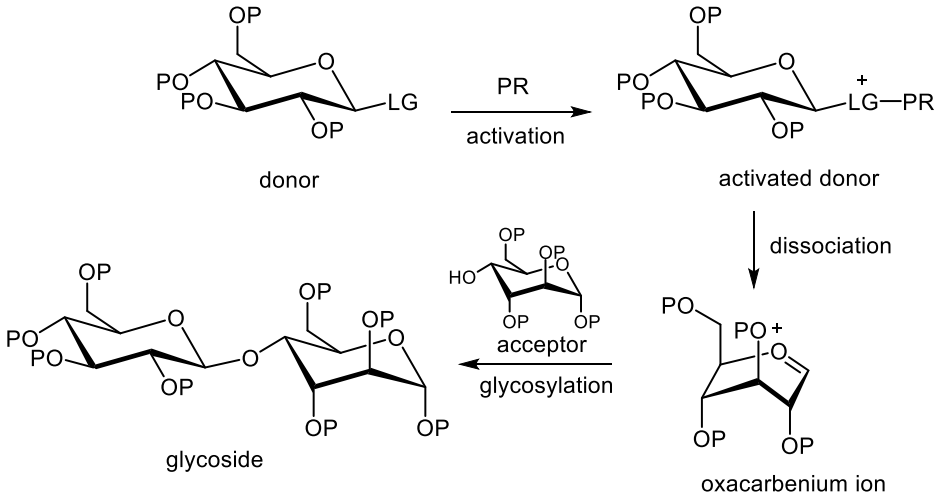


Figure 4.2. Glycosylation reaction mechanism; the sugar donor is firstly activated by a promoter (PR) resulting in the dissociation of the leaving group (LG) and formation of the oxocarbenium ion. The cation is then attacked by the acceptor nucleophilic OH resulting in the formation of

Hence, the chemical synthesis of a glycoside requires, prior to the glycosylation, the synthesis of the appropriate donor and acceptor derivatives and must then be followed by cleavage of the protecting groups. This synthetic procedure can be extremely time consuming and low yielding[26]. Yet, the biggest challenge in the glycosylation is controlling the anomeric stereochemistry of the linkage in order to afford solely the natural glycoside with minimal or no formation of the unnatural anomer[26]. In glycosylation reactions the anomeric selectivity is primarily controlled by the donor auxiliary group in position C2, which participate in the glycosylation by stabilising the oxocarbenium ion intermediate [26-27]. The stereochemistry can also be influenced by extra factors including reaction conditions (*e.g.* solvent, temperature), saccharide nature, protecting and leaving groups[26]. When performing a sialylation (*i.e.* addition of sialic acid) the equatorial glycoside, or α anomer, is the natural and desired one. However, in sialylations the stereochemistry is extremely hard to control due to the lack of an auxiliary group in position C3[26]. For this reason, the sialylation is the most challenging of all glycosylation reactions and rarely affords anomerically pure glycosides. In addition, the sialylation is also hampered by the presence of the electron withdrawing carboxylic acid on the anomeric carbon which renders the oxocarbenium ion unstable and thus susceptible to 2,3-elimination with formation of the corresponding glycal (Figure 4.3)[28].

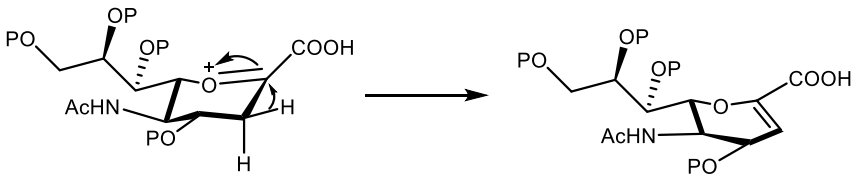


Figure 4.3. Oxocarbenium ion 2,3-elimination to afford the corresponding glycal.

The glycosylation is also sterically hindered by the tertiary anomeric centre and by the glycerol chain[28]. Chemical ways to control the sialylation stereochemistry have been developed and

consist of introducing a leaving group and an auxiliary group in position C2 and C3, respectively[26]. However, the synthesis of a suitable SA derivative and subsequent removal of the auxiliary group renders the synthesis of sialylated glycans exceptionally long and challenging.

To overcome these issues and achieve better anomeric purity enzymatic methods have been developed for the synthesis of sialylated glycans. The enzymatic synthesis exploits glycosyltransferase enzymes which provide high regio- and stereo- selectivity without requiring saccharide derivatisation[26].

4.1.2 Sialyltransferase enzymes

Glycosyltransferases are enzymes which catalyse the reaction between two sugar derivatives, a donor and an acceptor, to afford a glycoside. The donor is a sugar nucleotide, from which the sugar unit is transferred to the acceptor. Enzymes that transfer sialic acid from the sugar nucleotide, cytidine-5′-monophospho-sialic acid (CMP-sialic acid), to the acceptor are named sialyltransferases (STs) and they mediate the formation of α2-3 or α2-6 glycosidic linkages. The ST enzyme mediates the attack of the acceptor onto the donor anomeric carbon which is subject to an inversion of its configuration, from the β configuration of the CMP-sialic acid to the α configuration of the sialoside[29]. During the sialylation the phosphorylated nucleotide (CMP) is released and acts as an ST inhibitor limiting the glycosylation yield[30]. In order to avoid the enzymatic inhibition, Calf Intestinal Alkaline Phosphatase (CIAP) is generally added. This phosphatase hydrolyses the 5'-phosphate groups to afford the dephosphorilated nucleotide which does not display an inhibitory effect, resulting in improved yields[30].

Many sialyltransferases, either of bacterial or mammalian origin, have been employed in the enzymatic or chemoenzymatic synthesis of glycans[29]. Herein, we focus on *Pasteurella multocida* STs as these present significant advantages compared to other bacterial STs. The first sialyltransferase produced from the *Pasteurella multocida* was the recombinant multifunctional sialyltransferases PmST1, also known as tPm0188Ph[31]. The activity of this enzyme was found to be highly pH dependent with the α2-3 sialyltransferase activity being predominant at neutral to basic pH. Conversely, at acidic pH the PmST1 possess α2-6 sialyltransferase activity, although less efficient, and a sialidase activity which hydrolyses exclusively α2-3 linkages. The abovementioned activities are also displayed, to some extent, at neutral pH alongside the α2-3 sialyltransferase activity. In addition to this, the wild-type PmST1 catalyse the hydrolysis of the sugar donor CMP-sialic acid. All of the aforementioned side reactions significantly affect the α2-3 sialylation yield[32]. Nevertheless, the wild-type PmST1 present a series of advantages when compared to other bacterial STs, including the ability of acting as a catalyst on a broad variety of substrates, high expression levels in *Escherichia coli* and high kinetic constants[31-33]. The broad substrate specificity is a particularly attractive feature when glycans bearing unnatural sugar residues are required[31]. From the wild-type PmST1, two different mutants, the M144D and the P34H/M144L, have been engineered by the Chen group[31-32]. The single mutant PmST1 M144D had improved α2-3 sialylation yields due to the lower donor hydrolysis and lower α2-3 sialydase activity[32]. Whilst the double mutant PmST1 P34H/M144L was engineered to display exclusively α2-6 ST activity[31].

Compared to mammalian glycosyltransferases, the bacterial ones are more easily expressed, more stable and have broader acceptor specificity[34] [29]. Nevertheless, mammalian sialyltransferases have also been successfully employed in the enzymatic synthesis of sialoglycans[35-38]. Human ST3Gal enzymes found application in the synthesis of α2-3 sialylated

epitopes, as they mediate the transfer of a sialic acid residue from the CMP donor to the galactose residue of different acceptors[39]. In particular, the enzymes ST3Gal 3, 4 and 5 mediate the sialylation of Galβ1-4GlcNac structures, with the ST3Gal4 enzyme affording the highest conversion rate[39].

4.1.3 Expression of PmST1 M144D enzyme

The bacterial enzymes PmST1 M114D and PmST1 P34H/M144L as well as the mammalian ST3Gal4 were employed in the synthesis of sialylated epitopes. The enzymes ST3Gal4 and PmST1 P34H/M144L were previously expressed by the Boons group and stored at -80°C. Conversely, the PmST1 M144D was not available and was therefore expressed with the help of Boons team members following a published procedure[31]. A preculture of *E. coli* cells containing the plasmid with the genetic information for the expression PmST1 M144D was employed. The preculture of cells was added to lysogen broth rich medium to promote their growth. To this the antibiotic ampicillin was added to prevent the growth of other bacteria. The medium was incubated at 37 °C and the cell growth was monitored by UV every 30 mins. The cell growth was measured through the parameter OD_{600}. The OD_{600} is a measurement of the optical density of a sample at 600 nm and it is routinely employed to measure cell growth[40]. When the OD_{600} value reached a value above 0.6 the cell growth was stopped, and the protein expression induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG). The IPTG induces the expression of the recombinant protein from the plasmid in *E. coli*[41]. The cells were incubated overnight at room temperature whilst shaking and then harvested by centrifugation. The cells were then treated with a solution of Triton X-100 and addition of lysozyme and DNase in order to promote lysis of cells and cellular components[31]. This was followed by centrifugation and collection of the supernatant which was then loaded on nickel-nitrilotriacetic

acid affinity column (Ni^{2+}-NTA) to allow purification of the enzyme of interest. The enzyme, which was His-tagged, was retained by the solid phase whilst the rest of the component of the lysate were washed with MilliQ water. Then an imidazole buffer with increasing concentration, from 5 to 200 mM, was employed to elute the enzyme through the imidazole replacing the enzyme in binding the Ni^{2+} resin. The collected vials were analysed by UV to verify the presence of the enzyme. From the measured absorbance value it was possible to roughly estimate the concentration of the different vials by using the relationship relating the absorbance to the protein concentration; 1 A_{280} = 1 mg/ml[42]. According to this, the vials were found to have a concentration between 0.6 and 0.9 mg/ml. The purified enzyme was denatured and analysed by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) at 200 V for 30 minutes[31] and then stained with Coomassie Blue stain to reveal the presence of the PmST1 enzyme in the sampled vials. The PmST1 M144D was stored in a fridge at 4 °C and utilised within a month of expression.

4.1.4 Sialylation of β-methylgalactose

Cancer-associated sialylated glycans are polysaccharide chains terminating with a sialic acid unit[3]. In various cancerous conditions these display a α2-3 glycosidic bond between sialic acid and galactose residues. Conversely, the α2-6 isoforms are produced by non-altered cells. The simplest epitopes retaining the key discrepancy between the aforementioned glycosides are the disaccharide units consisting of sialic acid and galactose; SAα2-3Gal or SAα2-6Gal. These small epitopes can be employed to mimic the terminal end of sialylated glycans. Furthermore, in glycans the galactose residue is anchored to the oligosaccharide chain via a β glycosidic linkage between its the anomeric hydroxyl group and the third to last sugar unit[1]. To mimic this, the disaccharide epitopes must present the galactose anomeric hydroxylic group engaged

in a glycosidic linkage. The simplest β glycoside of galactose is the β-methylgalactose, in which the anomeric hydroxyl group is methylated. Hence, the two simplest sialylated epitopes consist of a sialic acid unit linked, α2-3 or α2-6, to β-methylgalactose. Herein, the synthesis of the sialylated disaccharides SAα2-3GalβMe (Figure 4.4a) and SAα2-6GalβMe (Figure 4.4b) is discussed.

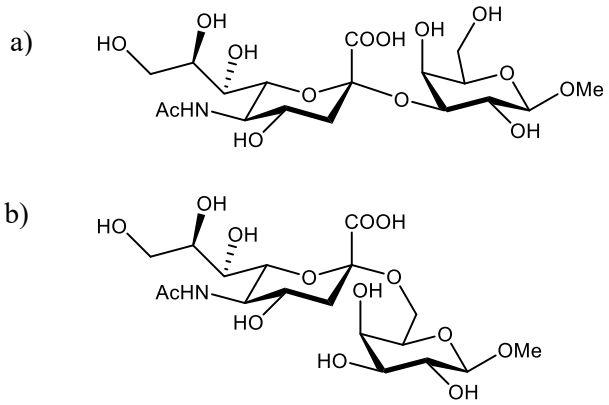


Figure 4.4. Sialylated disaccharides: a) SAα2-3GalβMe; b) SAα2-6GalβMe.

The enzymatic synthesis of the SAα2-6GalβMe epitope consist in the α2-6 sialylation of the acceptor β-methylgalactoside with the donor CMP-sialic acid (Figure 4.5). The bacterial sialyltransferase *Pasteurella multocida* PmST1 double mutant (P34H/M144L) was selected[33] for its broader acceptor promiscuity. In particular, the sialylation of GalβMe by this enzyme was not already validated. Hence, the ST activity of PmST1 P34H/M144L was tested with the donor GalβMe and the acceptor CMP-sialic acid. The enzymatic reaction was performed at pH 7.3 with addition of the cofactor $MgCl_2$ and the phosphatase CIAP. The reaction was incubated overnight at 37 °C and then analysed by LC-MS in positive ion mode.

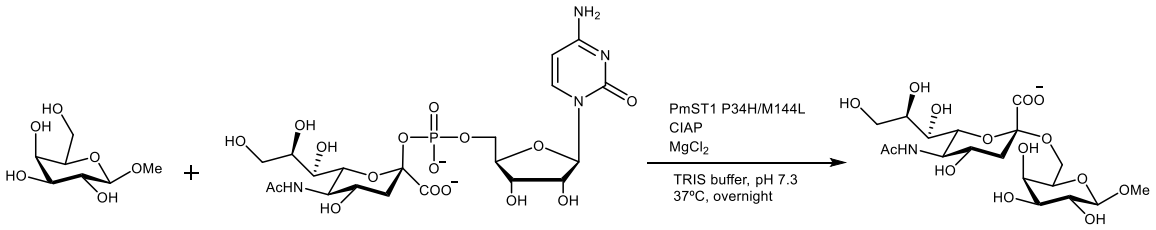


Figure 4.5. Enzymatic synthesis of SAα2-6GalβMe.

The LC-MS analysis allows identification of the species presented in the reaction mixture by analysis of their *m/z* ratio. If the enzymatic sialylation were successful the SAα2-6GalβMe disaccharide would be detected alongside with the unreacted CMP-sialic acid, which was added in excess. On the other hand, if any unreacted GalβMe was still present in the mixture this would not be observed by MS analysis as its molecular weight (MW = 194.2) is below the instrument's lower mass limit. The analysis of the reaction mixture showed solely the presence of the CMP-sialic acid with signals of *m/z* = 635.1 $[M + Na]^+$ and *m/z* = 681.1 $[M + 3Na]^+$, with no disaccharide being detected (MW = 484.4).

Furthermore, the same experimental conditions were employed when the testing of *Pasteurella multocida* PmST1 single mutant (M144D)[32] in the synthesis of the α2-3 epitope, SAα2-3GalβMe. The LC-MS analysis of this mixture in positive ion mode revealed solely the presence of the unreacted CMP-sialic acid without traces of the sialylated epitope. Consequently, it was concluded that despite the broader acceptor promiscuity of the PmST1 mutants they were not able to catalyse the sialylation of the unnatural β-methylgalactoside acceptor.

In spite of the limited acceptor substrate specificity of mammalian enzymes, the ST3Gal4 activity in the sialylation of β-methylgalactoside was tested (Figure 4.6). This reaction was performed at pH 7.5 in the presence of CIAP and BSA. BSA is added to prevent adhesion of the enzyme to the tube walls which would decrease its availability[43]. The reaction was

performed with an excess of donor and incubated at 37 ºC. The reaction progression was monitored by LC-MS. After 3 hours the sialylated disaccharide was observed, *m/z* = 508.2 $[M + Na]^+$ and *m/z* = 530.2 $[M + 2Na]^+$, alongside with the unreacted donor (*m/z* = 681.1 $[M + 3Na]^+$). The analysis in negative ion mode also confirmed the presence of the sialylated disaccharide, *m/z* = 484.1 $[M - H]^-$.

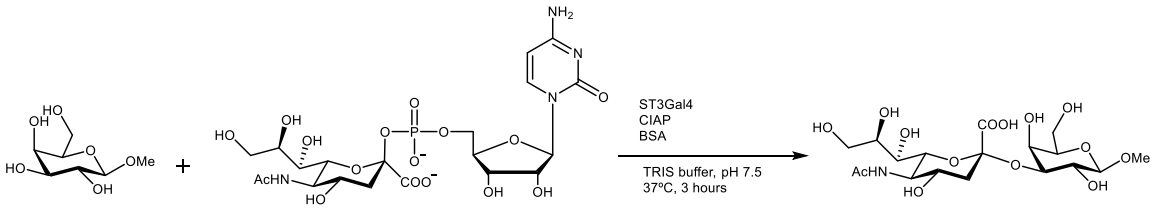


Figure 4.6. Enzymatic synthesis of SAα2-3GalβMe.

The unreacted GalβMe cannot be detected by MS analysis, hence in order to assess the rate of conversion the mixture was analysed by thin layer chromatography (TLC). Following staining with a 5% H_2SO_4 ethanolic solution a spot attributable to the disaccharide was revealed whist the acceptor GalβMe was not observed, thus implying full conversion (Figure 4.7).

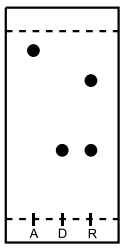


Figure 4.7. Representation of TLC plate run in 4:2:1 EtOAc:MeOH:water and stained with 5% H_2SO_4 in ethanol. A = acceptor, β-methylgalactoside; D = donor, CMP-sialic acid; R = reaction

From the above-described test reactions it was concluded that the SAα2-3GalβMe can be successfully synthesised with the mammalian enzyme ST3Gal4. Whilst the PmST1 double mutant did not display α2-6 ST activity towards the GalβMe acceptor. In addition, α2-6 sialylation of GalβMe with mammalian enzymes could not be explored as no α2-6 mammalian

STs were available to us. Consequently, a different strategy needed to be investigated as both α2-3 and α2-6 sialylated epitopes were required.

The enzymatic sialylation of unnatural sugar is particularly challenging as enzymes tend to favour natural acceptors. The use of galactose as acceptor, rather than its methylated counterpart, could offer greater compatibility with the STs. However, the corresponding epitopes, SAα2-3Gal and SAα2-6Gal, would present the galactose anomeric hydroxyl group exposed, thus diminishing the analogy to glycan terminal ends. Thus, it was decided to link galactose to another sugar unit, the N-acetylglucosamine in order to afford the disaccharide N-acetyllactosamine (LacNAc). This was then sialylated to give the SAα2-3LacNAc and SAα2-6LacNAc trisaccharide epitopes.

4.1.5 Synthesis of N-acetyllactosamine

The N-acetyllactosamine (LacNAc) disaccharide consists of a galactose unit linked β1-4 to N-acetylglucosamine (GlcNAc), Galβ1-4GlcNAc[38]. This epitope can be found in many glycans with its sialylated analogue observed in cancer-related glycans[44-45]. LacNAc is a natural sugar and therefore presents higher compatibility with sialyltransferases.

LacNAc was enzymatically synthesised from the N-acetylglucosamine acceptor and the uridine diphosphate galactose (UDP-Gal) donor (Figure 4.8). The galactosyltranferase B4GalT1 was employed for this reaction. B4GalT1 is a mammalian enzyme belonging to the β4-Gal-transferases family[34] which catalyse the transfer of galactose from UDP-Gal to GlcNAc, creating a glycosidic linkage between the galactose anomeric position and GlcNAc position 4. The enzymatic synthesis of LacNAc was performed at pH 7.5 in 50 mM TRIS buffer with the addition of CIAP, BSA and the cofactor $MnCl_2$. The UDP nucleoside released during the

reaction was dephosphorylated by the CIAP enzyme with the phosphate salt being visible as a white precipitate after overnight incubation at 37 ºC.

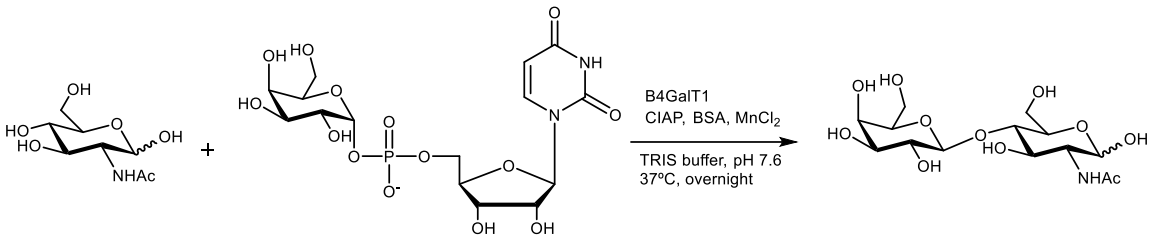


Figure 4.8. Enzymatic synthesis of LacNAc (Galβ1-4GlcNAc).

The LC-MS analysis of the mixture revealed presence of the LacNAc product (MW = 383.4) with *m/z* = 406.1 $[M + Na]^+$ and *m/z* = 789.3 $[2M + Na]^+$ corresponding to the monomer and dimer, respectively. However, due to the low molecular weight (MW = 221.2) of GlcNAc it was not possible to verify the rate of conversion for this reaction by mass spectrometry. Consequently, the mixture was analysed by TLC and following staining with 5% H_2SO_4 in ethanol no starting material was observed suggesting full conversion.

The enzymatic reaction was stopped by freeze drying the mixture. The crude was purified by size exclusion chromatography to separate the LacNAc from salt, enzyme, and the remaining sugar nucleotide. The collected fractions were analysed by TLC and the ones containing the disaccharide were also analysed by LC-MS to confirm their composition and purity. The fractions containing LacNAc, *m/z* = 406.1 $[M + Na]^+$ and *m/z* = 789.3 $[2M + Na]^+$, were collected and freeze dried overnight.

4.1.6 Synthesis of sialylated epitopes

The ST enzyme presents higher activity towards natural sugar acceptors, thus the sialylation of LacNAc was preferred to the unnatural GalβMe. In addition, the sialylation of a disaccharide

presents a practical advantage over monosaccharide sialylations. As described in Section 4.1.4, the rate of conversion for the sialylation of a monosaccharide is hard to monitor by LC-MS as the acceptor, which is the limiting reagent, has a low molecular weight. Thus, other methods, such as thin layer chromatography, are required to assess the conversion. On the other hand, the progression of the LacNAc sialylation can be easily monitored by LC-MS as its molecular weight is above the instrument's lower mass limit.

The α2-6 sialylation of LacNAc to afford the SAα2-6LacNAc epitope was performed with the PmST1 P34H/M144L enzyme and the CMP-sialic acid donor (Figure 4.9). In the mixture the cofactor $MgCl_2$ and the phosphatase enzyme CIAP were added, and the reaction was incubated for 3 hours at 37 °C and monitored by LC-MS.

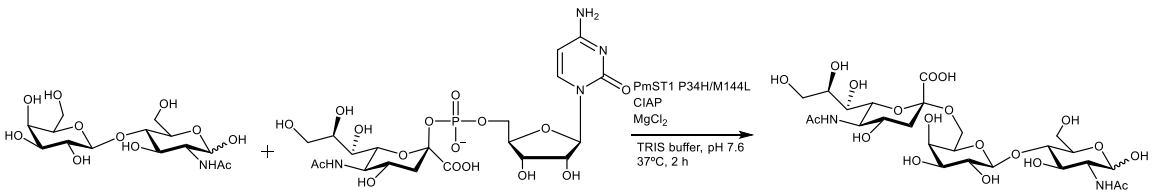


Figure 4.9. Enzymatic synthesis of SAα2-6LacNAc.

The positive MS spectra presented two signals at *m/z* = 697.2 and *m/z* = 384.1 corresponding respectively to the trisaccharide ($[M + Na]^+$) and the acceptor ($[M + H]^+$). From the relative intensity of the two signals, it was established that the conversion was near completion (80 – 90%) and therefore the reaction was stopped by freeze drying the mixture. The crude was dissolved in water and filtered through a spin filter to remove the enzyme from the mixture. The retrieved solution was freeze dried and then purified by ion exchange column chromatography. The column was initially eluted with MilliQ water allowing the elution of unreacted LacNAc, whilst the negatively charged sialyl-LacNAc was retrained by the column via interaction with the solid-phase cationic heads. Once the elution of LacNAc and salts was

complete the solvent system was changed to 20 – 50 mM ammonium carbonate buffer in order to disrupt the interaction between the sialyl-LacNAc and the amino groups of the stationary phase, hence allowing elution of the pure trisaccharide SAα2-6LacNAc (79.7 mg, 53.7%).

For the synthesis of the SAα2-3LacNAc epitope the mammalian enzyme ST3Gal4 was employed. The mammalian enzyme was preferred over the bacterial PmST1 M144D due to the significant sialidase side-activity that this display. Contrary to PmsST1, the mammalian enzyme ST3Gal4 does not display sialidase activity, offering a more efficient sialylation. The ST3Gal4 activity towards LacNAc of was previously confirmed by the Boons group[38]. The synthesis of SAα2-3LacNAc was performed with an excess of CMP-sialic acid and in the presence of CIAP and BSA in 50 mM TRIS buffer at pH 7.6 (Figure 4.10). The mixture was incubated at 37 °C for 3 hours and then analysed by LC-MS in positive and negative ion mode. In negative ion mode a signal of *m/z* = 673.4 $[M - H]^-$ confirmed the presence of the SAα2-3LacNAc. The trisaccharide was also observed in positive ion mode with *m/z* = 697.2 $[M + Na]^+$. However, the LacNAc was revealed to be the predominant component of the mixture with *m/z* = 406.1 $[M + Na]^+$ and *m/z* = 789.3 $[2 M + Na]^+$. In order to drive the reaction towards completion, a further 200 µl of ST3Gal4 were added. Following overnight incubation at 37 °C, the LC-MS analysis showed the predominant presence of product (*m/z* = 697.2 $[M + Na]^+$) and a small amount of acceptor (*m/z* = 406.1 $[M + Na]^+$).

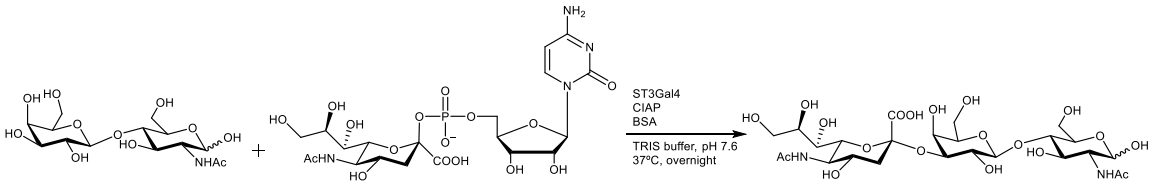


Figure 4.10. Enzymatic synthesis of SAα2-3LacNAc.

The mixture was freeze dried overnight and then purified by size exclusion chromatography, which was preferred to spin filtering as the latter can be extremely time consuming when a large amount of enzyme is present. The purification via size exclusion chromatography allowed separation of the enzyme and the sugar nucleotide from the mixture. The resulting solution, containing primarily LacNAc and sialylated-LacNAc, was initially freeze dried and then purified by ion exchange column chromatography to afford the pure trisaccharide SAα2-3LacNAc (55.4 mg, 23.7%).

4.1.7 NMR analysis of sialylated trisaccharides

The synthesised epitopes, SAα2-3LacNAc and SAα2-6LacNAc (Figure 4.11), were analysed by Nuclear Magnetic Resonance (NMR) analysis in order to confirm their structure and anomeric configuration. The NMR analysis of the synthesised trisaccharides comprised of ^{1}H, ^{1}H-^{1}H Correlation Spectroscopy (^{1}H-^{1}H COSY), ^{1}H-^{1}H Total Correlation Spectroscopy (^{1}H-^{1}H TOCSY) and ^{1}H-^{13}C Hetereonuclear Single Quantum Coherence Spectroscopy (^{1}H-^{13}C HSQC), for spectra see Appendix. The ^{1}H NMR alone is not very informative given the crowded region between 4.01 ppm and 3.55 ppm, and between 4.00 ppm and 3.48 ppm for the α2-3 and α2-6 sialosides, respectively. Due to the overlapping of signals, in the ^{1}H NMR only a few signals can be identified unequivocally. These include the acetamide CH_3 protons, the sialic acid H_{3eq} and H_{3ax} and the anomeric protons of galactose and N-acetylglucosamine. It must be noted that the configuration of GlcNAc anomeric carbon is not fixed, hence both anomers are present and can be observed in the NMR spectra. The anomeric proton (H_1) for the α-GlcNAc was found at 5.21 ppm and 5.22 ppm for the SAα2-3LacNAc and SAα2-6LacNAc, respectively. Whilst the H_1 signal for the GlcNAc β anomer was found at 4.73 ppm and 4.72 ppm, for the α2-3 and α2-6 glycans, respectively. Following the assignment of the

aforementioned signals, all other signals (Table 4.1 and Table 4.2) were assigned by analysis of the correlation with nearby protons via ^{1}H-^{1}H COSY and ^{1}H-^{1}H TOCSY NMR analysis.

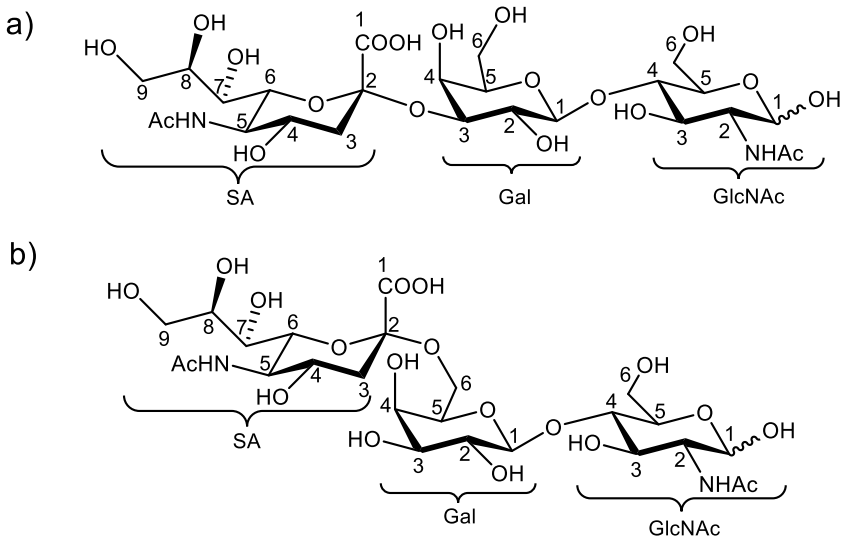


Figure 4.11. Structures of a) SAα2-3LacNAc, b) SAα2-6LacNAc.

The ^{1}H-^{1}H COSY NMR reveals the coupling of vicinal or geminal protons, whilst the ^{1}H-^{1}H TOCSY shows the coupling of protons which are part of the same chain even when not vicinal or geminal[46]. For instance, the COSY spectrum of SAα2-6LacNAc shows the correlation between galactose H_1, which was previously assigned, and a peak at 3.52 ppm which is therefore the vicinal H_2. The TOCSY NMR reveals H_1 correlating with the vicinal H_2 but also with two other signals at 3.65 ppm and 3.90 ppm (Figure 4.12). In order to assign these two H_3 and H_4, the cross peaks intensities were considered, since the correlation is higher for protons separated by fewer bonds. The more intense cross peak of the signal at 3.65 ppm was assigned to H_3, since this is in closer in the chain to H_1. Conversely, the signal at 3.90 ppm is weaker indicating that the proton is further away from H_1, thus this was ascribed to H_4.

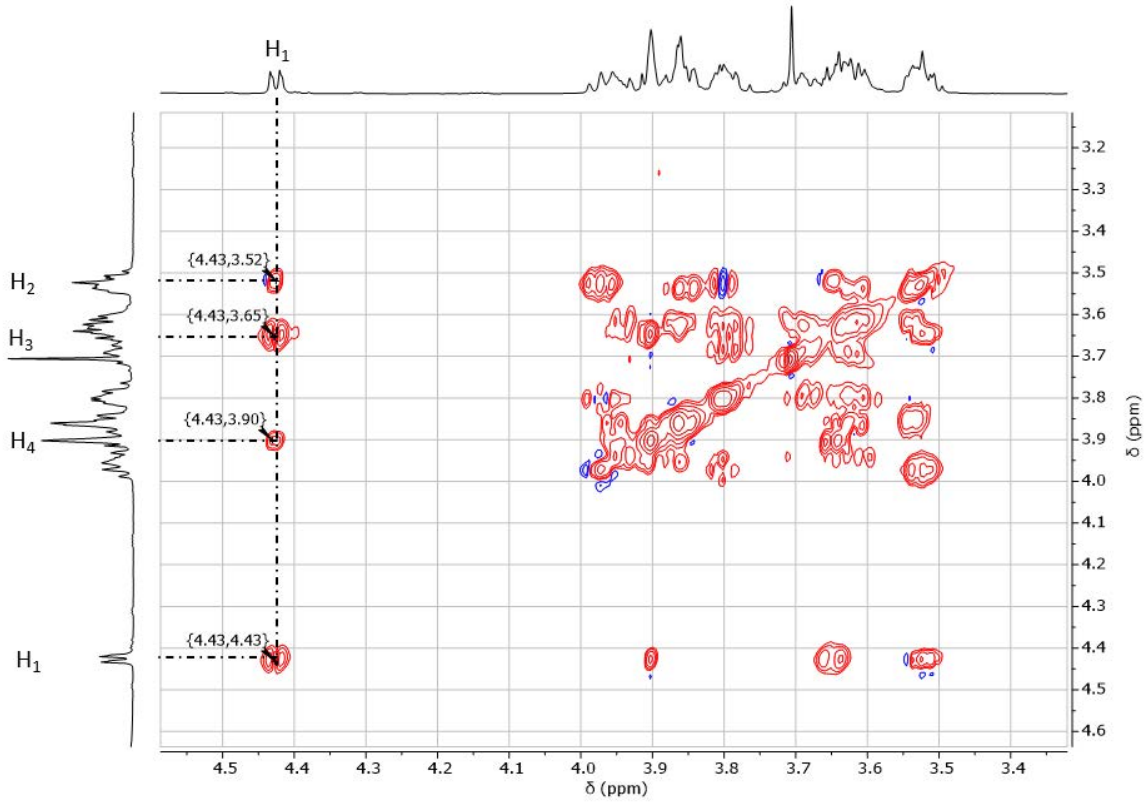


Figure 4.12. ^{1}H-^{1}H TOCSY NMR spectrum of SAα2-6LacNAc recorded in D_2O at 600 MHz for the assignment of galactose proton H_1, H_2, H_3 and H_4 by cross peak analysis.

The assignment of the peaks based on the 2D NMR spectra was applied for all other sugar units of both trisaccharides (Table 4.1 and Table 4.2). Furthermore, the ^{13}C signals were assigned via ^{1}H-^{13}C HSQC, which display the coupling between a given carbon and any protons directly bound to it (see Experimental section 5.2.3)[46]. It must be noted that in this fashion only non-tertiary carbon signals could be identified. The ^{13}C spectra were not recorded as the intensity of the signals would not been sufficient for accurate identification.

Table 4.1. ^{1}H NMR peak assignment of SAα2-3LacNAc recorded at 500 MHz in D_2O.

	H_1	H_2	H_3	H_4	H_5	H_6	H_7	H_8	H_9	NHAc
GlcNAc α	5.21 (d, J = 2.3 Hz, 1H)	3.91	3.76	3.98	3.91	3.90 3.85	-	-	-	2.04
GlcNAc β	4.73 (d, J = 7.9 Hz, 1H)	3.73	3.76	3.61	3.75	3.98 3.90	-	-	-	2.04
Gal	4.56 (dd, J = 7.9, 2.9 Hz, 1H)	3.59	4.13 (dt, J = 9.9, 3.2 Hz, 1H)	3.97	n/a	3.74	-	-	-	-
Neu5Ac	-	-	2.77 – eq (dd, J = 12.5, 4.5 Hz, 1H) 1.81 – ax (t, J = 12.7 Hz, 1H)	3.71	3.65	3.86	3.98	3.84	3.88 3.65	2.05

n/a = not assigned

Table 4.2. ^{1}H NMR peak assignment of SAα2-6LacNAc recorded at 600 MHz in D_2O.

	H_1	H_2	H_3	H_4	H_5	H_6	H_7	H_8	H_9	NHAc
GlcNAc α	5.22 (d, J = 2.7 Hz, 1H)	3.90	3.65	3.95	3.89	3.87 3.82	-	-	-	2.00
GlcNAc β	4.72 (d, J = 7.7 Hz, 1H)	3.70	3.60	3.80	3.95	3.95 3.87	-	-	-	2.00
Gal	4.43 (dd, J = 7.9, 2.1 Hz, 1H)	3.52	3.65	3.90	n/a	3.98 3.53	-	-	-	-
Neu5Ac	-	-	2.65 – eq (dd, J = 12.4, 4.7 Hz, 1H) 1.70 – ax (t, J = 12.2 Hz, 1H)	3.63	3.68	3.79	3.98	3.54	3.86 3.63	2.04

n/a = not assigned

The NMR analysis also provides crucial information on the glycosidic linkage anomeric configuration. The anomeric configuration is normally determined by the coupling constants $^{3}J_{H-1,H-2}$ and $^{1}J_{C-1,H-1}$, however this is not possible for sialylated glycans as the sialic acid anomeric position does not bear any hydrogen[26]. For sialylated glycans other parameters can be assessed to verify the anomeric configuration[26]. The most clear-cut parameter is the chemical shift of the equatorial proton in position 3 (H_{3eq}). Indeed, the protons in position 3 of sialic acid

can be unequivocally identified as they are at higher fields compared to the other proton signals. In α anomers the H_{3eq} is found downfield (δ = 2.67 – 2.72 ppm) compared to β anomers (δ = 2.25 – 2.40 ppm)[47]. In the α2-3 and α2-6 sialylated trisaccharides, the H_{3eq} has a chemical shift of 2.77 ppm and 2.65 ppm, respectively. Furthermore, the chemical shift of the sialic acid proton in position 4 (H_4) can also provide indication of the anomeric configuration. This appears between 3.60 ppm and 3.80 ppm for α anomers and between 3.90 ppm and 4.20 ppm for the β ones[48]. The H_4 signal is present in a crowded area of the spectra overlapping with other proton signals. Therefore, the identification of the H_4 requires analysis by its correlation with the vicinal H_3. The H_3 protons were found to correlate in the ^1H-^1H COSY spectrum with a signal at 3.71 ppm and 3.63 ppm for the α2-3 and α2-6 glycans, respectively. The chemical shift of the H_4 protons is consistent with the anticipated value for α anomers. Alternatively, the anomeric conformation can be evaluated via the proton signals of position 9, H_{9a} and H_{9b}. The two protons in position 9 provide two distinct signals as they are diastereotopic protons. The difference in the chemical shift of the two signals can offer information on the anomeric configuration. Literature states that α anomers present a difference in the chemical shift (Δδ) between H_{9a} and H_{9b} below 0.5 ppm, whilst this is about 1.00 ppm for the β configuration[49]. The H_{9a} and H_{9b} protons were identified via 2D NMR analysis. The ^1H-^{13}C HSQC spectrum reveals protons bound to their corresponding carbon atom. When two protons are bound to the same carbon atom this is observed in the HSQC spectrum. Each trisaccharide, considering both the presence of α and β GlcNAc, presents 4 CH_2 signals. Three of these signals were attributed to the α-GlcNAc, β-GlcNAc and Gal units via ^1H-^1H COSY and ^1H-^1H TOCSY analysis. The remaining signals at 3.88 ppm and 3.65 ppm, for the α2-3 glycan, and at 3.86 ppm and 3.63 ppm, for the α2-6 glycan, were ascribed to the H_{9a} and H_{9b} protons. Their assignment was also corroborated by ^1H-^1H COSY and ^1H-^1H TOCSY analysis. The Δδ of proton H_{9a} and H_{9b}

was calculated to be 0.23 ppm for both sialylated glycans and was therefore consistent with the α anomeric configuration. The NMR analysis therefore confirms the structure and anomeric configuration of the synthesised SAα2-3LacNAc and SAα2-6LacNAc epitopes.

4.2 Scaffold-based dynamic combinatorial chemistry

The target-driven dynamic combinatorial chemistry approach was explored for the development of highly selective carbohydrate receptors. This Section includes the design and synthesis of a peptidomimetic scaffold and the selection of templates and building blocks to be employed in future technique development.

4.2.1 Dynamic combinatorial chemistry overview

Dynamic combinatorial chemistry is a technique employed for the *in situ* generation of libraries of chemical compounds from which the best binder, for a given target, is selected (Figure 4.13)[21]. DCC is based on two sets of building blocks (BBs), which are formed by commercially available molecules, presenting complementary reactivity. The BBs react *in situ*, via a reversible reaction, to generate a library of adducts, known as dynamic combinatorial library (DCL)[21]. Alternatively to having two sets of small molecules, one of the two BB set can be replaced by a scaffold molecule[50]. The scaffold will react with a set of BBs and generate different appended scaffolds.

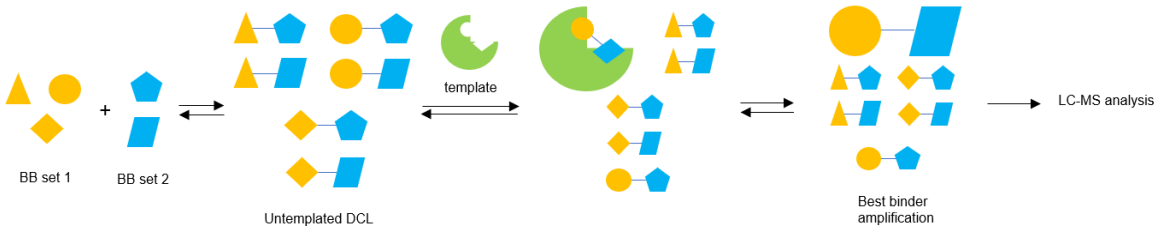


Figure 4.13. Schematic representation of the dynamic combinatorial chemistry (DCC)

A great variety of reversible reactions can be exploited in DCC including imine, hydrazone, oxime and disulfide exchanges, Michael addition, Dies-Alder, alkene metathesis and boronate esterification[22, 51]. Furthermore, the DCL composition (*i.e.* the concentration of each adduct) is governed by the thermodynamic equilibrium of the system and when this is perturbated by external stimuli the library composition will be consequently altered[21]. The addition of a target, known as template, to the library triggers a shift in the library equilibrium. The molecular recognition between the template and one (or more) adduct(s) will determine the amplification of these adducts amongst the others. Thus, the composition of the templated library will be significantly different to the untemplated, or unbiased, library. The DCLs are then analysed by chromatography and spectrometric techniques in order to assess the nature of the amplified adducts[52].

Herein, the DCC approach was applied to glycosensing with the aim of developing a universal method for sugar recognition which can be adapted to different biologically relevant glycans. It is anticipated, as also shown by Rauschenberg *et al.*[24], that the addition of different carbohydrates to a particular BB mixture will generate DCLs with distinct compositions and amplified adducts. The DCC method under development here consist of five main components: a scaffold molecule, a library of building blocks, a reversible reaction, a template and an analytical method. Instead of the classical methodology, which utilises two sets of small building blocks, a scaffold-based approach was adopted, due to the high degree of multivalency

offered by this system[50, 53]. A high degree of multivalency is particularly important for glycan recognition due to the weak nature of individual non-covalent interactions. Indeed, strong binding can only be achieved when many non-covalent interactions are formed, within the so-called "cluster effect"[16]. Among the possible scaffolds (*i.e.* calixarenes, peptides, dendrites and polymers)[20] peptidomimetic scaffolds were deemed more suitable for glycan recognition due to their structural analogy to lectins. In addition, libraries of building blocks were constructed to comprise of a variety of different small molecules containing sugar-binding functionalities.

The DCC approach was based on imine exchange. Imines, also known as Schiff bases[54], are formed by reaction of a carbonyl group, aldehyde or ketone, with a primary amine in a two-step mechanism. In the first step the addition of the amine to the aldehyde gives the hemiaminal intermediate which, in the second step, undergoes water elimination to yield the corresponding imine (Figure 4.14)[55].

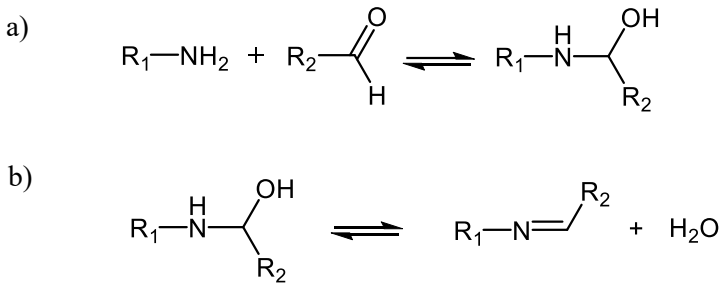


Figure 4.14. The imine formation consists of two steps; a) addition of the amine to the aldehyde with formation of the hemiaminal intermediate; b) elimination of water with formation of the imine.

When this reaction is performed in organic solvent the equilibrium is pushed towards the condensation product[56]. On the contrary, for DCC purposes, the imine exchange is performed in aqueous media to guarantee its reversibility[56]. The imine exchange can be performed between

pH 5.0 and 8.5[57]. Consequently, there is scope for the DCC to be performed at physiological conditions, where the receptors will find their ultimate application in diagnostics. Due to the commercial availability of the building blocks and the peptidomimetic synthetic strategy, a choice was made to assign the aldehyde moiety to the scaffold and for the building blocks to bear a primary amine.

Unfortunately, due to the thermodynamically instability of imines under aqueous conditions, the DCL cannot be directly analysed. In order to have detectable concentrations of the adducts for analytical purposes, the imine equilibrium must be frozen[57]. This is achieved via reductive amination to the corresponding secondary amine with sodium cyanoborohydrate[21]. It must be noted that in this process the information regarding the double bond conformation (*E* and *Z*) are lost and the resulting secondary amine could present different affinity for the template (Figure 4.15). Nevertheless, this issue is rarely encountered, and the amines are generally found to retain the activity of the corresponding imines[21].

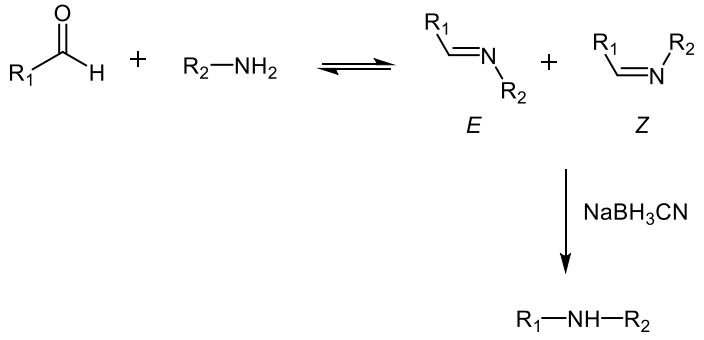


Figure 4.15. Formation of *E* and *Z* imines and consequent reductive amination with $NaBH_3CN$ to the corresponding secondary amine.

An analytical method is required to determine the composition of the DCL. In order to identify the amplified members a comparative approach must be adopted. This consists of the generation, under the same experimental conditions, of two libraries. One containing the template and the other being a control library without template addition[52]. The DCL analysis is

generally conducted by high performance liquid chromatography (HPLC) or liquid chromatography (LC) combined with mass spectrometry (MS) to allow assignment of the peaks. Comparison of the relative peak area (RPA) from the two libraries allows the determination of amplified peaks[52], whilst the MS allows identification of the adduct structure via their *m/z* ratio[22]. To allow comparison of the HPLC traces of biased and unbiased libraries the adduct-template complexes must be disrupted prior analysis[22]. Without disruption of the complex the most effective binders will not present amplified peak in the chromatogram. For protein templates the disruption is generally achieved by protein denaturation. Conversely, a glycan-adduct complex could be disrupted by a pH change perturbating both covalent and non-covalent interactions.

4.2.2 Scaffold-based DCC approach

The DCC approach presented herein is based on a linear scaffold which will form a library of adducts by reacting, via imine exchange, with a set of small building blocks. Linear scaffolds were preferred to cyclic ones as they can provide a higher degree of multivalency[20]. Each building block, when bound to the scaffold, represents a potential point of interaction for the glycan. The receptor must display at least one point of interaction per sugar unit[19]. Thus, the scaffold must contain a number of aldehyde groups at least equal to the number of template sugar units. For instance, if the template is an hexasaccharide the scaffold will require at least six points of interaction and the scaffold must bear at least six aldehyde groups and be appended by six building blocks. Consequently, the building blocks linked to the scaffold will act as interaction points for the hexasaccharide. It is pivotal that all the scaffold carbonyl groups display the same reactivity towards the imine exchange in order to avoid biases in the library composition. Therefore, the same scaffold cannot contain a combination of aliphatic and

aromatic aldehydes due to their different reactivities. Thus, the scaffold must be formed by equivalent units which can be defined as repeating unit. A repeating unit is a unit bearing the aldehyde moiety and it is repeated a number of times equal to the number of sugar units present in the glycan. In order to develop a universal system that can be tuned for different templates, the synthesis of scaffolds with different number of repeating units must be easy and straightforward. Polymers[50, 58], peptides[59] and oligonucleotides[60-61] can all be synthesised by well-established methodologies and have already found application in DCC, though not in the field of glycan recognition. Due to their analogy with lectins, peptides were selected as scaffolds for the target-driven DCC. The peptide backbone consists of aspartic acids residues to which benzaldehyde units are linked via the Asp side chain, thus forming the peptidomimetic scaffold. The aspartic acid backbone was preferred to other amino acids as it allows straightforward functionalisation with the aldehyde moiety via amide coupling.

The benzaldehyde moiety was preferred to aliphatic aldehydes as they provide more stable imines[57]. Indeed, aliphatic aldehydes tend not to react with amines, as in aqueous solution the hydration competes with the imine formation[56]. Therefore, the scaffold repeating unit consists of an aspartic acid residue bearing a benzaldehyde unit in α position. Moreover, it must be noted that the repeating units at the N and C termini differ slightly in their structures from the central repeating units. The N and C terminal residues are acetylated and esterified, respectively. The protection of the N terminal is required in order to prevent the amine group partaking in intermolecular and intramolecular imine exchange which would result in the formation of chains of peptides or cyclic peptides, respectively. Various stable protections in aqueous conditions are available at the N terminal, such as *t*-butyl and benzyl carbamate (Boc and Cbz, respectively), and benzylation (Bn). The acetylation of the N terminal was preferred to minimise the steric hindrance. Additionally, the C terminal carboxylic acid requires protecting

during the synthetic procedure to prevent side reactions at this end. Thus, the C terminal was methyl esterified as this group can be easily hydrolysed to reinstate the acid moiety, once the synthesis is complete. Alternatively, the esterified scaffold can be directly employed in the DCC experiment. It must be noted that due to the methyl ester being hydrolysed at pH ≥ 7.4, this group is only suitable for studies under acidic conditions. If an esterified scaffold is to be employed at neutral or basic pH the benzyl group is more appropriate. Other derivatisations of the C terminal are also possible, such as addition of a charged ammonium ion head which could improve the scaffold solubility if required.

