



UNIVERSITY OF
BIRMINGHAM

Discovery of selective saccharide receptors via Dynamic Combinatorial Chemistry

By

Miguel Ángel Aleñá Rodríguez

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

School of Chemical Engineering
College of Engineering and Physical Sciences
University of Birmingham
November 2023

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

The diagnosis of various diseases and pathological conditions can be accomplished by screening and detecting glycans in cells. Certain glycans serve as excellent biomarkers, being related to cell malfunctioning, while other structurally similar glycans perform completely different functions and are naturally present in healthy cells. Despite the theoretical feasibility of using glycans as biomarkers for early disease detection, our current inability to discriminate between them limits their use.

One promising approach to distinguishing between glycans is targeting their dissimilarities in saccharide chains. However, designing selective receptors for saccharides is challenging due to the complexity of these molecules. Their vast diversity, the fact that they exist in many interconvertible forms, their lack of recognisable functional groups, or the fact that they are normally heavily solvated in aqueous environments have made the design of receptors for saccharides a challenge that has kept the scientific community busy for the last 35 years. Although there have been ground-breaking discoveries in the field, improvements are needed to enhance our disease detection and risk stratification tools.

To address this challenge, we employed a technique known as Dynamic Combinatorial Chemistry (DCC). DCC enables the self-formation and self-selection of the best possible receptor for a given target from a pool or library of potentially good ligands. DCC has been effective for creating receptors for biomolecules such as DNA, RNA, and proteins, but its use for discovering sugar receptors is less explored. In this work, we filled this gap by implementing DCC for screening common saccharides (glucose, galactose, mannose, and

fructose) using small, simple, and inexpensive building blocks. Our results indicated that molecule **2DD**, which consists of a benzene ring with 2 units of amino acid aspartic acid derivatives connected in positions 1 and 3, is the best receptor in a library of very similar structures for the saccharides glucose, galactose, and mannose. For fructose, molecule **1P**, a benzene ring linked to just one unit of the amino acid phenylaldehyde, was appointed as the best receptor. The differential behaviour of fructose can provide insight into the relatively unknown processes behind molecular recognition of sugars.

Molecules **2DD** and **1P**, as well as some other library members as negative controls, were then synthesised for further testing and DCC results were then validated by Isothermal Titration Calorimetry (ITC) and NMR techniques, proving the effectiveness of DCC as a molecular recognition tool for the creation of receptors for saccharides. Moreover, molecule **1P** was found to have a high binding constant ($K_a = 1762 \text{ M}^{-1}$) and selectivity (50-100 times over other sugars) for fructose, which is surprisingly good considering the simplicity of the receptor.

A much more challenging approach was attempted by employing short peptides as scaffolds in DCC experiments. The benefits of using peptides are numerous but can be summarised in three bullet points: customisability, flexibility, and easiness in their synthesis. Unfortunately, we encountered many difficulties for the complete functionalisation of the peptides within the Dynamic Combinatorial Library (DCL) and this approach did not yield the desired results before the research project came to an end. However, we believe in its potential and the knowledge that we gained on the topic helped to establish the foundations on which new research will be carried out in the near future within the research group.

In summary, this thesis reports the development of a rapid methodology for the discovery of selective receptors for monosaccharides, employing a library of simple and inexpensive starting building blocks. While this was a proof-of-concept study, it can be scalable to larger library sizes and to target more complex biomolecules, becoming a useful tool that could accelerate the discovery of new molecules with biomedical applications.

Acknowledgements

First and foremost, I would like to express my gratitude to my main supervisor, Professor Paula Mendes. Her unwavering guidance and constant optimism helped me overcome challenging moments throughout this journey. Paula has been there for me from the very first day in the lab to assisting me in securing a job post-PhD. I feel incredibly fortunate to have had her as my supervisor and could not have asked for a better mentor.

Secondly, I would like to extend my sincere appreciation to my second supervisor, Dr. Marcos Fernandez-Villamarin, who is equally responsible (if not more) as myself for the successful completion of this thesis. Whether it was dealing with tedious paperwork or constructing a home-made reactor, his expertise and resourcefulness were instrumental in overcoming any obstacles that arose.

Thanks to Prof. Ignacio Alfonso and Dr Miriam Royo for giving me the opportunity to join their labs in Barcelona, and to Dr Carbajo and Dr Duran for their kind patience and valuable guidance during my brief yet fruitful stay.

I would also like to extend my heartfelt thanks to Francia, L. Brooks, and L. Buccoli, and to the rest of Mendes' group: Stephano, Josh, Craig, Setareh, Seyed, Barbara, Alice, Nasim, Dario, Pushpa, Charlie, Hannah, Daniel C., Daniel K., and Sohail. I cannot forget Dr Cécile Le Duff, Dr Chris Williams, Dr Allen Bowden, and all the technical staff at the Schools of Chemistry and Chemical Engineering. It has been a pleasure working alongside you guys.

I am deeply grateful to this country and to the city of Birmingham, for providing me with an opportunity that I could not find elsewhere. And I would like to thank the many friends I made here who made me feel at home in a place with little resemblance to the place where I was raised. If we ever played football, shared a pint at The Figure, or I ended up in your flat for pres (or an after): Thanks.

How lucky am I to have something that makes saying goodbye so hard.

Y por último pero no menos importante, gracias a mi familia por darme su apoyo y fuerza desde la distancia. Papá, Chiki, Álvaro, Gonzalo, abuela, tita Ani y tita Margot. Y a mi madre y abuelo que lo hicieron desde arriba. ¡Soy doctor!

Table of contents

Chapter 1 - Introduction.....	1
1.1 Glycans	2
1.1.1 Role of glycans in biological processes and importance in diseases.....	2
1.1.2 Structure & variety of glycans	3
1.1.3 Importance of glycans as biomarkers	7
1.1.4 Role of glycans in cancer detection.....	8
1.2 Saccharides.....	11
1.2.1 Complexity of monosaccharides	12
1.3 Saccharide recognition	16
1.4 Natural receptors	18
1.4.1 Lectins.....	18
1.4.2 Antibodies	23
1.5 Synthetic receptors	26
1.5.1 Non-covalent receptors.....	27
1.5.1.1 Aptamers	27
1.5.1.2 Macrocycles.....	30
1.5.1.3 Small-molecule acyclic receptors	35
1.5.2 Covalent receptors. Boronic Acids	38
1.6 DCC.....	48
1.6.1 Different types of DCC.....	52
1.6.2 Starting BBs in a DCC experiment	54
1.7 Conclusions	57
1.8 Aim of this work	59
Chapter 2 - Techniques	60
2.1 NMR.....	60
2.2 HPLC	66
2.3 LC-MS	70
2.4 ITC.....	72
Chapter 3 - Optimisation of Imine Formation Reaction for its use in DCC	81
3.1 Optimisation of imine formation reaction (IFR).....	81
3.1.1 Equilibration time and concentration of reagents in IFR	81
3.1.2 Influence of pH in IFR	87
3.2 Conclusions of Chapter 3.....	99
Chapter 4 - DCL- Small molecule approach.....	101

4.1	DCC Method development.....	101
4.1.1	LC-MS method development	101
4.1.2	Conclusions and remarks on LC-MS method development	112
4.1.3	DCC 1.0 – method development	113
4.1.4	Conclusions and remarks on DCC 1.0 experiment	117
4.2	DCC 2.0 - Small molecule approach	118
4.3	Conclusions of Chapter 4.....	127
Chapter 5 -	DCC Validation.....	130
5.1	Characterisation of the receptors. Binding studies.....	130
5.2	ITC.....	134
5.2.1	Optimisation of ITC experimental conditions	134
5.2.2	ITC to assess the binding properties of the amplified receptors	141
5.2.3	Conclusions on the ITC experiments	153
5.3	NMR titrations	155
5.4	Conclusions of Chapter 5.....	163
Chapter 6 -	DCL – Peptide approach	165
6.1	Benefits of using peptides as scaffold in DCC experiments	165
6.2	Selection of BBs for peptide-based DCC	168
6.3	Solid-phase peptide synthesis (SPPS) to make peptides for DCC 3.0 experiments	171
6.3.1	General considerations	172
6.3.2	SPPS resins.....	173
6.3.3	Conditioning of the resin and incorporation of the first amino acid	174
6.3.4	Loading determination by Fmoc quantification	176
6.3.5	Capping of unreacted positions	177
6.3.6	Kaiser test.....	179
6.3.7	Fmoc removal.....	180
6.3.8	On-resin elongation of the peptide.....	181
6.3.9	Alloc removal.....	183
6.3.10	Peptide cleavage	184
6.3.11	Manual SPPS with temperature	185
6.3.12	On-resin reductive amination	186
6.4	DCC 3.0	192
6.5	DCC 3.1 and 3.2	193
6.5.1	Synthetic protocol to afford KAK and KAAK by automated SPPS.....	196
6.6	Conclusions of Chapter 6.....	198
Chapter 7 -	Experimental	200

7.1	Materials & methods.....	200
7.1.1	NMR.....	201
7.1.2	MS.....	201
7.1.3	HPLC	201
7.1.4	ITC.....	202
7.1.5	SPPS.....	202
7.2	Synthesis of amine I for IFR optimisation experiments	202
7.3	NMR Studies.....	204
7.3.1	Buffer preparation.....	204
7.3.2	Preparation of solutions for NMR	204
7.4	Design of DCC experiment small-molecule approach.....	205
7.5	Synthesis of small-molecule receptors	206
7.5.1	Synthesis of monosubstituted receptors 1D and 1P.....	206
7.5.2	Synthesis of disubstituted receptors 2DD and 2PP.....	209
7.6	Solid-phase peptide synthesis (SPPS).....	211
7.6.1	Manual SPPS.....	211
7.6.1.1	Conditioning of the resin and incorporation of the first amino acid	211
7.6.1.2	Loading determination by Fmoc quantification	212
7.6.1.3	Capping of unreacted positions	212
7.6.1.4	Fmoc removal.....	212
7.6.1.5	On-resin elongation of the peptide.....	212
7.6.1.6	Alloc removal.....	212
7.6.1.7	Peptide cleavage	213
7.6.1.8	On-resin reductive amination	214
7.6.2	Automated SPPS.....	216
Chapter 8 -	Conclusions & Future work	218
	List of references.....	224
APPENDIX	267	

List of definitions & abbreviations

AA(s)	Amino acid(s)
Ac ₂ O	Acetic anhydride
Ala	L-Alanine
Alloc	Allyloxycarbonyl protecting group
Ar	Argon
Arg	L-Arginine
ARS	Alizarin Red S
Asn	L-Asparagine
B ₀	Magnetic field
BA(s)	Boronic Acid(s)
BB(s)	Building Block(s)
Boc	tert-Butoxycarbonyl protecting group
Bzl	Benzyl
Cys	L-Cysteine
δ	Chemical shift
D ₂ O	Deuterated water
Dab	L-2,4-Diaminobutyric acid
Dap	L-2,3-Diaminopropionic acid
DCC	Dynamic Combinatorial Chemistry

DCL	Dynamic Combinatorial Library
DCM	Dichloromethane
DIC	N,N'-Diisopropylcarbodiimide
DIPEA	N,N -Diisopropylethylamine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DP	Differential power
FA	Formic acid
Fmoc	Fluorenylmethyloxycarbonyl protecting group
Fuc	Fucose
γ	Gyromagnetic ratio
Gal	Galactose
Glc	Glucose
GlcNAc	N-acetyl glucosamine
Gly	Glycine
H	Enthalpy
h	Planck constant
H ₂ SO ₄	Sulfuric acid
HCl	Hydrochloric acid
His	L-Histidine

HPLC	High Performance Liquid Chromatography
I	Nuclear spin quantum number
IdoA	Iduronic acid
IFR	Imine Formation Reaction
ITC	Isothermal Titration Calorimetry
J	Coupling constant
K_a	Binding constant
K_d	Dissociation constant
LC	Liquid Chromatography
Lys	L-Lysine
Man	Mannose
MeOH	Methanol
MeCN	Acetonitrile
MS	Mass Spectrometry
m/z	Mass to charge ratio
ν	Frequency
n	Stoichiometry of binding, number of binding sites
NaOH	Sodium hydroxide
NMR	Nuclear Magnetic Resonance
O.N.	Overnight

Orn	L-Ornithine
PBA	Phenylboronic Acid
PCa	Prostate cancer
Phe	L-Phenylalanine
ppm	Parts per million
Pro	L-Proline
PSA	Prostate Specific Antigen
RPM	Revolutions per minute
rt	Room temperature
SA	Sialic Acid
Ser	L-Serine
SPPS	Solid Phase Peptide Synthesis
T	Temperature
TFA	Trifluoroacetic acid
Thr	L-Threonine
TIPS	Triisopropylhydrosilane
TLC	Thin Layer Chromatography
Trp	L-Tryptophan
UV	Ultraviolet
V ₀	Volume of sample cell

Chapter 1 -Introduction

Glycans are saccharide-containing molecules that play a key role in biological processes. Abnormal glycosylation patterns can be observed in cells in patients suffering from a large number of diseases. Therefore, glycans can be used as biomarkers becoming their study a major research topic to address disease risk stratification. The vast diversity of glycans in cells make their study a huge challenge as closely structural related glycans may be involved in completely different processes. It is believed that the key to tell these structures apart is to understand the differences in their saccharide chains. Thus, in order to design receptors for such biomarkers, the concepts of saccharide recognition must be contemplated.

There is no consensus regarding the key features needed for saccharide recognition. There are examples of powerful receptors for sugars from both natural and synthetic sources, relying on different intermolecular covalent and non-covalent interactions. However, scientist do agree on the necessity of multivalency to create potent receptors. Hence, Dynamic Combinatorial Chemistry (DCC) is an interesting approach for creating powerful receptors for saccharide recognition as it allows for the creation of multivalent and diverse systems exhibiting different chemical functionalities.

This chapter will begin with an extensive exploration of glycans, their role in biological processes, and their use as biomarkers for a wide variety of diseases including cancer. We will see the crucial role that sugars play in this regard and hence we will continue with a study of the field of saccharide recognition. We will explore the work that has been done so far, with a comprehensive literature review on the receptors for saccharides developed to date, and how a technique such as DCC could help to create more powerful and selective receptors.

1.1 Glycans

Glycans are usually referred as carbohydrate-based polymers. They can be composed of a variety of different monosaccharides and can be linked together in a number of different ways to form complex branching structures. Additionally, glycans can be decorated with a variety of different chemical groups, such as sulfo groups, phosphate groups, or methyl groups. They can also exist in combination with biomolecules like proteins or lipids via glycosylation. The diversity that these molecules can achieve is therefore huge. Moreover, they can assemble in a number of forms leading to different regio- and stereoisomers. They can also bind sequentially affording molecules with a wide range of molecular weights. In short, glycans are characterised by a remarkable structural diversity.¹

1.1.1 Role of glycans in biological processes and importance in diseases

Glycans play important roles in various biological processes including metabolism, cell adhesion or communication between cells.² They also provide physical and structural support and have protective, stabilizing, organizational, and barrier functions. The glycocalyx -a layer of glycans that covers all eukaryotic cells- provides a protective and adhesive coating. Protein-bound glycans are present in the nucleus and cytoplasm of cells, where they regulate cell signalling pathways. Glycans attached to matrix molecules such as proteoglycans help maintain tissue porosity, structure, and integrity.³ Additionally, N-glycosylation of membrane receptors is important for a wide variety of cellular processes including immune system function.⁴

Glycans are also implicated in disease development and progression, particularly in relation to neurodegenerative disorders,^{5,6} inflammatory bowel disease,⁷ and cancer.^{8–12} The specific roles that they play in this topic can vary widely, from promoting tumour growth and invasion to contributing to the toxic build-up of misfolded proteins in neurodegenerative diseases. Hence, there is an existing demand for research in this area.¹³

1.1.2 Structure & variety of glycans

Glycans are the most structurally diverse and complex biomolecules in nature.¹⁴ They can exist as a free polysaccharide chain, or they can be conjugated to protein or lipids. As mentioned in the beginning of this chapter, they play key roles in many biological processes. It is often that their carbohydrate motifs are the feature exposed on the cell surfaces acting as receptors for pathogens like bacteria and viruses.¹⁵ Hence, these are the centre of attention when studying the biological implications of glycans. While disregarding the rest of the molecule and focusing on the sugar chains may seem a simplification, the truth is that the structure of the saccharide chains alone is still a major source of complexity.

Carbohydrate chains in glycans are formed by different combinations of only 10 monosaccharides (Figure 1).¹⁶ 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).¹⁷

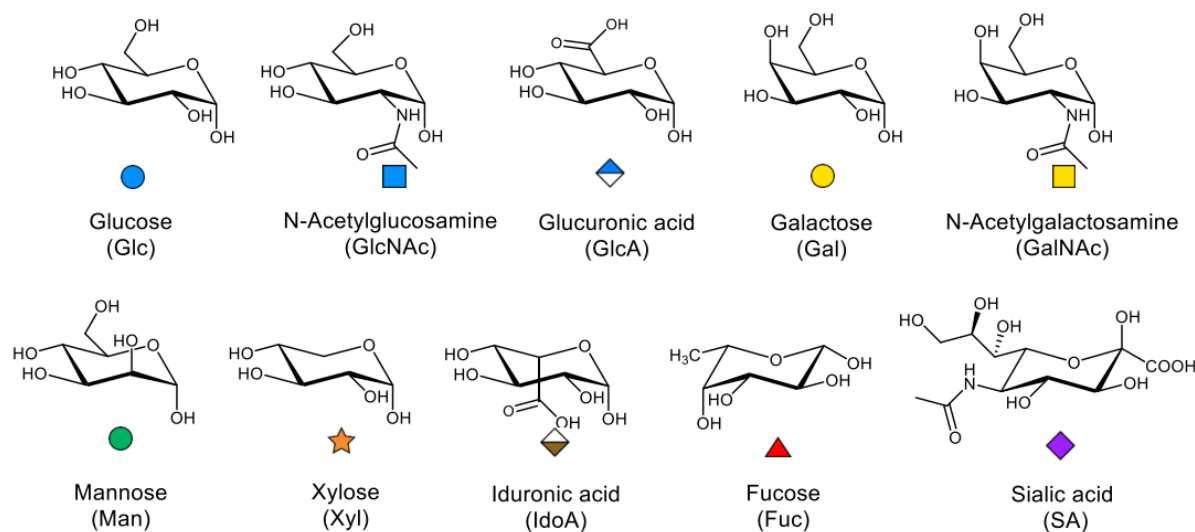


Figure 1: Structure of the monosaccharides found in glycans, their abbreviation and symbols.¹⁷

In a post-translational procedure known as glycosylation, the monosaccharides are joined together to produce di-, oligo-, or polysaccharides. Glycosyltransferases and glycosidases are the two major types of enzymes that catalyse glycosylation. Whereas glycosidases enzymes facilitate the linkage's hydrolysis, glycosyltransferases mediate its creation. The production of a specific glycosidic bond between sugar residues is catalysed by each of the more than 200 different kinds of glycosyltransferases.¹⁸ According to the stereochemistry of the anomeric carbon involved in the linkage, glycosidic bonds can be classified as either α or β as it will be explained later in Chapter 1.2, a section dedicated to saccharides.¹⁹

Different hydroxyl groups of a given sugar can be used to construct glycosidic bonds, leading to the possibility of several regioisomers. For instance, galactose can be connected to sialic acid at positions 3 or 6, resulting in the sialylated epitopes α 2-3 and α 2-6 (Figure 2).¹⁷ Moreover, the same sugar residue may form more than one glycosidic linkage, leading to the

formation of branched glycans. For example, a galactose (Gal) residue can attach to a GalNAc and a sialic acid (SA) unit at the same time through positions 1 and 3.¹⁷ After glycosylation, the saccharide units can go through changes and add groups like phosphate, acetyl, and sulphate. All of these factors combined are responsible for the immense variety of glycans expressed in cells and the huge amount of information that can therefore be stored on them. This is known as the 'sugar code'.²⁰

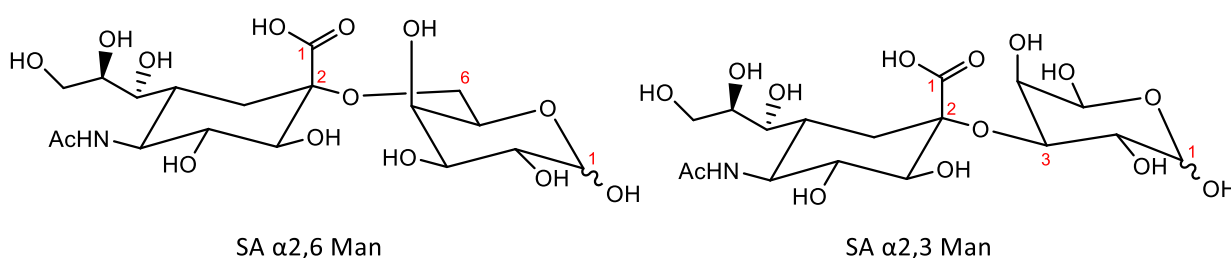


Figure 2: Sialic Acid (SA) connected to Mannose (Man) via α2,6 (left) and α2,3 (right). Stereochemistry of C1 in Mannose not specified.

Glycoproteins or glycolipids are created when glycans are attached to carrier molecules such as proteins or lipids. The sugar chain is attached to an asparagine residue in glycoproteins to create N-glycans, or alternatively to a serine or threonine residue to create O-glycans. N- and O-glycans can be linear and highly branched. There are three types of N-glycans: high-mannose, hybrid, and complex type (Figure 3). The core sequence of all three types is composed of Man and GlcNAc residues.²¹ The high-mannose type presents mostly mannose units linked to the core, whilst other N-glycans are more heterogenic. Hybrid N-glycans are mono- or bi-antennaries,²¹ while the complex type are extensively branched with up to six ramifications.²¹ N-glycans have important functions in the production of glycoproteins,

helping to ensure proper folding.²² On the other hand, O-glycans are mucins and proteoglycans. Proteoglycans are linear and play a role in cell-matrix interaction, proliferation, and inflammation,²³ whereas mucins are heavily branched and are implicated in cell adhesion and defence against pathogens.²⁴

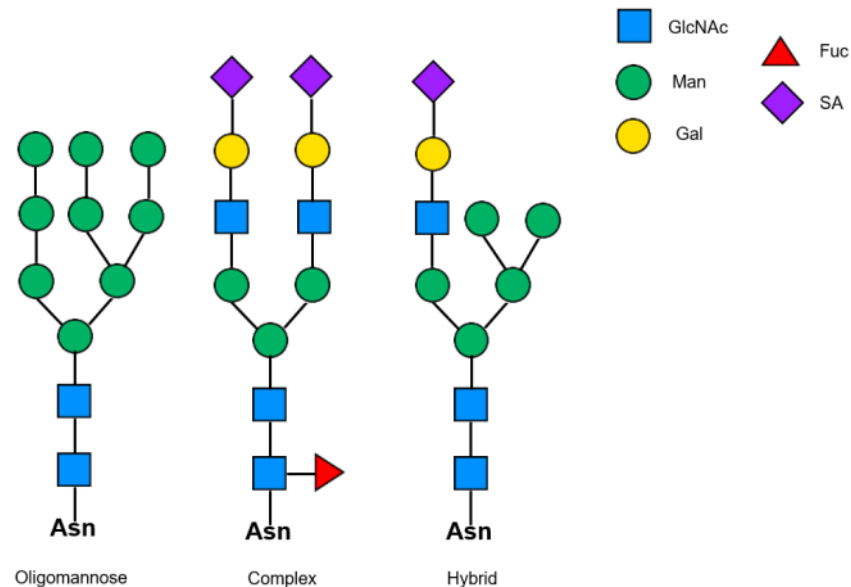


Figure 3: Structures of oligomannose, complex and hybrid N-glycans. Image reproduced from reference 21.

Glypolipids are glycans connected to lipidic scaffolds. Glycosphingolipids (GLS), which can be neutral or acidic, are glycolipids with a sphingoid or ceramide scaffold.²⁵ Globosides and cerebroside are examples of neutral GLS, whereas gangliosides and sulfatides are acidic GLS.²⁶ Gangliosides with sialic acid residues, such as the GM1, are particularly important in the central nervous system because they help with neuronal transmission and system maintenance as they participate in the repair of neuronal cells, memory formation and synaptic transmission.²⁵

1.1.3 Importance of glycans as biomarkers

In recent years, glycans have emerged as promising biomarkers for various diseases, including cancer, autoimmune diseases, and neurological disorders. The use of glycans as biomarkers has several advantages. First, saccharide chains provide glycans with structural stability, helping them maintain their tertiary three-dimensional shape in a wider range of pHs than proteins. Thus, glycans can be found in all sorts of body fluids such as blood and urine, making them attractive targets for non-invasive diagnostic tests. Second, glycans can be related to very specific biological processes, and therefore they can be used to distinguish between different disease states, providing the opportunity for a high level of accuracy in disease diagnosis and monitoring.

Several methods have been developed for detecting and analysing glycans as biomarkers. These include mass spectrometry, which can provide detailed information on the structures and quantities of glycans; lectin microarrays, which can be used to determine specific glycan-binding patterns;²⁷ and glycan microarrays, which can be used to identify novel glycan-binding proteins.²⁸

Glycan dysregulation has been identified in several neurodegenerative disorders. In Alzheimer's disease, the glycosylation profile of key disease-associated glycoproteins such as Amyloid Precursor Protein (APP) is altered.⁶ Aberrant glycosylation patterns of immunoglobulin G (IgG) have been identified as potential biomarkers for autoimmune diseases such as lupus and rheumatoid arthritis.²⁹

The role of glycosylation in inflammatory bowel disease (IBD) is not fully understood, but it is believed to affect various aspects of the disease. Glycosylation has been shown to contribute to the function and stability of the intestinal mucosal barrier, which is critical in maintaining gut homeostasis and preventing inflammation. Additionally, changes in glycosylation have been identified in various proteins and molecules involved in IBD pathology, such as cytokines and immunoglobulins.^{30,31}

1.1.4 Role of glycans in cancer detection

Glycosylation is also involved in cancer growth and progression, with reports of glycans promoting tumour development and contributing to the invasive properties of cancer cells. Altered glycans have been observed in cancer cells in patients suffering from prostate, breast, or ovarian cancer among others, and these changes have been associated with tumour progression and metastasis.^{32–34}

These modifications are caused by genetic and epigenetic dysregulation of glycozymes which control the expression of glycosyltransferases and glycosidases.¹⁵ The alteration of glycans in glycoproteins or glycolipids results in conformational changes of the whole macromolecule causing variation in the biological functions,¹⁷ including variation of cell to cell and cell to matrix interactions.¹⁵

As an example, in the case of prostate cancer (PCa), the analysis of PSA glycosylation patterns in blood emerged as a feasible tool. 8% of PSA total weight is due to carbohydrates in the form of a single N-oligosaccharide chain linked to Asn-45 amino acid.³⁵ Since oncogenic processes

are known to produce modifications in that chain pattern, it was considered a key point to research about. Among other kinds of glycosylation, sialylation grade has been extensively studied. This means the incorporation of the sialic acid (SA) monosaccharide by $\alpha 2,3$, $\alpha 2,6$ or $\alpha 2,8$ bond to the chain. An increase in $\alpha 2,3$ SA percentage of PSA in patients suffering from PCa has been reported, in comparison with healthy and diagnosed from benign conditions patients, allowing an accurate distinction between them (Figure 4).^{9,10,32,35–41}

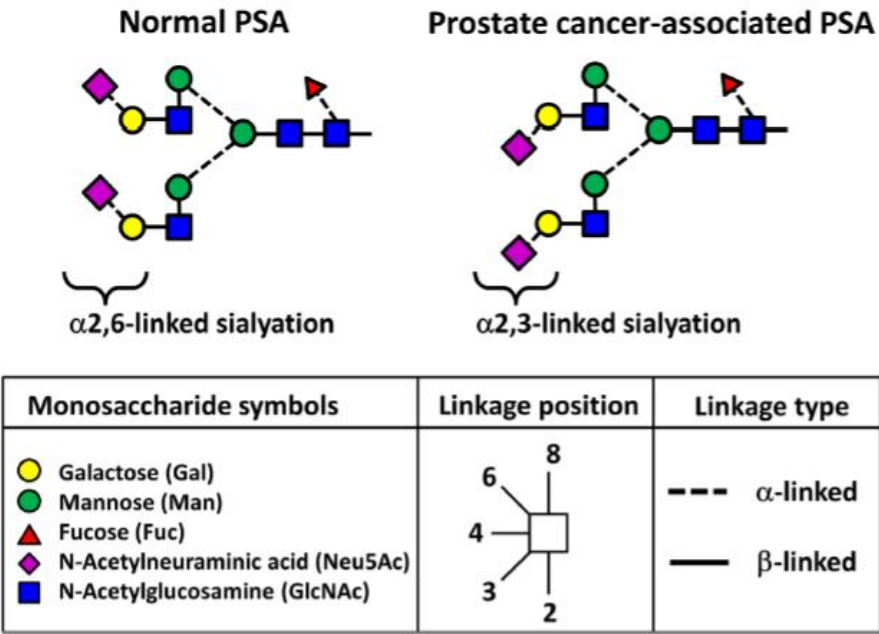


Figure 4: Anomalous glycosylation patterns as a result of PCa. Whereas PSA N-glycans in healthy patients mostly show their terminal sialic acid linked by $\alpha 2,6$ bond to a mannose residue, in patients suffering from aggressive PC it is found the $\alpha 2,3$ bond. Structures of both isomeric disaccharides can be found in Figure 2. Sialic acid (SA) and N-Acetylneuraminic acid (Neu5Ac) refer to the same molecule. Image reproduced from ref 42.

Glycans are thus an important area of study for researchers interested in understanding the biology of health and disease. And in some cases, the key to accurately identify a glycan and differentiate it from other closely structurally similar ones, relies on the determination of their constituent saccharides.

There is a growing understanding of the significance of saccharides in numerous biological processes and systems; while saccharide or carbohydrate sensing in medicine is frequently connected with the detection of glucose in diabetic patients, saccharides (on their own, or often as part of glycans) have shown to be important in a number of disease states.

Crucial roles in the initiation and development of cancer are played by saccharides, oligosaccharides, and their protein conjugates. For instance, aberrant N-glycosylation and post-translational O-glycosylation patterns, in which monosaccharides are covalently attached to certain proteins at asparagine, serine and threonine residues are frequently seen in metastatic tumours.⁴³

By studying these events not only we gain on understanding of cancer mechanisms but also, we can use it as a tool to monitor the disease.⁴⁴ As an example, Sialyl Lewis X is a cell-surface tetrasaccharide that, employing antibody microarrays, has been found to be overexpressed in cancerous tissues becoming a good indicator to evaluate the tumour malignancy.⁴⁵

As stated by others,⁴⁶ glycan biomarkers are proven to be more and more beneficial for disease stratification, staging, and monitoring in addition to early diagnosis. However, the

antigen-based technologies that are currently used to detect, recognise, and measure these intricate biological entities are costly, time consuming, or often, due to batch to batch variations, they fail to produce reliable results.⁴⁷ Thus, there is a significant demand for suitable chemosensors that can enhance current detection technologies for diseases such as cancer. The key to these new and improved glycan sensors could be the identification of the constituent saccharides.

1.2 Saccharides

Saccharides, also known as carbohydrates, are the most abundant biomolecules on Earth and are essential components of all living organisms. They have many functions in biological systems, including energy production, structural support, and cell signalling. Saccharides can be classified according to their size.

- **Monosaccharides:** Monosaccharides are simple sugars that cannot be hydrolysed into smaller units. The most commonly known monosaccharides are glucose, fructose, and galactose. Glucose is the primary fuel source for the human body and is used by cells to generate energy in the form of adenosine triphosphate (ATP) through the process of cellular respiration. Fructose and galactose are also used in the body for energy production.
- **Disaccharides:** Disaccharides are formed when two monosaccharides are joined together through a glycosidic bond. Examples of disaccharides include lactose, sucrose, and maltose. Lactose is found in milk and is made up of glucose and galactose. Sucrose (table sugar) is made up of glucose and fructose. Maltose is composed of two glucose molecules and is a product of the digestion of complex carbohydrates.

- Polysaccharides: Polysaccharides are complex carbohydrate molecules made up of many monosaccharide units. They are typically used for energy storage and structural support. Examples of polysaccharides include starch, glycogen, and cellulose. Starch is found in many plant-based foods and is broken down by enzymes in the body to produce glucose for energy. Glycogen is similar to starch but is found in animals, including humans, and is stored in the liver and muscles. Cellulose forms the structural components of plant cell walls and cannot be digested by human enzymes.

1.2.1 Complexity of monosaccharides

A simplification will be made and instead of studying glycans, their constituent saccharides will be the object of study within this thesis. More specifically, monosaccharides. These are structurally very complex molecules with a general formula of $(\text{CH}_2\text{O})_n$, where n can range from 3 to 7. Four of the most common monosaccharides are glucose, fructose, galactose, and mannose

They all possess the same molecular formula, $\text{C}_6\text{H}_{12}\text{O}_6$, but differ in the arrangements of their bonds which brings them different isomeric relationships with one another, and unique physical and chemical properties. Aldohexoses (sugars with a 6-Carbon backbone and a aldehyde group in C1 in its open form) have four stereocenters, and therefore there exists sixteen stereoisomers (Figure 5). Examples of aldohexoses are glucose, mannose, or galactose. D-Glucose and D-Mannose (definition of D and L sugars later on in this chapter section) are epimers on position C2, as they are enantiomers that differ only in the orientation of one hydroxyl group, in position C2. Likewise, D-Allose and D-Galactose are epimers of D-Glucose

in C3 and C4, respectively. Fructose is a ketohexose (sugar with a 6-carbon backbone and a ketone group at C2 in its open form), and its relationship of symmetry with the aldohexoses family is one of regioisomerism.

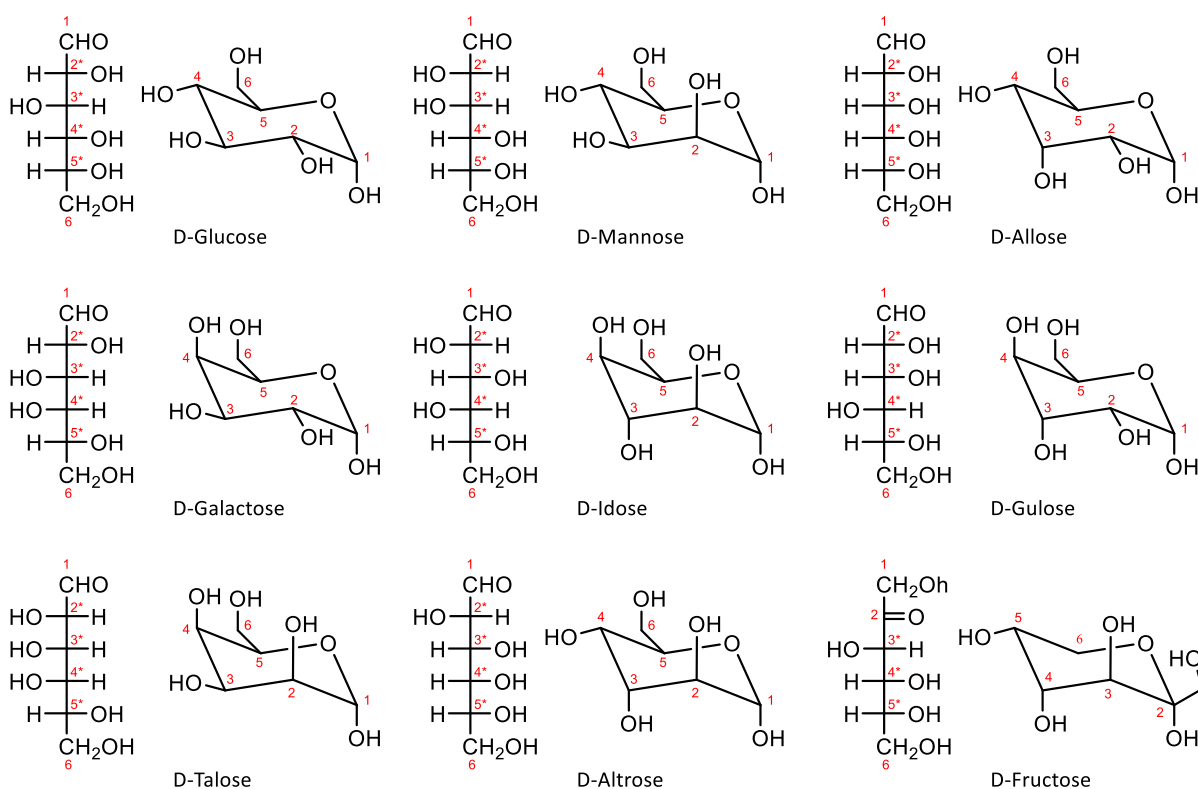


Figure 5: Linear (Fischer projections) and cyclic forms of aldohexoses (D-Glucose, D-Mannose, D-Allose, D-Galactose, D-Idose, D-Gulose, D-Talose and D-Altrose) and the ketohexose D-Fructose. Chiral carbons marked with a star () in the Fischer projections.*

The most common classification of sugars is the D and L nomenclature. This refers to the stereochemistry of the chiral carbon furthest from the carbonyl group in a monosaccharide. This carbon is usually the C5 (Figure 5). D- and L- configuration refers to the orientation of the hydroxyl group attached to this chiral carbon in a Fischer projection. If the hydroxyl group is

on the right side in a Fischer projection, the monosaccharide is classified as a D-sugar. If the hydroxyl group is on the left side, the monosaccharide is classified as an L-sugar.

It is important to note that D and L classification does not reflect the absolute configuration of the chiral carbon. Rather, it is a relative designation based on the orientation of the hydroxyl group. This means that if a monosaccharide is designated as a D-sugar, it does not necessarily mean that the absolute configuration of the chiral carbon is R. Similarly, an L-sugar does not necessarily have an absolute configuration of S.

Most naturally occurring monosaccharides are D-sugars. Although the reason for that is yet unclear, some researchers hypothesised that it could be due to evolutionary selection pressures, as enzymes that recognize D-sugars are more abundant and efficient than those that recognize L-sugars.⁴⁸ D-glucose, for example, is the primary source of energy for most organisms and is an essential component of many biological molecules such as DNA and RNA. L-glucose, on the other hand, is not metabolized by most organisms.⁴⁹ The D and L classification system is important in many areas of biochemistry and biology. For example, it is used to describe the stereochemistry of sugars in glycoproteins and glycolipids, or to distinguish between different types of sugars in carbohydrates with different biological functions and metabolic pathways.

However, the complexity of monosaccharides goes far beyond their relation of symmetry in their 6-membered cyclic forms. In aqueous solution, sugars suffer from mutarotation. This means that the hemiacetal carbon (C1 in aldohexoses, C2 in ketohexoses) opens up leading to a linear intermediate that can then close again causing interconversion between pyranose (6-

membered ring) and furanose (5-membered ring) forms with inversion of the configuration of the anomeric centre, affording alpha (α) and beta (β) tautomers (Figure 6).

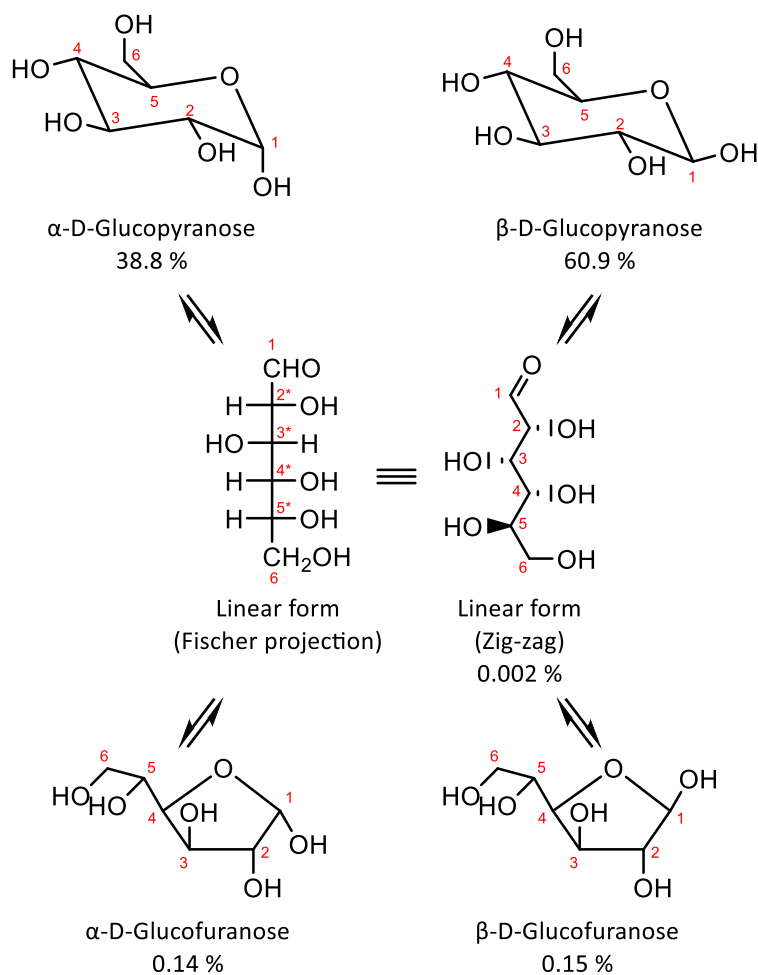


Figure 6: D-Glucose in its various tautomeric forms in solution, and the percentage of each form in equilibrium in D₂O at 27 °C. α -pyranose (38.8 %), β -pyranose (60.9 %), linear open chain form (0.002 %), β -furanose (0.15 %), and α -furanose (0.14 %).⁵⁰

1.3 Saccharide recognition

Like any other molecular recognition technique, saccharide recognition is based on the interaction between two entities: a receptor (or guest) and a substrate (or host). Selectivity and affinity between the host and the guest are requirements for a successful process. Hence, when artificial receptors are designed, aspects like geometric shape, functional group complementarity and electronic properties of the substrate must be taken in consideration.⁵⁰ Nature has spent millions of years developing tools to achieve molecular recognition of sugars mainly by the employment of lectins and antibodies. We as scientists have studied those processes and employ them to our benefit either using such biomolecules directly extracted from natural sources or implementing their mechanisms into synthetic receptors.

Since 1988 when the first synthetic saccharide receptor was designed (Figure 7a),⁵¹ a fascinating journey has taken place – and it is still taking place - to further develop this technique, usually by mimicking nature and/or with the help of computational studies. This led to modern and powerful saccharide recognition entities developed in recent years. These entities are often isolated from natural sources (Figure 7b),⁵² or designed and synthesised in the lab (Figure 7c).⁵³

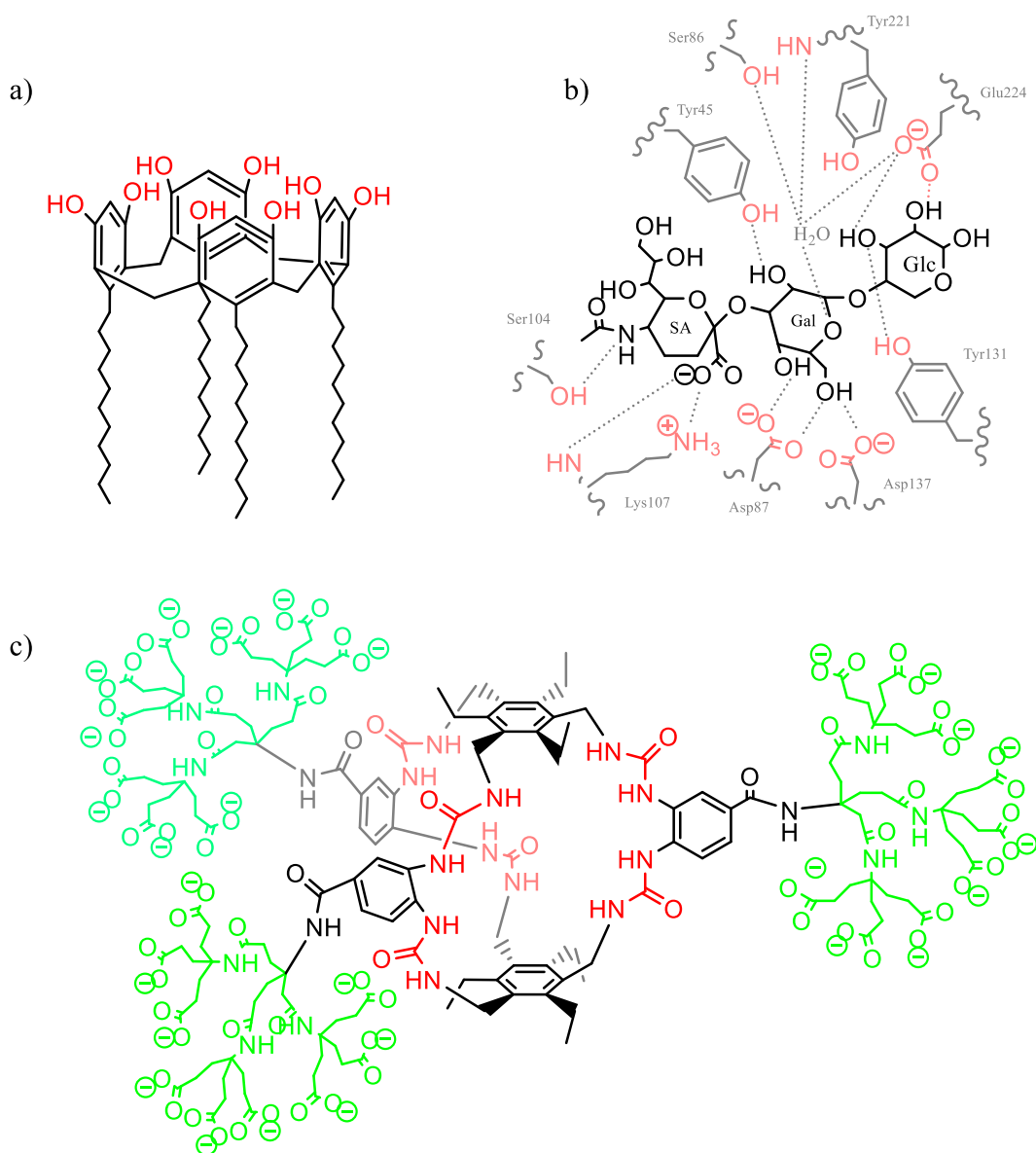


Figure 7: a) First saccharide receptor ever synthesised, with affinity for glucose. b) Model of interaction between *Maackia Amurensis* Lectin (MAL) with siallyllactose. c) in-lab desined lectin mimic receptor, with affinity for glucose. These examples will be further explained in section 1.5.1.2, page 30 (a and c) and section 1.4.1, page 18 (b).

The vast variability of saccharides makes their selective recognition and sensing a challenge, especially in the context of various biological settings and physiological environments, which can present a hurdle to achieve sensitivity.⁵⁴ Competition between different saccharides,⁵⁵ or

between solvent and receptor,⁵⁶ especially in water, can introduce entropic and enthalpic barriers toward successful receptor binding.^{57–60}

This thesis will detail the many biomolecules, tools and strategies that may be used to afford saccharide recognition.

1.4 Natural receptors

1.4.1 Lectins

Many living species, including plants, animals, and pathogens, express lectins. These are proteins or glycoproteins that bind carbohydrates.⁶¹ Lectins are involved in cell communication, the immune system of mammals and defence mechanisms in plants, and the invasion of cells by pathogens.⁶²

Lectins have been employed for the analysis and detection of glycans for a long time.⁶³ They bind carbohydrate residues via non-covalent interactions, especially hydrogen bonds and CH- π interactions between the sugars in the saccharide chain and the amino acids in the lectin binding pocket. Hydrogen bonds, highly directional, are key to acquire the correct fit between host and guest.⁶⁴

The carbohydrate's hydroxyl groups function simultaneously as a hydrogen bond donor and acceptor. Two lone pairs of electrons in the oxygen atoms can accept two hydrogen bonds, while the proton serves as a donor for one hydrogen bond.⁶⁵ This interaction affects all hydroxyl groups except for the anomeric one which is exclusively a donor. These cooperative effects produce strong hydrogen bonds.⁶⁴

Sugar chains may contain other groups that can participate in hydrogen bonding, such as the NH group and the carbonyl oxygen of the acetamide group in GlcNAc, GalNAc, or SA.⁶⁶ Moreover, the binding pockets in lectins are frequently rich in charged (such as histidine, arginine, or lysine) or bidentated amino acid residues (such as arginine, asparagine, or glutamine).⁶⁴ Whereas the latter produce a network of hydrogen bonds, the former generate charge-reinforced hydrogen bonds that are stronger than the neutral ones.^{64,66} In contrast, amino acids with side chains bearing hydroxyl groups (e.g. tyrosine, serine, threonine) are not frequently present in lectin binding pockets.⁶⁵ In the model of interaction of MAL with sialyllactose represented in Figure 7b it can be seen the polar groups responsible for the recognition event (red), as well as the H bonds they generate (dotted lines).

Hydrophobic interactions can also enrich the binding between lectins and saccharides. They are usually CH- π interactions taking place between aromatic amino acid residues (such as tryptophan, tyrosine, phenylalanine, and histidine) and the aliphatic C-H groups of sugars.⁶⁷ The tendency of the aromatic amino acids to create CH- π interactions also varies. The most propensity for CH- π bonds formation is shown by tryptophan, which possesses an aromatic system rich in electrons.⁶⁷

Tyrosine is preferred over phenylalanine, in forming CH- π interactions, due to the more electron-rich π system, whilst histidine rarely engages in these bonds.