The number of repeating units forming the scaffold is defined by the template and this consequently determines the number of adducts generated, for a given set of BBs. For instance, a set of five BBs binding a scaffold with two aldehyde units will afford 15 adducts (Figure 4.16), whilst if the number of aldehyde units was equal to three, 75 adducts will be formed.

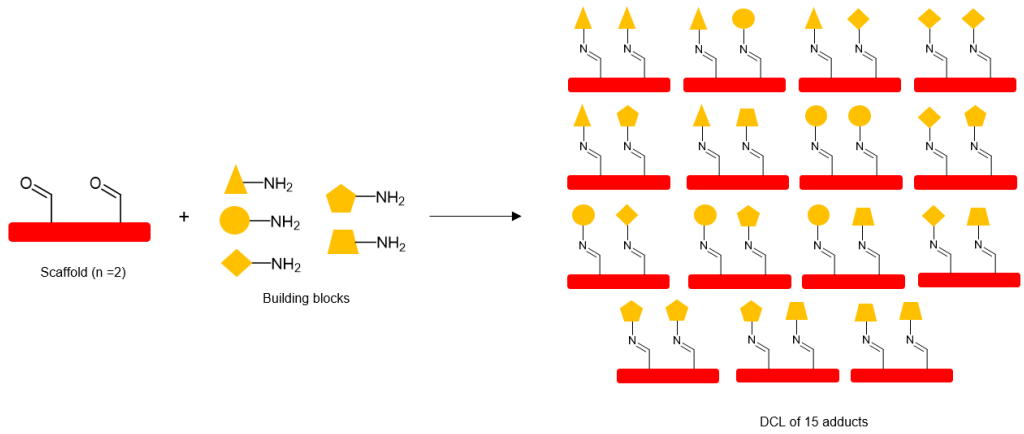


Figure 4.16. Schematic representation of a dynamic combinatorial library (DCL) resulting from the imine exchange of a scaffold with n = 2 and 5 different amine building blocks.

The number of adducts in the library significantly impact the DCL analysis by HPLC-MS. In order to evaluate the amplification of the adducts, via the relative peak area, a good separation

of the peaks is pivotal. Achieving a good separation in large libraries is particularly challenging, and usually only small libraries (5 - 50 adducts) are studied[22]. Moreover, the HPLC-MS analysis, with electrospray ionisation (ESI) provides a *m/z* ratio, corresponding to the parent ion for each adduct allowing the unequivocal identification of adducts with unique molecular weight. Given this, the structure of dipeptide adducts, which can be considered symmetric, can be established by their *m/z* ratio. However, scaffolds with more than three repeating units will generate adducts which are positional isomers presenting the same *m/z* ratio (Figure 4.17). Therefore, when n ≥ 3 additional studies, more advance techniques such as MS-MS analysis, are required to elucidate the structure of the adducts of interest by analysis of their fragmentation patterns[62].

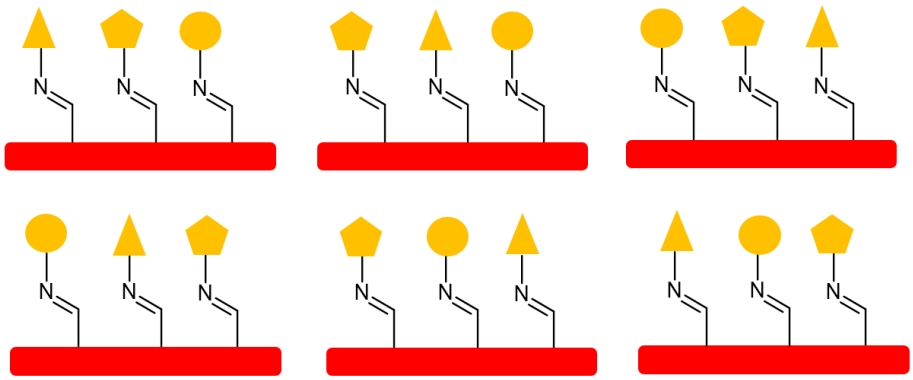


Figure 4.17. Schematic representation of adducts which are isomers with the same molecular weight.

These considerations go beyond the scope of this work, the aim of which is to validate the applicability of DCC for the target-driven generation of glycans receptors. For this reason, initial studies were performed on a simplified system to validate and investigate the optimisation of the proposed approach. Consequently, a dipeptide-based system was initially studied, as its adducts can be easily and unambiguously identified. For this reason, as a first

instance the SAα2-3LacNAc template was not employed, as this will require a tripeptide as scaffold, but a disaccharide template was employed instead (see Section 4.2.4).

4.2.3 Dipeptidomimetic scaffold synthesis

The dipeptidomimetic scaffold presented herein consists of a peptide backbone, formed by two aspartic acid residues, to which aromatic aldehydes were appended. This scaffold was entirely synthesised in solution, which is suitable for small peptide backbones, whilst larger peptides can be obtained via automatic solid-state synthesis and then derivatised in solution. The dipeptide backbone and the benzaldehyde derivatives were synthesised separately and then linked via amide coupling to give the peptidomimetic scaffold (Figure 4.18). In order to partake in the coupling, with the scaffold carboxylic acid groups, the benzaldehyde must bear a primary amine. In addition, the benzaldehyde presented the aldehyde moiety masked, as the corresponding acetal, to avoid side reactions during the synthesis. The aldehyde group was then restored only after coupling to the backbone.

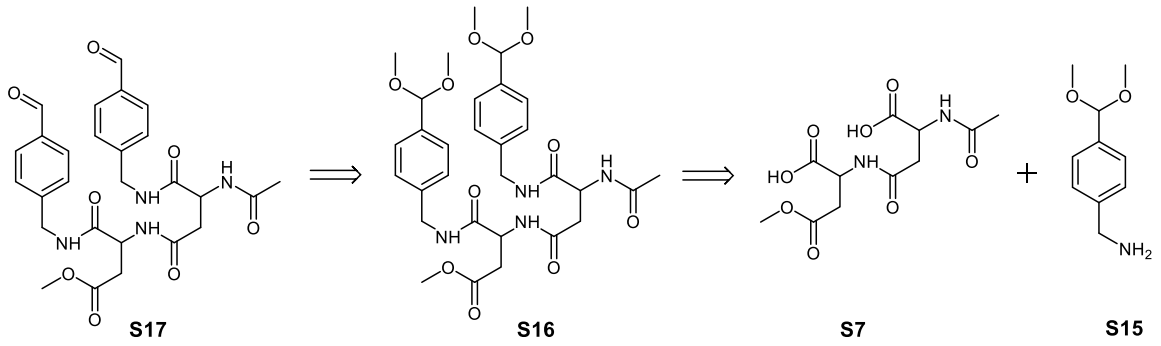


Figure 4.18. Retrosynthetic scheme for the synthesis of the dipeptidomimetic scaffold **S17**.

The peptide backbone was synthesised via amide coupling of two different aspartic acid derivatives, which are both benzylated in α position (Figure 4.19). The Asp derivative situated at the C terminal displayed the γ carboxylic acid methylated and the free amine group. On the

other hand, the N-terminal amino acid was acetylated in the amine position, whilst the γ carboxylic acid was unprotected. Consequently, the coupling of two aforementioned Asp derivatives affords the dipeptide backbone, which is debenzylated prior to benzaldehyde functionalisation.

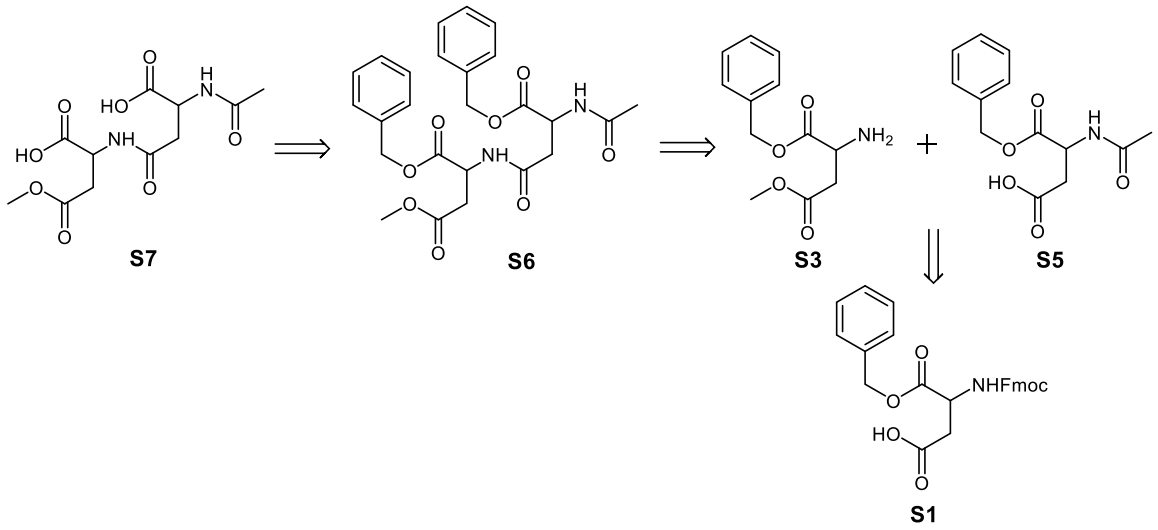


Figure 4.19. Retrosynthetic scheme for the synthesis of the dipeptide backbone **S7**.

The aspartic acid derivatives **S3** and **S5** were synthesised from the commercially available precursor N-α-Fmoc-L-aspartic acid α-benzyl ester (**S1**). The **S1** presented the α carboxylic acid and the amine groups protected with a benzyl and a Fmoc group, respectively. A different aspartic acid derivative, with the carboxylic acid protected with an allyl group, was initially trialled. However, the deallylation of the Fmoc-Asp(OAll)-OH was troublesome. Both a soluble and a resin-bound form of the tetrakis(triphenylphosphine)palladium(0) catalyst were trialled without affording full deprotection. Consequently, the Fmoc-Asp-OBzl (**S1**) was selected on the basis that debenzylation reactions, performed with hydrogen and the Pd/C catalyst, are usually a high yielding and clean.

From Fmoc-Asp-OBzl both **S3** and **S5** were derived. The γ carboxylic acid was esterified with methanol in the presence of a catalytic amount of DMAP to provide the pure

N-Fmoc-1-(phenylmethyl)-4-methylester-L-aspartic acid (**S2**) with 96.1% yield (Figure 4.20). Subsequently, the Fmoc group was cleaved in 20% piperidine in DMF and the resulting crude purified by column chromatography with 50% hexane and 50% ethyl acetate eluent system, to give the pure 1-(phenylmethyl)-4-methylester-L-aspartic acid (**S3**), 89.3%.

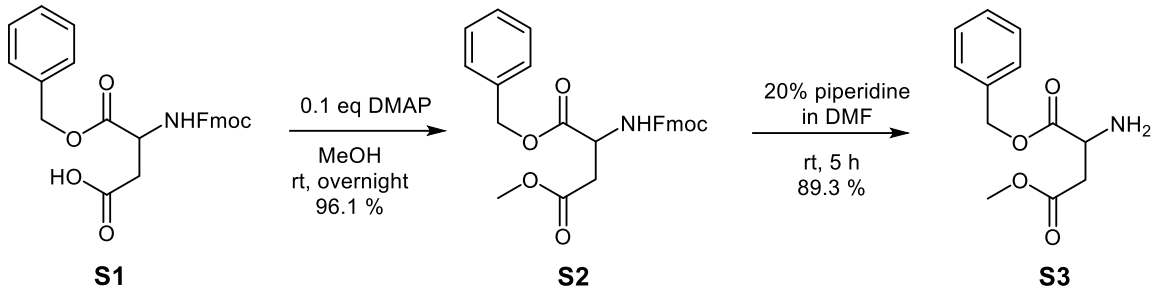


Figure 4.20. Synthesis of aspartic acid derivative **S3**.

The second aspartic acid derivative, **S5**, was also synthesised from the starting material **S1** (Figure 4.21). Initially, the Fmoc group was cleaved in 20% piperidine in DMF to afford the pure 1-(phenylmethyl)-L-aspartic acid **S4** (86.2%). Consequently, the free amine of **S4** was acetylated with acetic anhydride to quantitatively yield the pure N-acetyl-1-(phenylmethyl)-L-aspartic acid **S5**.

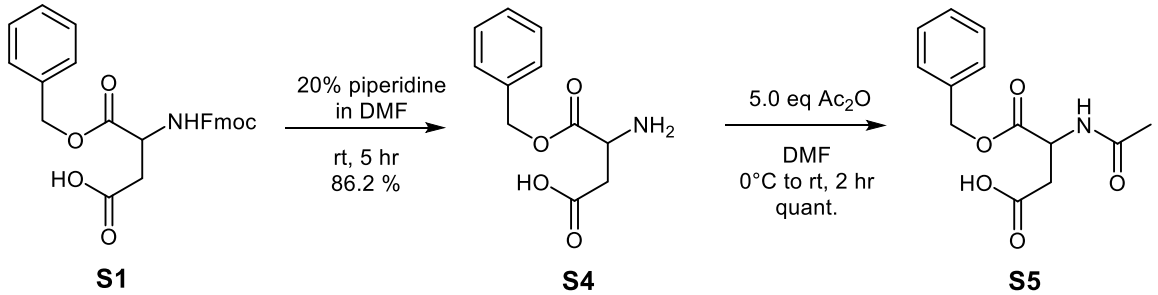


Figure 4.21. Synthesis of aspartic acid derivative **S5**.

The coupling of the two derivatives **S3** and **S5** was performed with standard coupling reagents, 1-hydroxybenzotriazole (HOBt) and 1-ethyl-3-(3′-dimethylaminopropyl)-carbodiimide hydrochloride (EDCI). The EDCI activates the carboxylic acid by firstly transferring a proton from the carboxylic acid and secondly forming a *O*-acylisourea intermediate with the carboxylate (Figure 4.22)[63]. EDCI was preferred to other carboimidines, such as N,N'-dicyclohexylcarbodiimide, as its urea by-product is water soluble and can be efficiently removed by aqueous washing. The HOBt auxiliary was added in order to activate the carboxylic acid and prevent the epimerization of the α-stereocenter[63]. In addition, HOBt also prevents the conversion of the active O-acylisourea into the inactive N-acylurea which is responsible for low yields.

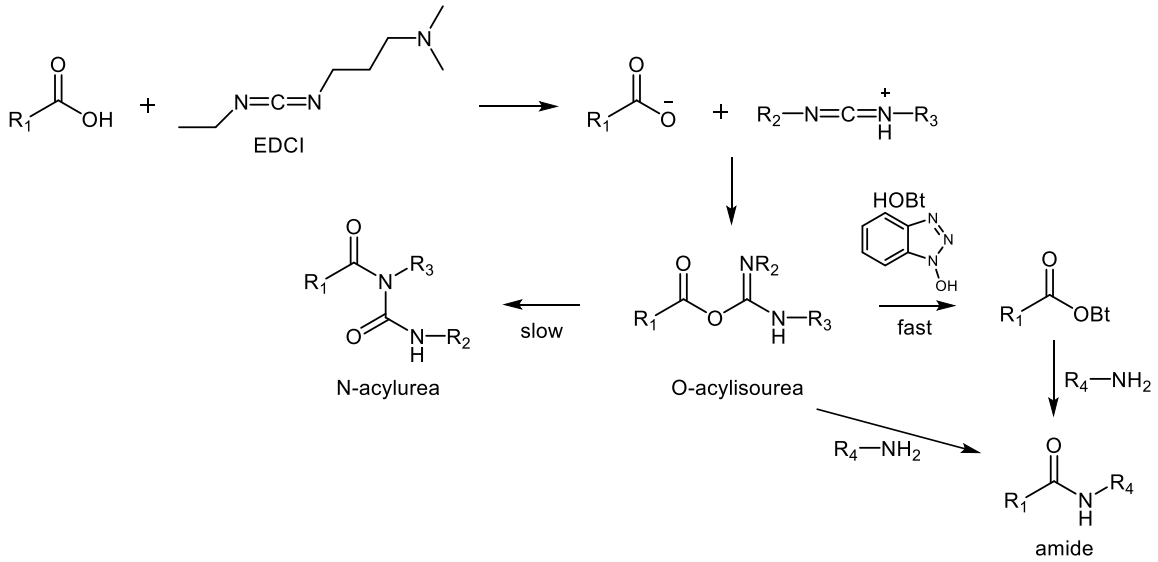


Figure 4.22. Amide coupling mechanism driven by EDCI and HOBt.

The crude resulting from the coupling reaction was purified by column chromatography with a gradient starting from 50% hexane 50% ethyl acetate to 35% hexane 65% ethyl acetate. The pure benzylated dipeptide **S6** was obtained with a yield of 92.5% (Figure 4.23). Compound **S6** was debenzylated with H_2 and Pd/C to afford the pure dipeptide backbone **S7** in quantitative yield.

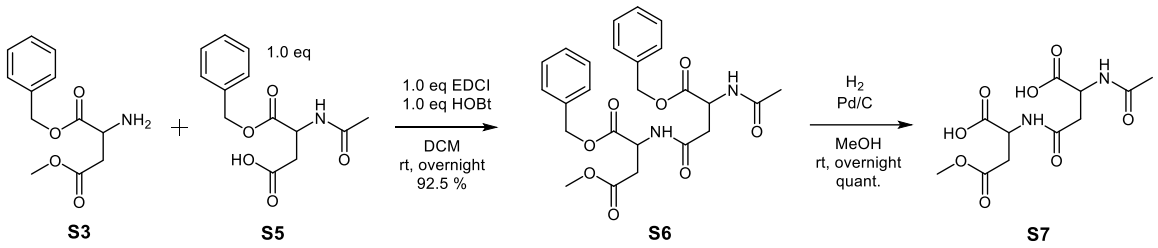


Figure 4.23. Synthesis of dipeptide backbone **S7**.

The benzaldehyde molecule, which was appended to the dipeptide, was also synthesised. This molecule must present a free amine group in order to be coupled to the backbone. In addition, the aldehyde group must be protected as an acetal to prevent any possible side reactions. The dimethyl acetal was selected, as the protecting group, due to easy cleavage under acidic condition.

Initially, the synthesis of 4-(dimethoxymethyl)benzenamine **S9** was attempted. The procedure for the synthesis of **S9** consisted of a two-step synthesis from the commercially available 4-nitrobenzaldehyde. The 4-nitrobenzaldehyde was firstly treated with methanol in the presence of a catalytic amount of *p*-toluenesulfonic acid to quantitatively afford the corresponding acetal **S8**. Compound **S9** was then hydrogenated in the presence of Pd/C in order to reduce the nitro group. Unexpectedly, the hydrogenation afforded 4-aminotoluene instead of the desired 4-(dimethoxymethyl)benzenamine **S10** (Figure 4.24).

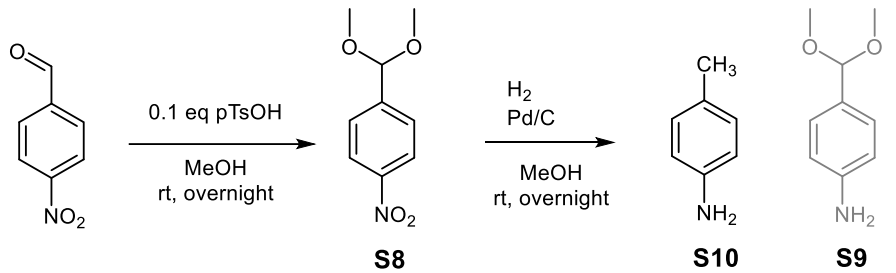


Figure 4.24. Attempted synthesis of 4-(dimethoxymethyl)benzenamine (**S9**) via hydrogenation resulting in the 4-aminotoluene (**S10**) being obtained.

Consequently, a different approach was designed concluding in the coupling of the 4-(dimethoxymethyl)benzylamine **S15** to the peptide backbone. Compound **S15** was synthesised in a 4-step synthetic route from the 4-(bromomethyl)benzonitrile **S11**. The nitrile group was reduced to the aldehyde with diisobutylaluminium hydride (DIBAL). DIBAL acts by partially reducing the nitrile group to the imine (Figure 4.25) which is then hydrolysed to give the corresponding aldehyde[64], to afford the pure **S12**, 63.2%.

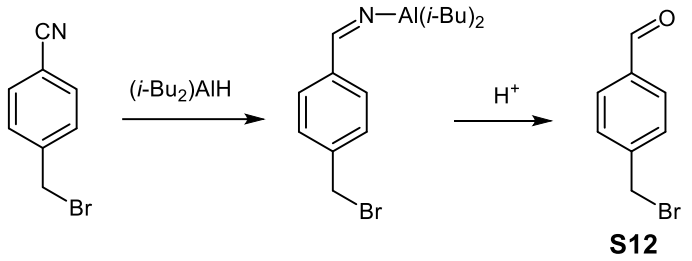


Figure 4.25. Reduction of nitrile with DIBAL to the corresponding imine and subsequent acid hydrolysis to afford the aldehyde group.

The aldehyde moiety was protected as previously described and the crude purified by column chromatography to afford the pure 1-(bromomethyl)-4-(dimethoxymethyl)benzene **S13**, 57.7% (Figure 4.26). Subsequently, the bromo halide was substituted, via an SN_2 reaction, with sodium azide to give the corresponding 1-(azidomethyl)- 4-(dimethoxymethyl)benzene **S14** (81.5%).

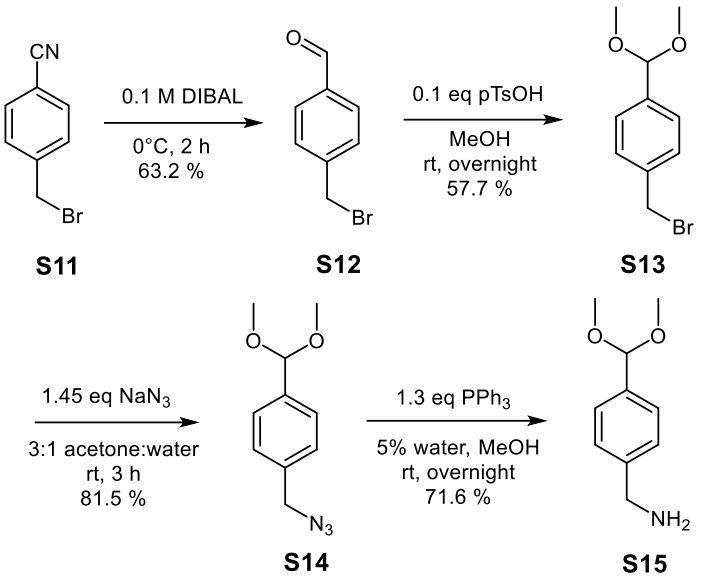


Figure 4.26. Synthesis of amino-functionalised benzaldehyde unit **S15**.

The azide group was reduced with triphenylphosphine via the Staudinger reaction[65]. The triphenylphosphine reacts with the azide in a mixture of methanol and water to generate a phosphazide which is followed by elimination of N_2 to give the iminophosphorane (Figure 4.27)[66]. The hydrolysation of iminophosphorane affords the primary amine.

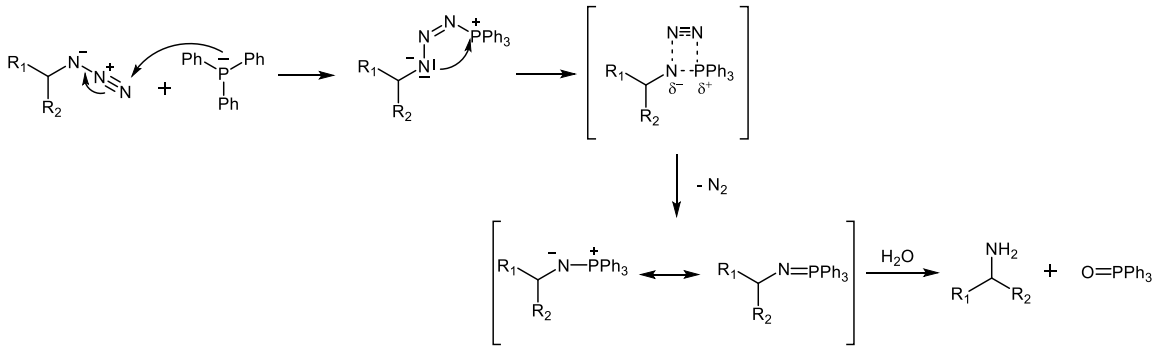


Figure 4.27. Staudinger reaction mechanism.

In order to separate the 4-(dimethoxymethyl)benzylamine **S15** from the unreacted triphenylphosphine the crude was diluted in chloroform and washed with pH 3 phosphate buffer. In this fashion, the protonated **S15** and entered the aqueous phase, whilst the PPh_3 was separated in the organic layer. The water phase was then basified to pH 12 to deprotonate the

amine and the 4-(dimethoxymethyl)benzylamine **S15** was extracted with chloroform and obtained as pure with a yield of 71.6%.

The 4-(dimethoxymethyl)benzylamine **S15** was then coupled to the peptide backbone (Figure 4.28). The coupling reaction was initially attempted with 1.10 equivalent of amine and 1.05 equivalent of HOBt and EDCI per carboxylic acid unit. Under these conditions only partial functionalisation was observed with unreacted **S7** still present. In order to promote the full functionalisation of the peptide the equivalents of **S15** employed were increased to 1.20 per COOH group. After overnight stirring, the reaction was analysed by TLC and the presence of unreacted carboxylic groups was still detected by staining with bromocresol green. Thus, 1.20 equivalent of N,N-diisopropylethylamine (DIEA) per COOH were added to the reaction mixture. DIEA acts as a proton scavenger to maintain the primary amine deprotonated and allow this to partake in the coupling with the acid[67]. Following overnight stirring, the TLC analysis showed no staining with bromocresol green suggesting full functionalisation of the carboxylic acid residues. The crude was taken up in dichloromethane and washed with water and brine. The resulting material was recrystallised in a mixture of ethyl acetate and methanol to give the pure acetal protected peptidomimetic **S16**, 84.9% yield. The acetal protection was then cleaved by acidic treatment with 2 M HCl in diethyl ether to quantitatively afford the pure peptidomimetic **S17**.

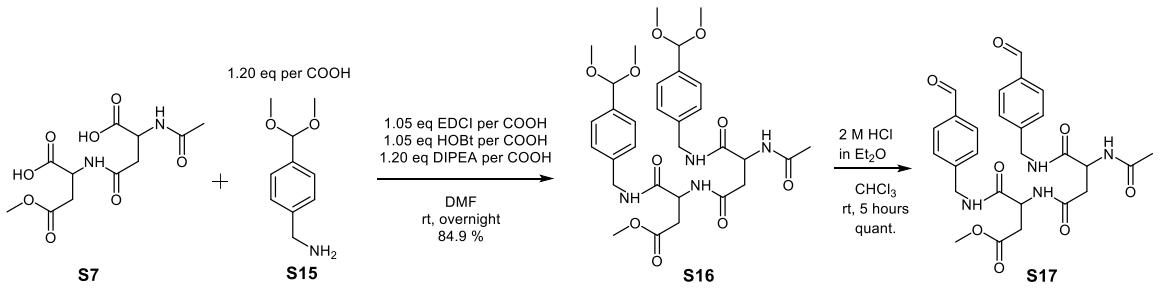


Figure 4.28. Synthesis peptide scaffold **S17**.

4.2.4 Template selection

The dipeptidomimetic scaffold detailed in Section 4.2.2 will provide two points of interaction for glycan recognition and therefore is suitable to target disaccharide templates. It is expected that, given the same set of BBs, different templates will provide different DCL compositions by amplifying different adducts[24]. In order to probe this, three disaccharide templates differing greatly in their structure were selected (Figure 4.29); lactose (**T1**), lactal (**T2**) and lactulose (**T3**).

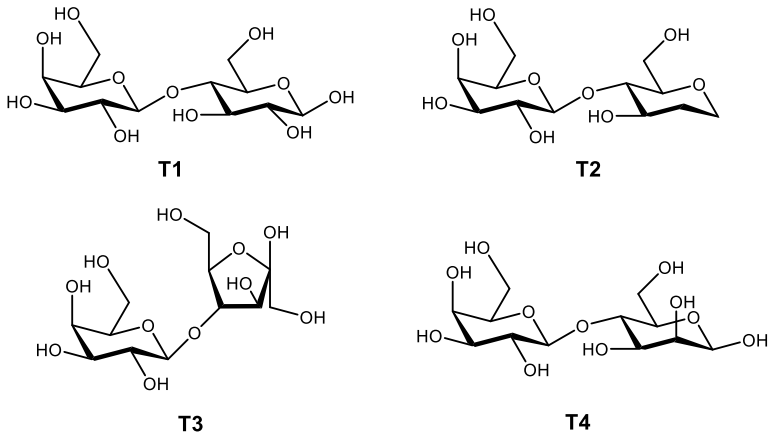


Figure 4.29. Structure of disaccharide templates; D-lactose (**T1**), D-lactal (**T2**), D-lactulose (**T3**), D-epilactose (**T4**).

The templates present a galactopyranoside unit β linked to a second sugar unit in which they differ greatly. In lactose the second unit consists of a glucose residue, whilst lactal displays a 1,2-dideoxy glucose derivative. Furthermore, lactose and lactal comprise exclusively of pyranose rings, whilst lactulose presents the galactopyranoside linked to the furanose ring of fructose. As the three templates differ greatly in their structure, they are expected to provide amplification of distinct adducts. For instance, the lack of two hydroxyl groups in the lactal might deprive this saccharide of certain non-covalent interactions, such as hydrogen bonds, that would otherwise be present in lactose, resulting in unique amplified binders. Furthermore, boronic acids, which are included in the building block sets (see Section 4.2.6) present different

reactivities towards pyranose and furanose rings and possibly determining different adduct amplification between lactose and lactulose. Moreover, the DCC approach should not only discriminate amongst templates with large structural differences but should also be appreciative of more subtle differences between templates. Discriminating between positional isomers, such as the cancer-related α2-3 sialylated glycans against the α2-6 isoforms, is of great relevance in the development of diagnostic probes. Given the challenges presented by employing the DCC platform for trisaccharide templates (see Section 4.2.2), the ability to discriminate positional isomers must be, as a first instance, tested on disaccharides. For this scope, lactose and its isomer epilactose (**T4**, Figure 4.29), were selected. These templates are C2' epimers differing solely in the stereochemistry of the hydroxyl group in position 2'.

Within the DCC experiments the template is generally added to the library in stoichiometric amounts to the scaffold concentration[22]. However, lower concentrations of template can enhance the competition between adducts increasing the prospect of affording the best binder for the given template[68]. This might be required for templates which are structurally very similar, such as lactose and epilactose. Furthermore, it is important to verify that the shift in the thermodynamic equilibrium is ascribable to the molecular recognition between template and amplified adduct. This can be verified by utilising a higher concentration of template which will result in a higher degree of amplification if adduct amplification is triggered by molecular recognition[69].

4.2.5 Building blocks selection

In the development of a DCC method, as described by Huc and Lehn, building blocks known for their affinity to the template are often employed[21]. In this fashion, the DCC method can be

optimised and validated. It is only after the optimisation of the method that different building blocks can be trailed to identify new binder species. Due to the novelty of the work presented herein, in the first instance it was important to validate the approach, thus BBs containing sugar binding units were selected. A series of BBs bearing sugar-binding groups and representing a wide chemical space were selected in order to maximise the molecular recognition with the template[21]. The selected building blocks included aromatic, heterocyclic, linear, hydrophobic, polar and charged compounds. In addition, boron-based molecules were also introduced as BBs (see Section 4.2.6).

All building blocks required a primary amine in order participate in the imine exchange with the aldehyde-appended scaffold. Aliphatic and aromatic amines present different reactivity towards the imine exchange[56, 70]. Aromatic amines are less reactive in the imine exchange, compared to the aliphatic counterpart, due to their lower nucleophilicity. More nucleophilic amines present higher formation constants as they afford more stable imines. However, some studies suggested that the opposite is true, with fully protonated aliphatic amines driving the imine hydrolysis[57]. Due to the contradicting evidence both set of amines, aromatic and aliphatic, were selected for the validation studies. In order to avoid biases, amines with different reactivity were segregated in different sets of BBs. Hence, two different sets of BBs were designed, one containing solely aliphatic amines and the other aniline-like molecules. The two libraries of BBs will be probed in separate DCC experiments.

In the BB selection process, there are additional factors that require consideration. For instance, building blocks must be soluble in aqueous media. If required, small percentages (5 - 10%) of organic solvents (*i.e.* DMSO) can be employed without altering the library composition[52, 71-72]. It is in fact crucial to maintain all the library components (*i.e.* scaffold, BBs and adducts) in solution as the precipitation of one or more compounds will shift the library equilibrium and

generate a bias[52]. In addition, for the library to be isoenergetic and unbiased, the building blocks must be added in equimolar concentrations with the equivalents of each BB being at least equal to the number of aldehyde groups[21],[52, 68]. For instance, the dipeptidomimetic scaffold requires at least 2 equivalents of each amine, one per aldehyde group, in order to allow the formation of all possible adducts. Furthermore, the building blocks of a given set must present different molecular weights in order to allow unequivocal identification of the adducts via their *m/z* ratio following HPLC-MS analysis[51]. The analysis by HPLC-MS is generally limited to small libraries (5 – 50 members) as it is more challenging to achieve a good separation of peaks in the chromatogram for larger libraries[22]. For the scope of this study, BBs were divided in sub-sets containing less than 5 compounds. These sub-sets can be independently tested in the DCC experiments. The BBs included in the amplified adducts for each sub-set will then be part of an advance sub-set which will be further tested. The above will be repeated until ultimately the best adduct for the given target will be obtained. This will then be independently synthesised and tested by isothermal titration calorimetry to assess the binding affinity.

The selected BBs, containing groups which are known to provide non-covalent interactions with carbohydrates, are presented and divided based on the aliphatic or aromatic nature of the amine group. Boronic acids and their analogues, which bind covalently to sugars will be part of their corresponding aromatic or aliphatic libraries along with the other building blocks and are described in Section 4.2.6. The BBs selection was inspired by lectins and synthetic receptors. Natural and synthetic lectins frequently comprise of aromatic groups, through which they generate CH-π interactions with the target[18]. For this reason, an array of aromatic compounds was selected as BBs. The BBs were divided into aromatic and aliphatic sets based solely on the nature of the amine group, and not for the presence or lack of an aromatic moiety. Indeed, aniline-like molecules will be segregated into the aromatic set, whilst aromatic or

non-aromatic compounds bearing an aliphatic amine will be part of the aliphatic set. The aliphatic and aromatic sets will then be divided in sub-sets to allow easier DCL analysis (Figure 4.30). Aliphatic and aromatic BBs will never be mixed in the same set or sub-set given the different reactivity of the two amines. Whenever possible, the two sets of libraries will contain analogous molecules. For instance, aniline (**AR1**) was selected to mimic lectins phenylalanine residues and provide CH-π interactions[14]. Consequently, the aliphatic analogue, benzylamine (**AL1**), was introduced in the aliphatic BBs set. Heterocyclic building blocks were introduced in the libraries, as these retain the hydrophobic nature but also offer an additional point of interaction via the heteroatom. The molecule 2-aminopyridine (**AR2**) was selected, as it was previously shown to promote interaction with saccharides by mimicking the asparagine and glutamine side chains of lectins[73]. Consequently, the aliphatic analogue 2-(aminomethyl)pyridine (**AL2**) was introduced in the aliphatic set. Bicyclic heteroaromatic compounds 2-quinolinamine (**AR3**) and 2-(aminomethyl)quinoline (**AL3**),were introduced in their respective sets as this moiety was shown to interact with sugars[74]. The analogues 2-naphthalenamine (**AR4**) and 1-(2-naphthyl)methanamine (**AL4**) were instead selected to assess the relevance of the nitrogen proton acceptor of quinoline in the sugar interaction. Indeed, building blocks which are not expected to offer advantageous interactions compared to other library members were also introduced to probe the role of certain functional groups in the binding. Furthermore, it is important that analogous molecules such as **AR3** and **AR4** will be part of the same initial sub-set in order to evaluate the above-mentioned consideration. The five-membered heterocyclic ring imidazole was included to mimic histidine[75].

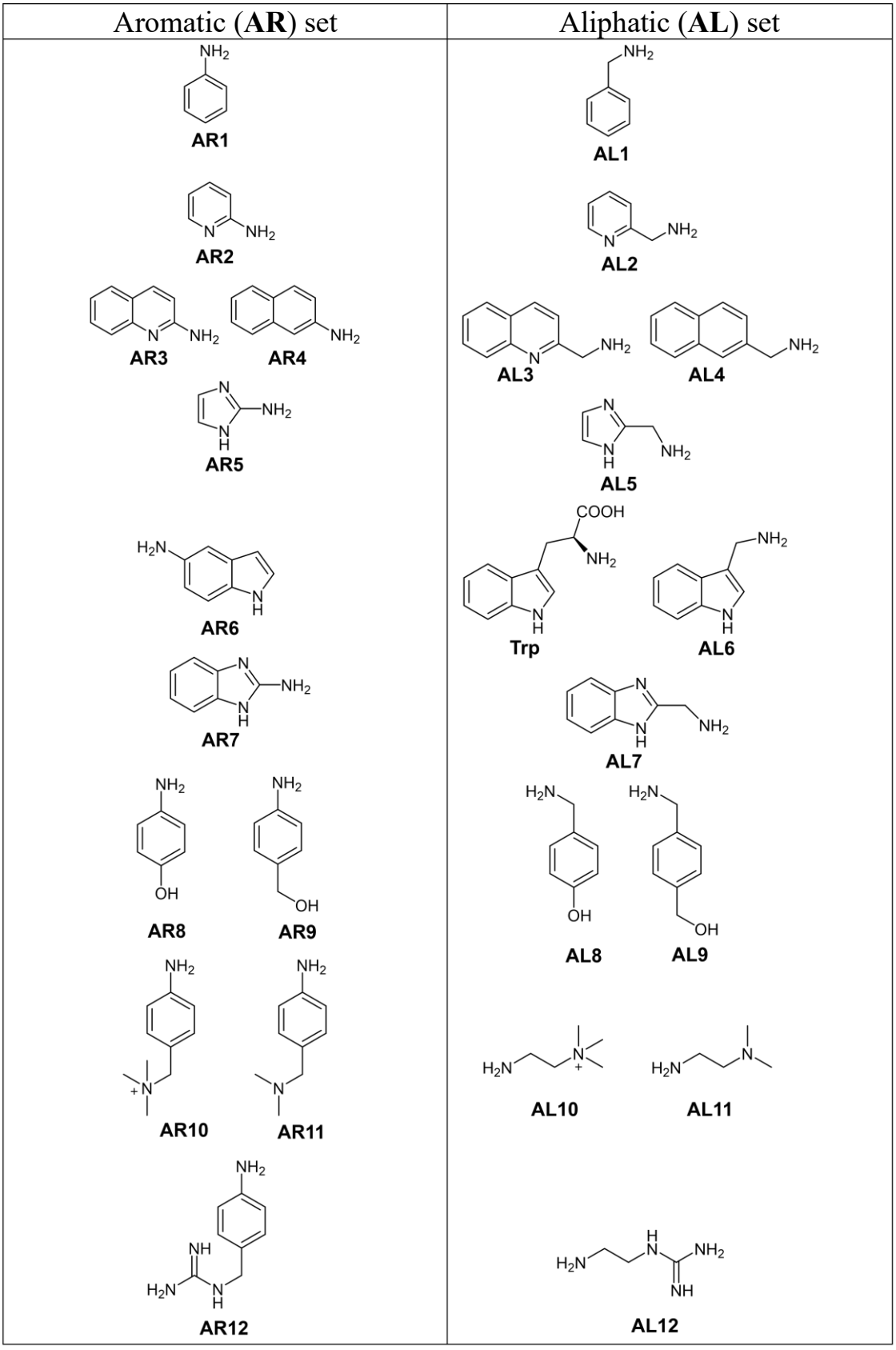


Figure 4.30. Aromatic (**AR**) and aliphatic (**AL**) non-covalent building blocks.

Two imidazole rings bearing a primary amine were selected; the 2-aminoimidazole (**AR5**) and the 2-(aminomethyl)imidazole (**AL5**). The bicyclic imidazole analogue, indole, was also introduced in the libraries. Indole represents the side chain of tryptophan (**Trp**), which is one of the main amino acid residues contributing to the binding of sugars in lectin pockets[14, 76-77]. However, when tryptophan is used as a building block it will present the free carboxylic acid group, unlike in lectins where this is engaged in the peptide backbone. Therefore, an analogue without the acid group, the 3-(aminomethyl)indole (**AL6**), was also included to mimic more closely the tryptophan structure in lectins. The corresponding aromatic analogue, 5-aminoindole (**AR6**), was introduced in the aromatic set. The indole analogues 2-aminobenzimidazole (**AR7**) and the 2-(aminomethyl)benzimidazole (**AL7**) were also included[78].

Additionally, polar and charged building blocks were also introduced. For instance, BBs functionalised with hydroxyl groups were selected to mimic tyrosine residues[79]. These included 4-aminophenol (**AR8**) 4-(hydroxymethyl)aniline (**AR9**) and their corresponding analogues 4-(aminomethyl)phenol (**AL8**) and 4-hydroxymethylbenzylamine (**AL9**). Positively charged building blocks were selected on the basis of their interactions with polar groups on the sugar, as shown in Chapter 3. However, molecules with an ammonium ion, NH_3^+, in addition to the primary amine cannot be introduced in the library as this will partake in the imine exchange and act as a linker between scaffold molecules. The trimethylammonium ion $N(CH_3)_3^+$ was therefore selected as an ammonium ion analogue. There exists, however, a significant difference between the two groups, with the ammonium ion being able to provide both charge-reinforced hydrogen bonds and electrostatic interactions whilst the trimethylammonium ion can only provide the latter. In particular, the 2-amino-N,N,N-trimethylethanaminium (**AL10**) was selected as one of the charged building blocks. This compound does not display

any other functional groups with its inclusion solely to evaluate the effect of a charged group in the DCL formation. The uncharged analogue, N,N-dimethylethylenediamine (**AL11**) was also introduced. The 2-amino-N,N,N-trimethylethanaminium (**AL10**) was synthesised through a 3-step synthetic route (Figure 4.31). The N,N-dimethylethylenediamine (**AL11**) was firstly Boc protected to the tert-butyl 2-(dimethylamino)ethylcarbamate and then methylated to the corresponding trimethylammonium ion in presence of iodomethane to give tert butyl 2-(trimethylamino)ethylcarbamate as iodide salt. The Boc group was then removed in acidic conditions with the addition of acetyl chloride in methanol to afford the pure **AL10** as the dihydrochloride salt (overall yield across 3 steps of 57.0%).

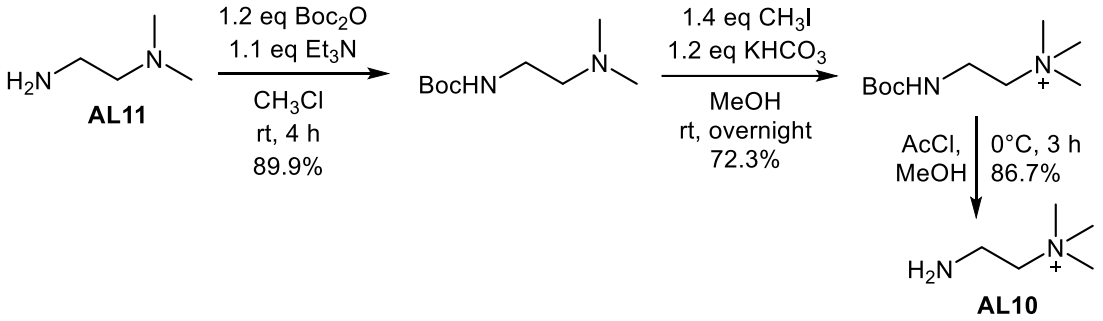


Figure 4.31. Synthesis of aliphatic building block **AL10** from the precursor **AL11**.

Similarly, an aromatic analogue of **AL10**, the 4-amino-N,N,N-trimethyl-benzenemethanaminium (**AR10**), was introduced into the library alongside its uncharged precursor, the 4-amino-N,N-dimethyl-benzenemethanamine (**AR11**). The 4-amino-N,N,N-trimethyl-benzenemethanaminium can be obtained from its precursor in a similar fashion to **AL10**. Furthermore, guanidino analogues to **AL10** and **AR10** were also considered; the N-(2-aminoethyl)-guanidine (**AL12**) and 1-(4-aminobenzyl)guanidine (**AR12**). The guanidino group is particularly relevant in sugar binding as it can offer strong charge-reinforced hydrogen bonds[80-81], as seen in Chapter 3.

4.2.6 Boron-based building blocks

Generally, target-driven DCC is based solely on cooperative non-covalent interactions with the target, with only few examples of adducts binding covalently to the template[82-84]. In carbohydrate recognition covalent interactions can be of great value in sugar recognition, as demonstrated in Chapter 3. For this reason, boron-based compounds were introduced in the BB sets. The boronate esterification has previously found applications in DCC but always as the reversible reaction for the generation of adducts and never as point of interaction with the template[85-89]. For instance, boronic acids were employed as BBs, along with diol BBs, to generate a series of adducts via boronate esterification for protein inhibition[87]. On the other hand, within this work the boron-based BBs are employed to provide molecular recognition with the template, whilst the DCC is based on a different reversible reaction (*i.e.* imine exchange). Nevertheless, the boronate esterification between building blocks and template will constitute a secondary, or orthogonal, equilibrium (Figure 4.32). Therefore, when boron-based compounds are introduced as building blocks two different, orthogonal, equilibria will be present; the imine exchange and the boronate esterification. In this fashion, the number of species present in the DCC mixture will be greater. Each library component, such as scaffold, BB and template, will be present as free species but also engaged with other library components. The adducts generated by imine exchange between the scaffold and the BBs, defined herein as scaffold-based adducts, will be present both in the free form and bound to the template via non-covalent interactions. Furthermore, the scaffold-based adducts containing a boron unit will be covalently bound to the template and therefore governed by both the imine and boronate equilibrium. In addition, free boronic acids, not linked to a scaffold molecule, can independently bind the sugar template generating boronate esters. Nonetheless, these small boronate esters will be less favourable than scaffold-based adducts as the latter provide a higher

degree of multivalency and therefore it is expected that the presence of these aforementioned small boronate esters will be negligible. Despite the many species formed in solution, this will not result in the analysis of highly complex mixtures, as the number of species will be significantly reduced prior analysis. Indeed, the imine exchange will be frozen, via reductive amination, obtaining stable forms of the scaffold-based adducts. Following this, the complexes between the scaffold-based adducts and the sugar will be disrupted to allow more straightforward analysis. This can be performed by changing the pH of the mixture and will result not only in the disruption of the scaffold-based adducts covalently and non-covalently bound to the template, but also of the small boronate ester complexes, if present in significant amounts.

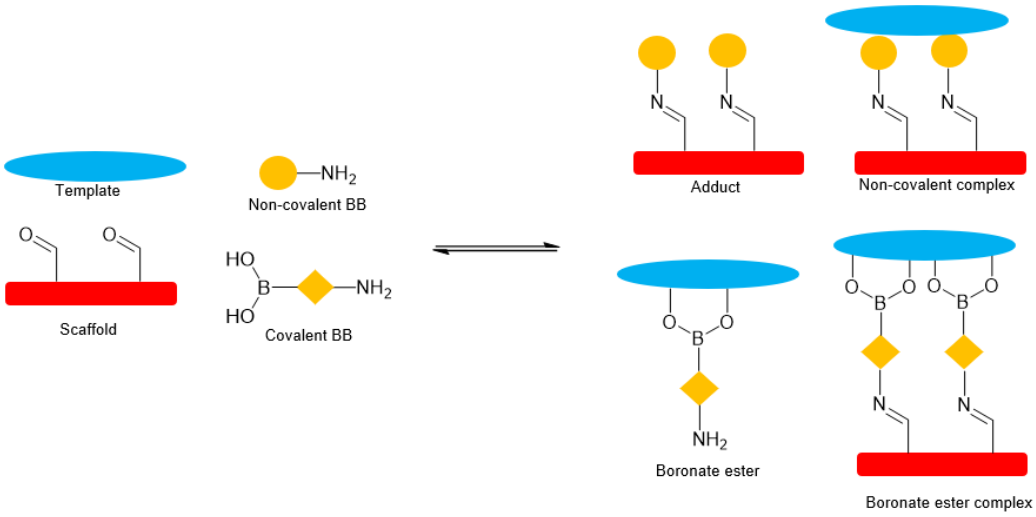


Figure 4.32. Representation all the possible species in solution in a DCC presenting two equilibria; imine exchange and boronate esterification.

Another important consideration when combining the imine exchange and boronate ester formation in the same system, is with regards to the pH. The imine DCC are typically performed at pH between 5.0 and 8.5[57], whilst the boronate equilibrium requires neutral to basic pH[16]. Indeed, the boronate esterification occurs at pH values above the pK_a of the boronic acid

(phenylboronic acid pK_a = 8.86)[90], with the only exception being sialic acid[81]. Thus, a pH > 7.4 is necessary for the boron-based BBs to engage with the disaccharide templates and with sialylated targets, such as SAα2-3LacNAc and SAα2-6LacNAc. In Chapter 3 it was demonstrated that boron-based receptors cannot bind sialic acid in sialosides and therefore the covalent BBs will bind to the other sugar residues for which they require neutral to basic pH. Consequently, studies are necessary to determine the pH at which both the imine exchange and boronate esterification occur significantly in order to avoid biased libraries. If the DCC were performed at acidic pH the adducts will be formed via imine exchange, but the amplified ones will not display boron units as, under these conditions, they do not provide molecular recognition. On the other hand, if a basic pH above 8.5 were employed no scaffold adducts will be obtained via imine exchange, rendering the DCC unfeasible.

In order to participate in the imine exchange with the aldehyde-appended scaffold the boron-based building blocks must contain a primary amine. Two main class of boron-based compounds were considered, phenylboronic acids and benzoboroxoles comprising of both aromatic and aliphatic amines (Figure 4.33).

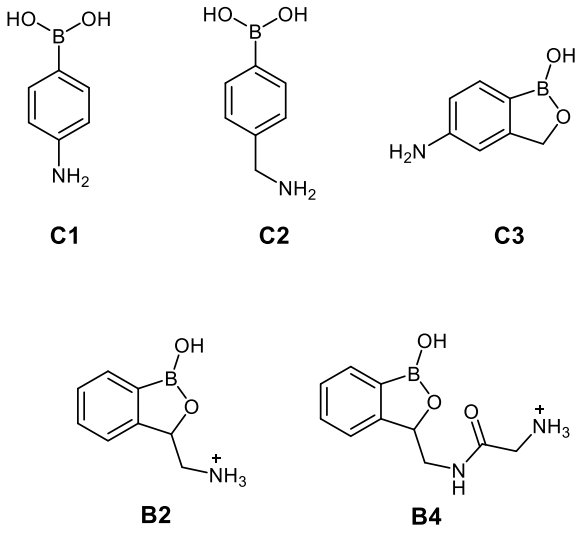


Figure 4.33. Covalent building blocks.

The 4-aminophenylboronic acid (**C1**) and the aliphatic analogue analogues 4-(aminomethyl)phenylboronic acid (**C2**) were selected. Moreover, benzoboroxole compounds were also included, as they are more advantageous than phenylboronic acid derivatives due to their higher binding constants and ability to bind pyranose rings[91-92]. In particular the aromatic 6-aminobenzoboroxole (**C3**) and the aliphatic **B2** were selected. Receptor **B2** presents the amine functionality in close proximity to the boron atom, which could result in steric hindrance when this is bound to the scaffold, for this reason the benzoboroxole **B4** was introduced into the library to overcome the potential steric issues. The phenylboronic acid derivatives **C1** and **C2** and the 6-aminobenzoboroxole **C3** are commercially available, whilst the **B2** and **B4** were synthesised as described in Chapter 3.

4.3 Summary and outlook

The detection of specific cancer-related glycans holds a great potential in diagnostics, however tools to discriminate between closely related glycoforms are limited. In particular, herein the focus is on developing a tool to selectively discriminate the cancer-related α2-3 sialylated glycans over α2-6 glycoforms. It is therefore pivotal, for the tool development, primarily to obtain the sialoside of interest and the corresponding control glycoform. A small epitope consisting of three saccharide units, SAα2-3LacNAc, was identified as representing the terminal end of cancer-related α2-3 sialylated glycans and was obtained via enzymatic synthesis. The enzymatic synthesis consisted of 2 steps with the first affording the disaccharide LacNAc via a B4GalT1 catalysed reaction between the acceptor N-acetylglucosamine and the donor UDP-Galactose. In the second step LacNAc was sialylated in α2-3 and α2-6 fashion with the ST3Gal4 and PmST1 P34H/M144L enzymes, respectively, to obtain the corresponding epitopes SAα2-3LacNAc and SAα2-6LacNAc. The selective recognition of the α2-3

trisaccharide requires scaffold-based receptors, as these can provide a high degree of multivalency with the sialoside. The generation of scaffold-based receptors for carbohydrate recognition was pursued by dynamic combinatorial chemistry, a technique which allows *in situ* generation and selection of the best binder for a given template. Due to the novelty of the DCC application in glycan recognition a series of preliminary studies and constructs were detailed. In particular, a peptidomimetic scaffold was designed and synthesised. A series of building blocks were selected on the basis of their sugar-binding properties and their reactivity towards the aldehyde-appended scaffold. In addition, different templates presenting different degrees of similarity in their structure were selected to be employed in the evaluation of the method selectivity. Reaction and analysis conditions, including possible limitations and issues, were also reviewed. This work can be considered the foundation for the future development of glycan diagnostic tools by dynamic combinatorial chemistry means.

List of references

1. Stanley P, C. R., Structures Common to Different Glycans. *Essentials of Glycobiology. 2nd edition* **2009,** *Chapter 13*.

2. Stanley, P.; Taniguchi, N.; Aebi, M., N-Glycans. *Essentials of Glycobiology* **2017**.

3. Orntoft, T. F.; Vestergaard, E. M., Clinical aspects of altered glycosylation of glycoproteins in cancer. *Electrophoresis* **1999,** *20* (2), 362-371.

4. Ugorski, M.; Laskowska, A., Sialyl Lewis(a): a tumor-associated carbohydrate antigen involved in adhesion and metastatic potential of cancer cells. *Acta Biochim. Pol.* **2002,** *49* (2), 303-311.

5. Cazet, A.; Julien, S.; Bobowski, M.; Krzewinski-Recchi, M.-A.; Harduin-Lepers, A.; Groux-Degroote, S.; Delannoy, P., Consequences of the expression of sialylated antigens in breast cancer. *Carbohydr. Res.* **2010,** *345* (10), 1377-1383.

6. Ohyama, C.; Hosono, M.; Nitta, K.; Oh-eda, M.; Yoshikawa, K.; Habuchi, T.; Arai, Y.; Fukuda, M., Carbohydrate structure and differential binding of prostate specific antigen to Maackia amurensis lectin between prostate cancer and benign prostate hypertrophy. *Glycobiology* **2004,** *14* (8), 671-679.

7. Perez-Garay, M.; Arteta, B.; Pages, L.; de Llorens, R.; de Bolos, C.; Vidal-Vanaclocha, F.; Peracaula, R., alpha 2,3-Sialyltransferase ST3Gal III Modulates Pancreatic Cancer Cell Motility and Adhesion In Vitro and Enhances Its Metastatic Potential In Vivo. *PLoS One* **2010,** *5* (9), 11.

8. Kilpeläinen, T. P.; Tammela, T. L. J.; Roobol, M.; Hugosson, J.; Ciatto, S.; Nelen, V.; Moss, S.; Määttänen, L.; Auvinen, A., False-positive screening results in the European randomized study of screening for prostate cancer. *Eur. J. Cancer* **2011,** *47* (18), 2698-2705.

9. Catalona, W. J.; Smith, D. S.; Ratliff, T. L.; Dodds, K. M.; Coplen, D. E.; Yuan, J. J. J.; Petros, J. A.; Andriole, G. L., Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. *N. Engl. J. Med.* **1991,** *324* (17), 1156-1161.

10. Thompson, I. M.; Pauler, D. K.; Goodman, P. J.; Tangen, C. M.; Lucia, M. S.; Parnes, H. L.; Minasian, L. M.; Ford, L. G.; Lippman, S. M.; Crawford, E. D.; Crowley, J. J.; Coltman, C. A., Prevalence of prostate cancer among men with a prostate-specific antigen level <= 4.0 ng per milliliter. *N. Engl. J. Med.* **2004,** *350* (22), 2239-2246.

11. Catalona, W. J.; Richie, J. P.; Ahmann, F. R.; Hudson, M. A.; Scardino, P. T.; Flanigan, R. C.; Dekernion, J. B.; Ratliff, T. L.; Kavoussi, L. R.; Dalkin, B. L.; Waters, W. B.; Macfarlane, M. T.; Southwick, P. C., Comparison of Digital Rectal Examination and Serum Prostate

Specific Antigen in the Early Detection of Prostate Cancer: Results of a Multicenter Clinical Trial of 6,630 Men. *J. Urol.* **1994,** *151* (5), 1283-1290.

12. Quiocho, F. A., Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions. *Annu. Rev. Biochem.* **1986,** *55*, 287-315.

13. Weis, W. I.; Drickamer, K., Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* **1996,** *65*, 441-473.

14. Hudson, K. L.; Bartlett, G. J.; Diehl, R. C.; Agirre, J.; Gallagher, T.; Kiessling, L. L.; Woolfson, D. N., Carbohydrate-Aromatic Interactions in Proteins. *J. Am. Chem. Soc.* **2015,** *137* (48), 15152-15160.

15. Badjic, J. D.; Nelson, A.; Cantrill, S. J.; Turnbull, W. B.; Stoddart, J. F., Multivalency and cooperativity in supramolecular chemistry. *Accounts Chem. Res.* **2005,** *38* (9), 723-732.

16. Wu, X.; Li, Z.; Chen, X. X.; Fossey, J. S.; James, T. D.; Jiang, Y. B., Selective sensing of saccharides using simple boronic acids and their aggregates. *Chem. Soc. Rev.* **2013,** *42* (20), 8032-8048.

17. Mazik, M., Molecular recognition of carbohydrates by acyclic receptors employing noncovalent interactions. *Chem. Soc. Rev.* **2009,** *38* (4), 935-956.

18. Davis, A. P.; Wareham, R. S., Carbohydrate recognition through noncovalent interactions: A challenge for biomimetic and supramolecular chemistry. *Angew. Chem. Int. Ed.* **1999,** *38* (20), 2978-2996.

19. Tommasone, S.; Tagger, Y. K.; Mendes, P. M., Targeting Oligosaccharides and Glycoconjugates Using Superselective Binding Scaffolds. *Adv. Funct. Mater.* **2020**, 2002298.

20. Ulrich, S.; Dumy, P., Probing secondary interactions in biomolecular recognition by dynamic combinatorial chemistry. *Chem. Commun.* **2014,** *50* (44), 5810-5825.

21. Huc, I.; Lehn, J. M., Virtual combinatorial libraries: Dynamic generation of molecular and supramolecular diversity by self-assembly. *Proc. Natl. Acad. Sci. U. S. A.* **1997,** *94* (6), 2106-2110.

22. Frei, P.; Hevey, R.; Ernst, B., Dynamic Combinatorial Chemistry: A New Methodology Comes of Age. *Chem. Eur. J.* **2019,** *25* (1), 60-73.

23. Li, J.; Nowak, P.; Otto, S., Dynamic Combinatorial Libraries: From Exploring Molecular Recognition to Systems Chemistry. *J. Am. Chem. Soc.* **2013,** *135* (25), 9222-9239.

24. Rauschenberg, M.; Bomke, S.; Karst, U.; Ravoo, B. J., Dynamic Peptides as Biomimetic Carbohydrate Receptors. *Angew. Chem. Int. Ed.* **2010,** *49* (40), 7340-7345.

25. Yang, D.; von Krbek, L. K. S.; Yu, L.; Ronson, T. K.; Thoburn, J. D.; Carpenter, J. P.; Greenfield, J. L.; Howe, D. J.; Wu, B.; Nitschke, J. R., Glucose Binding Drives Reconfiguration of a Dynamic Library of Urea-Containing Metal–Organic Assemblies. *Angew. Chem. Int. Ed.* **2021,** *60* (9), 4485-4490.

26. Boons, G. J.; Demchenko, A. V., Recent advances in O-sialylation. *Chem. Rev.* **2000,** *100* (12), 4539.

27. Ranade, S. C.; Demchenko, A. V., Mechanism of Chemical Glycosylation: Focus on the Mode of Activation and Departure of Anomeric Leaving Groups. *J. Carbohydr. Chem.* **2013,** *32* (1), 1-43.

28. Komura, N.; Kato, K.; Udagawa, T.; Asano, S.; Tanaka, H.-N.; Imamura, A.; Ishida, H.; Kiso, M.; Ando, H., Constrained sialic acid donors enable selective synthesis of α-glycosides. *Science* **2019,** *364* (6441), 677-680.

29. Schelch, S.; Zhong, C.; Petschacher, B.; Nidetzky, B., Bacterial sialyltransferases and their use in biocatalytic cascades for sialo-oligosaccharide production. *Biotechnol. Adv.* **2020,** *44*, 107613.

30. Unverzagt, C.; Kunz, H.; Paulson, J. C., High-efficiency synthesis of sialyloligosaccharides and sialoglycopeptides. *J. Am. Chem. Soc.* **1990,** *112* (25), 9308-9309.

31. Yu, H.; Chokhawala, H.; Karpel, R.; Yu, H.; Wu, B. Y.; Zhang, J. B.; Zhang, Y. X.; Jia, Q.; Chen, X., A multifunctional Pasteurella multocida sialyltransferase: A powerful tool for the synthesis of sialoside libraries. *J. Am. Chem. Soc.* **2005,** *127* (50), 17618-17619.

32. Sugiarto, G.; Lau, K.; Qu, J.; Li, Y.; Lim, S.; Mu, S.; Ames, J. B.; Fisher, A. J.; Chen, X., A Sialyltransferase Mutant with Decreased Donor Hydrolysis and Reduced Sialidase Activities for Directly Sialylating Lewisx. *ACS Chem. Biol.* **2012,** *7* (7), 1232-1240.

33. McArthur, J. B.; Yu, H.; Zeng, J.; Chen, X., Converting Pasteurella multocida alpha 2-3-sialyltransferase 1 (PmST1) to a regioselective alpha 2-6-sialyltransferase by saturation mutagenesis and regioselective screening. *Org. Biomol. Chem.* **2017,** *15* (7), 1700-1709.

34. Brockhausen, I., Crossroads between Bacterial and Mammalian Glycosyltransferases. *Front Immunol* **2014,** *5*, 492-492.

35. Broszeit, F.; Tzarum, N.; Zhu, X.; Nemanichvili, N.; Eggink, D.; Leenders, T.; Li, Z.; Liu, L.; Wolfert, M. A.; Papanikolaou, A.; Martínez-Romero, C.; Gagarinov, I. A.; Yu, W.; García-Sastre, A.; Wennekes, T.; Okamatsu, M.; Verheije, M. H.; Wilson, I. A.; Boons, G. J.; de Vries, R. P., N-Glycolylneuraminic Acid as a Receptor for Influenza A Viruses. *Cell Rep* **2019,** *27* (11), 3284-3294.e6.

36. t Hart, I. M. E.; Li, T.; Wolfert, M. A.; Wang, S.; Moremen, K. W.; Boons, G.-J., Chemoenzymatic synthesis of the oligosaccharide moiety of the tumor-associated antigen disialosyl globopentaosylceramide. *Org. Biomol. Chem.* **2019,** *17* (31), 7304-7308.

37. Gagarinov, I. A.; Li, T.; Toraño, J. S.; Caval, T.; Srivastava, A. D.; Kruijtzer, J. A. W.; Heck, A. J. R.; Boons, G.-J., Chemoenzymatic Approach for the Preparation of Asymmetric

Bi-, Tri-, and Tetra-Antennary N-Glycans from a Common Precursor. *J. Am. Chem. Soc.* **2017,** *139* (2), 1011-1018.

38. Liu, L.; Prudden, A. R.; Capicciotti, C. J.; Bosman, G. P.; Yang, J.-Y.; Chapla, D. G.; Moremen, K. W.; Boons, G.-J., Streamlining the chemoenzymatic synthesis of complex N-glycans by a stop and go strategy. *Nature Chemistry* **2019,** *11* (2), 161-169.

39. Gupta, R.; Matta, K. L.; Neelamegham, S., A systematic analysis of acceptor specificity and reaction kinetics of five human α(2,3)sialyltransferases: Product inhibition studies illustrate reaction mechanism for ST3Gal-I. *Biochem. Biophys. Res. Commun.* **2016,** *469* (3), 606-612.

40. Maia, M. R. G.; Marques, S.; Cabrita, A. R. J.; Wallace, R. J.; Thompson, G.; Fonseca, A. J. M.; Oliveira, H. M., Simple and Versatile Turbidimetric Monitoring of Bacterial Growth in Liquid Cultures Using a Customized 3D Printed Culture Tube Holder and a Miniaturized Spectrophotometer: Application to Facultative and Strictly Anaerobic Bacteria. *Front Microbiol* **2016,** *7*, 1381-1381.

41. Gomes, L.; Monteiro, G.; Mergulhão, F., The Impact of IPTG Induction on Plasmid Stability and Heterologous Protein Expression by Escherichia coli Biofilms. *Int. J. Mol. Sci.* **2020,** *21* (2).

42. Noble, J. E., Chapter Two - Quantification of Protein Concentration Using UV Absorbance and Coomassie Dyes. *Methods Enzymol.* **2014,** *536*, 17-26.

43. Suelter, C. H.; DeLuca, M., How to prevent losses of protein by adsorption to glass and plastic. *Anal. Biochem.* **1983,** *135* (1), 112-119.

44. Trinchera, M.; Aronica, A.; Dall'Olio, F., Selectin Ligands Sialyl-Lewis a and Sialyl-Lewis x in Gastrointestinal Cancers. *Biology-Basel* **2017,** *6* (1), 18.

45. Lan, Y.; Hao, C.; Zeng, X.; He, Y.; Zeng, P.; Guo, Z.; Zhang, L., Serum glycoprotein-derived N- and O-linked glycans as cancer biomarkers. *Am. J. Cancer Res.* **2016,** *6* (11), 2390-2415.

46. Manfred Hesse, H. M., Bernd Zeeh, *Spectroscopic Methods in Organic Chemistry*. 2008.

47. Dabrowski, U.; Friebolin, H.; Brossmer, R.; Supp, M., 1H-NMR studies at N-acetyl-D-neuraminic acid ketosides for the determination of the anomeric configuration II. *Tetrahedron Lett.* **1979,** *20* (48), 4637-4640.

48. Haverkamp, J.; Van Halbeek, H.; Dorland, L.; Vliegenthart, J. F. G.; Pfeil, R.; Schauer, R., High-Resolution 1H-NMR Spectroscopy of Free and Glycosidically Linked O- Acetylated Sialic Acids. *Eur. J. Biochem.* **1982,** *122* (2), 305-311.

49. Okamoto, K.; Kondo, T.; Goto, T., Glycosylation of 4,7,8,9-Tetra-O-acetyl-2-deoxy-2β,3β-epoxy-N-acetylneuraminic Acid Methyl Ester. *Bull. Chem. Soc. Jpn.* **1987,** *60* (2), 637-643.

50. Mahon, C. S.; Jackson, A. W.; Murray, B. S.; Fulton, D. A., Templating a polymer-scaffolded dynamic combinatorial library. *Chem. Commun.* **2011,** *47* (25), 7209-7211.

51. Mondal, M.; Hirsch, A. K. H., Dynamic combinatorial chemistry: a tool to facilitate the identification of inhibitors for protein targets. *Chem. Soc. Rev.* **2015,** *44* (8), 2455-2488.

52. Hartman, A. M.; Gierse, R. M.; Hirsch, A. K. H., Protein-Templated Dynamic Combinatorial Chemistry: Brief Overview and Experimental Protocol. *Eur. J. Org. Chem.* **2019,** *2019* (22), 3581-3590.

53. Mahon, C. S.; Fulton, D. A., Templation-induced re-equilibration in polymer-scaffolded dynamic combinatorial libraries leads to enhancements in binding affinities. *Chem. Sci.* **2013,** *4* (9), 3661-3666.

54. Schiff, H., Mittheilungen aus dem Universitätslaboratorium in Pisa: Eine neue Reihe organischer Basen. *Justus Liebigs Annalen der Chemie* **1864,** *131* (1), 118-119.

55. Giuseppone, N.; Lehn, J.-M., Protonic and temperature modulation of constituent expression by component selection in a dynamic combinatorial library of imines. *Chemistry* **2006,** *12* (6), 1715-1722.

56. Godoy-Alcantar, C.; Yatsimirsky, A. K.; Lehn, J. M., Structure-stability correlations for imine formation in aqueous solution. *J. Phys. Org. Chem.* **2005,** *18* (10), 979-985.

57. Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J.-L.; Sanders, J. K. M.; Otto, S., Dynamic Combinatorial Chemistry. *Chem. Rev.* **2006,** *106* (9), 3652-3711.

58. Fulton, D. A., Dynamic Combinatorial Libraries Constructed on Polymer Scaffolds. *Org. Lett.* **2008,** *10* (15), 3291-3294.

59. Busschaert, N.; Thompson, S.; Hamilton, A. D., An α-helical peptidomimetic scaffold for dynamic combinatorial library formation. *Chem. Commun.* **2017,** *53* (2), 313-316.

60. Bugaut, A.; Bathany, K.; Schmitter, J.-M.; Rayner, B., Target-induced selection of ligands from a dynamic combinatorial library of mono- and bi-conjugated oligonucleotides. *Tetrahedron Lett.* **2005,** *46* (4), 687-690.

61. Ura, Y.; Beierle, J. M.; Leman, L. J.; Orgel, L. E.; Ghadiri, M. R., Self-Assembling Sequence-Adaptive Peptide Nucleic Acids. *Science* **2009,** *325* (5936), 73-77.

62. Nazarpack-Kandlousy, N.; Nelen, M. I.; Goral, V.; Eliseev, A. V., Synthesis and Mass Spectrometry Studies of Branched Oxime Ether Libraries. Mapping the Substitution Motif via Linker Stability and Fragmentation Pattern. *J. Org. Chem.* **2002,** *67* (1), 59-65.

63. Dunetz, J. R.; Magano, J.; Weisenburger, G. A., Large-Scale Applications of Amide Coupling Reagents for the Synthesis of Pharmaceuticals. *Org Process Res Dev* . **2016,** *20* (2), 140-177.

64. Carey, F. A.; Sundberg, R. J., Nucleophilic Substitution. In *Advanced Organic Chemistry: Part A: Structure and Mechanisms*, Carey, F. A.; Sundberg, R. J., Eds. Springer US: Boston, MA, 2007; pp 389-472.

65. Staudinger, H.; Meyer, J., Über neue organische Phosphorverbindungen III. Phosphinmethylenderivate und Phosphinimine. *Helv. Chim. Acta* **1919,** *2* (1), 635-646.

66. Lin, F. L.; Hoyt, H. M.; van Halbeek, H.; Bergman, R. G.; Bertozzi, C. R., Mechanistic Investigation of the Staudinger Ligation. *J. Am. Chem. Soc.* **2005,** *127* (8), 2686-2695.

67. Sorgi, K. L., Diisopropylethylamine. In *Encyclopedia of Reagents for Organic Synthesis*.

68. Herrmann, A., Dynamic combinatorial/covalent chemistry: a tool to read, generate and modulate the bioactivity of compounds and compound mixtures. *Chem. Soc. Rev.* **2014,** *43* (6), 1899-1933.

69. Zameo, S.; Vauzeilles, B.; Beau, J. M., Direct composition analysis of a dynamic library of imines in an aqueous medium. *Eur. J. Org. Chem.* **2006,** *2006* (24), 5441-5444.

70. Kulchat, S.; Chaur, M. N.; Lehn, J.-M., Kinetic Selectivity and Thermodynamic Features of Competitive Imine Formation in Dynamic Covalent Chemistry. *Chem. Eur. J* **2017,** *23* (46), 11108-11118.

71. Bunyapaiboonsri, T.; Ramström, H.; Ramström, O.; Haiech, J.; Lehn, J.-M., Generation of Bis-Cationic Heterocyclic Inhibitors of Bacillus subtilis HPr Kinase/Phosphatase from a Ditopic Dynamic Combinatorial Library. *J. Med. Chem.* **2003,** *46* (26), 5803-5811.

72. Bunyapaiboonsri, T.; Ramstrom, H.; Ramstrom, O.; Haiech, J.; Lehn, J. M., Generation of Bis-cationic heterocyclic inhibitors of Bacillus subtilis HPr kinase/phosphatase from a ditopic dynamic combinatorial library. *J. Med. Chem.* **2003,** *46* (26), 5803-5811.

73. Mazik, M.; Kuschel, M., Amide, amino, hydroxy and aminopyridine groups as building blocks for carbohydrate receptors. *Eur. J. Org. Chem.* **2008,** *2008* (9), 1517-1526.

74. Mazik, M.; Geffert, C., 8-Hydroxyquinoline as a building block for artificial receptors: binding preferences in the recognition of glycopyranosides. *Org. Biomol. Chem.* **2011,** *9* (7), 2319-2326.

75. Mazik, M.; Kuschel, M., Highly effective acyclic carbohydrate receptors consisting of aminopyridine, imidazole, and indole recognition units. *Chem. Eur. J.* **2008,** *14* (8), 2405-2419.

76. Crocker, P. R.; Vinson, M.; Kelm, S.; Drickamer, K., Molecular analysis of sialoside binding to sialoadhesin by NMR and site-directed mutagenesis. *Biochem. J.* **1999,** *341*, 355-361.

77. Haji-Ghassemi, O.; Blackler, R. J.; Young, N. M.; Evans, S. V., Antibody recognition of carbohydrate epitopes. *Glycobiology* **2015,** *25* (9), 920-952.

78. Mazik, M.; Cavga, H., Molecular recognition of N-acetylneuraminic acid with acyclic benzimidazolium- and aminopyridine/guanidinium-based receptors. *J. Org. Chem.* **2007,** *72* (3), 831-838.

79. Imberty, A.; Gautier, C.; Lescar, J.; Perez, S.; Wyns, L., An unusual carbohydrate binding site revealed by the structures of two Maackia amurensis lectins complexed with sialic acid-containing oligosaccharides. *J. Biol. Chem.* **2000,** *275* (23), 17541-17548.

80. Geffert, C.; Kuschel, M.; Mazik, M., Molecular Recognition of N-Acetylneuraminic Acid by Acyclic Pyridinium- and Quinolinium-Based Receptors in Aqueous Media: Recognition through Combination of Cationic and Neutral Recognition Sites. *J. Org. Chem.* **2013,** *78* (2), 292-300.

81. Di Pasquale, A.; Tommasone, S.; Xu, L.; Ma, J.; Mendes, P. M., Cooperative Multipoint Recognition of Sialic Acid by Benzoboroxole-Based Receptors Bearing Cationic Hydrogen-Bond Donors. *J. Org. Chem.* **2020,** *85* (13), 8330-8338.

82. Zhou, Y.; Ye, H.; You, L., Reactivity-Based Dynamic Covalent Chemistry: Reversible Binding and Chirality Discrimination of Monoalcohols. *J. Org. Chem.* **2015,** *80* (5), 2627-2633.

83. Leung, I. K. H.; Brown Jr, T.; Schofield, C. J.; Claridge, T. D. W., An approach to enzyme inhibition employing reversible boronate ester formation. *MedChemComm* **2011,** *2* (5), 390-395.

84. Cancilla, M. T.; He, M. M.; Viswanathan, N.; Simmons, R. L.; Taylor, M.; Fung, A. D.; Cao, K.; Erlanson, D. A., Discovery of an Aurora kinase inhibitor through site-specific dynamic combinatorial chemistry. *Bioorg. Med. Chem. Lett.* **2008,** *18* (14), 3978-3981.

85. Martin, A. R.; Barvik, I.; Luvino, D.; Smietana, M.; Vasseur, J. J., Dynamic and Programmable DNA-Templated Boronic Ester Formation. *Angew. Chem. Int. Ed.* **2011,** *50* (18), 4193-4196.

86. Leung, I. K. H.; Brown, T.; Schofield, C. J.; Claridge, T. D. W., An approach to enzyme inhibition employing reversible boronate ester formation. *MedChemComm* **2011,** *2* (5), 390-395.

87. Demetriades, M.; Leung, I. K. H.; Chowdhury, R.; Chan, M. C.; McDonough, M. A.; Yeoh, K. K.; Tian, Y.-M.; Claridge, T. D. W.; Ratcliffe, P. J.; Woon, E. C. Y.; Schofield, C. J., Dynamic Combinatorial Chemistry Employing Boronic Acids/Boronate Esters Leads to Potent Oxygenase Inhibitors. *Angew. Chem. Int. Ed.* **2012,** *51* (27), 6672-6675.

88. Leung, I. K. H.; Demetriades, M.; Hardy, A. P.; Lejeune, C.; Smart, T. J.; Szöllössi, A.; Kawamura, A.; Schofield, C. J.; Claridge, T. D. W., Reporter Ligand NMR Screening Method for 2-Oxoglutarate Oxygenase Inhibitors. *J. Med. Chem.* **2013,** *56* (2), 547-555.

89. Golovanov, I. S.; Mazeina, G. S.; Nelyubina, Y. V.; Novikov, R. A.; Mazur, A. S.; Britvin, S. N.; Tartakovsky, V. A.; Ioffe, S. L.; Sukhorukov, A. Y., Exploiting Coupling of Boronic Acids with Triols for a pH-Dependent "Click-Declick" Chemistry. *J. Org. Chem.* **2018,** *83* (17), 9756-9773.

90. Nakatani, H.; Hiromi, K., Binding of m-nitrobenzeneboronic acid to the active site of subtilisin BPN′. *Biochim. Biophys. Acta* **1978,** *524* (2), 413-417.

91. Norrild, J. C.; Eggert, H., Evidence for Mono- and Bisdentate Boronate Complexes of Glucose in the Furanose Form. Application of 1JC-C Coupling Constants as a Structural Probe. *J. Am. Chem. Soc.* **1995,** *117* (5), 1479-1484.

92. Stones, D.; Manku, S.; Lu, X. S.; Hall, D. G., Modular solid-phase synthetic approach to optimize structural and electronic properties of oligoboronic acid receptors and sensors for the aqueous recognition of oligosaccharides. *Chem. Eur. J.* **2004,** *10* (1), 92-100.

Chapter 5 – Experimental

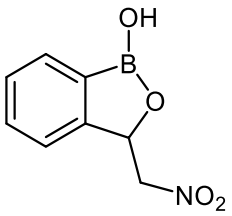
5.1 Synthetic procedures

5.1.1 Materials and methods

Reagents were purchased from Sigma-Aldrich, Acros Organics and Alfa Aesar. N-acetylneuraminc acid and 2-O-methyl-α-D-N-acetylneuraminic acid were purchased from Carbosynth Limited. Fmoc-Asp-OBzl was purchased from Fluorochem Ltd. All the reagents were used without further purification. ^{1}H NMR and ^{13}C NMR spectra were recorded at room temperature on the following spectrometers: Bruker AVIII at 300 MHz, Bruker AVIII400 at 400 MHz and 101 MHz, Bruker AVANCE NEO at 400 MHz and 101 MHz and Bruker AVANCE NEO at 500 MHz and 126 MHz. Thin layer chromatography (TLC) was performed using commercially available Macherey-Nagel aluminium backed plates coated with a 0.20 mm layer of silica gel 60 Å with fluorescent indicator UV254. TLC plates were visualized using ultraviolet light of 254 nm wavelength. TLC plates were stained with the following stains; permanganate, ninhydrin, coumarin and bromocresol green. Silica gel column chromatography was carried out using Sigma-Aldrich 60 Å silica gel (35-70 μm). Mass spectra were recorded with a Bruker Daltonics MicroTOF-Q II and Waters Xevo G2-XS spectrometer.

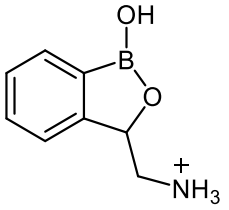
5.1.2 Synthesis of benzoboroxole receptors

Synthesis of **B8**[1]



2-formylphenylboronic acid (1.92 g, 12.82 mmol) was dissolved in water (12.78 ml). The solution was cooled to 0 °C and nitromethane was added (2.08 ml, 38.33 mmol) followed by the addition of sodium hydroxide (0.54 g, 13.50 mmol). The reaction mixture was stirred for 3 hours, diluted with water and acidified with 2 M HCl. The precipitate was filtered to give the pure **B8** as a white-yellow solid (2.37 g, 92.0%). ^{1}H NMR (300 MHz, DMSO-d_6, 298 K) δ ppm 9.51 (s, 1 H), 7.74 (dt, J = 7.3 Hz, 1.1 Hz, 1 H), 7.53 (m, 2 H), 7.42 (m, 1 H), 5.78 (dd, J = 9.2 Hz, 2.8 Hz, 1 H), 5.34 (dd, J = 13.4 Hz, 2.8 Hz, 1 H), 4.85 (dd, J = 13.5 Hz, 9.2 Hz, 1 H). ^{1}H NMR is in agreement with literature[2].

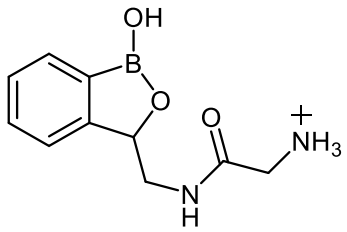
*Synthesis of receptor **B2***[3]



A solution of **B8** (0.50 g, 2.59 mmol) in anhydrous methanol (20 ml) was cooled to 0 °C. $(Boc)_2O$ (1.20 ml, 5.22 mmol) and $NiCl_2 \cdot 6\ H_2O$ (0.62 g, 2.96 mmol) were added and the mixture was stirred under argon for 20 min. $NaBH_4$ (0.59 g, 15.59 mmol) was added and the mixture was stirred overnight at room temperature under argon. The solvent was evaporated

under reduced pressure and the crude dissolved in ethyl acetate and filtered through Celite. The solution was concentrated under reduced pressure and the crude deprotected overnight with 2 M HCl in diethyl ether (15 ml). The precipitate was filtered and triturated with diethyl ether to give derivative **B2** as an off-white solid (0.40 g, 78.0%). ^{1}H NMR (300 MHz, DMSO-d_6, 298 K) δ ppm 9.55 (s, 1 H), 8.16 (s, 3 H), 7.82 (dt, J = 7.3 Hz, 1.1 Hz, 1 H), 7.53 (m, 2 H), 7.42 (m, 1 H), 5.35 (dd, J = 9.3 Hz, 2.8 Hz, 1 H), 3.51 (m, 1 H), 2.79 (m, 1 H). ^{1}H NMR is in agreement with literature[3].

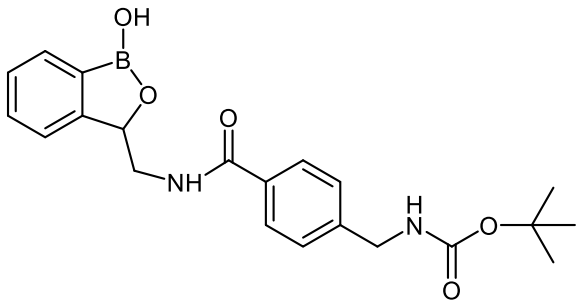
*Synthesis of receptor **B4***



Derivative **B2** (0.10 g, 0.50 mmol) was dissolved in DMF (5 ml). DMAP (0.15 g 1.25 mmol) was added, followed by the addition of Boc-Gly-OH (0.11g, 0.6 mmol) and EDCI (0.12 g, 0.60 mmol). The reaction mixture was stirred overnight, and the solvent was evaporated under reduced pressure. The crude was dissolved in ethyl acetate and washed with 1 M HCl (3 × 5 ml) and the with sat. $NaHCO_3$ (3 x 5 ml). The organic layer was dried with $MgSO_4$ and concentrated under reduced pressure. The resulting oil was stirred overnight with 2 M HCl in diethyl ether (10 ml). The solution was filtered, and the precipitate triturated with diethyl ether to give **B4** the as an off-white solid. (40.2 mg, 31.4%). ^{1}H NMR (400 MHz, DMSO-d_6, 298 K) δ ppm 9.38 (s, 1 H), 8.67 (t, J = 5.5 Hz, 1 H), 8.18 (s, 3 H), 7.79 (d, J = 7.3 Hz, 1 H), 7.49 (m, 2 H), 7.38 (m, 1 H), 5.19 (dd, J = 7.4 Hz, 4.0 Hz, 1 H), 3.68 (ddd, J = 9.9 Hz, 5.6 Hz, 4.3 Hz, 1 H), 3.53 (q, J = 16.3 Hz, 1 H), 3.24 (m, 1 H). ^{13}C NMR (100 MHz, DMSO-d_6, 298 K) δ ppm 166.1,

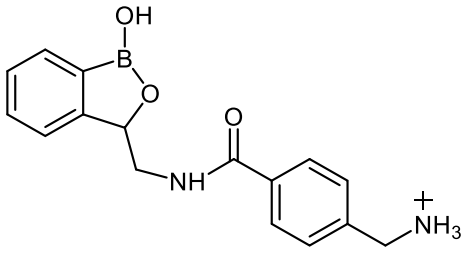
154.0, 132.2, 130.7, 130.7, 127.5, 121.7, 78.6, 44.4, 40.1. ESI+MS *m/z* 221.11 $[M]^+$. HRMS (ESI) *m/z* $[M]^+$ calculated for $C_{10}H_{14}BN_2O_3$ 221.1092, found 221.1091.

*Synthesis of derivative **B9***



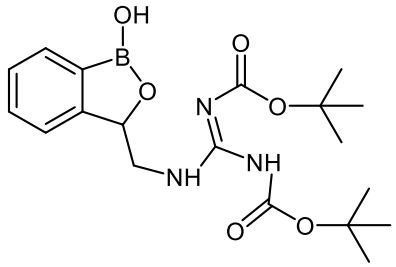
Derivative **B2** (0.30 g, 1.52 mmol) was dissolved in DMF (5.0 ml). DMAP (0.47 g, 3.82 mmol) was added, followed by the addition of 4-(Boc-aminomethyl)benzoic acid (0.38 g, 1.52 mmol) and EDCI (0.28 g, 1.83 mmol). The reaction mixture was stirred overnight and concentrated under reduced pressure. The crude was then dissolved in ethyl acetate and washed with 1 M HCl (3 × 5 ml) and then with $NaHCO_3$ (5 ml). The organic layer was dried with $MgSO_4$ and concentrated under reduced pressure. The crude was recrystallised with hexane/ethyl acetate to give the pure **B9** as an off-white solid (0.184 g, 30.6%). ^{1}H NMR (400 MHz, DMSO-d_6, 298 K) δ ppm 9.27 (s, 1 H), 8.73 (t, *J* = 5.60, 1 H), 7.81 (d, *J* = 8.20, 2 H), 7.73 (dt, *J* = 7.24, 1.04, 1 H), 7.46 (m, 3 H), 7.37 (m, 1 H), 7.31 (d, *J* = 8.12, 2 H), 5.32 (dd, *J* = 7.84, 4.52, 1 H), 4.17 (d, *J* = 6.20, 2 H), 3.69 (dt, *J* = 13.6, 5.08, 1 H), 3.35 (1 H), 1.40 (s, 9 H). ^{13}C NMR (100 MHz, DMSO-d_6, 298 K) δ ppm 166.30, 155.86, 154.65, 143.55, 132.78, 130.67, 130.60, 127.47, 127.28, 126.69, 121.77, 77.96, 78.76, 45.43, 43.16, 28.27. ESI+MS *m/z* 419.17 $[M + Na]^+$. HRMS (ESI) *m/z* $[M + Na]^+$ calculated for $C_{21}H_{25}BN_2O_5Na$ 419.1749, found 419.1721.

Synthesis of receptor ***B6***



Derivative **B9** (184.50 mg, 0.47 mmol) was dissolved in ethyl acetate (5.0 ml). 2 M HCl in diethyl ether (10.0 ml) was added and the mixture was stirred overnight. The solvent was removed under vacuum, to give the pure **B6** as an off-white solid (0.155 mg, quant.). ^{1}H NMR (400 MHz, DMSO-d_6, 298 K) δ ppm 9.30 (s, 1 H), 8.85 (dt, J = 5.6 Hz, 1 H), 8.52 (s, 3 H), 7.90 (d, J = 8.3 Hz, 2 H), 7.76 (dt, J = 7.3 Hz, 1.1 Hz, 1 H). 7.58 (m, 2 H), 7.45 (m, 2 H), 7.36 (dt, J = 7.0 Hz, 1.10 Hz, 1 H), 5.34 (dd, J = 7.7 Hz, 4.40 Hz, 1 H), 4.07 (s, 2 H), 3.71 (dt, J = 5.0 Hz, 13.64 Hz, 1 H), 3.40 (m, 1 H). ^{13}C NMR (100 MHz, DMSO-d_6, 298 K) δ ppm 165.9, 154.6, 137.2, 134.2, 130.7, 130.6, 128.8, 127.4, 121.7, 78.7, 45.4, 41.8. ESI+MS *m/z* 297.14 $[M]^+$. HRMS (ESI) *m/z* $[M]^+$ calculated for $C_{16}H_{18}BN_2O_3$ 297.1405, found 297.1402.

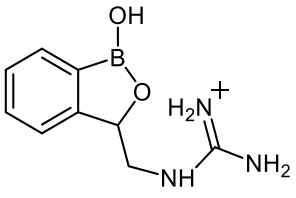
Synthesis of derivative ***B10***



Derivative **B2** (0.20 g, 1.00 mmol) was dissolved in methanol (15 ml). N,N′-Di-Boc-1H-pyrazole-1-carboxamidine (0.27 g, 1.1 mmol) and triethylamine (324 µl) were added and the mixture was stirred overnight. The solvent was removed under vacuum,

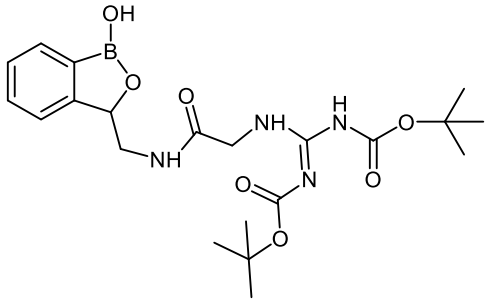
and the crude purified by column chromatography 85% hexane 15% ethyl acetate to give **B10** as an off-white solid (0.10 g, 24.6%). ^{1}H NMR (400 MHz, DMSO-*d*$_6$, 298 K) δ ppm 11.48 (s, 1 H,), 9.38 (s, 1 H), 8.46 (t, *J* = 5.2 Hz, 1 H), 7.73 (dt, *J* = 7.3 Hz, 1.1 Hz, 1 H), 7.48 (m, 2 H), 7.39 (td, *J* = 7.1 Hz, 1.3 Hz, 1 H), 5.32 (dd, *J* = 8.2 Hz, 3.5 Hz, 1 H), 3.97 (ddd, *J* = 13.7 Hz, 6.1 Hz, 3.6 Hz, 1 H), 3.26 (ddd, *J* = 13.7 Hz, 8.2 Hz, 4.5 Hz, 1 H), 1.47 (s, 9 H), 1.39 (s, 9 H). ^{13}C NMR (100 MHz, DMSO-*d*$_6$, 298 K) δ ppm 163.0, 155.5, 153.7, 152.2, 130.8, 130.7, 127.7, 121.7, 83.2, 78.4, 78.3, 45.8, 28.0, 27.6. ESI+MS *m/z* 406.22 $[M + H]^+$. HRMS (ESI) *m/z* $[M + H]^+$ calculated for $C_{19}H_{29}BN_3O_6$ 406.2144, found 406.2164.

*Synthesis of receptor **B3***



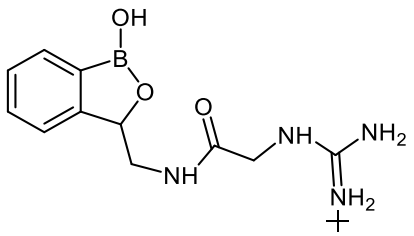
Derivative **B10** (0.10 g, 0.25 mol) was dissolved in a mixture of ethyl acetate (5.0 ml) and methanol (2.0 ml). The mixture was cooled to 0 °C and acetyl chloride (0.50 ml) was added dropwise. The reaction was stirred at room temperature for 48 hours. The mixture was concentrated under vacuum to give **B3** as an off-white hygroscopic solid (59.6 mg, quant.). ^{1}H NMR (400 MHz, DMSO-*d*$_6$, 298 K) δ ppm 9.43 (s, 1 H), 7.78 (dt, *J* = 7.2 Hz, 1.0 Hz, 1 H), 7.61 (t, *J* = 6.0 Hz, 1 H), 7.52 (m, 2 H), 7.39 (m, 2 H), 6.96 (s), 5.23 (dd, 1 H, *J* = 7.7 Hz, 3.3 Hz), 3.77 (ddd, *J* = 14.2 Hz, 6.0 Hz, 3.4 Hz, 1 H), 3.28 (ddd, *J* = 13.8 Hz, 7.6 Hz, 5.9 Hz, 1 H). ^{13}C NMR (100 MHz, DMSO-*d*$_6$, 298 K) δ ppm 157.2, 153.2, 130.8, 130.7, 127.7, 121.8, 78.6, 46.0. ESI+MS m/z 206.11 [M]+. HRMS (ESI) m/z $[M]^+$ calculated for $C_9H_{13}BN_3O_2$ 206.1095, found 206.1099.

Synthesis of derivative ***B11***



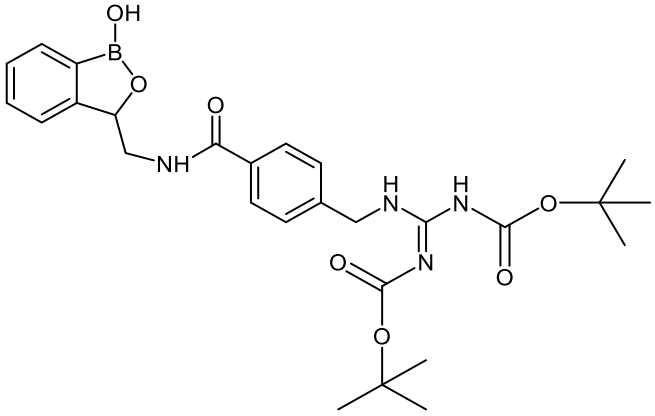
Derivative **B4** (0.21 g, 0.83 mmol) was dissolved in methanol (15.0 ml). N,N′-Di-Boc-1H-pyrazole-1-carboxamidine (0.26 g, 0.84 mmol) and triethylamine (259 µl) were added and the mixture was stirred overnight. The solvent was removed under vacuum, and the crude purified by gradient column chromatography, 85% hexane 15% ethyl acetate to 98% ethyl acetate 2% methanol to give **B11** as an off-white solid (0.232 g, 60.7%). ^{1}H NMR (400 MHz, DMSO-d_6, 298 K) δ ppm 11.43 (s, 1 H), 9.26 (s, 1 H), 8.69 (t, J = 4.8 Hz, 1 H), 8.35 (t, J = 5.6 Hz, 1 H), 7.72 (dt, J = 11.4 Hz, 5.6 Hz, 1 H), 7.45 (m, 2 H), 7.36 (dt, J = 7.2 Hz, 1.5 Hz, 1 H), 5.17 (dd, J = 7.1 Hz, 4.4 Hz, 1 H), 3.93 (dd, 1 H, J = 9.6 Hz, 4.8 Hz, 2H), 3.56 (ddd, J = 13.7 Hz, 5.8 Hz, 4.5 Hz, 1 H), 3.24 (ddd, J = 13.3 Hz, 7.2 Hz, 5.7 Hz, 1 H), 1.48 (s, 9 H), 1.38 (s, 9 H). ^{13}C NMR (100 MHz, DMSO-d_6, 298 K) δ ppm 167.9, 162.9, 155.0, 154.3, 151.9, 130.6, 130.5, 127.4, 121.8, 83.0, 78.7, 78.3, 44.5, 43.5, 28.0, 27.6. ESI+MS *m/z* 463.24 $[M + H]^+$. HRMS (ESI) *m/z* $[M + H]^+$ calculated for $C_{21}H_{32}BN_4O_7$ 463.2359, found 463.2378.

Synthesis of receptor ***B5***



Derivative **B11** (0.23 g, 0.50 mol) was dissolved in a mixture of ethyl acetate (10.0 ml) and methanol (5.0 ml). The mixture was cooled to 0 °C and acetyl chloride (1.0 ml) was added dropwise. The reaction was stirred at room temperature for 48 hours. The mixture was concentrated under vacuum and the crude recrystallised with a mixture of methanol and diethyl ether to give the **B5** as an off-white solid (45.4 mg, 30.4%). ^{1}H NMR (400 MHz, DMSO-d_6, 298 K) δ ppm 9.27 (s, 1 H), 8.73 (t, J = 5.6 Hz, 1 H), 7.81 (d, J = 8.2 Hz, 2 H), 7.73 (dt, J = 7.2 Hz, 1.0 Hz, 1 H), 7.46 (m, 3 H), 7.37 (m, 1 H), 7.31 (d, J = 8.1 Hz, 2 H), 5.32 (dd, J = 7.8 Hz, 4.5 Hz, 1 H), 4.17 (d, J = 6.2 Hz, 2 H), 3.69 (dt, J = 13.6 Hz, 5.1 Hz, 1 H), 3.35 (1 H), 1.40 (s, 9 H). ^{13}C NMR (100 MHz, DMSO-d_6, 298 K) δ ppm 166.3, 155.9, 154.7, 143.6, 132.8, 130.7, 130.6, 127.5, 127.3, 126.7, 121.8, 78.8, 78.0, 45.4, 43.2, 28.3. ESI+MS *m/z* 263.12 $[M]^+$. HRMS (ESI) *m/z* $[M]^+$ calculated for $C_{11}H_{16}BN_4O_3$ 263.1310, found 263.1312.

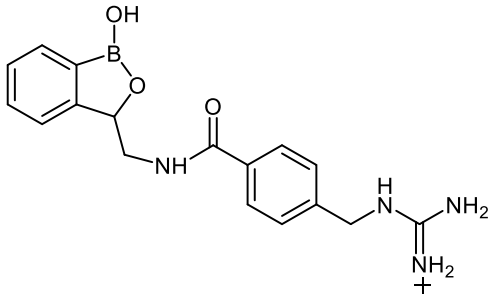
*Synthesis of derivative **B12***



Derivative **B6** (0.16 g, 0.47 mmol) was dissolved in methanol (15.0 ml). N,N′-Di-Boc-1H-pyrazole-1-carboxamidine (0.16 g, 0.52 mmol) and triethylamine (163 µl) were added and the mixture was stirred overnight. The solvent was removed under vacuum and the crude purified by column chromatography, 80% hexane 20% ethyl acetate to 100% ethyl

acetate, to give **B12** as an off-white solid (0.139 g, 55.1%). ^{1}H NMR (400 MHz, DMSO-d_6, 298 K) δ ppm 11.54 (s, 1 H), 9.27 (s, 1 H), 8.80 – 8.71 (m, 2 H). 7.83 (m, 2 H), 7.73 (dt, J = 7.3 Hz, 1.1 Hz, 1 H), 7.46 (m, 2 H), 7.39 – 3.34 (m, 3 H), 5.32 (dd, J = 7.8 Hz, 4.6 Hz, 1 H), 4.58 (d, J = 6.0 Hz, 2 H), 3.69 (dt, J = 13.6 Hz, 5.1 Hz, 1 H), 3.40 (m, 1 H), 1.48 (s, 9 H), 1.37 (s, 9 H). ^{13}C NMR (100 MHz, DMSO-d_6, 298 K) δ ppm 166.2, 163.1, 155.6, 154.5, 152.0, 141.8, 133.1, 130.7, 130.6, 127.5, 127.4, 126.9, 121.8, 83.0, 78.7, 78.4, 45.4, 43.3, 28.0, 27.7. ESI+MS *m/z* 539.24 $[M + H]^+$. HRMS (ESI) *m/z* $[M + H]^+$ calculated for $C_{27}H_{36}BN_4O_7$ 539.2672, found 539.2697.