On the other hand, amino acids with aliphatic side chains are less frequently found in binding pockets because they do not significantly interact with sugars.

Furthermore, electrostatic interactions can happen when the carbohydrate ligand is charged.⁶⁵ For instance, hydrogen bonds or electrostatic interactions are common with anionic groups, such as the carboxylate group of sialic acid. Hydrogen bonds are employed to bind the N-acetyl group, often by serine residues, whilst electrostatic interactions occur between the carboxylate and arginine or lysine side chains (Figure 7b).^{52,68}

Sialic acid-specific lectins are particularly interesting because of the importance of this monosaccharide in several diseases. These lectins are used as analytical and diagnostic instruments and come from either plant or animal sources.^{69,70} A few instances of SA-specific lectins will be addressed here, along with the specifics of their binding models. Various lectins that target sialic acid can show selectivity for a specific neuraminic acid derivative (such as SA, Neu5Gc, or deaminated neuraminic acid (KDN)), a specific glycosidic linkage (such as α 2-3, α 2-6), or for a particular carbohydrate sequence containing sialic acid (also known as neuraminic acid).⁶¹

The Limulin lectin and the Limax Flavus Agglutinin (LFA) are two examples of lectins that are specific to certain neuraminic acid derivatives. The former lectin is found in the American horseshoe crab (*Limulus polyphemus*), and it has selectivity for SA and N-glycolylneuraminic acid, two neuraminic acid derivatives. The Limulin lectin has been used in both optical and electron microscopy to identify sialylated glycans.⁶¹ The LFA, derived from the slug *Limax Flavus*, is selective for SA and binds any sialylated glycan regardless of the glycosidic linkage type.⁷¹ According to reports, LFA has a binding affinity of $K_a = 3.8 \times 10^4 \text{ M}^{-1}$ for SA, with the glycerol chain serving as a crucial point of contact.⁷¹

Moreover, several plant lectins, such as the Sambucus Nigra lectin (SNL), Maackia amurensis Leukoagglutinin (MAL, Figure 7b), and Wheat Germ Agglutinin (WGA), can distinguish between various glycosidic connections. The Wheat Germ Agglutinin (WGA) binds to SA residues when coupled 2-3 to lactose sequences.⁷² WGA mainly targets N-acetylglucosamine and β 1-4 GlcNAc-containing sequences, which are structurally similar.^{65,73} They both display comparable H bonds interactions between their N-acetyl and 4-OH groups, and the binding pocket of WGA.⁷³

There is also H bond interaction between the SA carboxylate group and a serine residue, as well as polar interactions among the glycerol chain and a tyrosine.⁷³

The MAL and MAH lectins, from the Maackia amurensis seeds, are selective for glycans with sialic acid residues connected via α 2–3 bond to galactose (K_a in the μ M range), whilst showing no binding to α 2–6 structures.⁷⁴ Whereas MAH interacts with disialylated epitopes in O-glycans, MAL specifically binds the sialylactosamine structures in N-glycans.⁷⁵ The Maackia amurensis lectins have been used to analyse the glycan phenotype in prostate cancer patients using lectin affinity chromatography and surface plasmon resonance.⁹ For the detection of α 2-3 sialylated glycans in biological samples, MAL has also been used in sandwich-type biosensors,⁷⁶ and in carbon nanotube-based electrochemical sensors.⁷⁷ Despite having a low limit of detection, lectin-based biosensors have disadvantages that limit their clinical use.

Their generally complex structures make their synthesis a challenge and often they suffer from batch-to-batch variations in their affinity to glycan targets.⁷⁸ Moreover, they can react off-site causing cross-reactivity.⁷⁹

On the other hand, the *Sambucus nigra* (SNA), a lectin derived from elderberry-bark, has preference for sialylated glycans with α 2-6 links to Gal or GalNAc units over the α 2-3 isomers.^{80,81} It binds SA- α 2-6-lactitol with $K_a = 3.9 \times 10^5 \text{ M}^{-1}$, with the sialic acid glycerol chain playing a key role in the interaction.⁸² Yet, SNA interacts not only with sialylated glycans but also with Gal or GalNAc ending glycans,⁸³ which accounts for the lectin's lack of specificity. SNA has been used in light and electron microscopy to analyse the glycan phenotype of cells,⁸³ and in lectin affinity chromatography to separate 2-6 sialylated glycans from the 2-3 isomers.⁸¹ It has also been used in biosensors for low-limit detection of glycans.^{84,85} Unfortunately, because to their difficult manufacture and poor stability, as well as the aforementioned common problems of cross-reactivity and batch-to-batch variations, they are not suitable for clinical use.

Human lectins that target sialic acid have been discovered too. Human lectins called siglecs (SA-binding Immunoglobulin-like Lectins) mediate cell signalling and communication.⁸⁶ Based on the preferred sialic acid derivative and glycosidic bond, they are separated into 11 groups. Siglec-1, also known as sialoadhesin, was demonstrated to bind the SA's α anomeric form with a $K_d = 3 \text{ mM}$.⁸⁷ Siglec-1 prefers α 2-3 epitopes over α 2-6 isomers. It binds them through Van der Waals and hydrophobic forces between the N-acetyl group and the glycerol chain with tryptophan residues.⁸⁷ Moreover, there are electrostatic interactions between the carboxylate group and an arginine residue as well as hydrogen bonds between the hydroxyl groups and polar amino acids.

Many lectins have been discovered and used as sialic acid detection probes; however, they pose a number of drawbacks that restrict their usefulness. Lectins suffer from batch-to-batch variations,⁷⁸ they must be stored at 4 °C and may only be kept for 6 to 12 months.⁶¹ Moreover, they only remain stable and active within a narrow pH range (pH 7–9).⁶¹ They need to be isolated from natural sources, often with very low yields.⁷² Furthermore, lectins only engage in non-covalent interactions with saccharides, resulting in poor binding and cross reactivity,⁷⁹ as they frequently target numerous glycan epitopes.⁸⁵ All of these make lectin-based biosensors unsuitable for clinical application.^{77,84}

1.4.2 Antibodies

Antibodies are a type of protein produced by the immune system that plays a vital role in protecting the body against infections and diseases. These proteins are specifically designed to recognize and bind foreign substances, such as viruses, bacteria, and other pathogens, and trigger an immune response to eliminate them from the body.

Antibodies are formed by two heavy chains and two light chains. The heavy and light chains are linked by disulfide bonds, forming a Y-shaped molecule with two arms that can bind to antigens. Each arm of the Y contains a variable region that is unique to the specific antibody, allowing it to recognize and bind to a particular antigen (Figure 8).

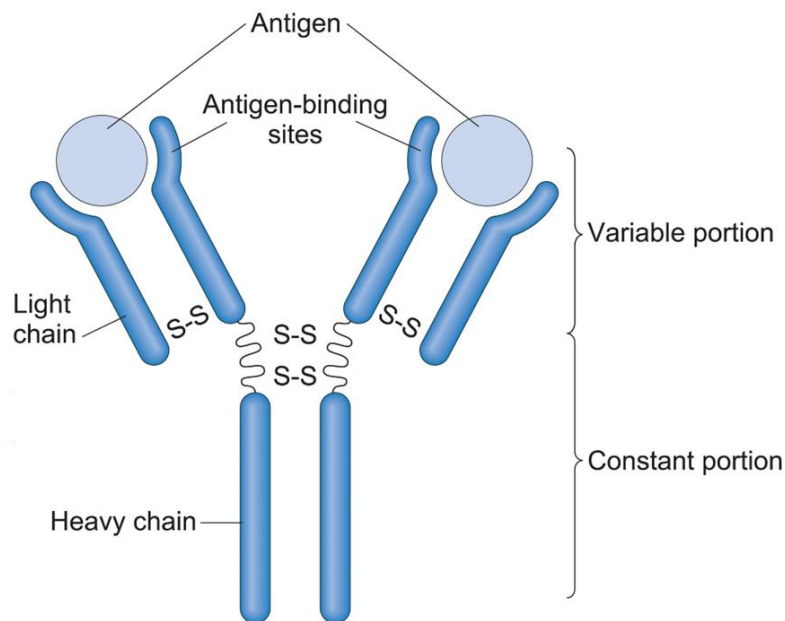


Figure 8: General structure of antibodies.

There are five main classes of antibodies based on the structure of their constant portion: IgG, IgM, IgA, IgD, and IgE. These classes differ in their structure, function, and distribution throughout the body.

Antiglycan antibodies (AGAs) are monoclonal antibodies, mainly IgM or IgG, and they target carbohydrates (glycans) present on the surface of cells or extracellular matrix molecules.⁸⁸ These antibodies are produced by the immune system in response to the presence of glycans that are recognized as foreign or altered in some way. They act by recognising glycan sequences or the entire glycosylated macromolecule.⁸⁸ A large collection of anti-glycan antibodies (AGAs) are present in human serum, playing critical functions in many immune processes. They can be used as biomarkers in the diagnosis of some conditions such as Crohn's disease,⁸⁹ or even cancer.^{33,90} In the case of cancer, although there are high expectations on their use for diagnosis, this AGAs cannot activate the immune system to eliminate malignant

cells, and therefore their use in immunotherapy against cancer is limited.^{91–93} Naturally occurring AGAs in human sera also detect foreign glycan epitopes, such as galactose- α -1,3-galactose and N-glycolylneuraminic acid (Neu5GC). These antigens are responsible for the rejection of xenotransplanted organs in humans from pigs.^{94,95}

Recent discoveries combining modern advances in carbohydrate chemistry with the latest knowledge in understanding immunological mechanisms afforded the creation of synthetic high affinity AGAs, that showed nanomolar affinity for their glycan targets.^{96,97} This is a developing field that is expected to bring successful results in the future.

Similar to lectins, antibodies bind glycans by non-covalent mechanisms. The hydroxyl groups of the carbohydrate chain and the polar amino acids of the binding pocket form hydrogen bonds. There are also hydrophobic interactions between the non-polar sugar face and aromatic residues like tyrosine and tryptophan.⁹⁸ Furthermore, positively charged functions are found in the binding sites of AGAs that target anionic glycans, enabling ionic interactions.⁹⁹ Some AGAs exhibit high affinity for naturally occurring glycans, although often they lack selectivity for a specific carbohydrate sequences, and may bind similarly closely related structures.¹⁰⁰ An exception to this would be the AGAs that are able to discriminate between the similar glycan structures of antigens A, B, and O in blood.¹⁰¹

Nevertheless, glycan-binding antibodies often have lower affinities than protein-specific antibodies, with equilibrium dissociation constant (K_d) values in the micromolar range for the

first and nanomolar for the latter.^{96,98} AGAs also have problems with availability and specificity since glycans have a low immunogenicity.⁸⁸

Because of this, there are few antiglycan antibodies available. The number of AGAs is only approximately 1000, but the number of glycan epitopes has been estimated to be over 7000.⁸⁸ Moreover, many of these antibodies share the same target, thus the number of different antigens for which an antibody exists is below 250. AGA frequently have low specificity since many of them recognise multiple structurally related glycans¹⁰² but in particular, the AGAs that target N-glycans have the least specificity and therefore there is no commercial availability.⁸⁸

1.5 Synthetic receptors

Learning from Nature and the principles behind molecular recognition in natural receptors, a great variety of synthetic receptors have been developed as recognition tools for saccharides and glycans. The advantage of synthetic receptors over lectins or antibodies is the total control over their structures allowing interaction with the biomolecules of interest through multiple and different functionalities. The formation of multiple interactions with the target, known as 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 an overall affinity receptor-target that is greater than the sum of the single interactions.⁵⁹

Lectins also benefit from this phenomenon, called the “cluster effect” to bind carbohydrates.⁵⁹ For instance, the binding affinity of lectins to monosaccharides is in the mM to μ M range,

whilst for oligosaccharides, which offer more sites for interactions, is up to nM.¹⁰³ These units of affinity refer to the value of dissociation constant, or K_d , between the receptor and the target and, the lower the value, the higher the affinity.

Synthetic receptors can be divided into receptors acting solely through non-covalent interactions and receptors able to form covalent bonds with the target.

1.5.1 Non-covalent receptors

As non-covalent bonding to accomplish saccharide recognition is the strategy employed by Nature,¹⁰⁴ a lot of work has been done in this field.^{105,106} These research studies are mainly focused on designing appropriate binding points, surfaces or even structures which the target could fit into, so the guest ended up completely encapsulated. These processes are mainly based on π - π , polar or hydrogen-bonding interactions.

1.5.1.1 Aptamers

Aptamers are short, single-stranded oligonucleotides that are capable of binding to target molecules via non-covalent interactions (mostly H bonds) with high specificity and affinity. These molecules are often compared to antibodies due to their ability to bind to specific targets,¹ but they have several advantages over antibodies, including their small size, ease of synthesis, and stability.¹⁰⁷

Aptamers are typically generated using an *in vitro* selection process known as systematic evolution of ligands by exponential enrichment (SELEX).¹⁰⁸ In this process, a library of random oligonucleotides is exposed to the target molecule of interest, and those oligonucleotides that

bind to the target are isolated and amplified. This process is repeated several times, with the library being modified between rounds to enrich for aptamers with higher binding affinities. After several rounds of selection, the aptamers with the highest binding affinities are identified and characterized (Figure 9).

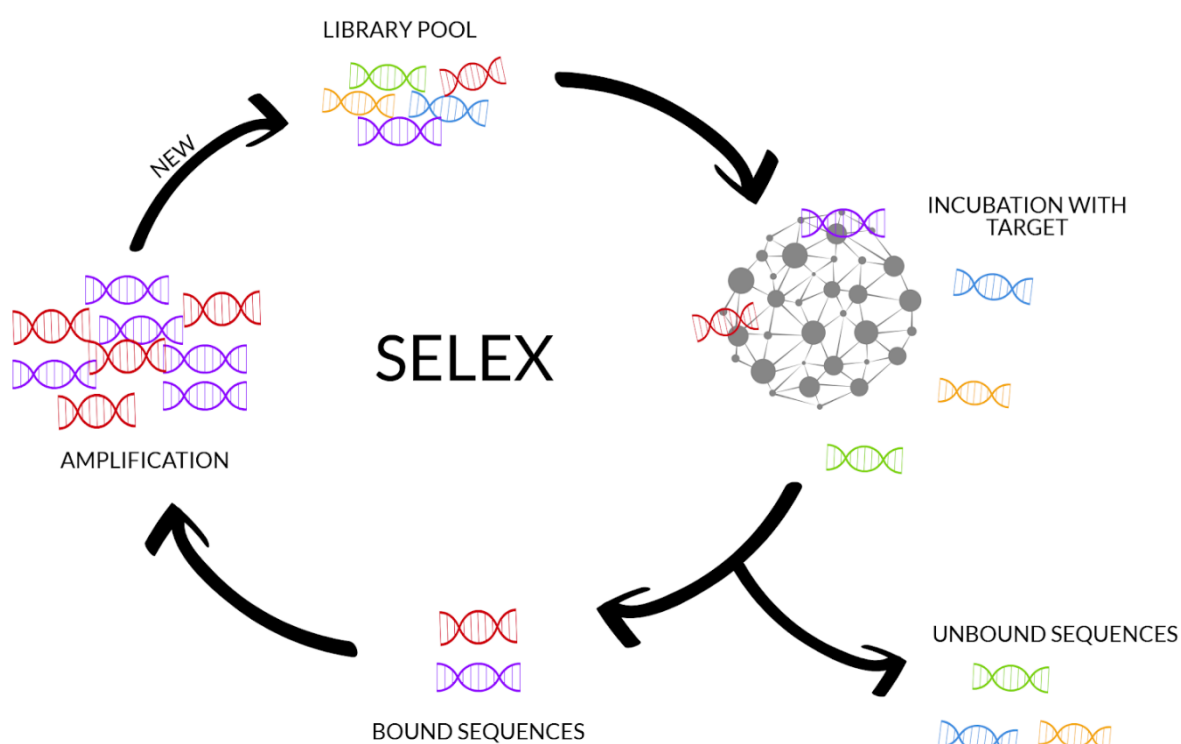


Figure 9: Diagram representing the SELEX process. Figure reproduced from reference 109.

One of the most significant advantages of aptamers is their ability to bind to a wide range of targets, including small molecules such as metal ions^{110–112} or amino acids,^{113,114} peptides,^{115,116} proteins,^{117,118} and even whole cells.^{119,120} This versatility makes aptamers useful in a variety of applications, including diagnostics, therapeutics, and biosensors.¹²¹

An advantage of aptamers is their ease of synthesis. Unlike antibodies, which require the use of animals for production, aptamers can be synthesized chemically, making them more accessible and cost-effective. Additionally, aptamers are generally more stable than antibodies and can be stored for longer periods of time without losing their binding properties.¹⁰⁷

Despite their many advantages, there are also some limitations to the use of aptamers. One of the most significant limitations is their susceptibility to nuclease degradation. Since aptamers are made of DNA or RNA, they can be degraded by nucleases in the bloodstream or other biological fluids, which can limit their efficacy as therapeutics. Additionally, in the particular case of glycan recognition, aptamers often exhibit low affinity and specificity, and therefore they can be susceptible to off-target binding, which can result in unwanted side effects.

The main shortcomings of aptamers are the low affinity and low specificity, causing cross reactivity with other glycan epitopes. This is a consequence of the lack of recognition tools that aptamers naturally offer. They have insufficient aromatic moieties or charged functionalities and thus can only interact through hydrogen bonding with carbohydrates. There are examples of RNA-like aptamers that bind SA in the low nM range, however they fail to differentiate α 2-3 and α 2-6 isomers or among different Lewis antigens.^{122,123} To date, no aptamers have been reported to recognise, let alone distinguish from each other, the most simple uncharged monosaccharide glucose and other hexoses.¹²⁴

To overcome these limitations, additional functional groups such as boronic acids or ionic groups have been added to aptamers to enhance their binding properties.^{125,126} Some examples are DNA-like aptamers modified to contain charged amino groups to target sialylactose, that resulted in a considerable enhancement of its binding constant (μM range) when compared with uncharged analogues.¹²⁵

1.5.1.2 Macrocycles

A lot of work has been carried out in the development of macrocyclic structures for the molecular recognition of saccharides. From the first saccharide receptor ever synthesised in 1988 that consisted in a fairly simple calixarene-type structure that binds glucose (with many limitations though, Figure 10a),⁵¹ to apparently not too complicated 3D cages developed as recent as in 2019 for which a pharmaceutical company invested hundreds of millions of pounds (Figure 10b).^{53,104}

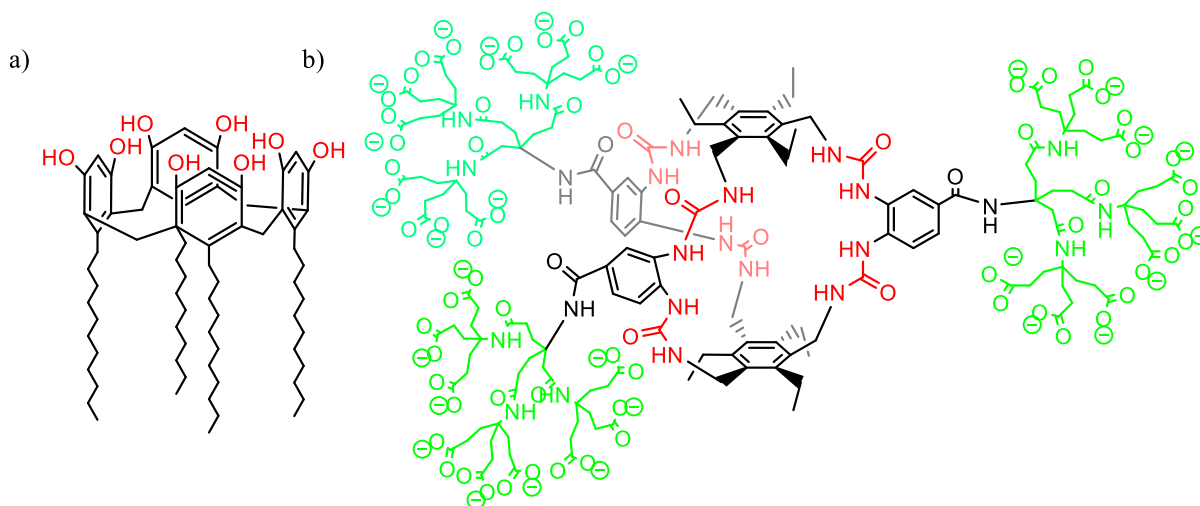


Figure 10: a) Aoyama's saccharide receptor with affinity for glucose. b) Davis' lectin mimic saccharide receptor, with affinity for glucose. Functional groups responsible for forming H bonds are coloured in red, while those in black are capable of creating hydrophobic interactions. Dendrimeric structures in green serve to solubilise the molecule in aqueous solvents.

In the first case, the receptor showed affinity for D-Glucose. The ring formed by the calixarene framework was not large enough to encapsulate the sugars, but it served as a rigid structure to permit 'face to face' recognition between hydroxyls groups of both receptor and substrate. NMR experiments showed that the affinity relied mainly on hydrogen bond interactions, which was evident when the affinity was lost upon the acetylation of receptors' OH groups. However, in order to bind saccharides, they required high concentrations of them. Furthermore, they failed to acquire selectivity for glucose over other saccharides, and since the receptor was not soluble in water and sugars normally are, they had to work with complex biphasic systems.

On the other hand, the modern lectin mimic receptor synthesised 30 years later could overcome most of these obstacles. First of all, peripheral nonacarboxylates (green) helped to maintain the molecule solubility in water. Second, the bicyclic cage formed by urea groups attached in -orto to a benzene ring, and the triethylmethylene units enriched the system with both polar hydrogen and hydrophobic π - π interactions. This allowed the receptor to efficiently and selectively bind the β -pyranose form of glucose, predominant one in water, with binding constant $K_a = 18 \times 10^3 \text{ M}^{-1}$, and 2 to 3 orders of magnitude lower for other saccharides such as D-galactose ($K_a = 180 \text{ M}^{-1}$), D-mannose ($K_a = 140 \text{ M}^{-1}$), D-ribose ($K_a = 220 \text{ M}^{-1}$), D-fructose ($K_a = 60 \text{ M}^{-1}$) or D-cellobiose ($K_a = 30 \text{ M}^{-1}$).⁵³ Nevertheless, when this system was tested in biological contexts, the affinity showed dropped. The presence of Ca^{2+} and Mg^{2+} ions caused a weaker binding, and hence, a different approach – perhaps covalent interactions - had to be considered.

Cyclic structures rely on the basic principles of saccharide recognition already outlined in this thesis in the section dedicated to the topic, but with a number of advantages.

Affinity: Macrocyclic structures can pre-organize functional groups in a way that is complementary to the target saccharide, allowing a tighter fitting towards the target molecule and therefore better binding interactions.

Selectivity: Macrocyclic structures can be designed to selectively bind to specific saccharide targets by incorporating binding sites that are optimized for the size and/or shape of particular saccharide structures.

Stability: Macrocyclic structures can provide enhanced stability for the saccharide complex, due to the constrained nature of the macrocycle, which can prevent unwanted conformational changes or degradation.

There are of course disadvantages associated with them and situations in which it may be more adequate working with non-cyclic receptors. They can be difficult and expensive to synthesise. They are often rigid, and therefore they may not be able to achieve effective binding towards targets that are flexible or have different conformations. But perhaps the most negative aspect of macrocyclic receptors that clearly marked their history is their generally poor solubility in water. This was obviously a major point of concern considering that most sugars are polar and therefore non-particularly soluble in organic solvents.

These molecules would rely on the basic principles of saccharide recognition, employing polar groups such as OH, NH, N, or even phosphate groups for H bonding as well as aromatic units

to promote π -CH interactions, in a pre-organised way to maximise the molecular groups complementarity.

Since Aoyama's first macrocyclic receptor (Figure 10a),⁵¹ many attempted to effectively bind saccharides with similar structures. In 1990 Davis' group reported **1** (Figure 11), a macrocyclic steroid-based structure that bound a β -glucoside (a lipophilic derivative of glucose) with a binding constant $K_a = 1.7 \times 10^3 \text{ M}^{-1}$.¹²⁷ Soon after, Diederich's group would develop **2** (Figure 8), a far more complex macrocyclic structure with 10 times more affinity for the same glycoside.¹²⁸ Davis again in 1998 proposed **3** (Figure 8), a more sophisticated tricycle that would take into account the all-equatorial stereochemistry of the saccharide of interest (still the same β -glucoside) to generate a more fitting geometry in its cavity, resulting in a much higher affinity for the saccharide with $K_a = 3 \times 10^5 \text{ M}^{-1}$.¹²⁹

In 2016 Abe and Inouye set a new record with their semiplanar crown-like receptor **4** (Figure 11) that showed an affinity for glycosides in the order of 10^6 M^{-1} .¹³⁰

It is also worth highlighting the bicyclic cages designed by Roelens, **5** (Figure 11), that bound the β -glucoside with $K_a = 5 \times 10^4 \text{ M}^{-1}$, but with selectivity over the α anomer,¹³¹ or the series of bicyclic cryptophanes **6** (Figure 11) developed by Martinez and Dutasta showing not only affinity for saccharides but also tunable selectivities.¹³²

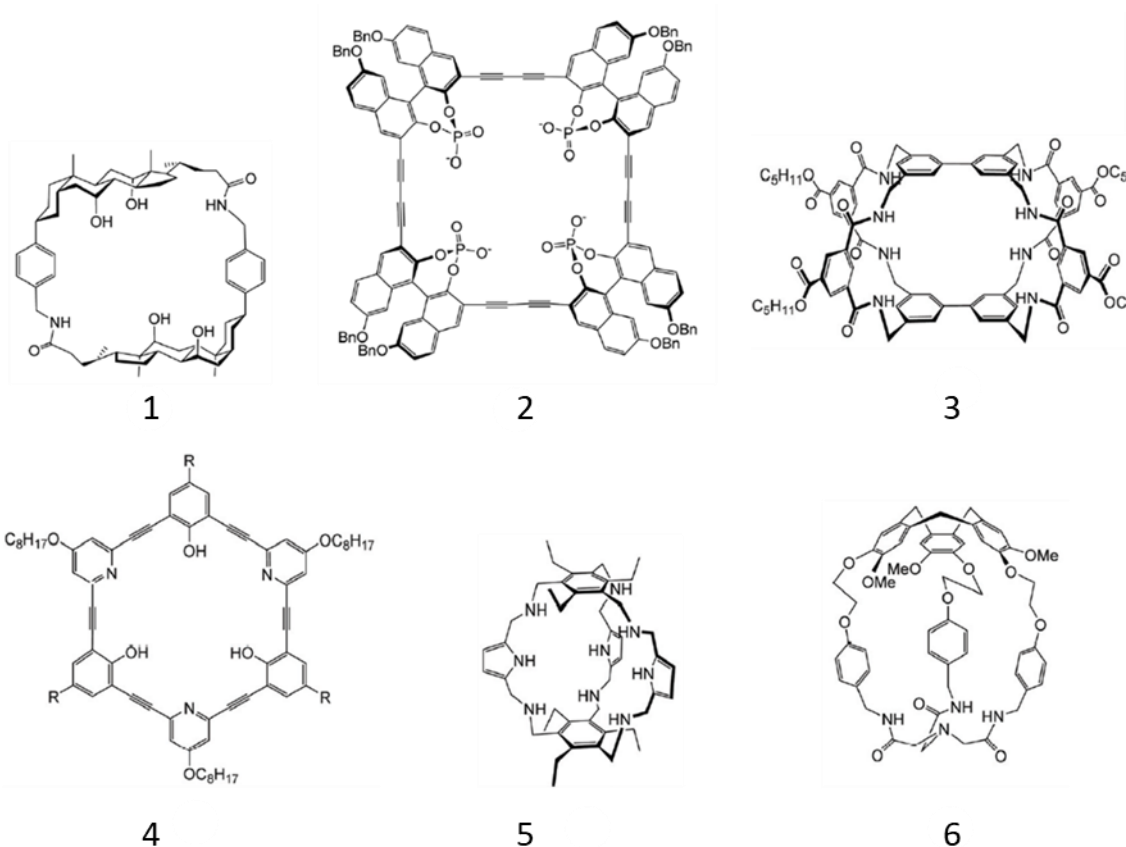


Figure 11: Series of macrocyclic receptors for saccharides developed by different research groups over the last years. 1: Davis; steroid-based structure. 2: Diederich's planar macrocycle. 3: Davis' three-dimensional receptor. 4: Abe's semiplanar cyclic structure. 5: Roelens' bicyclic cage. 6: Martinez and Dutasta's cryptophanes.

It was not until 2005 that Davis reported the first receptor that could bind saccharides in water environments.¹³³ It was an evolution of its own molecule created seven years before, now incorporating four dendrimer-like tricarboxylate groups as solubilising agents. A proof of the challenge of binding saccharides in water environments is that the affinity dropped from $3 \times 10^5 \text{ M}^{-1}$ for their previous design, in organic solvent, to just 9 M^{-1} this time in water.

This line of research continued over the years with new macrocyclic receptors offering better affinities and selectivities for different monosaccharides,^{134–139} and even disaccharides,^{140,141} until 2019 when Davis et. al. once again designed a C3 symmetric hexa-urea cage (Figure 10b) that could encapsulate monosaccharides in its interior, with excellent functional group complementarity, as an evolution of their previously reported saccharide receptors. This time, they achieved a binding affinity for glucose in water of $K_a = 18 \times 10^3 \text{ M}^{-1}$, which is higher than most lectin-monosaccharide interactions,¹⁴² and representing a 70-fold increase over previous water-friendly receptors, and a remarkable selectivity over other common sugars such as galactose, mannose, ribose, fructose and cellobiose. Other all-equatorial substrates were found to have a notorious affinity for the receptor, but always with at least 25-fold decrease in comparison with glucose.⁵³

1.5.1.3 Small-molecule acyclic receptors

Small-molecule receptors, of course, rely as well on the same basic principles of molecular recognition. This subtype of receptors has also experienced major improvements over the years although perhaps they have been less explored than cyclic structures.

These molecules are in general easier and less expensive to make. Open structures offer more flexibility than their cyclic analogues. They can accommodate saccharides of different sizes and shapes and therefore they are more versatile, although this may play against their ability to offer selectivity for a particular target.

In general, small molecules are easier to work with and to design and synthesise to be soluble in water. However, the enormous challenges of saccharide recognition in aqueous

environments forced the development of these receptors, same as for cyclic receptors, to work mostly in organic solvents.

One popular strategy in the design of these receptors is to employ a central aromatic platform with potential for non-polar interactions, with polar 'arms' offering multivalent connectivity via electrostatic or H bonds. This open structure with no roof should be compatible with saccharides bearing axial substituents.¹⁰⁴

Figure 12 shows the structure of some small acyclic receptors for saccharides developed over the years. Receptor **7** (Figure 12) demonstrated that simple receptors can achieve great affinities. It showed binding for β -glucosides as high as $26 \times 10^3 \text{ M}^{-1}$, in organic solvent.¹⁴³ Other examples relying on the same principle but with a more complex structure are the bis-phenanthroline **8** (Figure 12), that bound the same substrate with $K_a = 10^5 \text{ M}^{-1}$,¹⁴⁴ or the hexamine **9** (Figure 12) with affinity for octyl α -D-mannoside of $K_a = 10^4 \text{ M}^{-1}$,¹⁴⁵ both in organic solvent.

As per receptors that can bind saccharides in water environments, receptor **10** (Figure 12) bound the important oligosaccharide Sialyl Lewis X with $K_a = 100 \text{ M}^{-1}$.¹⁴⁶ However, this complex substrate containing polar functionalities and a larger skeleton is probably less difficult to target than simple monosaccharides. The benzene derivative **11** (Figure 12) exhibited affinity for the disaccharide cellobiose with $K_a = 305 \text{ M}^{-1}$.¹⁴⁷ Receptor **12** (Figure 12), with certain resemblance to anthracene **10** (Figure 12), was found to bind quite strongly the biomedically relevant target heparin, and therefore has potential for biosensing applications in medicine.

Another example of small-molecule receptor for heparin is the molecule **13** (Figure 12). This had a slightly different design as a non-aromatic linear molecule as spermine served as scaffold, while the aromatic units were in the 'arms'.^{148,149} The most interesting aspect of **13** is that it was 'self-designed' via Dynamic Combinatorial Chemistry (DCC). DCC for the recognition of biomolecules will be discussed in depth in a separate section of this thesis.

Some last examples of this category are the pyrenes **14** and **15** (Figure 12), designed to target substrates with charged axial groups. Anionic **14** bound mannosamine with $K_a = 3 \times 10^3 \text{ M}^{-1}$, and cationic **15** bound sialoside with $K_a = 1.3 \times 10^3 \text{ M}^{-1}$.¹⁵⁰

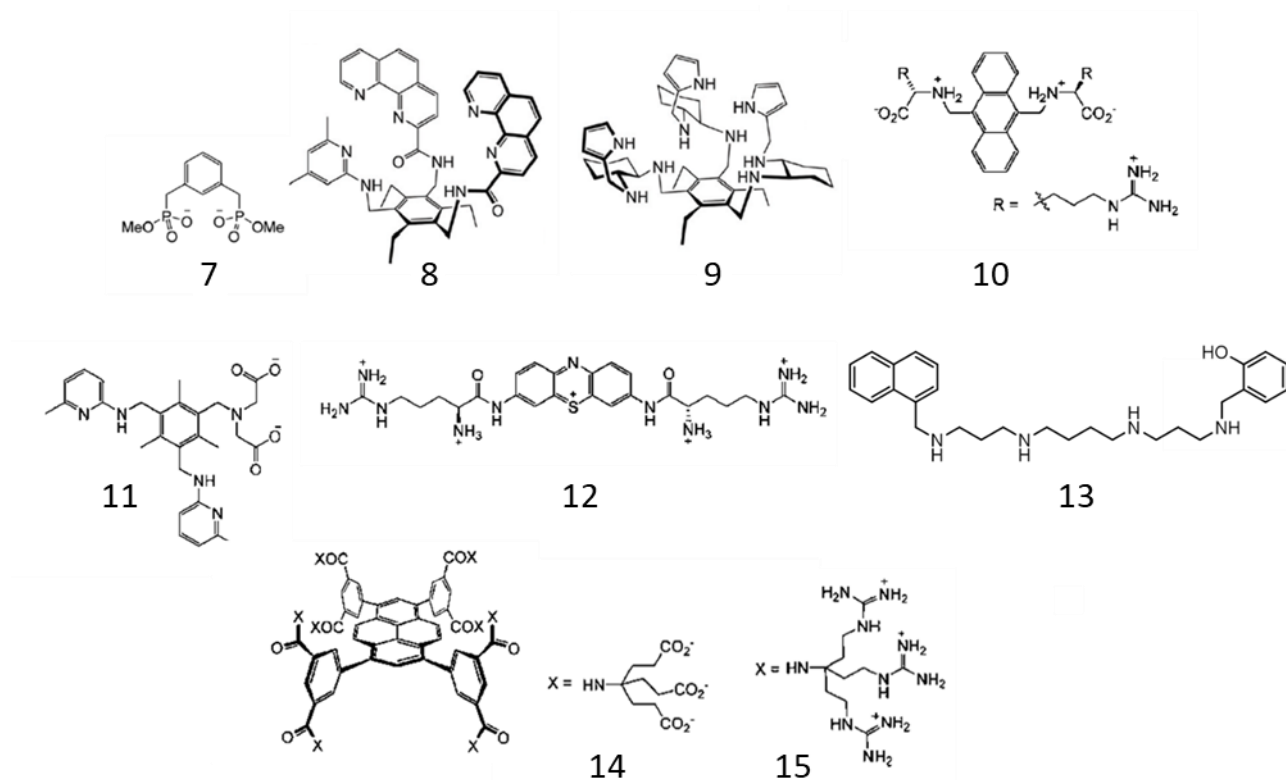


Figure 12: Series of small acyclic receptors for saccharides developed by different research groups over the last years. 7: Hamilton's phosphate derivate. 8: Mazik's bis-phenanthroline. 9: Liekens' aminopyrrolic receptor. 10: Grundberg's anthracene-amino acid. 11: Cavga's receptor. 12: Grundberg's second anthracene-amino acid. 13: Alfonso's spermine-based receptor. 14 and 15: Davis' pyrene-based ionic receptors.

1.5.2 Covalent receptors. Boronic Acids

It has been demonstrated with numerous examples the challenges arising from attempting molecular recognition of saccharides, especially in aqueous environments where these are heavily solvated. In these circumstances, receptors have to compete with molecules of water for the binding sites, and non-covalent host guest interactions are often not strong enough to achieve binding affinity.¹⁵¹ While working in organic solvents is an alternative that yielded positive results in the past, saccharides are often insoluble in organic media and therefore O-alkylated sugar derivatives are needed.^{105,106} This, unfortunately, reduces enormously the usability of such receptors for real-life applications.

To overcome this handicap, receptors with the ability to covalently bind saccharides are a good alternative. Covalent bonds are generally stronger than non-covalent ones, and therefore they can better displace molecules of water from the binding sites affording higher binding affinities. In this context, receptors containing boronic acids (BAs) apply as appropriate candidates offering the possibility to covalently bind sugars in aqueous environments. Boronic acids are Lewis acids employed in carbohydrate recognition due to the ability of their boron atom to form stable cyclic 5 and 6 membered boronate esters with cis-1,2 and 1,3-diols, which are present in saccharides (Figure 13).¹⁵²

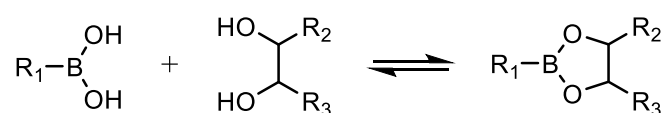


Figure 13: Generic boronate ester formation reaction between a generic BA and a 1,2-diol.

The formation of the boronate ester with ligands is pH dependent. The optimal pH for binding is determined by the pK_a of both the BA and the target as defined by Equation 1.1

.¹⁵³

$$\text{Optimal pH} = \frac{pK_{a_{BA}} + pK_{a_{target}}}{2} \quad \text{Equation 1.1}$$

The most commonly employed BA is the Phenylboronic acid (PBA). This aromatic BA has a pK_a of 8.86,¹⁵⁴ and it is known to bind sugars at alkaline pHs.⁵⁹ According to Equation 1.1

, for PBA to bind Fructose ($pK_a = 12.1$), the theoretical optimal pH would be 10.5.¹⁵⁵

The reason why PBA binds sugars at basic pHs is that the boron center electronic structure changes upon the pH. At pH values below the pK_a the boron centre is in the neutral and trigonal form (1 in Figure 14) whilst, at pH values above the pK_a the boron is negatively charged in the tetrahedral form (2 in Figure 14). Therefore, the binding of boronic acid to saccharides at basic pH occurs through the anionic tetrahedral boronate, to give a negatively charged boronate ester (4 in Figure 14).¹⁵²

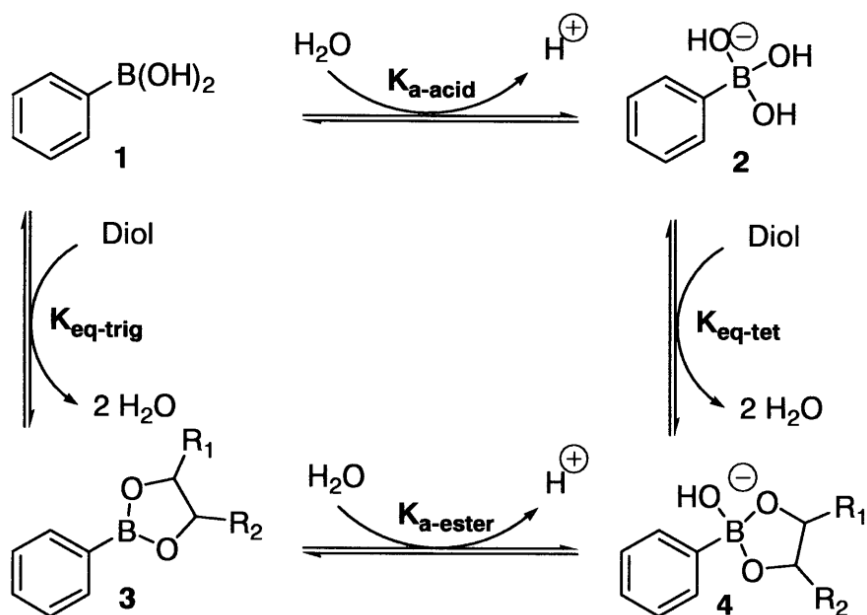


Figure 14: Phenylboronic acid (PBA) in its neutral trigonal form (1) and anionic tetrahedral form (2), as well as the relationships between their diol esters at pH below PBA's pK_a (1 to 3) and pH above PBA's pK_a (2 to 4).¹⁵⁵

At pHs below its pK_a , the trigonal form of BA can also bind diols. However, the resulting boronate ester (3 in Figure 14) is less stable and gets hydrolysed readily. In other words, $K_{eq-tet} > K_{eq-trig}$. As a result, neutral carbohydrates (i.e. not charged) are effectively bound by PBA (or BA in general) only at pHs above the BA's pK_a .⁵⁹ Whilst the majority of carbohydrates generally have a high pK_a (and therefore can only be bound by PA at alkaline pHs), there are a few exceptions. The dye Alizarin Red S (ARS) ($pK_a = 5.5$) and sialic acid ($pK_a = 2.6$) can be bound by the trigonal form of the BA at neutral and acidic pH, respectively.^{155,156}

However, these are normally exceptions and since the binding affinity of BA to neutral monosaccharides is significantly diminished at physiological pH (7.4), BA receptors are only partially useful in the study of biological fluid.¹⁵⁵ At pH 7.4, PBA has the greatest affinity for ARS and catechol ($K_a > 800 \text{ M}^{-1}$); and it decreases significantly for carbohydrates with the

following trend; sorbitol, fructose, mannitol ($K_a > 100 \text{ M}^{-1}$) > arabinose, sialic acid, glucuronic acid, galactose, mannose ($K_a > 10 \text{ M}^{-1}$) > glucose, lactose ($< 10 \text{ M}^{-1}$).¹⁵⁵ For instance, at pH 8.5 PBA binds fructose with $K_a = 560 \text{ M}^{-1}$, but it drops to 160 M^{-1} at pH 7.4.¹⁵⁵

A solution to this problem would be increasing the acidity of the BA derivative. In the case of PBA (Figure 15a), this can be done by making the boron atom electron deficient via the addition of electron-withdrawing groups in the aromatic ring.¹⁵⁷ The introduction of a nitro group in meta with respect to the BA in PBA results in a drop in the pK_a from 8.86 from PBA¹⁵⁴ to 6.9 in m-nitrophenylboronic acid (Figure 15b).¹⁵⁷ In response to this, the affinity of the nitro derivative for fructose is $K_a = 1350 \text{ M}^{-1}$, which means an 8-fold enhancement in comparison with the affinity exhibited by PBA.

Another strategy to increase the acidity of the boron atom in PBA is the introduction of a tertiary amine in ortho position.¹⁵⁸ These structures are called Wulff-type boronic acids and the reduction in pK_a is believed to occur due to the increment in stability of the boronate ester formed when binding saccharides. In aprotic solvents, this stability comes from the formation of a B-N intramolecular bond between them whilst ion pairing forces are the responsible for the effect in protic solvents.¹⁵⁹ The Wulff-type BA 2-(dimethylaminomethyl)phenylboronic acid (Figure 15c), with pK_a of 5.3, can be employed at acidic pH for the detection of neutral monosaccharides.¹⁶⁰

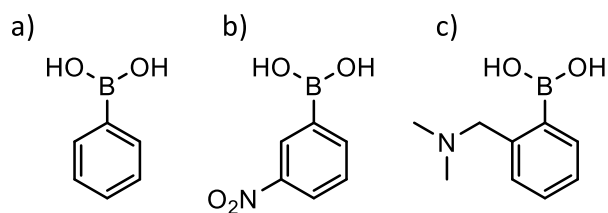


Figure 15: Structures of a) Phenylboronic acid (PBA), b) m-Nitrophenylboronic acids, and c) o-(dimethylaminomethyl)phenylboronic acid.

In summary, the detection of monosaccharides has made considerable use of boronic acid receptors, whose binding affinities change depending on the selected ligand. For instance, PBA has the maximum affinity for fructose at pH 8.5 ($K_a = 560 \text{ M}^{-1}$) while galactose ($K_a = 80 \text{ M}^{-1}$) and glucose ($K_a = 11 \text{ M}^{-1}$) both have much lower affinities.¹⁵⁵ At physiological pH, the affinity values decrease but the trend remains.

As mentioned previously in this chapter (Figure 6), perhaps the major challenge for the molecular recognition of saccharides is the existence of different tautomeric forms in equilibrium in solution, and this is also a challenge when employing boronic acids. BA preferentially bind the furanose forms of saccharides.^{161,162} In particular, the stability of the boronate ester (and therefore the affinity BA-ligand) is greatest for the endocyclic *vic*-diol furanose form. It is up to 30 times lower for exocyclic-1,2-diol pyranose, an even lower for other carbohydrate diols, in the order exocyclic-1,2-diol furanose > *cis-vic*-diol pyranose > exocyclic 4,6-diol pyranose (Figure 16).¹⁵³

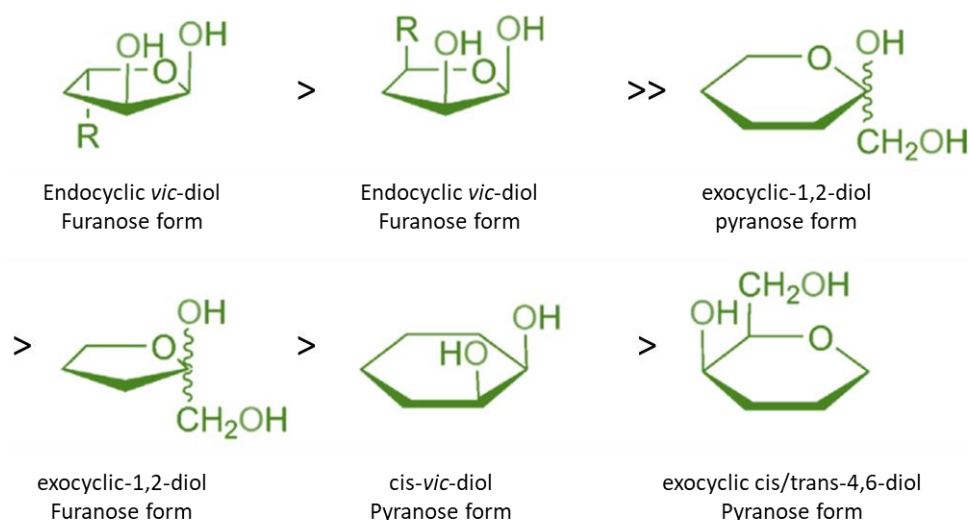


Figure 16: Different tautomeric forms of monosaccharides, in order from highest to lowest affinity for BA receptors.

While mono, di, and oligosaccharides produced by furanose units can be recognised by BA receptors, the use of boronic acids in the identification of glycans displaying just (or mostly) pyranose sugar residues is constrained by their preferred affinity to the furanose form.

Yoon et al. created the first boronic acid receptor for carbohydrate recognition, which was an anthracene-based boronic acid (Figure 17a) used for monosaccharide fluorescence studies.¹⁶³ Since then, several functionalized boronic acids with one or more boronic acid units and polar or apolar functional groups have been created. By enabling different points of contact with the ligand (multivalency), the affinity and selectivity of the probe is enhanced. For instance, by adding two correctly spaced boronic acid units to the receptor design, Tsukagoshi et al. created the first receptor selective for glucose (Figure 17b)¹⁶⁴. Fifteen years later Fossey et. al. studied the rational design of synthetic receptors. They synthesised probes similar to Tsukagoshi's, but with a linker of variable length between the BA groups to study the effect of

the distance between recognition groups in the affinity and selectivity shown for carbohydrates.¹⁶⁵ While all the di-BA tested (Figure 17d) outperformed the mono-BA derivative (Figure 17c), proving the importance of multivalency, they also demonstrated a strong distance-dependence with affinities almost 10 times apart from the worst ($n=3$, $K_a = 103 \text{ M}^{-1}$) to the best one ($n=6$, $K_a = 962 \text{ M}^{-1}$) observed, employing glucose as ligand.

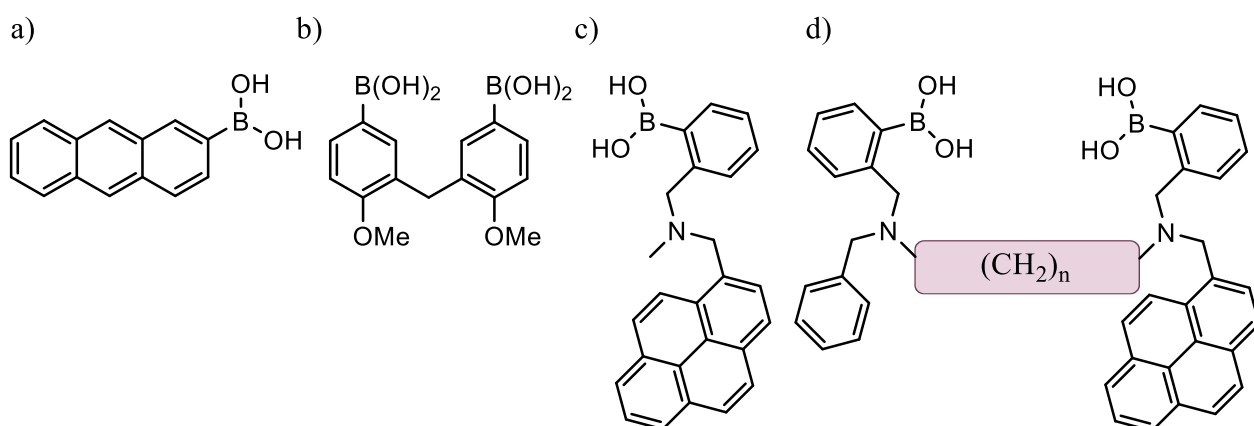


Figure 17: A series of BA derivatives developed for carbohydrate recognition that exemplifies not only the benefits of two BA groups over just one (multivalency), but also the importance of the distance between those groups.