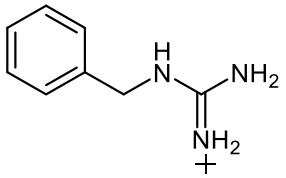
*Synthesis of receptor **B7***



Derivative **B12** (0.14 g, 0.26 mol) was dissolved in a mixture of ethyl acetate (10.0 ml) and methanol (5.0 ml). The mixture was cooled to 0 °C and acetyl chloride (1 ml) was added dropwise. The reaction was stirred at room temperature for 72 hours. The mixture was concentrated under vacuum to give the **B7** as an off-white solid (87.4 mg, quant.). ^{1}H NMR (400 MHz, DMSO-d_6, 298 K) δ ppm 9.28 (s, 1 H), 8.80 (t, J = 5.5 Hz, 1 H), 8.25 (t, J = 6.3 Hz), 7.88 (m, 2 H), 7.76 (dt, J = 7.3 Hz, 1.0 Hz, 1 H), 7.46 (m, 2 H), 7.37 (m, 3 H), 5.33 (dd, J = 7.8 Hz, 4.5 Hz, 1 H), 4.46 (d, J = 6.3 Hz, 2 H), 3.70 (dt, J = 13.6 Hz, 5.0 Hz, 1 H), 3.39 (m, 1 H). ^{13}C NMR (100 MHz, DMSO-d_6, 298 K) δ ppm 166.0, 157.2, 154.6, 140.6, 133.5, 130.6, 130.6,

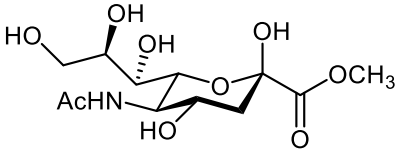
127.5, 127.4, 126.9, 121.7, 78.7, 45.4, 43.60. ESI+MS *m/z* 339.16 $[M]^+$. HRMS (ESI) *m/z* $[M]^+$ calculated for $C_{17}H_{20}BN_4O_3$ 339.1623, found 339.1614.

Synthesis of receptor ***B13***



Benzylamine (0.20 g, 1.87 mmol) dissolved in methanol (10 ml). N,N′-Di-Boc-1H-pyrazole-1-carboxamidine (0.64 g, 2.06 mmol) and triethylamine (210 µl) were added and the mixture was stirred overnight. The precipitate was filtered and dissolved in a mixture of ethyl acetate (5 ml) and methanol (2 ml). The mixture was cooled to 0 °C and acetyl chloride (1 ml) was added dropwise and the mixture stirred at room temperature for 24 hours. The solvent was removed under vacuum to afford compound **B13** as a white solid (51.6 mg, 18.0%). ^{1}H NMR (400 MHz, DMSO-d_6, 298 K) δ ppm 8.23 (t, J = 5.9 Hz, 1 H), 7.38 (m, 2 H), 7.31 (m, 2 H), 4.39 (d, J = 6.2 Hz, 2 H). ^{13}C NMR (100 MHz, DMSO-d_6, 298 K) δ ppm 157.2, 137.3, 128.5, 127.5, 127.2, 43.9. ESI+MS *m/z* 150.09 $[M]^+$. HRMS (ESI) *m/z* $[M]^+$ calculated for $C_8H_{12}N_3$ 150.1026, found 150.1024.

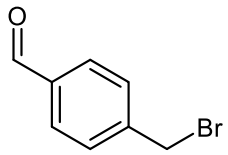
5.1.3 Synthesis of sialic acid methyl ester, S3 [4]



Sialic acid (**S1**) (5.00 g, 16.7 mmol) was suspended in anhydrous methanol (125 ml). TFA (1.50 ml, 19.40 mmol) was added and the reaction was stirred at room temperature for 72 hours. The solvent was evaporated under reduced pressure. The crude was triturated with diethyl ether to give the sialic acid methyl ester (**S3**) as white solid (5.05 g, 97.0%). ^{1}H NMR (300 MHz, DMSO-d_6, 298 K) δ ppm 4.02 (m, 2 H), 3.90 (d, J = 10.2 Hz, 1 H), 3.80 (m, 4 H), 3.69 (ddd, J = 9.8 Hz, 6.2 Hz, 2.49 Hz, 1 H), 3.57 (dd, J = 11.7 Hz, 6.2 Hz, 1 H), 3.51 (dd, J = 9.2 Hz, 0.7 Hz, 1 H), 2.27 (dd, J = 13.14 Hz, 4.38 Hz, 1 H), 2.00 (s, 3 H), 1.87 (dd, J = 13.0 Hz, 11.7 Hz, 1 H). ^{1}H NMR is in agreement with literature[4].

5.1.4 Synthesis of peptidomimetic scaffold

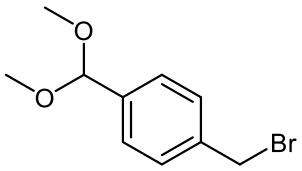
Synthesis of 4-(bromomethyl)benzaldehyde, ***S12*** [5]



4-(bromomethyl)-benzonitrile (1.00 g, 5.10 mmol) was dissolved in anhydrous toluene (10 ml). The mixture was cooled to 0 °C and stirred under argon. A solution of 0.1 M DIBAL in hexanes (6.6 ml) was added dropwise and the reaction was stirred to 0 °C for 1 hour. Chloroform (13 ml) and 10% HCl (33 ml) were added and the reaction stirred for 1 hour. The two phases were separated and the organic layer washed with water (3 x 15 ml), dried over $MgSO_4$ and concentrated under vacuum to give the 4-(bromomethyl)benzaldehyde **S12** (0.64 g, 63.2%) as an off-white solid. ^{1}H NMR (400 MHz, DMSO-d_6, 298 K) δ ppm 10.00 (s, 1 H), 7.90 (dt, J = 8.2 Hz, 1.8 Hz, 2 H), 7.67 (dt, J = 8.3 Hz, 2.0 Hz, 2 H), 4.78 (s, 2 H). ^{13}C NMR (100 MHz,

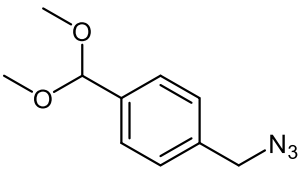
DMSO-d_6, 298 K) δ ppm 192.7, 144.6, 135.8, 130.0, 129.9, 33.16. ESI+MS *m/z* 198.98 $[M + H]^+$. HRMS (ESI) *m/z* $[M + H]^+$ calculated for C_8H_8BrO 198.9759, found 198.9758.

Synthesis of 1-(bromomethyl)-4-(dimethoxymethyl)-benzene, ***S13***



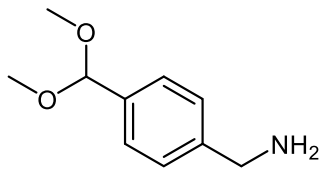
4-(bromomethyl)benzaldehyde **S12** (0.64 g, 3.22 mmol) was dissolved in anhydrous methanol (35 ml) and *p*-toluenesulfonic acid (61.3 mg, 0.32 mmol) was added. The reaction was stirred overnight under argon. The solvent was removed under reduced pressure and the crude dissolved in diethyl ether and washed with of sat. $NaHCO_3$ (3 x 25 ml) and water (3 x 25 ml). The crude was purified by column chromatography, 94% hexane 6% ethyl acetate. The solvent was removed under vacuum to afford the pure 1-(bromomethyl)-4-(dimethoxymethyl)-benzene **S13** (0.46 g, 57.7%) as a colourless oil. ^{1}H NMR (400 MHz, DMSO-d_6, 298 K) δ ppm 7.46 (dt, *J* = 8.2 Hz, 1.8 Hz, 2 H), 7.37 (m, 2 H), 5.37 (s, 1 H), 4.71 (s, 2 H), 3.24 (s, 6 H). ^{13}C NMR (100 MHz, DMSO-d_6, 298 K) δ ppm 138.3, 138.1, 129.1, 126.9, 102.3, 52.6, 34.1. ESI+MS *m/z* 245.00 $[M + H]^+$. HRMS (ESI) *m/z* $[M + H]^+$ calculated for $C_{10}H_{14}BrO_2$ 245.0001, found 245.0000.

Synthesis of 1-(azidomethyl)-4-(dimethoxymethyl)-benzene, ***S14***



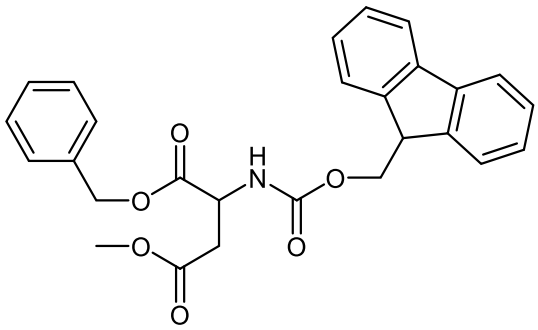
1-(bromomethyl)-4-(dimethoxymethyl)-benzene **S13** (0.46 g, 1.86 mmol) was dissolved in a 3:1 acetone:water mixture (20 ml) to which sodium azide (0.18 g, 2.70 mmol) was added. The reaction was stirred for 2 hours. Dichloromethane (40 ml) was added and the reaction was stirred for 1 hour. The mixture was extracted, washed with of water (3 x 20 ml) and dried over $MgSO_4$. The solvent was removed under vacuum to afford the pure 1-(azidomethyl)-4-(dimethoxymethyl)-benzene **S14** (0.31 g, 81.5%) as a colourless oil. ^{1}H NMR (400 MHz, DMSO-*d*$_6$, 298 K) δ ppm 7.42 (m, 2 H), 7.38 (m, 2 H), 5.39 (s, 1 H), 4.46 (s, 2 H), 3.24 (s, 6 H). ^{13}C NMR (100 MHz, DMSO-*d*$_6$, 298 K) δ ppm 138.1, 135.7, 128.2, 126.9, 102.4, 53.3, 52.6.

Synthesis of 1-(aminomethyl)-4-(dimethoxymethyl)-benzene, ***S15***



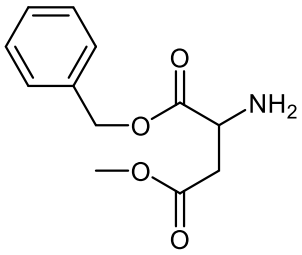
To a solution of 1-(azidomethyl)-4-(dimethoxymethyl)-benzene **S14** (0.31 g, 1.52 mmol) in methanol with 5% water (20 ml), triphenylphosphine (0.52 g, 1.97 mmol) was added and the mixture stirred overnight. The solvent was removed under vacuum and the crude dissolved in chloroform (30 ml) which was then washed with 0.1 M PB buffer pH 3 (4 x 20 ml). The aqueous phase was basified to pH 11 with 5 M NaOH and then extracted with chloroform (4 x 20 ml). The organic phase was concentrated under reduced pressure to give the 1-(aminomethyl)-4-(dimethoxymethyl)-benzene **S15** (0.20 g, 71.6%) as a colourless oil. ^{1}H NMR (400 MHz, DMSO-*d*$_6$, 298 K) δ ppm 7.32 (m, 2 H), 7.30 (m, 2 H), 5.35 (s, 1 H), 3.71 (s, 2 H), 3.30 (s, 6 H). ^{13}C NMR (100 MHz, DMSO-*d*$_6$, 298 K) δ ppm 144.5, 136.0, 126.7, 126.3, 102.6, 52.4, 45.4. ESI+MS *m/z* 181.09 $[M]^+$.

Synthesis of N-Fmoc-1-(phenylmethyl)-4-methylester-L-aspartic acid, ***S2***



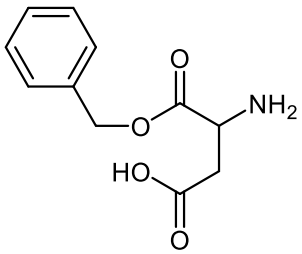
Fmoc-Asp-OBzl (**S1**) (0.50 g, 1.12 mmol) was dissolved in methanol (20 ml). DMAP (13.7 mg, 0.11 mmol) and EDCI (0.26 g, 1.34 mmol) were added and the reaction was stirred overnight at room temperature. The solvent was removed under vacuum and the crude dissolved in ethyl acetate (50 ml). The solution was washed with 1 M HCl (4 x 15 ml) and sat. $NaHCO_3$ (3 x 15 ml). The solvent was removed under vacuum to give the pure N-Fmoc-1-(phenylmethyl)-4-methylester-L-aspartic acid (**S2**) (0.49 g, 96.1%). ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ ppm 7.76 (d, *J* = 7.7 Hz, 2 H), 7.59 (d, *J* = 7.4 Hz, 2 H), 7.40 (dq, *J* = 7.5 Hz, 0.6 Hz, 2 H), 7.35 – 7.28 (m, 7 H), 5.82 (d, *J* = 8.6 Hz, 1 H), 5.23 (d, *J* = 12.2 Hz, 1 H), 5.19 (d, *J* = 12.2 Hz, 1 H), 4.69 (dt, *J* = 8.6 Hz, 4.6 Hz, 1 H), 4.42 (dd, *J* =10.6 Hz, 7.3 Hz, 1 H) 4.34 (dd, *J* =10.6 Hz, 7.3 Hz, 1 H), 4.22 (t, *J* = 7.2 Hz, 1 H), 3.64 (s, 3 H) 3.07 (dd, *J* = 17.1 Hz, 4.6 Hz, 1H), 2.88 (dd, *J* = 17.1 Hz, 4.6 Hz, 1H). ^{13}C NMR (100 MHz, $CDCl_3$, 298 K) δ ppm 171.4, 170.7, 156.1, 143.83, 141.44,135.3, 128.7, 128.6, 128.4, 127.9, 127.2, 125.3, 125.3, 120.1, 120.1, 67.8, 67.4, 52.2, 50.6, 47.2, 36.6. ESI+MS *m/z* 482.16 $[M + Na]^+$. HRMS (ESI) *m/z* $[M + H]^+$ calculated for $C_{27}H_{26}NO_6$ 460.1760, found 460.1768.

Synthesis of 1-(phenylmethyl)-4-methylester-L-aspartic acid, ***S3***



Compound **S2** (0.49 g, 1.08 mmol) was dissolved in 20% piperidine in DMF (15 ml) and stirred at room temperature for 5 hours. The solvent was evaporated under reduced pressure and the crude purified by column chromatography, 50% hexane 50% ethyl acetate to give the pure 1-(phenylmethyl)-4-methylester-L-aspartic acid (**S3**) as a yellow oil (0.23 g, 89.3%). ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ ppm 7.35 (m, 4 H), 5.19 (d, *J* = 12.2 Hz, 1 H), 5.15 (d, *J* = 12.2 Hz, 1 H), 3.88 (dd, *J* = 7.2 Hz, 4.7 Hz, 1 H), 3.65 (s, 3H), 2.84 (dd, *J* =16.6 Hz, 4.7 Hz, 1H), 2.74 (dd, *J* = 16.5 Hz, 7.2 Hz, 1 H). 1.98 (s, 3 H). ^{13}C NMR (100 MHz, $CDCl_3$, 298 K) δ ppm 173.9, 171.7, 135.5, 128.8, 128.6, 128.5, 67.3, 52.0, 51.4, 38.65. ESI+MS *m/z* 238.11 $[M + H]^+$. HRMS (ESI) *m/z* $[M + H]^+$ calculated for $C_{12}H_{16}NO_4$ 238.1079, found 238.1083.

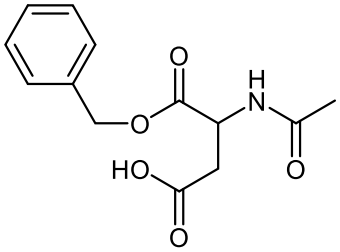
Synthesis of 1-(phenylmethyl)- L-aspartic acid, ***S4***



Fmoc-Asp-OBzl (**S1**) (0.50 g, 1.12 mmol) was dissolved in 20% piperidine in DMF (15 ml) and the reaction was stirred at room temperature for 5 hours. The solvent was removed under reduced pressure and the crude triturated with diethyl ether to give the pure 1-(phenylmethyl)-

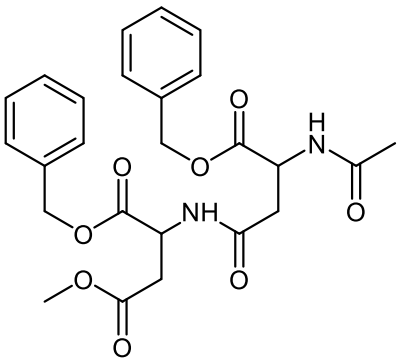
L-aspartic acid (**S4**) as a white-off solid (0.22 g, 86.2%). ^{1}H NMR (400 MHz, DMSO-d_6, 298 K) δ ppm 7.37 (d, J = 4.4 Hz, 4H), 7.33 (m, 1H), 5.12 (s, 2H), 3.74 (dd, J = 7.2 Hz, 5.4 Hz, 1H), 2.53 (dd, J = 16.4 Hz, 5.4 Hz, 1H), 2.44 (dd, J = 16.3, 7.2 Hz, 1H). ^{13}C NMR (101 MHz, DMSO-d_6, 298 K) δ ppm 173.9, 172.7, 136.5, 128.9, 128.4, 128.2, 66.3, 51.4, 38.8. ESI+MS *m/z* 224.09 $[M + H]^+$. HRMS (ESI) *m/z* $[M + H]^+$ calculated for $C_{11}H_{14}NO_4$ 224.0923, found 245.0923.

Synthesis of N-acetyl-1-(phenylmethyl)- L-aspartic acid, ***S5***



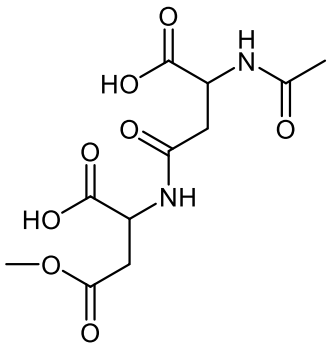
Compound **S4** (0.22 g, 0.96 mmol) was dissolved in DMF (15 ml) and the mixture was cooled to 0 °C. Acetic anhydride (0.46 ml, 4.81 mmol) was added dropwise and the reaction stirred at room temperature for 2 hours. The solvent was removed under vacuum to give N-acetyl-1-(phenylmethyl)- L-aspartic acid (**S5**) as a colourless oil (0.26 g, quant.). ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ ppm 7.33 (m, 5 H), 6.69 (d, J = 8.1 Hz,1 H), 5.20 (d, J = 12.3 Hz, 1 H), 5.17 (d, J = 12.3 Hz, 1 H), 4.89 (dt, J = 8.4 Hz, 4.3 Hz, 1 H), 3.09 (dd, J = 17.6 Hz, 4.3 Hz, 1 H), 2.87 (dd, J = 17.6 Hz, 4.3 Hz, 1 H), 2.03 (s, 3 H). ^{13}C NMR (100 MHz, $CDCl_3$, 298 K) δ ppm 174.6, 170.7, 170.5, 135.3, 128.7, 128.6, 128.3, 67.7, 48.7, 36.2, 23.2. ESI+MS *m/z* 266.11 $[M + H]^+$. HRMS (ESI) *m/z* $[M + H]^+$ calculated for $C_{13}H_{16}NO_5$ 266.1087, found 266.1028.

Synthesis of benzylated peptide scaffold, ***S6***



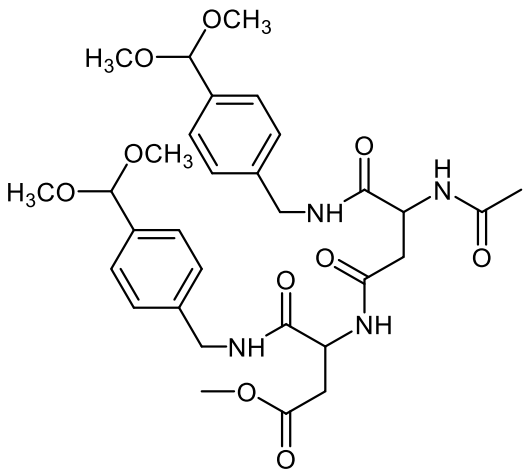
Compound **S3** (0.23 g, 0.96 mmol) and **S5** (0.25 g, 0.96 mmol) were dissolved in dichloromethane (20 ml). HOBt (0.13 g, 0.96 mmol) and EDCI (0.18 g, 0.96 mmol) were added and the reaction mixture was stirred at room temperature overnight. The mixture was washed with water (4 x 15 ml) and brine (3 x 5 ml). The organic layer was concentrated under reduced pressure and the crude purified by column chromatography, 50% hexane 50% ethyl acetate to 35% hexane 65% ethyl acetate. The solvent was removed under vacuum to give the pure **S6** as an off-white solid (0.43 g, 92.5%). ^{1}H NMR (500 MHz, $CDCl_3$, 298 K) δ ppm 7.33 (m, 10 H), 6.76 (d, *J* = 8.2 Hz, 1 H), 6.58 (d, *J* = 8.3 Hz, 1 H), 5.17 (m, 4 H), 4.89 (dt, *J* = 8.4 Hz, 4.3 Hz, 1 H), 4.84 (dt, *J* = 8.6 Hz, 4.5 Hz, 1 H), 3.60 (s, 3 H), 2.99 (dd, *J* = 17.0 Hz, 4.5 Hz, 1 H), 2.95 (dd, *J* = 15.7 Hz, 4.7 Hz, 1 H), 2.76 (dd, *J* = 15.7 H, 4.4 Hz), 2.71 (dd, *J* = 17.3 Hz, 4.6 Hz, 1 H), 1.99 (s, 3 H). ^{13}C NMR (126 MHz, $CDCl_3$, 298 K) δ ppm 171.4, 170.9, 170.4, 170.3, 170.0, 135.5, 135.2, 128.8, 128.7, 128.7, 128.5, 128.5, 128.4, 67.9, 67.6, 52.2, 49.1, 48.7, 37.7, 36.0, 23.2. ESI+MS *m/z* 485.19 $[M + H]^+$. HRMS (ESI) *m/z* $[M + Na]^+$ calculated for $C_{25}H_{28}N_2O_8Na$ 507.1743, found 507.1755.

Synthesis of deprotected peptide scaffold, ***S7***



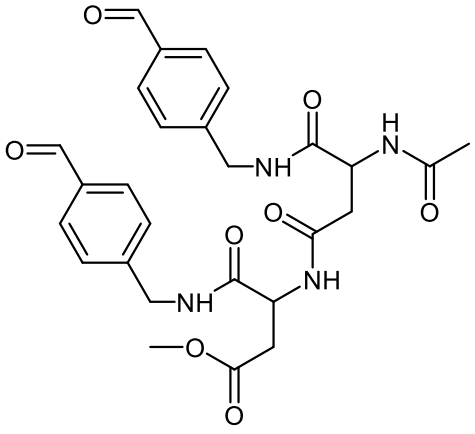
Compound **S6** (0.32 g, 0.65 mmol) was dissolved in anhydrous methanol (10 ml) and Pd/C (41.5 mg) was added. The flask was connected to a hydrogen balloon and the reaction was stirred at room temperature overnight. The reaction mixture was filtered over Celite and the solvent evaporated under reduced pressure to afford the pure **S7** (0.16 g, 87.4%). ^{1}H NMR (400 MHz, DMSO-d_6, 298 K) δ ppm 8.27 (d, J = 8.0 Hz, 1 H), 7.98 (d, J = 8.0 Hz, 1 H), 3.59 (s, 3 H), 2.71 (dd, J = 16.3 Hz, 6.1 Hz, 1 H), 2.63 (dd, J = 16.4 Hz, 7.0 Hz, 1 H), 2.57 (dd, J = 15.2 Hz, 5.4 Hz, 1 H), 1.99 (s, 3H).

Synthesis of functionalised-peptide scaffold, ***S16***



To a solution of **S7** (0.15 g, 0.49 mmol) in DMF (10 ml), **S15** (0.21 g, 1.18 mmol) was added. HOBt (0.14 g, 1.03 mmol), EDCI (0.20 g, 1.03 mmol) and DIEA (0.20 ml, 1.18 mmol) were added and the reaction was stirred overnight at room temperature. The solvent was removed under reduced pressure and the crude diluted with dichloromethane (50 ml) and washed with water (4 x 15 ml) and brine (3 x 10 ml). The organic phase was concentrated under vacuum and the crude was purified by recrystallisation with a mixture of ethyl acetate and methanol. The solid was filtered to afford the pure **S16** (0.26 g, 84.9%). ^{1}H NMR (500 MHz, DMSO-*d*$_6$, 298 K) δ ppm 8.44 (t, *J* = 6.2 Hz, 1 H), 8.40 (t, *J* = 6.1 Hz, 1 H), 8.32 (d, *J* = 8.3 Hz, 1H), 8.15 (d, *J* = 8.3 Hz, 1H), 7.28 (dd, *J* = 8.2 Hz, 3.1 Hz, 4 H), 7.22 – 7.19 (m, 4 H), 5.33 (s, 1 H), 5.32 (s, 1 H), 4.69 – 4.64 (m, 2 H), 4.29 – 4.19 (m, 4 H), 3.58 (s, 3 H), 3.21 (s, 6 H), 3.21 (s, 6 H), 2.78 (dd, *J* = 16.3 Hz, 6.1 Hz, 1 H), 2.65 (dd, *J* = 14.6 Hz, 7.2 Hz, 1 H), 2.59 (dd, *J* = 16.2 Hz, 7.5 Hz, 1 H), 2.34 (dd, *J* = 14.5 Hz, 7.2 Hz, 1 H), 1.83 (s, 3 H). ^{13}C NMR (126 MHz, DMSO-*d*$_6$, 298 K) δ ppm 171.8, 171.2, 170.7, 170.1, 169.9, 139.9, 139.9, 137.1, 127.3, 127.2, 126.8, 103.0, 52.9, 52.0, 50.5, 49.8, 42.5, 42.4, 38.4, 36.5, 23.2. ESI+MS *m/z* 653.28 $[M + Na]^+$. HRMS (ESI) *m/z* $[M + Na]^+$ calculated for $C_{31}H_{42}N_4O_{10}Na$ 653.2796, found 653.2795.

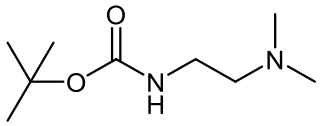
Synthesis of aldehyde-functionalised peptide scaffold, ***S17***



Compound **S16** (49.4 mg, 0.08 mmol) was dissolved in chloroform (5 ml) and 2 M HCl in diethyl ether (8 ml) was added and the rection stirred for 5 hours. The solvent was removed under reduced pressure to afford the pure **S17** (42.2 mg, quant.). ^{1}H NMR (500 MHz, DMSO-d_6, 298 K) δ ppm 9.95 (s, 1 H), 9.94 (s, 1 H), 8.54 (t, J = 6.0 Hz, 2 H), 8.41 (d, J = 8.2 Hz, 1 H), 8.22 (d, J = 8.1 Hz, 1 H), 7.81 – 7.78 (m, 4 H), 7.40 (t, J = 8.0 Hz, 4 H), 4.71 – 4.64 (m, 2 H), 4.37 – 4.26 (m, 4 H), 3.58 (s, 3 H), 2.79 (dd, J = 16.3 Hz, 6.5 Hz, 1 H), 2.69 (dd, J = 14.5 Hz, 7.2 Hz, 1 H), 2.61 (dd, J = 16.2 Hz, 7.5 Hz, 1 H), 2.37 (dd, J = 14.6 Hz, 7.2 Hz, 1 H), 1.84 (s, 3 H). ^{13}C NMR (126 MHz, DMSO-d_6, 298 K) δ ppm 193.1, 193.1, 172.0, 171.3, 170.9, 170.2, 170.0, 147.0, 146.9, 135.4, 135.4, 129.9, 128.0, 127. 9, 52.0, 50.5, 49.9, 42.6, 42.6, 38.4, 36.3, 23.2. ESI+MS *m/z* 539.21 $[M + H]^+$. HRMS (ESI) *m/z* $[M + H]^+$ calculated for $C_{27}H_{31}N_4O_8$ 539.2142, found 537.2141.

5.1.5 Synthesis of building blocks

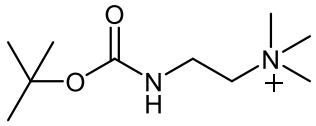
Synthesis of tert-butyl 2-(dimethylamino)ethylcarbamate [6]



Triethylamine (0.97 ml, 6.96 mmol) and Boc_2O (1.66 g, 7.61 mmol) were added to chloroform (15 ml). N,N-Dimethylethylenediamine (**AL11**) (0.65 ml, 5.95 mmol) was added and the mixture was stirred for 4 hours at room temperature. The reaction mixture was washed with sat. $NaHCO_3$ (3 x 5 ml) and water (3 x 5 ml). The solvent was removed under vacuum to give the pure tert-butyl 2-(dimethylamino)ethylcarbamate as a yellow solid (1.01 g, 89.9%). ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ ppm 5.02 (s, 1 H), 3.20 (q, J = 5.8 Hz, 2 H), 2.39 (t, J = 6.1 Hz, 2 H), 2.23 (s, 6 H), 1.44 (s, 9 H). ^{13}C NMR (100 MHz, $CDCl_3$, 298 K) δ ppm 156.2,

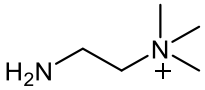
79.2, 58.5, 45.3, 38.0, 28.6. ESI+MS *m/z* 189.16 $[M + H]^+$. HRMS (ESI) *m/z* $[M + H]^+$ calculated for $C_9H_{21}N_2O_2$ 189.1603, found 189.1608.

Synthesis of tert-butyl 2-(trimethylamino)ethylcarbamate [6]



To a solution of tert-butyl 2-(dimethylamino)ethylcarbamate (1.01 g, 5.35 mmol) in methanol (15 ml), iodomethane (0.43 ml, 6.96 mmol) and potassium bicarbonate (0.61 g, 6.01 mmol) were added and the reaction stirred overnight at room temperature. The solvent was removed under vacuum and the crude dissolved in chloroform and filtered. The filtrate was evaporated under reduced pressure and the solid washed with diethyl ether to give tert-butyl 2-(trimethylamino)ethylcarbamate iodide salt (1.28 g, 72.3%). ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ ppm 5.90 (t, *J* = 5.9 Hz, 1H), 3.82 (t, *J* = 5.9 Hz, 2H), 3.72 (d, *J* = 6.3 Hz, 2H), 3.47 (s, 9H), 1.42 (s, 9H). ^{13}C NMR (101 MHz, $CDCl_3$, 298 K) δ ppm 156.41, 80.61, 65.99, 54.68, 35.70, 28.50. ESI+MS *m/z* 203.18 $[M]^+$. HRMS (ESI) *m/z* $[M]^+$ calculated for $C_{10}H_{23}N_2O_2$ 203.1759, found 203.1762.

Synthesis of 2-amino-N,N,N-trimethyl-ethanaminium **AL10**[6]



The tert-butyl 2-(trimethylamino)ethylcarbamate iodide salt (1.28 g, 3.88 mmol) was dissolved in anhydrous methanol (20 ml) and the solution cooled to 0 °C. Acetyl chloride (4 ml) was added dropwise and the reaction was stirred at 0 °C for 3 hours. The solvent was removed under

vacuum and the crude was washed with diethyl ether to give the pure 2-amino-N,N,N-trimethyl-ethanaminium dihydrochloride salt **AL10** as a dark yellow solid (0.59 g, 86.7%). ^{1}H NMR (400 MHz, DMSO-d_6, 298 K) δ ppm 8.67 (s, 3H), 3.67 – 3.60 (m, 2H), 3.33 – 3.26 (m, 2H), 3.16 (s, 9H). ^{13}C NMR (101 MHz, DMSO-d_6, 298 K) δ ppm 61.29, 52.82, 32.45. ESI+MS *m/z* 103.12 $[M]^+$. HRMS (ESI) *m/z* $[M]^+$ calculated for $C_5H_{15}N_2$ 103.1235, found 103.1230.

5.2 Enzymatic synthesis

5.2.1 Materials and methods

Reagents stock solutions were prepared in 50 mM TRIS buffer (pH 7.3 – 7.6). The pH of the buffer was adjusted with 1 M NaOH and 1 M HCl. Methyl-β-D-galactoside (4.9 mg) was dissolved in TRIS buffer (400 µl) to give a 63 mM stock solution. CMP-sialic acid (9.5 mg) was dissolved in TRIS buffer (500 µl) to give a 30.9 mM stock solution. A stock solution of $MgCl_2$ (10 mM) was prepared from dilution of a 200 mM solution. A stock solution of $MgCl_2$ (10 mM) was prepared from dilution of a 1 M solution. The enzymes PmST1 M144D, PmST1 P34H/M144L and ST3Gal4, B4GalT1, CIAP were available as solutions of approximate concentration of 1 mg/ml. The reaction mixture was analysed by LC-MS with a Shimadzu LCMS-IT-TOF instrument with a HILIC column (XBridge® Amide 5 µm, 4.6 mm x 250 mm column, Waters) with a binary gradient 95% acetonitrile 5% water to 50-50% and a flow rate of 0.40 ml/min. The MS was recorded in positive and negative ion mode. Thin layer chromatography (TLC) was performed on Silicagel 60 F254 plates with a 4:2:1 mixture of ethyl acetate:methanol:water as eluent and detection by UV light (254 nm) and staining with 5% H_2SO_4 in ethanol followed by heating. The purification by size exclusion chromatography was performed with a Bio-Gel P-2 column and MilliQ water as eluent. The purification by ion exchange was performed on a DEAE Sepharose FF column with MillQ water and ammonium

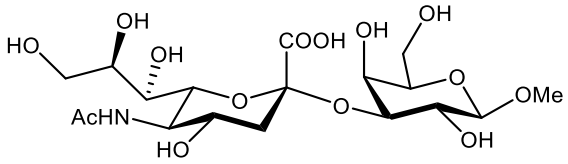
carbonate buffer (20 – 50 mM) as eluents. The NMR spectra (^{1}H, COSY, TOCSY, HSQC) were recorded on a Bruker AVANCE Neo 600 MHz and on a Bruker AVANCE NEO 500 MHz in D_2O.

5.2.2 Expression of the PmST1 M144D enzyme

The enzyme PmST1 M144D was expressed following a published procedure[7]. A preculture of *E. coli*. BL21 (DE3) (3 ml) containing the recombinant plasmid, was added to a to lysogen broth (LB) rich medium (400 ml). Ampicillin (100 µg/mL) was added and the medium incubated at 37 °C under 200 RPM shaking and cell growth monitored by UV (600 nm) every 30 mins. After 4 hours the OD_{600} reached 0.69 and a stock solution of isopropyl β-D-1-thiogalactopyranoside was added (20 µl, 0.1 mM) and the cells were incubated overnight at room temperature under shaking. The cells were centrifuged at 4 °C at 3900 rpm for 50 mins and then separated from the media. The cells were treated with a solution of Triton X-100 in TRIS buffer (5 ml), lysozyme (10 µl) and DNase (4 µl) were also added. This was incubated for 50 min at 37 °C under shaking and then centrifuged at 5 °C for 35 minutes at 7000 RPM. The collected supernatant was purified with a Ni^{2+}-NTA column with MilliQ water and imidazole buffer (5 – 200 mM) as eluent. The collected vials were analysed by UV (340 nm) and their concentration was comprised between 0.6 and 0.9 mg/ml. Samples (10 µl) from each vial were denatured at 100 °C for 10 min and analysed by SDS-PAGE gel run in buffer at 200 V for 30 minutes. The gel was placed in water for 20 mins and then stained with Coomassie Blue stain whilst shaking for 1 hour. The excess of stain was removed by shaking the gel in water for 1 hour and the gel was then analysed confirming the presence of the PmST1 M144D enzyme.

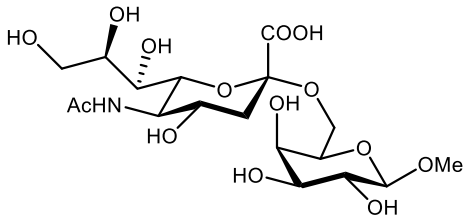
5.2.3 Enzymatic synthesis of sialylated glycans

Synthesis of SAα2-3GalβMe with bacterial enzyme PmST1 M144D



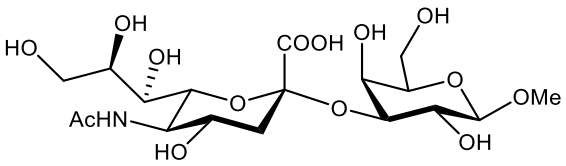
12 µl of methyl-β-D-galactoside stock solution (0.5 µmol) and 22.7 µl stock solution of CMP-sialic acid (1.5 eq.) were added in TRIS buffer (428 µl). The enzymes PmST1 M144D (10 µL) and CIAP (2 µl) were added to the solution followed by the addition of the $MgCl_2$ (25 µl) stock solution. The vial was incubated overnight at 37 °C. The mixture was analysed by LC-MS and TLC, no product was detected.

Synthesis of SAα2-6GalβMe with bacterial enzyme PmST1 P34H/M144L



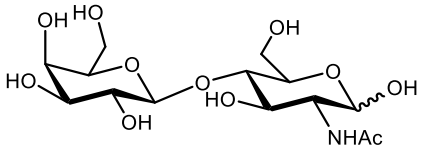
12 µl of methyl-β-D-galactoside stock solution (0.5 µmol) and 22.7 µl stock solution of CMP-sialic acid (1.5 eq.) were added in TRIS buffer (428 µl). The enzymes PmST1 P34H/M144L (10 µL) and CIAP (2 µl) were added to the solution followed by the addition of the $MgCl_2$ (25 µl) stock solution. The vial was incubated overnight at 37 °C. The mixture was analysed by LC-MS and TLC, no product was detected.

Synthesis of SAα2-3GalβMe with mammalian enzyme ST3Gal4



2.9 µl of methyl-β-galactoside stock solution (0.5 µmol) and 7.7 µl stock solution of CMP-sialic acid (1.5 eq.) were added in TRIS buffer (33.4 µl). BSA (0.5 µl) was added to the solution, followed by the addition of the enzymes ST3Gal4 (5 µL) and CIAP (0.5 µl). The vial was incubated for 3 hours at 37 °C. The LC-MS analysis in positive mode showed peaks of the product *m/z* = 508.2 $[M + Na]^+$ and *m/z* = 530.2 $[M + 2Na]^+$ and peaks of the unreacted CMP-sialic acid *m/z* = 681.1 $[M + 3Na^+]$. In negative mode a peak of the product *m/z* = 484.1 $[M - H]^-$ was observed, in addition to unreacted CMP-sialic acid *m/z* = 635.1 $[M + Na]^+$. The TLC analysis showed no traces of starting material indicating full conversion.

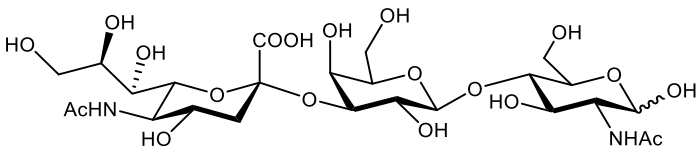
Synthesis of LacNAc



GlcNAc (50.0 mg, 0.23 mmol) and UDP-Gal (165 mg, 0.29 mmol) were added to TRIS buffer (50 mM, 10 ml) at pH 7.5. The 100 µl of stock solution of $MnCl_2$ (10 mM) was added, followed by the addition of B4GalT1 (200 µl) and CIAP (100 µl). The mixture was incubated overnight at 37 °C. The mixture was analysed by LC-MS and the product was observed, *m/z* = 406.1 $[M + Na]^+$ and *m/z* = 789.3 $[2M + Na]^+$. The TLC analysis did not show the presence of starting material. The reaction was freeze dried and purified by size exclusion with a P2 column eluting with MilliQ water. The collected fractions were analysed by TLC and the fractions containing

material were analysed by LC-MS and the ones containing products were freeze dried to afford LacNAc (84.3 mg, quant.) as a white solid, which was used without further purification.

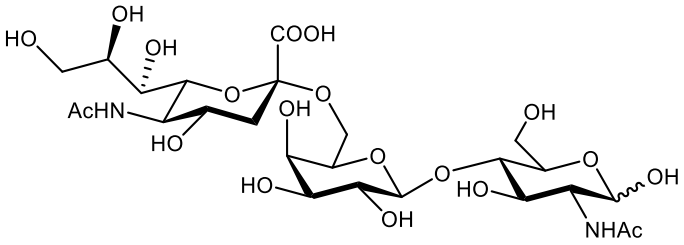
Synthesis of SAα2-3LacNAc



LacNAc (132.9 mg, 0.35 mmol) and CMP-sialic acid (243 mg, 0.39 mmol) were added to the TRIS buffer (50 mM, 11.7 ml) at pH 7.6. BSA (10 mg) was added, followed by the addition of ST3Gal4 (400 µl) and CIAP (100 µl). The vial was incubated overnight at 37 °C. The LC-MS analysis confirmed the presence of product *m/z* = 697.2 $[M + Na]^+$. The mixture was freeze dried and the resulting solid purified by size exclusion chromatography on a P2 column eluting with MilliQ water. The collected fractions were analysed by TLC and the ones presenting material were analysed by LC-MS. The fractions containing product were freeze dried. The solid was dissolved in water and loaded on a DEAE Sepharose FF column. The column was initially eluted with MilliQ water and then with 20 – 50 mM ammonium carbonate buffer. The collected fractions were analysed by TLC and the ones presenting material were analysed by LC-MS. The fractions containing pure product were freeze dried to afford the pure SAα2-3LacNAc as a white solid (55.4 mg, 23.7%). ^{1}H NMR (500 MHz, D_2O, 298 K) δ ppm 5.21 (d, *J* = 2.3 Hz, 1 H), 4.73 (d, *J* = 7.9 Hz, 1 H), 4.56 (dd, *J* = 7.9, 2.9 Hz, 1 H), 4.13 (m, 1 H), 3.98 (m, 1 H), 3.98 (m, 1 H), 3.98 (m, 1 H), 3.97 (m, 1 H), 3.91 (m, 1 H), 3.91 (m, 1 H), 3.90 (m, 1 H), 3.90 (m, 1 H), 3.88 (m, 1 H), 3.85 (m, 1 H), 3.86 (m, 1 H), 3.84 (m, 1 H), 3.76 (m, 1 H), 3.76 (m, 1 H), 3.75 (m, 1 H), 3.74 (m, 1 H), 3.73 (m, 1 H), 3.71 (m, 1 H), 3.65 (m, 1 H), 3.65 (m, 1 H), 3.61 (m, 1 H), 3.59 (m, 1 H), 2.77 (dd, *J* = 12.5, 4.5 Hz, 1 H), 2.05 (s,

3 H), 2.04 (s, 3 H), 2.04 (s, 3 H), 1.81 (t, J = 12.7 Hz, 1 H). ^{13}C NMR (126 MHz, D_2O, 298 K) δ ppm 102.6, 94.9, 90.5, 78.5, 78.4, 75.5, 75.2, 74.9, 72.9, 70.2, 70.2, 70.2, 69.2, 69.2, 68.7, 68.3, 67.5, 67.4, 62.6, 59.9, 59.9, 56.2, 51.7, 39.6, 22.0, 22.0, 22.0. ESI+MS *m/z* 673.23 $[M - H]^-$. HRMS (ESI) *m/z* $[M - H]^-$ calculated for $C_{25}H_{41}N_2O_{19}$ 673.2303, found 673.2297.

Synthesis of SAα2-6LacNAc



LacNAc (84.3 mg, 0.22 mmol) and CMP-sialic acid (154.0 mg, 0.24 mmol) were added to the TRIS buffer (50 mM, 13.9 ml) at pH 7.6. 1 ml of stock solution of $MgCl_2$ (1.0 ml) was added, followed by the addition of PmST1 P34H/M144L (5.0 ml) and CIAP (100 µl). The vial was incubated at 37 °C for 3 hours. The LC-MS analysis confirmed the presence of product *m/z* = 697.2 $[M + Na]^+$ and *m/z* = 384.1 $[M + H]^+$. The mixture was freeze dried and the resulting solid dissolved in water (3.0 ml) and filtered through a spin filter at 3000 rpm for 9 hours. The resulting solution was freeze dried and then the solid was dissolved in water and loaded on a DEAE Sepharose FF column. The column was initially eluted with MilliQ water and then with 20 – 50 mM ammonium carbonate buffer. The collected fractions were analysed by TLC and the ones presenting material were analysed by LC-MS. The fractions containing pure product were freeze dried to afford the pure SAα2-6LacNAc as a white solid (79.7 mg, 53.7%). ^{1}H NMR (600 MHz, D_2O, 298 K) δ ppm 5.22 (d, J = 2.7 Hz, 1 H), 4.72 (d, J = 7.7 Hz, 1 H), 4.43 (dd, J = 7.9, 2.1 Hz, 1 H), 3.98 (m, 1 H), 3.98 (m, 1 H), 3.95 (m, 1 H), 3.95 (m, 1 H), 3.95 (m, 1 H), 3.90 (m, 1 H), 3.90 (m, 1 H), 3.89 (m, 1 H), 3.87 (m, 1 H), 3.87 (m, 1 H), 3.86

(m, 1 H), 3.82 (m, 1 H), 3.80 (m, 1 H), 3.79 (m, 1 H), 3.70 (m, 1 H), 3.68 (m, 1 H), 3.65 (m, 1 H), 3.63 (m, 1 H), 3.63 (m, 1 H), 3.60 (m, 1 H), 3.54 (m, 1 H), 3.53 (m, 1 H), 3.52 (m, 1 H), 2.65 (dd, *J* = 12.4, 4.7 Hz, 1 H), 2.04 (s, 3 H) 2.00 (s, 3 H), 2.00 (s, 3 H), 1.70 (t, *J* = 12.2 Hz, 1 H). ^{13}C NMR (151 MHz, D_2O, 298 K) δ ppm 103.5, 94.6, 90.5, 83.4, 74.6, 73.6, 72.6, 72.5, 71.7, 70.7, 70.1, 69.9, 68.6, 68.4, 68.2, 68.1, 63.3, 62.6, 60.3, 60.1, 51.7, 53.5, 73.6, 40.1, 22.1, 22.0, 22.0. ESI+MS *m/z* 673.23 $[M - H]^-$. HRMS (ESI) *m/z* $[M - H]^-$ calculated for $C_{25}H_{41}N_2O_{19}$ 673.2303, found 673.2299.

5.3 Binding studies

5.3.1 NMR and MS binding studies

^{1}H NMR spectra were recorded at room temperature on Bruker AVANCE NEO at 400 MHz at 400 MHz. Samples were dissolved in deuterated 0.1 M CD_3COOD buffer and 0.1 M PB buffer at pD 5.1 and 7.0, respectively. The deuterated 0.1 M CD_3COOD buffer was prepared with D_2O and acetic acid-d_4. The deuterated 0.1 M PBS buffer was prepared by dissolving K_2HPO_4 and KH_2PO_4 in D_2O, the solution was the dried under vacuum and the resulting solid re-dissolved int D_2O. The pH of the buffer was adjusted with a 5 M NaOD solution in D_2O prepared following the same procedure described for the 0.1 M PB buffer. MS spectra were recorded with a Waters Xevo G2-XS spectrometer in ESI negative mode with the samples in 0.1 M acetate buffer.

5.3.2 Isothermal titration calorimetry binding studies

ITC experiments were performed with 2 mM receptor solution (**B1** and **B2** – **B7**) and 80 mM solution of ligand (**S1**, **S2**, **S3**). Five different buffers were used: 0.1 M acetate buffer (pH 5.5),

0.1 M phosphate buffer (pH 6.5), 0.1 M phosphate buffer (pH 7.4), 0.1 M ammonium acetate buffer (pH 8.5) and 0.1 M ammonium acetate (pH 10). The buffer was degassed prior to solution preparation. The solution pH was adjusted with 5 M HCl or 5 M NaOH. The solutions were filtered and degassed prior to use. ITC experiments were performed using a VP ITC MicroCalorimeter with the parameters reported in Table 5.1 by titrating the ligand into the receptor solution.

Table 5.1. ITC parameters for the binding studies with VP-ITC MicroCalorimeter.

Number of injections	35
Cell temperature	25 °C
Reference Power	10 – 25 µcal s^{-1}
Initial delay	60 s
Stirring	307 RPM
Volume of each injection	7.0 µl
Duration of each injection	4.0 s
Spacing	280 – 480 s
Filter	2

The experiments were performed at 25 °C. The volume of the first injection for each experiment was 2 µl and it was discarded, whilst the volume of all other injections was set at 7 µl. The spacing and the reference power selected for each experiment were based on the amount of heat released. For experiments with a small heat release (< 2.5 µcal s^{-1}), the spacing was set at 280 s and the reference power at 10 µcal s^{-1}. For titrations with large heat release (> 10 µcal s^{-1}), the spacing was set at 480 s and the reference power at 25 µcal s^{-1}. Experiments with

intermediate heat release had a spacing between 300 and 380 s and a reference power of 15 - 20 μcal s^{-1}. Each experiment consists of 3 titrations and the heat of dilution was measured and subtracted for each experiment. The data was analysed with Origin software for ITC, fixing the number of binding sites to 1.

List of references

1. Hernandez, V.; Crepin, T.; Palencia, A.; Cusack, S.; Akama, T.; Baker, S. J.; Bu, W.; Feng, L. S.; Freund, Y. R.; Liu, L.; Meewan, M.; Mohan, M.; Mao, W. M.; Rock, F. L.; Sexton, H.; Sheoran, A.; Zhang, Y. C.; Zhang, Y. K.; Zhou, Y.; Nieman, J. A.; Anugula, M. R.; Keramane, E.; Savariraj, K.; Reddy, D. S.; Sharma, R.; Subedi, R.; Singh, R.; O'Leary, A.; Simon, N. L.; De Marsh, P. L.; Mushtaq, S.; Warner, M.; Livermore, D. M.; Alley, M. R. K.; Plattner, J. J., Discovery of a Novel Class of Boron-Based Antibacterials with Activity against Gram-Negative Bacteria. *Antimicrob. Agents Chemother.* **2013,** *57* (3), 1394-1403.

2. Shapiro, A. B.; Gao, N.; Hajec, L.; McKinney, D. C., Time-dependent, reversible, oxaborole inhibition of Escherichia coli leucyl-tRNA synthetase measured with a continuous fluorescence assay. *Anal. Biochem.* **2012,** *431* (1), 48-53.

3. Hernandez, V. S. Benzoboroxole derivatives for treating bacterial infections. 2012. WO12/033858.

4. Carter, T. S.; Mooibroek, T. J.; Stewart, P. F. N.; Crump, M. P.; Galan, M. C.; Davis, A. P., Platform Synthetic Lectins for Divalent Carbohydrate Recognition in Water. *Angew. Chem. Int. Ed.* **2016,** *55* (32), 9311-9315.