Another interesting aspect when comparing mono and di-BA is that, while PBA (mono-BA) binds preferentially fructose over glucose, this selectivity is swapped over in di-BAs. This is due to the fact that glucose poses 2 binding sites which can be simultaneously accessed by di-boronic acids, while fructose only has one.⁵⁹ Di-boronic acids have also been employed as receptors for oligosaccharides. An example would be the one developed for the sialyl Lewis epitope.¹⁶⁶

Other than increasing the number of BA units in the receptor, different strategies have been followed to afford multivalency. Polar and non-polar groups, in a charged or neutral form, has been used for the detections of monosaccharides. Or in other words, covalent interactions can be complemented by non-covalent forces to better tailor the needs of each particular ligand thus affording more powerful receptors. As examples, the urea and thiourea (Figure 18a) modified boronic acids cooperatively bind sialic acid by hydrogen bonding and boronate esterification.¹⁶⁷ PBA receptors are frequently modified with charged moieties to recognise negatively charged saccharides. PBA receptors containing the guanidino group have been created to detect D-glucarate and SA (Figure 18b).^{168,169} In addition to the boronate ester, the guanidino group interacts electrostatically with the carboxylate group of glucarate, causing a two-point contact and a fivefold increase in affinity. Due to a similar cooperative effect, heparin, an anionic polysaccharide, is bound by triboronic acid receptors functionalized with amino groups showing binding constants up to 10^8 M^{-1} (Figure 18c).¹⁷⁰

Additionally, 5-indolylboronic acid (Figure 18d), which resembles the tryptophan residues in lectins, binds preferentially glucose residues by engaging in hydrophobic and CH- π interactions with oligosaccharides.¹⁷¹

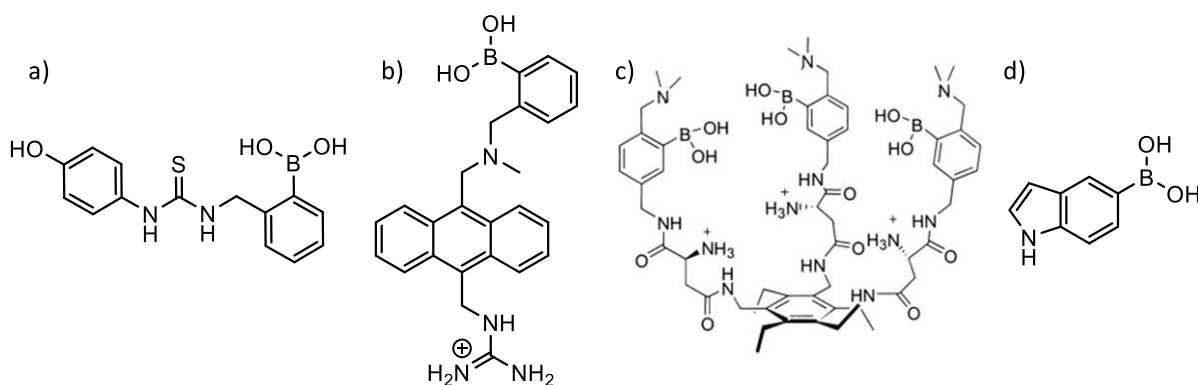


Figure 18: Boronic acid derivatives designed for saccharide recognition posing different chemical functions besides the BA: a) Thiourea group, b) guanidino group, c) three BA + guanidino group, and d) indol group.

The boronic acid group can also be introduced in peptides to enhance the natural lectin-saccharide interactions. Charged amino acid residues like aspartic and lysine, polar residues like tyrosine, and aromatic residues like phenylalanine and tryptophan are frequently found in this peptidomimetic compounds.^{172–174}

Different amino acid sequences produce peptides with various selectivities and affinities for carbohydrates.¹⁷⁴ In order to improve affinity through the creation of covalent compounds, boronic acid may also be added to nucleic acids and polymers.¹⁷⁵ For example, when comparing the affinity of aptamers containing and lacking BA towards glycosylated proteins, it was shown a clear enhancement for the first ones.¹²⁶

PBA-functionalized hydrogels have proven to detect glucose at mM levels,¹⁷⁶ and micelles containing PBA have been used to detect cells with high sialylation levels as a novel way for the administration of anticancer drugs.¹⁷⁷ Another novel applicability of BA for detecting carbohydrates is their use in modified surfaces or nanoparticles. There are examples of modified golden surfaces with the ability to detect glucose,¹⁷⁸ sialic acid,¹⁷⁹ and

glycoproteins.^{180,181} Boronic acids have also been used in molecularly imprinted surfaces, where the binding pocket of the ligand of interest is formed on the surface, offering great specificity and selectivity for such ligand. With a K_d value as low as 1.8 μM ,^{182,183} it was shown that molecularly imprinted BA-functionalized gold surfaces showed high affinity and selectivity for glycosylated PSA over other glycosylated and non-glycosylated proteins. In addition, mono- or oligosaccharides can be imprinted on nanoparticles containing boronic acid binding sites, with reported binding affinities in the low mM range.¹⁸⁴

Their preferential binding to furanose tautomers is a major side back restricting the use of BA as receptors for, for instance, cancer related glycans which often exhibit a major presence of sugars in their pyranose form. On top of that, the limited pH range in which BA often can effectively bind saccharides limits their applicability as probes in biological fluid analysis. Although the use of Wulff-type and electro-deficient boronic acids helps to partially resolve this issue, many of these BA derivatives are not very soluble in water, hence co-solvents are frequently necessary. In summary, while BA derivatives have their limitations in the field of saccharide recognition, they are present in numerous receptors that proved to bind sugars successfully and they have proven to be a group worth studying in more detail.

As it has been demonstrated, there are numerous factors involved in the design of molecules that bind sugars. Functional groups to incorporate, the distance and geometry between them, or the best intermolecular forces to consider, are still unclear. There is only consensus regarding the importance of multivalency. In this context, Dynamic Combinatorial Chemistry (DCC) poses as an excellent approach for the design of powerful and selective saccharide receptors.

1.6 DCC

Dynamic Combinatorial Chemistry (DCC) is a powerful approach used for designing complex and diverse chemical systems. DCC involves the creation of a mixture of reactants (Dynamic Combinatorial Library, DCL) that can react with one another via a reversible reaction under thermodynamic control, which eventually reaches the equilibrium state. This enables the system to respond to an external stimulus by shifting its equilibrium composition.¹⁸⁵ The products generated from these reactions are a function of the reaction conditions and the composition of the starting mixture. DCC is a versatile method that can be used to generate complex systems such as foldamers^{186–188} or supramolecular assemblies^{189–191} as receptor molecules. The method is particularly useful for creating receptors for biological molecules, such as saccharides,^{188,192} peptides,¹⁹³ and proteins.^{194–200} In this context, DCC allows for the creation of dynamic libraries of receptors that are capable of binding to specific biomolecules with high selectivity and affinity.

Therefore, DCC can be used as a molecular recognition technique to synthesise and detect the best binders for a given target. Its potential relies on the reversible reactions among members of the DCL and the ability for some of these members or building blocks (BB) to interact via covalent or non-covalent bonds with a potential target.²⁰¹ Once a member of the DCL (a ligand, L) binds the target (T) and forms a different species (complex L-T), the equilibrium will shift towards the formation of more and more of the missing ligand at the expense of weak binders. Therefore, if a screening of the library is performed before and after the addition of the target, the library member that bound the target will normally appear amplified, or in other words, in a higher presence in the screening of the library containing the template. Alternatively, two

separate DCLs can be formed in parallel, and the target is added only to one of them while the other is used as a blank (Figure 19). It is worth mentioning that in some particular cases, the concentration of the best binder within the DCL will actually be reduced in the presence of the template. This is the case when the target is, for instance, an active enzyme that destroys the high affinity ligands.^{202–204}

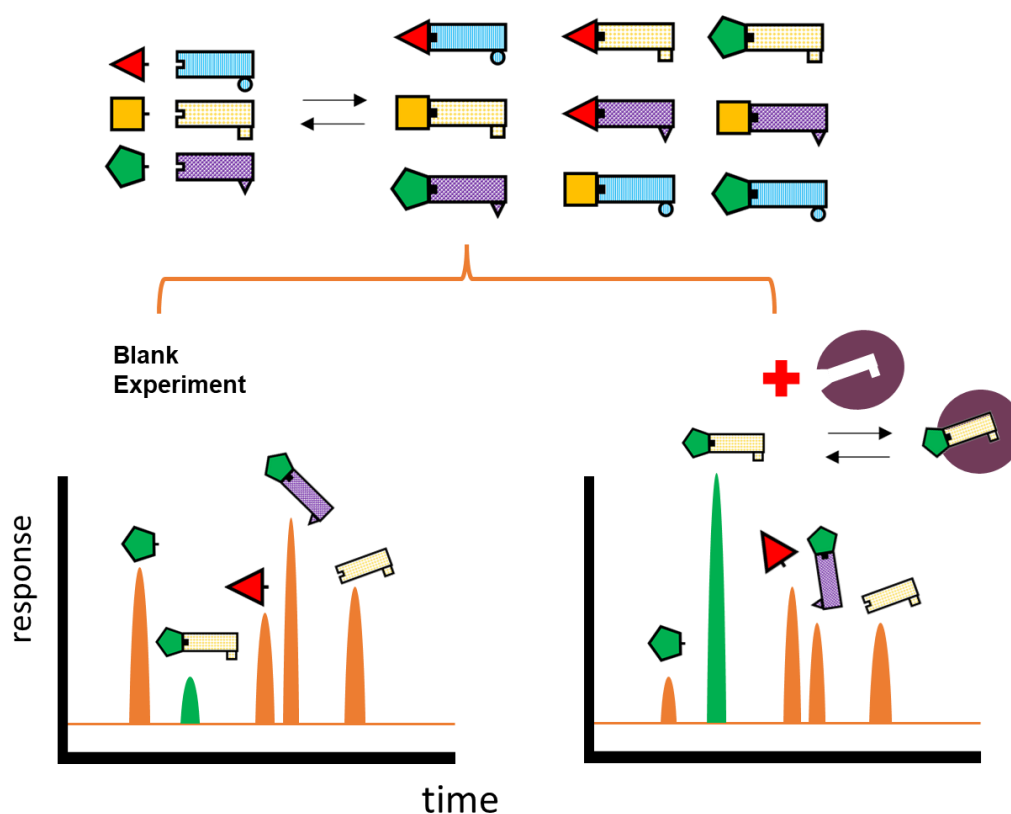


Figure 19: Schematic representation of the DCC protocol followed in this thesis. A DCL is generated making use of a reversible chemical reaction. Once the reactions have reached a thermodynamic equilibrium, the mixture is split in two. To one half, a template (biomolecule) is added. This generates a new situation of equilibrium in which the molecules with the highest affinity for such template will appear as amplified when compared to the non-templated library.

The ability to discover good binders makes DCC ideally suited for drug discovery.^{194,195,198–200,205–207} DCC has also been applied in the fields of material sciences, catalyst development, and receptor molding.^{201,208–224} Regarding the biomolecules employed as templates, there are published examples using DNA, RNA,^{225–232} and peptides,²³³ but the targets that have been explored the most in the field have been proteins.^{194,195,198–200,205,207} Despite the potential of DCC for molecular recognition, there are very few examples of its use for the detection of saccharides, with only a couple of examples found in the literature.^{188,192}

Rauschenberg et.al.¹⁸⁸ developed a DCL of linear and cyclic peptides connected by disulfide exchange, targeting α -methylated monosaccharides (galactose, glucose, mannose and fucose), as well as N-acetyl neuraminic acid (NANA) and a couple of disaccharides (sucrose and trehalose). In the DCC experiments, they observed amplification of one of the DCL members for NANA and another member for trehalose, and they confirmed such binding by ITC. While these were ground-breaking results an opened the door for more scientific research in the field, same aspects were not considered in their investigation. They did not carry out replicates to ensure the reproducibility of the results and they only tested by ITC the most amplified molecule of each library against the template presumably causing such amplification, and a different template as negative control. There is not enough data to confidently affirm that there is correlation between amplification and affinity. Moreover, they did not carry out enough ITC experiments with the amplified molecules and the rest of the saccharides considered to discuss selectivity in their receptors, and the saccharide that they could bind with the strongest affinity, NANA, is perhaps a less challenging sugar to detect in comparison with common monosaccharides due to its N-acetyl and carboxylate groups.

Yang et. al.¹⁹² published a highly interesting approach utilising metal organic cages capable of generating H bonds with guest molecules as scaffolds in their DCC experiments. They tested common monosaccharides (D-Glucose, L-Glucose, D-Fructose, L-Arabinose, and D-Xylose among others) as templates, and reported a surprising self-amplification effects, as one of the helicate members of the library got amplified in presence of β -D-Glucose and, at the same time, the concentration of β -D-Glucose got increased as the cage induced the mutarotation of α -D-Glucose into its β anomer. They also demonstrated certain selectivity of one of the helicates for β -D-Glucose over its α anomer, as well as over the α and β anomers of the 1 O-methylated glucosides. However, the selectivity did not extend to most of the other monosaccharides tested. Moreover, this DCC experiments were carried out in organic solvent (MeCN), most likely due to the hydrophobicity of the library constituents, which limits the applicability of these receptors. Hence, there is room for new knowledge in the field of DCC for the discovery of new receptors for saccharides, and there is a gap in the literature that we intend to fill.

DCC poses as an excellent tool to overcome most of the difficulties already discussed in the field of molecular recognition of saccharides. A cautiously designed DCC experiment can create a complex and diverse library of compounds posing multivalency, by incorporating multiple binding sites into a single molecule, and a rich variety of functionalities susceptible to bind the sugar target via covalent and/or non-covalent interactions. The combination of these two factors enables the creation of receptors with enhanced overall binding affinity for any given template, including saccharides.

Another advantage of DCC is its ability to generate complex mixtures of molecules with potentially desirable biological activity, that are difficult to synthesize using traditional methods. For example, DCC has been used to generate libraries of cyclic peptides,^{187,188} which are challenging to synthesize using traditional solid-phase peptide synthesis methods.

DCC is a rather complex technique with many factors to be considered. It has been widely employed and there are several reviews on the topic covering every detail to be considered when designing a DCC experiment: different types of DCC,^{195,201} selection of starting building blocks (BBs),¹⁹⁵ reversible reactions that can be employed,^{199,207,221,224} or the different species that can serve as templates.^{195,201,205} This thesis will not attempt to cover in detail all of these aspects but instead will summarise and give examples of them.

1.6.1 Different types of DCC

DCC experiments can be classified in three different groups. They all have a common first step of generation of reaction in equilibrium but differ in the screening or selection phase.¹⁹⁵

Adaptative DCC. The DCL is generated in the presence of the target. The dynamic characteristic of the system allows the adaptation of the library to the template, the subsequent amplification of the best-bound ligands, and the modification of the system to respond to disturbances and external triggers, such as changes in pH or temperature.

For example, if one DCL member interacts with the target better than the others, this element will be removed from the equilibrating pool, and all of the constituent BBs that make up this

element will likewise be masked by the binding. The system will then change in order to create more of this member at the expense of the other species in the library due to the equilibrium condition. In comparison with the blank experiment (no target molecule was introduced), the most active component (the best binder) will consequently suffer some degree of amplification upon re-equilibration. There are many examples in the literature of Adaptive DCLs,^{189,234,235} and this will be the approach followed in this thesis as is the simplest to implement that provides actual amplification values.

Combinatorial chemistry. Named by some authors as pre-equilibrated DCC, in this instance, the combinatorial library is built under reversible settings that are analogous to the adaptive scenario but the screening step is carried out under static conditions. The reaction in equilibrium is stopped (change in pH, temperature, addition of external agent) before the addition of the biomolecule target. Therefore, no amplification can take place. Then, deconvolution protocols can be used to identify the active components. This protocol is very helpful when dealing with delicate and sensitive biological target species that are not available in large quantities or when the conditions for the equilibrium reaction are incompatible with the biological target. There are numerous examples in the literature employing this protocol.^{187,193,236–250}

Iterative DCC. In this example, the DCL is produced in one compartment under reversible conditions, and after that, in a later phase, its components are permitted to interact with the target species either in the same reaction chamber or in a separate one. The species that bound the target need to be separated from the unbound ones. This can be done, for example,

by utilising target entities that are immobilised in a surface. Following this, the unbound species are returned to the reaction chamber, recombined, and given another chance to engage with the binding site. This "dynamic panning protocol" can be repeated numerous times, and the accumulated active species can then be examined.²⁵¹

1.6.2 Starting BBs in a DCC experiment

The selection of the most appropriate starting BBs is as well a crucial step when designing a DCC experiment, as it will define the ability for the DCL to create potent ligands.

First and most importantly, the building blocks must possess functional groups capable of reversible exchange. Additionally, these units should cover the potential target sites in both geometry and functionality, which can be achieved with recognition groups designed to interact with the binding site.¹⁸⁵ The selection of these groups may be determined through experience or rational design based on analysis of the target species' structure. The approach that will be followed in this thesis, as it will be further explained in a separate chapter, is the enrichment of the system with a variety of different functionalities that will ensure the existence of all sorts of interactions (covalent and non-covalent) so the template can 'choose' the interactions that suit it the best. For optimal binding, these recognition groups should be organized geometrically, using organizational units, which may be based on molecular scaffolds that can undergo dynamic decoration through reversible reactions. These scaffolds will establish the core geometry and multivalency of the DCL members. The molecule of isophthalaldehyde, as well as some short peptides, will be employed in this thesis to meet this requirement.

The first requirement that has to be met for the starting BBs is the possession of a functional group capable of reversible exchange. There are a number of chemical groups that can achieve this and have been used in DCC experiments, as represented in Table 1.

Table 1: List of chemical groups employed for DCC experiments.

Reaction	Equilibrium	References
imine formation (+ reduction)	$R_1-\overset{\text{O}}{\parallel}{C}-H + R_2-NH_2 \rightleftharpoons R_1-\overset{\text{N}^{\text{R}_2}}{\parallel}{C}-H \rightarrow R_1-CH_2-NH^{\text{R}_2}$	244,252–260
acylhydrazone formation	$R_1-\overset{\text{O}}{\parallel}{C}-R_2 + H_2N-\overset{\text{O}}{\parallel}{C}-R_3 \rightleftharpoons R_1-\overset{\text{R}_2}{\underset{\text{N}}{\parallel}}-\overset{\text{H}}{\underset{\text{N}}{\parallel}}-\overset{\text{O}}{\parallel}{C}-R_3$	236,243,245,261–264
thiol–disulfide exchange	$R_1-SH + R_2-SH \rightleftharpoons R_1-S-S-R_2$	265–272
boronate <i>trans</i> -esterification	$R_1-\underset{\text{OR}_3}{\overset{\text{OR}_2}{\text{B}}} + R_4-OH \rightleftharpoons R_1-\underset{\text{OR}_4}{\overset{\text{OR}_4}{\text{B}}}$	273,274
Michael-type addition to enones	$R_1-\overset{\text{O}}{\parallel}{C}-CH=CH_2 + R_2-X \rightleftharpoons R_1-\overset{\text{O}}{\parallel}{C}-CH_2-CH_2-X-R_2$ X = OH, SH	275,276
hemithioacetal formation	$R_1-\overset{\text{O}}{\parallel}{C}-H + R_2-SH \rightleftharpoons R_1-\overset{\text{OH}}{\underset{\text{S}^{\text{R}_2}}{\text{C}}}-H$	277
metal–ligand coordination	coordination of pyridines to Fe ^{II}	278
enzymatic reaction	N-acetylneuraminic acid aldolase	279,280

The reactions responsible for reversibility must take place in conditions that do not affect the target activity or structure. This means, for most of biological targets, that the equilibrium reaction has to be compatible with physiological conditions. The starting materials cannot produce off-site reactions with the target, and the products have to be stable enough to be analysed in a later step. To accomplish this, it is necessary to switch off the reversibility of reactions used during library generation, resulting in pseudo-static libraries that cannot react

to external stimuli and are more amenable to analysis. There are only a limited number of reactions that meet these requirements, and they are those summarised in Table 1.

The first ever example reported of DCC was carried out in 1997 and utilised the reversible formation and exchange of imines.²⁵² Since then, there have been numerous reported examples of this chemistry, most of the times employing amines and aldehydes as starting BBs,^{244,252–258,260} although there is an example employing ketones.²⁵⁹ To ensure the conversion to imines and also to prevent potential side reactions of the starting aldehyde with amino-containing targets (proteins, for instance), an excess of starting amine is used. The imines generated are labile and require reduction to form stable amines. Sodium cyanoborohydride is the most commonly used reducing agent as it selectively reduces imines over other common unsaturated groups such as aldehydes.^{252–255} The reduction step does not alter the composition of the iminic library, and, even though the reduction from an imine to a secondary amine group may cause an alteration in the affinity of the BBs for the target, it is a compromise that has to be made with imine chemistry. However, we do not anticipate this behaviour as we expect that the ‘recognition units’ in the potential receptors would be far away from the imine/secondary amine position. If activity loss occurs in the reduced amines, a solution that some researchers proposed was the synthesis of the amide analogues for follow-up investigations, as their affinities better aligned with the amplification observed by imines during the DCC experiment.²⁵⁴ It is essential to note that upon reduction of the C=N double bond, information on the E or Z conformation of the imines is lost, and the potentially higher affinity of a single isomer cannot be traced back.

Due to their biocompatibility, easy to control chemistry, and the fact that there are numerous and diverse interesting BBs commercially available that contain the amine and aldehyde groups, imines are the most common functional group employed as core of the reversible exchange in DCC experiments and they will be used for the experimental work of this thesis.

In conclusion, DCC is a powerful method for generating complex and diverse chemical systems, including receptor molecules for biological targets. The method is versatile and can be used to generate a wide range of molecules that are difficult to synthesize using traditional methods. The iterative process of amplification and screening allows for the generation of highly selective and potent receptor molecules. The ability to generate multivalent receptors is a significant advantage of DCC, allowing for the creation of highly potent receptors for biological targets.

1.7 Conclusions

Glycans play key roles in biological processes and therefore they could be used as biomarkers for disease detection and risk stratification.^{2,13,29–34,43–46} However, this use is limited by the current limitations to detect exclusively the ones involved in disease development and not those present naturally in healthy cells.^{9,10,32,35–41} In this context, trying to recognise their saccharide side chains seem a feasible approach to effectively tell them apart.¹⁵

The molecular recognition of saccharides, however, is not exempt of challenges. Sugars are chemically difficult to detect in physiological conditions for a number of reasons: They are

structurally very similar to one another, they present different interconvertible forms, they are heavily solvated, and they lack easy-to-recognise functional groups.^{54–60}

1.8 The major improvements in the field over the last years with the discovery of ligands from natural sources and the creation of rationally designed synthetic ligands is reported herein in the section dedicated to Saccharide recognition

. However, there is still room for improvement, and we are in need of new and enhanced saccharide receptors.

The use of Dynamic Combinatorial Chemistry (DCC) has proven to be an effective strategy in the field of drug discovery,^{194,195,198–200,205–207} for the development of potent binders for biomolecules such as DNA, RNA,^{225–232} and proteins^{194,198–200,205,207}. However, its application in the discovery of saccharide receptors has been far less explored.

Considering the burgeoning demand within the field of biomedical sciences for novel and enhanced receptors for glycans, together with the saccharides' pivotal roles in discerning among these entities, the application of DCC to identify saccharide receptors (a domain relatively underexplored compared to DCC's application with other relevant biomolecules) emerges as a highly pertinent research capable of filling existing knowledge gaps in the scientific literature.

1.9 Aim of this work

This work main focus is the development and implementation of DCC for its use in the discovery of selective receptors for saccharides. While achieving that goal, other aspects need to be addressed:

- Optimisation of imine formation reaction (IFR) as it is pivotal for the correct functioning of DCC experiments (Chapter 3).
- Design of a small-molecule based DCC experiment and optimisation of the analytical procedures (LC-MS) for its screening (Chapter 4).
- Synthesis of the receptors for their further analysis (Chapter 7).
- Validation of the DCC results with analytical techniques (ITC and NMR). Study the mechanisms behind molecular recognition (Chapter 5).
- Explore the feasibility of more complex DCC systems, employing peptides as scaffolds (Chapter 6).

Chapter 2 -Techniques

This chapter will describe in detail the main techniques employed throughout this thesis for analysis (NMR, LC-MS, ITC) and purification (HPLC). ^1H NMR and ^{13}C NMR will be used for the characterisation of the molecules synthesised after DCC experiments: **1D**, **1P**, **2DD**, and **2PP**. ^1H NMR will also be used to carry out titration experiments with **1P** and Fructose, in order to gain a better understanding of their binding mechanisms. HPLC will be employed for the purification of the aforementioned synthesised molecules. LC-MS is needed for the screening of the Dynamic Combinatorial Libraries (DCLs), in order to calculate the amplification factors of the library members for each one of the sugar templates tested. Finally, ITC will be used in order to obtain the binding parameters (K_a and K_d) of the receptors of interest for the saccharide templates.

2.1 NMR

Nuclear Magnetic Resonance (NMR) Spectroscopy is an analytical technique used to determine the structure of organic molecules.²⁸¹ This technique relies on the quantum property of spin possessed by atomic nuclei. The spin is visualized as the spinning direction of the nucleus around its axis and gives rise to a non-zero magnetic moment ($\mu \neq 0$) for atoms with an odd total number of proton and neutrons, such as ^1H , ^{11}B , ^{13}C , ^{15}N , ^{17}O , ^{19}F and ^{31}P .^{282,283} When a magnetic field (B_0) is applied, the magnetic dipole of these nuclei aligns along the axis of the applied field.²⁸² The number of possible orientations (S) is defined by Equation 2.1.

$$S = 2I + 1 \quad \text{Equation 2.1}$$

The nuclear spin quantum number (I) is fixed for each atom and represents the spin of the atom's nucleus.²⁸⁴ Protons have a nuclear spin quantum number of $I = \frac{1}{2}$, which means they can exist in two different orientations in the presence of an external magnetic field: parallel and antiparallel to the applied field B_0 .²⁸² The two orientations give rise to energy level differences (ΔE), which depend on the applied magnetic field B_0 and Planck's constant h , as described by Equation 2.2.²⁸³

$$\Delta E = \gamma \frac{h}{2\pi} B_0 \quad \text{Equation 2.2}$$

The gyromagnetic ratio (γ) is a nuclear property that is defined by the type of nucleus being studied. When electromagnetic radiation of frequency ν is supplied, corresponding to an energy equal to ΔE , the nuclei are excited to higher energy levels, which results in a spin inversion.²⁸³ The relationship between ΔE and $h\nu$ can be used to obtain Equation 2.3, and subsequently, Equation 2.4 can be derived.

$$h\nu = \Delta E = \gamma \frac{h}{2\pi} B_0 \quad \text{Equation 2.3}$$

$$\nu = \frac{\gamma}{2\pi} B_0 \quad \text{Equation 2.4}$$

After excitation, the nuclei relax and emit at their resonant frequency ν , producing an interferogram of superimposed waves known as FID (free induction decay).²⁸³ This signal is then processed by Fourier transformation to produce the NMR spectrum. In the NMR spectrum, each resonant frequency for each nucleus appears as a signal, after transformation into a chemical shift (δ) and quoted in parts per million (ppm).²⁸³ This is a crucial step in the

analysis of an NMR spectrum as it enables the identification of the different types of nuclei present in the sample being examined.

The chemical shift (δ) is a measure of the influence of the chemical environment on the resonant frequency of a specific nucleus. It is defined by the frequency in Hz of the nucleus (ν_i), the reference frequency (ν_R), and the frequency of the instrument in MHz (ν_0), as described by Equation 2.5.²⁸⁴

$$\delta = \frac{(\nu_i - \nu_R)}{\nu_0} \quad \text{Equation 2.5}$$

The shielding of the nucleus is determined by the electrons near it, which provide an opposing magnetic field to B_0 , thus reducing the effective magnetic field seen by the nucleus, which results in a lower resonance frequency. According to Equation 2.4, a higher electron density results in greater shielding and a smaller chemical shift, shifting the signal to higher fields. In contrast, when the nucleus is de-shielded, the signal is shifted to lower fields, resulting in larger δ values. Figure 20 shows the ^1H NMR spectrum of a simple organic molecule, 4-ethoxyacetophenone. It can be observed that different ^1H signals with unequal chemical environments (a-e) appear at different regions of the spectrum. The numbers in blue underneath the peaks correspond to their integral.

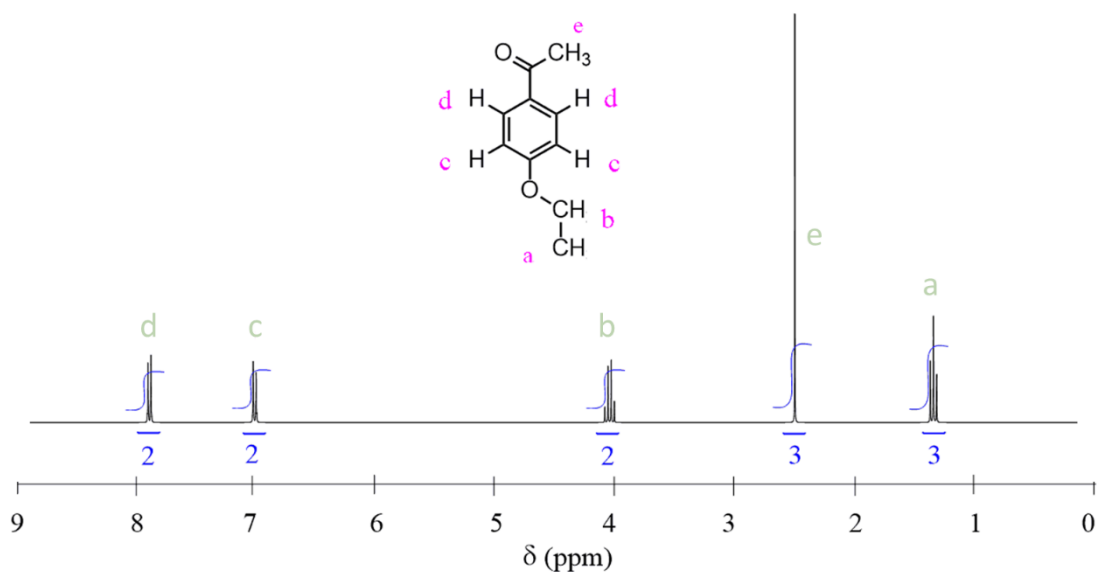


Figure 20: ^1H NMR spectra of 4-ethoxyacetophenone as an example. Assignment of peaks is done with letters (a-e), and integrals are showed in blue under each peak.

The signal multiplicity, coupling constant (J), and signal integral are additional key parameters for the interpretation of NMR spectra.

The multiplicity (sometimes called peak splitting) provides information on the number of NMR-active nuclei that are within a certain distance (usually three bonds away from a given nucleus). The intensities of the resulting peaks in which an original signal will split is given by the Pascal triangle (Figure 21). Hence, if a signal is split in 3, according to the distribution given by the Pascal triangle (1-2-1), the intensities of the resulting peaks will be 25%-50%-25%.

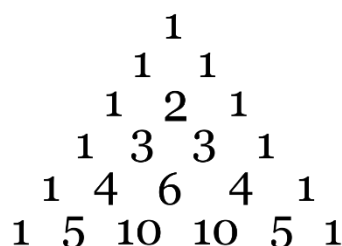


Figure 21: The Pascal triangle can be used to determine the intensities of the peaks for any given splitting pattern.

A signal that appears as a doublet suggests the presence of a nucleus closer with one equivalent proton, while a signal as a triplet suggests the presence of a nucleus with two equivalent protons, and so on. Figure 22 shows some common peak splitting patterns in ^1H NMR.

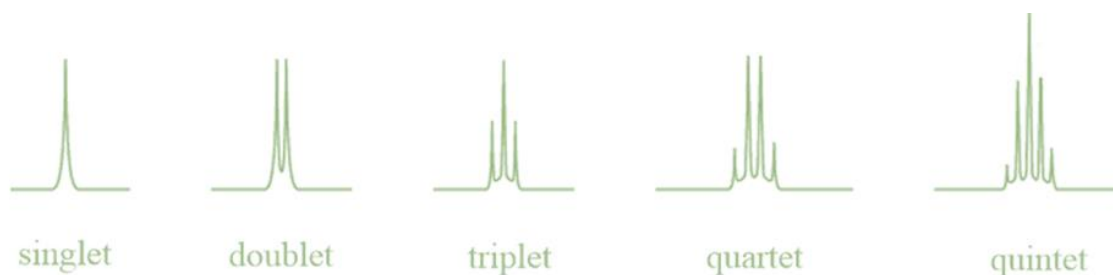


Figure 22: common multiplicity or splitting patterns observed in ^1H NMR.

The integrated area of a peak gives important information too, as it is directly proportional to the total number of protons producing the signal.

Coupling constants J (Hz) provide insights into the chemical bonding of atoms, with different J values corresponding to different types of chemical bonds. Understanding these various NMR parameters can greatly assist in the characterization of molecules by NMR spectroscopy.

In Figure 23 are represented different ^1H - ^1H coupling constants for hydrogens in different chemical arrangements in organic molecules.

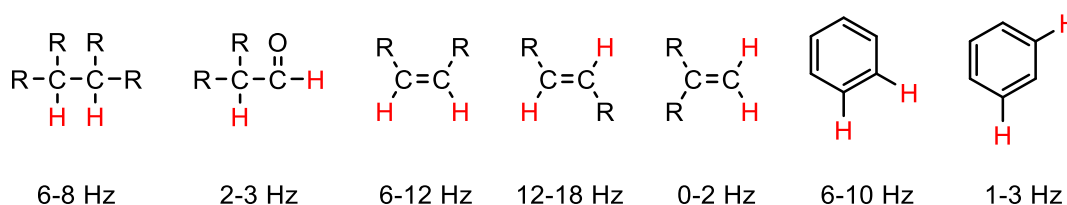


Figure 23: H-H Coupling constant (J) values in common chemical arrangements in organic molecules.

In addition to ^1H , other nuclei such as ^{13}C are also employed in NMR for the identification of organic molecules. However, due to its lesser abundance, the sensitivity of C NMR is much lower than H NMR (Table 2).

Table 2: Properties of most common H and C nucleotides and their natural abundance.

Nuclide	Number of Protons	Number of Neutrons	Spin (I)	Natural Abundance
^1H	1	0	$\frac{1}{2}$	99.9885%
^2H	1	1	1	0.0115%
^{12}C	6	6	0	98.9000%
^{13}C	6	7	$\frac{1}{2}$	1.0700%

Two-dimensional (2D) experiments, such as COSY (CORrelated SpectroscopY) and HSQC (Heteronuclear Single Quantum Correlation), can provide additional information on the molecular structure. COSY refers to the ^1H - ^1H correlation spectroscopy and displays the correlation between H nuclei separated by a maximum of three bonds. The correlation

between protons is displayed as cross peaks in the 2D spectra. Similarly, ^1H - ^{13}C HSQC is used to observe the correlation of protons attached to a specific carbon.²⁸⁵ In HSQC, the signal from the proton is correlated to the signal from the attached carbon, making it useful in identifying the position of protons in a molecule. The use of these 2D NMR techniques can provide more detailed information about molecular structure and aid in the identification of molecules. There are also other 2D NMR techniques available, such as NOESY (Nuclear Overhauser Effect Spectroscopy) and TOCSY (Total Correlation Spectroscopy) which provide further information about the spatial arrangement of atoms in a molecule.

2.2 HPLC

Liquid chromatography (LC) and high-performance liquid chromatography (HPLC) are both widely used chromatography techniques for separating complex mixtures.²⁸⁶ HPLC is a significant improvement over traditional liquid chromatography, as it utilizes pressurized liquid solvent (50-350 bar) and smaller adsorbent particles to provide superior resolving power for sample separation. These techniques rely on two phases, a stationary phase, and a mobile phase, to separate analytes based on their affinities. In this discussion, we will focus on reverse phase chromatography, commonly employed in organic chemistry. In reverse phase chromatography, an apolar stationary phase such as a C18 column is packed inside a chromatography column, while a polar mobile phase consisting of a mixture of water and acetonitrile, or water and methanol is passed through it. Typical HPLC column dimensions are 2.1–4.6 mm in diameter and 30–250 mm in length. The HPLC instrument also consists of a degasser, sampler, pumps, detector, and a column oven (Figure 24).

The degasser is responsible for removing dissolved gases (such as oxygen and carbon dioxide) from the solvent before it enters the HPLC column. This is important because dissolved gases can interfere with the separation process and cause baseline drift in the chromatogram.

The sampler introduces the sample to be analysed into the HPLC system. The sample is typically injected onto the head of the HPLC column, where it is then carried through the column by the mobile phase.

The pumps deliver the mobile phase (or eluent) to the HPLC column at a constant flow rate. In most HPLC systems, the mobile phase is composed of a mixture of solvents (such as water and acetonitrile) that are blended together in precise proportions.

The detector is responsible for detecting and measuring the analyte (or analytes) of interest as they elute from the HPLC column. There are many different types of detectors used in HPLC, including UV-Vis, fluorescence, and mass spectrometry detectors.

The column oven keeps constant the temperature of the HPLC column during the chromatographic separation process. The column is typically held at a temperature between 20-40°C to ensure reproducible separations. Additionally, some analyses may require the column to be heated or cooled during the separation process to optimize separation efficiency or selectivity.

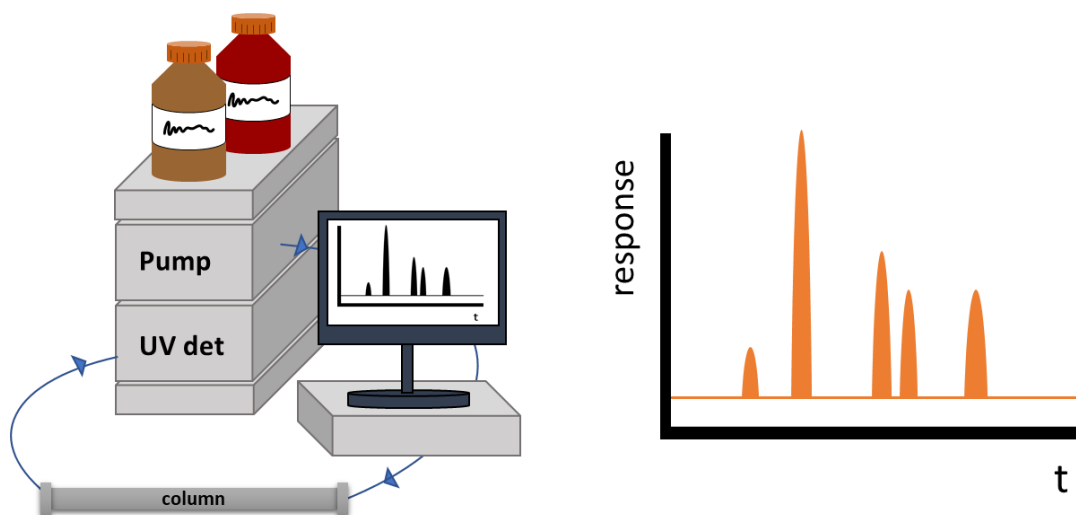


Figure 24: Schematic representation of a HPLC system.

During the chromatographic run, the composition of the mobile phase can be altered to aid in separating similar species, which is known as gradient elution. When a mixture is loaded onto the column, the different components of the mixture are separated based on their affinities for the stationary phase versus the mobile phase. Each component is eluted out of the column at a specific time, known as its retention time. In most common scenarios, when apolar chromatography columns are employed, polar components will have weak affinity for the stationary phase and therefore they will elute more quickly and thus present shorter retention times, while apolar components take longer to elute and present longer retention times. Each component is represented in the chromatogram as a peak with a corresponding retention time. In the schematic chromatogram shown in Figure 24, assuming that a non-polar column was employed, polarity of eluted analytes would decrease from left to right.

The selection of the most appropriate chromatography column is perhaps the most important aspect for a successful separation. There are a number of different columns for reverse phase chromatography available on the market, each one of them with particular specifications:

C18 Columns: The most widely used and available column for reverse phase (RP)-HPLC is the C18 column. It has an octadecylsilane (ODS) stationary phase, which is highly hydrophobic and interacts strongly with analytes through van der Waals forces and hydrophobic interactions. C18 columns offer excellent separation efficiency and peak symmetry, making them suitable for a wide range of applications, from pharmaceutical analysis to environmental monitoring. Moreover, C18 columns come in different particle sizes and pore sizes, allowing for tailored separation performance.

C8 Columns: These columns are similar to C18 columns, but with a shorter alkyl chain length. They offer less retention for hydrophobic analytes, resulting in faster elution times, but they also provide less separation efficiency and resolution. C8 columns are useful for the analysis of less hydrophobic compounds, such as small organic molecules and natural products.

C4 Columns: Following the trend, C4 columns have an even shorter alkyl chain length than C8 columns, resulting in even faster elution times and less retention for hydrophobic analytes. However, they also offer less separation efficiency and resolution than C18 or C8 columns. C4 columns are typically used for the analysis of very hydrophilic compounds, such as polar amino acids and peptides.

Phenyl Columns: Phenyl columns have a phenyl stationary phase, which provides an alternative selectivity to the more commonly used alkyl-bonded phases. Phenyl columns offer a more aromatic selectivity, which can be useful for separating compounds that contain aromatic rings.

Cyano Columns: As their name suggest these columns have a cyano functionalized stationary phase, which provides an alternative selectivity to the more commonly used alkyl-bonded phases. They are particularly useful for the analysis of polar compounds, such as peptides and nucleic acids. Cyano columns also have the advantage of being more stable than other columns, making them suitable for high-temperature applications.

As it was said, compounds are identified by their retention time in HPLC columns, which are visualized as chromatograms and their amount quantified by the area under the peak.²⁸⁷ The type of detection methodology employed can vary between mass spectrometry, refractive index, or UV detection, among others. HPLC can be classified into two types: analytical and preparative, based on their primary function. Analytical HPLC is used for both qualitative and quantitative sample analysis, as it separates individual sample components. Preparative HPLC isolates and purifies specific substances from mixtures. Analytical HPLC can also be used to determine the optimal separation conditions, which can then be applied to preparative HPLC.

2.3 LC-MS

Liquid Chromatograph – Mass Spectrometry (LC-MS) is an analytical technique that combines the resolving power of Liquid Chromatography with the detection specificity of via Mass Spectrometry analysis.²⁸⁸

After HPLC chromatographic separations, compounds are eluted off the column and desolvated into the gas phase before being ionized and introduced into the MS instrument for mass analysis. The resulting ions are separated through an electromagnetic field in a mass analyser based on their mass to charge ratio (m/z).^{289,290} For small molecules ionized by

electrospray, singly charged ions are usually generated with minimal fragmentation. As a result, the obtained m/z ratio reflects the analyte's molecular weight as an ion. In the ionization process, a proton or cation is typically added to the analyte for ionization in positive ion mode. Therefore, the m/z values correspond to the analyte's mass plus a proton $[M + H]^+$ or plus a cation like sodium $[M + Na]^+$. Conversely, if the ionization is done in negative ion mode, a proton is lost from the analyte with the m/z value corresponding to the analyte's mass minus a proton $[M - H]^-$.

LC-MS provides molecular weight and structural information on the separated compounds, allowing for identification and quantification. Diode array detectors can be used to acquire data on selected UV and visible wavelengths and spectra, while MS can be used with non-UV responsive analytes. The use of both a UV detector and MS detector can be employed simultaneously to identify and quantify compounds (Figure 25). Additionally, UV-diode array detection can help identify the presence of non-ionizable compounds as long as they can absorb in the ultraviolet (UV) wavelengths range.²⁹¹

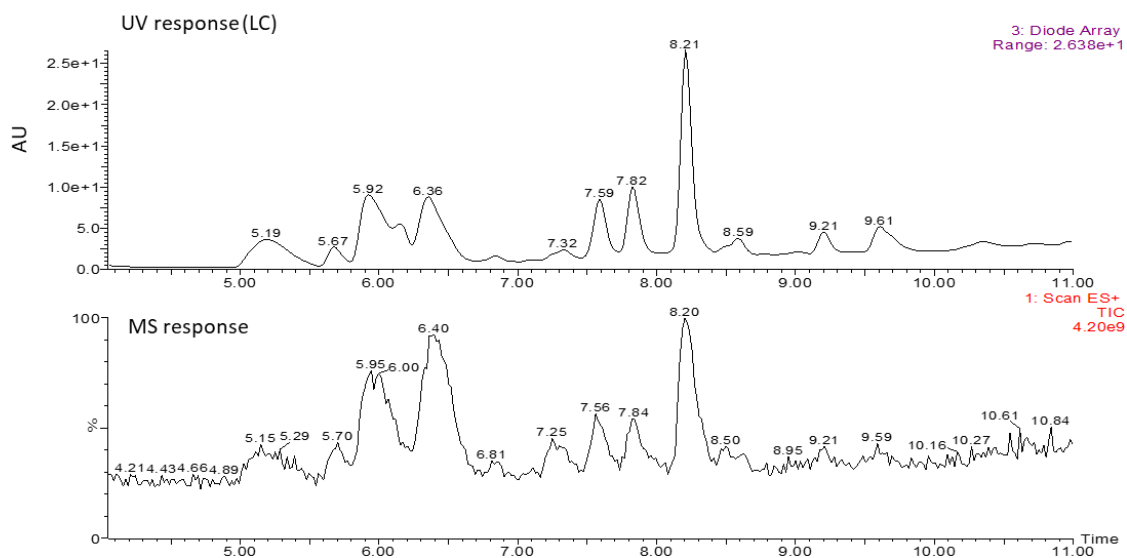


Figure 25: Example of a fraction of a LC-MS chromatogram displaying UV absorbance response (top) and MS response (bottom). Each peak corresponds to a different compound.

2.4 ITC

Isothermal Titration Calorimetry (ITC) is a powerful technique used to measure the heat released or absorbed during a chemical reaction in solution.²⁹² It is based on the principles of thermodynamics, which dictate that a reaction between two molecules will result in a change in heat. By measuring this heat variation change in heat, ITC can provide valuable information about the thermodynamics of the reaction, including the enthalpy, entropy, and Gibbs free energy changes. The binding event is generally represented as the equilibrium between a receptor (R) and a ligand (L) to form the corresponding complex (RL) (Equation 2.6).



The ITC instrument consists of two cells, one containing a solution of the molecule to be titrated (the “sample” cell), and the other containing a solution of the titrant (the “reference” cell). The two cells are kept at the same temperature, hence the term isothermal. The titrant is injected into the sample cell in small increments while the heat released or absorbed (for exothermic and endothermic processes, respectively) is continuously measured by a highly sensitive calorimeter providing a thermodynamic profile of the interaction.

The heat associated with the binding event, Q (cal), is defined by Equation 2.7, 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.²⁹³

$$Q = [RL]V_0\Delta H \quad \text{Equation 2.7}$$

The heat change after each injection is proportional to the amount of complex RL formed, or in other words, to the amount of receptor that binds the ligand. As the number of injections increases, there is less and less of receptor available to be bound by injected ligand. As a result, there is a decrease in the net amount of heat change after each injection that ends up with the saturation of the system. The ITC system utilises the data collected during the titration experiment and calculates the thermodynamic parameters.

The data collected during the titration are integrated and fitted to provide the thermodynamic parameters associated with the binding event such as the binding constant K_a (M^{-1}) or dissociation constant K_d (M) between receptor and ligand (Figure 26, Equation 2.8, Equation 2.9).

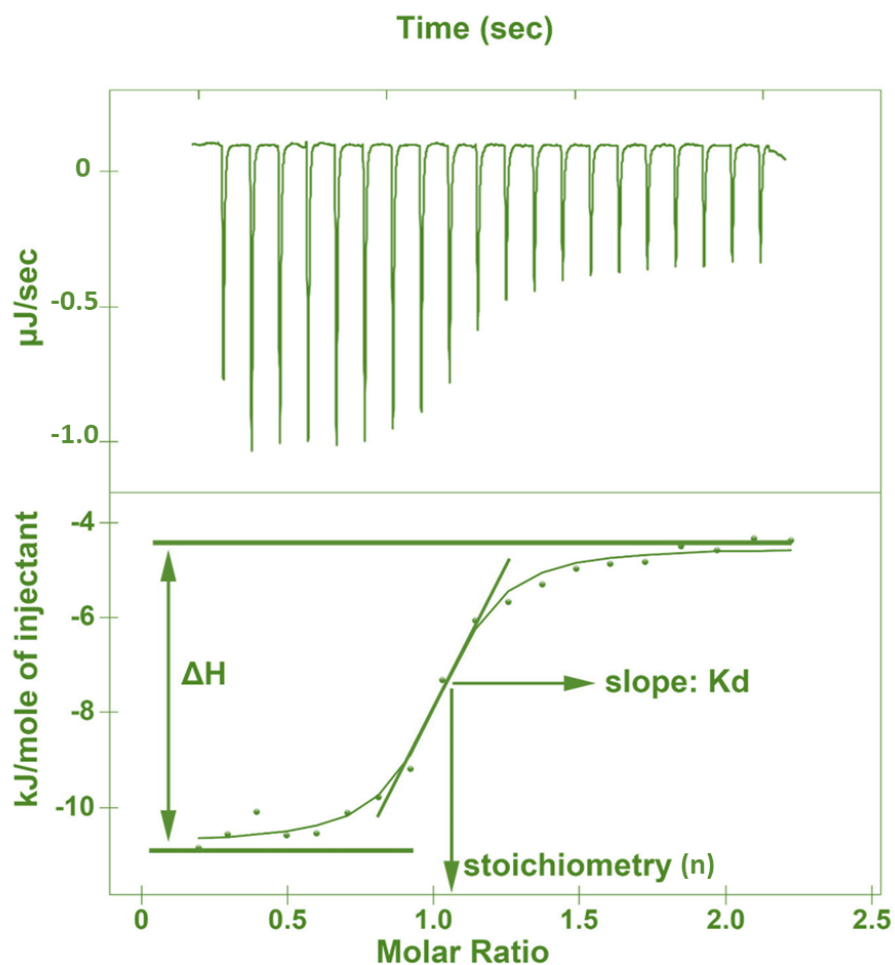


Figure 26: Titration graph for an exothermic binding event (top), and fitting of the data in the titration curve to obtain the relevant thermodynamic parameters.

$$K_a = \frac{[RL]}{[R][L]} \quad \text{Equation 2.8}$$

$$K_d = \frac{1}{K_a} \quad \text{Equation 2.9}$$

ITC also gives information regarding the stoichiometry of binding (n), the enthalpy of binding (ΔH , cal mol⁻¹) and entropy of binding (ΔS , cal mol⁻¹ deg⁻¹). With these data, the free energy of binding (ΔG , cal mol⁻¹) can be extracted from Equation 2.10, with the molar gas constant R being 1.987 cal mol⁻¹ K⁻¹ and T in K.

$$\Delta G = -RT \ln K_a = \Delta H - T\Delta S \quad \text{Equation 2.10}$$

ITC is a useful method for studying the thermodynamics of various binding events since most interactions are associated with heat changes. This technique is commonly used to investigate the binding of proteins to other macromolecules, carbohydrates, or small molecules.²⁹⁴ It can also be applied to other systems, such as the binding between small molecules,^{169,295} or these with nanoparticle systems.²⁹⁶

One of the significant advantages of ITC over other methods is that it doesn't require any chemical modifications or immobilization of the receptor.^{294,297}

An ITC instrument typically consists of a sample cell that contains one of the binding partners, and a syringe that contains the other binding partner. The syringe is attached to a motor-driven piston, which allows for the precise delivery of small amounts of ligand solution into the sample cell. The sample cell is placed in an adiabatic shield along with a reference cell (Figure 27). During the experiment, the ligand is injected into the receptor solution in the sample cell while the reference cell contains only the solvent. If the reaction solvent is organic, the reference cell must be filled with the same solvent. If it is aqueous (buffers for instance), the reference cell can contain just deionised water (properly filtered and degassed). The temperature is selected before the start of the experiment and can range from 2 to 80 °C,²⁹⁸ but is usually set at 25 °C. The volume of the sample cell may vary depending on the instrument used but typically ranges from 300 µl to 3 ml.

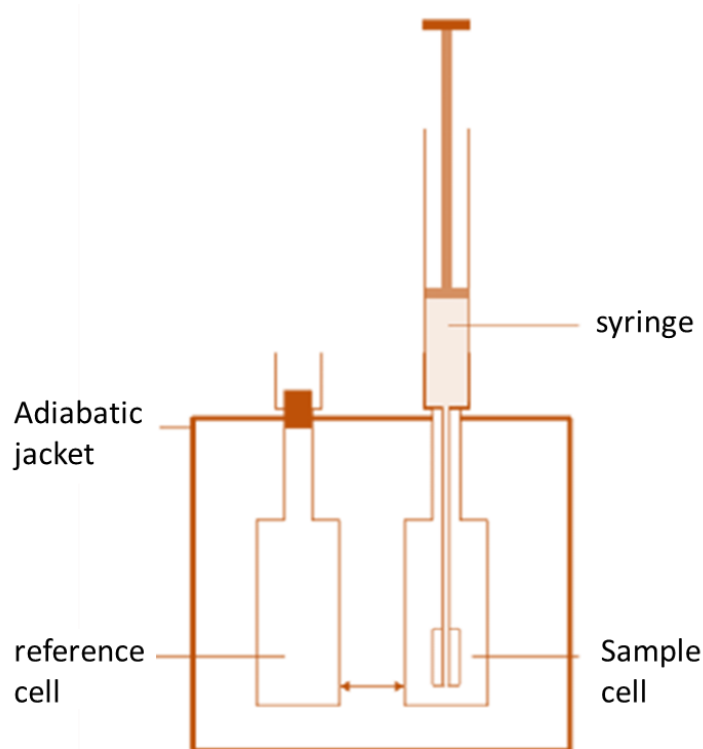


Figure 27: Schematic representation of an ITC instrument.

A thermoelectric device is placed between the sample cell and reference cell to measure the temperature difference (ΔT). A constant power (≤ 1 mW) is provided to the reference cell during the experiment, which activates a feedback circuit to maintain the $\Delta T = 0$ °C between the cells by supplying power to the sample cell.²⁹² The differential power (DP), also known as feedback power, provided to the sample cell remains constant when there is no binding event occurring, and it serves as the experiment's baseline.²⁹² However, when a binding event occurs, the heat change associated with it causes a variation in the sample cell temperature. For exothermic reactions, the heat released provides some of the heat that the DP would have supplied, causing the DP to decrease as less power is required to maintain the ΔT at 0 °C. On the other hand, for endothermic reactions, the DP increases as more power is needed to keep the ΔT at 0 °C. The experimental data recorded during the titration is the DP supplied to

the sample cell (DP, $\mu\text{cal s}^{-1}$), which can be plotted against time. For exothermic reactions, the resulting plot shows negative peaks as the DP decreases compared to the baseline, while endothermic events produce positive peaks. It is worth noting that the thermoelectric device measures the temperature difference between the two cells and not the absolute temperature of the sample cell, which is why the baseline value of DP is important.

The heat change resulting from the injection, represented by ΔH (cal mol^{-1}), is calculated by integrating 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.²⁹⁹ This means that the area under the peak represents the ΔH of the injection. The ΔH is then plotted against the molar ratio to generate the calorimetric binding curve, which provides information about the binding process.³⁰⁰ The shape of the binding curve is determined by a parameter known as the c-value, which depends on the binding stoichiometry 'n', binding constant ' K_a ', and concentration of the receptor in the sample cell ' $[R]_{\text{cell}}$ ' as given by Equation 2.11

, which can be simplified to Equation 2.12

when $n=1$.²⁹⁷ Accurate determination of the c-value is crucial for the proper fitting of the binding curve and for the accurate calculation of thermodynamic parameters.³⁰⁰

$$c = n[R]_{\text{cell}}K_a \quad \text{Equation 2.11}$$

$$c = [R]_{\text{cell}}K_a \quad \text{Equation 2.12}$$

In order to obtain a well-defined sigmoidal binding curve, the c-value of a system should ideally fall between 10 and 500 (Figure 28).³⁰⁰ However, high affinity systems with $K_a > 10^7 \text{ M}^{-1}$ cannot be easily studied using ITC due to their high c-values, which result in an almost vertical

graph shape towards the saturation point, from which thermodynamic parameters cannot be calculated. In the same way, too low c -values cause a flat isotherm curve that makes difficult to determine such parameters.^{299,301}

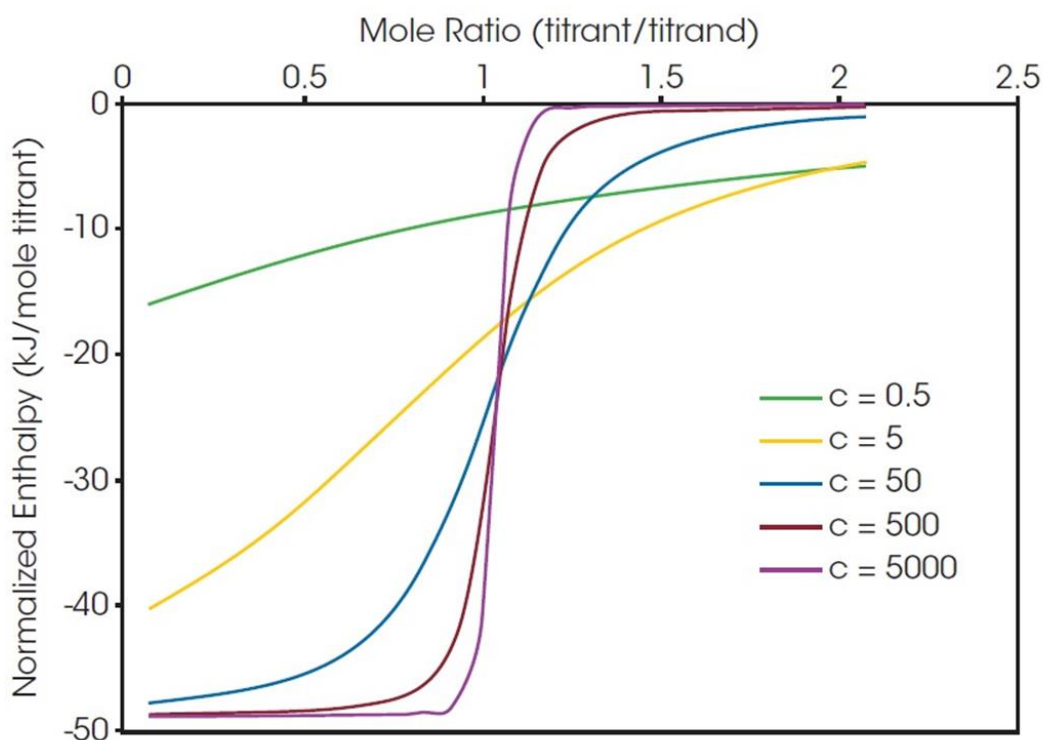


Figure 28: Influence of c value on the shape of the binding isotherm. Thermodynamic values K_d , n , and ΔH are identical for all the isotherms displayed.³⁰²

To decrease the c -value, the concentration of the receptor in the sample cell must be lowered, but concentrations below 10^{-7} M could produce signals that are too weak to detect.^{299,303} One way to study high affinity systems is through competition studies.³⁰⁴ Low affinity systems with $K_a < 10^4 \text{ M}^{-1}$ require high concentrations of the ligand to reach saturation and generally have c -values lower than 10. At low c -values ($c \sim 1.0$), the data of the first part of the sigmoidal curve cannot be collected, while at extremely low c -values ($c < 0.1$), it is difficult to accurately

determine all the thermodynamic parameters.²⁹⁹ To overcome this, the binding stoichiometry can be fixed using prior knowledge from literature data or previous studies, allowing experiments with extremely low c-values to still provide a reliable K_a value.²⁹⁷

The c-value can also be used to calculate the required experimental receptor concentration when K_a and n are known, as shown by Equation 2.13.²⁹⁹

$$[R]_{cell} = \frac{c}{nK_a} \quad \text{Equation 2.13}$$

Once the optimal concentration of receptor has been calculated, the concentration of the ligand can also be determined. It 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 with a molar ratio receptor to ligand of 1 to 2 or 1 to 3.³⁰⁵ The molar ratio is the ratio of concentrations between receptor and ligand at the end of the titration in the sample cell.³⁰⁶

To study systems with low c-values ($c < 10$), a high molar ratio of ligand is required (up to 1:10),²⁹⁷ and the ligand concentration can be calculated from the required molar ratio and the total injected volume.²⁹⁹

The ligand solution is injected sequentially into the sample cell, usually over 10 to 60 injections, with each injection consisting of a small volume of titrant (1-10 μL)³⁰⁷ selected to provide a detectable heat change.²⁹⁹

It is important to note that each instrument may have different sensitivity and reference power requirements. The sensitivity of the VP-ITC instrument used in this work is of 0.1

μcal .^{292,298} The time interval between injections is set to ensure t and the system returns to the baseline prior to the next injection. It is normally set to be between 180 and 480 s.²⁹⁹

The injection syringe is also responsible for mixing the cell content through the stirring paddle attached to its end.²⁹² The stirring rate (usually 250-300 RPM)³⁰⁵ is also selected prior to the experiment.²⁹⁸

A control titration is conducted to establish the heat of dilution, which involves injecting the ligand solution into the reaction solution to determine the heat associated with dilution.

The heat of dilution can then be subtracted from the titration data to obtain accurate results.

After collecting the data, it is analysed using software provided by the manufacturer. To analyse the data, an appropriate binding model must be chosen, such as one-binding site, two or more independent binding sites, or a cooperative sites model.³⁰³

Once the model is selected, the software estimates the thermodynamic parameters K_a , ΔH , and n (unless n is fixed) and uses these preliminary values to calculate the heat associated with each injection (ΔQ). The experimental and calculated ΔQ values are compared, and the thermodynamic parameters are refined through iterative least-squares fitting until the final values for n , K_a , and ΔH are obtained.

Chapter 3 -Optimisation of Imine Formation Reaction for its use in DCC

As stated in the introduction of this thesis, in Chapter 1.6, imine formation reaction (IFR) will be in the core of the DCC experiments of this work. Parameters such as reaction time, concentration of reagents and reaction pH will be investigated by ^1H NMR in order to maximise the conversion to products. This is a vital step to observe a reliable template effect in the future DCC experiments.

3.1 Optimisation of imine formation reaction (IFR)

3.1.1 Equilibration time and concentration of reagents in IFR

Imine formation is an equilibrium reaction. Therefore, the first parameter investigated was the equilibration time (i.e., time required for the reaction to reach a state in which concentration of species do not change over time) in the conditions that we anticipated they would be the reaction conditions further on in the DCC experiments.

IFR will be studied in aqueous buffered solvents as these are the conditions that will be employed in a later stage during DCC experiments. A compromise had to be found regarding the molar concentration of reagents. Too high concentrations led to the formation of precipitates, which cannot be tolerated when performing DCC experiments as the desired thermodynamic equilibrium would be affected by the solubility constants of the insoluble library members, phenomena named kinetic trap, biasing the experiment.³⁰⁸ On the other hand, If the concentrations are too low, the reaction may not take place or the products may

be formed to an extent below the sensitivity of analytical instruments and therefore IFR could not be implemented in DCC experiments.

In order to study the optimal range of concentrations for the use of IFR in DCC, a series of experiments were carried out with concentrations of reagents ranging from 0.5 – 2 mM of starting aldehyde **2**, and 1 – 50 mM of starting amine **E**. These molecules were selected for the study as they are aldehydic scaffold, and one of the amine BBs that will be employed during DCC experiments. This reaction would afford the monosubstituted product **2Ei**, and the disubstituted product **2EiEi** (subscript 'i' indicates that it is the iminic product) (Figure 29). A minimum of 2 equivalents of amine is required for the complete functionalisation of **2**, but an excess of it would help to displace the reaction equilibrium towards the formation of the disubstituted product.

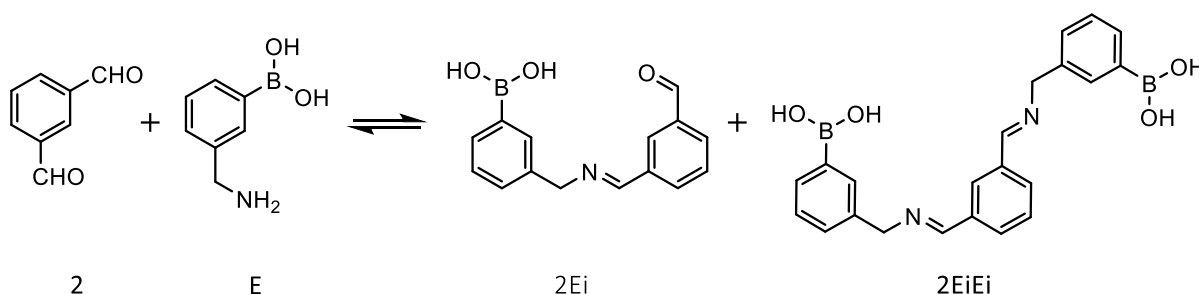


Figure 29: Reaction between **2** and **E** to investigate the optimal concentration of reagents for IFR.

The influence of the pH will be studied in a separate experiment later in this chapter, but pH 9.0 was selected at this stage as an alkaline pH was thought to be beneficial for IFR as it increases the nucleophilicity of the primary amine in **E**. As solvent it was selected 100mM ammonium hydroxide. The buffer was prepared from ND₄OD, D₂O and acetic acid-d₄ to adjust the pH. It must be clarified that, since we are working in deuterated conditions, we were

actually measuring pD and not pH. Values pD were converted to pH using the equation: $pD = pH + 0.4$.³⁰⁹

A certain amount of organic cosolvent was needed to maintain the condition of solubility due to the high hydrophobicity of the starting aldehyde **2**. The use of organic cosolvents is a common practice in DCC experiments. However, as we are interested in the study of receptor-target interactions in aqueous media, the percentage of organic solvent had to be kept as low as possible. MeCN was selected as cosolvent as it is compatible with analytical instruments that need to be used to analyse the Dynamic Combinatorial Library (DCL) in a future step. Different percentages of organic cosolvent (0 – 10%) were tested.

When selecting the range of concentrations of reagents to be tested, a compromise had to be made. Concentrations needed to be low to prevent the formation of precipitates, while the concentration of iminic products being formed was high enough to be accurately measured by ¹H NMR. The optimal reaction conditions we found were Aldehyde **2** 1mM + Amine **E** 20mM in 100mM ammonium hydroxide buffer pH 9.0 90%-10% MeCN-d₃ (Table 3). It must be pointed out that the addition of MeCN would have an impact on the overall solvent pH, but that deviation will be subtle and equal in all the experiments.

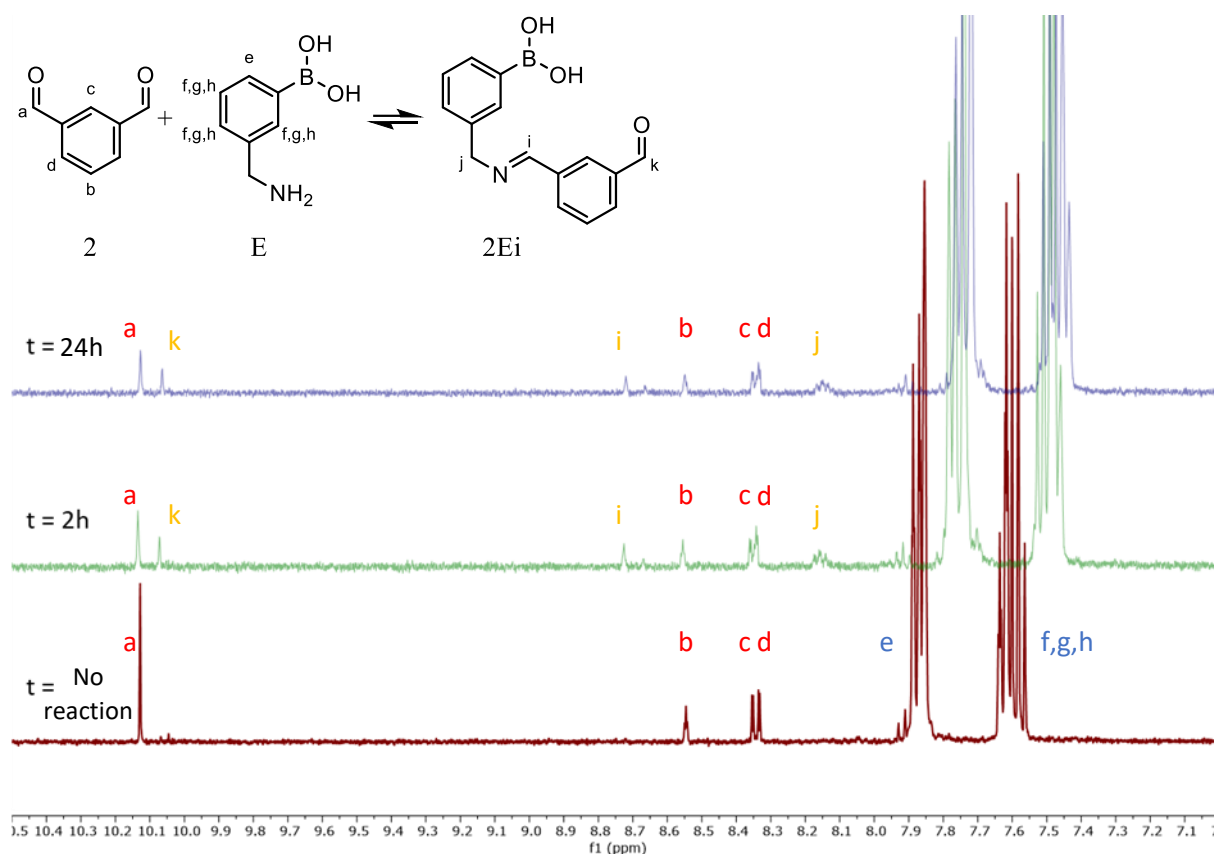
Cells that read 'Insol.' (red) mean that the combination indicated was not soluble or generated a precipitate. 'Sol (Invis.)' (yellow) means that the combination was soluble (no precipitates were observed to the naked eye), but no new products were detected by ¹H NMR. 'Sol (Visible)' (green) means that the combination was soluble (naked eye), and a new peak corresponding to an iminic product was detected by ¹H NMR.

Table 3: Combinations of concentration of reagents and percentage of organic solvent tested to find the best optimal conditions for IFR. Each experiment was conducted once.

			[E] (mM)					
			1	2	5	10	20	50
[2] (mM)	0% MeCN	0.5	Insol.	Insol.	Insol.	Insol.	Insol.	Insol.
		1	Insol.	Insol.	Insol.	Insol.	Insol.	Insol.
		1.5	Insol.	Insol.	Insol.	Insol.	Insol.	Insol.
		2	Insol.	Insol.	Insol.	Insol.	Insol.	Insol.
	5% MeCN	0.5	Insol.	Insol.	Insol.	Insol.	Insol.	Insol.
		1	Insol.	Insol.	Insol.	Insol.	Insol.	Insol.
		1.5	Insol.	Insol.	Insol.	Insol.	Insol.	Insol.
		2	Insol.	Insol.	Insol.	Insol.	Insol.	Insol.
	10% MeCN	0.5	Sol. (Invis.)	Sol. (Invis.)	Sol. (Invis.)	Sol. (Invis.)	Sol. (Invis.)	Insol.
		1	Sol. (Invis.)	Sol. (Invis.)	Sol. (Invis.)	Sol. (Invis.)	Sol. (Visible)	Insol.
		1.5	Insol.	Insol.	Insol.	Insol.	Insol.	Insol.
		2	Insol.	Insol.	Insol.	Insol.	Insol.	Insol.

With these conditions, an ^1H NMR study over time with **2** + **E** at pH 9.0 was conducted (Figure 30). The appearance and integration of the peak in the spectrum at 8.72 ppm, corresponding to the iminic proton in **2Ei** (Figure 29) was used to keep track of the reaction progress.

The reaction was analysed at t = No reaction; t = 2h and t = 24h.



*Figure 30: Plot of ¹H NMR spectrums corresponding to reaction between **2** and **E**. Peaks corresponding to starting aldehyde **2**, starting amine **E**, and product **2Ei** are flagged with letters a-d (red), e-h (blue) and i-k (yellow).*

No change was observed after 2h, meaning that the equilibrium had been reached. Nevertheless, to ensure this state regardless of the library composition and because of the limited availability of the NMR instrument, in future experiments we will always leave the reactions stirring 24h.

For logistic reasons, NMR spectrum at time 0 could not be recorded. In order to compare the experiments at 2h and O.N. with an experiment in which no imine has been formed, we use a spectrum of the reaction **2** + **E** in 100 mM phosphate buffer pH 6.5 (Figure 30, bottom). At this pH, the amino group in **E** is protonated and it is not nucleophile enough to react with aldehyde

2. The acidic pH alters the chemical environment of **E**, which explains the shift in the peaks of its aromatic protons (7.4 – 7.9 ppm, e-h, Figure 30).

It is noticeable how the peak at 10.12 ppm (a, Figure 30) corresponding to the CHO groups of the starting aldehyde gets smaller, but do not disappear completely. This is a proof that IFR is indeed an equilibrium reaction and the equilibrium point, at least under these reaction conditions, is far from the 100% conversion to imines (quantitative analysis will be carried out later in a separate experiment). A new peak is formed at approximately 10.05 ppm (k, Figure 30). In this region, it can only be an aldehydic proton too. It can be assigned to the CHO group in the molecule **2E_i**, the product of condensation of only one molecule of aldehyde with one molecule of amine. The different chemical environment and lack of symmetry of **2E_i** makes the remaining aldehydic proton shift slightly upfield.

The peak at 8.72 ppm (i, Figure 30) represents the iminic proton being formed, and the peak appearing at 8.13 ppm corresponds to the benzyl CH₂ that in **2E_i** is next to the imine group.

Quantitative analysis will be carried out in a separate experiment in which IFR will be tested at different pH and therefore there will be more data to analyse the composition of the mixture and the amount of each one of the species in the equilibrium, but this experiment served to prove that IFR can take place under the reaction conditions and the equilibrium is reached relatively fast (<2h).

3.1.2 Influence of pH in IFR

IFR is a reaction in equilibrium. A molecule of water is formed as a side product, meaning that the reaction is naturally disfavoured in an aqueous environment (Figure 30).

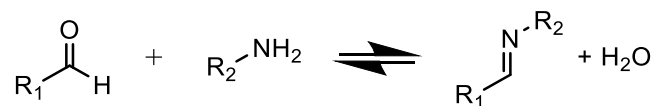


Figure 30: Imine formation reaction exchange under thermodynamic control.

Yet accepting that 100% of conversion may never be obtained in aqueous environments, due to the fact that IFR generates a molecule of water as a side product, it is crucial to maximise this value as much as possible as only appreciable amounts of product can cause a reliable and consistent template effect in DCC experiments. To this aim, a set of experiments were carried out to test IFR with a number of different BBs at different pHs from slightly acidic (pH 6.5) to highly alkaline ones (pH 10). 10% of acetonitrile-d₃ was needed in all cases to maintain the condition of solubility. The solvents of these reactions were completely deuterated and therefore the pH meter read the concentration of deuterium. As the acidity of deuterium (D) is not equal to the one of protons (H), values pD were converted to pH using the equation: pD = pH + 0.4.³⁰⁹

The different buffer compositions employed to afford the required pHs were the following. The preparation process will be discussed in the experimental chapter of this thesis (Chapter 7.3.1):

- pH 6.5 - 8: 100mM phosphate buffer. Prepared from KD_2PO_4 , K_2DPO_4 and D_2O .
- pH 8.5, 9: 100mM ammonium hydroxide buffer. Prepared from ND_4OD , D_2O and acetic acid- d_4 to adjust the pH.
- pH 9.5, 10: 100mM carbonate buffer. Prepared from Na_2CO_3 , D_2O , and acetic acid- d_4 to adjust the pH.

With this broad spectrum of pHs we studied how the system behaves when switching from physiological pH - which would likely be the optimal for target biomolecules in DCC experiments (i.e., sugars, glycans) - to alkaline pH - where amines are usually more nucleophilic and therefore, they are expected to be more reactive.

We must bear in mind that in order to favour the BA-diol complexation reaction occurring between the library members and the potential target, alkaline pHs are encouraged as the BA group must be in its tetrahedral form as discussed in the introduction chapter (Covalent receptors. Boronic Acids).^{59,154,310}

Some amines that we considered to be interesting to include in DCC experiments, according to the criteria already discussed in this thesis in the section dedicated to saccharide recognition, were tested. They were mixed with isophthalaldehyde **2** in pairs aldehyde + amine

and analysed by ^1H NMR. To keep track of the evolution of the reactions, resulting imine peaks (chemical shifts of 8.5 - 8.9 ppm on ^1H NMR spectroscopy) were integrated and the percentage of the species were calculated. As an example, the plotted spectrums corresponding to the reaction between **2** and **E** is shown (Figure 31).

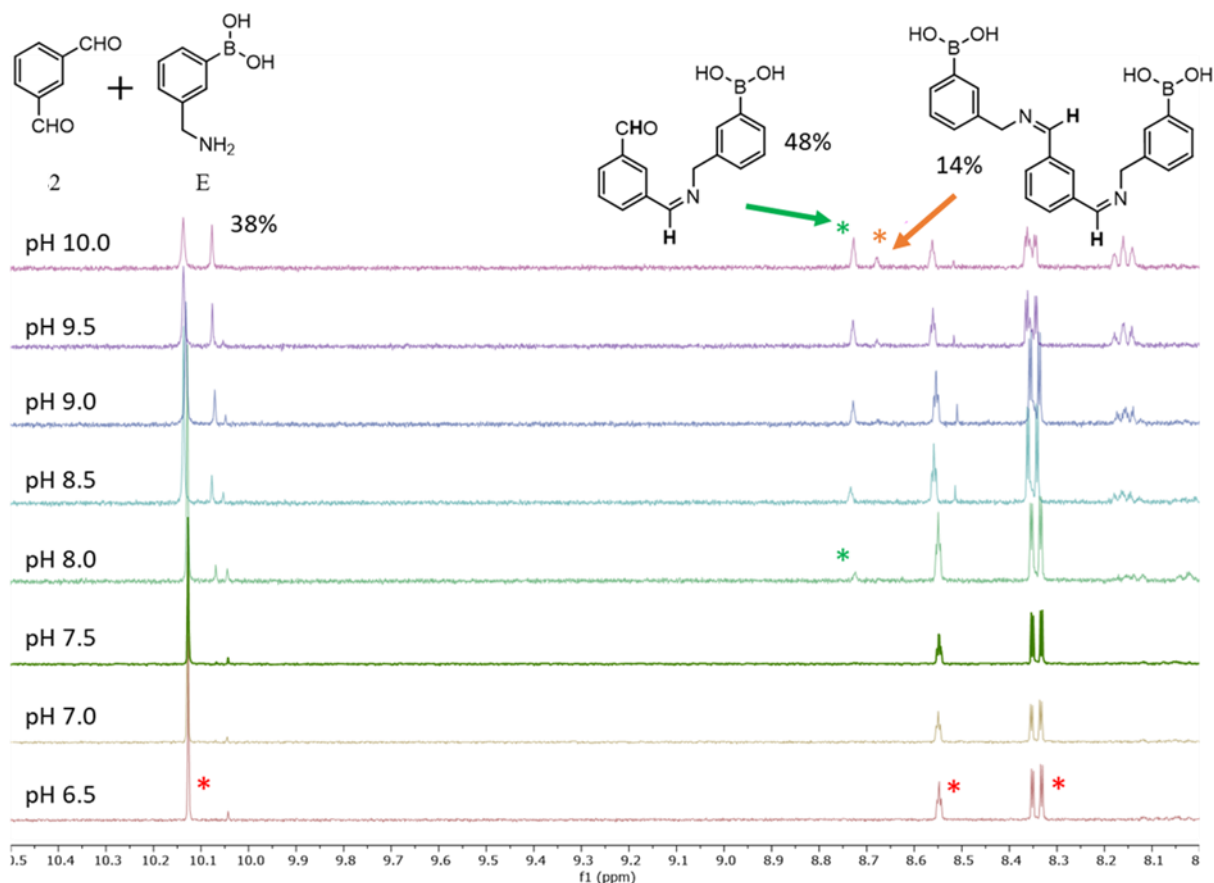


Figure 31: Plot of ^1H NMR spectrums corresponding to reaction between **2** and **E** in the range of pHs 6.5-10. Peaks flagged with a red star correspond to the starting aldehyde **2**. Green and orange ones correspond to the iminic protons of the mono and di-substituted products being formed, respectively.

It was evident that, the more alkaline the pH was, the greater the conversion obtained. As it will be demonstrated in future experiments, this seems to be a trend for most of aldehyde - amine combinations studied.

Interestingly, the di-substituted product **2E_iE_i** is visible only at pH 9.5 and pH 10 (8.67 ppm). The remaining of the peaks present in the spectrum were already identified in the previous experiment at pH 9 (Figure 30), except for a new small peak present at 10.05 ppm in spectrums at pH 8 and 8.5. That peak could not be identified but we understand that it must correspond to some intermediate and not to a final iminic product, as the peak is not found anymore at pH 9.5 or 10.

The percentages of the resulting products were calculated from the area under the peak (integral, I) of their iminic protons using Equation 3.1:

$$\% \text{ Product } i = \frac{I_i/nH_i}{\frac{I_{A2}}{nH_{A2}} + \frac{I_i}{nH_i} + \frac{I_{ii}}{nH_{ii}}} \times 100 \quad \text{Equation 3.1}$$

Where I_i is the Integral value of product i (mono-substituted product in this case), I_{ii} is Integral value of product ii (di-substituted product) and I_{A2} is the Integral value corresponding to remaining aldehyde 2. nH is the number of H assigned to corresponding peak.

For example, in case of reaction at pH 10, integrals of starting aldehyde (I_{A2}), mono (Ii), and di-substituted (Iii) products were 2, 1.27, and 0.74, respectively. The number of H behind the respective peaks are 2, 1, and 2. Inserting these values in Equation 3.1, the percentage of each of the species in the equilibrium are:

$$\% \text{ Mono-susbtituted product } \mathbf{2E_i} = \frac{1.27/1}{\frac{2}{2} + \frac{1.27}{1} + \frac{0.74}{2}} \times 100 = 48\%$$

$$\% \text{ di-susbtituted product } \mathbf{2E_iE_i} = \frac{0.74/2}{\frac{2}{2} + \frac{1.27}{1} + \frac{0.74}{2}} \times 100 = 14\%$$

$$\% \text{ Starting aldehyde } \mathbf{2} = \frac{2/2}{\frac{2}{2} + \frac{1.27}{1} + \frac{0.74}{2}} \times 100 = 38\%$$

At pH 10, only a 38% of the starting aldehyde remained unreacted. It had been formed a considerable 48% of the mono and 14% of the di-substituted products $\mathbf{2E_i}$ and $\mathbf{2E_iE_i}$. LC-MS experiments carried out in a later stage confirmed the formation of these products.

The same experimental protocol and data analysis was then applied to a number of BBs that met the criteria to be part of our DCL (Figure 32). These requirements will be explained in more detail in a separate chapter of this thesis, but the most fundamental ones are that the BBs must contain an amine group so they can react in a reversible manner with aldehyde $\mathbf{2}$, they have to be functionally diverse, and remain in solution under the DCC reaction conditions. The ^1H NMR spectra of these combinations with the calculated percentage of each species are included in Annex 1. The results are included in Table 4, and represented in Figure 33. These results are also classified in different groups attending to the nature of the starting amine employed (-).

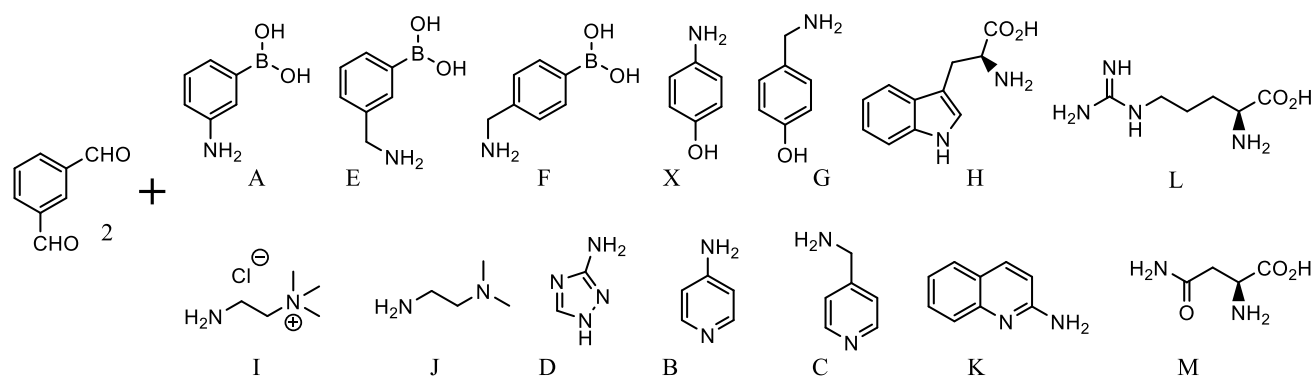


Figure 32: List of BBs to be studied on this chapter.

Table 4: Conversion (%) to imine (mono + di-substituted products) for each one of the combinations and pHs tested. NMR Experiments were conducted once.

Reactions	pH							
	6.5	7	7.5	8	8.5	9	9.5	10
2+A	29	29	29	29	31	31	37	41
2+D	0	0	0	0	0	0	0	0
2+B	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
2+K	0	0	0	0	0	0	0	0
2+X	-	-	-	-	-	-	-	-
2+E	0	0	0	25	35	42	54	62
2+F	0	0	0	33	37	48	63	63
2+G	0	0	0	0	26	38	59	66
2+H	0	0	0	0	0	0	56	67
2+C	-	-	-	-	-	-	-	-
2+I	41	47	74	76	73	68	34	25
2+J	19	35	40	47	71	66	50	53
2+L	19	35	40	47	71	66	50	53
2+M	19	35	40	47	71	66	50	53

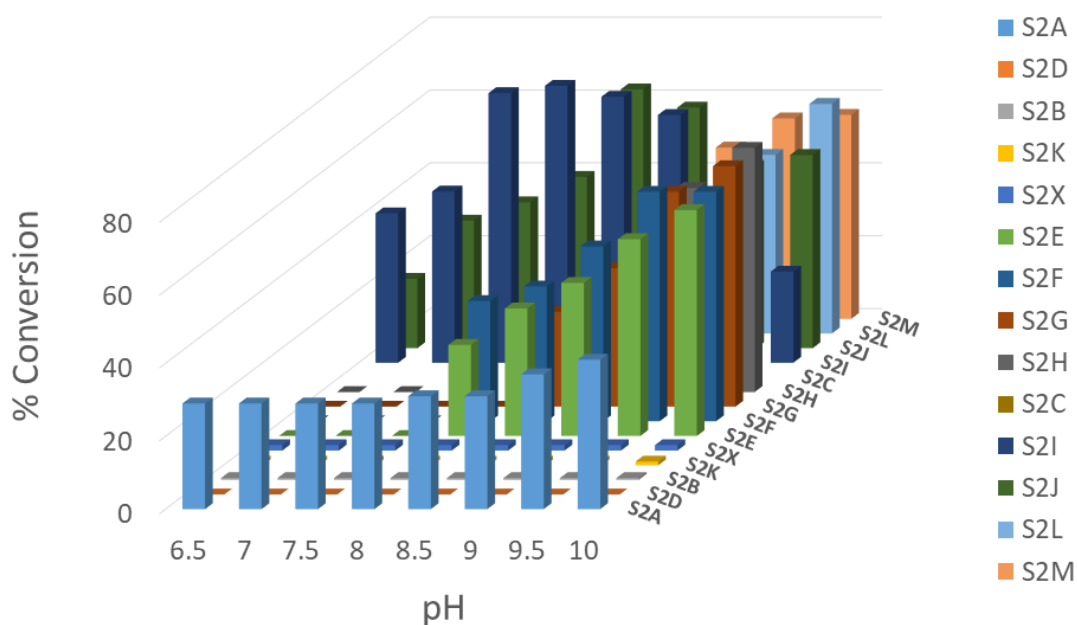


Figure 33: Bar chart showing conversion values (mono + disubstituted products) for IFR between aldehyde 2 and each one of the amines studied. Results showing a dash '-' could not be performed due to solubility reasons.

Aromatic amines

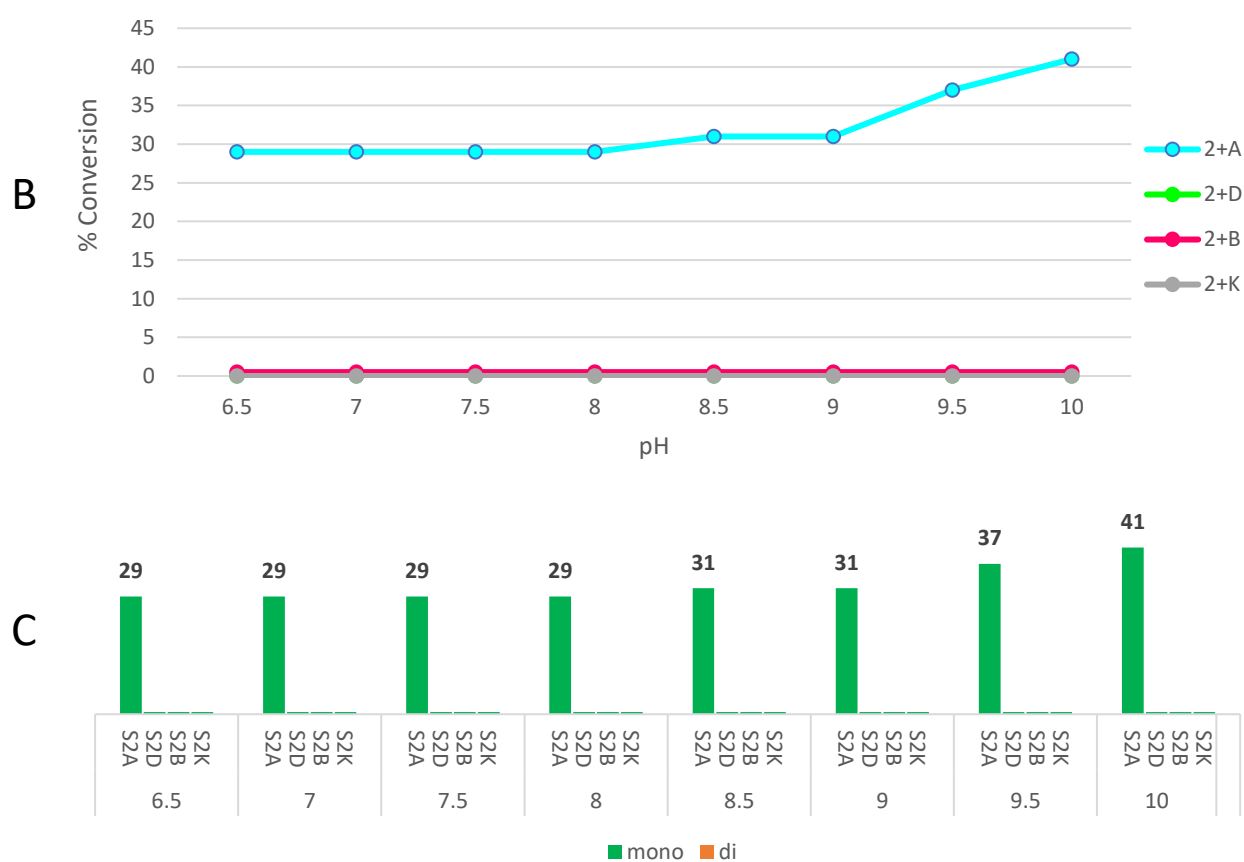
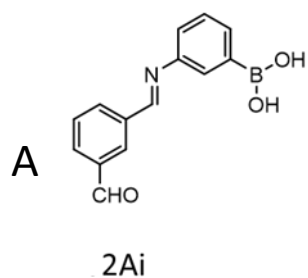


Figure 34: Reaction products being formed in the group 'Aromatic amines' (A). Line graph showing conversion values (B), and bar chart showing the same values but also indicating the percentage of mono and di-substituted (when applicable) products being formed (C). 'S' before the name of the molecule stands for 'sample'.

Aliphatic amines

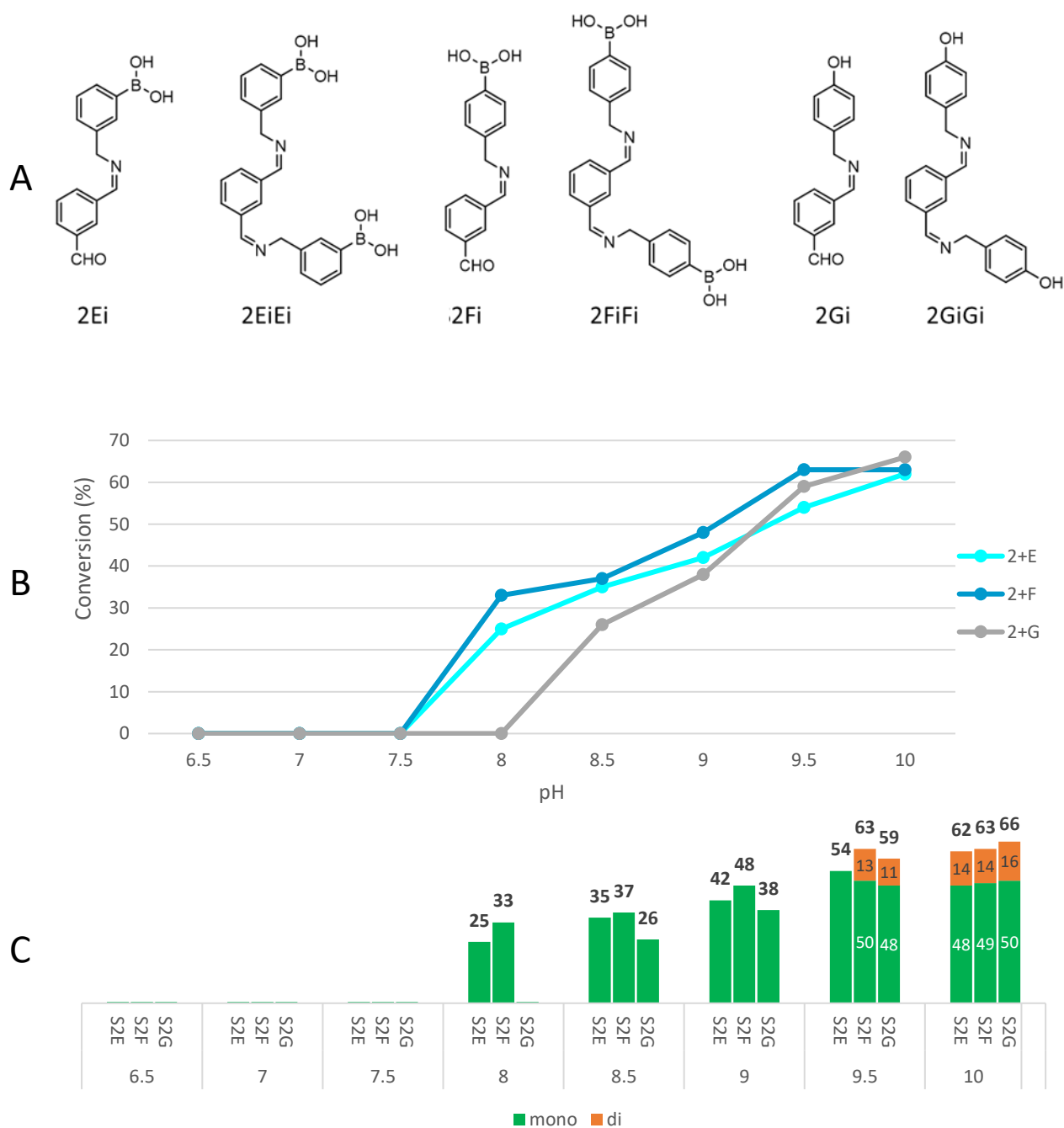


Figure 35: Reaction products being formed in the group 'Aliphatic amines' (A). Line graph showing conversion values (B), and bar chart showing the same values but also indicating the percentage of mono and di-substituted (when applicable) products being formed (C). 'S' before the name of the molecule stands for 'sample'.

Amino acids

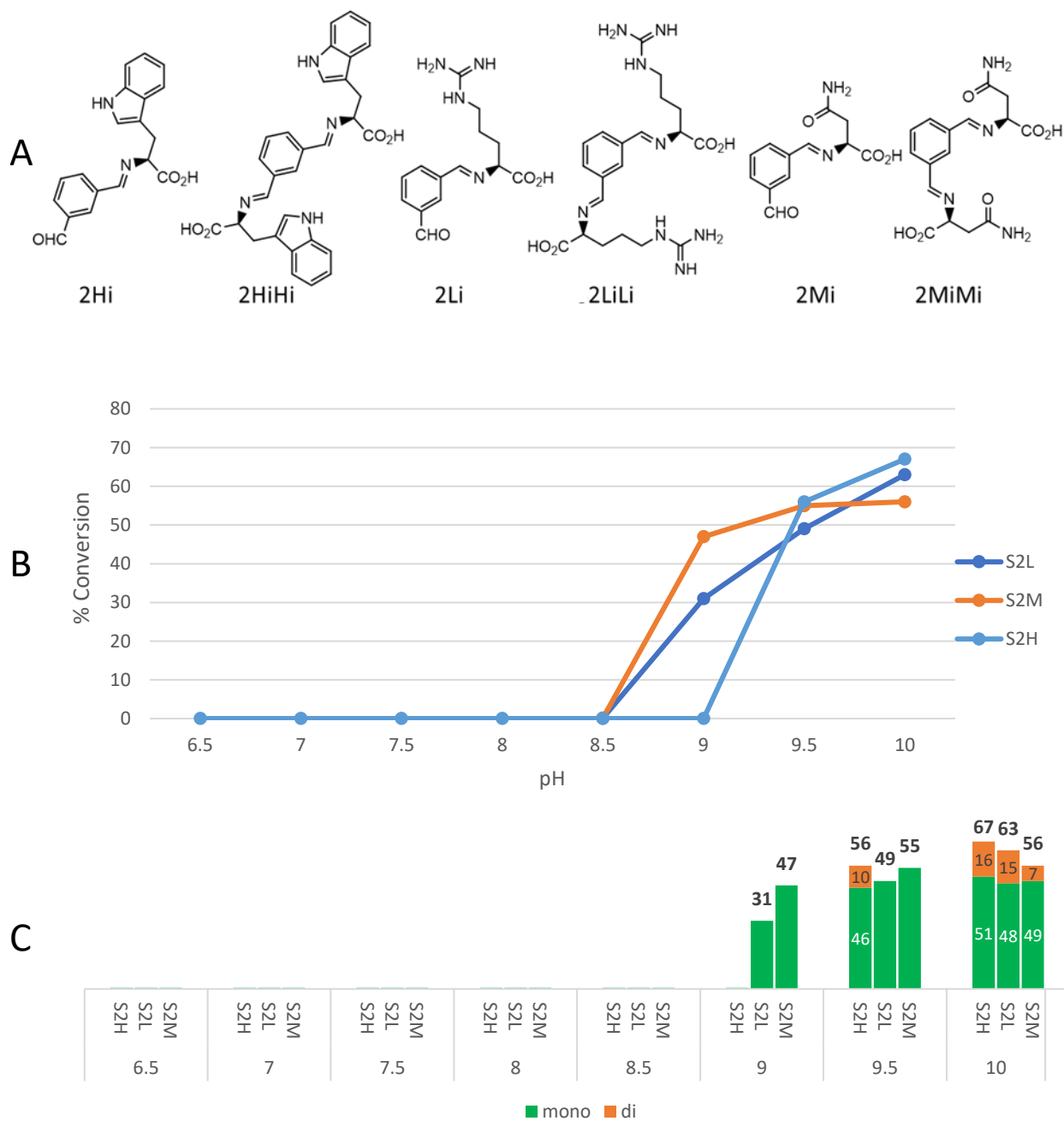


Figure 36: Reaction products being formed in the group 'Amino acids' (A). Line graph showing conversion values (B), and bar chart showing the same values but also indicating the percentage of mono and di-substituted (when applicable) products being formed (C). 'S' before the name of the molecule stands for 'sample'.