5. Wen, L. Q.; Li, M.; Schlenoff, J. B., Polyporphyrin thin films from the interfacial polymerization of mercaptoporphyrins. *J. Am. Chem. Soc.* **1997,** *119* (33), 7726-7733.

6. Torde, R. G.; Therrien, A. J.; Shortreed, M. R.; Smith, L. M.; Lamos, S. M., Multiplexed Analysis of Cage and Cage Free Chicken Egg Fatty Acids Using Stable Isotope Labeling and Mass Spectrometry. *Molecules* **2013,** *18* (12), 14977-14988.

7. Yu, H.; Chokhawala, H.; Karpel, R.; Yu, H.; Wu, B. Y.; Zhang, J. B.; Zhang, Y. X.; Jia, Q.; Chen, X., A multifunctional Pasteurella multocida sialyltransferase: A powerful tool for the synthesis of sialoside libraries. *J. Am. Chem. Soc.* **2005,** *127* (50), 17618-17619.

Chapter 6 - Conclusions and future work

Alteration in the expression of carbohydrates in cells usually indicates underlaying pathological conditions. Healthy and cancerous cells express different arrays of glycans with certain key modifications being linked to a diverse range of tumours. One of the main alterations is higher degree of sialyation in the terminal end of glycans produced by cancerous cells. Furthermore, high levels of sialic acid, in both its free and unbound forms, are also observed in other pathological conditions such as alcoholism and sialic acid storage diseases, which are neurodegenerative diseases characterised by high secretion of sialic acid. Sialylated glycans can be exploited as biomarkers in cancer, whilst the monosaccharide sialic acid can be employed as a diagnostic marker for alcoholism-related liver dysfunction and the aforementioned neurodegenerative diseases. The detection of this negatively charged monosaccharide is therefore of significant value in diagnostics.

In Chapter 3, benzoboroxole receptors, boronic acid analogues, were developed for sialic acid sensing. Boronic acids are a class of compounds known for their ability to bind sugar's diol forming covalent boronate ester complexes. Benzoboroxoles are superior to their acyclic analogues as they display improved solubility and binding profile (*i.e.* pH, and binding to pyranose rings). Nevertheless, only a few examples of benzoboroxole receptors for sialic acid can be found in the literature. One of the main concerns regarding the use of boron-based receptors in sialic acid recognition involves the debate surrounding the binding site. The majority of publications detailing boron-based receptors for sialic acid based their binding model on the receptor binding to the diols of the glycerol chain. On the other hand, a few publications seem to disregard the aforementioned binding model and propose the

α-hydroxyacid group as either the sole point of interaction or as an additional binding site. Consequently, it was pivotal to firstly identify the binding site unequivocally. Calorimetric and analytical studies were conducted, on a wide pH range, to assess the binding between non-functionalised benzoboroxole **B1** and different sialic acid derivatives. Binding was observed exclusively with the free form of sialic acid, the only sugar derivative displaying the α-hydroxyacid moiety unaltered. Therefore, boron-based receptors bind sialic acid exclusively via the α-hydroxyacid moiety, with no binding occurring with the glycerol chain at any pH value. The studies conducted by isothermal titration calorimetry (ITC) were also able to confirm that the highest binding affinity between non-functionalized benzoboroxole **B1** and sialic acid is observed at acidic pH with $K_a = 51.2 \pm 1.2\ M^{-1}$, as was previously detailed in the literature. With the binding site being established an explanation for this anomalous binding was proposed. Boron-based receptors bind neutral sugars (*e.g.* fructose, glucose, galactose) preferentially at basic pH via the negative and tetrahedral boron centre. However, with sialic acid the binding at basic pH is greatly reduced due to the repulsion between the boron centre and the α-hydroxyacid group which are both negatively charged. Conversely, the binding at acidic pH is more favourable due to the lack of significant repulsive interactions between the neutral trigonal boron and the negatively charged α-hydroxyacid, at this pH.

Following this, a series of positively charged functionalised benzoboroxole receptors **B2** – **B7** were designed and synthesised for sialic acid recognition. Additional functional groups were introduced to provide additional non-covalent interactions and determine affinity enhancement. Positively charged groups, amino and guanidino, were introduced to afford charge-reinforced hydrogen bonds with sialic acid acceptors which can provide a significant contribution to the binding affinity. The charged groups were linked to the oxaborole ring of the benzoboroxole via chains of different length. In addition, for two of the six functionalised receptors, receptor

B6 and **B7**, an aromatic ring was also introduced to provide hydrophobic and CH-π interactions, as observed in lectins. The designed receptors **B2** – **B7** were synthesised in moderate to good yields and their binding profile to sialic acid measured by ITC for a wide pH range. Functionalised receptors **B2** – **B7** present their highest binding affinity at pH 5.5, as observed for **B1**, and their lowest K_a at basic pH. This is consistent with the binding profile of boron-based receptors to sialic acid, as the main binding event is the esterification, which is more favourable at acidic pH as detailed above. Furthermore, the functionalised receptors **B2** – **B7** displayed higher binding constants than non-functionalised benzoboroxole **B1** across the whole pH range. This is due to the multipoint interaction offered by non-covalent interactions between the additional functional groups and sialic acid. At acidic pH, the main contribution to the binding is offered by charge-reinforced hydrogen bonds between the charged amino and guanidino groups and sialic acid hydroxyl groups. In particular, guanidino receptors **B3**, **B5** and **B7** display enhanced binding affinity, when compared to their corresponding amino receptors **B2**, **B4** and **B6**, as the guanidino group can create a wider network of hydrogen bonds providing further stabilisation of the complex. In addition, ion paring between the charged head and the boron centre are also present and contribute to the binding. The highest binding affinity at pH 5.5 is achieved by the amino receptor **B2** (K_a = 150.4 ± 7.9 M^{-1}) and its corresponding guanidino receptor **B3** (K_a = 234.3 ± 8.0 M^{-1}) which present a 3- and 4.7-fold increase in the binding affinity, respectively, when compared to non-functionalised benzoboroxole **B1**. For receptors **B4 – B7**, which present longer side chains, the affinity is decreased at pH 5.5, when compared to **B2** and **B3**, as they cannot provide comparable interactions and stabilisation due to the charged group being further away from the boron centre. Conversely, at more basic pH, receptors **B4 – B7** displayed higher affinity than shorter derivatives **B2** and **B3**. This is due to the charge-reinforced hydrogen bonds being less prevalent at higher pH values with other

interactions prevailing, such as neutral hydrogen bonds and CH-π interactions. Nevertheless, the contribution from these interactions is lower and this, combined to the repulsion observed at basic pH, results in a decreased affinity at pH above 5.5. It was therefore established that charge-reinforced hydrogen bonds play a key role in enhancing the affinity of receptors to sialic acid at acidic pH. Furthermore, studies with a control molecule (**B13**) lacking the boron unit displayed no significant binding. The boronate esterification is therefore pivotal for the binding and it is only after this covalent event that additional non-covalent interactions can be formed and provide stabilisation to the complex, highlighting the cooperative nature of the interaction. Furthermore, the selectivity against neutral and anionic monosaccharide at acidic pH was also probed, with receptor **B3** presenting both the highest affinity and selectivity for sialic acid. These findings suggest that receptor **B3** can be employed for sialic acid sensing as part of diagnostic testing. Future work to enable this would include tethering of the receptor to surfaces or nanoparticles to allow the measurement of sialic acid levels in biological samples by analytical techniques such as surface plasmon resonance and fluorescence.

In Chapter 4, studies involving dynamic combinatorial chemistry were initiated with the aim of applying this technique in the generation of highly selective receptors for glycans in the future. Sialosides are particularly relevant biomarkers and they have been linked to several cancerous conditions. Glycans terminating with a sialic acid unit bound in the α2-3 fashion have been linked to cancer, whilst their α2-6 isomers are expressed by healthy cells. For the future development of receptors able to discriminate between the two sialosides it is pivotal to firstly synthesise the epitopes of interest. Sialylated oligosaccharide, SAα2-3LacNAc and SAα2-6LacNAc, were enzymatically synthesised to be representative of the sialosides terminal ends. The disaccharide LacNAc was enzymatically obtained and then sialylated in α2-3 and α2-6 fashion exploiting both bacterial and mammalian enzymes to afford the final epitopes

SAα2-3LacNAc and SAα2-6LacNAc, respectively. In order to develop a highly selective receptor for the α2-3 epitope, the initial steps in the application of dynamic combinatorial chemistry in the glycosensing field were taken. A series of considerations were discussed concerning the application of DCC in this field, including the reversible reaction at its basis (*i.e.* imine exchange) and various reaction conditions and parameters. Moreover, the detection of glycans requires a high degree of multivalency (*e.g.* cluster effect) to be displayed by the receptor in order to achieve high affinity and selectivity. Thus, a scaffold-based approach was selected to be employed in the DCC. A dipeptidomimetic backbone (**S7**) was synthesised in solution and appended with aldehyde groups (**S15**) to give the scaffold **S17** which will find application in the DCC experiments. A series of amine-bearing building blocks containing sugar binding moieties were selected, to be reversibly bound to the scaffold, and divided into different sets to be independently used in the DCC experiments. The amine building blocks will reversibly bind the aldehyde- appended scaffold (**S17**) via imine exchange to give a dynamic combinatorial library to which a template can be added to generate selective receptors. Disaccharides **T1**, **T2** and **T3** were selected as templates to probe the DCC application in the development of highly selective receptors prior to its application to the cancer-related sialosides. This preliminary work sets the basis for the future application of dynamic combinatorial chemistry in glycosensing. Preliminary studies will be conducted to optimise the experimental conditions (*e.g.* pH and concentration) under which the dipeptidomimetic scaffold and the building blocks generate a library under thermodynamic control. Following this the disaccharide templates **T1**, **T2** and **T3** will be employed to gauge whether they can produce distinct DCLs and to optimise the experimental conditions under which these are achieved. The analysis of the libraries by chromatographic techniques will require development of new methods in order to resolve the adducts' peaks and allow identification of these by mass

spectrometry. Following this, the optimised methodology will be applied to generate highly selective receptors for biologically relevant glycans.

Appendix

NMR spectra of benzoboroxole receptors

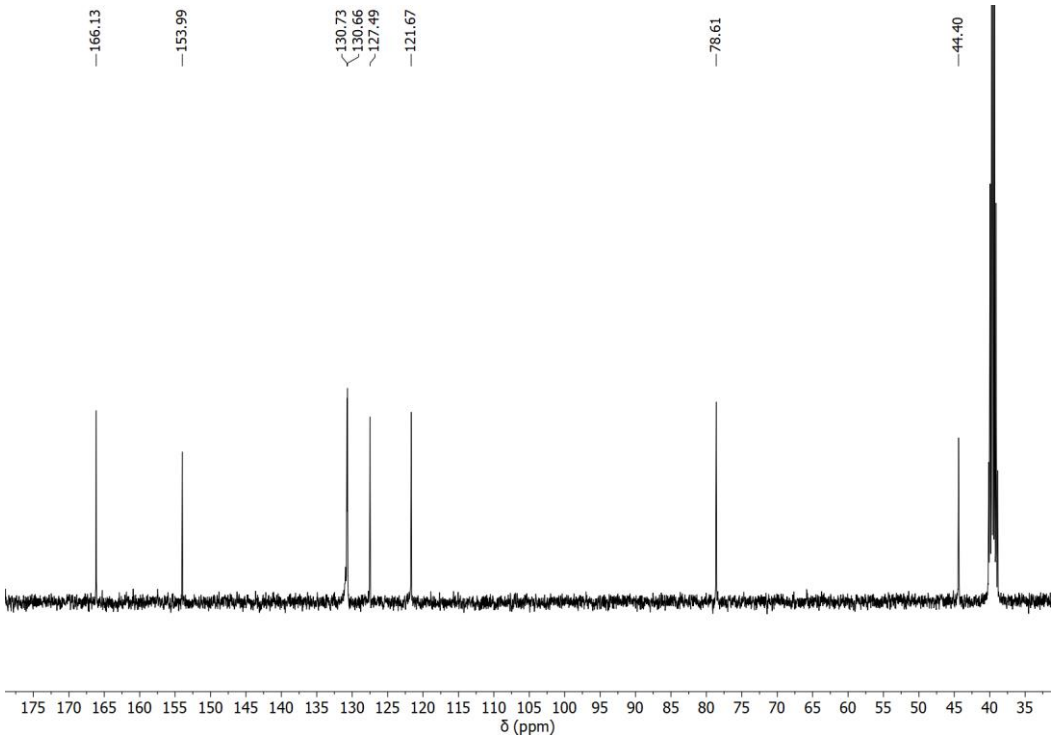


Figure 1. ^{13}C NMR spectrum of **B4** (101 MHz, DMSO-d_6, 298 K).

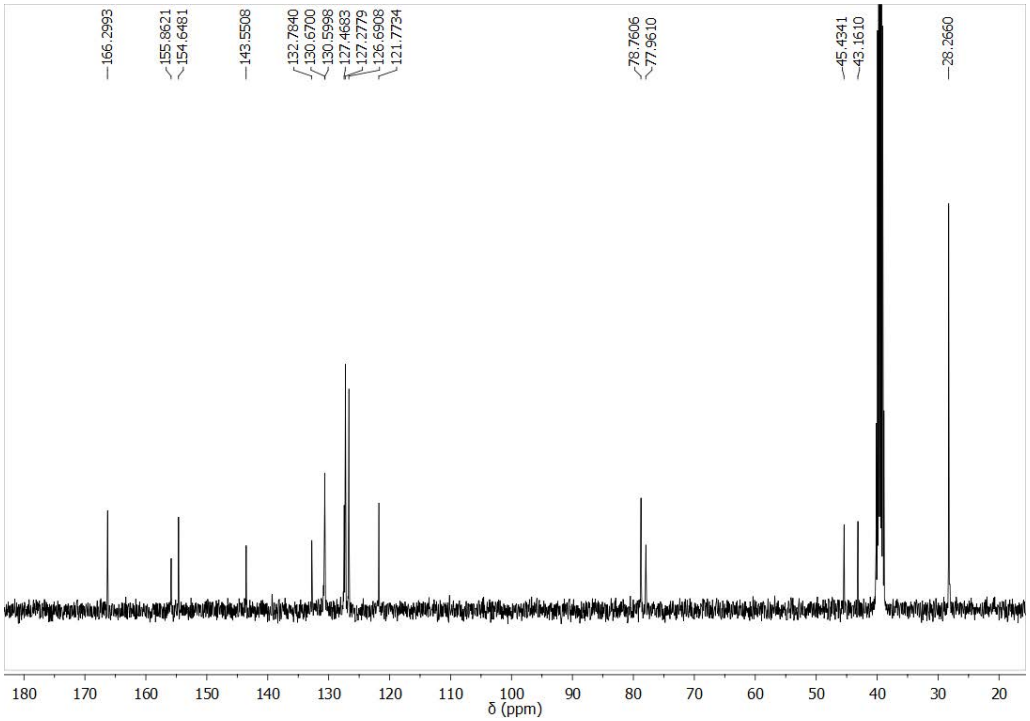


Figure 2. ^{13}C NMR spectrum of **B9** (101 MHz, DMSO-d_6, 298 K).

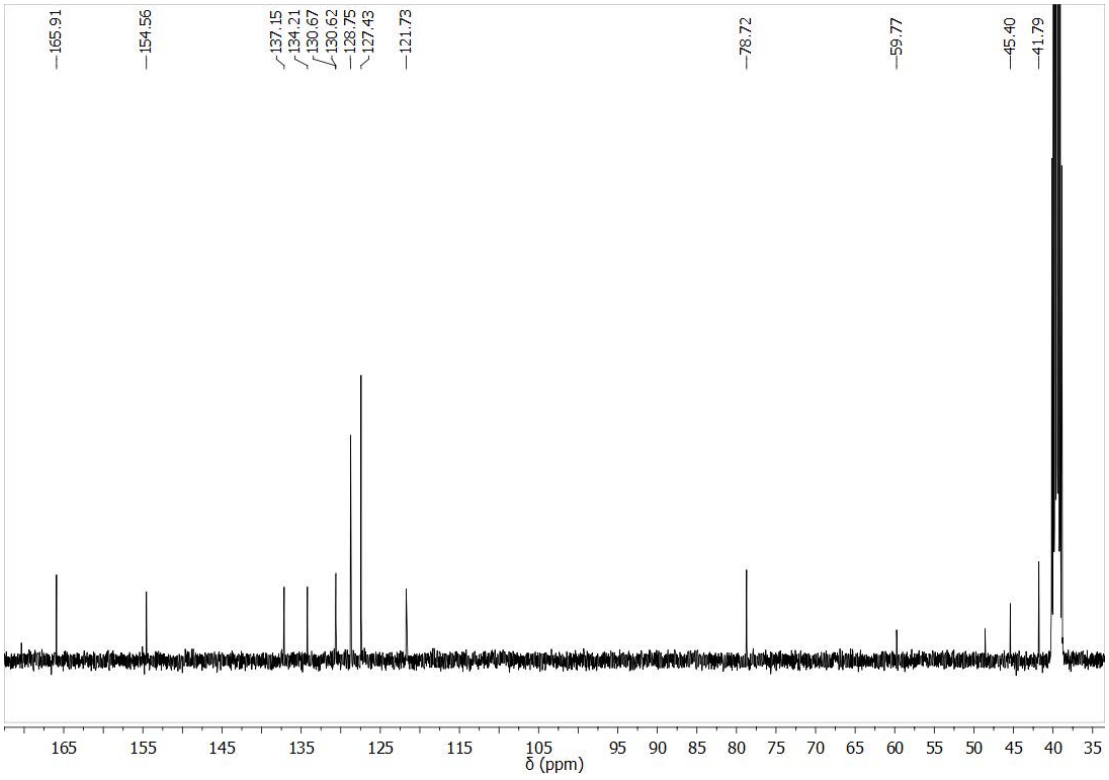


Figure 3. ^{13}C NMR spectrum of **B6** (101 MHz, DMSO-d_6, 298 K).

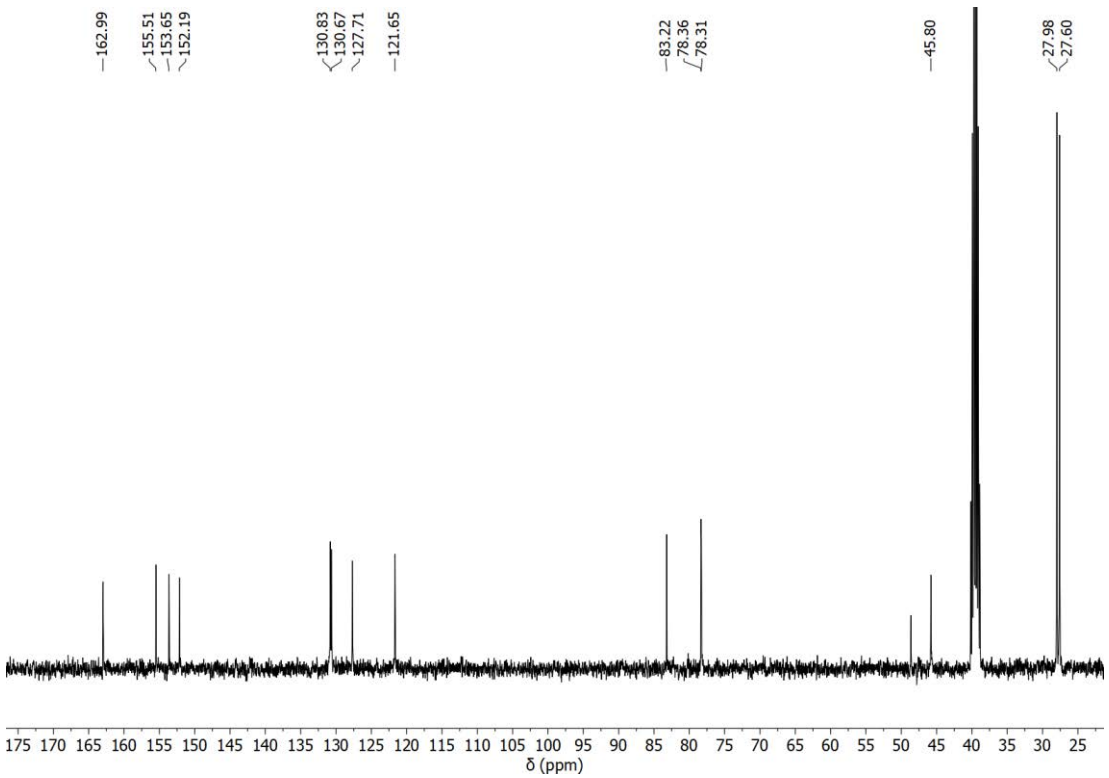


Figure 4. ^{13}C NMR spectrum of **B10** (101 MHz, DMSO-d_6, 298 K).

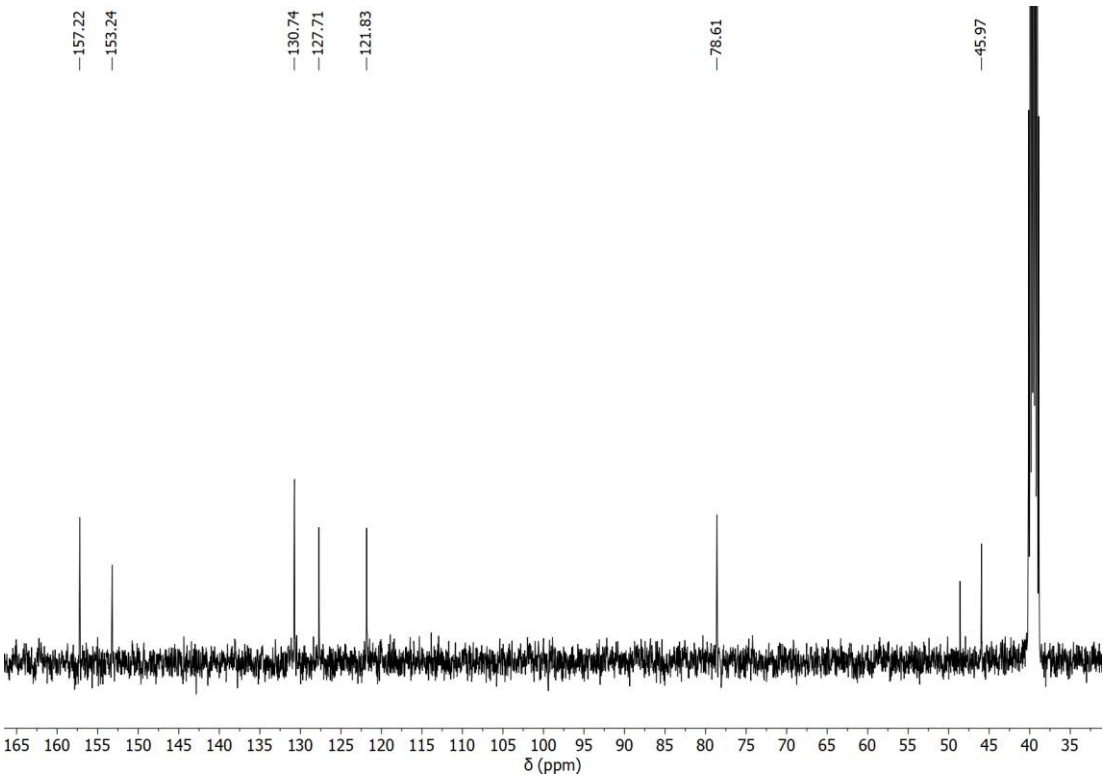


Figure 5. ^{13}C NMR spectrum of **B3** (101 MHz, DMSO-d_6, 298 K).

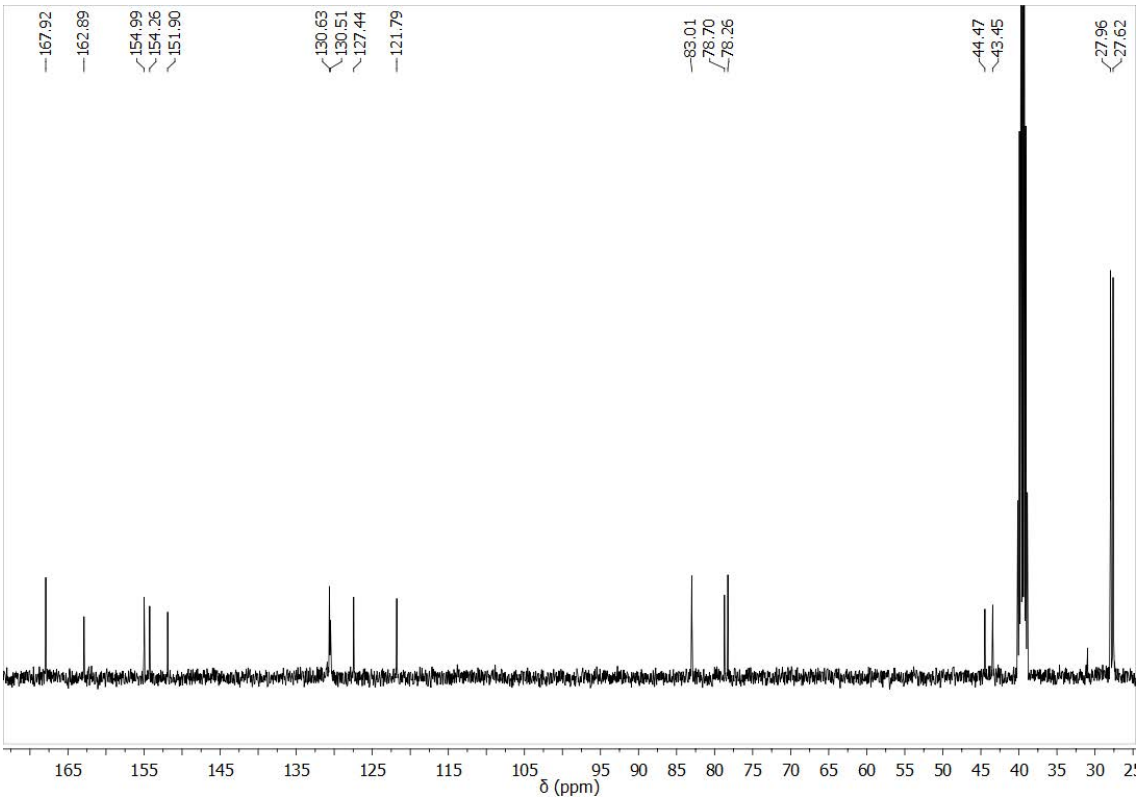


Figure 6. ^{13}C NMR spectrum of **B11** (101 MHz, DMSO-d_6, 298 K).

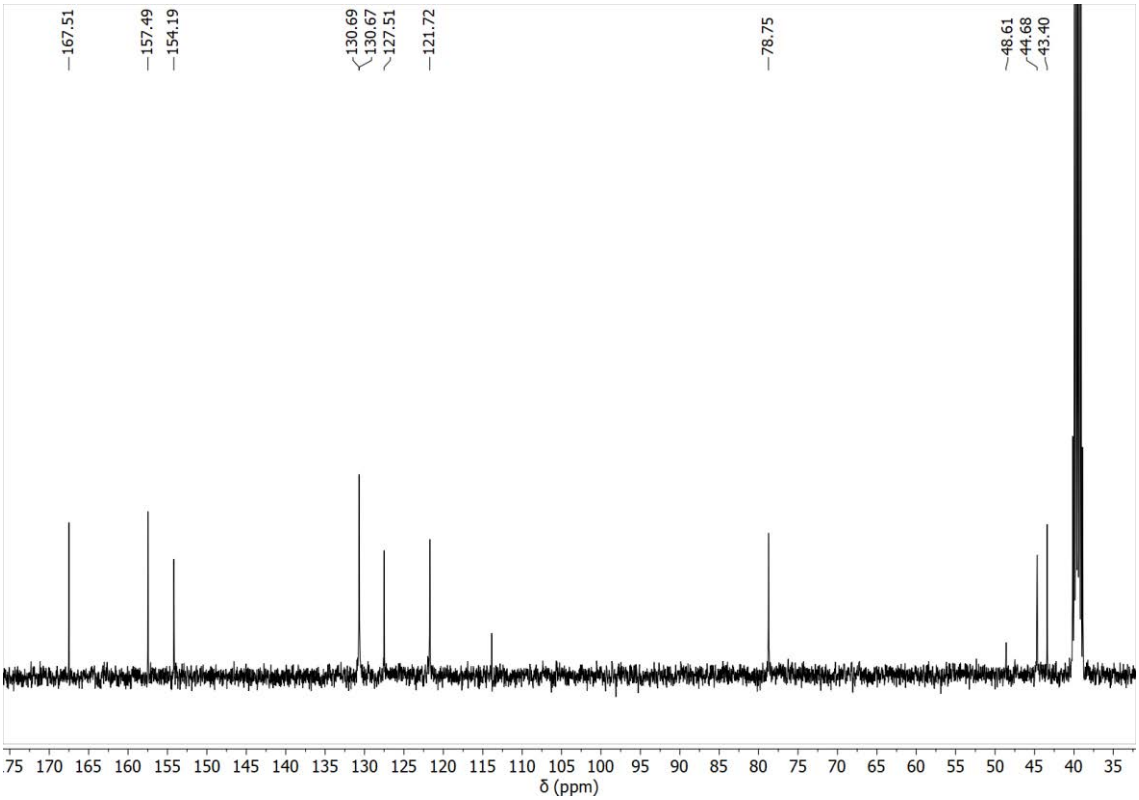


Figure 7. ^{13}C NMR spectrum of **B5** (101 MHz, DMSO-d_6, 298 K).

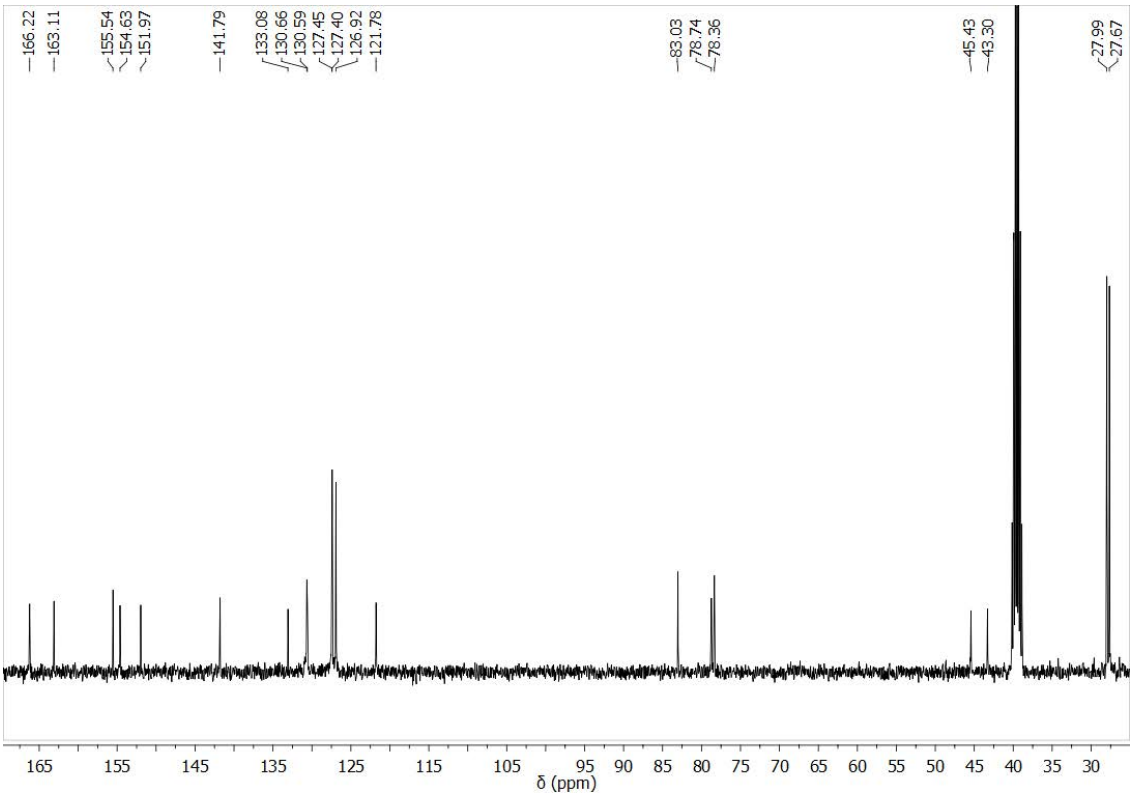


Figure 8. ^{13}C NMR spectrum of **B12** (101 MHz, DMSO-d_6, 298 K).

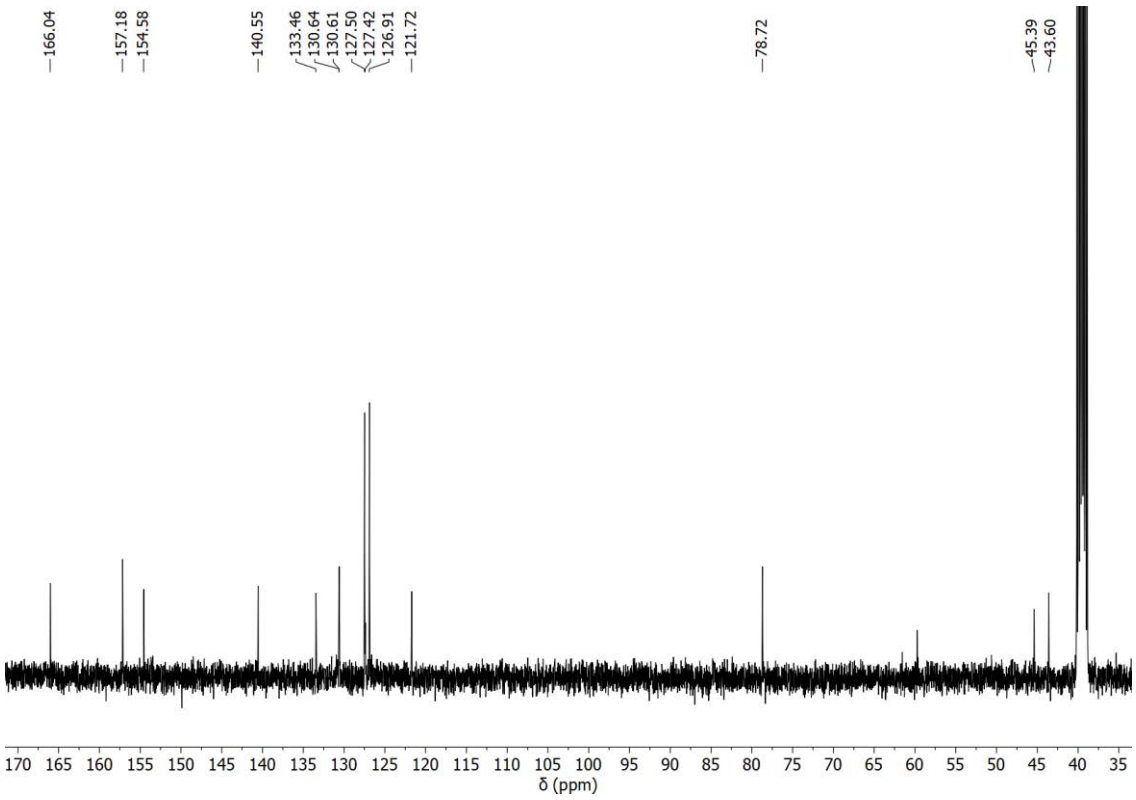


Figure 9. ^{13}C NMR spectrum of **B7** (101 MHz, DMSO-d_6, 298 K).

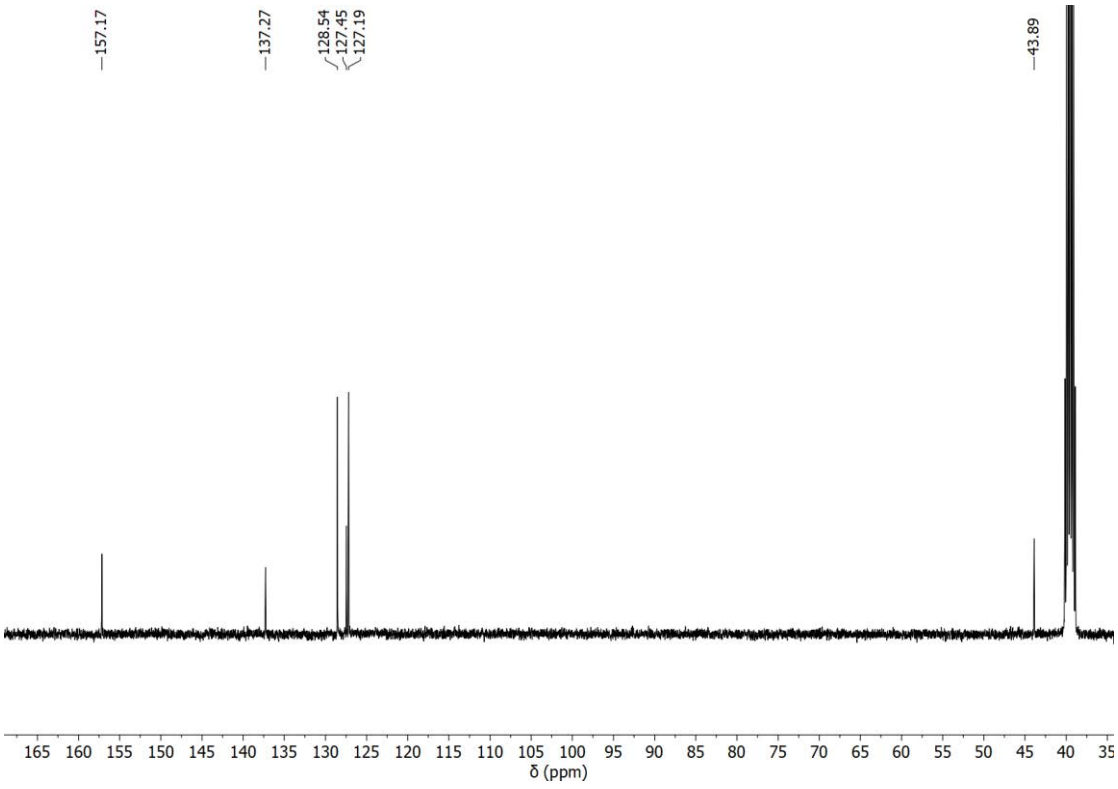


Figure 10. ^{13}C NMR spectrum of **B13** (101 MHz, DMSO-d_6, 298 K).

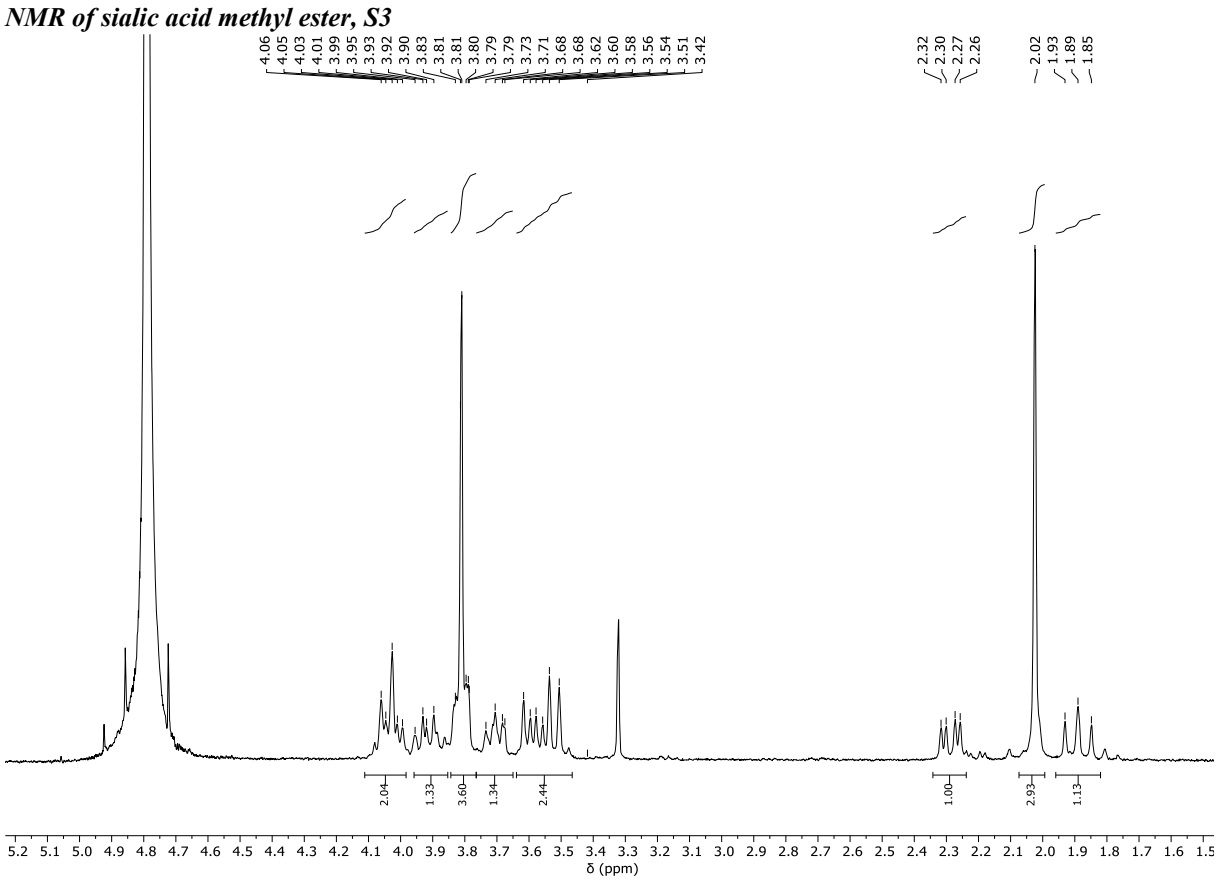


Figure 11. ^{1}H NMR spectrum of **S3** (300 MHz, D_2O, 298 K).

NMR spectra of peptidomimetic scaffold related molecules

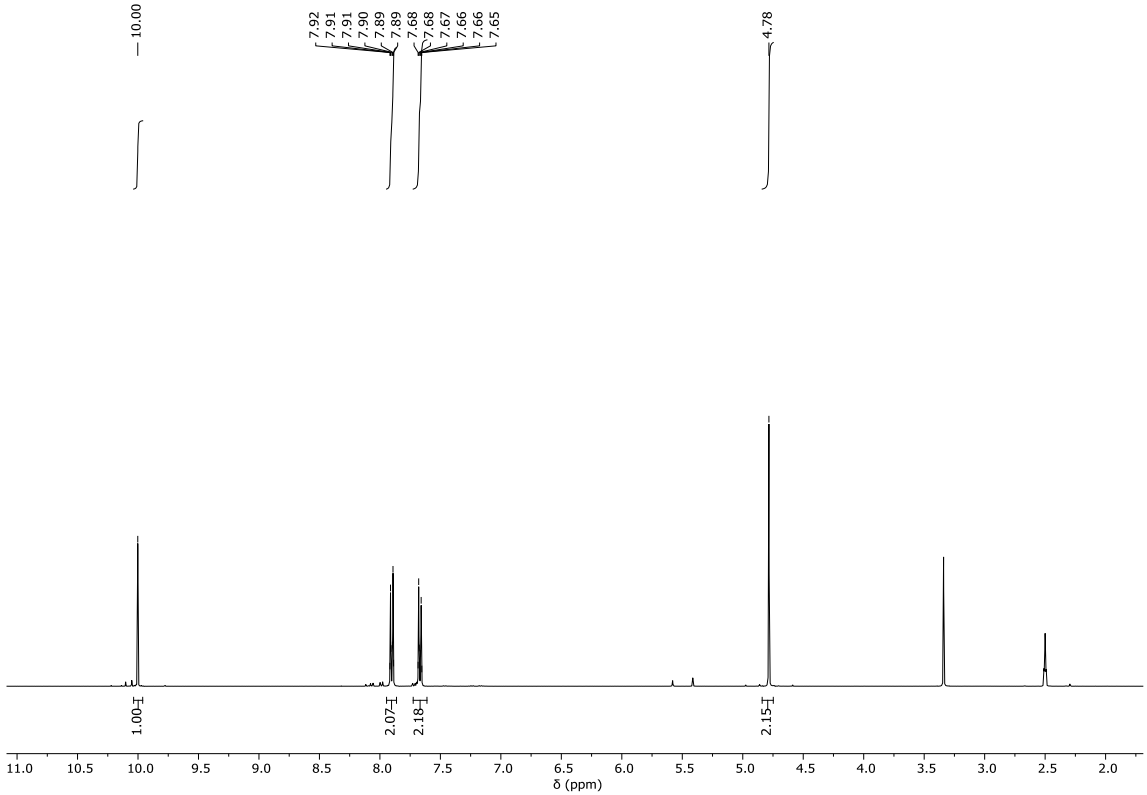


Figure 12. ^{1}H NMR spectrum of **S12** (400 MHz, DMSO-d_6, 298 K).

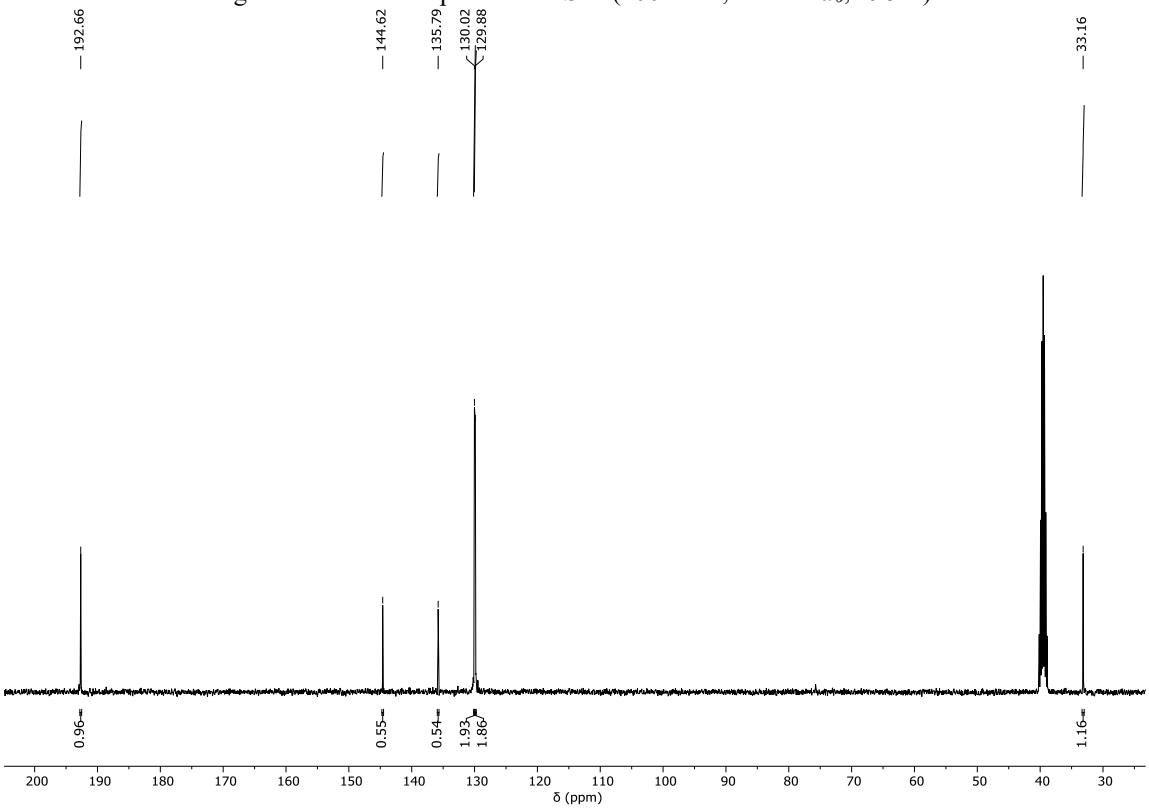


Figure 13. ^{13}C NMR spectrum of **S12** (101 MHz, DMSO-d_6, 298 K).

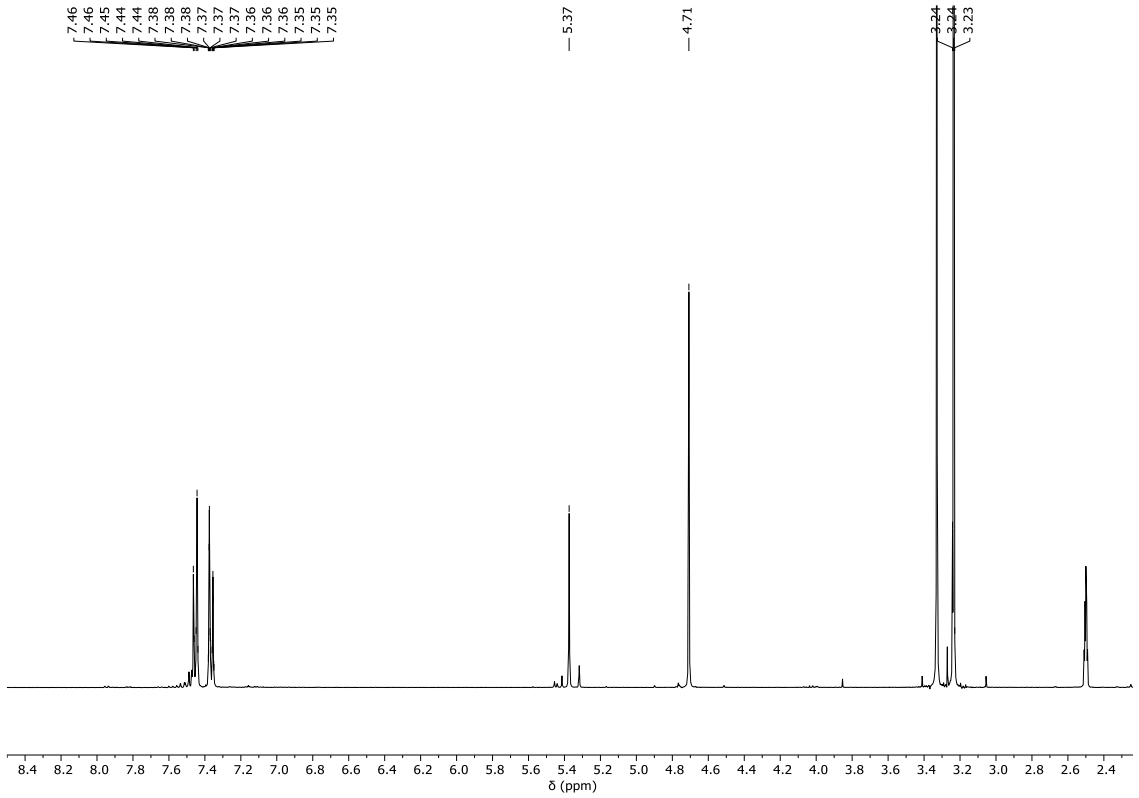


Figure 14. ^{1}H NMR spectrum of **S13** (400 MHz, DMSO-d_6, 298 K).