Linear amines

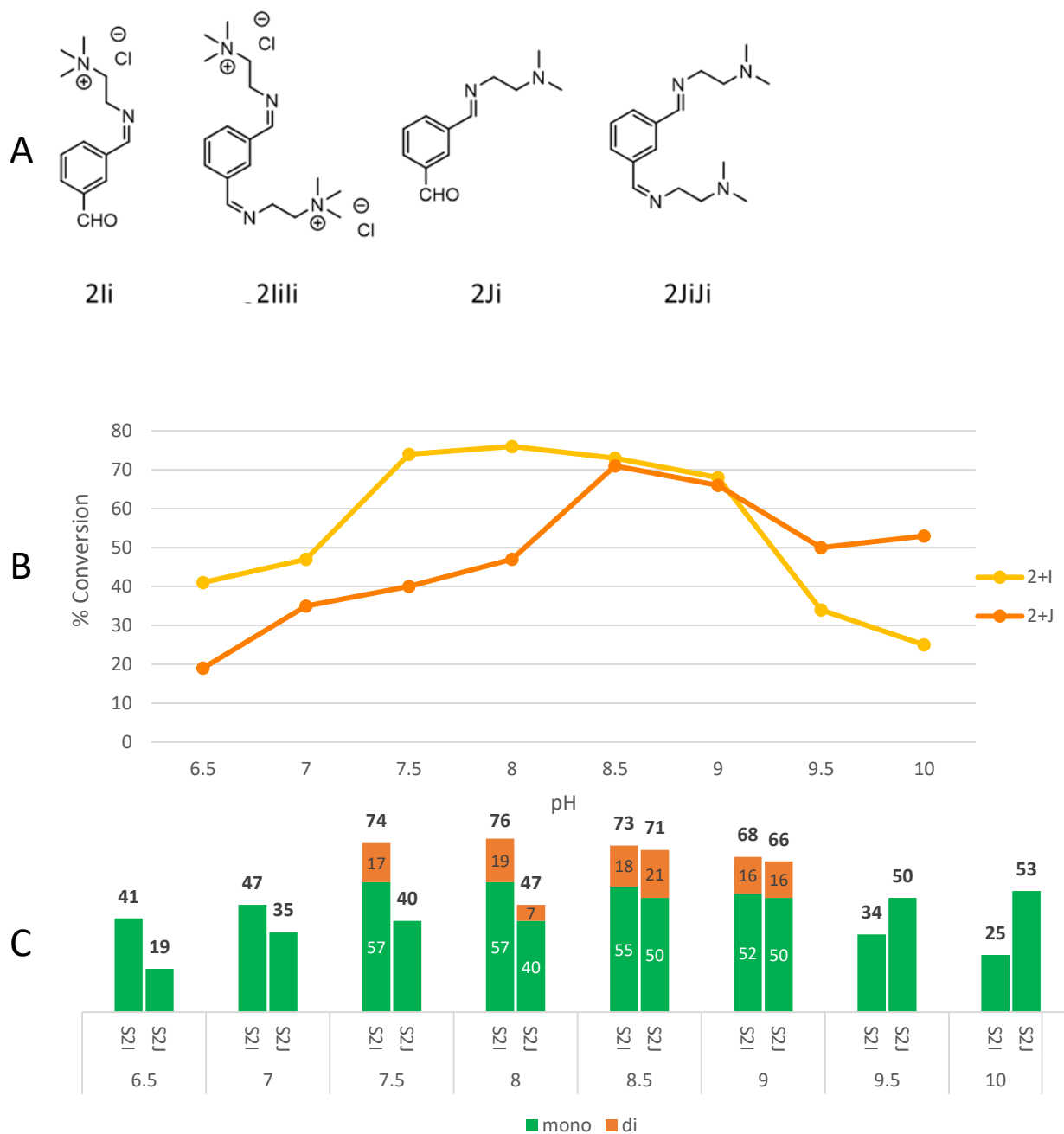


Figure 37: Reaction products being formed in the group 'Linear amines' (A). Line graph showing conversion values (B), and bar chart showing the same values but also indicating the percentage of mono and di-substituted (when applicable) products being formed (C). 'S' before the name of the molecule stands for 'sample'.

Aromatic amines did not perform very well (). In fact, only **A** formed appreciable amount of imine (maximum of 41% at pH 10). Di-substituted product was not formed. This can be explained by the poor nucleophilicity of amines directly attached to aromatic rings. **X** was not completely soluble and therefore had to be ruled out.

Aliphatic amines, more nucleophilic, gave better results (). They too reached their higher conversion (62-66%) at pH 10. **C** was not fully soluble in the buffer, so it was not considered. Di-substituted products were not formed below pH 9.5, and the highest yields reached were 14-16% at pH 10 in all the three cases.

Amino acids tested, as the aliphatic amines that they are, behaved similarly (). The best one was **H** (L-Tryptophan, 67%), followed by **L** (L-Arginine, 63%), and **M** (L-Asparagine, 56%). **H** was the only forming disubstituted product at pH 9.5. They all formed the di-substituted product at pH 10 to similar extents as the other aliphatic amines (7-16%).

Linear amines **I** and **J** are also aliphatic. They were just grouped separately as they are smaller molecules, and they lack aromaticity (which may have implications in solubility and reactivity). They escaped to the trend, and they reached excellent conversion rates of 76 and 71% at pH 8 and 8.5, respectively, to then decrease to lower values at pH 10. Di-substituted products were formed in the range from pH 7.5 to 9, reaching the highest conversion rates in the whole set of experiments with 19-21% (). While the higher overall reactivity of amines **I** and **J** can be explained by their greater nucleophilicity (they are smaller and therefore they offer less steric

hindrance), we cannot find an explanation for the decrease in reactivity at pH > 8 and pH > 8.5 for **I** and **J**, respectively.

3.2 Conclusions of Chapter 3

To summarise the chapter dedicated to the optimisation of IFR we can conclude:

Imine formation reaction between aldehyde **2** and amine **E**, in alkaline medium (100mM ammonium hydroxide buffer pH 9 90%-10% MeCN) reaches the thermodynamic equilibrium in less than 2h. While it is likely that the equilibration time would be similar for the rest of combinations aldehyde + amine considered in this thesis, it was decided to fix the reaction time to O.N (approximately 24h) for future DCC experiments to ensure the condition of equilibrium.

The equilibrium state was never shifted towards the complete formation of imines with any of the amines tested (**A**, **B**, **C**, **D**, **E**, **F**, **G**, **H**, **I**, **J**, **K**, **L**, **M**, **X**) and in any of the pH tested (6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10.0). Some amines can simply not be considered due to solubility reasons (**C** and **X**) and quantitative formation of products cannot be achieved, most likely due to the presence of water in the medium promoting the hydrolysis of imines and holding back the conversion rates. Nevertheless, some factors proved to be critical in the yield of the reaction. Conversion rates were a direct function of the nucleophilicity of the starting amine in the order: aromatic amines < bulky aliphatic amines < linear aliphatic amines.

Moreover, in concordance with results published by others,³¹¹ we proved a strong dependency of IFR with pH. For most cases, the more alkaline the pH was, the higher the conversion with rates obtained ranging from 0% to 76%. This is a consequence of the decrease in the nucleophilicity of the starting amine at more acidic pHs. Amines **I** and **J** were an exception to this trend, as they reached their highest conversion rates at pH 8 and 8.5 respectively. These conversion rates (76% and 71% respectively) were the highest observed throughout this experiment for any combination aldehyde-amine at any pH tested. While these high conversion rates can be explained by the fact that amines **I** and **J** are the most nucleophilic ones among the set, we fail to find an explanation for the decay in conversion rates at more alkaline pHs that these amines showed, in contradiction with the results exhibited for the rest of amines.

Formation of di-substituted products was always poor. 10-20% of conversion to these molecules was the highest we could observe for the best amines (most nucleophilic ones) at their optimal pH. Steric and entropic factors may be playing against the formation of these products but thankfully, as per the aims and objectives of this research, that amount of multivalent products might be sufficient as we only need enough of them in the Dynamic Combinatorial Library (DCL) so they can be quantified by the analytical instrument of our choice.

Given these results, amines **E**, **G** and **H** were selected to further develop the DCC protocol. They have similar and acceptably good behaviour at pH 10, and they will enrich the system with a broad array of functionalities (BAs, polar groups, proton-donor and proton-accepting groups, and aromatic systems). This will consequently increase the probability of finding good receptors for the given biomarkers.

Chapter 4 -DCL- Small molecule approach

DCC is undoubtedly the core of this work. This chapter will investigate the optimal conditions to perform DCC experiments. Special attention will be placed on the challenges of analysing DCLs by LC-MS and the criteria used to select the BBs. First, a small-scale DCC 1.0 experiment was successfully carried out employing small molecules as BBs. Then, a more complex DCC 2.0 experiment was accomplished as well employing simple monosaccharides as templates. Some molecules were clearly amplified in the experiments, suggesting that they could be the best receptors for such templates within the DCL.

4.1 DCC Method development

4.1.1 LC-MS method development

In a Dynamic Combinatorial Library (DCL), a large number of different species may be present. This number increases rapidly with the number of starting BBs, the number of reacting positions they have, and the possibility of isomers being formed. The complexity of such libraries often makes their analysis a difficult task, and it could become a bottleneck or the limiting step in the design of a DCC experiment.

DCLs will be screened by LC-MS as it is an effective way to not only isolate but also identify and quantify each one of the library members.

In order to carry out the method development, we designed a pilot DCC experiment, or DCC 1.0, with only one starting aldehyde, **2** (isophthalaldehyde), and three amines (**E**, **G**, **H**) so the total library size would not be too large (13 members), simplifying its analysis.

A priori we did not know much in advance about the mechanism that would afford the desired affinity and selectivity. All we could do in advance was to provide the system with enough tools or recognition units so it could self-assemble the most powerful receptor available. Therefore, we incorporated into the system the following entities:

BA groups, capable of reversibly and covalently bind 1,2 and 1,3 cis diols.⁵⁹ While glycans (and specially saccharides) are rich in OH groups and even diols, not all of them meet the geometrical criteria for them to react with the BA group and therefore this might be a key selective tool. Amine **E** (Figure 38) incorporates this feature.

Polar groups capable of creating H bonds. While these are weak interactions, they are often numerous and the sum of them can become a major force. Saccharides are heavily solvated in water which will weaken the contribution of these interactions but in combination with other energies they can play an important role. In fact, since the early 90s is well known their importance in molecular recognition,³¹² and most of natural and synthetic receptors for biomarkers rely, at least partially, on this interaction. Primary and secondary amines present in all the recognition units, the hydroxyl group in **G**, and the carboxylate group in amino acid **H** (negatively charged at the reaction pH) (Figure 38) can cause this effect.

Hydrophobic groups that can bring π -H interactions. In an aqueous environment, the ability to create hydrophobic pockets for non-polar molecules (or at least non-polar parts of the molecule) is a precious feature.³¹³ This can benefit the aim of selectivity as

saccharides show many isoforms in solution (tautomers) and there might be only some of them that can fit in such pockets. Hydrophobic groups are routinely found as well in synthetic and natural receptors, often in the form of benzene rings. Since the scaffold **2** contains an aromatic ring, most of the library members will count with this feature. **E**, **G**, and **H** (Figure 38) incorporate another ring so on the DCL there will be members with 0, 1, 2, and 3 of these motives.

None of these groups by themselves may be able to afford effective and most importantly selective binding, but the summatory of them -phenomena named multivalency- could be the key to it.

Amines **G** (4-Hydroxybenzylamine), **E** (3-Aminomethylphenylboronic acid), and **H** (L-Tryptophan) meet most of these criteria and also performed well -i.e., they gave high conversion rates in IFR with **2**- in the NMR studies carried out in the previous chapter.

2 bears two aldehyde groups, meaning that there are two reactive positions for IFR to occur. Symmetry operations will restrict the formation of isomers ($2EG = 2GE$). There are therefore 9 different products that can be formed that, together with the 4 starting building blocks, add up to 13 different compounds that can be present in the library (Figure 38).

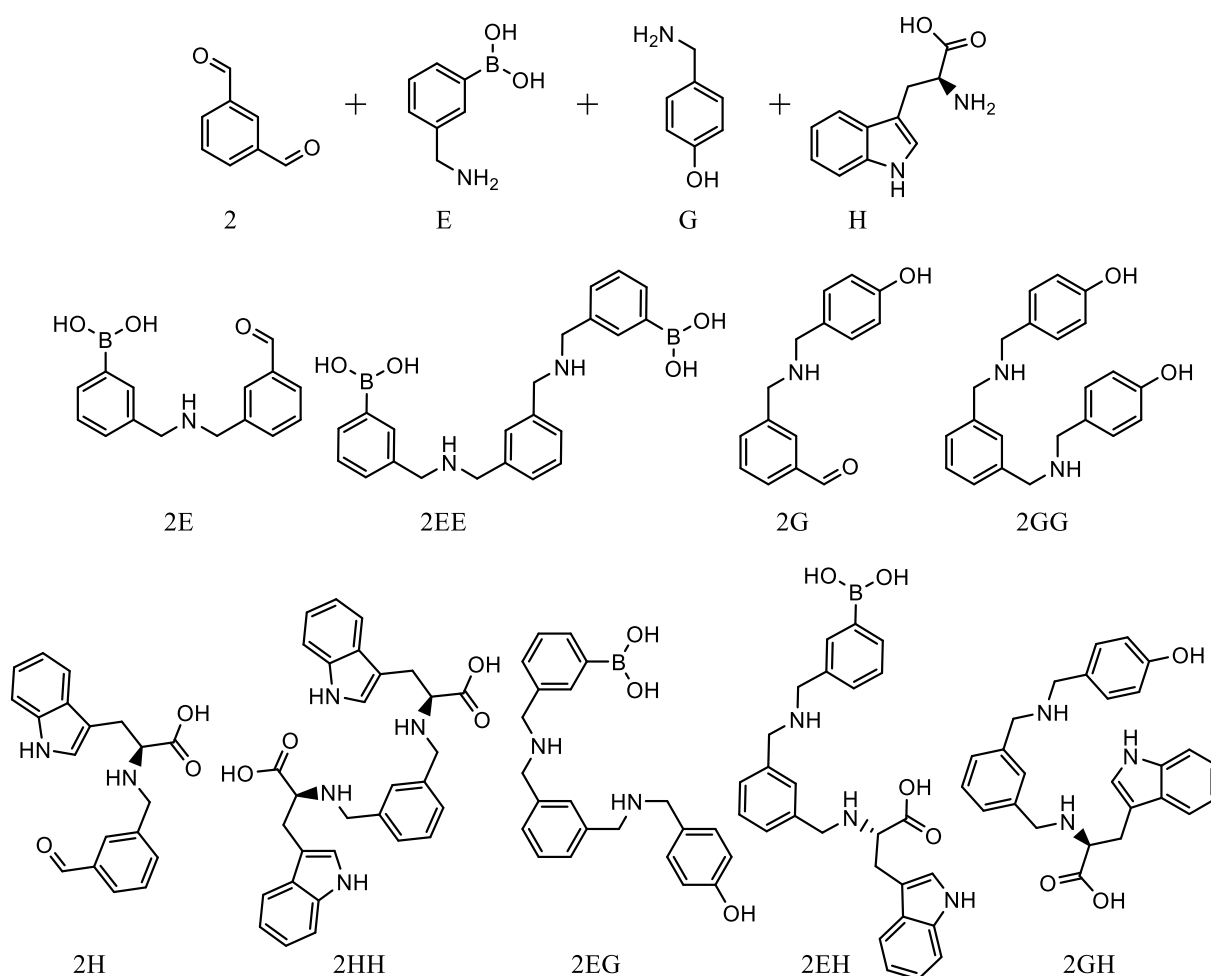


Figure 38: Library members in the experiment DCC 1.0.

An LC-MS instrument consists of a liquid chromatography (LC) apparatus that allows the separation of the analytes in a mixture according to their physicochemical properties (polarity, size...) connected to a mass spectrometry detector (MS) that ionize the eluted compounds and detect their mass to charge (m/z) ratio.

The protocol that will be followed throughout this thesis to examine and screen DCLs by LC-MS will be the following: the samples will be injected in the LC-MS instrument. Its constituent analytes will be retained in the chromatography column with a strength according to their

affinity for the stationary phase of the column. Depending on the effectiveness of the column and the running conditions employed, all the analytes in the mixture may or may not be isolated (resolved). The eluent coming off the column will then be ionised in the MS chamber, and analytes will be characterised according to their m/z profile.

Effectively running LC-MS requires previous method development of both pieces of equipment. While the mass spectrometer functioned well with common default settings, the LC instrument required much deeper study and consideration.

The key component of any LC system is the chromatography column. We first tried the most commonly employed C18 columns. They are known as all-purpose columns as they normally work well to isolate organic compounds. However, they did not work well for us as they could not isolate many of the library members under any running conditions tested.

Our library consists of molecules with a wide array of chemical groups such as primary and secondary amines, carboxylic acids, or aromatic structures. This makes challenging the complete isolation of their components and a comprehensive study needed to be carried out in order to find the most suitable chromatography column.

Columns may differ in size (length and width), packing material and its particle size, being the material the most crucial factor. A set of columns with different packing materials were tested.

Running conditions were based on those found in the literature and further optimised according to the needs of each particular case.

1) C18 columns.

Octadecyl groups retain non-polar compounds mainly through hydrophobic interactions.

Available with many different modifications to cover a wider range of uses. If not stated otherwise, running solvents and conditions were the following gradient:

Solvent A: Water + 0.1%FA

Solvent B: MeCN + 0.1%FA

100%A for 1 min, then to 5%A over 10 min.

1a) Raptor C18 2.7 μm RP (150x2.1 mm). A standard chromatography column.

Different compositions of mobile phases and running conditions were tested. From highly polar mixtures (100%A to 80%A over 10 minutes) to highly non-polar ones (20%A to 5%A over 10 minutes), and a range of different compositions in between were tested but none of them was able to isolate most of the analytes. It was concluded that a different column with another stationary phase was required.

1b) Kinetex Core-Shell Polar C18 2.6 μm 100A (50x2.1 mm). Core-shell based particle, designed to provide increased efficiency compared to traditionally fully porous particles, with C18 alkyl phase and a polar modified surface to gain retention for polar compounds.

The level of separation obtained was not adequate, with 3 pairs of the library members coeluting, i.e., when the protocol detailed above for the analysis of the DCL by LC-MS was followed, the m/z of three pairs of compounds (**2EG + 2G**, **2EE + 2E**, and **2EH + 2H**) were found under the same peak, at the same retention time (Figure 39).

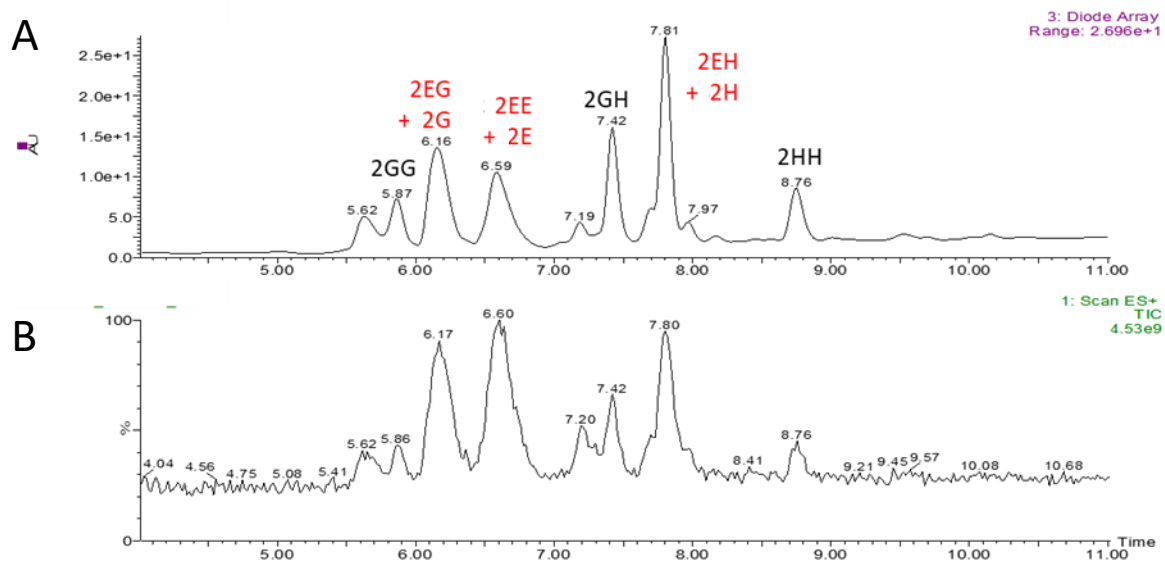


Figure 39: Distribution of the DCL members in the LC-MS chromatogram when Kinetex Polar C18 column and conditions 1b were employed. UV absorbance chromatogram at the top (A), MS response at the bottom (B). Only a partial representation of the chromatogram is shown (minute 5 to minute 11). Starting BBs not seen as they eluted before minute 5. For schematic representation of molecules labelled here go to Figure 38.

1c) Kinetex XB C18 Core-Shell 2.6 μm 100A (50x2.1 mm). C18 alkyl chains and Di-isobutyl side chains with slightly modified core-shell particles to make the column a great hydrogen acceptor. Designed to improve peak shape and retention times for acidic compounds.

The level of separation obtained was good, but not the best, with **2EE** and **2G** being eluted together and **2GG** and **2EG** partially coeluting (Figure 40).

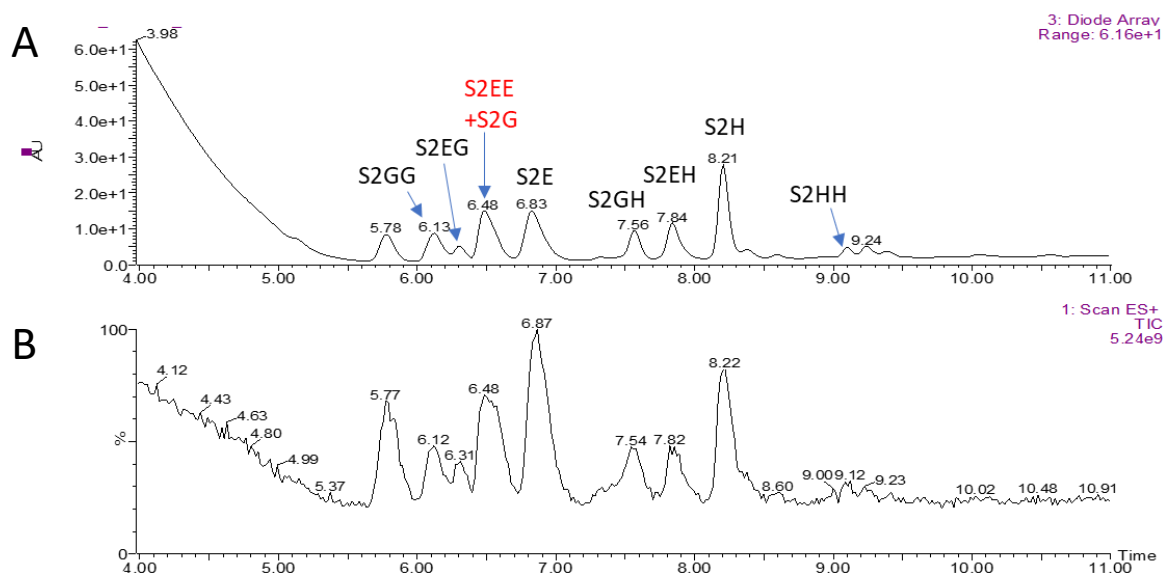


Figure 40: Distribution of the DCL members in the LC-MS chromatogram when Kinetex XB C18 column and conditions 1c were employed. .UV absorbance chromatogram at the top (A), MS response at the bottom (B). Only a partial representation of the chromatogram is shown (minute 5 to minute 11). Starting BBs not seen as they eluted before minute 5. For schematic representation of molecules labelled here go to Figure 38.

1d) Kinetex XB-C18 - long 5 μ m 100A (250x4.6 mm). Same packing than 1c) but larger particle (5 μ m vs. 2.6 μ m) and overall size (250x4.6 mm vs. 50x2.1 mm) to improve peak separation.

Identical result as with column 1c.

In conclusion none of the C18 columns provide the necessary separation. In all the cases, a number of molecules were eluted off the column together.

2) Pentafluorophenyl (PFP) columns.

Silica modified with PFP groups to provide a better retention for electron-rich compounds such as amines or alcohols, which could be beneficial considering the presence of these groups in many of the library members.

2.a) Raptor FluoroPhenyl (PFP) 2.7 μm (50x3.0 mm). It can work both in Reverse Phase (RP) – as usual, non-polar compounds retained the most – and in HILIC mode – polar compounds retained the most –.

2.a.1) HILIC mode.

A: Water+0.1%FA; B: MeCN+0.1%FA. 5-60%A over 10 min.

2.a.2) Reverse Phase (RP) mode.

A: Water+0.1%FA; B: MeOH+0.1%FA. 95-5%A over 10 min. Ammonium Acetate 0.1 M pH 7 instead of water tried as well.

For both 2a.1 and 2a.2 most of the analytes coeluted in the first 2 minutes. This indicates no retention at all of any of the analytes as 2 minutes is roughly the 'dead time' or the time that it takes for the eluent solvent to go through the column. No retention of any of the compounds is rare so the column was probably malfunctioning.

2.b) Kinetex F5 2.6 μm 100A (50x2.1 mm). Modification of standard PFP column. The manufacturer claims that they provide five interaction mechanisms: hydrophobic, aromatic, electrostatic, steric, and hydrogen bonding to provide retention for all types of compounds.

A: Water+0.1% FA; B: MeOH+0.1%FA. 100%A for 1 min, then switch to 5%A over 10 min.

The level of separation obtained was not adequate as two pairs of compounds (**2EG** + **2G**, and **2EE** + **2E**) coeluted (Figure 41).

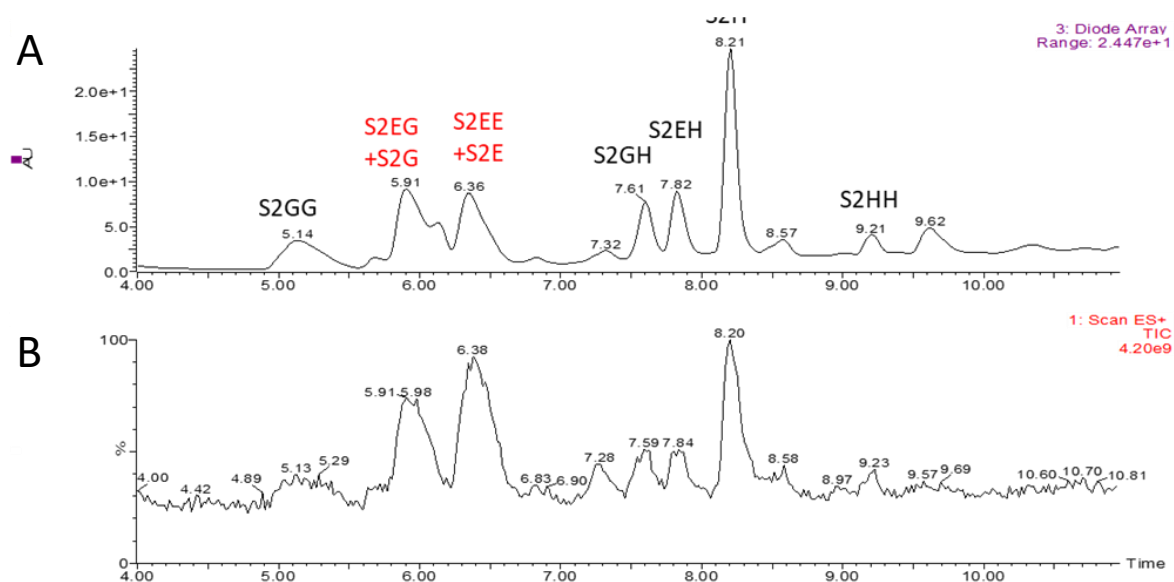


Figure 41: Distribution of the DCL members in the LC-MS chromatogram when Kinetex F5 column and conditions 2b were employed. UV absorbance chromatogram at the top (A), MS response at the bottom (B). Only a partial representation of the chromatogram is shown (minute 5 to minute 11). Starting BBs not seen as they eluted before minute 5. For schematic representation of molecules labelled here go to Figure 38.

3) Aromatic columns.

Packing containing aromatic rings to provide not only hydrophobic forces, but also π - π interactions with the analytes. We suspected these could work for us due to the high abundance of aromatic structures in the library composition. Running solvents and compositions were the following:

A: Water+0.1%FA; B: MeOH+0.1%FA. 100%A for 1 min, then switch to 5%A over 10 min.

3.a) Raptor Biphenyl 2.7 μm (50x2.1 mm). Biphenyl rings in the packing cause the retention of aromatic structures present in the analytes.

3.b) Kinetex Phenyl-Hexyl - long 5 μm 100A (250x4.6 mm). Phenyl-Hexyl modification causes slightly different interaction than Biphenyl rings.

3a and 3b, performed similarly but 3b acquired the best separation overall with only 2 products (**2EE** and **2E**) coeluting and all the others perfectly resolved (Figure 42).

Kinetex Phenyl-Hexyl column and conditions 3b will therefore be employed in future DCL screening analysis.

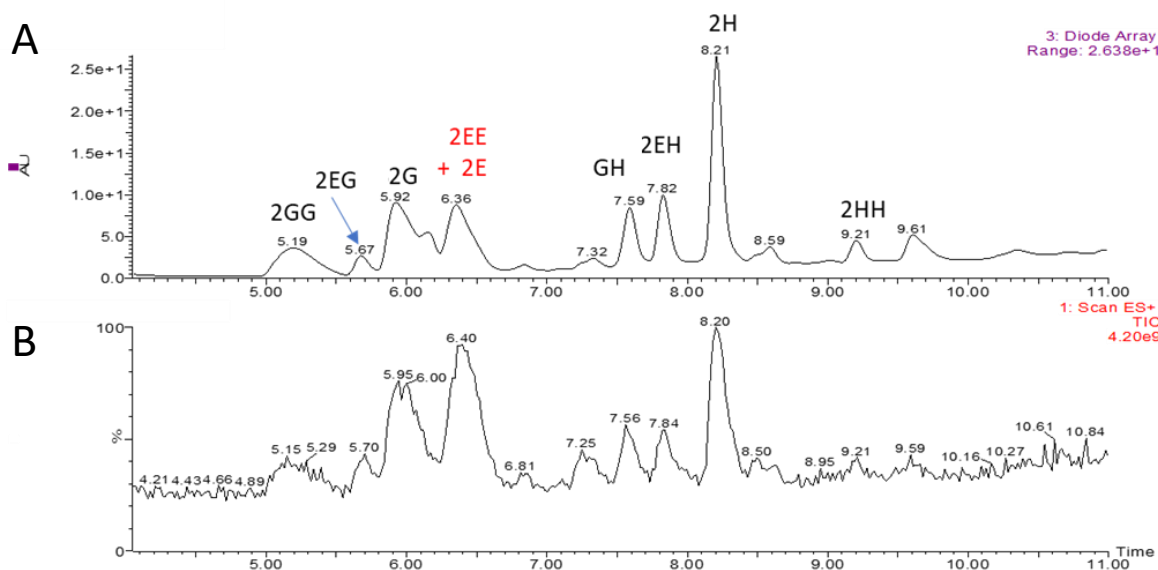


Figure 42: Distribution of the DCL members in the LC-MS chromatogram when Phenyl-Hexyl column and conditions 3b were employed. UV absorbance chromatogram at the top (A), MS response at the bottom (B). Only a partial representation of the chromatogram is shown (minute 5 to minute 11). Starting BBs not seen as they eluted before minute 5. For schematic representation of molecules labelled here go to Figure 38.

4.1.2 Conclusions and remarks on LC-MS method development

When comparing the peak size of different compounds in , Figure 41, and , and , it is remarkable how distinct they are among them. **2H** showed peak sizes, especially in UV chromatograms, 2 to 3 times larger than its competitors. Yet peak size is directly related to the UV absorbance (for UV chromatogram) and ionizing properties (for MS chromatogram) of each compound, it is also a direct function of the concentration of each molecule in the mixture.

In our library, UV and MS response should not be that disparate among the members, so we believe these differences are a signal of their unequal distribution in the equilibrium. In other words, library members are present in the equilibrium with different concentrations.

BBs were selected as they performed similarly in previous NMR experiments. At pH 10, the combination of aldehyde **2** and amine **E**, **G**, and **H** yielded percentages of imines of 62, 66, and 67% with similar distribution of mono and di-substituted products. However, those experiments tested individual reactions **2** + one single amine at a time. LC-MS experiments concluded that in a competitive environment where amines must compete for a limited number of aldehydes, preconceived similar reactivity does not apply. Nevertheless, as long as there is enough amount of each library member in the mixture so we can measure their peak areas, this uneven distribution should not be an issue and DCC experiments can be carried out.

Moreover, we learnt that columns containing aromatic structures as stationary phase (packing) are the most suitable for our DCL. Among other factors, phenyl groups of the

stationary phase interact via π - π bonds with the aromatic entities present in the analytes, affording an effective separation. Therefore, further studies will be carried out with these types of columns and particularly with Kinetex Phenyl-Hexyl column (column 3b).

4.1.3 DCC 1.0 – method development

In the chapter dedicated to the optimization of IFR, parameters such as the concentration of reagents, solvent composition and reaction time were already discussed. Another key factors to consider in order to perform a successful DCC experiment are reaction temperature and concentration of template.

Most DCLs reported in the literature were formed at room temperature. In general, the reversible exchange is efficient under these conditions and the equilibrium is reached quickly enough. However, in some cases, due to the instability of the template employed (often proteins) it was necessary to perform the experiment at low temperatures.³¹⁴ In DCC 1.0 and the rest of DCC experiments described in this thesis, temperature will be set to room temperature as we demonstrated by ^1H NMR that the equilibrium can be reached fast (<2h), and the templates to be employed are perfectly stable.

Regarding the concentration of template needed, there are conflicted ideas in the literature. While some authors claim substoichiometric concentrations are needed to ensure the competition between the DCL species,³⁰⁸ other researchers have used comparative concentrations or even excess of template to design their DCC experiments.²⁶¹

The optimal concentration of template may depend on the specific system being studied and the desired outcome. It will be influenced by factors such as the library size, or the strength of the interaction receptor-template. Since this last parameter is difficult to predict in advance, the optimal concentration of template may need to be determined empirically for each particular case.

Hence, we decided that we would investigate this parameter ourselves employing D-(+)-Glucose (glucose from now on) as template.

The concentration of starting reagents was previously fixed to be 1 mM for starting aldehyde and 20 mM for the sum of amines, meaning that the concentration of potential receptors would be in the range of 0-1 mM. Therefore, we decided to investigate the use of 0.1 mM, 2 mM, and 20 mM of template to cover three different scenarios. A schematic representation of the DCC protocol can be seen in Figure 43.

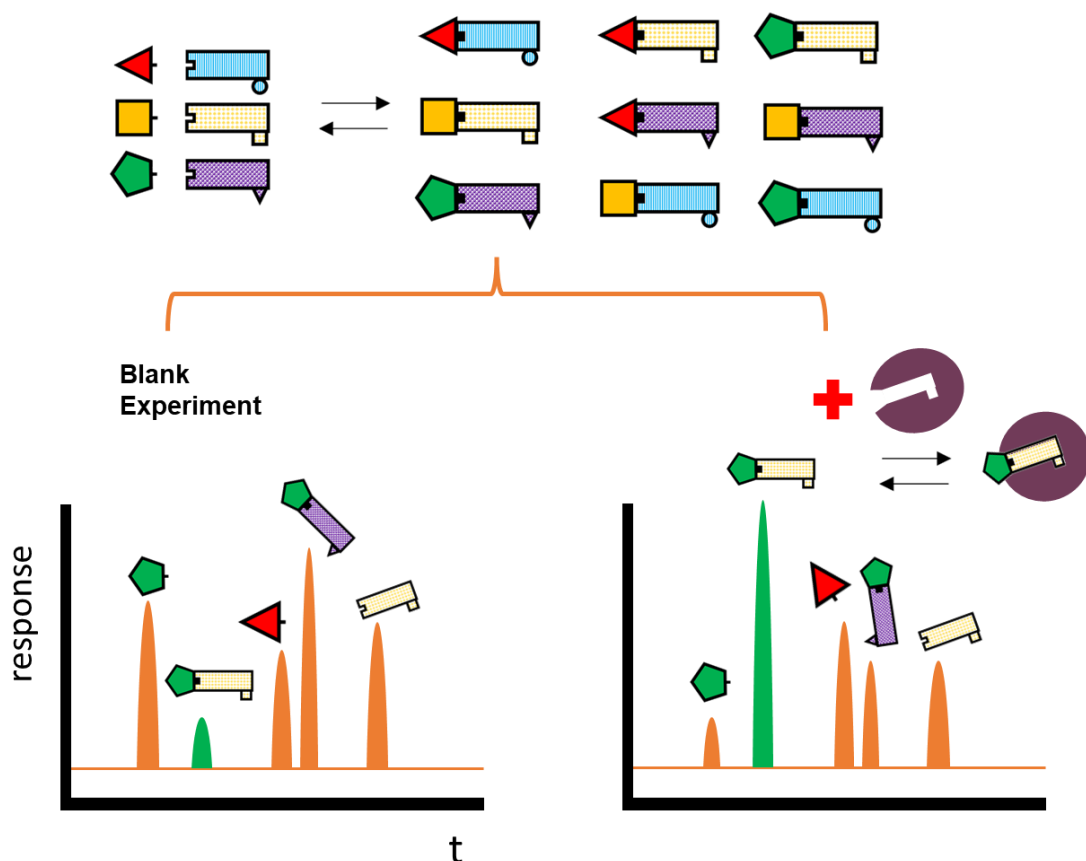


Figure 43: Schematic representation of DCC experiments protocol. The mixture of starting BBs (part of keys) afforded a library of compounds in equilibrium with the starting materials. Upon addition of a template, the best binder for the template stands out (when comparing with a blank experiment) in LC-MS analysis.

The area under the peak of all the species was quantified in the MS chromatograms. The division of that value in an experiment with presence of glucose and the value in the blank experiment was utilised to calculate the amplification value (A). Therefore $A > 1$ means a positive amplification (i.e., an increment in the concentration of the molecule in the presence of the given template) while $A < 1$ corresponds to negative amplification (i.e., the concentration of the molecule decreases in the presence of the given template). Negative amplification

values occur because library members have to compete for the limited molecules of scaffold **2**, which is in shortage (1 mM) with respect to the starting amines (20/3 mM each). Therefore, amplification of some library members (those with the highest affinity for the template) can only occur to the detriment of others. The resulting A values for the reaction **2** + **E**, **G**, **H** are shown in Figure 44. For clarity, an example of the process carried out to calculate and plot A values can be found in Annex 2.

The most reliable results were obtained when employing comparative amount of glucose as template (2 mM) and potential receptors (up to 1 mM) (Figure 44). When using 0.1 mM of glucose, no template effect could be observed (all A values close to 1). When employing 20 mM of glucose, the results were random, mostly negative, and did not seem to follow any trend (Annex 3). A possible explanation could be that the high concentration of template could induce the formation of aggregates that could either form precipitates or interfere with the formation of ions in the ionization chamber of the MS instrument. Therefore, from now on we will be employing 2mM of template in future DCC experiments.

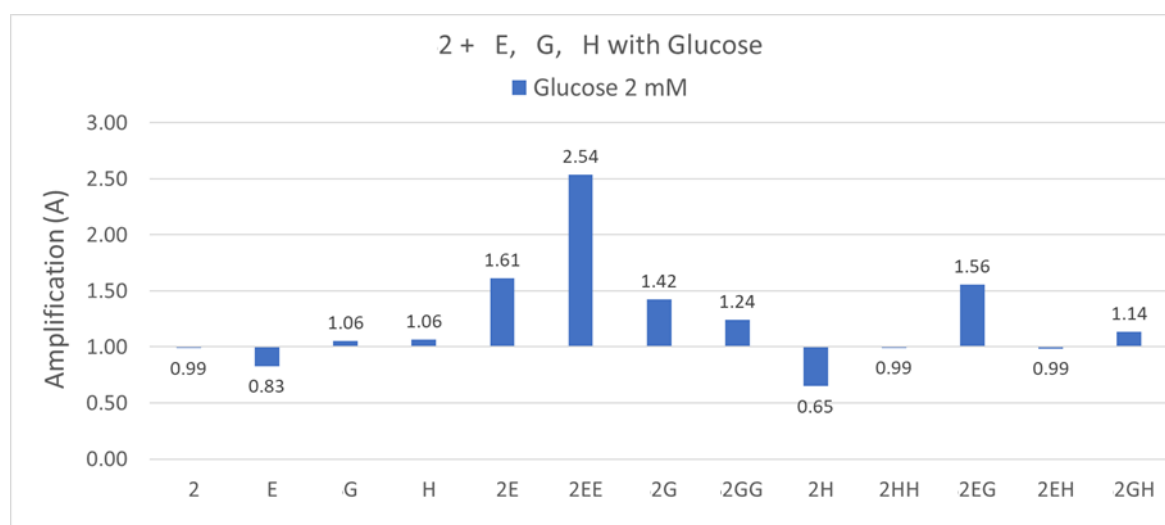


Figure 44: Amplification values for the DCL formed with **2**, **E**, **G**, **H**, and glucose 2mM as template.

In this experiment, **2EE** (aldehyde **2** functionalised with two molecules of boronic acid **E** in meta) emerged as a good receptor for glucose with an amplification value of $A = 2.54$, meaning that its concentration more than doubled due to the presence of glucose. This can be explained by the well-known ability for boronic acids (BA) to covalently bind cis diols (Figure 13) and the presence of these patterns in glucose. However, further experiments will have to be done to confirm this hypothesis.

Since the receptor–target complexation takes place in the ‘core’ of a thermodynamic equilibrium, the amplification of some species must take place to the detriment of others. In this experiment, molecules of starting **E** were needed to amplify **2E** and especially **2EE**, which explains the negative amplification value of **E** ($A = 0.83$). Moreover, since there is a limited number of molecules of **2**, in order to amplify any library member containing this structure, some other must get negatively affected. In this experiment, the concentration of **2H** got reduced in presence of the template so other molecules can get amplified (mainly **2E**, **2EE**, **2G**, and **2EG**).

4.1.4 Conclusions and remarks on DCC 1.0 experiment

With this experiment we learnt that comparative amounts of template work the best (in our experimental conditions) to prepare DCC experiments. Substoichiometric concentrations of template may not cause a template effect, and excessive amounts of it may lead to unpredictable and apparently meaningless scenarios.

We also successfully carried out and analysed a small-scale DCC 1.0 experiment employing all the information gathered previously regarding the optimal conditions for IFR and the analytical procedures needed to screen DCLs.

Finally, we also found a potentially good receptor for glucose. **2EE** poses two BA units that can explain its affinity for saccharides. The validity of these results, the affinity of **2EE** for glucose and its possible selectivity for some saccharides over others will be assessed in future experiments carried out in this thesis.

4.2 DCC 2.0 - Small molecule approach

With experimental conditions already optimised, we were prepared to carry out a complex DCC experiment or DCC 2.0.

Aldehyde **2** would remain as the 'scaffold' for the system and previously employed amines **E**, **G** and **H** will be complemented by amino acids **P** (L-Phenylalanine) and **D** (L-Aspartic acid). It was decided to include these molecules as they were reported by other authors to have excellent binding properties for glycans.³¹⁵ The incorporation of just these two new BB meant an increment in the library size from 13 to 26 members (Figure 45), increasing the complexity of the system and the difficulty of its screening, but also its capacity to create better receptors.

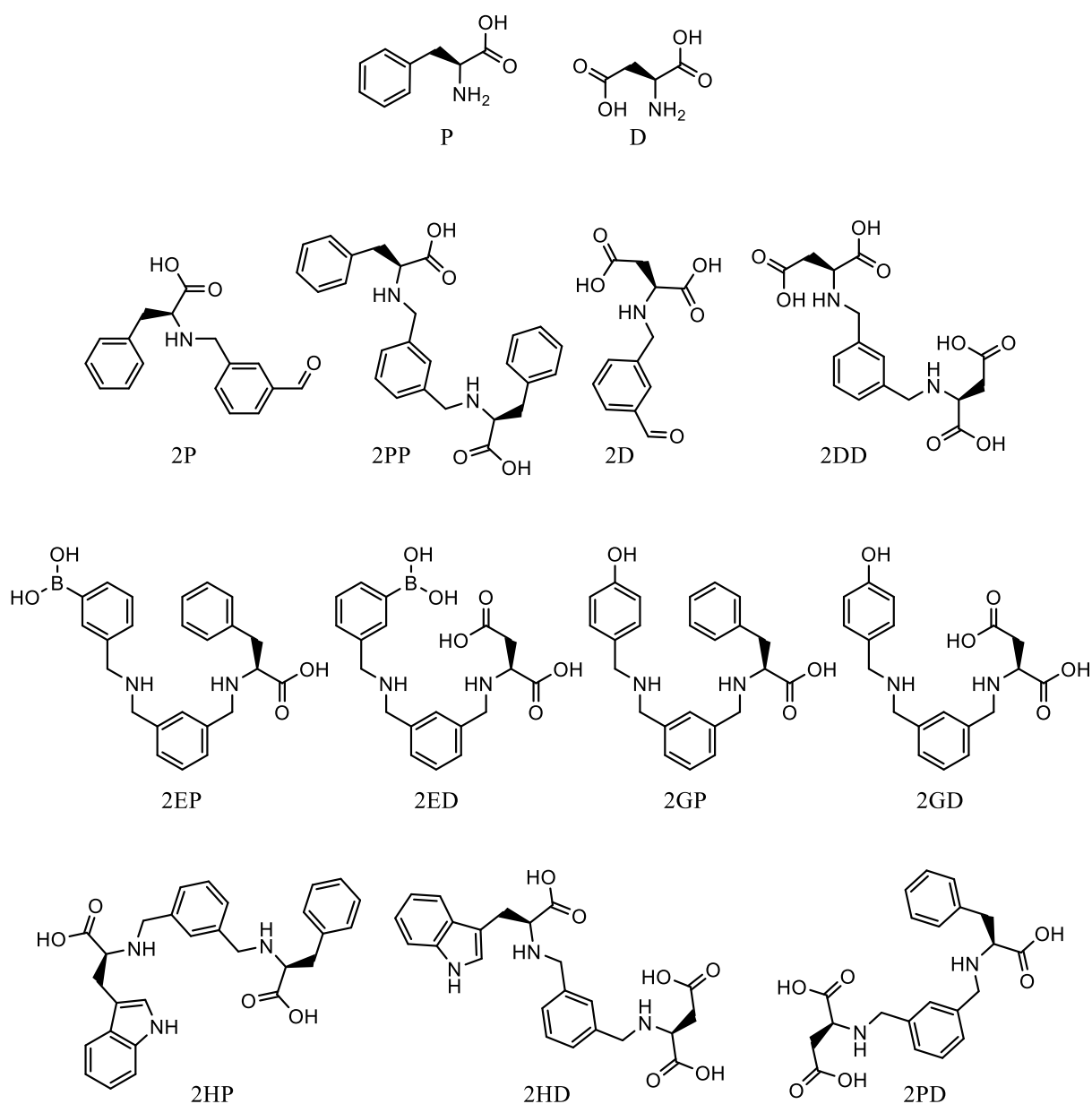


Figure 45: New library members incorporated into DCC 2.0. The rest of the library members, those that were also part of DCC 1.0, are represented in Figure 38.

We wanted to explore the ability for this DCL to self-synthesise and identify receptors that can bind molecules of huge relevance from a biological perspective such as saccharides.