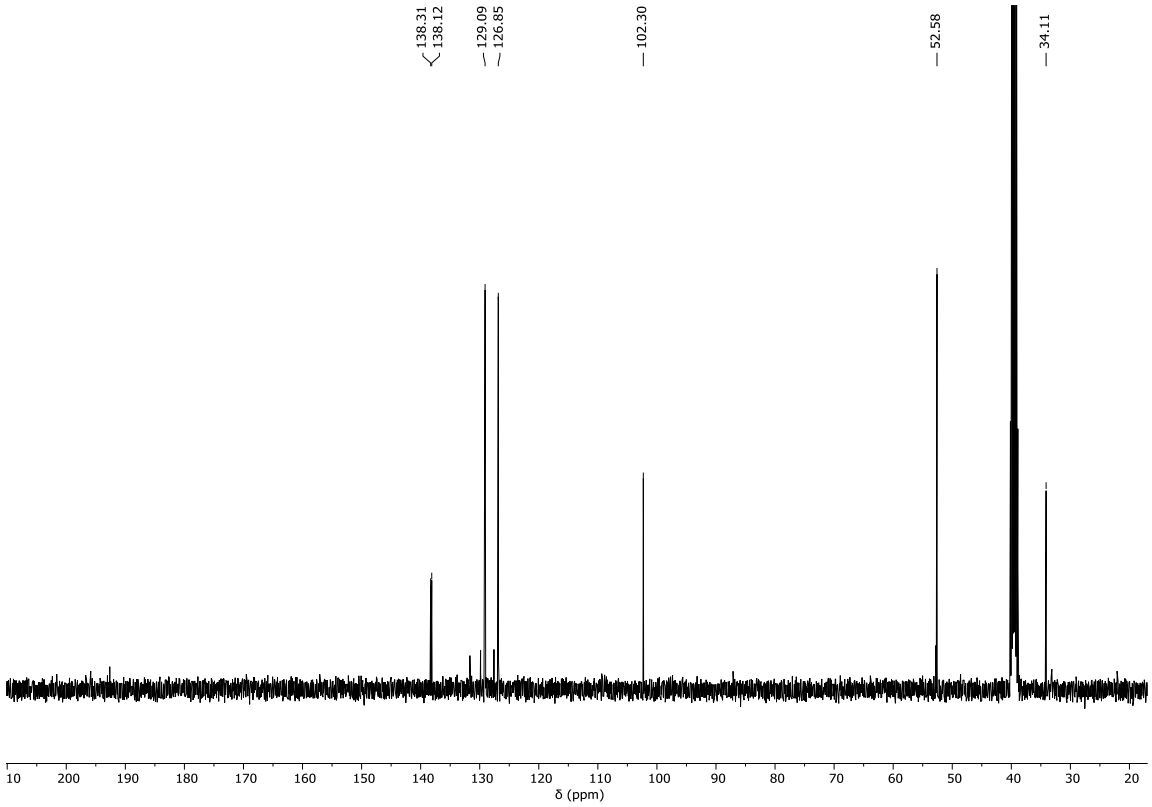


Figure 15. ^{13}C NMR spectrum of **S13** (101 MHz, DMSO-d_6, 298 K).

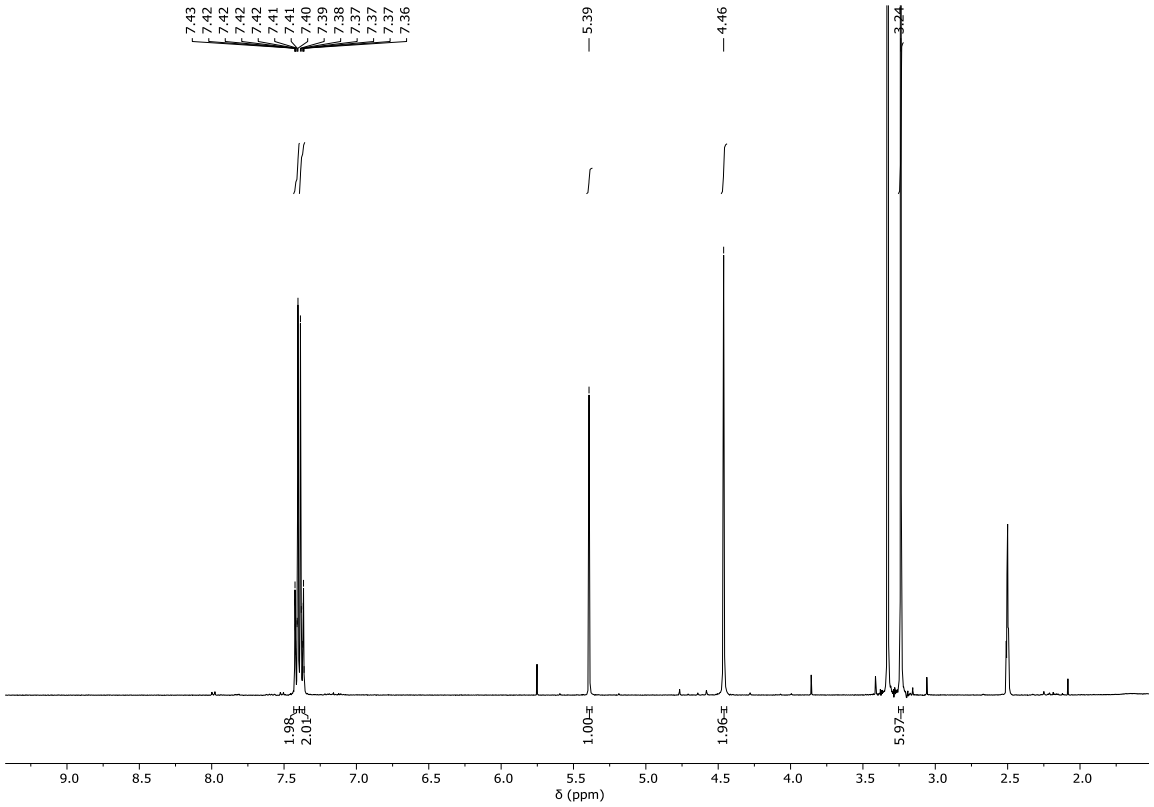


Figure 16. ^{1}H NMR spectrum of **S14** (400 MHz, DMSO-d_6, 298 K).

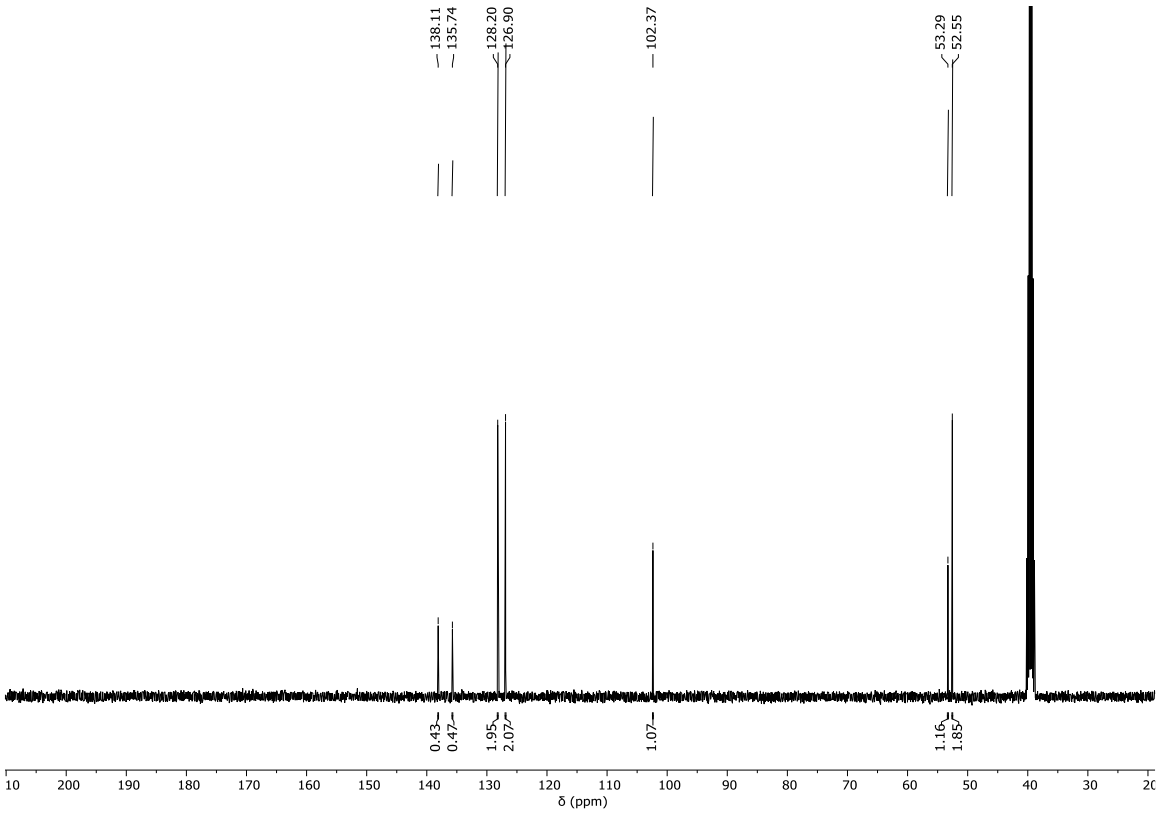


Figure 17. ^{13}C NMR spectrum of **S14** (101 MHz, DMSO-d_6, 298 K).

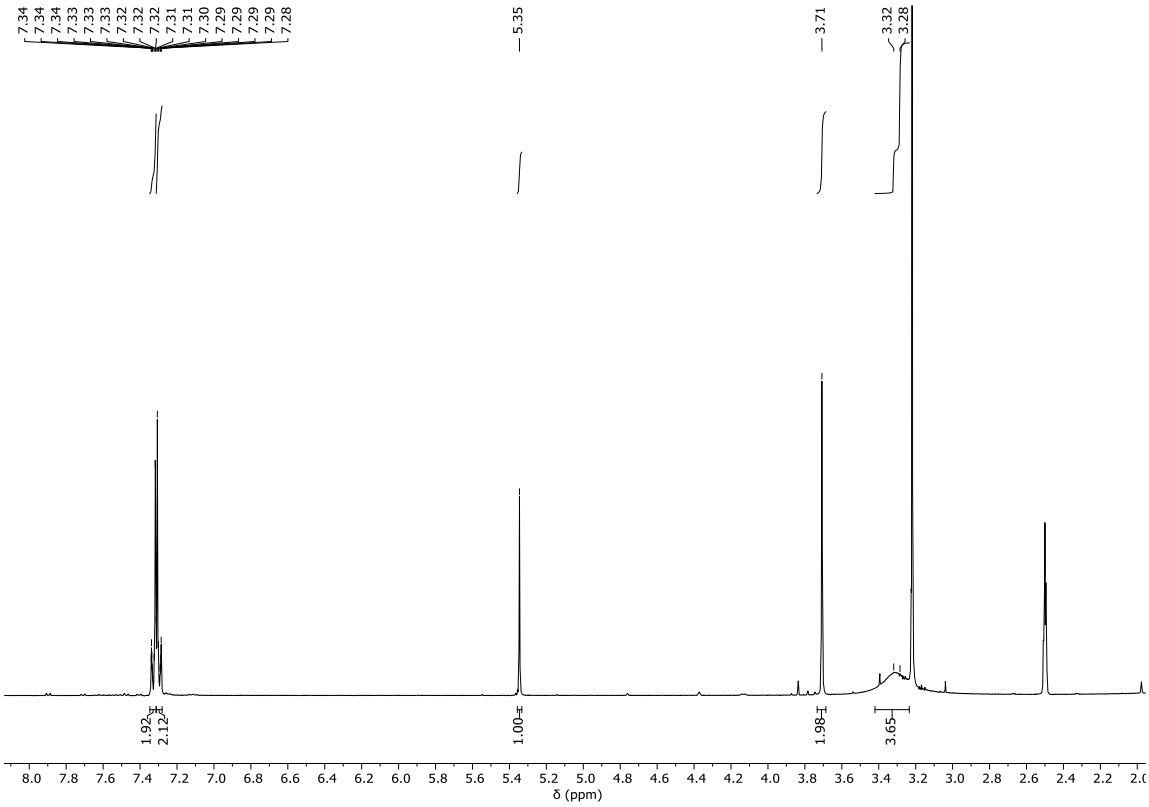


Figure 18. ^{1}H NMR spectrum of **S15** (400 MHz, DMSO-d_6, 298 K).

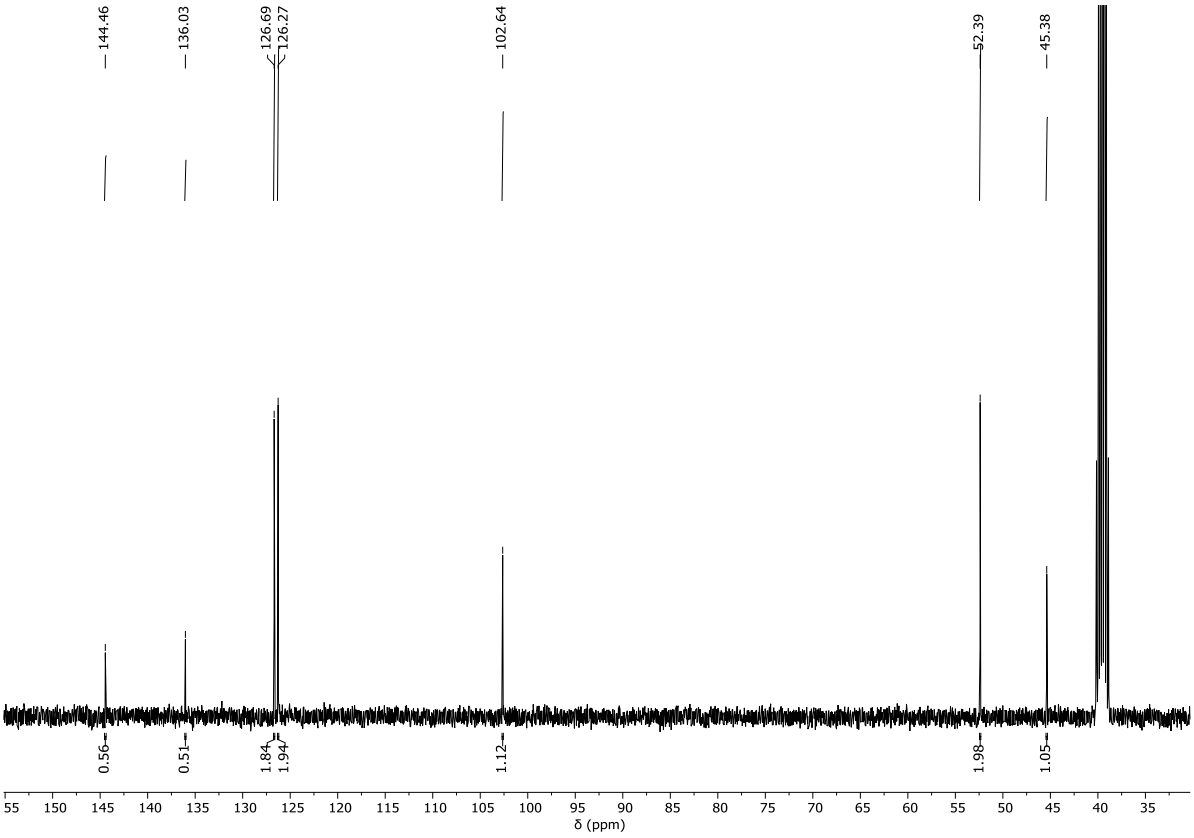


Figure 19. ^{13}C NMR spectrum of **S15** (101 MHz, DMSO-d_6, 298 K).

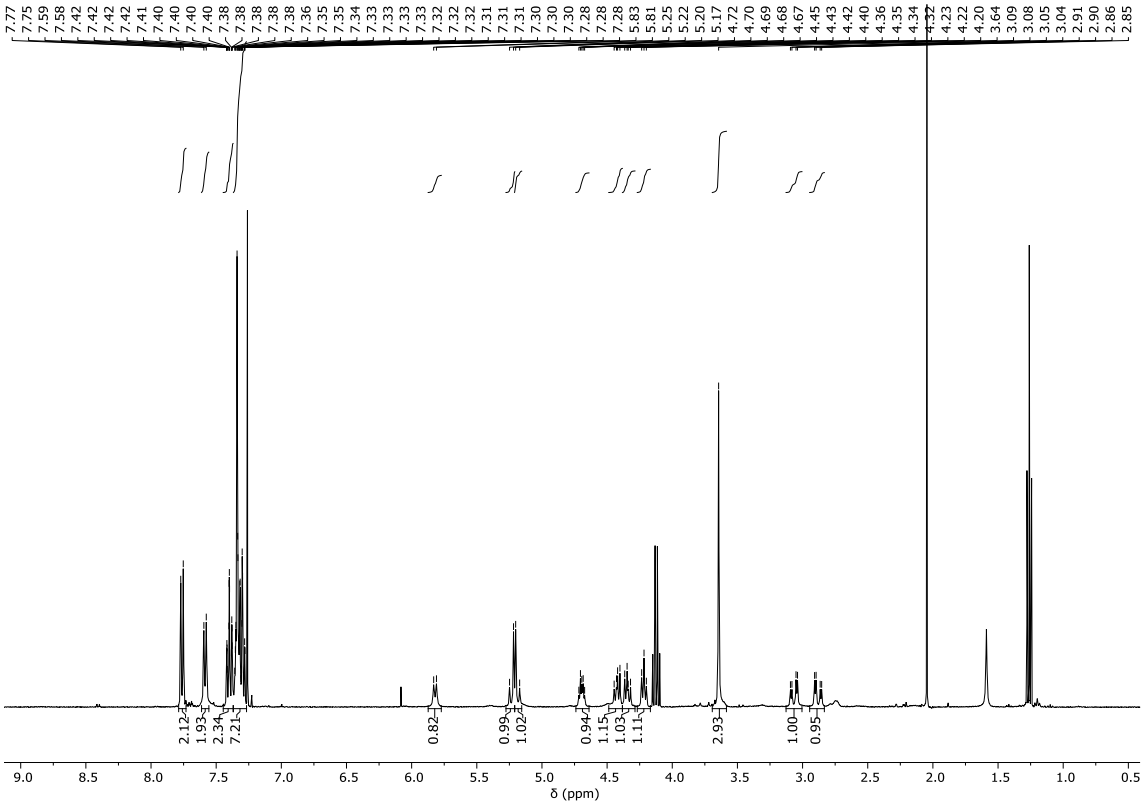


Figure 20. ^{1}H NMR spectrum of **S2** (400 MHz, $CDCl_3$, 298 K).

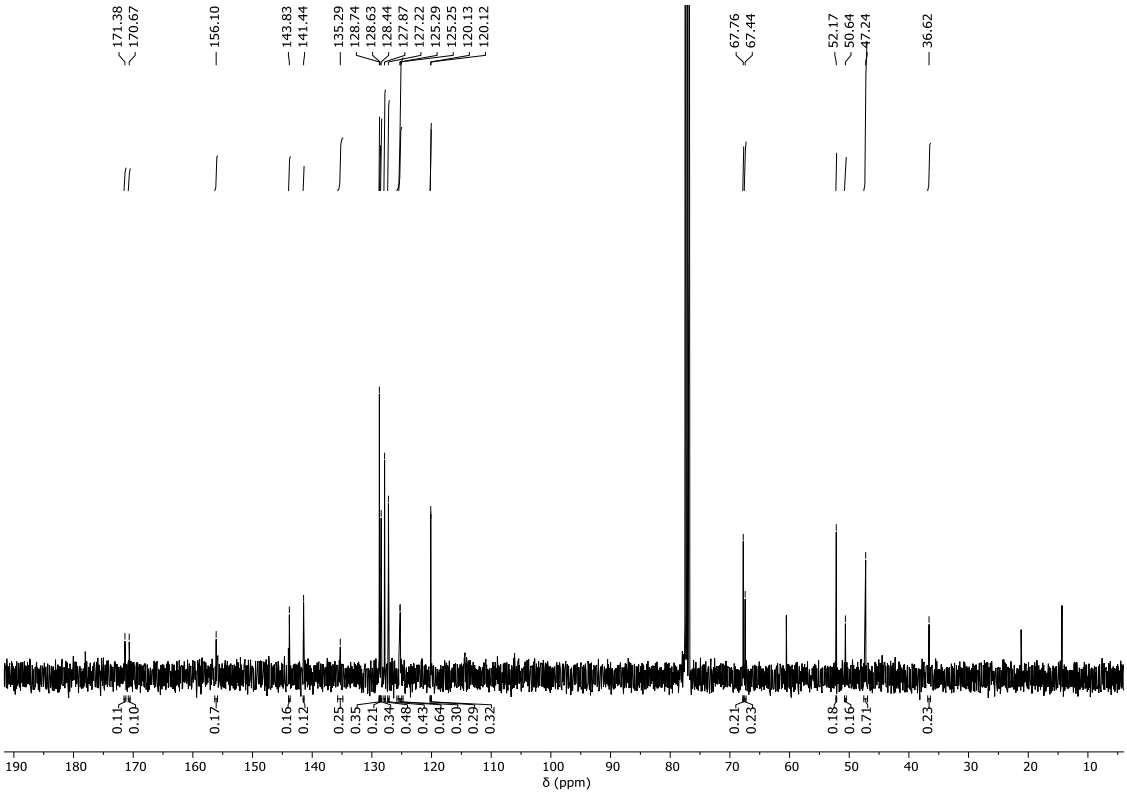


Figure 21. ^{13}C NMR spectrum of **S2** (101 MHz, $CDCl_3$, 298 K).

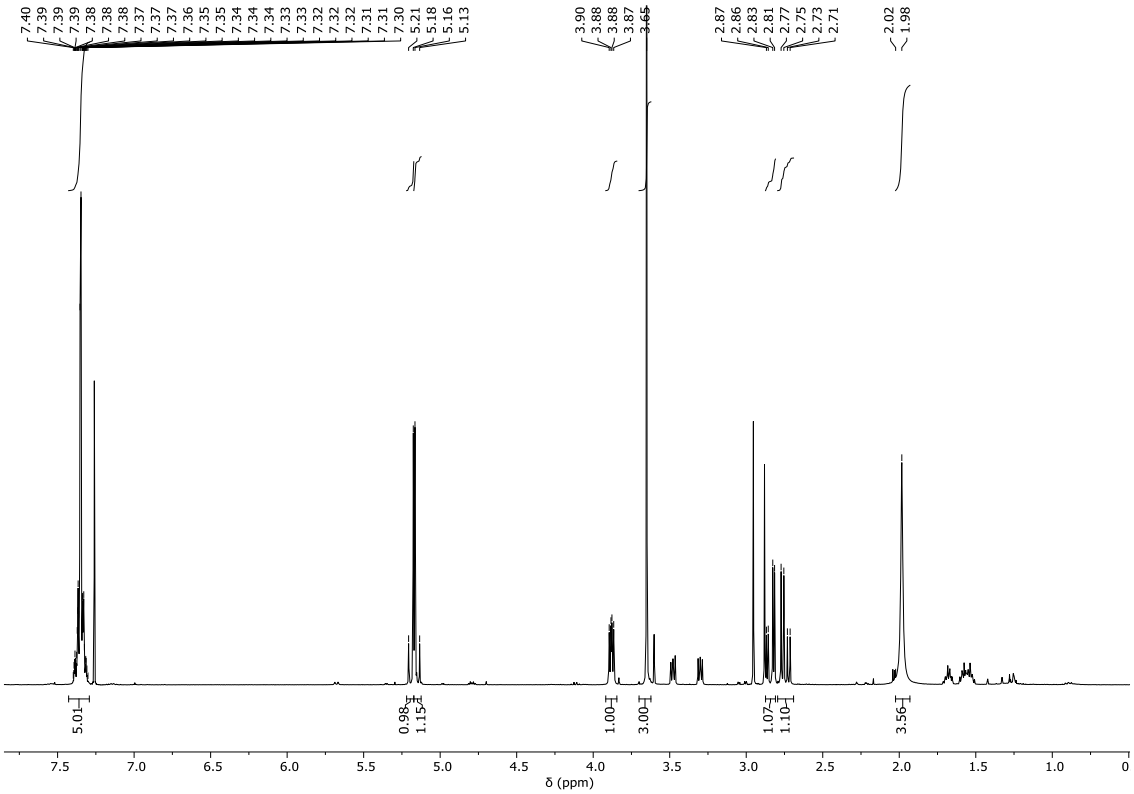


Figure 22. ^{1}H NMR spectrum of **S3** (400 MHz, $CDCl_3$, 298 K).

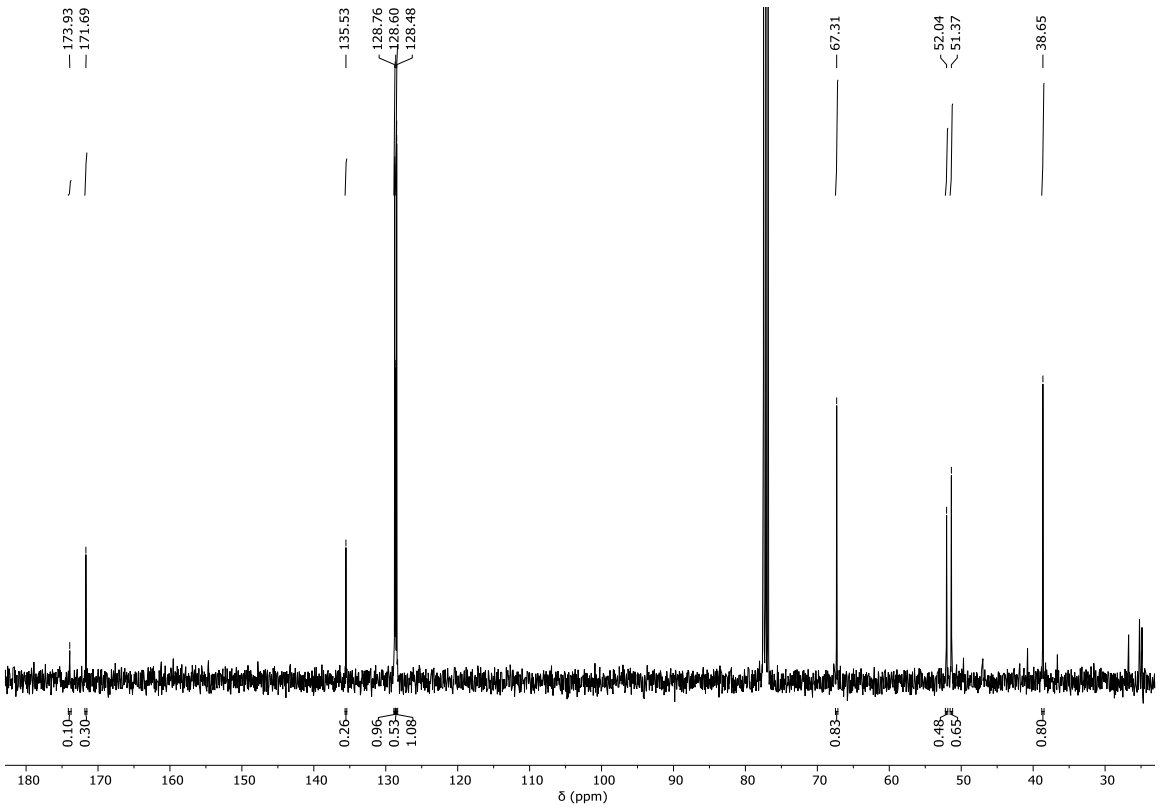


Figure 23. ^{13}C NMR spectrum of **S3** (101 MHz, $CDCl_3$, 298 K).

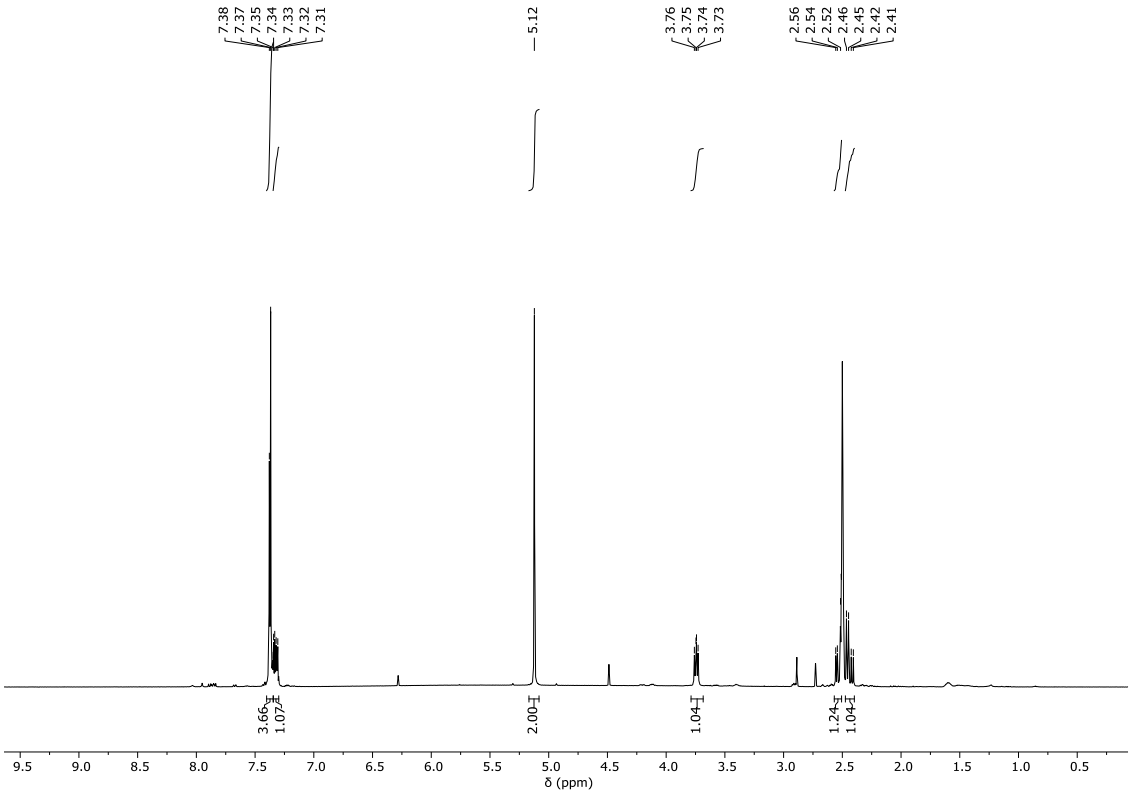


Figure 24. ^{1}H NMR spectrum of **S4** (400 MHz, $CDCl_3$, 298 K).

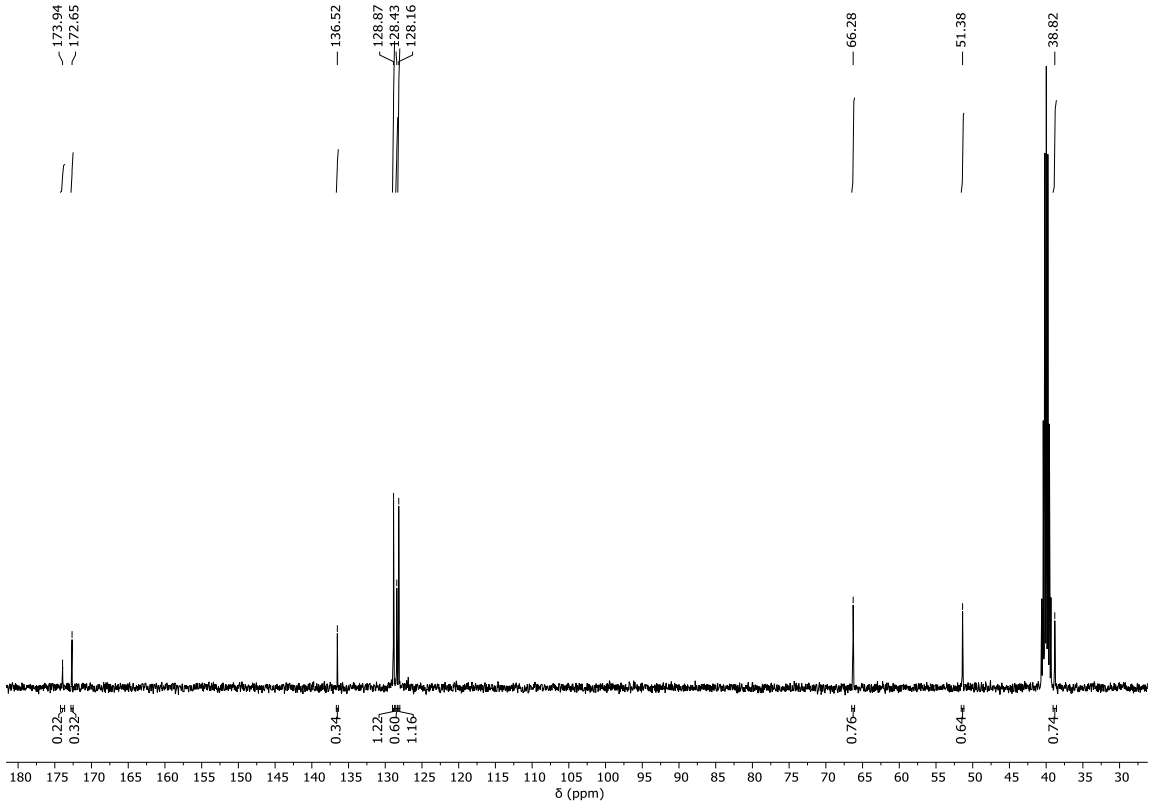


Figure 25. ^{13}C NMR spectrum of **S4** (101 MHz, $CDCl_3$, 298 K).

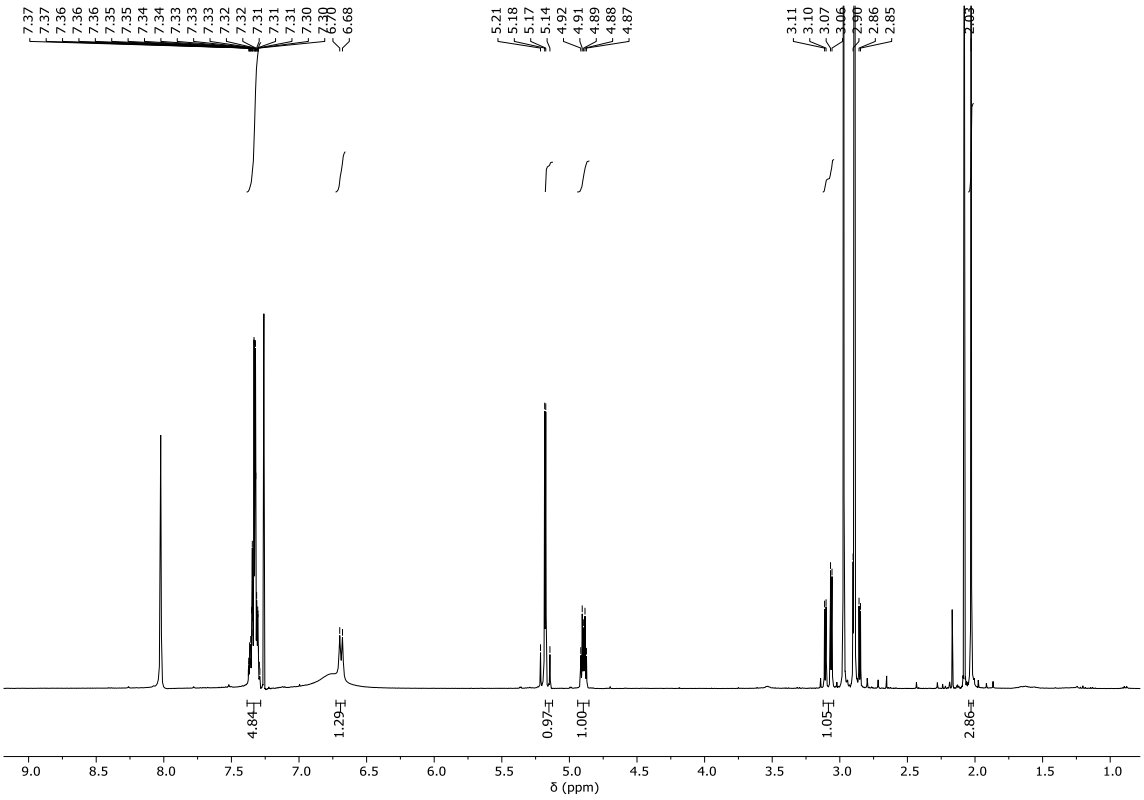


Figure 26. ^{1}H NMR spectrum of **S5** (400 MHz, $CDCl_3$, 298 K).

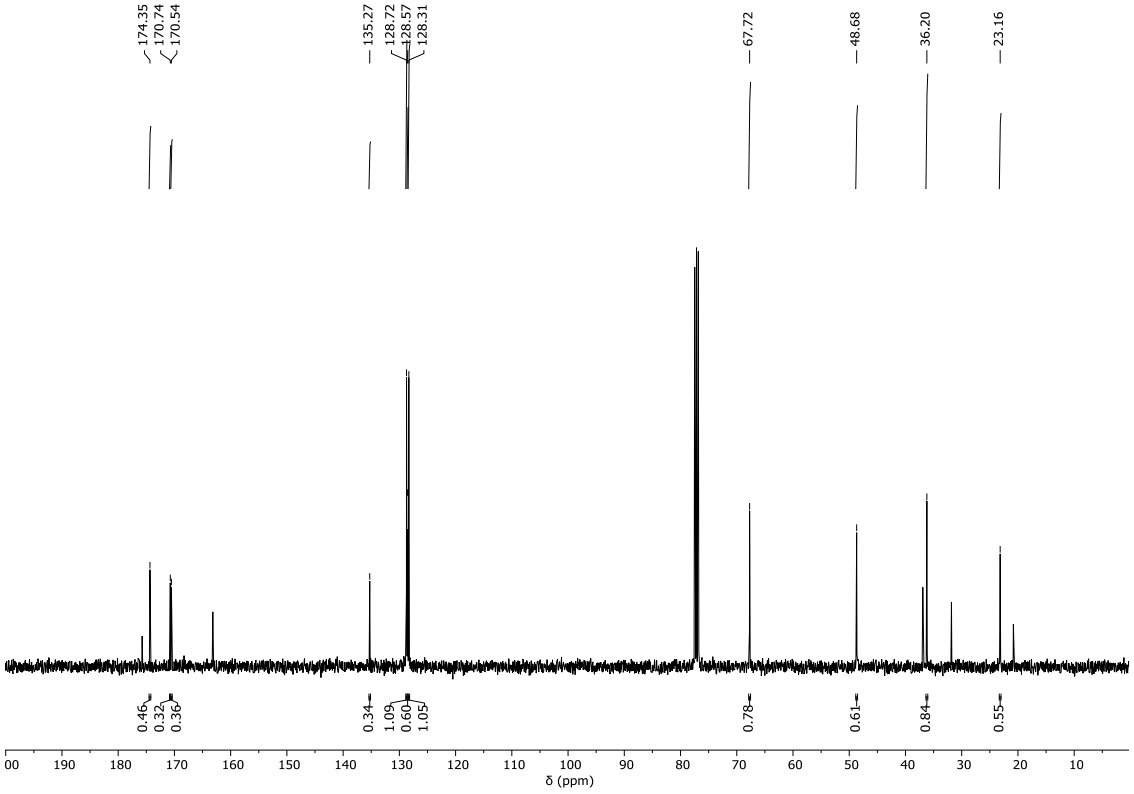


Figure 27. ^{13}C NMR spectrum of **S5** (101 MHz, $CDCl_3$, 298 K).

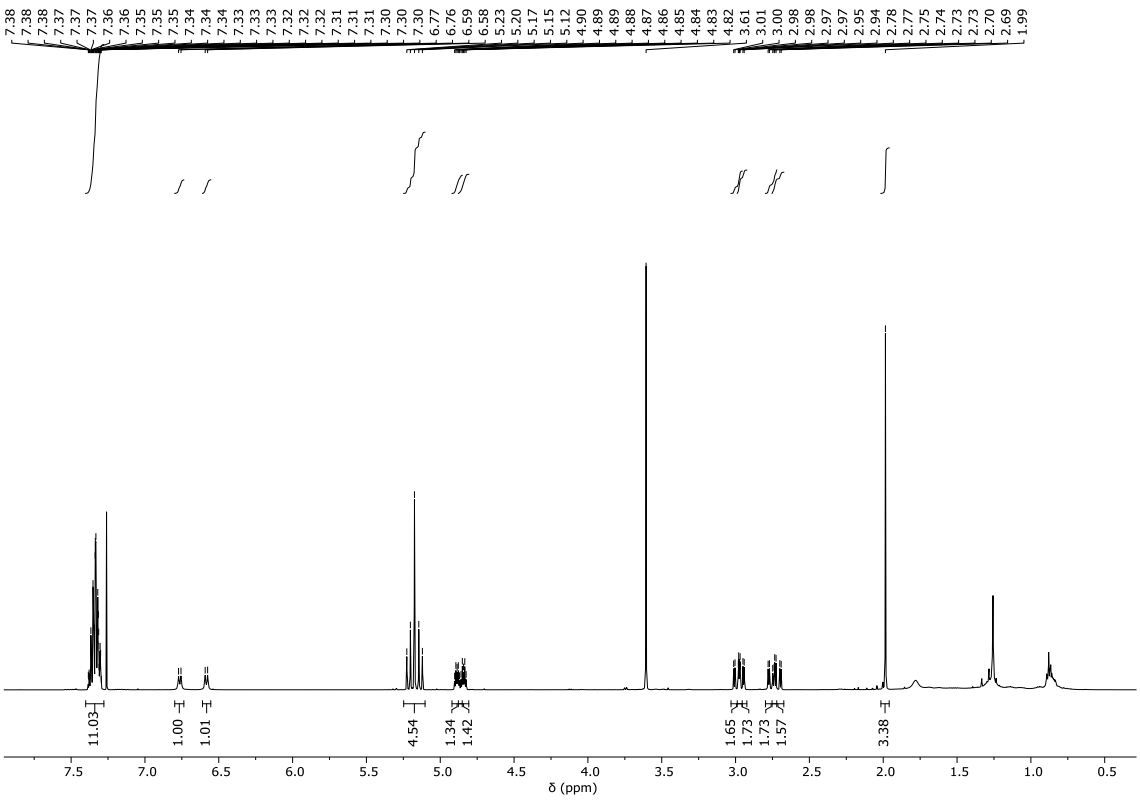


Figure 28. ^{1}H NMR spectrum of **S6** (500 MHz, $CDCl_3$, 298 K).

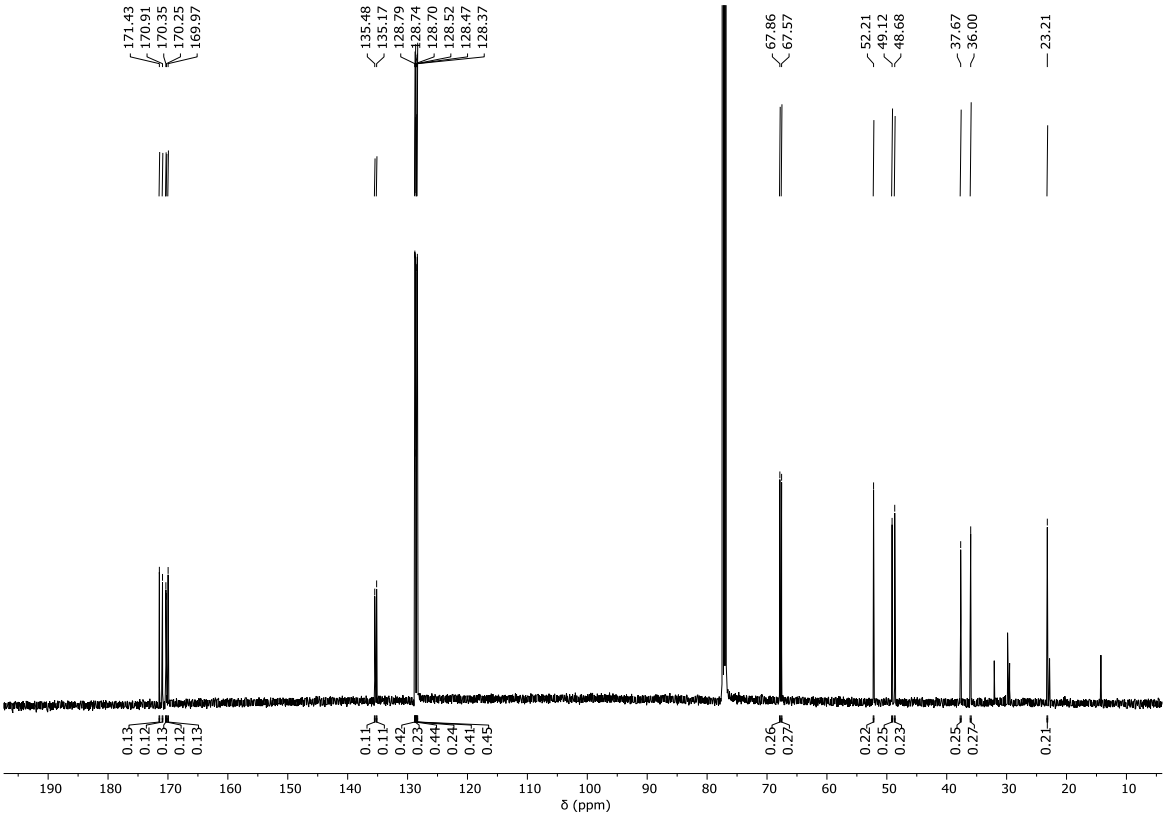


Figure 29. ^{13}C NMR spectrum of **S6** (126 MHz, $CDCl_3$, 298 K).

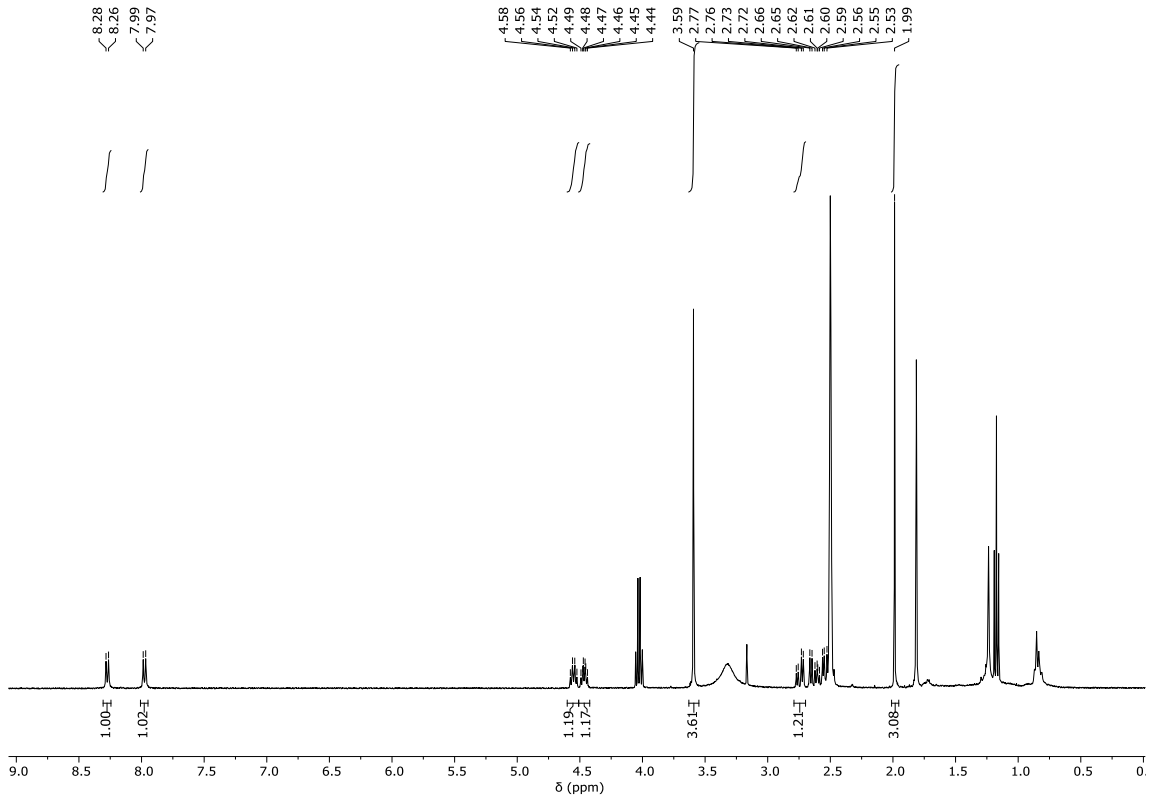


Figure 30. ^{1}H NMR spectrum of **S7** (400 MHz, DMSO-d_6, 298 K).