Simply finding a receptor that can bind monosaccharides effectively in aqueous media would be a major accomplishment worth investigating. As discussed in the introduction of this thesis,

saccharides play a key role in the differentiation of glycans, and glycans have major implications in processes such as molecular recognition, early detection of diseases or disease risk stratification.^{4,6,13,30,31}

Unfortunately, saccharides are not an easy target. They are structurally very complex, exhibiting different interconverting tautomers and they are heavily solvated in aqueous media, which means that the potential receptor must compete with the molecules of solvent that are present to a much higher extent. But perhaps the most challenging aspect is the existence of many isomeric structures with different functions and roles in biological processes.

We wanted to take on this tough task and test the ability for DCC to discriminate among a series of common isomeric monosaccharides, while advancing on the limited current knowledge on the binding processes involved in molecular recognition. This could open the doors for new scientific research in the fields of drug design or molecular medicine towards the synthesis of more effective receptors for the diagnosis or treatment of diseases.

With this aim, we tested 4 sugars as templates in our DCC experiment: D-(+)-Glucose, D-(+)-Mannose, D-(+)-Galactose and D-(-)-Fructose (glucose, mannose, galactose, and fructose from now on). They all have a molecular weight of 180.16 g/mol and differ only in the orientation of their OH groups (Figure 46).

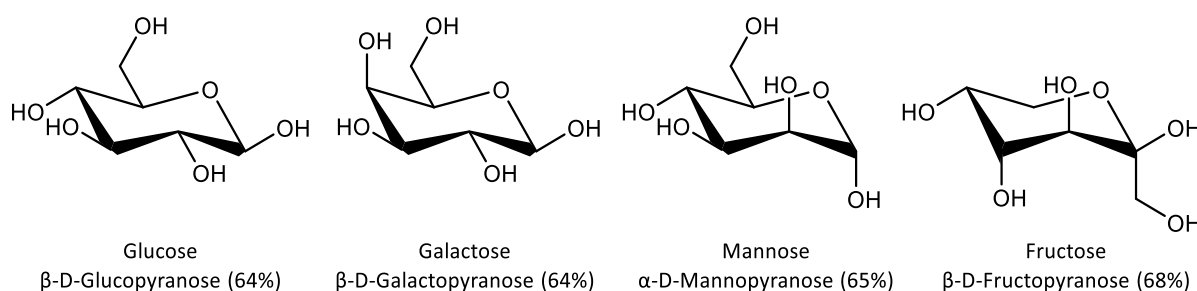


Figure 46: Chemical structure of the predominant tautomeric forms in D_2O for the 4 saccharides employed in the DCC 2.0. The percentage of each tautomer is indicated in brackets.

If DCL members behaved differently towards the different saccharides, that would imply some sort of preference or selectivity. Then, we could hypothesise about the reasons behind the disparity.

Experimentally, DCC 2.0 was carried out as previously in DCC 1.0 but conducting a total of three replicates for each sugar. Moderate errors are a common feature in DCC methodologies, due to the low quantities of analytes to quantify and the consequently large signal-to-noise ratios to deal with. Taking the average amplification value (A) of the 3 measurements, with their corresponding standard deviations would cancel out these errors.

Each combination BB + sugar needs to be studied individually and A values from different combinations cannot be compared. Amplification values are a direct function of selectivity and not necessarily of affinity. In other words, a large A value in a DCC experiment does not necessarily correspond with a great affinity library member – receptor. It only means that the library member is a better receptor than those with lower A values in that particular DCL. In

the same way, a great receptor may be assigned a small A value just because there are other good receptors in the DCL competing for the template.

In DCC 2.0 experiment with glucose as template (DCC 2.0 – G) a DCL member, **2DD**, emerged as a good receptor with A=3.43 (Figure 47). Or to be more precise, it emerged as a *better* receptor for glucose than its competitors. Further studies need to be done to confirm its binding properties.

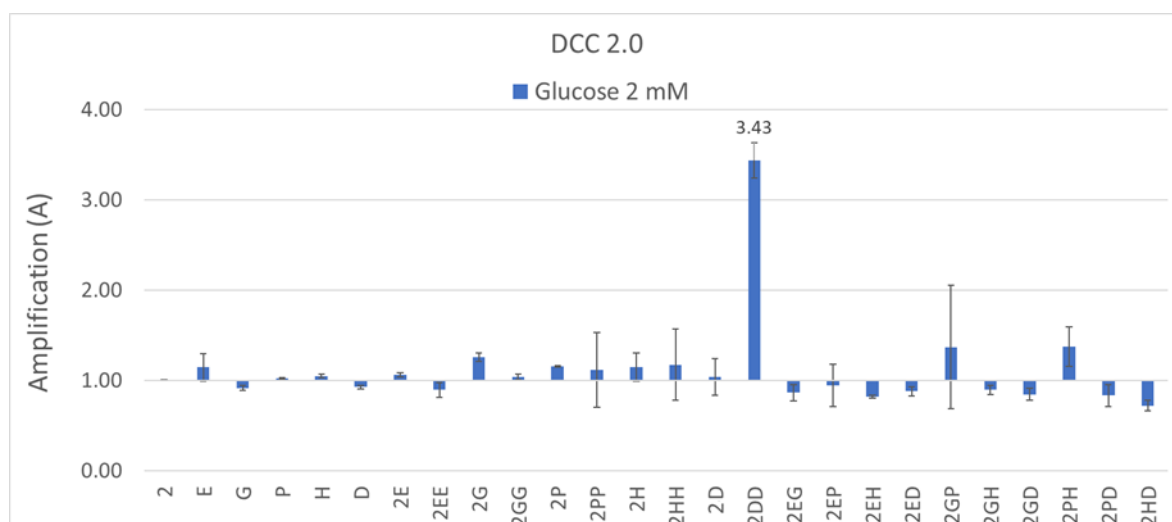


Figure 47: Amplification (A) values obtained for DCC 2.0 experiment employing glucose as template.

As discussed in the chapter dedicated to the LC-MS method development, library members are unequally distributed in the DCL even in the absence of a template. This is due to their differences in relative stabilities. The least stable members are present to a lower extent, and therefore when measuring their peak area in the MS chromatogram, there is a higher signal-to-noise ratio. As a result, the error, or the difference in peak area values across the 3 replicate experiments, is higher in these cases. Nevertheless, all the library members were adequately

identified, and error bars are not an obstacle to appreciate the general trend that the experiment is clearly indicating.

It is worth highlighting that **2EE**, that appeared to be a good receptor according to DCC 1.0 experiment, is now completely outplayed by **2DD** that seems to be, comparatively, a much better receptor for glucose. It could be hypothesised that the ionic interactions and H bonding that can offer **2DD** play a more important role in binding glucose than the BA-diol complexation offered by **2EE**. As explained before for DCC 1.0 experiment, due to the competitive nature of DCC experiments and the limited aldehyde molecules available, for one molecule to get amplified, some others must shrink. **D**, **2EH**, or **2GD** seem to be these ones but with all the library members (other than **2DD**) showing A values very close to 1, one would not draw conclusions about their affinity for glucose.

When adding mannose as template (DCC 2.0 – M) a similar behaviour is observed. **2DD** emerged as the best receptor with A = 5.97 (Figure 48). Again, it cannot be concluded yet that its binding affinity for mannose is remarkably good, or even better than for glucose.

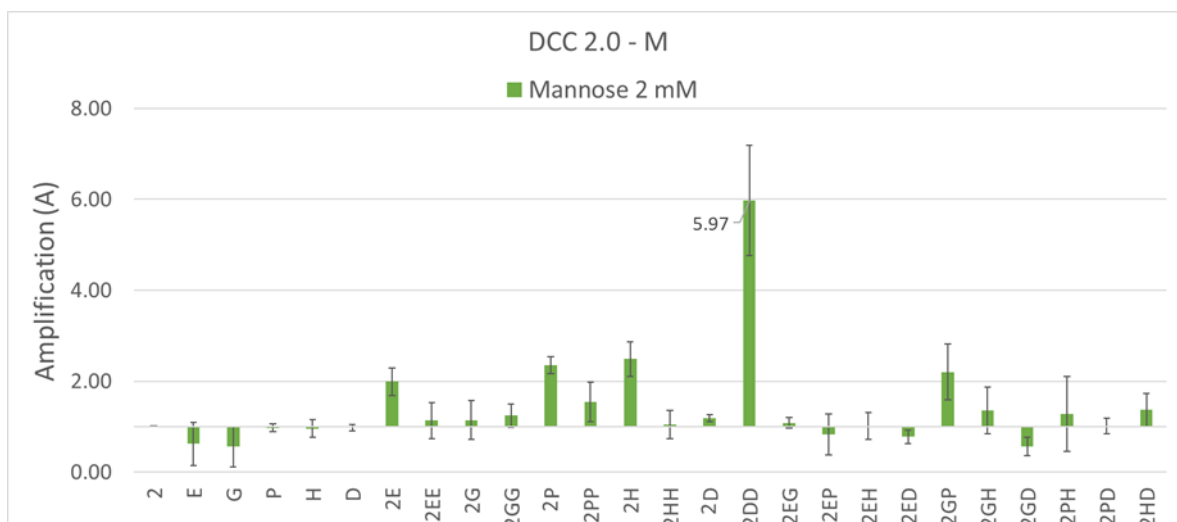


Figure 48: Amplification (A) values obtained for DCC 2.0 experiment employing mannose as template.

Despite the large A value exhibited by **2DD**, three more molecules are above A = 2, which must mean that they also have affinity for Mannose. These are **2P**, **2H**, and **2GP**. Same as most library members, these 3 molecules can offer both H bonding and π -H interactions, but again, it seems like the H binding and ionic interactions offered by **2DD** played a more important role. The most disfavoured ones in this experiment were **E**, **G**, and **2GD**.

In DCC 2.0 experiment with galactose as template (DCC 2.0 – Ga) even though the standard deviation across the different replicates ended up being larger than in previous experiments, the trend indicated the same pattern and **2DD** was again the most amplified molecule with A = 2.96 (Figure 49).

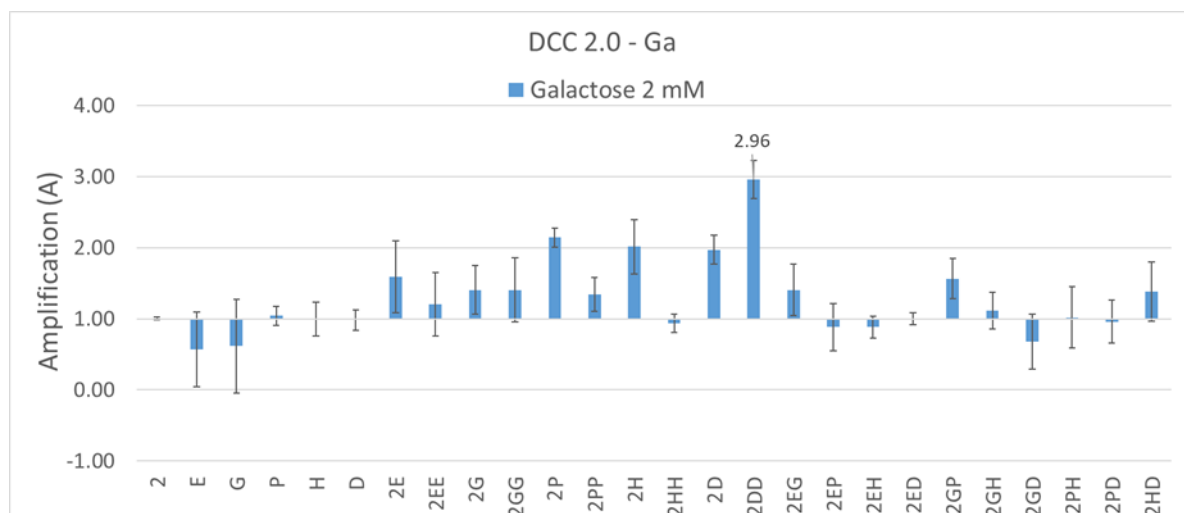


Figure 49: Amplification (A) values obtained for DCC 2.0 experiment employing galactose as template.

The considerably large error bars observed in this experiment limit the possibility of drawing any more conclusions besides the fact that **2DD**, due to the same reasons already explained for the previous experiments, seems to be the best binder for galactose.

The last sugar tested was fructose. Fructose, even though it is as well an isomer to the previously tested saccharides, it is comparatively different to them. Its predominant tautomeric form in solution is, as well as for the other three monosaccharides, the 6-membered ring pyranose form (Figure 46). However, fructose shows an abnormally high 28.59% of its 5-membered ring furanose forms (Figure 50).³¹⁶ The sum of both α and β furanose tautomers only represents a 0.14% for glucose, 0.9% for mannose, and 6% for galactose.³¹⁷ Therefore, we anticipated a different behaviour for fructose in the DCC experiments.

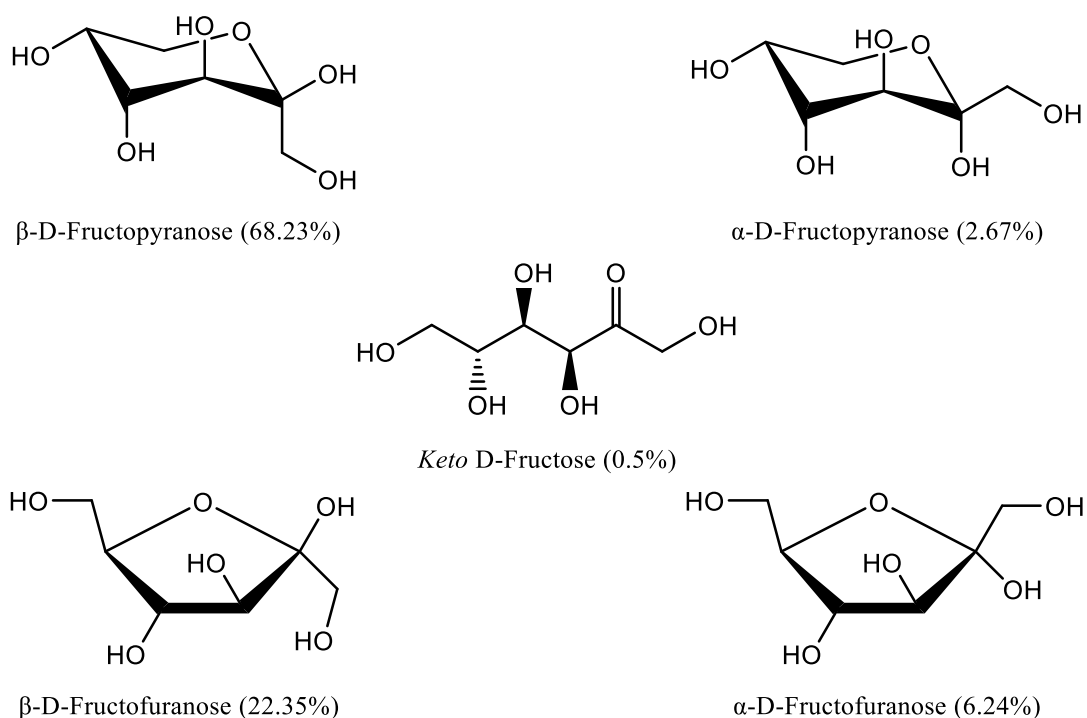


Figure 50: Different tautomeric forms of Fructose in wate (D_2O , 20 °C).³¹⁶

As anticipated, DCC 2.0 experiment with fructose as template (DCC 2.0 – F) showed a different trend (Figure 51). In this case, **2DD** did not experience enhancement and remained with an A value close to 1. Instead, **2P** (A = 2.11) and **2D** (A = 1.91) were the most amplified molecules. These more discrete values could be a signal of higher competitiveness within the library members for the template.

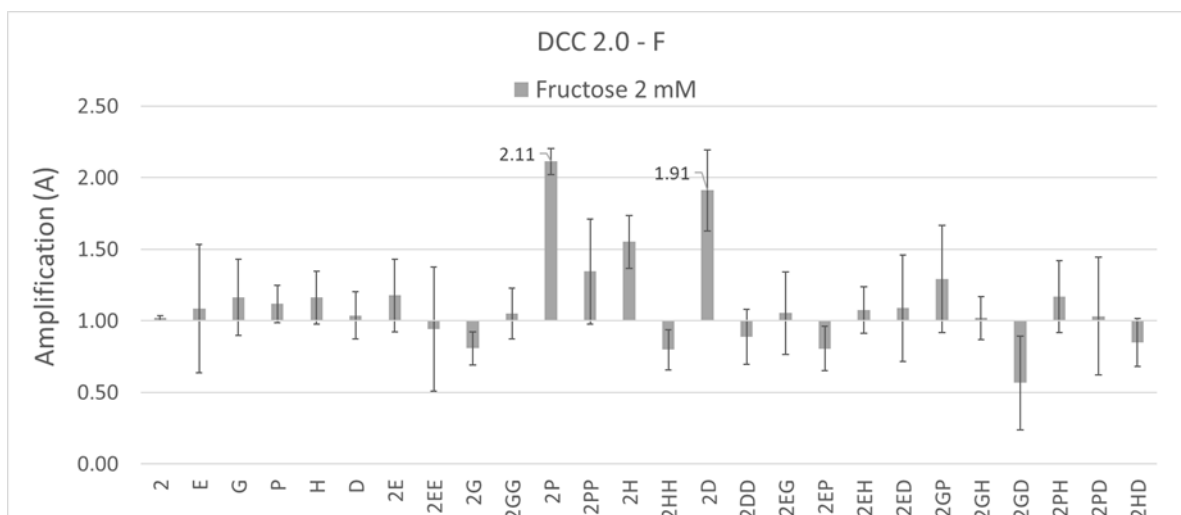


Figure 51: Amplification (A) values obtained for DCC 2.0 experiment employing fructose as template.

With these results, it was decided to individually synthesise and purify some of the library members for further testing and characterisation. Isothermal Titration Calorimetry (ITC) will be utilised to measure the heat liberated when these receptors bind the templates, as a form to obtain their binding constant K_d or K_a . The molecules to be tested were **P**, **2P**, **2PP**, **D**, **2D**, and **2DD**.

4.3 Conclusions of Chapter 4

A DCC methodology was developed employing aldehyde **2** as scaffold and amines **E** (3-Aminomethylphenylboronic acid), **G** (4-Hydroxybenzylamine), and **H** (L-Tryptophan) as starting BBs. This afforded a DCL of 13 members, with a rich variety of functionalities. The analytical procedures to screen the DCLs by LC-MS were optimised. Eight different chromatography columns and conditions were tested and columns containing aromatic

packaging, in particular the Kinetex Phenyl-Hexyl column, gave the best results as it was able to isolate the majority of the library members.

DCC 1.0 experiments explored the use of different concentrations of glucose as template to find the optimal one. Substoichiometric (0.1 mM) amounts did not seem to produce a template effect and the amplification factor (A) for most of library members was very close to 1. When employing excess of template (20 mM), the results seemed to be erratic, mostly negative, and did not seem to follow any trend. The most reliable results were obtained when using comparative concentrations of template (2 mM). This time, library member **2EE** (posing two BA groups) was amplified the most with $A = 2.54$. This could be explained by the well-known ability for the BA groups to bind cis diols, and the presence of these in glucose.

Then, amino acids **D** and **P** (L-Aspartic acid and L-Phenylalanine, respectively, were added to the DCL so the library size and complexity was increased from 13 members in DCC 1.0 experiment to 26 members in DCC 2.0. Their affinity toward the common saccharides: glucose (G), mannose (M), galactose (Ga), and fructose (F) was tested. The amplification profile of G, M, and Ga was very similar, with the library member **2DD** (starting scaffold 2 posing 2 aspartic acid residues in meta positions) standing out as the best receptor for such sugars. This comparable behaviour did not come as a surprise since those sugars are isomers with very similar distribution of their different tautomers in solution.

In the experiment with fructose, however, **2DD** did not get amplified and was outperformed by molecules **2D** and **2P** (aldehydic scaffold **2** functionalised with only one molecule of aspartic acid and phenylalanine, respectively).

However, it must be kept in mind that amplification values on their own cannot confirm good binding, they rather suggest that the amplified molecules have more affinity for the template than the ones that did not get amplified, with no idea of the scale of binding affinity they may show. While these are indeed promising results, the relevant library members need to be synthesised and further testing must be carried out.

The differences in behaviour in the experiment with fructose could be explained by the unequal distribution of its tautomers in solution, with a higher presence of the 5-membered furanose form in comparison with the other three monosaccharides tested, although further testing should be carried out to confirm this hypothesis.

Chapter 5 -DCC Validation

As stated before, DCC results can be indicative or valid just for comparative purposes. They can only suggest which ones among the library members tested in an experiment have the greatest affinity for the target employed, but they give no information regarding the scale of binding affinity that they may have for such target. Further analytical techniques need to be carried out to assess this. Moreover, DCC experiments often give misleading information as amplification values can be influenced not only by binding affinity as we expect, but also by other factors such as the formation of precipitates or complexes. Therefore, the binding properties of the amplified library members, and consequently the validity of DCC as a tool to predict potential receptors for biomarkers needs to be further confirmed by other techniques. To this aim, ITC and NMR titrations have been carried out.

5.1 Characterisation of the receptors. Binding studies

In order to perform further testing molecules need to be synthesized and purified, because the DCL complex mixture is not suitable to obtain quantitative data. The molecules that performed the best in DCC 2.0 experiment and therefore had to be tested were the disubstituted receptor **2DD** and monosubstituted **2D** and **2P**. Disubstituted molecule **2PP** was synthesised as well, and tested as a negative control together with the commercially available amino acids **D** and **P**.

Mono-substituted molecules **2D** and **2P** were not synthesised exactly as they are present in the DCL. Functionalising only one of the two aldehydes of starting isophthalaldehyde **2** would have been a rather complex procedure that would have involved several reaction steps. Instead, we pursued a much simpler approach using benzaldehyde (**1**), the benzene ring

posing only one aldehyde group, as starting material. This would afford the mono substituted products similar to **2D** and **2P**, without the extra CHO group in position 3 of the benzene ring. These molecules will be called **1D** and **1P**, respectively. The missing aldehyde group is far away from the 'recognition positions' of the receptor and therefore their absence are not expected to affect the overall binding properties of the molecule. In other words, the affinity of **1D** and **1P** for the target saccharides are expected to be the same as that of their analogues in the DCC experiment, **2D** and **2P**. The set of molecules to be tested are shown in Figure 52a.

All receptors will be synthesised via reductive amination reaction between isophthalaldehyde or benzaldehyde and the corresponding amine **D** or **P** (Figure 52b). In the same way as in the chemistry used for the DCC, the aldehyde reacted spontaneously with amine to obtain an imine that is readily reduced to secondary amine. After simple workup and purification by HPLC, pure compounds were obtained. Synthetic protocols are explained in detail in the experimental chapter of this thesis (Chapter 7 -).

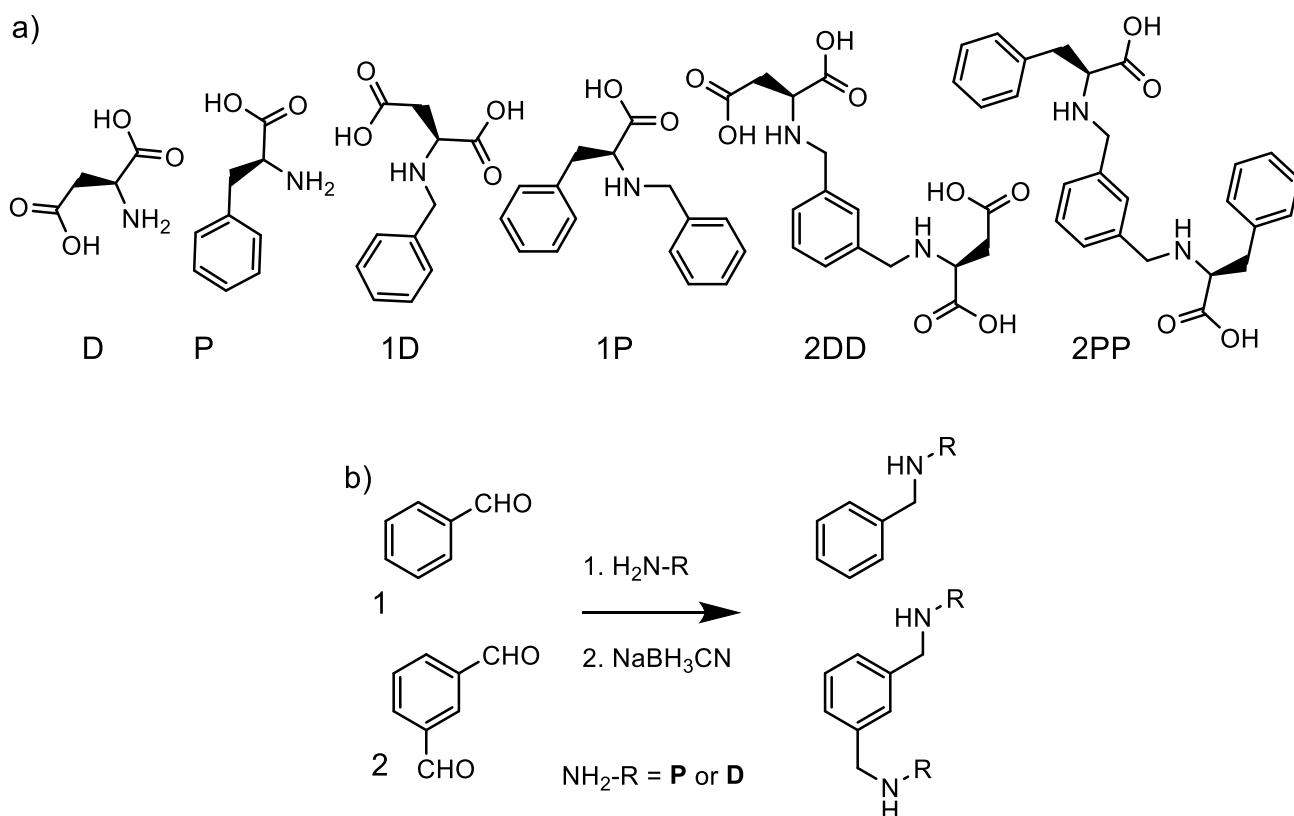


Figure 52: a) Set of molecules to be tested in order to obtain binding data with the saccharides of interest. b) Reductive amination reaction to afford molecules to be tested.

It was interesting to us not only to prove that these molecules are indeed good receptors for the sugars tested, but even more importantly, to demonstrate that the trend of amplification factors obtained during the DCC experiment match real binding data, hence validating the method, and proving DCC to be an efficient and powerful technique for the prediction of receptors for biomarkers. To this aim, the binding data of starting BBs **D** and **P**, and synthesised molecules **2PP** -that did not get especially amplified during DCC 2.0 experiment- was recorded as well as negative controls.

Proving a successful way to test large libraries of potential receptors to find the best one among them in a single experiment could save time and resources in processes such as the development of drug candidates.

The information learnt during a DCC experiment regarding the necessary features present in a molecule to accomplish molecular recognition could be employed to re-design future optimised DCLs (Figure 53).

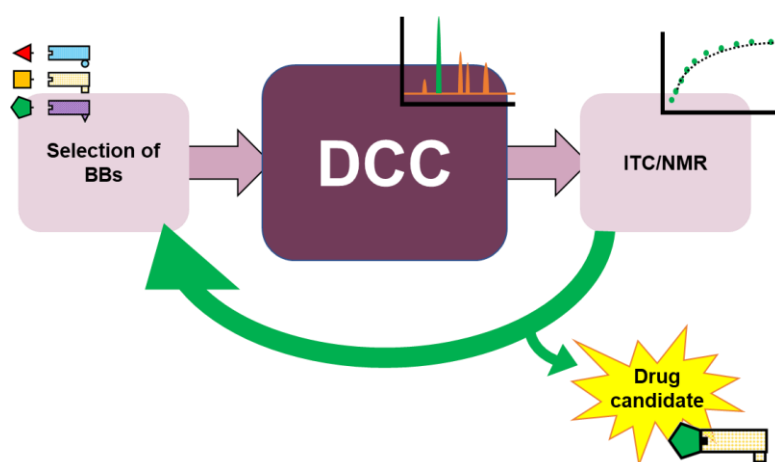


Figure 53: Schematic representation of the circular process attempted in this thesis with the aim of creating receptors for biomarkers with potential applications in medicinal chemistry, for instance, as drugs.

The key analytical experiment carried out to calculate the binding properties of our candidates was Isothermal Titration Calorimetry (ITC). ^1H NMR titrations experiments will be also carried out to obtain more information regarding the binding process itself.

5.2 ITC

isothermal titration calorimetry (ITC) allows the determination of the heat involved (generated or consumed) in a binding event. With this it can be calculated the following thermodynamic parameters: n (number of binding sites), K_a (binding constant, M^{-1}), ΔH (change in enthalpy, $Kcal\ mol^{-1}$), and S (change in entropy, $Kcal\ mol^{-1}\ K^{-1}$).²⁹² ITC does not require labelling or immobilization of the receptor, which is a benefit in comparison with similar techniques as it avoid the possibility of artifacts generated due to the derivatisation of the receptor.³¹⁸ ITC was used to evaluate the interaction between the molecules **D**, **P**, **1D**, **1P**, **2DD**, and **2PP** with the saccharides glucose (G), galactose (Ga), mannose (M), and fructose (F).

5.2.1 Optimisation of ITC experimental conditions

The ITC experiment entails titrating the monosaccharide ligand solution into a receptor solution. The receptor-containing cell receives the ligand in tiny aliquots (μl). When the ligand binds to the receptor to form the complex receptor-sugar, heat is released from the system and a negative heat change is registered. A significantly larger quantity of heat is generated during the first few injections since there is more accessible number of receptors. In the last injections, when a lesser fraction of the receptor is free, the amount of heat emitted gradually decreases until saturation is attained.

The experimental setup had to be optimized before conducting the binding studies.

First, the parameter c-value had to be found. This is connected to the experimental conditions (concentrations, heat change, and shape of the isotherm).³⁰⁰ The c-value should ideally

be ≥ 10 so the titration data can be fitted with a sigmoidal shape which allow all the thermodynamic parameters (n , K_a , ΔH and ΔS) to be calculated with accuracy. c -value can be calculated from Equation 5.1.

$$c = n[R]_{cell}K_a \quad \text{Equation 5.1}$$

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 5.2, provided K_a and n are known values.

$$[R]_{cell} = \frac{c}{nK_a} \quad \text{Equation 5.2}$$

The binding constant (K_a) of our receptors (**D**, **P**, **1D**, **1P**, **2DD**, **2PP**) to the monosaccharides (G, Ga, M, F) was unknown. Thus, for a preliminary estimation of the $[R]_{cell}$, since we could not find in the library reported values of binding constants for similar molecules, the binding constant of phenyl boronic acid (PBA) to glucose at pH 7.4 was considered ($K_a = 10 \text{ M}^{-1}$).¹⁵⁵ The number of binding sites, n , was set to 1 as the small size of the receptors could not bind any more than one monosaccharide.

Due to the small size and simplicity of our receptors, and more importantly because of the high solvation of the target saccharides in water environments, we were not expecting strong interactions. Weak affinity systems ($K_a < 100$) such as ours, tend to present low c -values.²⁹⁷

For $c = 10$, the molar ratio receptor-ligand as per Equation 5.3 at the end of the titration should at least be 1:2 (Figure 28).³⁰⁶

$$\text{Molar ratio} = \frac{[R]}{[L]} \quad \text{Equation 5.3}$$

If $c = 10$, then $[R]_{\text{cell}}$ equals to 1 M. Unfortunately, these conditions could not be met due to the poor solubility of the receptors and a lower c value had to be chosen. Even with $c = 1$, $[R]_{\text{cell}}$ would be 100 mM which is still far above the solubility of the receptor. Hence, we arbitrarily select the value of 2 mM as it was employed by others in ITC binding studies of Fructose.³¹⁹

In this circumstances, c value becomes smaller than 0.1 and therefore some considerations had to be taken to obtain reliable binding data: The number of binding sites, n , must be fixed during the fitting.³⁰⁶ We will therefore be fixing it to $n=1$ for all the ITC experiments shown in this thesis. Moreover, since saturation of the receptor is difficult to achieve in systems with such as extremely low c value, a high molar ratio of ligand is required.²⁹⁷

Even with all these precautions, the success of the experiment will be determined by the range of operation of the ITC instrument and the minimum amount of heat that can be measured to acquire heat peaks that are considerably higher than the background noise. With the instrument employed this value rests around $3 \mu\text{Js}^{-1}$.²⁹⁸

With this in mind, some preliminary experiments were carried out employing glucose as titrant. The library member **2DD** that exhibited the best amplification value in the majority of the DCC experiments and was therefore expected to be a good receptor for glucose, was employed as ligand. The concentration of the ligand was previously decided to be 2 mM for

solubility reasons, but we had to optimise parameters such as concentration of titrant, volume of injections and number of injections to ensure a sufficient amount of heat generated that can be measured by the instrument but also a final ratio receptor-ligand enough to saturate the receptor.

Initially, a molar ratio of 1 : 7 was selected resulting in a 2 mM solution of 2DD and an 40 mM solution of glucose for a total of 40 x 1.5 μL injections. These settings generated too small peaks, with heat released $>1.5 \mu\text{J s}^{-1}$. We could have increased the injection volume, but this could induce higher injection noise and spacing between injections. Instead, we decided to increase the concentration of the sugar, which is unexpensive and perfectly soluble, to 80 mM. This time the heat liberated was $>2.5 \mu\text{J s}^{-1}$ (Figure 54), that is measurable by the instrument, and resulted in a final molar ratio receptor : sugar of 1 : 15.4 (Table 5).

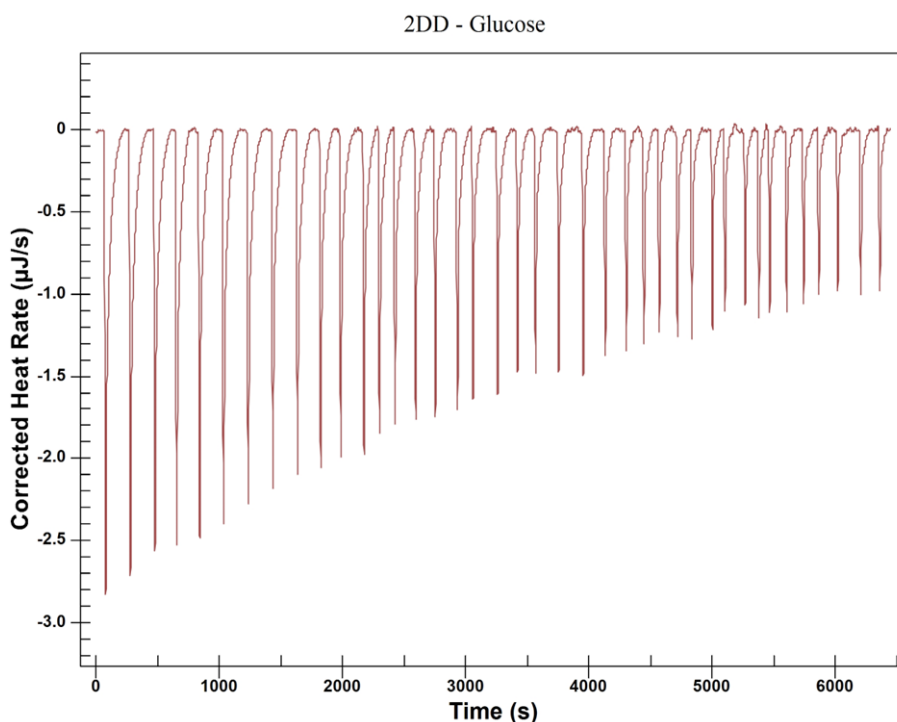


Figure 54: Titration graph corresponding to the experiment with glucose (80 mM) and 2DD (2 mM).

Table 5: ITC results for the titration experiment with glucose (80 mM) as titrant and 2DD (2 mM) as titrate.

injection	Q (μ J)	Corrected Q (μ J)	inj volume (μ L)	moles titrant	moles titrate	moles titrant / moles titrate	inj interval (s)
1	-94.04	-89.70	1.5	1.20E-07	3.67E-07	0.33	160
2	-86.02	-81.45	1.5	2.39E-07	3.64E-07	0.66	159
3	-78.45	-74.01	1.5	3.57E-07	3.61E-07	0.99	147
4	-72.13	-68.82	1.5	4.74E-07	3.58E-07	1.32	119
5	-72.09	-68.39	1.5	5.90E-07	3.55E-07	1.66	135
6	-69.03	-65.50	1.5	7.06E-07	3.52E-07	2.00	140
7	-63.89	-60.20	1.5	8.20E-07	3.50E-07	2.35	134
8	-60.45	-56.86	1.5	9.33E-07	3.47E-07	2.69	139
9	-56.76	-53.64	1.5	1.05E-06	3.44E-07	3.04	134
10	-52.63	-51.04	1.5	1.16E-06	3.41E-07	3.39	114
11	-51.79	-48.17	1.5	1.27E-06	3.38E-07	3.75	136
12	-49.64	-46.94	1.5	1.38E-06	3.36E-07	4.11	115
13	-43.56	-41.10	1.5	1.49E-06	3.33E-07	4.47	115
14	-42.68	-40.54	1.5	1.59E-06	3.30E-07	4.83	122
15	-41.55	-39.45	1.5	1.70E-06	3.27E-07	5.20	112
16	-39.23	-36.13	1.5	1.81E-06	3.25E-07	5.56	104
17	-40.05	-40.11	1.5	1.91E-06	3.22E-07	5.94	111
18	-36.52	-34.73	1.5	2.02E-06	3.20E-07	6.31	107
19	-37.58	-33.39	1.5	2.12E-06	3.17E-07	6.69	129
20	-32.52	-31.00	1.5	2.22E-06	3.14E-07	7.07	106
21	-30.99	-28.60	1.5	2.33E-06	3.12E-07	7.46	88
22	-30.80	-27.62	1.5	2.43E-06	3.09E-07	7.85	95
23	-32.26	-29.40	1.5	2.53E-06	3.07E-07	8.24	103
24	-26.84	-24.18	1.5	2.63E-06	3.04E-07	8.63	82
25	-25.35	-22.92	1.5	2.73E-06	3.02E-07	9.03	83
26	-27.95	-27.82	1.5	2.82E-06	2.99E-07	9.43	94
27	-22.61	-20.93	1.5	2.92E-06	2.97E-07	9.83	74
28	-25.29	-22.65	1.5	3.02E-06	2.95E-07	10.24	97
29	-24.41	-20.30	1.5	3.11E-06	2.92E-07	10.65	102
30	-23.35	-19.38	1.5	3.21E-06	2.90E-07	11.07	91
31	-19.19	-15.74	1.5	3.30E-06	2.87E-07	11.48	67
32	-19.20	-16.39	1.5	3.39E-06	2.85E-07	11.90	78
33	-20.79	-16.93	1.5	3.49E-06	2.83E-07	12.33	88
34	-19.44	-15.61	1.5	3.58E-06	2.81E-07	12.76	68
35	-20.45	-18.58	1.5	3.67E-06	2.78E-07	13.19	84
36	-21.00	-19.36	1.5	3.76E-06	2.76E-07	13.62	101
37	-18.16	-16.12	1.5	3.85E-06	2.74E-07	14.06	99
38	-17.89	-15.84	1.5	3.94E-06	2.72E-07	14.50	97
39	-18.99	-15.36	1.5	4.03E-06	2.69E-07	14.95	90
40	-16.89	-13.77	1.5	4.11E-06	2.67E-07	15.40	84

Other experimental parameters such as the duration of each injection, stirring speed, temperature, or the interval between injections were set as advised by the manufacturer (Table 6).²⁹⁸

Table 6: Isothermal titration calorimetry experimental parameters.

Ligand concentration	80 mM
Receptor concentration	2 mM
Number of injections	40
Injection volume	1.5 μ L
Duration of each injection	4.0 s
Initial cell volume	185 μ L
Cell temperature	20 $^{\circ}$ C
Initial delay	60 s
Stirring speed	150 RPM
Spacing	280 - 480 s

As recommended by the instrument manual, the stirring speed was selected to be 150 RPM and the filter period to 2 s. The differential power (DP) signal is the experimental data captured during a titration.²⁹² When heat is generated during an exothermic event, DP decreases. On the titration graph, the drop in DP is seen as a negative peak (Figure 55), and when there is no heat change, the DP value is at zero.

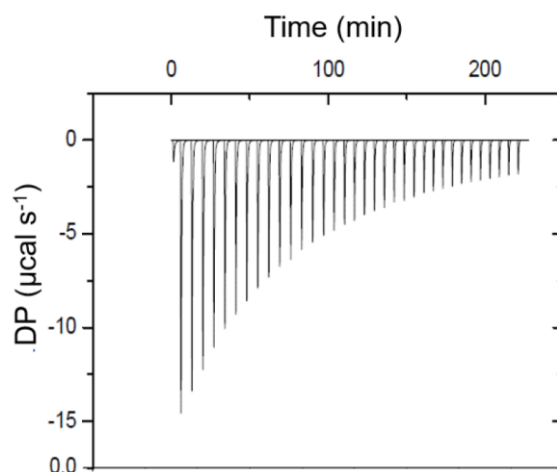


Figure 55: Titration graph with negative peaks corresponding to a decrease in the DP value determined by an exothermic binding event.

The spacing, or time gap, needed between injections depends on the magnitude of heat change. The interval must enable the DP to revert to the pre-injection baseline value. The spacing was chosen at 280 s to 480 s. The temperature during the analysis was set at 20°C.

In addition, the baseline must be flat, and the DP value must be steady for the experiment to begin. Only when the temperature difference between the cells and the adiabatic jacket temperature is 0 ± 0.05 °C and the DP is stable at a value equal to the chosen reference power (1 cal s^{-1}) the experiment begins. There is also a 60 second initial delay when the experiment begins during which the device gathers enough data points to create a baseline.

In addition to collecting receptor-template binding information, the heat of dilution was also measured. The dilution of the ligand into a larger volume in the sample cell typically results in a heat change when titrating highly concentrated solutions. This heat, known as heat of dilution, must be subtracted from the titration data of the binding experiment.²⁹⁸ Using the

same conditions as the binding experiment, the heat of dilution was calculated by titrating an 80 mM ligand solution into the buffer. The thermodynamic parameters were then derived from the fitting after the heat of dilution was removed from each titration dataset.

In conclusion, for all ITC studies described in this chapter, 40 1.5 μ l injections of 80 mM monosaccharide solutions are injected to the sample cell at 20 °C to titrate 2 mM solutions of receptor. The heat of dilution is measured and subtracted. The binding stoichiometry n is set to 1 when fitting the data. The average of the three titrations, with an error equal to twice the standard deviation, is employed to calculate the thermodynamic parameters K_a , K_d , ΔH , and ΔS .

5.2.2 ITC to assess the binding properties of the amplified receptors

According to the previous DCC experiments, molecule **2DD** could be a good receptor for glucose (G), mannose (M), and galactose (Ga). In the study with fructose (F), **2DD** did not get amplified. Instead, **2D** and **2P** seemed to be the best receptors although in a more competitive scenario with other molecules exhibiting similar amplification values.

ITC was used to evaluate the interaction between the amino acids **D** and **P**, as well as the synthesised -potential- receptors **1D**, **1P** (analogues of **2D** and **2P**, respectively), **2DD**, and **2PP** with the saccharides G, Ga, M, and F. For comparability of results, the ITC titrations took place in the same solvent used for the DCC experiments: 100 mM carbonate buffer pH 10. ITC experimental parameters employed were the indicated and optimised in the previous section of this thesis, as summarised in Table 6.

All the titrations in this chapter were performed 3 times with the heat of dilution (obtained from ligand into buffer experiments) subtracted. The values displayed are the mean of the 3 replicates, with the error calculated as 2 times the standard deviation of the 3 measurements. Obtained values are shown in Table 7.

*Table 7: Binding constants (K_a , M^{-1}) measured by ITC titrating 80 mM solution of the saccharides D-Glucose, D-Galactose, D-Mannose, and D-Fructose into 2mM solution of receptors **D**, **P**, **1D**, **1P**, **2DD**, and **2PP** at pH 10 100 mM carbonate buffer. The titrations were carried out 3 times and the values displayed are the average value \pm the standard deviation. Heat of dilution was subtracted. Cells displaying a zero '(0)' correspond to titrations that generated a heat variation too small for the instrument to measure it. It can be considered that $K_a = 0 M^{-1}$ for such combinations.*

	D	P	1D	1P	2DD	2PP
D-Glucose	(0)	(0)	(0)	(0)	44.5 \pm 1.8	(0)
D-Galactose	(0)	(0)	(0)	17.9 \pm 2.6	56.7 \pm 0.3	(0)
D-Mannose	10.7 \pm 1.2	(0)	11.7 \pm 0.5	31.9 \pm 1.8	47.7 \pm 3.2	(0)
D-Fructose	170.7 \pm 39.6	1136.0 \pm 148.3	1548.3 \pm 98.1	1762.0 \pm 162.4	840.7 \pm 35.8	289.0 \pm 21.0

Studying the results sugar by sugar it is noticeable that G, Ga, and M show a similar profile. They display a discrete binding constant with **2DD** (K_a around 50 M^{-1} for the three of them), and lower to no interaction with the other sugars tested. As indicated previously, it is reasonable that this set of monosaccharides show similar properties as they are isomers differing only in the spatial orientation of one of their OH groups (Figure 46).

Interestingly enough, all the molecules exhibited a much higher binding for fructose. Being fructose also an isomer to the other tested sugars, this mismatch may arise from the unequal

distribution of its tautomers. Fructose presents in water a considerably higher amount of its furanose forms in comparison with the others (28.6 % vs 0.1-6 %).^{316,317} Even though all the molecules showed considerable binding for fructose, there was a clear best one. The K_a of **1P** for fructose was 1762 M^{-1} ($K_d = 5.7 \times 10^{-4} \text{ M}$). This is remarkable considering the small size and simplicity of the ligand, and the fact that the binding occurs in aqueous environment, when the saccharides are heavily solvated.

While there are reported receptor with higher affinities for fructose, these are more complex molecules and can only work in organic solvents that are unable to create H bonds with the saccharide target (or they do it less effectively than water). An example of this is the helical foldameric oligomer developed by Chandramouli et. al. that binds fructose with $K_a = 165 \times 10^3 \text{ M}^{-1}$ (5:95 DMSO/ CHCl_3).³²⁰ The few examples of receptors for fructose published in the literature that are small and simple as **1P**, and can work in aqueous environments as well, showed mostly poorer affinity for fructose. These are the cases of the phenyl boronic acid (PBA), with a maximum K_a for fructose of 560 M^{-1} ,¹⁵⁵ its m-nitro derivative with $K_a = 1350 \text{ M}^{-1}$, or the ortho-amino PBA with $K_a = 1640 \text{ M}^{-1}$.¹⁵⁷ There is one example however of a receptor of this type that outperformed **1P** and this is the bis-boronic acid reported by Fossey et al. that binds D-Fructose with $K_a = 130 \times 10^3 \text{ M}^{-1}$, with selectivity over D-Glucose ($K_a = 5 \times 10^3 \text{ M}^{-1}$).³²¹

Monosubstituted receptors **1D** and **1P** showed more affinity for fructose than their disubstituted analogues **2DD** and **2PP**. By contrast, bidentate ligand **2DD** is better for glucose than mono-functionalised receptors.

But more interesting conclusions can be drawn when interpreting the results given by all the receptors sugar by sugar, in comparison with previously obtained amplification data from DCC experiments. We represented the binding data obtained in a bar graph and overlapped it with the previously created amplification charts.

In the case of glucose, it can be seen how the charts matched almost perfectly. The only molecule that was notoriously amplified during the DCC experiments, **2DD** ($A = 3.43$), was indeed the only molecule that showed binding for Glucose (Figure 56).

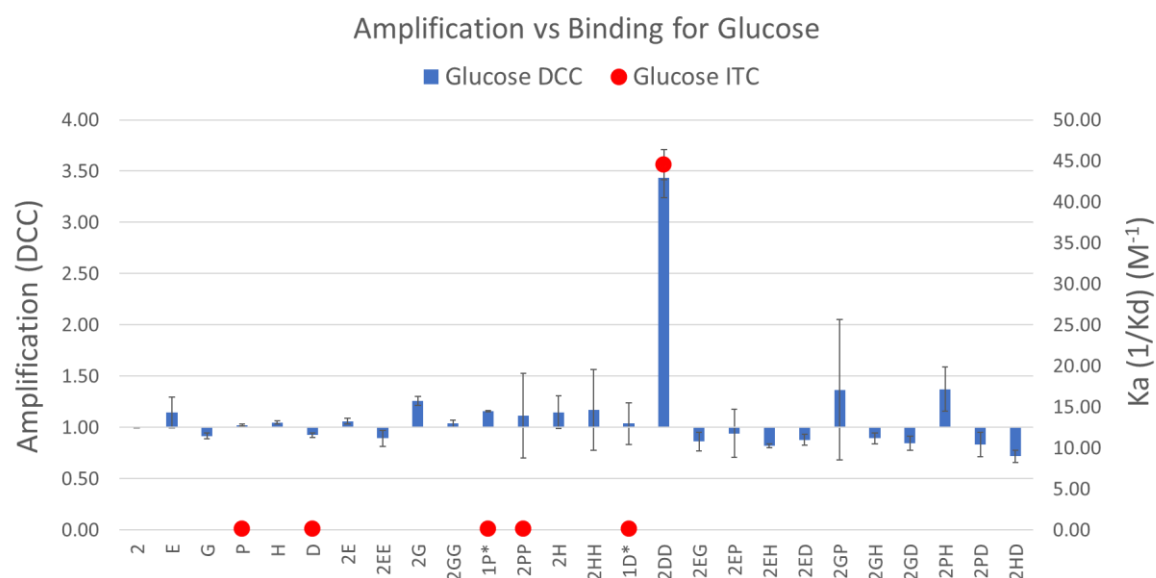


Figure 56: Amplification values obtained in DCC experiment (y axis, left-hand side) and K_a constants obtained in the ITC titrations (y axis, right-hand side) employing D-Glucose as template/titrant. It must be kept in mind that molecules **2D** and **2P** formed in the DCL were not tested by ITC. Instead, their analogues **1D** and **1P** did.

While **P**, **D**, **1P**, and **1D** showed no binding profile in the ITC titrations, **2DD** displayed values of 42.5, 45.0, and 46.1 M^{-1} in the 3 replicates carried out, giving an average binding constant of

$K_a = 44.5 \pm 1.8 \text{ M}^{-1}$ ($K_d = 22.5 \text{ mM}$). Figure 57 and Figure 58 show the ITC binding profile of **2DD** and **D** for glucose, respectively. Figure 59 represents both results overlapped so the difference in the magnitude of their binding profiles can be appreciated.

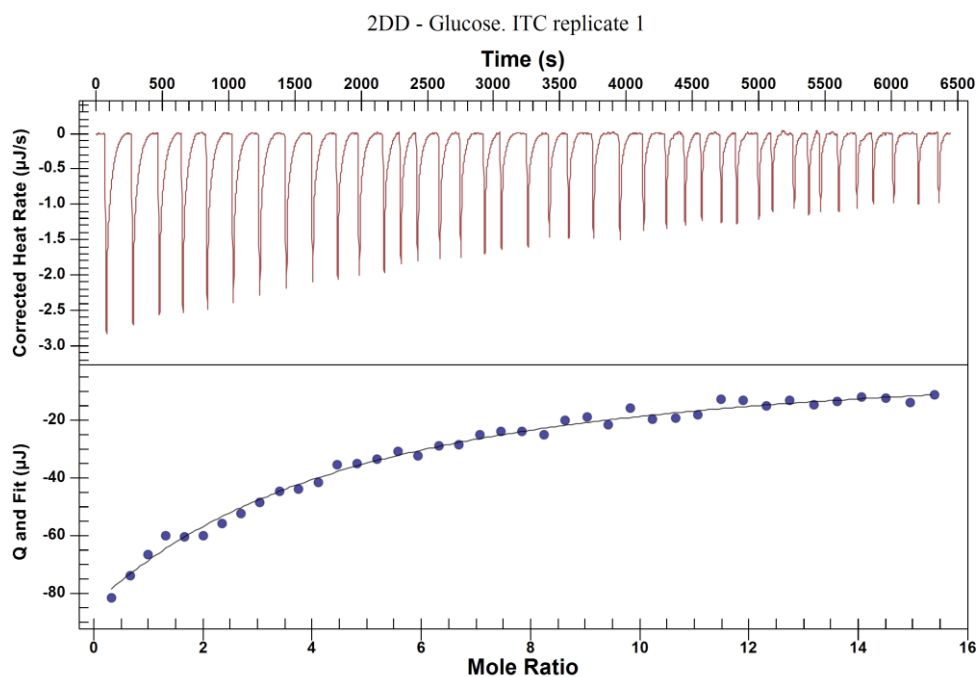


Figure 57: Titration graph corresponding to the binding event between D-Glucose and the molecule **2DD** (top). Negative peaks correspond with an exothermic process. Successful fitting of the binding data with the ITC model: Independent (bottom).

Although a binding constant of 44.5 M^{-1} may not seem impressive It is worth mentioning that for a long time the best water-friendly glucose synthetic receptors reported were not much better, with binding constants in the range of $K_a = 250 \text{ M}^{-1}$ and these are more complex 3D cages that completely encapsulate the sugar.¹⁰⁴

Only after decades of comprehensive structure design, Davis' group was able to develop a substantially better receptor for glucose that can work in water environments, with $K_a = 18000$

M^{-1} .⁵³ Moreover, biological receptors for glucose that have been employed commercially do not exhibit a much higher affinity either. An example is the lectin concanavalin A with a K_a of just $520 M^{-1}$.¹⁴²

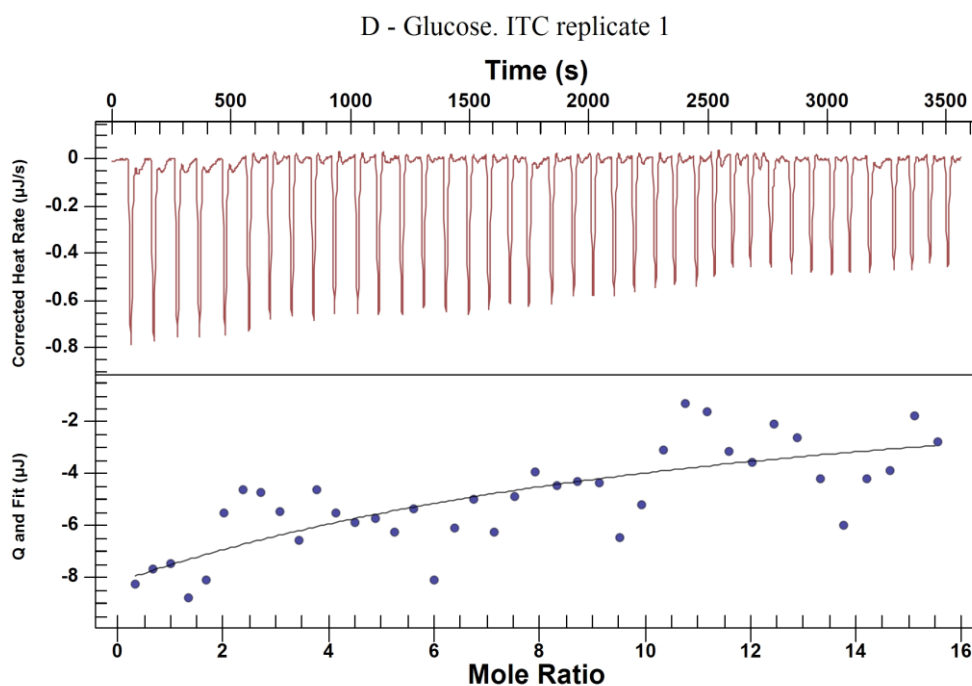


Figure 58: Titration graph corresponding to the binding event between D-Glucose and the molecule **D** (top). Negative peaks correspond with an exothermic process. Unsuccessful fitting of the binding data with the ITC model: Independent (bottom).

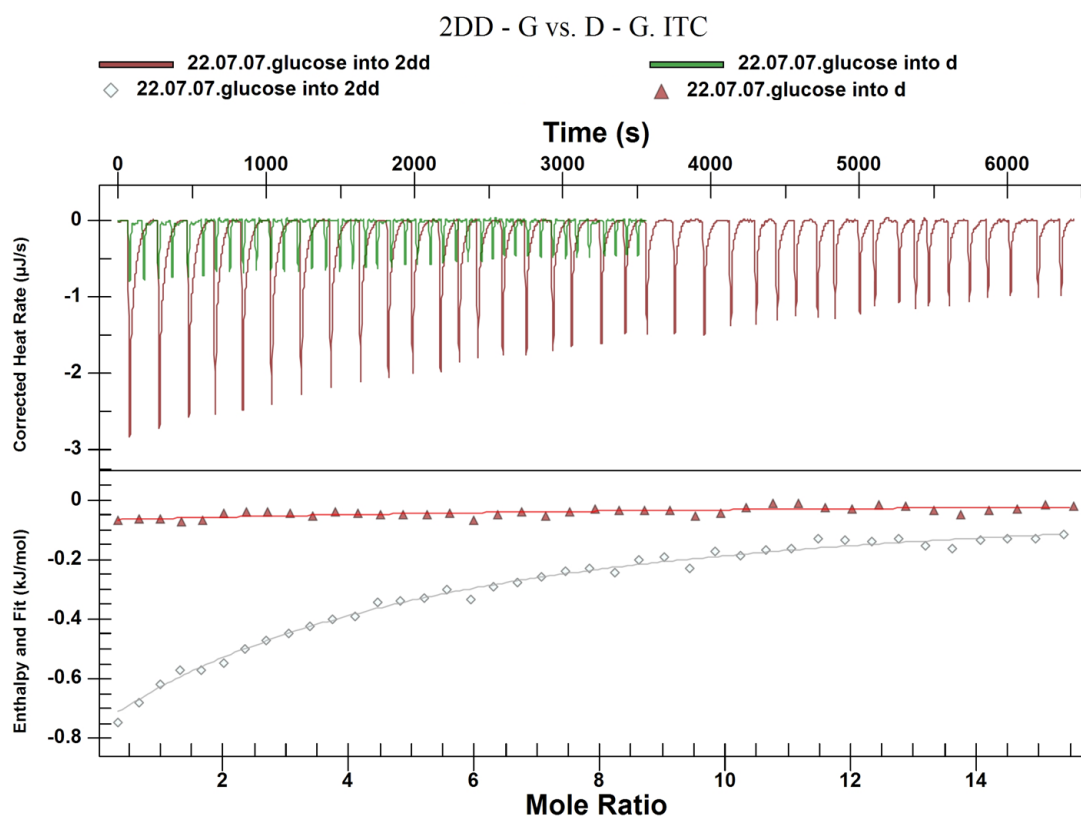


Figure 59: Overlap of the titration graphs and models fitting for both 2DD and D with glucose.

Similar results were obtained for galactose and mannose. When overlapping the amplification values (DCC experiment) and the binding constant (ITC experiment) of the receptors, they match reasonably good. In the case of the mannose experiment, **2DD** was again the best receptor in both experiments, with $A = 5.97$ and $K_a = 47.7 \pm 3.2 \text{ M}^{-1}$ ($K_d = 21.0 \text{ mM}$). **2P** showed a much lower but still notorious $A = 2.35$, and its analogue **1P**, a $K_a = 31.9 \pm 1.8 \text{ M}^{-1}$ ($K_d = 31.3 \text{ mM}$). **P**, **D**, **2PP**, and **1D** showed smaller or no binding, in concordance with their more discrete A values (Figure 60). ITC binding curves included in Annex 4.

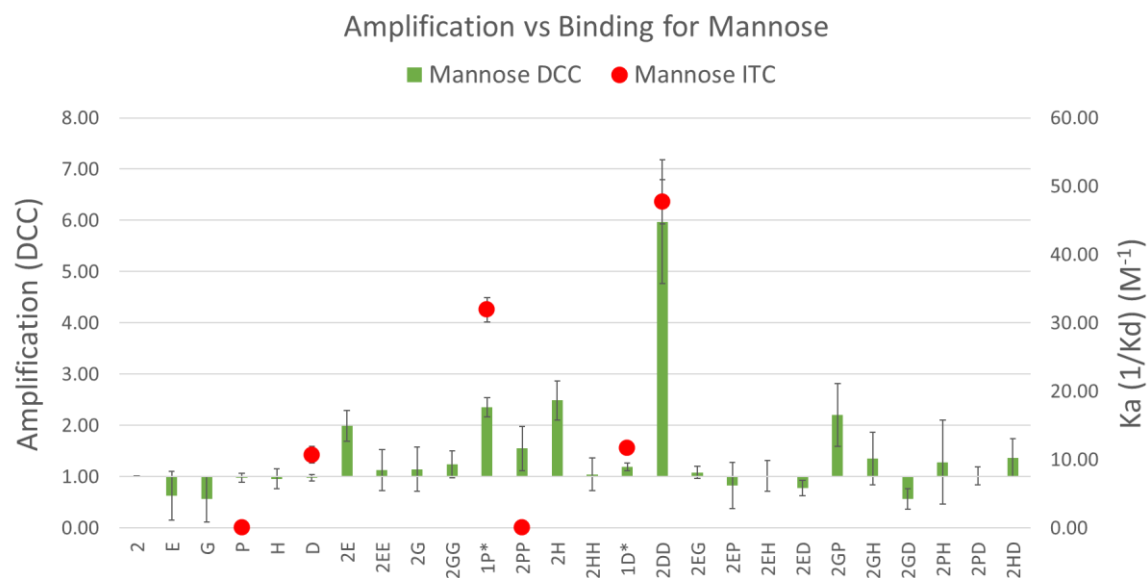


Figure 60: Amplification values obtained in DCC experiment (y axis, left-hand side) and K_a constants obtained in the ITC titrations (y axis, right-hand side) employing D-Mannose as template/titrant. It must be kept in mind that molecules **2D** and **2P** formed in the DCL were not tested by ITC. Instead, their analogues **1D** and **1P** did.

In the experiment with galactose, **2DD** was once again the best receptor in both experiments, with $A = 2.96$. As discussed in the sub-chapter dedicated to the DCC experiments, this experiment showed much higher deviation among the replicates, hence the large error bars. Nevertheless, ITC results still matched the DCC data, being **2DD** the best receptor with $K_a = 56.7 \pm 0.3 \text{ M}^{-1}$ ($K_d = 17.6 \text{ mM}$), followed by **1P** ($K_a = 17.9 \pm 2.6 \text{ M}^{-1}$; $K_d = 55.9 \text{ mM}$) (Figure 61). ITC binding curves included in Annex 4.

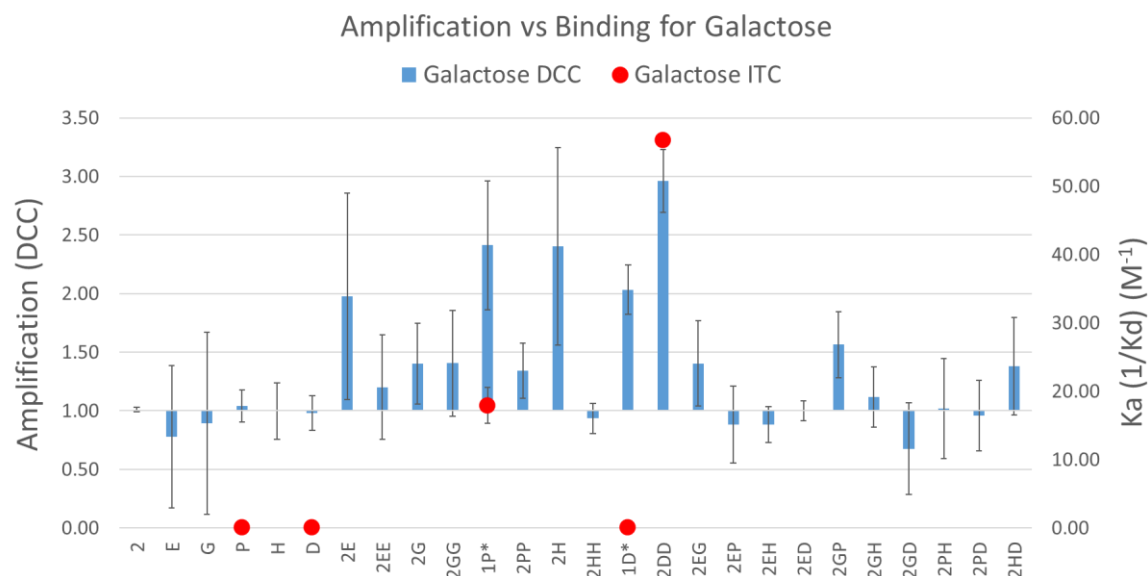


Figure 61: Amplification values obtained in DCC experiment (y axis, left-hand side) and K_a constants obtained in the ITC titrations (y axis, right-hand side) employing D-Galactose as template/titrant. It must be kept in mind that molecules **2D** and **2P** formed in the DCL were not tested by ITC. Instead, their analogues **1D** and **1P** did.

As we learnt with DCC 2.0 experiment, the results for fructose had to be studied separately. The distribution of its tautomers in solution, that is significantly different than in the other monosaccharides, led to a distinct amplification profile with mono-functionalised products **1P** ($A = 2.11$) and **1D** ($A = 1.91$) as the most amplified library members, completely outplaying **2DD** ($A = 0.89$). ITC studies were to determine whether this behaviour was indeed indicative of a completely different binding situation, or it was just a mistake and DCC experiment failed to predict the best receptors.

In contrast with the DCC results obtained for other sugars (especially for Glucose and Mannose), this time there was not a clear most amplified molecule, but two, and many other

library members were close behind them. This suggests a much more competitive scenario in which many molecules will have similar binding affinities for the target fructose.

However, Amplification values alone cannot give information regarding the actual binding affinity of the receptors for the targets, and even if the ITC results stick to the trend suggested by the DCC experiment, it could happen that no receptors have considerably good affinity for fructose.

Fortunately, not only they did bind fructose, but they did it with K_a values around 30 times higher than those showed by **2DD** for G, Ga, and M (Figure 62).

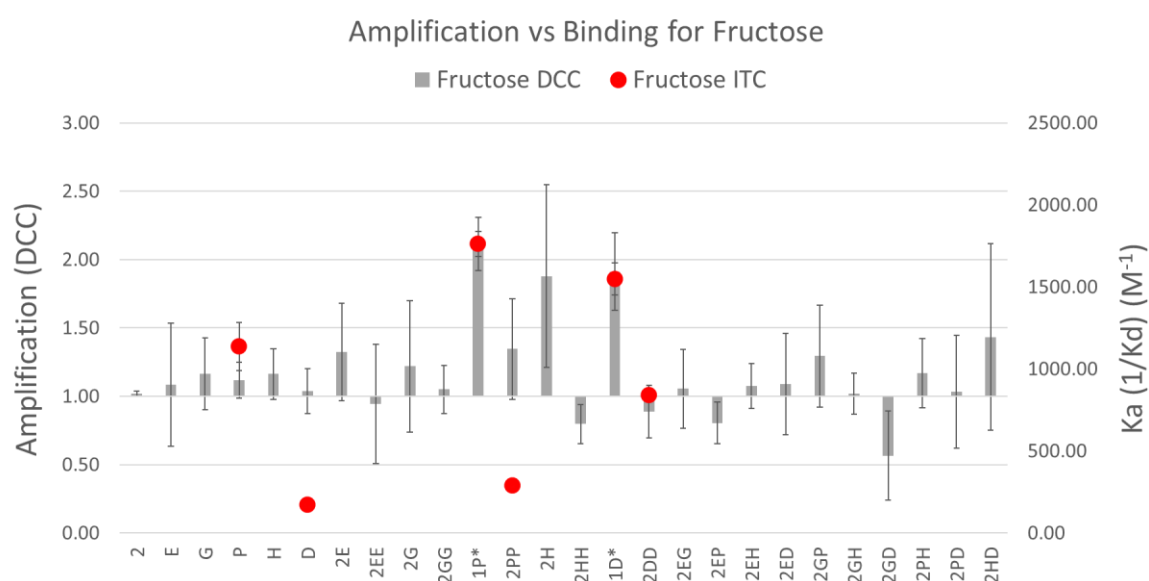


Figure 62: Amplification values obtained in DCC experiment (y axis, left-hand side) and K_a constants obtained in the ITC titrations (y axis, right-hand side) employing D-Fructose as template/titrant. It must be kept in mind that molecules **2D** and **2P** formed in the DCL were not tested by ITC. Instead, their analogues **1D** and **1P** did.

The best binders were, as predicted by DCC, **1P** ($K_a = 1762.0 \pm 162.4 \text{ M}^{-1}$; $K_d = 0.5 \text{ mM}$) and **1D** ($K_a = 1548.3 \pm 98.1 \text{ M}^{-1}$; $K_d = 0.65 \text{ mM}$). It was indeed a much more competitive situation as other library members tested displayed considerably high binding constants: **P** ($K_a = 1136.0 \pm 148.3 \text{ M}^{-1}$) or **2DD** ($K_a = 840.7 \pm 35.8 \text{ M}^{-1}$). What is perhaps more remarkable, is that the order of affinity was almost perfectly predicted by DCC: **1P** > **1D** > **P** > others.

D and **2DD**, with A values very close to 1; or **2PP** that exhibited a large error along the different DCC experiments replicates, did not fit the model as well. Nevertheless, DCC experiments successfully 'predicted' the best binders which was in the end the goal of the technique. ITC binding curves included in Annex 4.

Not only the ITC results matched the DCC predictions, but also, we discovered a remarkable ligand, **1P**, with affinity and selectivity for fructose in the range of best small-molecule receptors reported to date that work in aqueous media. While there have been reported synthetic receptors that show higher affinity for fructose, these are far more complex than **1P**. An example of it are the linear oligomers developed by Chamdramouli et. al. that fold to create binding sites with the pyranose forms of fructose with $K_a = 30\,000 \text{ M}^{-1}$.³²⁰ However, these systems can only work in organic media (4 : 1 CDCl_3 /[D6]-DMSO).

We also demonstrate a 50 to 100-fold selectivity of **1P** for fructose vs. other saccharides tested (Table 8). The affinity of **1P** for glucose was too small to be accurately calculated by ITC. This means that it is probably smaller than the measured for galactose and therefore the selectivity of **1P** for fructose over glucose must be greater than 98.4.

Table 8: Affinity of 1P for D-Fructose vs. D-Glucose, D-Galactose, and D-Mannose. Measurements carried out three times. Values shown are the a of the three measurements \pm standard deviation. Selectivity of 1P for D-Fructose vs. D-Glucose, D-Galactose, and D-Mannose calculated dividing their affinities (K_a values).

	Affinity of 1P for saccharides (K_a , M^{-1})	Selectivity of 1P for Fructose
D-Glucose	-	-
D-Galactose	17.9 ± 2.6	98.4
D-Mannose	31.9 ± 1.8	55.2
D-Fructose	1762.0 ± 162.4	1

More valuable information can be extracted from this experiment. **1P** and **1D** did not bind (or they did it with very low K_a) glucose, mannose, or galactose. Considering – as explained several times within this thesis- that the main difference between fructose and the other isomeric monosaccharides is its different tautomeric distribution (Figure 50), with higher abundance of the pyranose form (28.6 % vs 0.1 – 6.0 %) it can be concluded that **1P** and **1D** must ‘like’ the 5-membered furan form of the sugars. This could potentially be exploited to design even better receptors that are selective for fructose over other common saccharides.

These experiments could also throw some light on the binding mechanism between **2DD** and G, F, and Ga. The pyranose tautomeric form is the predominant for the 4 tested sugars is, with a similar distribution percentage too. Hence, the similar binding constant of **2DD** for G, Ga, and M (around $50 M^{-1}$ for them all) may be a proof of a similar binding process. The only feature that these 3 sugars share is the equatorial position of the hydroxyl group in C3 (Figure 63). Further studies need to be carried out, but we can hypothesise that C3 may be involved in the binding with **2DD**.

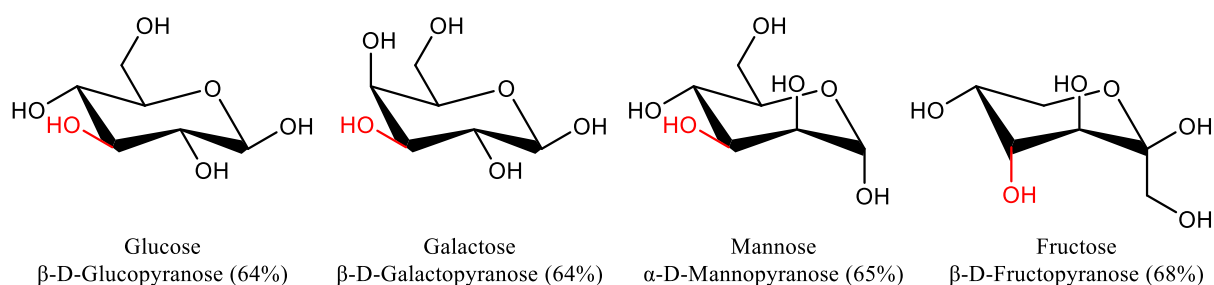


Figure 63: Chemical structure of the predominant tautomeric forms in solution for the four saccharides employed in the DCC 2.0. The percentage of each tautomer is indicated in brackets and the OH in C3 is highlighted in red.

The binding of **2DD** with fructose is much larger ($K_a = 840.7 \text{ M}^{-1}$) despite the axial position of the hydroxyl group in C3. However, as previously discussed, fructose is to be studied separately, and this mismatch in binding constants may arise from the abnormally high abundance of the furanose forms of fructose in solution.

5.2.3 Conclusions on the ITC experiments

ITC experiments served to demonstrate the validity of DCC as a robust, rapid, and resources-saving methodology to synthesise and screen a library of potential receptors against a particular biological target.

2DD was found to be a decent receptor for glucose, galactose, and mannose ($K_a = 44.5 \pm 1.8$, 56.7 ± 0.3 , and $47.7 \pm 3.2 \text{ M}^{-1}$, respectively). Although it must be said that there are reported synthetic receptors with much better affinity for such monosaccharides.

1P was found to be a considerably good receptor for fructose, with a $K_a = 1762.0 \pm 162.4 \text{ M}^{-1}$, value in the range of the best yet small-molecule receptors for fructose that can work in aqueous environments.

1P not only shows a good binding constant for fructose, but also it is selective for it over other monosaccharides tested.

Given the results obtained by ITC, some hypothesis can be proposed regarding the binding mechanism between the receptors and saccharides. The excellent selectivity of **1P** for fructose vs the other saccharides tested (glucose, mannose, and galactose) may arise from the differential distribution of their tautomers in water. Fructose presents an abnormally high percentage of its pyranose forms (28.6%), while this percentage is much lower for the other sugars tested (0.1 – 6%). It can then be hypothesised that **1P** must 'like' the 5-membered furan form of the monosaccharides. The chemical groups of **1P** responsible for the recognition event must fit better with this form of the sugar, hence the higher affinity shown.

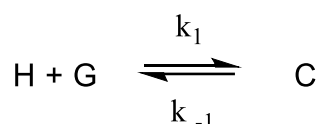
Finally, the differential behaviour of **2DD** with the set of saccharides could help us understand its binding mechanisms. **2DD** binds glucose, galactose, and mannose with a similar constant (K_a around 50 M^{-1}). Other than the 6-membered ring skeleton and the C6-OH, the only feature that the major tautomeric form of these 3 isomers share is the equatorial position of the hydroxyl group in C3. It can then be suggested that the equatorial OH (or perhaps, the axial H) in position 3 is directly involved in the binding event. Further ITC experiments could be carried out to study the binding of **2DD** with 3-O-Methyl-D-glucopyranose and D-Allose to corroborate this hypothesis. 3-O-Methyl-D-glucopyranose has the OH in C3 methylated, which should prevent it from acting as a proton donor and therefore should negatively affect its binding

affinity with **2DD**. D-Allose presents an axial OH in C3 which again should negatively affect its binding with **2DD**.

This information could potentially be exploited to design better receptors for monosaccharides in the future.

5.3 NMR titrations

NMR spectroscopy is a powerful technique to understand dynamic molecular processes. It can be employed in the field of molecular recognition to study the formation of host-guest complexes (Equation 5.4).



Equation 5.4

The strength of the complexation, or in other words, the binding constant, can be calculated from Equation 5.5.

$$K = \frac{k_1}{k_{-1}} = \frac{[C]}{[H][G]} = \frac{nc}{nH[G]} = \frac{nc}{(1-nc)[G]}$$

Equation 5.5

Where nH is the mole fraction of free (uncomplexed) host and nC is the mole fraction of complexed host.

Modern NMR instruments allow good quality spectra with even sub-millimolar concentrations of titrants,³²² which for us was a necessary requirement due to the poor solubility of some of

our ligands, and K_a values of up to 10^5 M^{-1} ,³²² which also falls within our scale according to previous ITC experiments.

To use ^1H NMR to study the complexation phenomena, at least one proton in the free host (or guest) molecule must produce a peak in the spectrum that is different from the same proton in the complexed molecule. This difference may be a change in the chemical shift (δ), the intensity (integral), or even its symmetry. In the most common scenario when there is a change in the chemical shift of the peaks involved in the binding (the peak could move either upfield or downfield), the magnitude of this difference ($\Delta\delta = \delta_{\text{H}} - \delta_{\text{G}}$) can give information about the structure of the complex and may be as large as several ppm (Figure 64).^{323,324}

Complex H-G

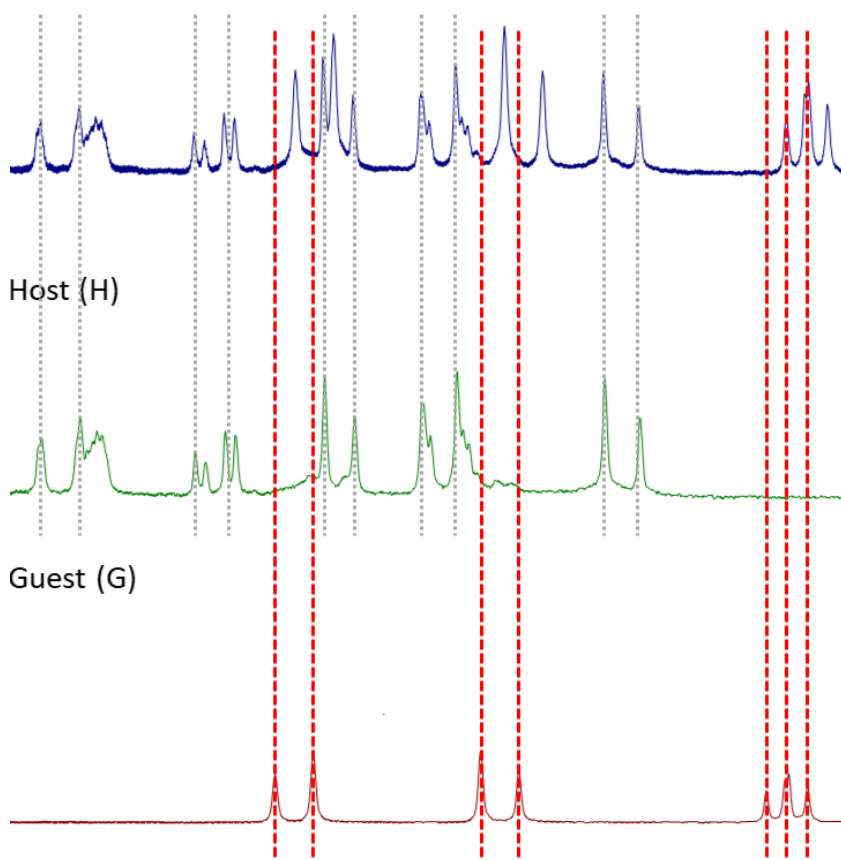


Figure 64: Real example of titration experiment carried out in this thesis. Peaks coming from the guest molecule (red, dashed line) shift upfield when forming the complex host-guest. By contrast, peaks coming from the host (grey, dotted line) did not move. This is titration 3 experiment and will be explained in more detail later in this chapter.

In order to re-validate the binding between **1P** and fructose and to obtain more insights regarding its mechanism, a ¹H NMR titration experiment was prepared.

Mixtures with different ratios of concentrations of fructose and **1P** were prepared and analysed by ¹H NMR. The solvent employed was the same as in the DCC and ITC experiments,

but completely deuterated: 100 mM carbonate buffer pH 10 prepared from Na₂CO₃, D₂O, and DCl to adjust the pH.

Employing the obtained binding constant for **1P** and Fructose, and utilising HySS software, it was calculated that the optimal concentration of reagents to perform the NMR titration experiment, in order to obtain the highest concentration of complex 1P-Fructose. The result was 10 mM for both **1P** and fructose. Under these circumstances, the conversion to complex Fructose-1P was >90% with [complex] = 9.38 mM (Figure 65).

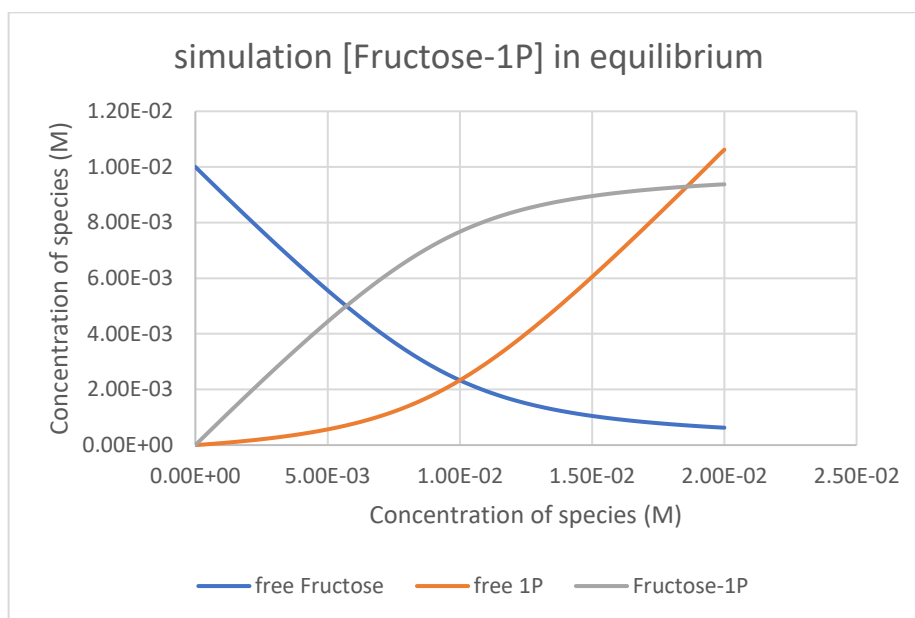


Figure 65: Simulation generated with HySS software to calculate the optimal concentration of reagents for a ¹H NMR titrations study.

However, similar to the issue we faced when designing the ITC protocol, concentrations had to be readjusted due to solubility limitations of the receptor.

1P could be perfectly solubilised at 1.8 mM. We wanted to attempt titration experiments with ratios 1:1, 2:1, and 1:2 to observe their differences. For comparability reasons, ^1H NMR spectrums of the individual species alone were recorded as well. Therefore, the conditions for the NMR titrations experiments were as represented in Table 9.

Table 9: Concentration of species for ^1H -NMR Titration experiments.

	[F] (mM)	[1P] (mM)	Ratio
Titration 1	0	0.9	-
Titration 2	0.9	0	-
Titration 3	0.9	0.9	1:1
Titration 4	0.9	1.8	1:2
Titration 5	1.8	0.9	2:1

When comparing Titrations 1, 2, and 3 (Figure 66 and Figure 67) some differences can be appreciated. The group of peaks from 7.44 to 7.20 ppm corresponding to the aromatic protons is slightly shifted upfield in the 1:1 titration in comparison with the receptor **1P** alone ($\Delta\delta = 7.4033 - 7.3977 = 0.0056$ ppm) (Figure 66a). While this is indeed indicative of binding, such a small shift may suggest that the aromatic rings are not especially involved in the binding and may not be the crucial factor behind molecular recognition. Signals coming from fructose, although not present in this fraction of the spectrum, did not move at all. Interestingly, there seems to be as well a change in the intensity of some peaks (Figure 66b). Disregarding the apparent higher intensity of all peaks in titration 1:1 (blue spectrum in Figure 66), which is simply due to a different visualisation setting in MestreNova software, four peaks in this aromatic region change their relative intensity with one another when comparing the

spectrum of **1P** and 1:1 titration. This is indicative of unequal relevance for the protons behind those peaks in the molecular recognition process. Unfortunately, the complexity of this region of the spectrum prohibits a deeper analysis of the situation.

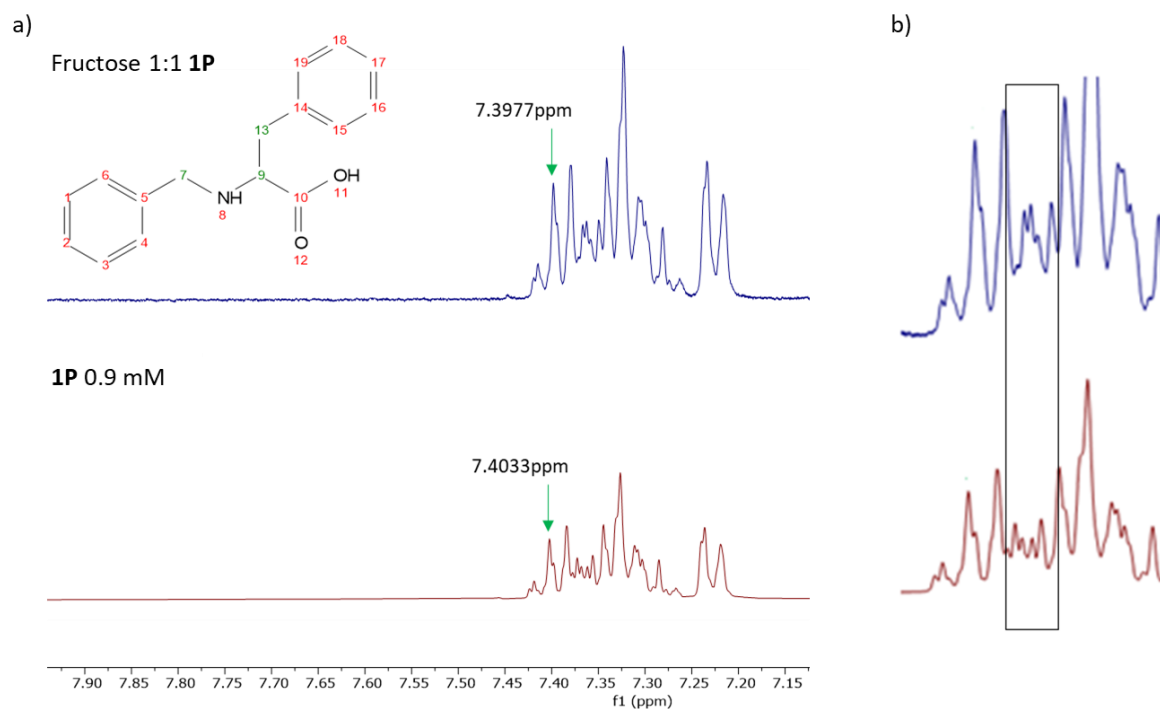
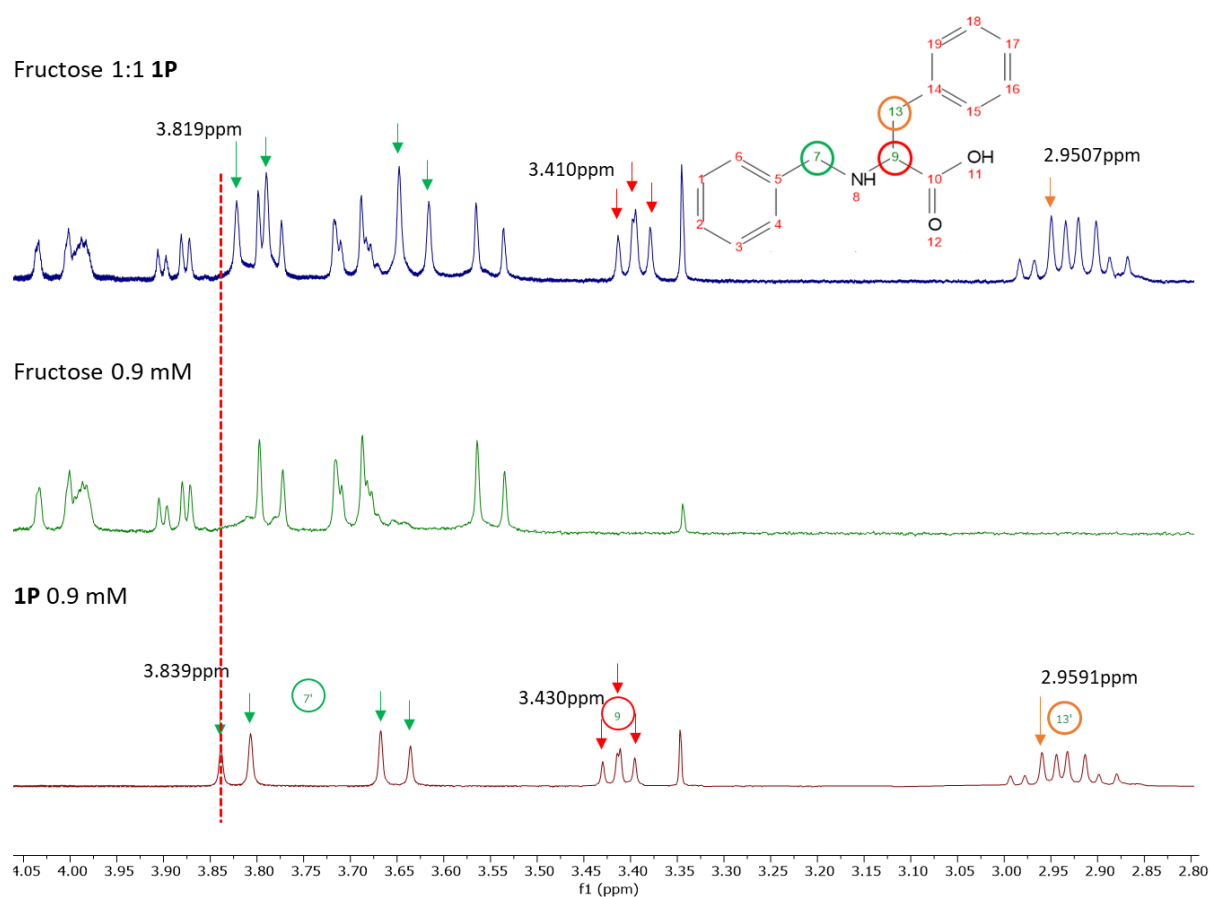


Figure 66: a) Overlapped fraction (7.9 - 7.15 ppm) of the ^1H NMR spectrums of receptor **1P** (bottom, red) and the titration experiment with ratio of concentrations 1:1 fructose:**1P** (top, blue). NMR spectrum of fructose alone not shown in this figure as fructose presented no peaks in this region of the spectra. b) Zoom in the region of the spectrum where the change in the relative intensity of the peaks is observed.

Other signals in **1P** were more significantly affected. The group of protons next to the secondary amino group were both the most affected in the molecule. Peaks at 3.83, 3.80, 3.66, and 3.63 coming from the benzylic protons on one side of the secondary amine and the signal at 3.43 ppm corresponding to the proton on the tertiary carbon on the other side of the

amino group, were both shifted upfield $\Delta\delta = 0.02$ ppm (Figure 67). This indicates a direct implication in the binding process. Lastly, the group of peaks at 2.99 – 2.87 ppm suffered a less abrupt shift upfield of $\Delta\delta = 0.008$ ppm (Figure 67).



*Figure 67: Overlapped fraction (4.1-2.8 ppm) of the ^1H NMR spectrums of receptor **1P** (bottom, red), fructose (middle, green), and the titration experiment with ratio of concentrations 1:1 fructose:**1P** (top, blue).*

In the fraction of the chromatogram displayed in Figure 67 it can be observed that the peaks coming from fructose did not shift at all.

In titration 5 experiment (ratio 2:1, excess of fructose) a similar behaviour is observed. By contrast, in titration 4 experiment (ratio 1:2, excess of **1P**) the shift in the signals is no longer observed and both the signals coming from **1P** and fructose match perfectly with those observed in the spectra of themselves alone (titrations 1 and 2, respectively) (Spectra included in Annex 5). This can be explained as, with an excess of **1P** - and keeping in mind that at the range of concentrations that we are using there is a low conversion to complex - most of the molecules of the receptor are in its free (uncomplexed) form and therefore that is the one visible by NMR.

Overall, the variation in the chemical shifts observed in titration 3 and titration 5 experiments is not large, but it must be kept in mind that due to the poor solubility of **1P**, we had to work with roughly one fifth of the optimal concentration of reagents (according to Hyss software calculations), hence the percentage of the complex **1P**-fructose in solution will not be large and the effects of molecular recognition in the chemical environment of the protons will be attenuated.

All of this suggests that the region near the tertiary amino group is the key motif responsible for the molecular recognition phenomenon (Figure 68). Even though the aromatic rings do not seem to be massively affected in the NMR titrations experiments, it should not be ruled out the possibility of them playing an important role. The fact that **1P**, with two aromatic rings, showed a greater binding constant by ITC than other similar molecules with only one aromatic structures (**P** or **1D**) may indicate that they are in fact needed in some way, perhaps by creating non-polar pockets that facilitate the recognition. It is noticeable as well that the analogue with

three aromatic rings, **2PP**, exhibited a binding constant six times lower than **1P**, suggesting that two is the optimal number of aromatic structures to achieve the best binding affinity for fructose with these types of receptors.

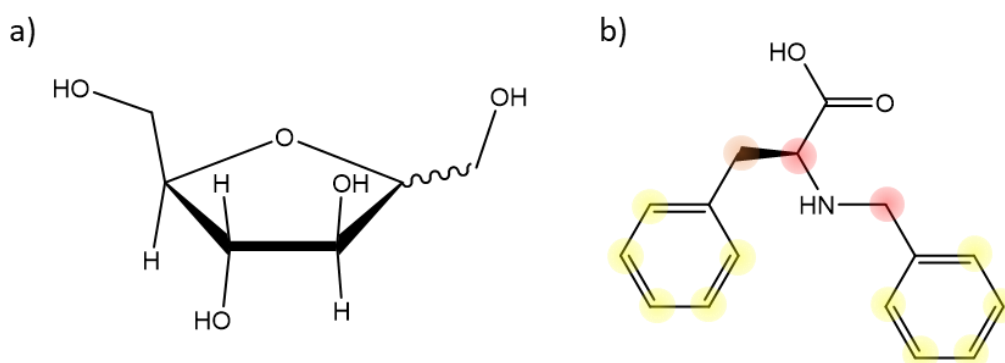


Figure 68: a) Molecule of fructose in its furanose form. The β tautomer would have the CH_2OH group pointing up, and the α tautomer would have it pointing down. b) Molecule **1P**. Hydrogens involved in the binding process with fructose, according to the titration experiments, are highlighted in red (most involved, $\Delta\delta = 0.02$ ppm), orange (less involved, $\Delta\delta = 0.008$ ppm), and yellow (least involved, $\Delta\delta = 0.0056$ ppm).

5.4 Conclusions of Chapter 5

The work carried out in this chapter serves as a validation of the previously obtained DCC results, demonstrating that DCC is a reliable technique that can save time and resources in the design of powerful receptors for saccharides. This is especially relevant as the synthesis of receptors for sugars, despite their huge biological relevance, has been far less explored than the synthesis of receptors for other relevant biomolecules such as proteins or peptides.

ITC results demonstrated that **2DD** is a good receptor for glucose, galactose, and mannose, with binding constants of around $K_a = 50 \text{ M}^{-1}$ for the three of them. Moreover, **1P** was found to be a magnificent receptor for fructose. It showed an affinity for it that is higher than most of receptors published to date ($K_a = 1762 \text{ M}^{-1}$) and selectivity over the other monosaccharides tested. With this information, hypotheses regarding the binding mechanisms sugar-receptor can be suggested. **1P** may bind tighter the 5-membered ring furanose form of saccharides than their 6-membered pyranose forms. With respect to receptor **2DD**, it is plausible that the equatorial position of the hydroxyl group in C3 (or perhaps, the axial H) in glucose, galactose, and mannose, is directly involved in their binding with such receptor.

NMR titrations experiments worked as another proof, in addition to ITC experiments, to corroborate the molecular recognition event between **1P** and fructose. In titration 3 and titration 5 experiments, when there is a detectable amount of complex **1P**-fructose, a displacement on the chemical shift of some signals was observed at a maximum of $\Delta\delta = 0.02$ ppm. This is not a large shift due to the impossibility to form more amount of complex because of the solubility limitations of receptor **1P**, but it is still a proof of binding. NMR titrations experiments also threw some light on the binding mechanism, as the group of protons near the secondary amide group of **1P** were significantly more affected than the others in the molecule, suggesting that they are directly responsible for the molecular recognition event.

Chapter 6 -DCL – Peptide approach

Encouraged by the results obtained in DCC 2.0 experiment and their validation with ITC and NMR techniques, a much more complex and ambitious approach was attempted. For this experiment, the scaffold of the DCL would be a short peptide and the reacting BBs would be small molecules containing potentially good recognition units. The peptide will pose primary amines in the side chain of some of its constituent amino acids, and the reacting building blocks will incorporate an aldehyde group so they can react via IFR.

This peptide work was carried out at the Institute of Advanced Chemistry of Catalonia (IQAC), in Barcelona, Spain. There, I carried out the synthesis and derivatisation of peptides as a member of Dr Miriam Royo's group, as they are experts in the synthesis of peptides for biomedical applications.^{325–327} While in IQAC, I also worked with Prof. Ignacio Alfonso, an expert on the use of DCC for the discovery of novel receptors for biological targets,^{148,149} to test the peptides in real DCC experiments.

6.1 Benefits of using peptides as scaffold in DCC experiments

The benefits of using peptides are numerous. They are easy to make via solid phase peptide synthesis (SPPS). SPPS methodologies allow the creation of peptides in short times (2-3 h per amino acid coupling), and without the need of purification procedures in between couplings. Only one purification step – usually by HPLC - is necessary in the very last step, after the cleavage of the final peptide from the solid support. There are also well-established methodologies that allow the synthesis of peptides in high yields both manually (normally used for short peptides, <5 amino acids) and by automated methods (≥5 amino acids) using instruments called peptide synthesisers.

SPPS also carries some disadvantages as it normally requires larger amounts of reagents and solvent, but the positive aspects clearly outweigh the negatives.

Peptides are also very easily customisable. Modern peptide synthesisers incorporate technologies that allow the efficient synthesis of peptides, and/or the parallel synthesis of a number of them. During my PhD I had access to the CEM LibertyBlue synthesiser that incorporate microwave technology to improve the efficiency of the peptide synthesis and to the CEM MultiPep 1 instrument that can make up to 196 different peptides in parallel at the same time. This opens the doors to a whole new category of molecules to be included in DCLs for DCC experiments.