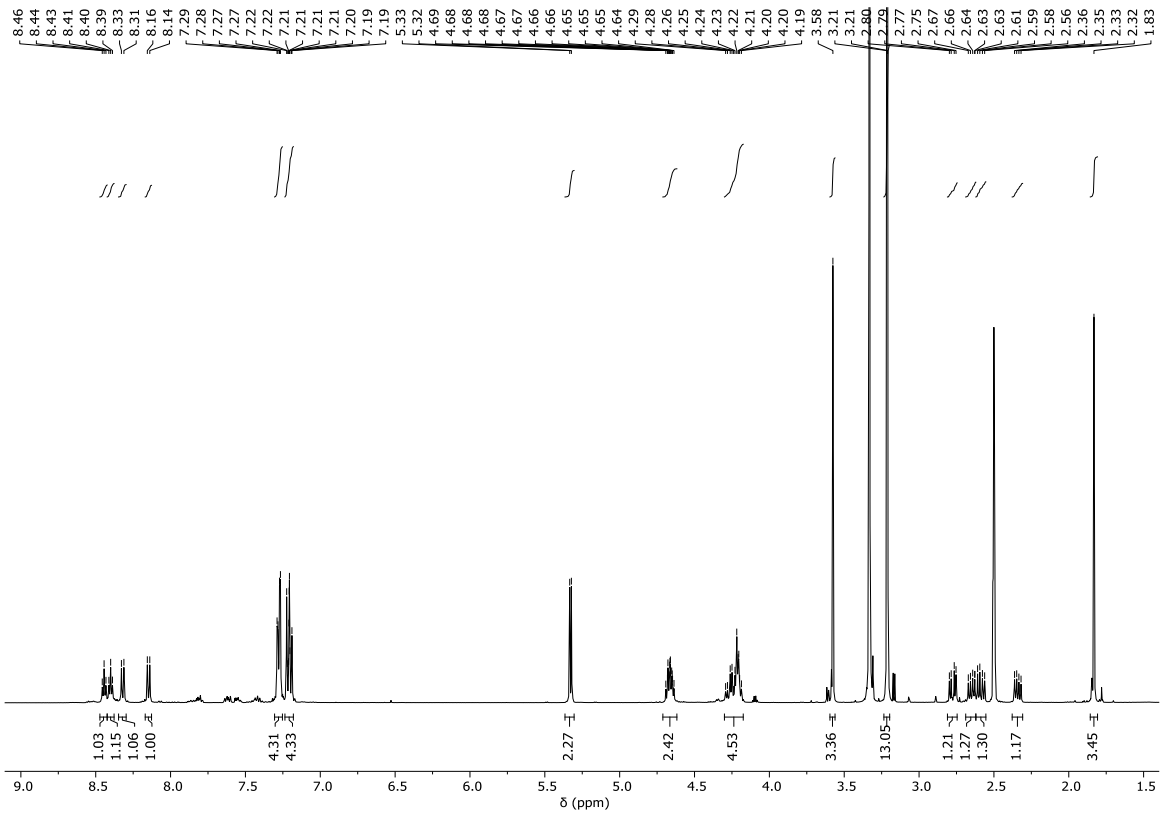


Figure 31. ^{1}H NMR spectrum of **S16** (500 MHz, DMSO-d_6, 298 K).

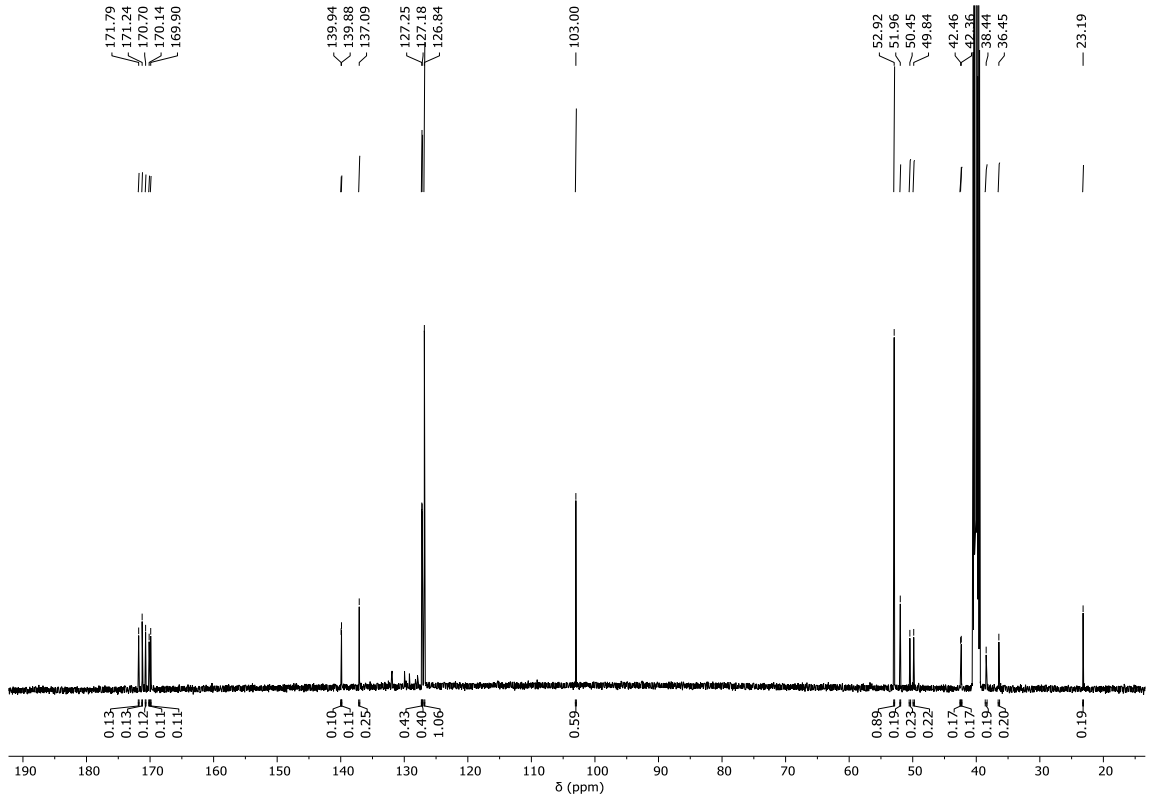


Figure 32. ^{13}C NMR spectrum of **S16** (126 MHz, DMSO-d_6, 298 K).

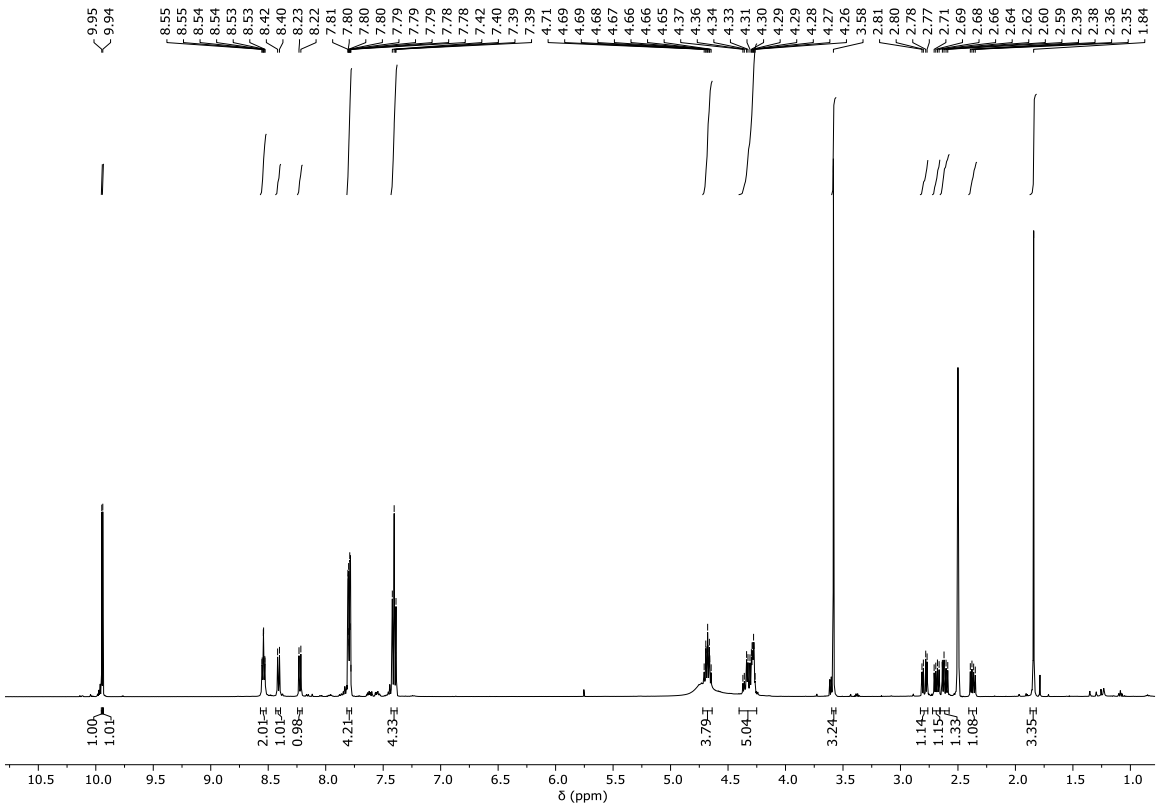


Figure 33. ^{1}H NMR spectrum of **S17** (500 MHz, DMSO-d_6, 298 K).

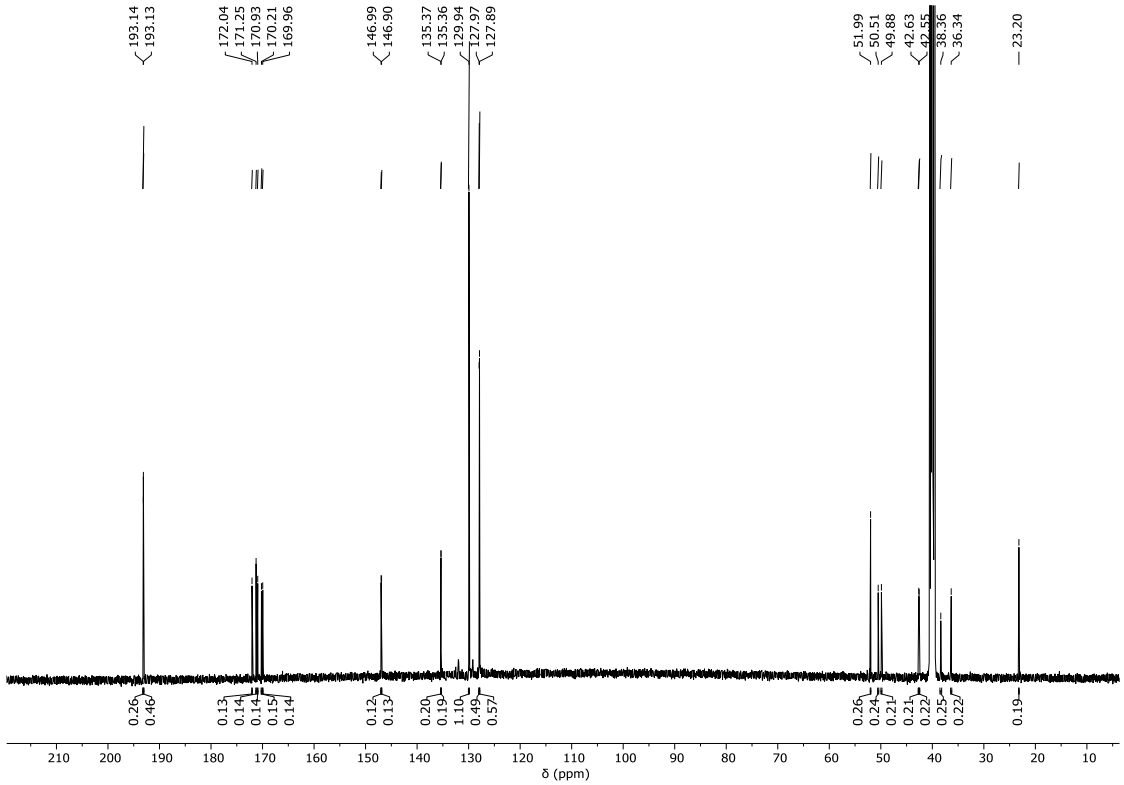


Figure 34. ^{13}C NMR spectrum of **S17** (126 MHz, DMSO-d_6, 298 K).

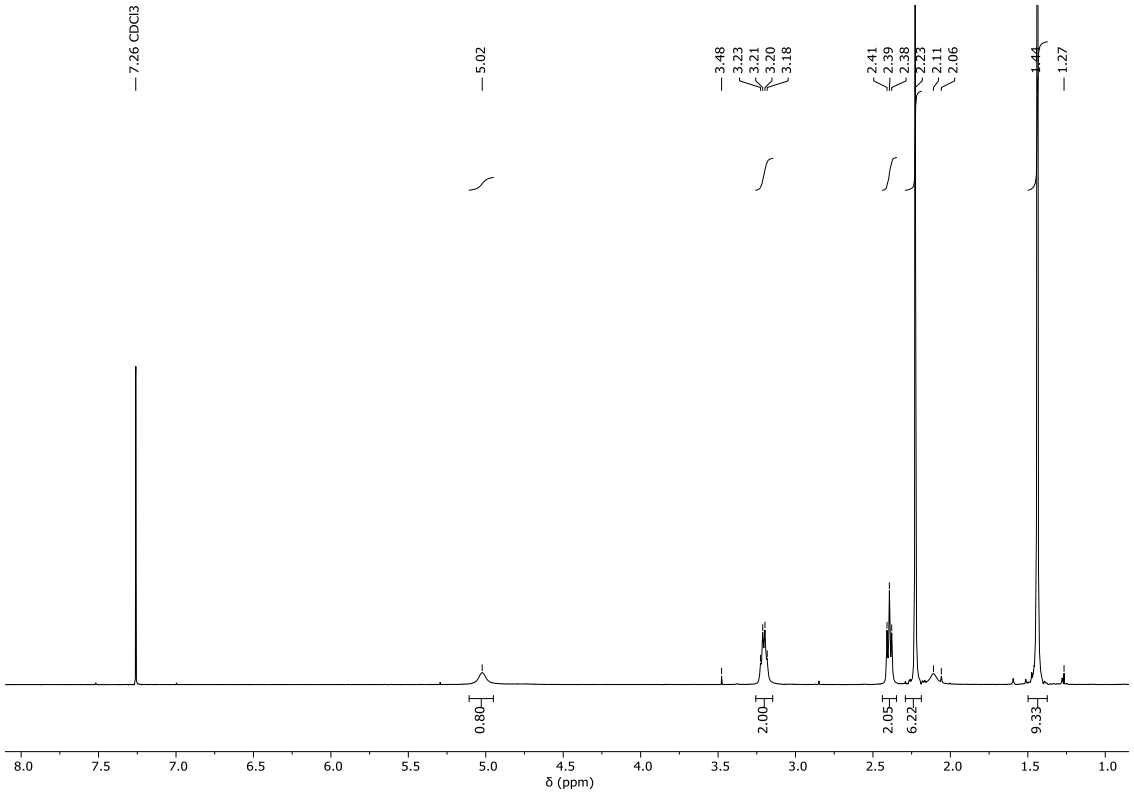


Figure 35. ^{1}H NMR spectrum of tert-butyl 2-(dimethylamino)ethylcarbamate (400 MHz, $CDCl_3$, 298 K).

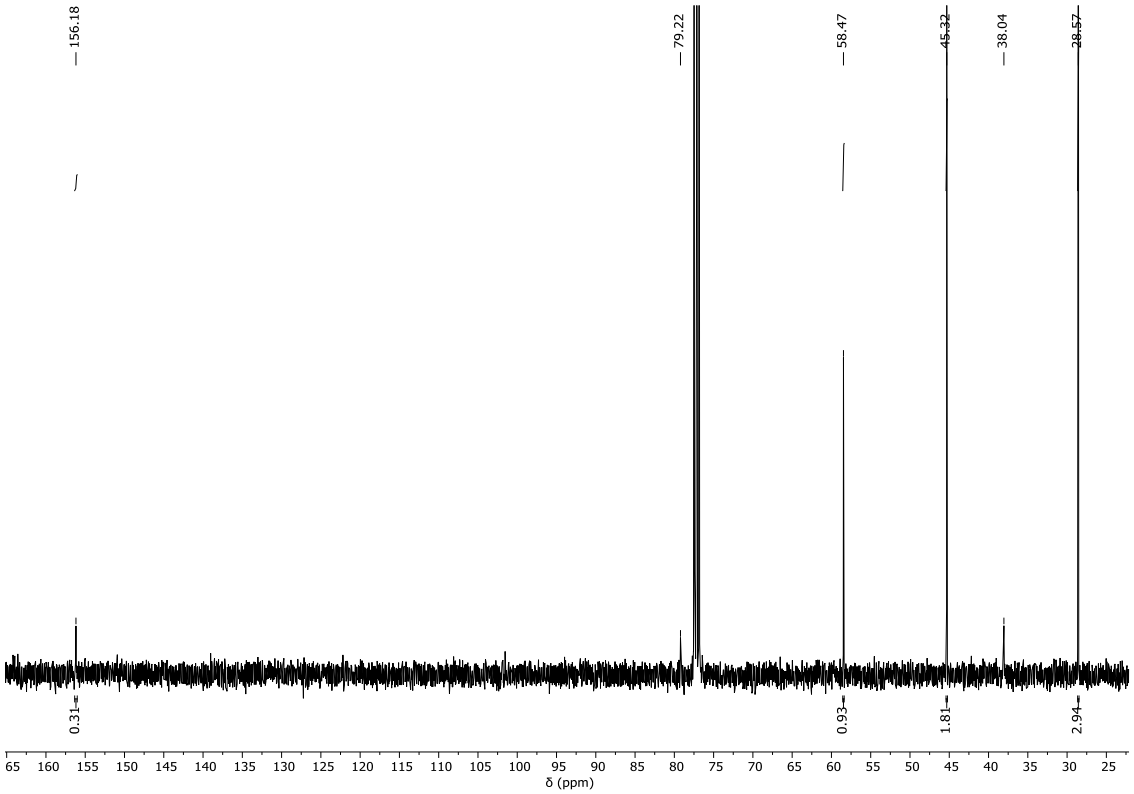


Figure 36. ^{13}C NMR spectrum of tert-butyl 2-(dimethylamino)ethylcarbamate (101 MHz, $CDCl_3$, 298 K).

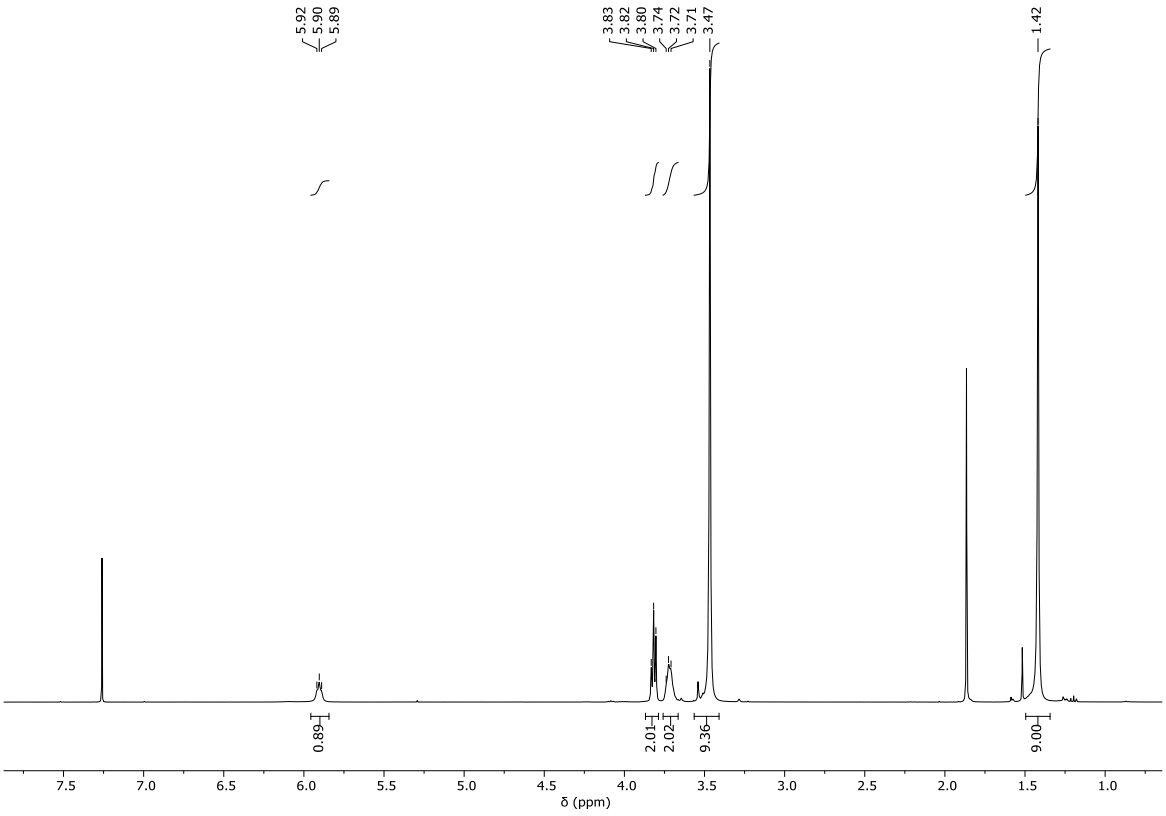


Figure 37. ^{1}H NMR spectrum of tert-butyl 2-(trimethylamino)ethylcarbamate (400 MHz, $CDCl_3$, 298 K).

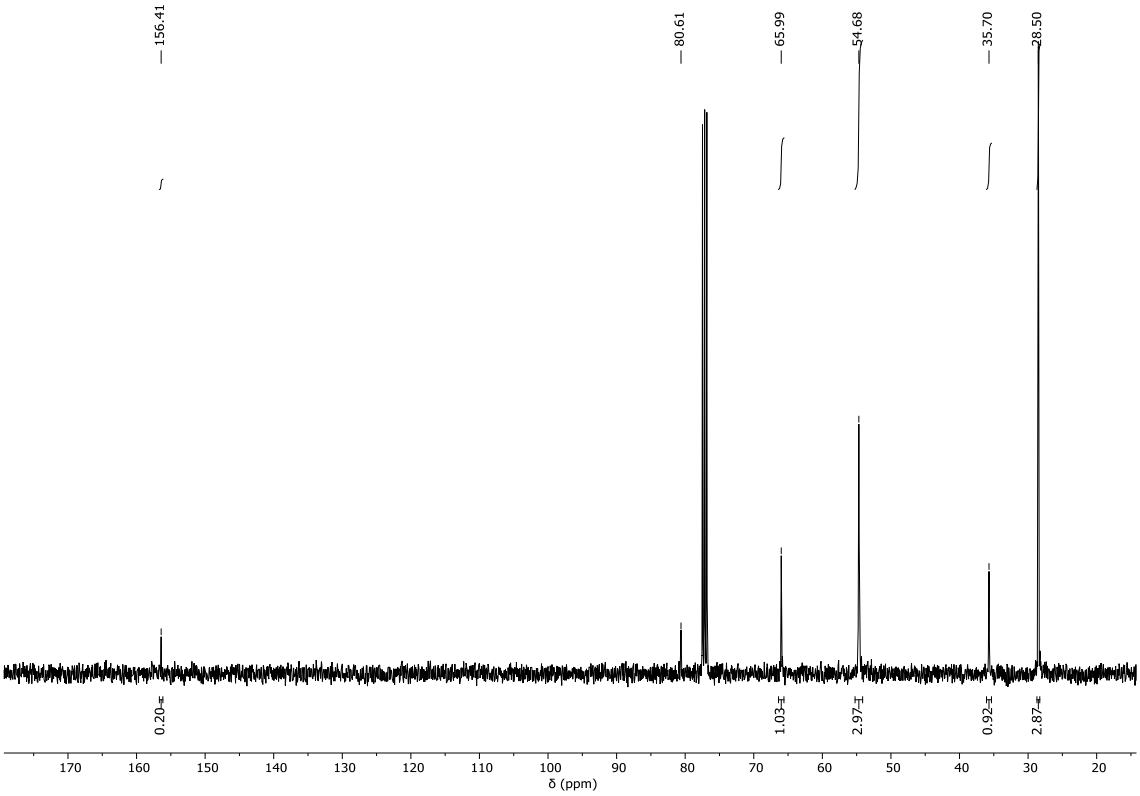


Figure 38. ^{13}C NMR spectrum of tert-butyl 2-(trimethylamino)ethylcarbamate (101 MHz, $CDCl_3$, 298 K).

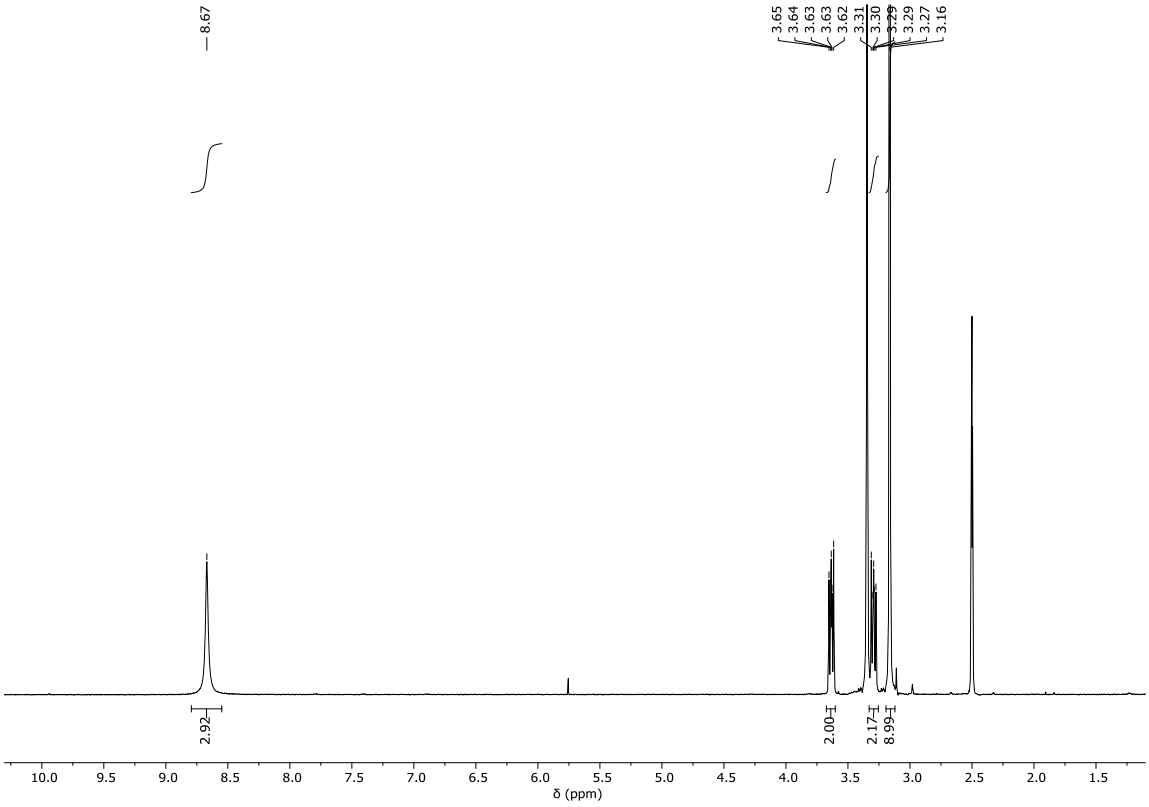


Figure 39. ^{1}H NMR spectrum of **AL10** (400 MHz, DMSO-d_6, 298 K).

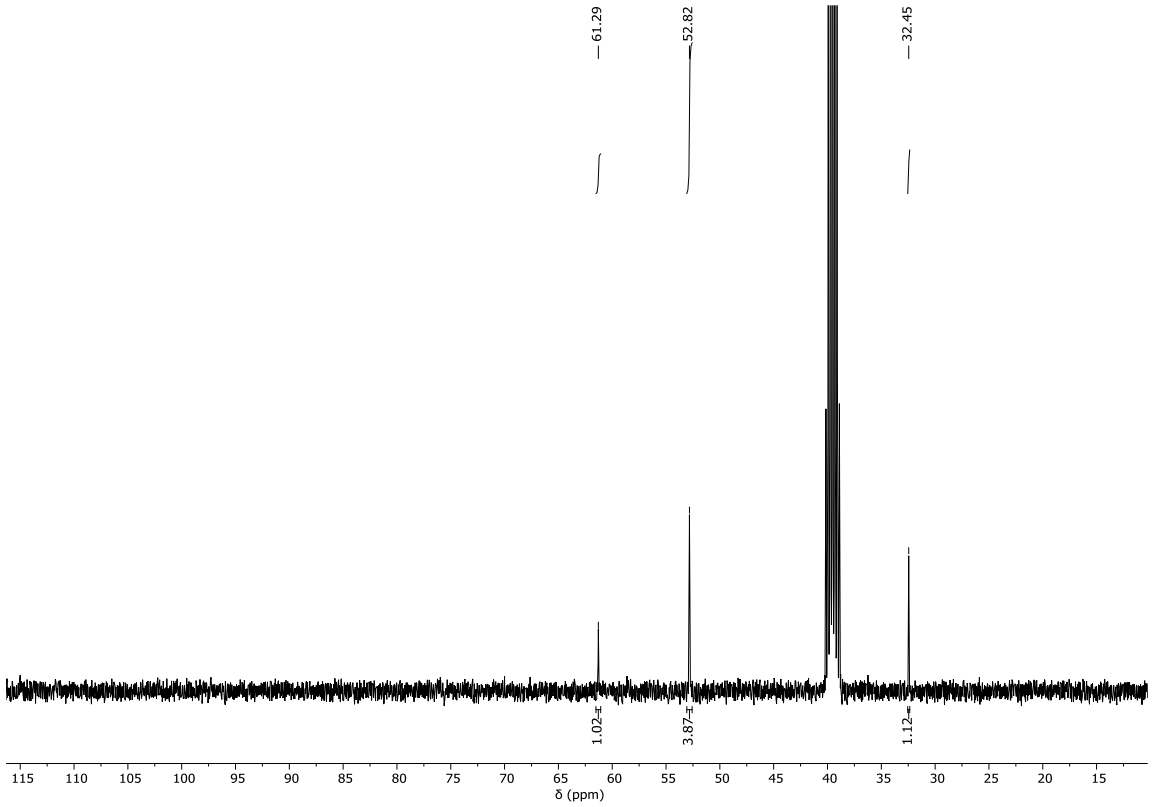


Figure 40. ^{13}C NMR spectrum of **AL10** (101 MHz, DMSO-d_6, 298 K).

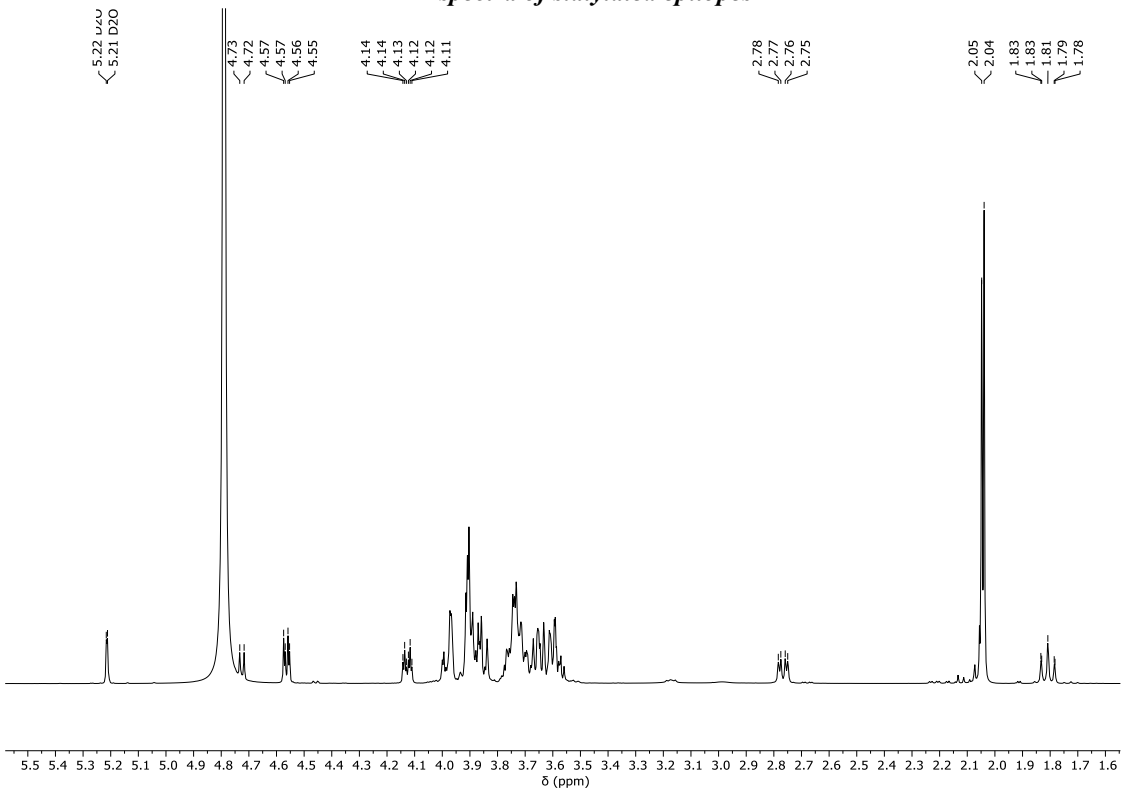


Figure 41. ^{1}H NMR spectrum of SAα2-3LacNAc (500 MHz, D_2O, 298 K).

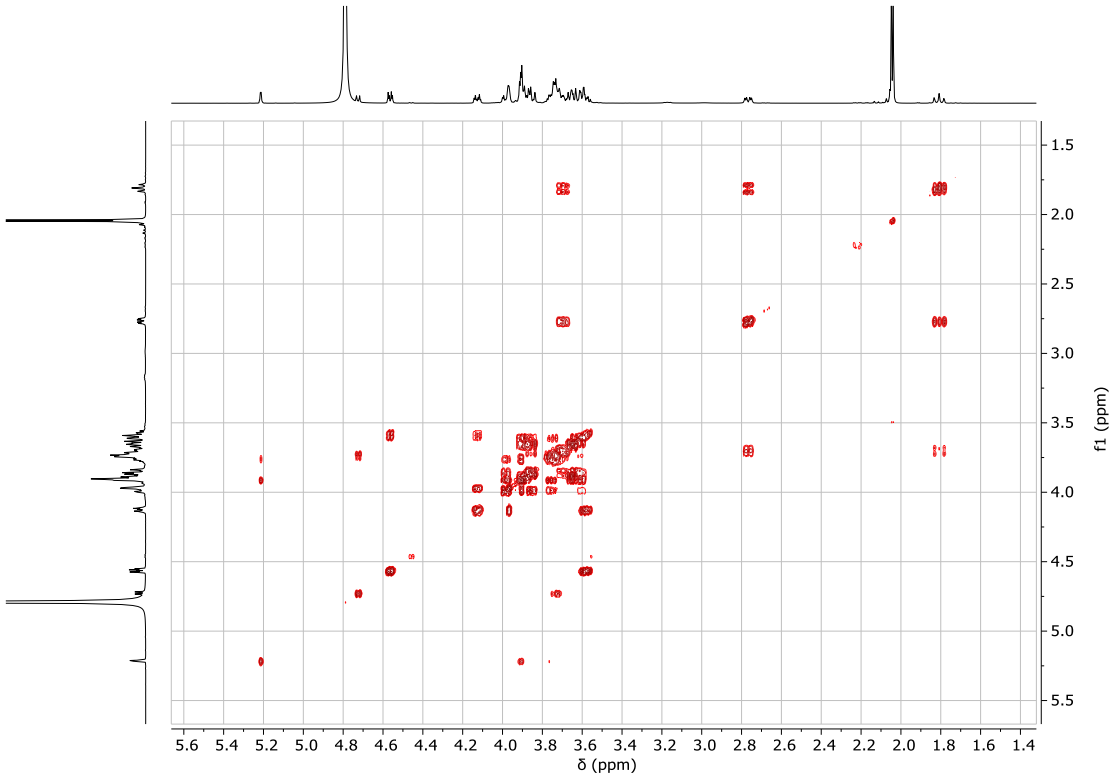


Figure 42. ^{1}H-^{1}H COSY NMR spectrum of SAα2-3LacNAc (500 MHz, D_2O, 298 K).

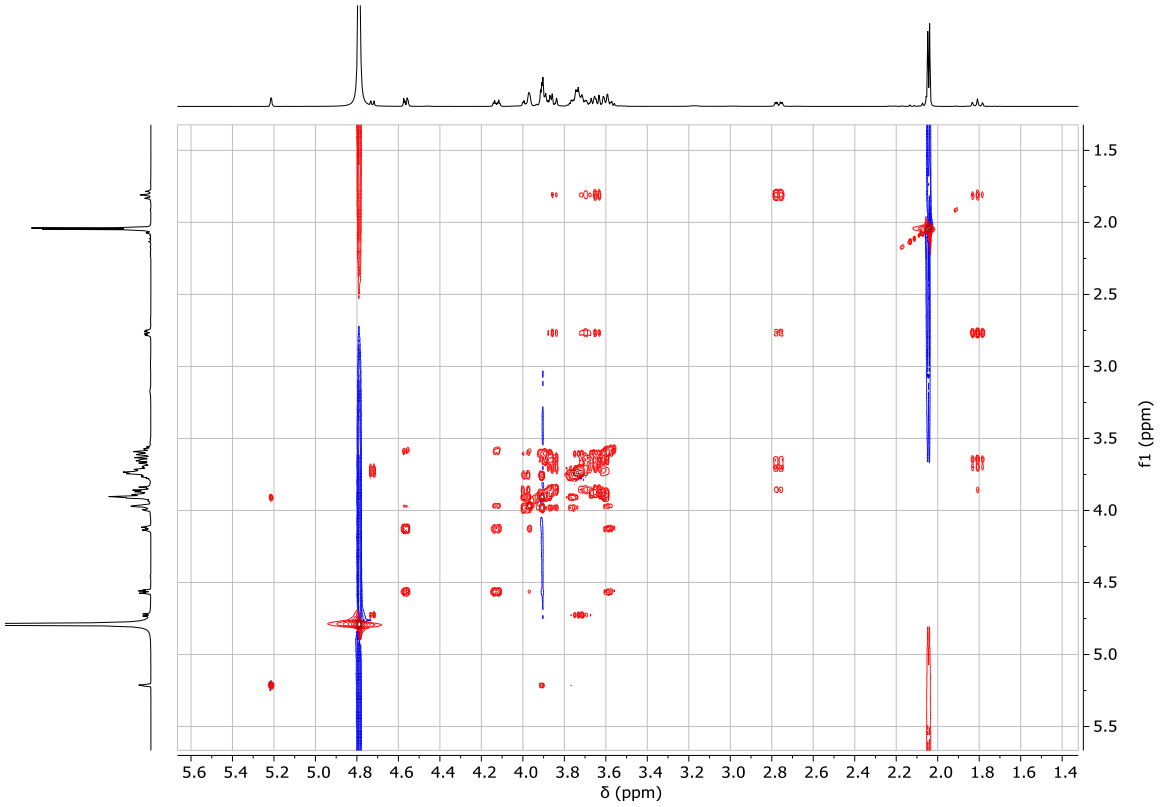


Figure 43. ^{1}H-^{1}H TOCSY NMR spectrum of SAα2-3LacNAc (500 MHz, D_2O, 298 K).

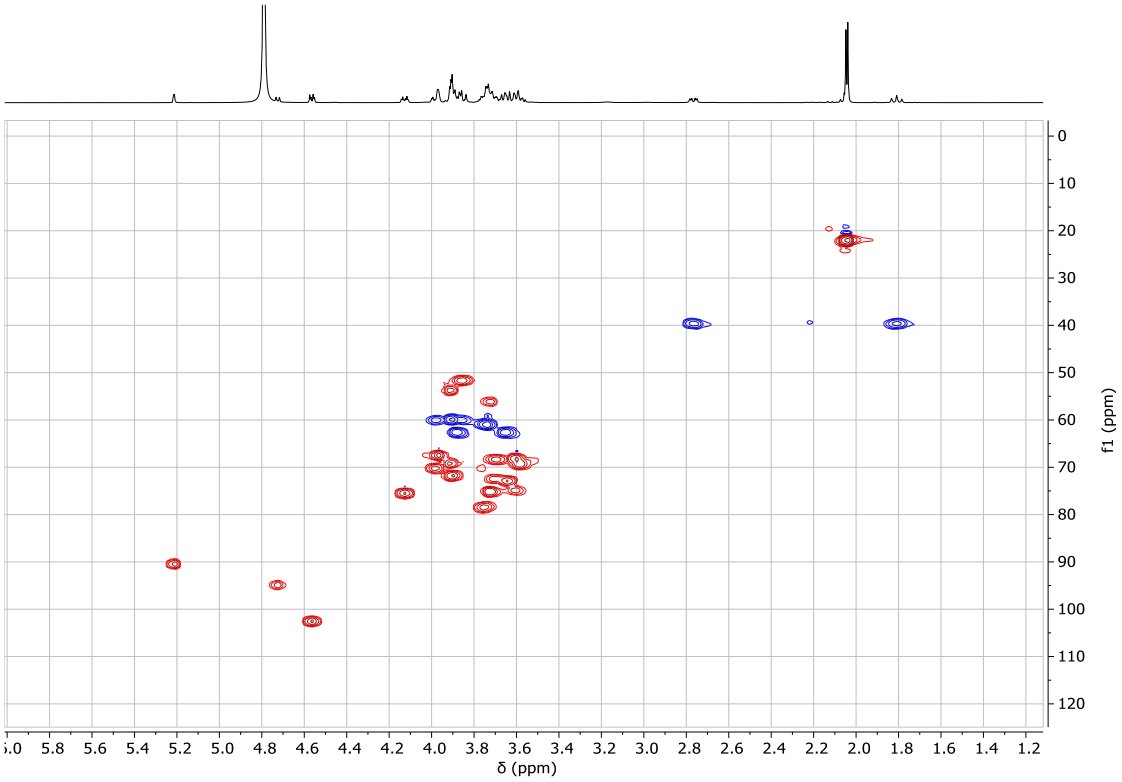


Figure 44. ^{1}H-^{13}C HSQC NMR spectrum of SAα2-3LacNAc (500 MHz, D_2O, 298 K).

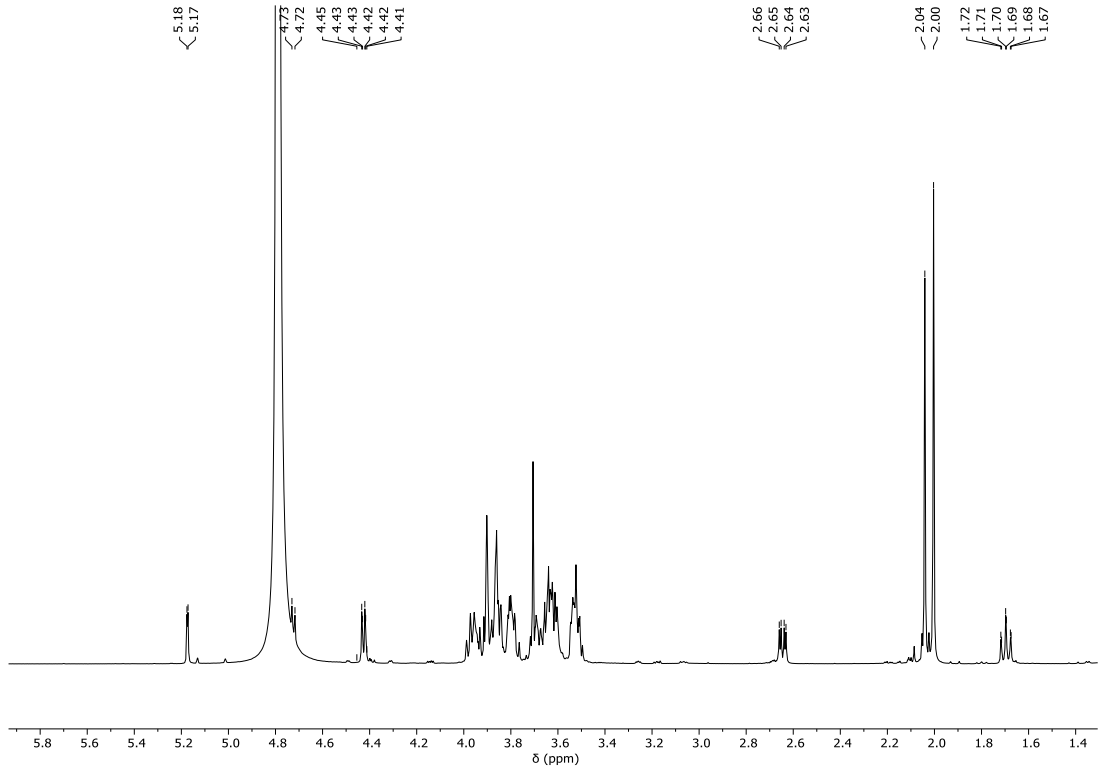


Figure 45. ^{1}H NMR spectrum of SAα2-6LacNAc (600 MHz, D_2O, 298 K).

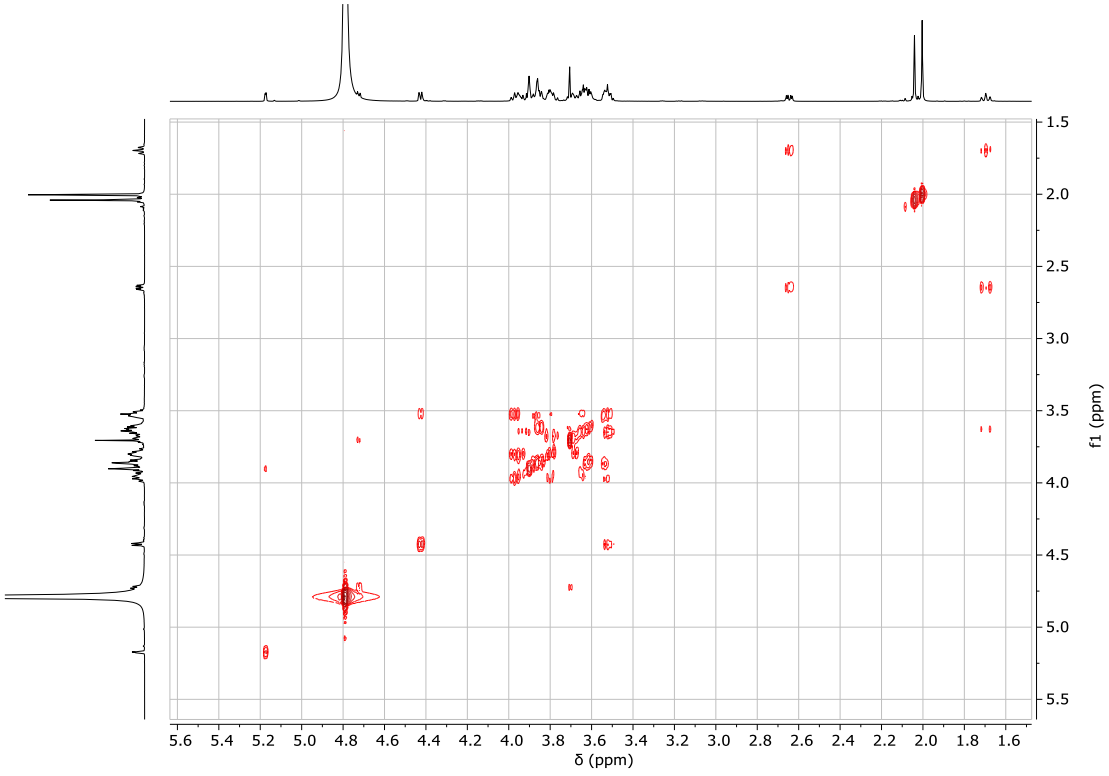


Figure 46. ^{1}H-^{1}H COSY NMR spectrum of SAα2-6LacNAc (600 MHz, D_2O, 298 K).

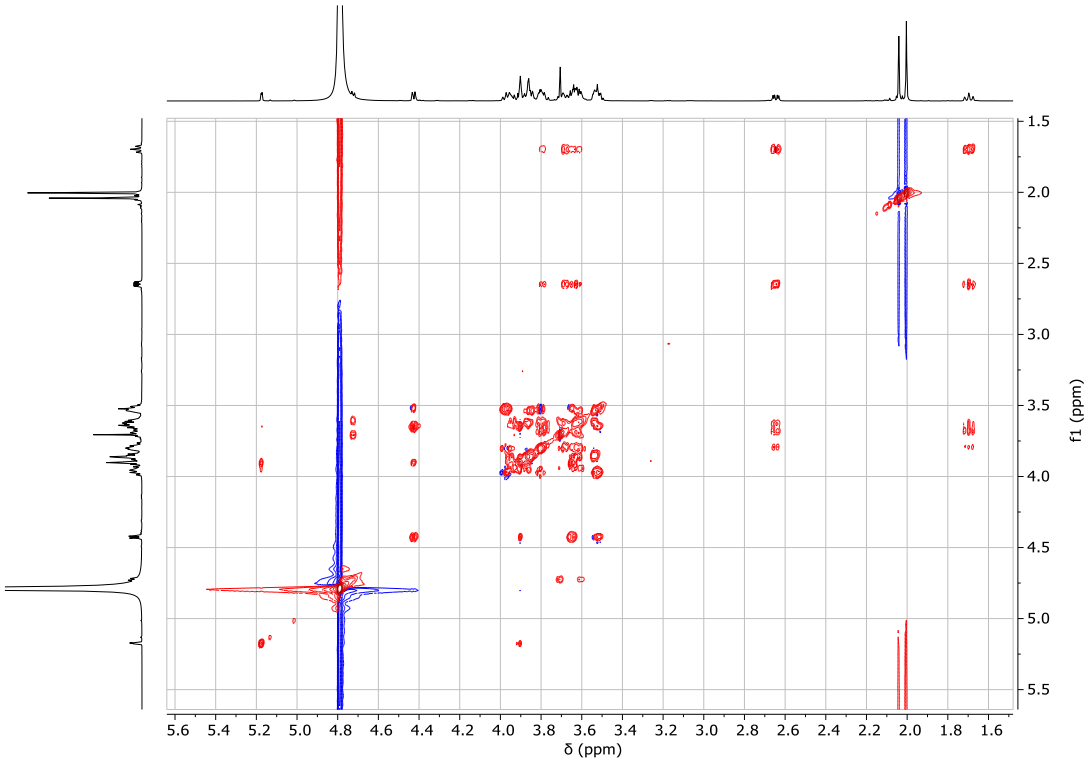


Figure 47. ^{1}H-^{1}H TOSY NMR spectrum of SAα2-6LacNAc (600 MHz, D_2O, 298 K).

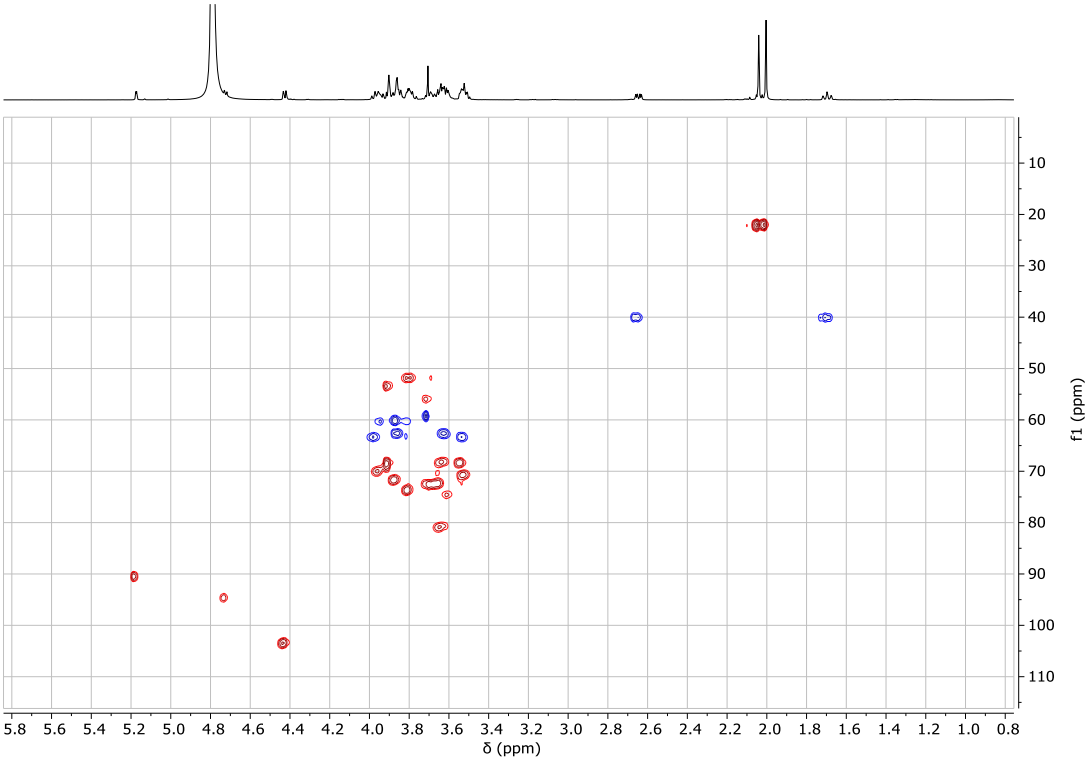


Figure 48. ^{1}H-^{13}C HSQC NMR spectrum of SAα2-6LacNAc (600 MHz, D_2O, 298 K).

ITC binding studies

For each ITC experiments, which consist of 3 titrations, only one titration graph is presented herein.

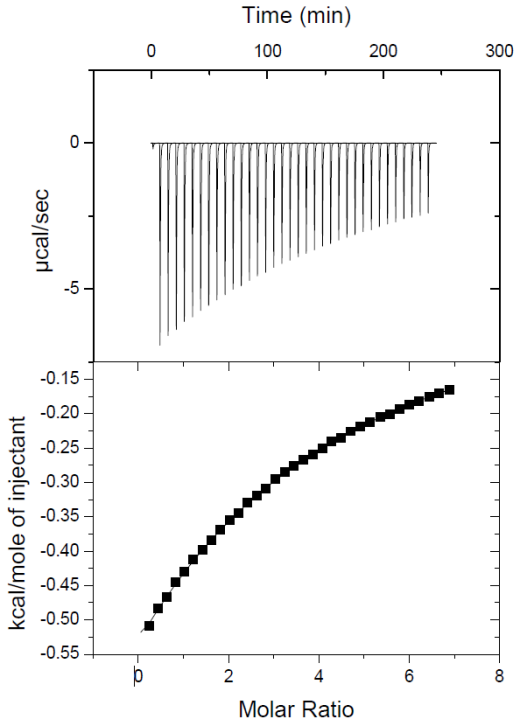


Figure 49. ITC titration graph of **B1** and **S1** at pH 5.5.

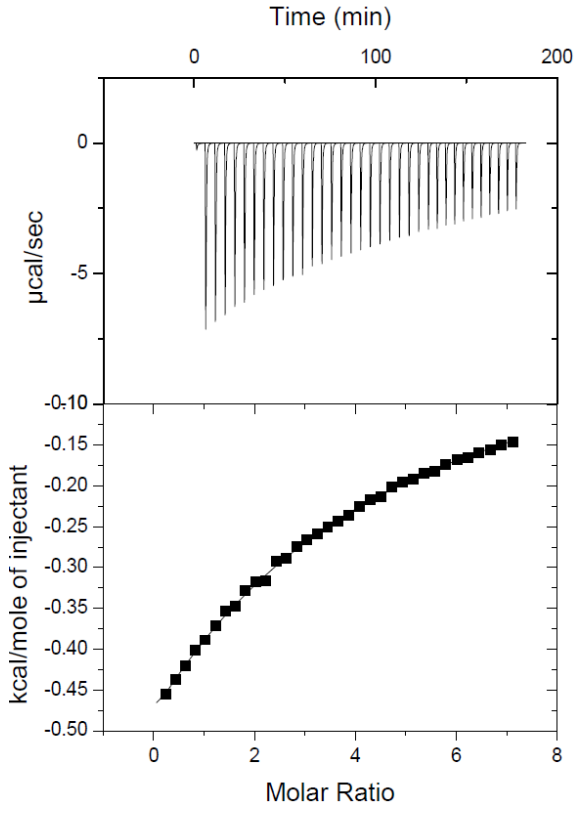


Figure 50. ITC titration graph of **B1** and **S1** at pH 6.5.

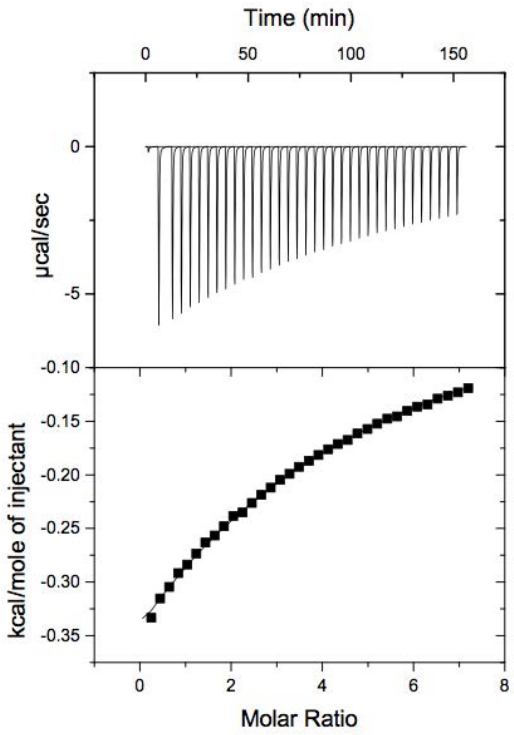


Figure 51. ITC titration graph of **B1** and **S1** at pH 7.4.

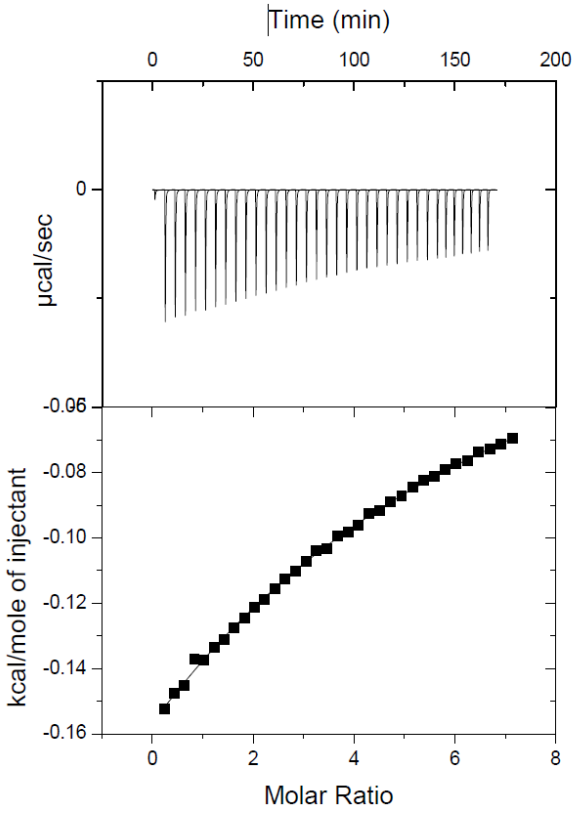


Figure 52. ITC titration graph of **B1** and **S1** at pH 8.5.

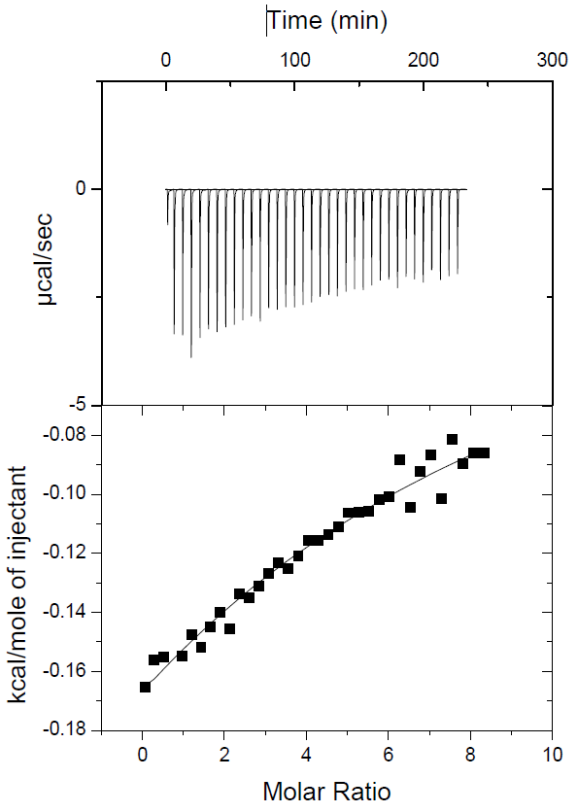


Figure 53. ITC titration graph of **B1** and **S1** at pH 10.

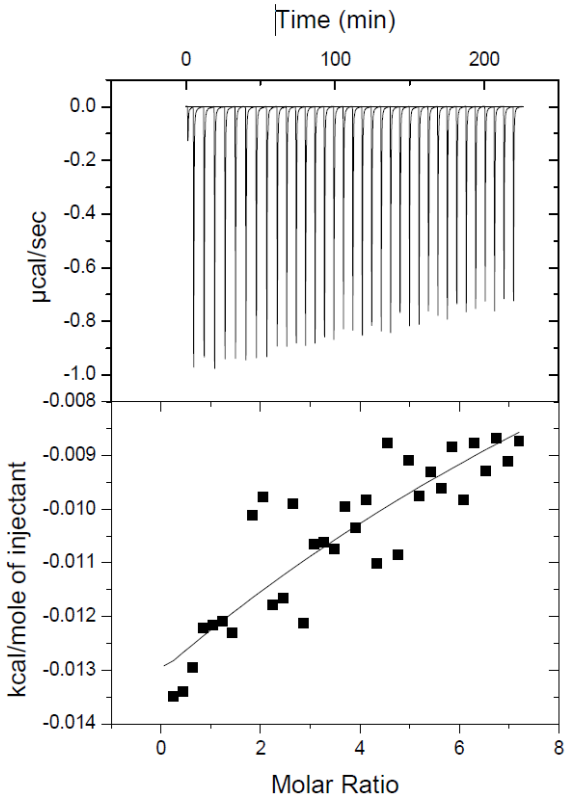


Figure 54. ITC titration graph of **B1** and **S2** at pH 5.5.

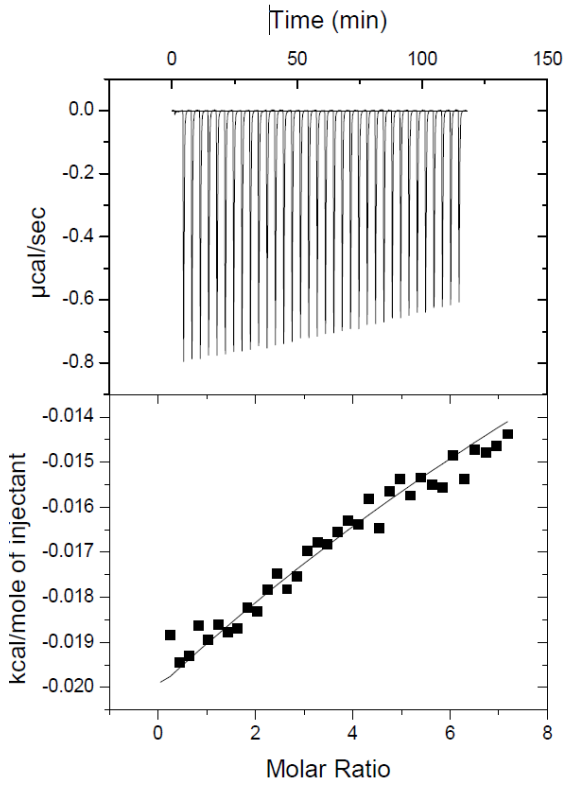


Figure 55. ITC titration graph of **B1** and **S2** at pH 7.4.

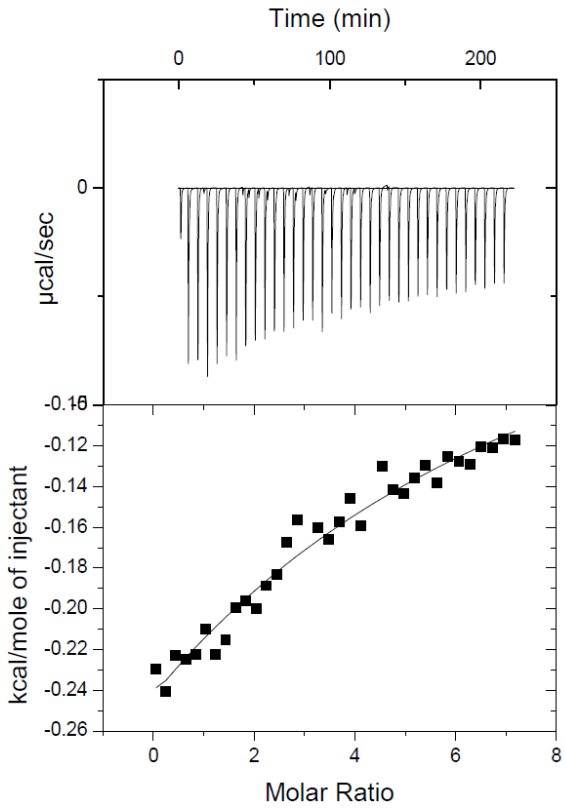


Figure 56. ITC titration graph of **B1** and **S3** at pH 5.5.

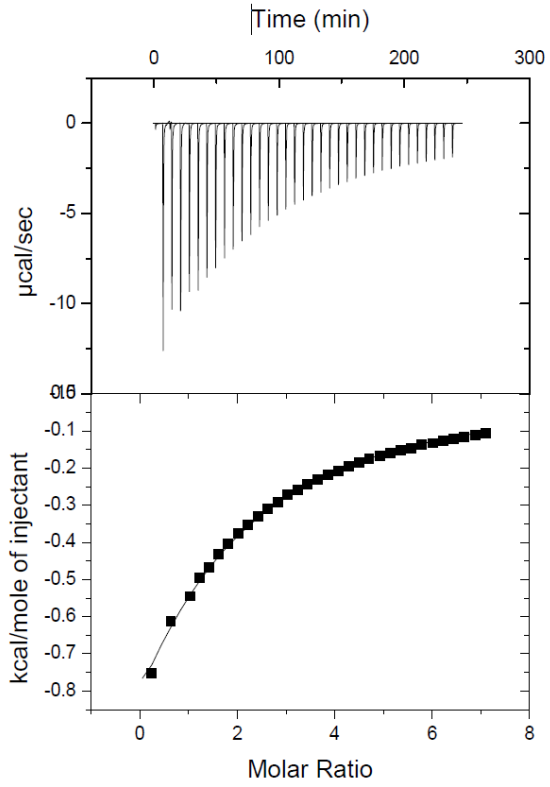


Figure 57. ITC titration graph of **B2** and **S1** at pH 5.5.

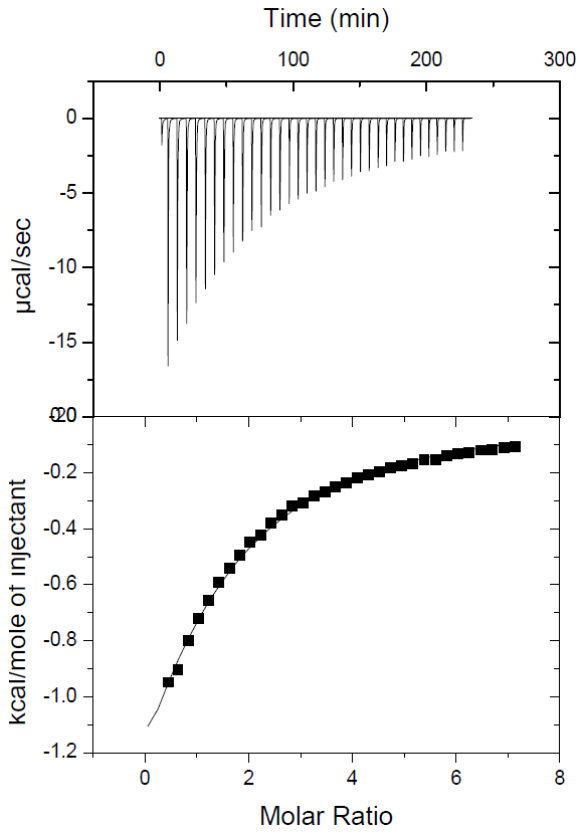


Figure 58. ITC titration graph of **B3** and **S1** at pH 5.5.

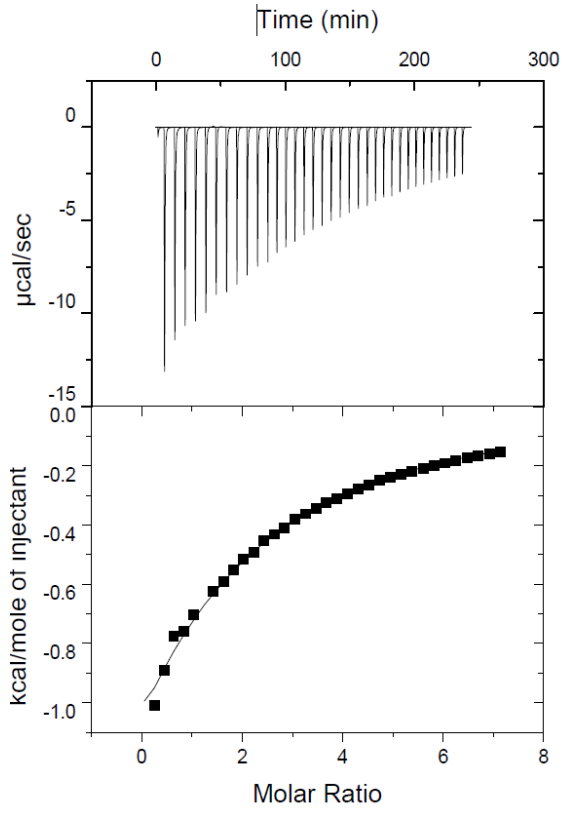


Figure 59. ITC titration graph of **B4** and **S1** at pH 5.5.

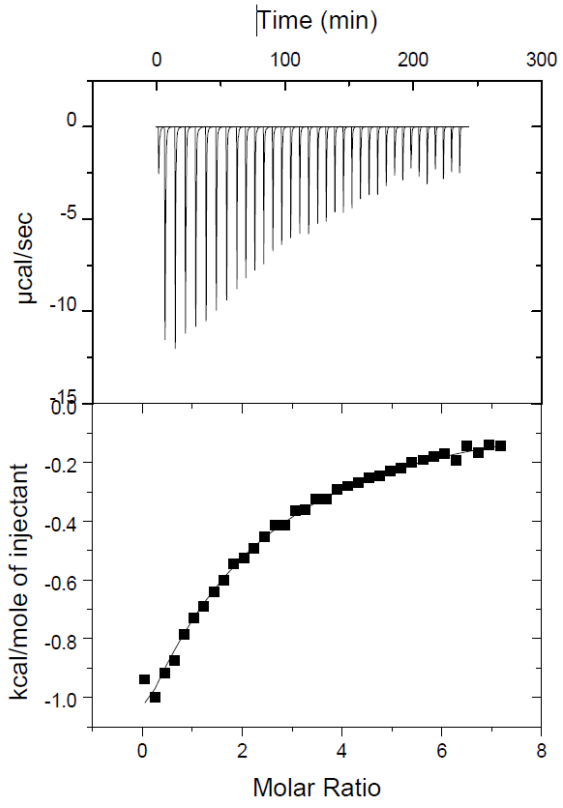


Figure 60. ITC titration graph of **B5** and **S1** at pH 5.5.

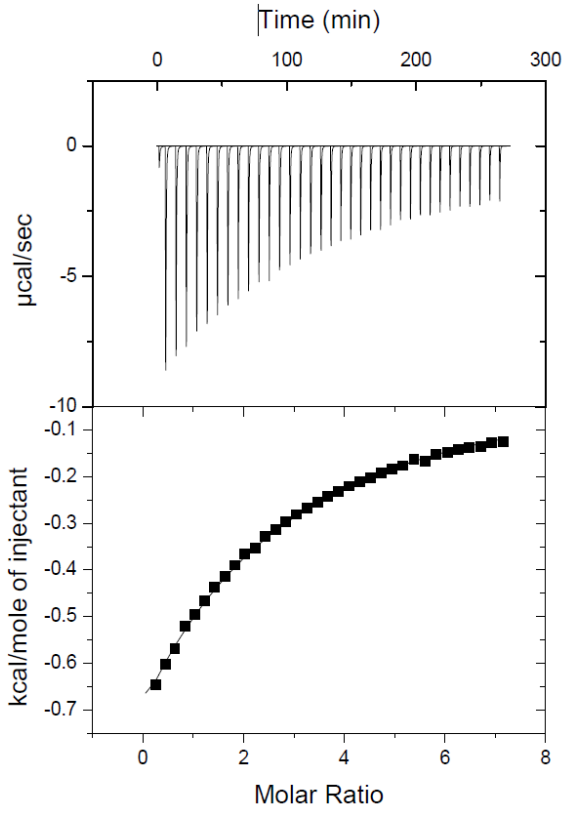


Figure 61. ITC titration graph of **B6** and **S1** at pH 5.5.

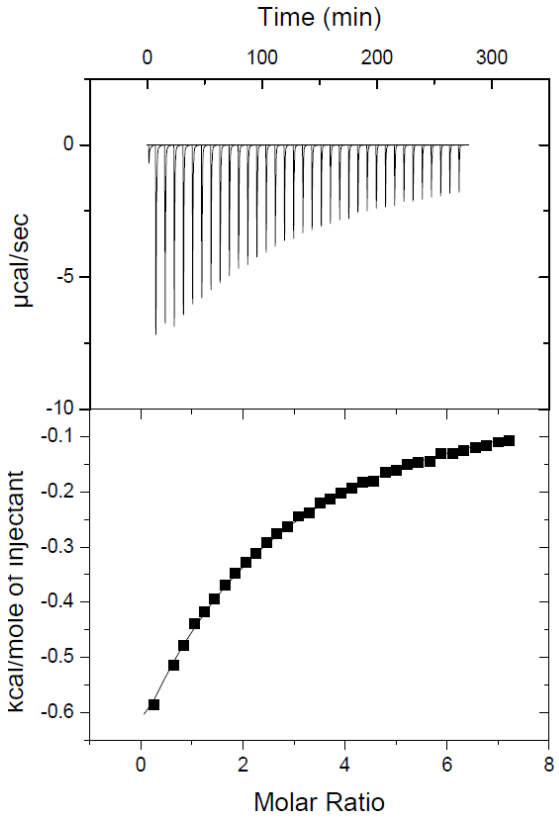


Figure 62. ITC titration graph of **B7** and **S1** at pH 5.5.

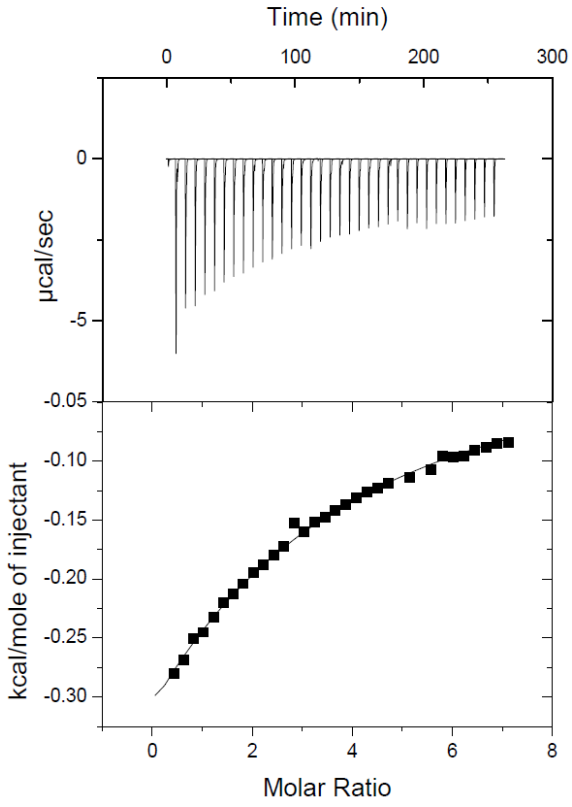


Figure 63. ITC titration graph of **B2** and **S1** at pH 6.5.

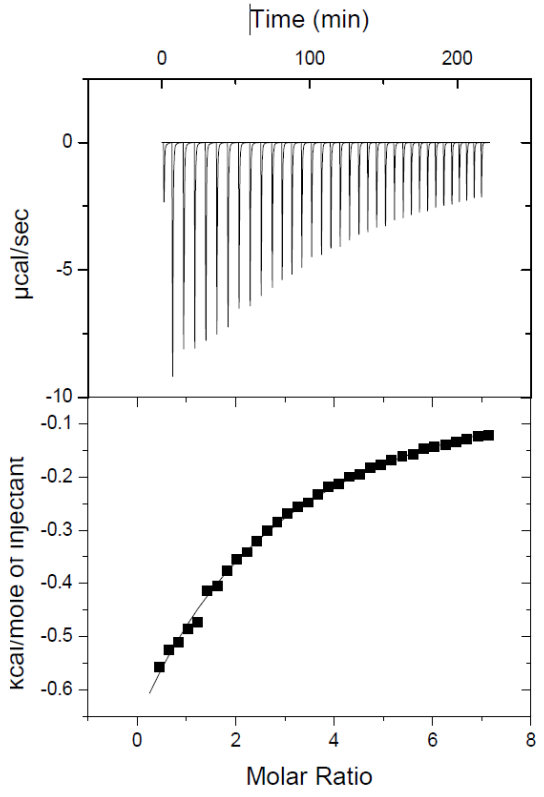


Figure 64. ITC titration graph of **B3** and **S1** at pH 6.5.

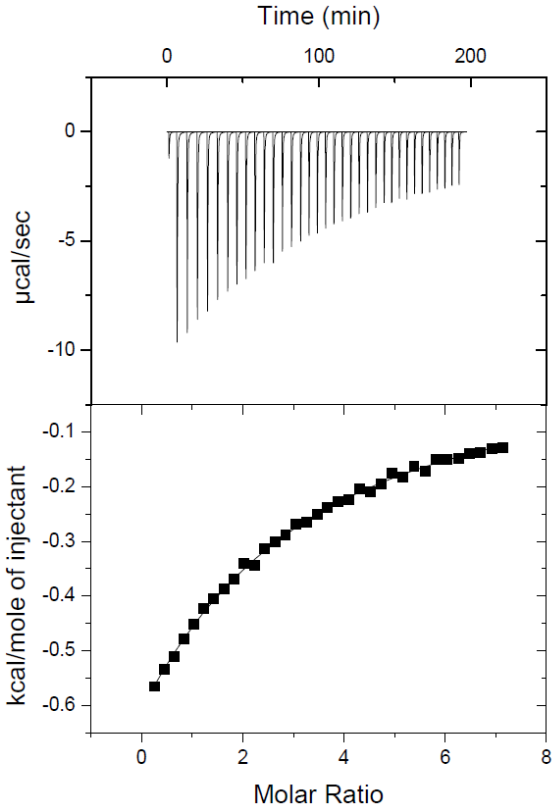


Figure 65. ITC titration graph of **B4** and **S1** at pH 6.5.

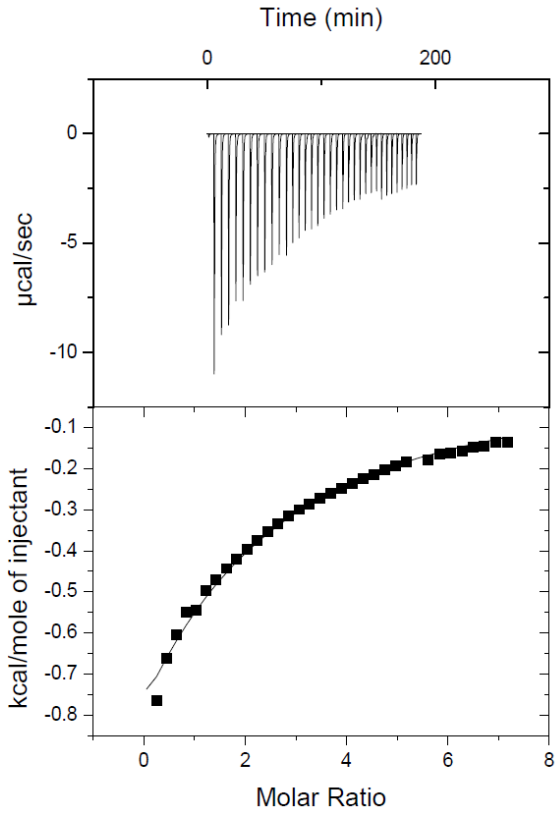


Figure 66. ITC titration graph of **B5** and **S1** at pH 6.5.

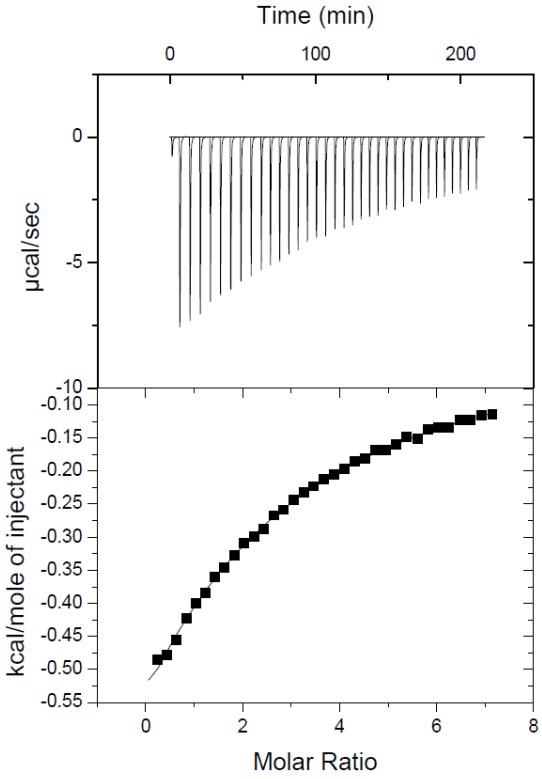


Figure 67. ITC titration graph of **B6** and **S1** at pH 6.5.

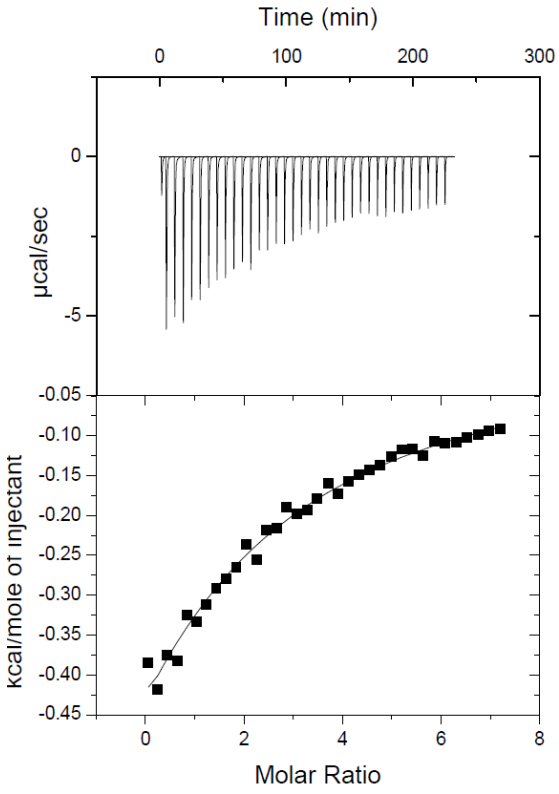


Figure 68. ITC titration graph of **B7** and **S1** at pH 6.5.

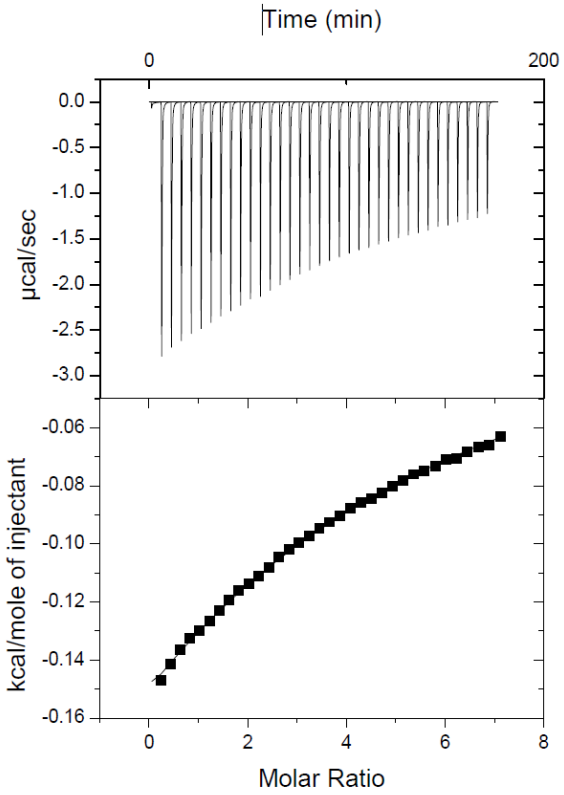


Figure 69. ITC titration graph of **B2** and **S1** at pH 7.4.

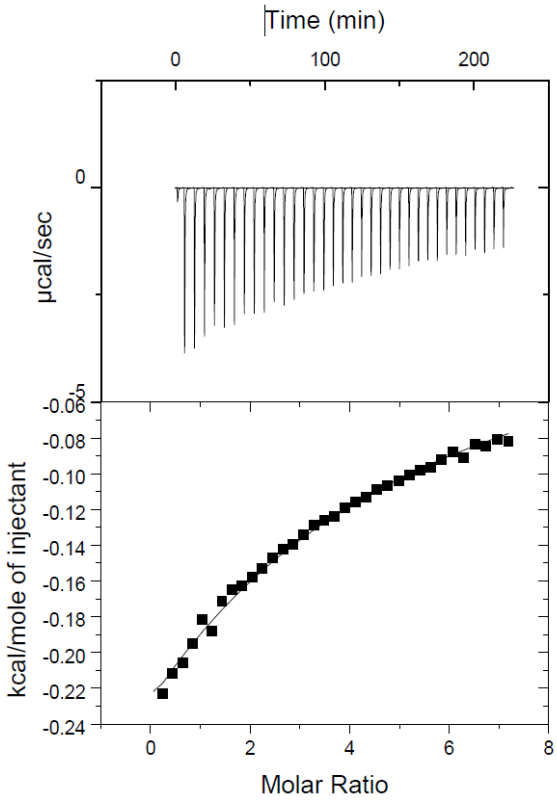


Figure 70. ITC titration graph of **B3** and **S1** at pH 7.4.

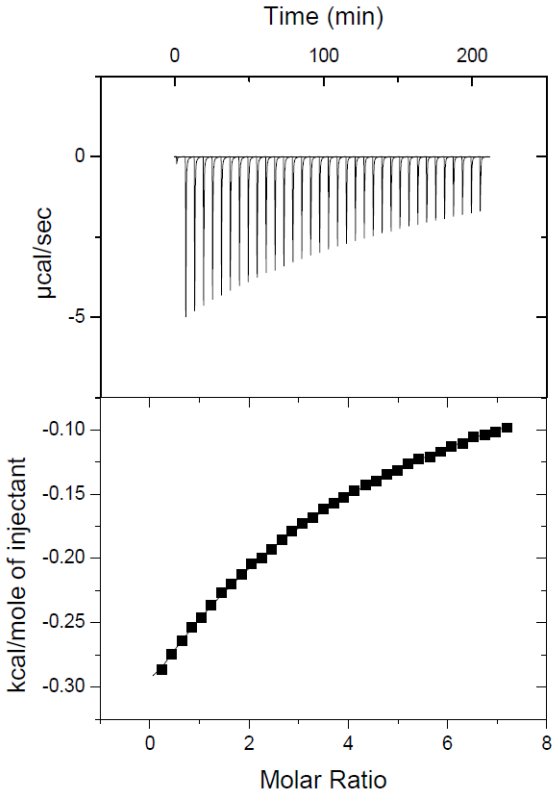


Figure 71. ITC titration graph of **B4** and **S1** at pH 7.4.

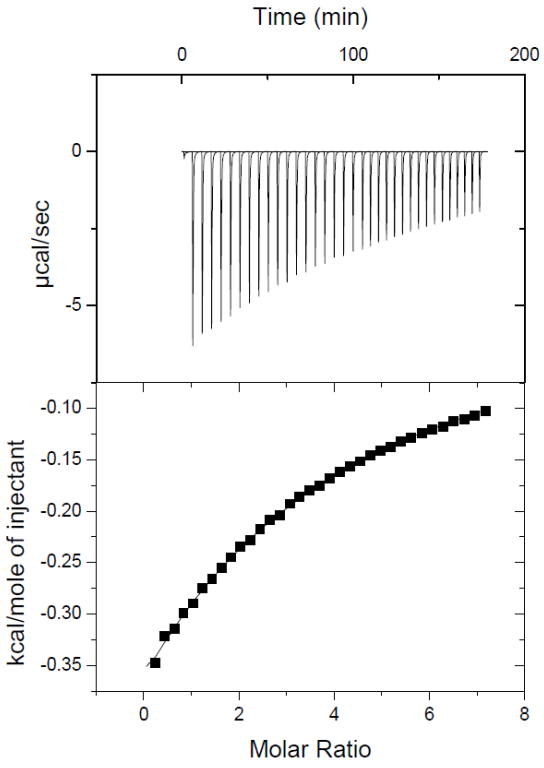


Figure 72. ITC titration graph of **B5** and **S1** at pH 7.4.

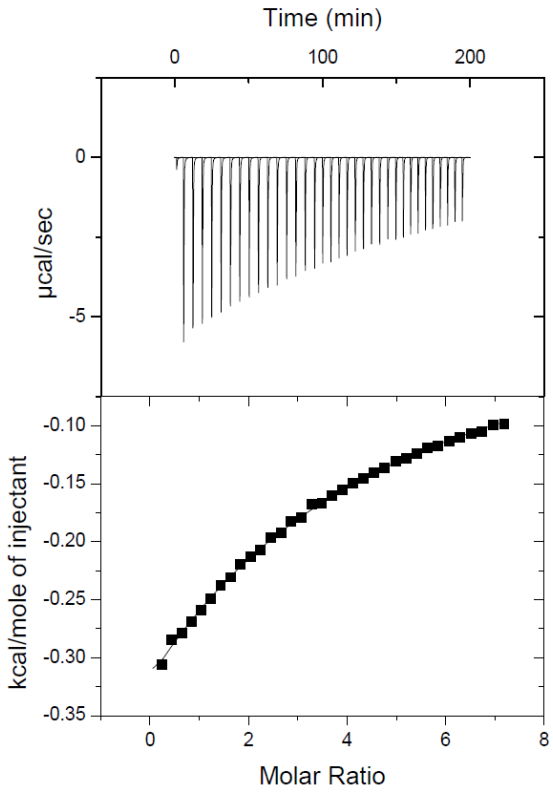


Figure 73. ITC titration graph of **B6** and **S1** at pH 7.4.

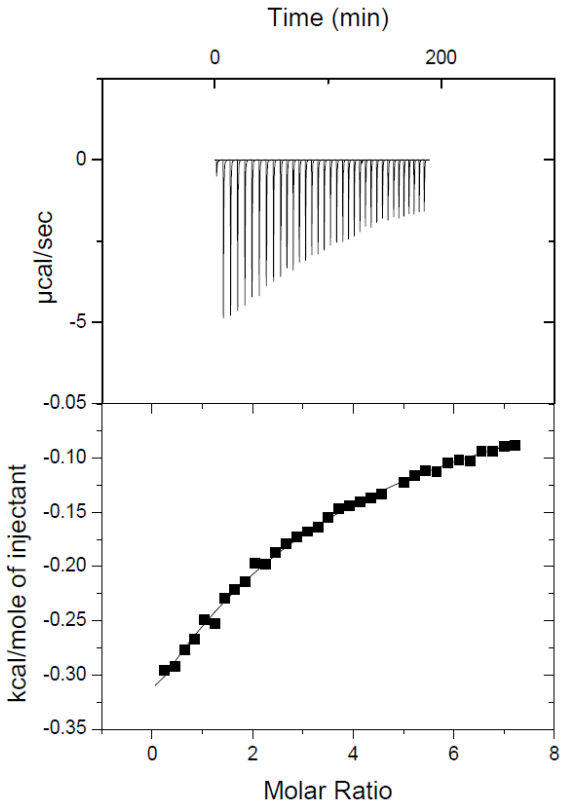


Figure 74. ITC titration graph of **B7** and **S1** at pH 7.4.

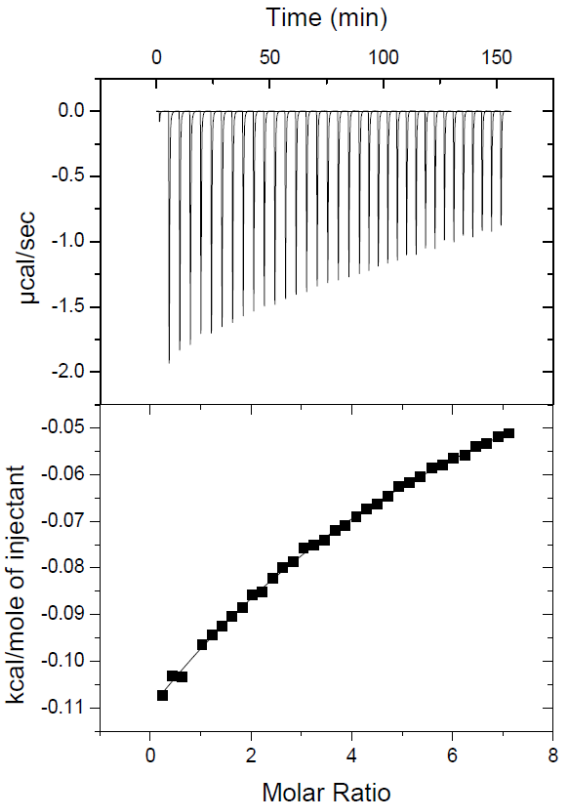


Figure 75. ITC titration graph of **B2** and **S1** at pH 8.5.

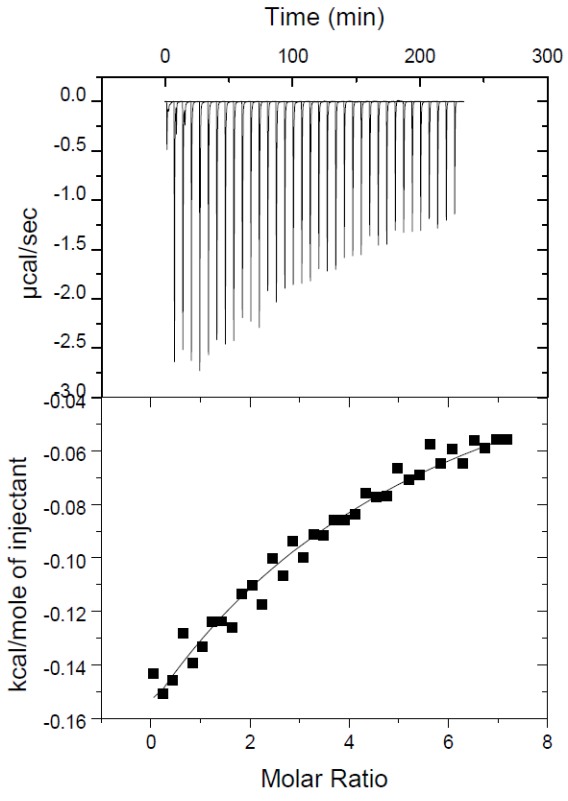


Figure 76. ITC titration graph of **B3** and **S1** at pH 8.5.

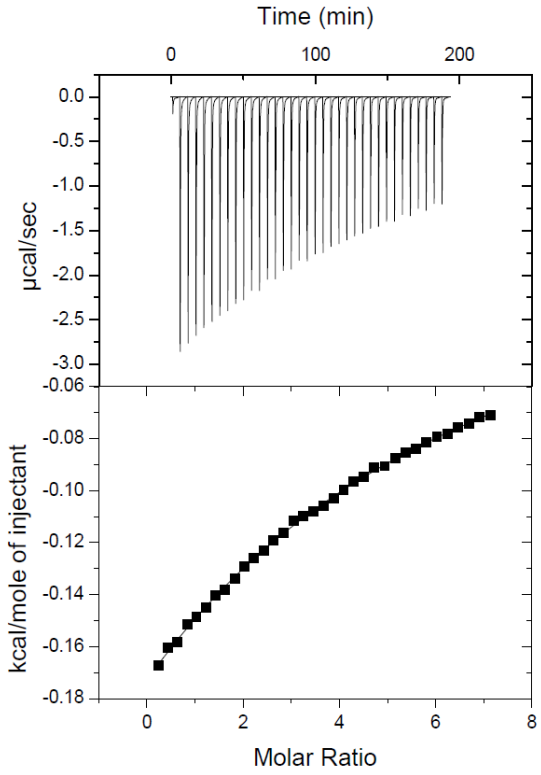


Figure 77. ITC titration graph of **B4** and **S1** at pH 8.5.

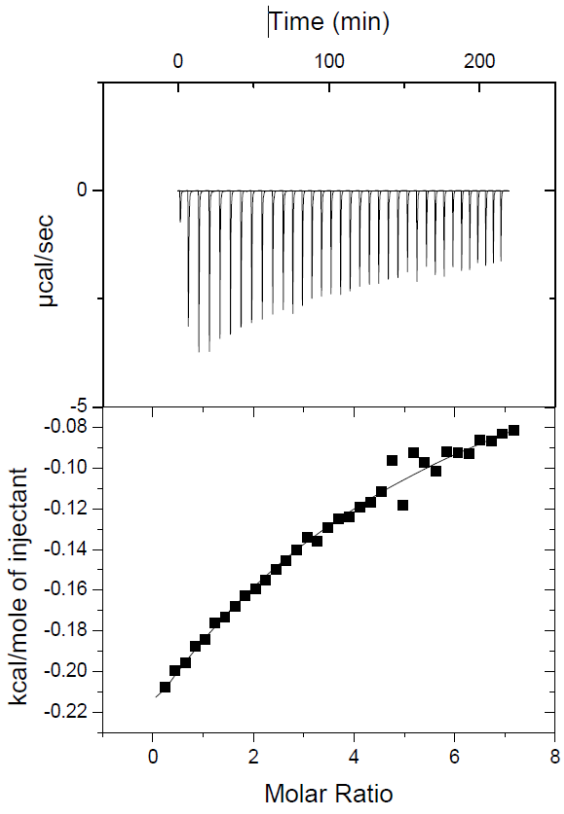


Figure 78. ITC titration graph of **B5** and **S1** at pH 8.5.

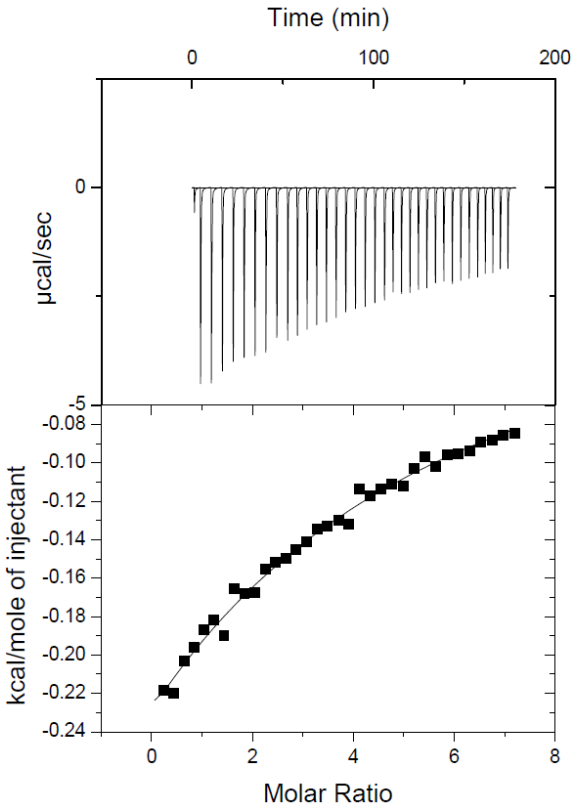


Figure 79. ITC titration graph of **B6** and **S1** at pH 8.5.

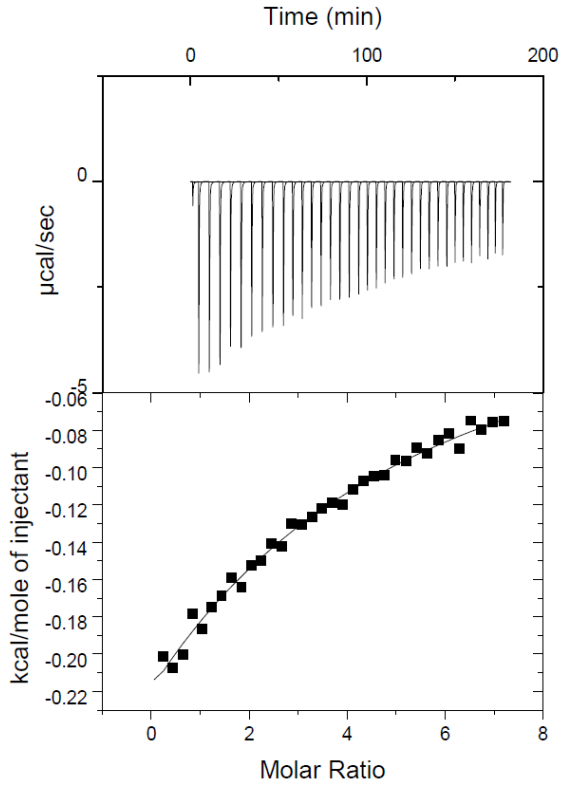


Figure 80. ITC titration graph of **B7** and **S1** at pH 8.5.