In the simplest scenario in which only 2 different amino acids (A and B, as an example) are considered to build a peptide, one can play with the length of the peptide (AB, ABA, ABAB, etc), the position of the amino acids in the chain (ABAB, AABB, ABBA, etc), or the times that a particular amino acid is present in the structure (AAAB, AABB, ABBB, etc). It is also worth reminding that peptides have a direction. They are not symmetrical (ABAB and BABA are 2 different molecules with potentially different properties).

Including more different amino acids (C, D, etc) would increment exponentially and make virtually infinite the total number of combinations that can be synthesised and tested in a DCC experiment.

Peptides can be flexible. It seems to be commonly accepted that rigidity plays in favour of affinity when designing a receptor. Most of the synthetic receptors found in the literature contain indeed highly rigid structures such as the calixarenes⁵¹ or cucurbiturils,^{328,329} cages with a defined shape,³³⁰ or even switchable molecular grippers.^{331–334}

However, some authors claim that having flexibility in receptors may play an important role in molecular recognition too. Forrey et. al. developed an *in silico* model to study the effect of molecular flexibility on non-covalent binding, and ultimately binding constant K_a .³³⁵ They concluded that for rigid targets, a decrease in rigidity in a ligand indeed affects negatively its binding constant for such target. On the other hand, they found that an increase in the flexibility of the receptor plays in favour of its affinity for more flexible targets.

In most cases the binding sites or the binding mechanisms of the targets are unknown and therefore a flexible receptor that can fold and adapt to the receptor's epitope could be beneficial. The degree of flexibility of a peptide can also be predetermined by the cautious selection of its amino acids. Small and linear amino acids (Gly, Ala, Ser, Thr, Cys, Pro) are generally more flexible than bulkier ones (Phe, Tyr, Trp, His, Arg). We hypothesised that an ideal receptor should have a flexible backbone with more rigid regions that work as recognition units, responsible for the binding with a target.

6.2 Selection of BBs for peptide-based DCC

As the placement in Spain was set to be of only 3 months, we designed and synthesised a small and apparently simple peptide for it to be then tested in a DCC experiment.

The design of the peptide was purely based on two parameters: simplicity and feasibility. It was decided to employ a short five amino acid peptide, with only two different amino acids. One of them had to contain a reactive primary amine in the side chain, and the other would just act as a spacer to keep the reactive points separated and avoid congestion that could cause steric hindrance. Lysine would be the obvious choice for the first case, but the length of the side chain (4 $-CH_2-$ groups) would induce too much flexibility into the system. We want the peptidic backbone to be flexible, not the recognition units. The analogue with the shortest side chain, diaminoacetic acid would not be the best choice either as the reactive primary amine would be too close to the peptidic backbone and that could hinder its reactivity (Figure 69).

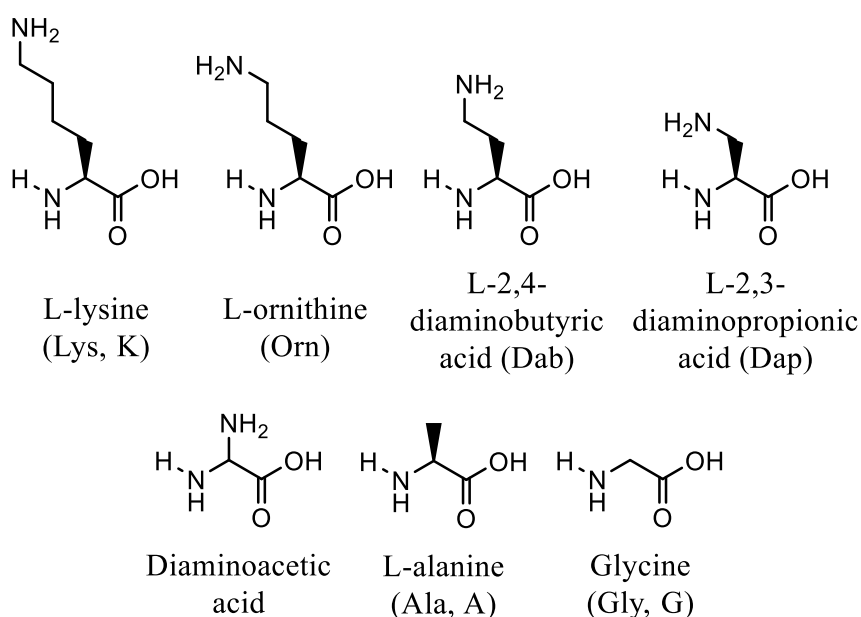


Figure 69: Set of amino acids considered for the peptidic scaffold of DCC 3.0 experiment. Dap and Gly were finally selected.

In this context, the unnatural amino acid L-2,3-diaminopropionic acid (Dap) was selected as it was a good compromise between reactivity and rigidity. As spacer, we selected Glycine (Gly) as it is the simplest natural amino acid. The peptide to synthesise and test would therefore be: (NH₂)Dap-Gly-Dap-Gly-Dap(CONH₂) (Dap-G-Dap-G-Dap, for short, or **4**). Note that this peptide has a total of four reactive points as there are four primary amines susceptible of IFR in the presence of aldehydes. One in each Dap side chain, plus another one in the N-terminal position. The C-terminal position can be in the form of an amide (peptide **4**) or a carboxylic acid (peptide **4.1**), depending on the characteristic of the resin used for the solid synthesis of the peptide. For the purpose of conducting DCC experiments it does not make a difference which one of those two peptides is employed as amides, less nucleophilic than free amines, do not undertake IFR and there are therefore only four reactive positions in both cases.

As we could not know in advance which peptide would be easier to synthesise, we decided to make both.

As aldehyde-containing BBs, it was selected the molecules 1-Hydroxy-1,3-dihydro-2,1-benzoxaborole-7-carbaldehyde (**Q**), 5-imidazolecarboxaldehyde (**R**), benzaldehyde (**S**), and Indole-3-carboxaldehyde (**T**) (Figure 70). These were selected for similar reasons as detailed in the chapter dedicated to DCC 2.0. Molecule **Q** contained a benzoboroxole group, that is able to bind diols in a similar way as the boronic acid group does, but at pH closer to physiological. **R**, **S**, and **T** were included in an attempt to mimic Nature as they, once attached to the peptidic scaffold, would resemble to amino acids Histidine, Tyrosine, and Tryptophan respectively, and these amino acids are often found in natural receptors.

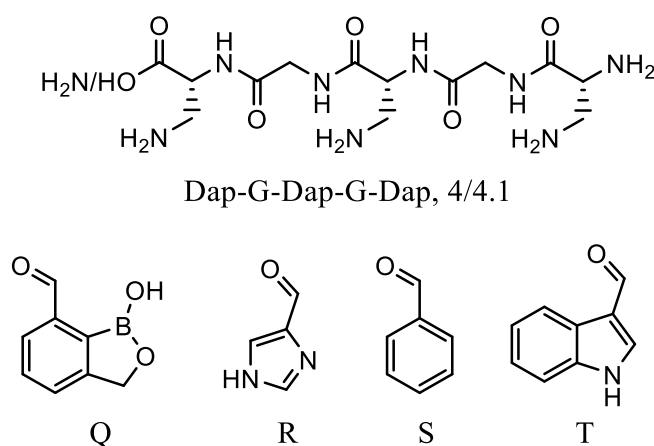


Figure 70: Starting BBs of DCC 3.0 experiment.

Syntheses of peptide **4** and **4.1** was carried out by solid-phase peptide synthesis (SPPS).

6.3 Solid-phase peptide synthesis (SPPS) to make peptides for DCC 3.0 experiments

Peptides **4** (CONH₂ group in C-terminal position) and **4.1** (COOH group in C-terminal position) were synthesised manually by SPPS at room temperature and at 40°C. A schematic representation of the synthetic route is shown in Figure 71.

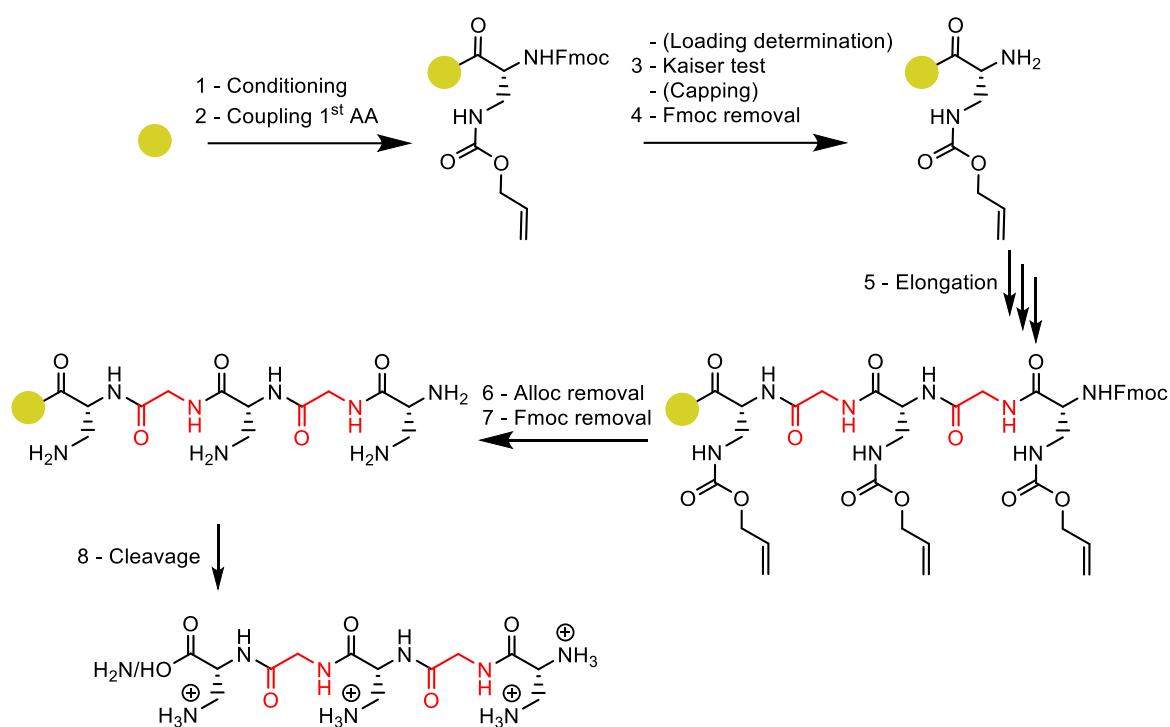


Figure 71: Diagram showing the synthetic route to afford peptides **4.1** and **4.2** via SPPS.

The first amino acid was attached to the polymeric support by its C-terminal end, and a standard C→N strategy was followed. The protecting groups employed were 9-fluorenylmethyloxycarbonyl (Fmoc) for the N-terminal end and allyloxycarbonyl (Alloc) for the amino group in the amino acid side chains. These groups are orthogonal as one can be selectively removed over the other (Figure 72).

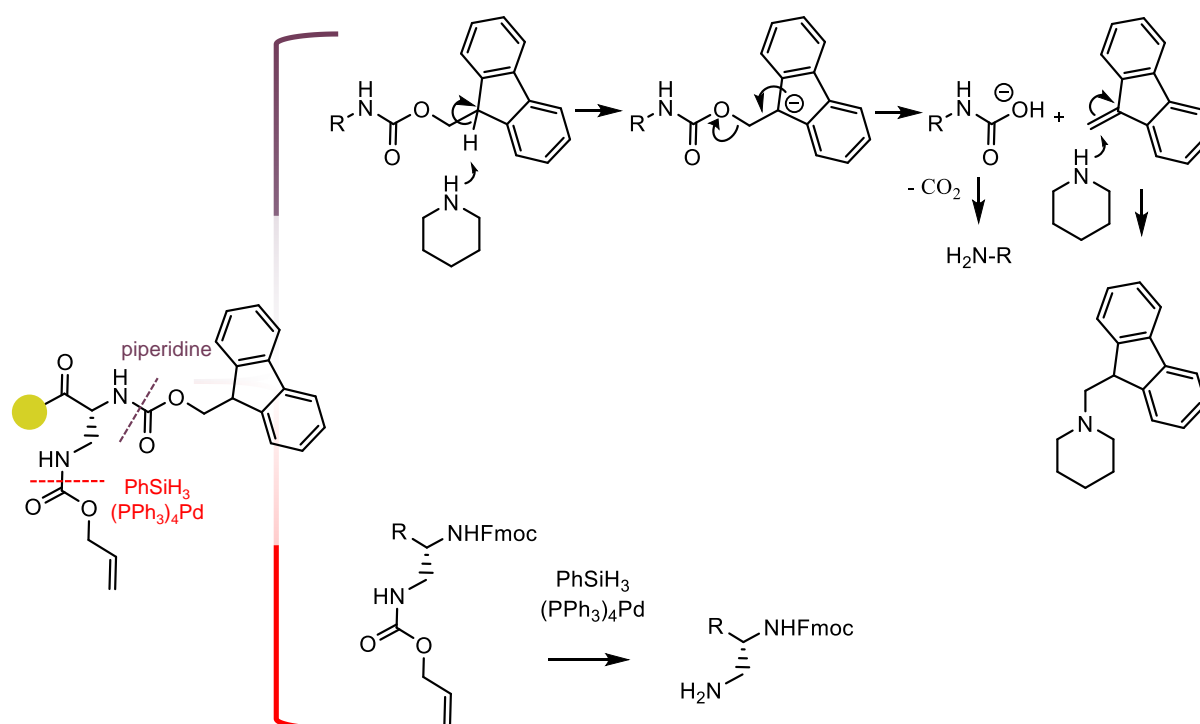


Figure 72: Orthogonality of Alloc and Fmoc protecting groups.

6.3.1 General considerations

Solid-phase synthesis was carried out manually in polypropylene syringes fitted with polyethylene filter discs or, when the reactions were conducted other than at room temperature, in a 250 mL double layer jacketed glass reactor vessel equipped with a porous glass filter disc. When using a syringe, stirring was done with a Teflon bar or in an orbital shaker; while using a reactor vessel, mixing was done with a flow of Ar gas. The Fmoc/Alloc protection method was used to synthesize each and every peptide. All solvents and soluble chemicals were drawn out using vacuum suction (Figure 73).

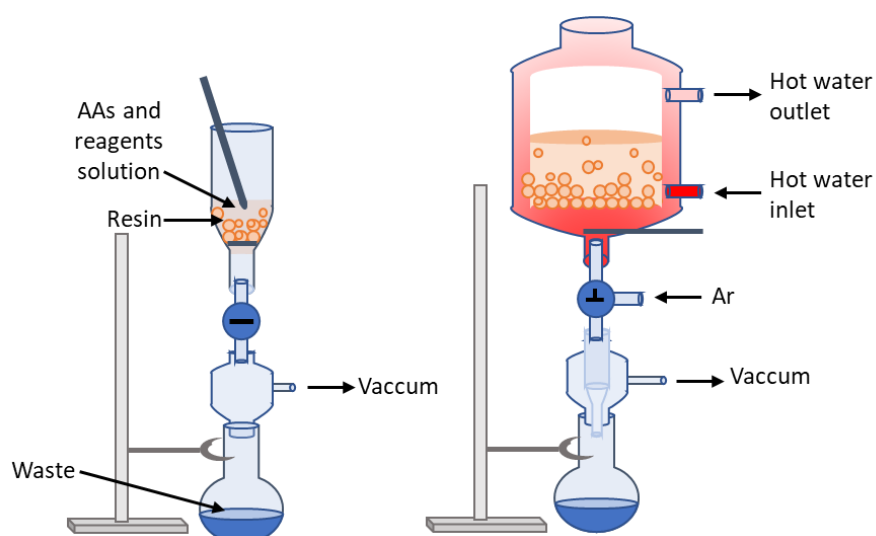


Figure 73: Schematic representation of the instrumentation used to carry out manual SPPS at room temperature (left) and with heating (right).

6.3.2 SPPS resins

Two different resins were employed in this thesis: the rink amide resin was used for the synthesis of **4**, and 2-chlorotrityl chloride (2-CTC) resin and rink amide resin was employed to make **4.1**.

Rink amide resin is a polystyrene resin functionalized with a 4-methylbenzhydrylamine linker and the primary amine protected with a Fmoc group. Hence, it is needed first to remove the Fmoc group so the free amine can react and form an amide with the carboxylic acid of the amino acid. This resin is typically used for the synthesis of peptides with C-terminal amides (Figure 74a).

2-CTC resin is a polystyrene resin functionalized with chlorotriptyl chloride groups. It is widely used in SPPS for the synthesis of peptides with C-terminal carboxylic acids. The attachment of the peptide to the resin occurs through a nucleophilic substitution reaction of the amino group in the amino acid and the chlorotriptyl chloride group of the resin (Figure 74b).

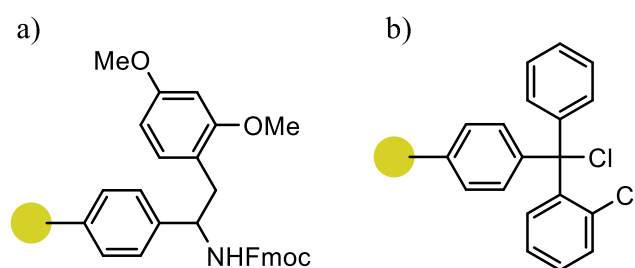


Figure 74: Structure of a) rink amide and b) 2-CTC resins.

6.3.3 Conditioning of the resin and incorporation of the first amino acid

Two different resins were employed in this thesis at a scale of 250 mg with each one: Rink amide and 2-CTC resins. The conditioning and loading of the resins were performed differently for each one:

Rink amide resin (100-200 mesh. Loading 0.5 mmol/g of resin): 260 mg of resin were washed with DCM to induce the swelling of the resins (1 x 60 min, 1 x 1 min), DMF (2 x 1 min), piperidine 20% in DMF to remove the Fmoc protecting group present in the resin hence activating it (2 x 4 min) and DMF (2 x 1 min). For the loading, the first Fmoc-protected amino acid was added together with Oxyma (3 equiv. each). Then DMF is added (The minimum amount of it so the mixture can be stirred). Next, DIC (3 equiv.) was added, and the mixture is

stirred for 1h (Figure 75). Finally, the solvent, excess of reagents, and side products were vacuum- suctioned, and the resin washed with DMF (2 x 1 min):

- DCM 1 x 60 min ; 1 x 1 min
- DMF 2 x 1 min
- Piperidine 20% in DMF 2 x 4 min
- DMF 2 x 1 min
- AA, Oxyma, DMF, DIC 1 x 60 min
- DMF 2 x 1 min

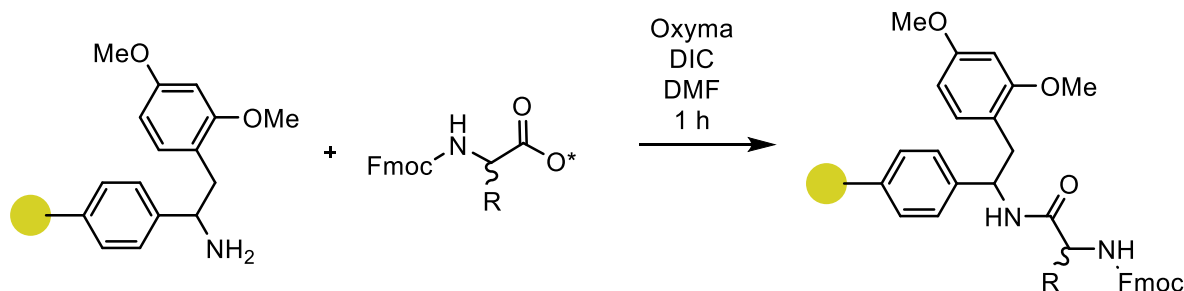


Figure 75: Schematic representation of the incorporation of the first amino acid on a rink amide resin.

Notice that the Fmoc group in the resin has already been removed. The star (*) on the amino acid carboxylic group represents that it has already been activated by the combination of oxyma and DIC.

2-CTC resin (100-200 mesh, 1% DVB. Loading 1.6 mmol/g of resin): 300 mg of resin were washed with DCM (1 x 30 min, 1 x 1 min), DMF (2 x 1 min) and DCM (2 x 1 min). For the first amino acid to be attached, a solution of the Fmoc-protected amino acid (1.5 equiv.) in DCM was added to the resin along with just enough DIPEA to dissolve the amino acid. Following the slow addition of more DIPEA (3.0 equiv.), the reaction mixture was agitated for 1.5 hours

(Figure 76). Then, 0.8 mL of MeOH/g of resin was added and stirred for 30 min to cap the unreacted positions. The solvent was removed by vacuum suction to then perform a DCM wash (3 x 1 min) and a MeOH wash (1 x 1 min). The syringe with the 2-CTC resin was left drying overnight in a vacuum oven so it was prepared for the quantitative analysis of Fmoc in the next step.

- DCM 1 x 30 min ; 1 x 1 min
- DMF 2 x 1 min
- DCM 2 x 1 min
- AA, DCM, DIPEA 1 x 90 min
- MeOH
- DCM 3 x 1 min
- MeOH 1 x 1 min

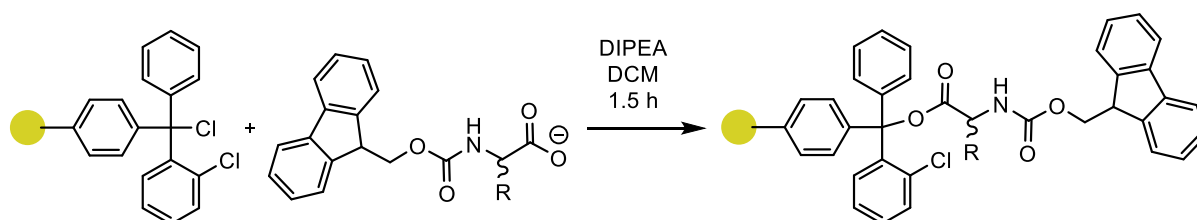


Figure 76: Schematic representation of the incorporation of the first amino acid on a 2-CTC resin.

6.3.4 Loading determination by Fmoc quantification

When using a 2-CTC resin it is common an uncomplete loading of first amino acid (hence the capping step). This means that the loading capacity given by the manufacturer may not correspond with the actual amount of amino acid that gets attached to the resin. In order to

have more certainty on this and correct the equivalents of reagents employed in future steps, it is wise to perform a loading determination.

This was determined using UV - Vis spectrophotometry by quantifying the dibenzofulvene adduct formed after the Fmoc removal of the resin. Thus, two small aliquots (around 10 mg) were taken from the resin that was previously dried. They were weighed and placed into two different 50 mL volumetric flasks and then diluted to the mark with 20% piperidine in DMF. The solutions were left in the ultrasonic bath for 30 min. Then, 1 mL of each one was filtered and their UV absorbance (A) at 301 nm was measured. The Lambert-Beer law (Equation 7.1) was used to determine the concentration of the dibenzofulvene adduct (concentration = c ; $\epsilon = 7800 \text{ M}^{-1} \text{ cm}^{-1}$ at 301 nm with a $l = 1 \text{ cm}$).

$$A = c * \epsilon * l \quad \text{Equation 7.1}$$

By dividing the total number of mmol of amino acids that were attached into the resin (calculated from the quantity of dibenzofulvene adduct) by the amount of g of resin that was weighed, the loading of the aliquot can be determined with Equation 7.2:

$$\text{Loading determination} = \frac{\text{mmol of amino acid}}{\text{g of resin}} \quad \text{Equation 7.2}$$

6.3.5 Capping of unreacted positions

In some cases, during the loading of the resin or in a coupling cycle, the functionalisation is incomplete. This could lead to errors in the peptide growing (deletions) and therefore the formation of undesired products. This effect gains more importance the longer the proposed

peptide is. When the length of the suggested peptide rises, the overall yield decreases because the concentration of full-length peptide produced is inversely associated with the proposed peptide's length. To prevent this from happening, the best tool that we have is to repeat the loading or cycle, perhaps with different reagents that can of course achieve the same goal. If this fails, a capping must be carried out to stop the elongation. When capping a position that has already been functionalised before (i.e., there was an ongoing peptide there), such peptide will be cleaved in the final step together with the desired peptide, hence becoming an impurity. However, it is important to cap it as soon as the failure occurs as the smaller the impurity is in comparison with the target peptide, the easier it is to be removed during the purification process.

There are two widely common strategies to carry out a capping. When capping a resin such as 2-CTC resin, 0.8mL of MeOH/g of resin and DIPEA (3.0 equiv.) for 30 min in DCM is usually employed to methylate the unreacted positions (Figure 77)

- MeOH (0.8 mL/g resin) + DIPEA (3 equiv.) in DCM, 1 x 30 min

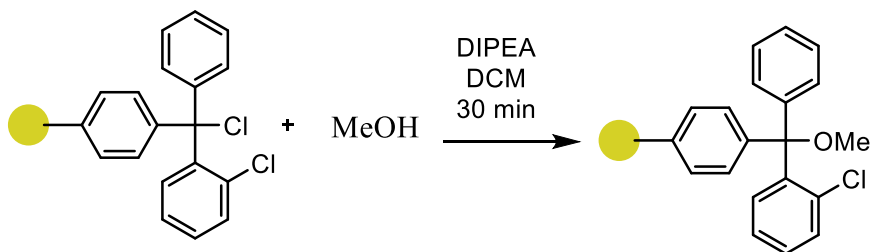


Figure 77: MeOH capping of a 2-CTL resin.

When capping a growing peptide while following standard C→N growing strategy, one must block the N-terminal free amino group. An effective way of doing it is its acetylation: 1 x 30 min wash with acetic anhydride and DIPEA (10 equiv. each) in the minimum amount of DCM that allows the effective stirring of the mixture (Figure 78).

- Ac₂O (10 equiv.) + DIPEA (10 equiv.) in DCM, 1 x 30 min

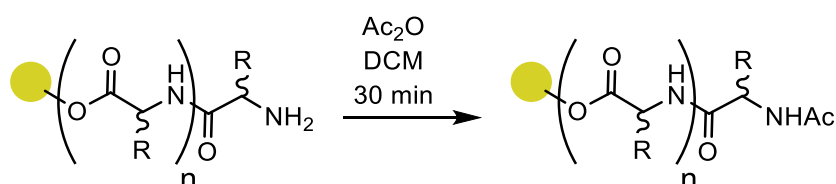


Figure 78: Ac₂O capping of the N-terminal position of a peptide.

6.3.6 Kaiser test

Whenever it is necessary to check the absence of free amines, a Kaiser test can be performed. This could be needed, for instance, after a coupling or capping step in SPPS as an indication of reaction completion. The Kaiser or Ninhydrin test is a colorimetric technique based on the reaction of ninhydrin with a primary amine coming from the unfunctionalised peptide to produce Ruhemann's purple, a dark blue chemical (Figure 79).

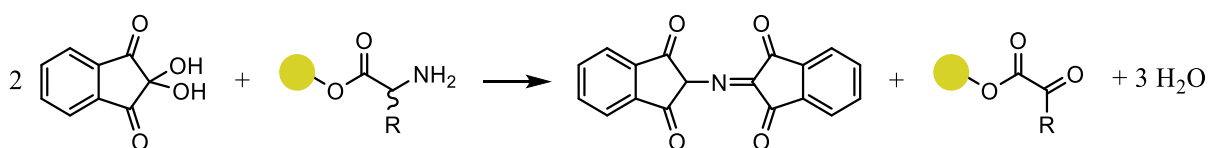


Figure 79: Kaiser test reaction between ninhydrin and primary amines.

To perform the assay, three solutions were prepared:

Reagent A: 5 g of ninhydrin in 100 mL of EtOH.

Reagent B: 80 g of phenol in 20 mL of EtOH.

Reagent C: 2 mL of KCN 0.001 M in 98 mL of pyridine distilled over ninhydrin.

The peptidyl resin is first washed with DMF to get rid of any extra reagents, and then it is dried with MeOH. Two drops of each reagent are added to a tiny aliquot of the resin in a test tube. The tube is then heated for 3 minutes at 110 °C. To compare colours, a blank sample using only the chemicals and no resin is carried out simultaneously. If the sample turns dark blue (positive result), that indicates the presence of free amines and therefore an incomplete coupling. If it is pale yellow -as the control test- (negative result) it is signal of complete functionalisation.

6.3.7 Fmoc removal

Previous to a coupling cycle, the N-terminal of the peptide that is protected with Fmoc must be deprotected. The methodology employed in this thesis involved the use of 2 x 5 min washes with 20% piperidine in DMF as base, followed by 2 x 1 min washed with DMF (Figure 80).

- Piperidine 20% in DMF 2 x 5 min
- DMF 2 x 1 min

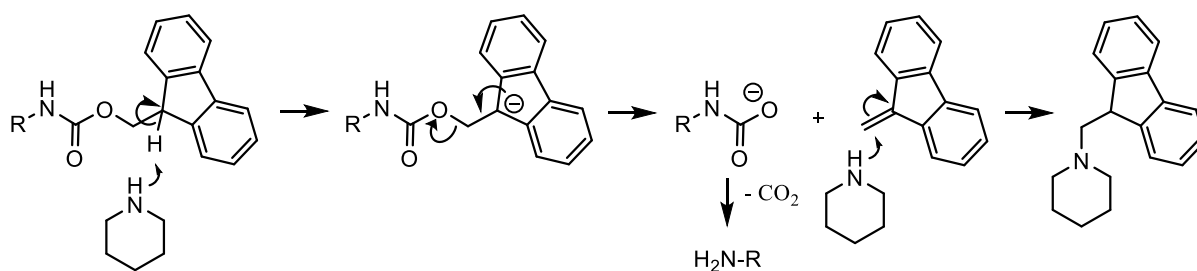


Figure 80: Fmoc removal using piperidine.

6.3.8 On-resin elongation of the peptide

Our first approach to the synthesis relied on the use of Fmoc-protected amino acid and the conventional and economic DIC and Oxyma (3 equiv. each) as coupling agents (Figure 81). K-Oxyma was used instead in the synthesis with the 2-CTL resin, as the subtle acidity of regular Oxyma can induce partial cleavage of the peptide when using that resin. Gly proved to be less reactive, and it was needed the more powerful reagent, PyBOP (3 equiv.) with DIPEA as base (6 equiv.), for the coupling. Both reaction mechanisms for the activation of the amino acid and coupling are shown in Figure 81.

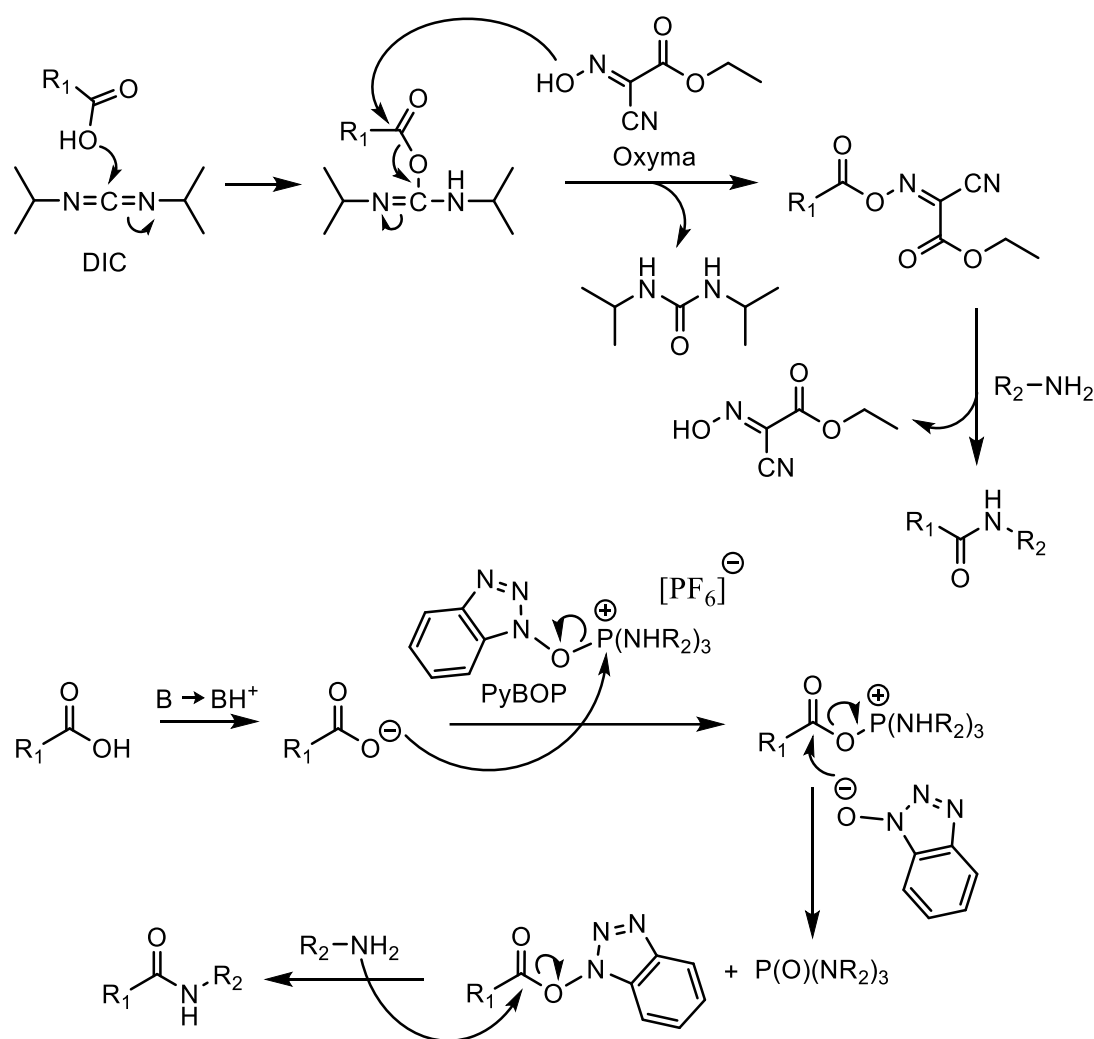


Figure 81: Reaction mechanisms for the amino acid coupling when employing DIC/Oxyma (top) and PyBOP/Base (bottom) as activating agents.

The reactions were monitored with the ninhydrin test, and 2-3 cycles were needed in all the cases. In the reactions of incorporation of the 3rd and 4th amino acids onto the peptide chain (Dap and Gly, respectively), coupling did not go to completion and the colorimetry test would result positive (dark blue). Therefore, capping steps with acetic anhydride had to be performed. Fmoc deprotection reactions were carried out with 20% piperidine in DMF as explained in a previous chapter. Figure 82 shows a schematic representation of the SPPS protocol employed.

- AA, (K) Oxyma, DIC, DMF 2 x 1 h // AA, PyBOP, DIPEA, DMF 2 x 1 h
- DMF 2 x 1 min
- Kaiser test

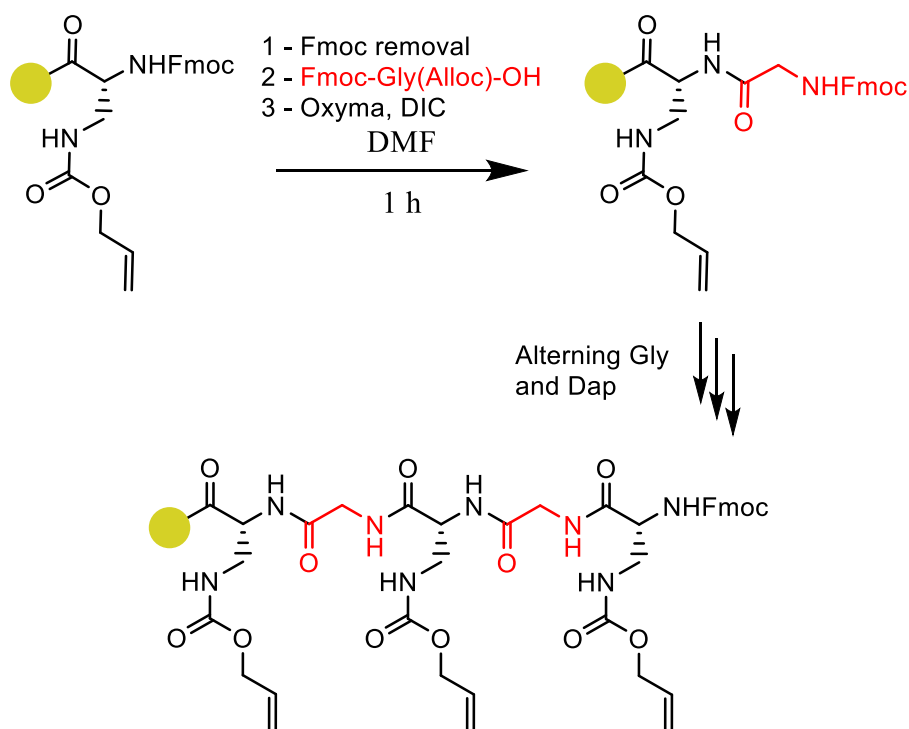


Figure 82: Schematic representation of the SPPS protocol.

6.3.9 Alloc removal

After the last coupling step, Alloc protecting groups were to be eliminated. They were removed with the combination of $\text{Pd(PPh}_3)_4$ (0.1 equiv.) in the presence of phenyltrihydrosilane (PhSiH_3) (10 equiv.) as a neutral ally group scavenger.³³⁶ The reaction mechanism consisted of a palladium-catalyzed transfer of the allyl unit to phenyltrihydrosilane in the presence of a proton source. Three treatments of 10 min each were needed with the minimum amount of DCM as solvent to allow an efficient mixing.

On a last step, a Fmoc removal reaction was carried out following previously detailed methodologies (Figure 83).

- Pd(PPh₃)₄, PhSiH₃, DCM 3 x 10 min
- DCM 2 x 1 min
- DMF 2 x 1 min
- Piperidine 20% in DMF 2 x 4 min
- DMF 2 x 1 min

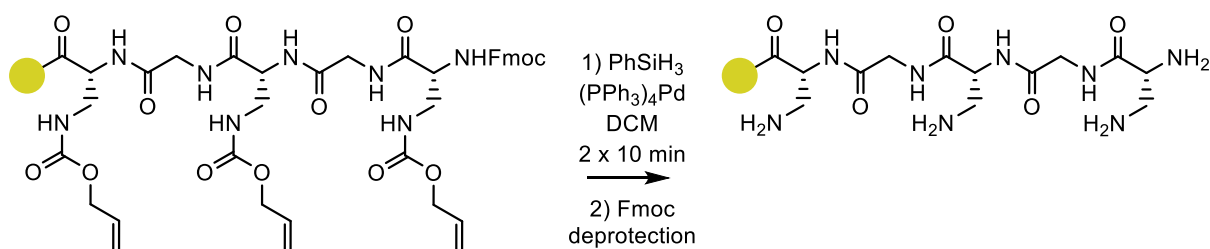


Figure 83: Alloc and Fmoc deprotection steps.

6.3.10 Peptide cleavage

In the Fmoc/Alloc strategy, TFA, a relatively weak acid, is used for the break of the peptide-resin bond and consequently cleavage of the peptide. This acidic cleavage produces highly reactive carbocation species and therefore it is needed to employ scavenger species to minimise side reactions produced by these species with sensitive amino acids, or disulfide bonds formation in Cysteine containing peptides. A scavenger commonly used is triisopropyl silane (TIPS) that acts as a hydride donor under acidic conditions.

The peptide chain was removed from the resin and dissolved by using the acidolytic mixture TFA/TIPS/H₂O (95:2.5:2.5, v/v) 1 x 30 min; 1 x 60 min; 1 x 1 min (Figure 84). Due to the nature of the resins, they would afford a slightly different peptide. The 2-CTC resin would produce **4.1**, with a carboxylate group in the C-terminal position while the rink amide resin would generate **4**, with an amide group in such position.

- TFA/TIPS/H₂O 1 x 30 min ; 1 x 60 min ; 1 x 1 min

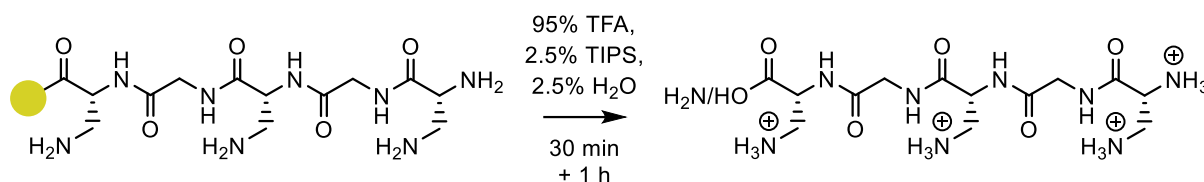


Figure 84: Cleavage reaction step.

6.3.11 Manual SPPS with temperature

The capping procedures needed during the formation of the peptides might be the major cause behind the low yields obtained. As some researchers claim that temperature can enhance the efficiency of the coupling reactions,³³⁷ we decided to do the synthesis of **4.1** once again (i.e. using the CTL-2 resin), now carrying out the coupling steps at 40°C. If this could avoid the need of capping along the synthetic route, it could perhaps lead to a more efficient SPPS. Only the coupling steps, involving the use of DMF as solvent, were carried out under temperature. The apparatus utilised to carry out the synthesis is shown in Figure 73 (General considerations section).

Unfortunately, despite this time all the coupling reactions went to completion after 2-3 cycles without the need of capping the peptides, the overall yield of the formation of peptide **4.1** was not considerably higher than in the previous synthesis at room temperature (38% this time vs. 32% at room temperature).

6.3.12 *On-resin reductive amination*

A DCC experiment with a peptide as scaffold and aldehyde-containing BBs or recognition units would afford a variety of peptide derivatives library members. Those with the best amplification for a target, must be individually synthesise for further testing. This synthesis can be done on solid phase following different methods:

Method 1. In the simplest scenario when the target molecule is a fully functionalised peptide (tetra-substituted product) with only one type of recognition unit (homopeptide, 4AAAA), the product can easily be made by carrying out the IFR on-resin in the end of the synthetic route, right before the cleavage of the resin (Figure 85). Different reducing agents and conditions were explored, and the best results were obtained with the following protocol: Adding the recognition unit (5 equiv. per amino group = 20 equiv.) in the minimum amount of DMF + 1% Acetic Acid (v/v); mixing for 1 h with an orbital shaker; adding the reducing agent (NaCNBH_3 , 5 equiv. per amino group = 20 equiv.) and mixing for another 30 min.

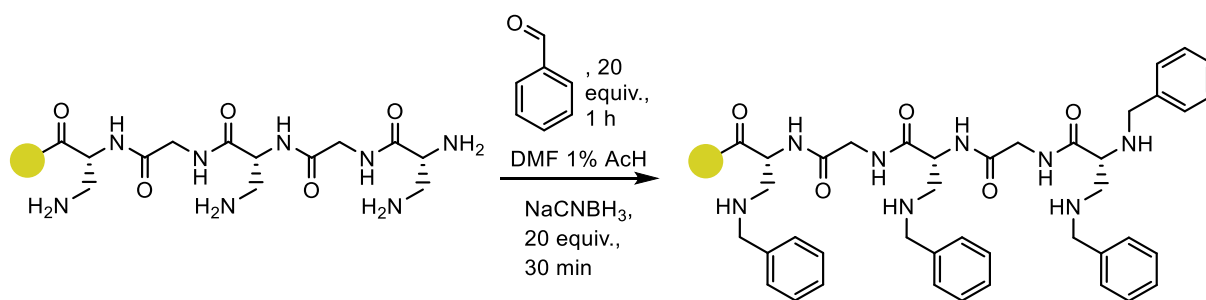


Figure 85: On-resin IFR protocol for tetra-functionalised homopeptides.

When non-fully functionalised peptides and/or peptide derivatives with more than one recognition unit attached were to be synthesised, the synthetic route had to be more elaborated.

Method 2. To afford fully functionalised peptides with more than one type of recognition unit and the same unit in positions 3 and 4 (tetra-substituted heteropeptides, 4ABCC), the Dap groups in the peptide were to be functionalised individually as soon as they are attached onto the growing chain.

Firstly, we attempted to couple Fmoc-Dap(NH₂)-OH. Having the amino group on the side chain free would avoid the need of the Alloc deprotection step prior IFR (Figure 86).

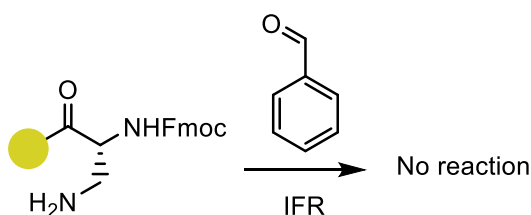


Figure 86: On-resin IFR protocol for heteropeptides. Unsuccessful approach.

However, this protocol did not afford the desired peptide. Perhaps, the steric hindrance caused by the bulky Fmoc group close the reactive amine would prevent the reaction from taking place.

A new approach was then considered. The next AA (Gly) was coupled on the chain before the IFR so the bulky Fmoc group is further away (Figure 87). For the functionalisation of the last Dap amino acid, a Fmoc deprotection step must be carried out before the IFR and twice the equivalents of reagents will be employed so positions 3 and 4 are functionalised at the same time. This reaction proved to be successful with phenylaldehyde **S** and indole derivative **T**.

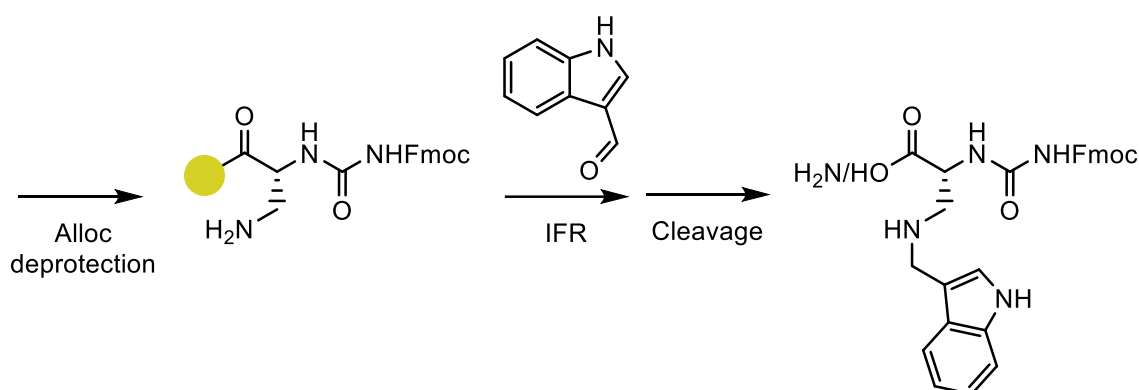


Figure 87: First steps for the On-resin IFR protocol for tetra-functionalised heteropeptides.

Method 3. The most complex scenario would be to synthesise peptide derivatives with 2 different functionalisations in positions 3 and 4 -the amino groups of the last N-terminal Dap- (4ABCD). We did not experimentally attempt this synthesis, but we will propose 2 viable synthetic routes. The first one would be the same as in method 2 except for the fact that the last amino acid Fmoc-Dap(NH₂)-OH would undergo IFR on the position 3 (side chain of last Dap amino acid) before removing the Fmoc in position 4. This proved to be unsuccessful in the past, probably due to the steric hindrance of the adjacent Fmoc group. However, longer

reaction times and/or temperature might be sufficient to overcome this impediment. Another approach would be to use the amino acid PG-Dap(NH₂)-OH, where PG is a different and smaller orthogonal protecting group: Boc, for instance. This could facilitate the functionalisation of position 3. Then, Boc PG would be removed (mild acidic conditions, 50% TFA in DCM) and IFR can be carried out on position 4 (Figure 88).

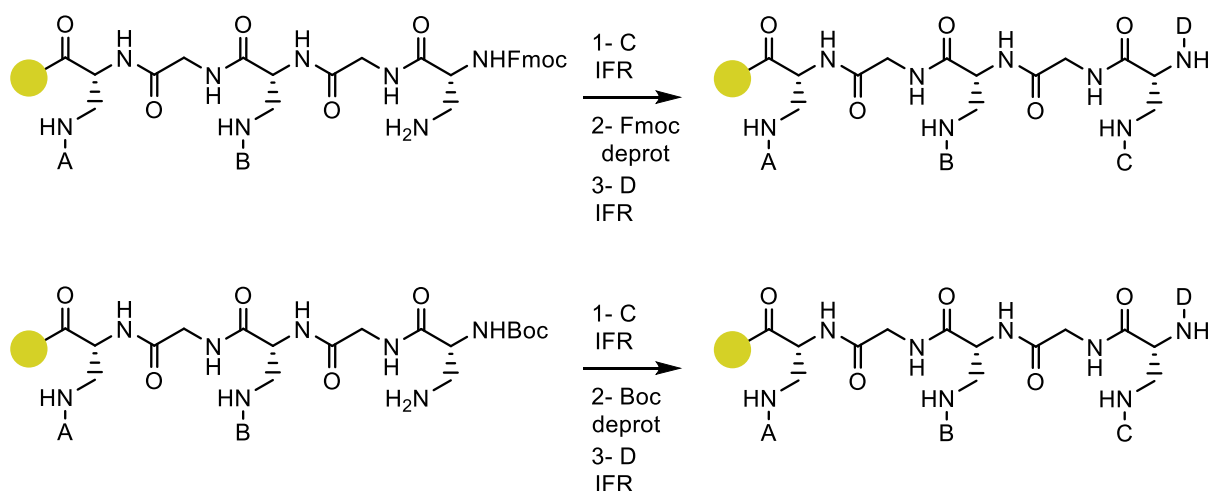


Figure 88: Two different proposed strategies to afford tetra-functionalised heteropeptides with different groups in positions 3 and 4.

Method 4. The last possible scenario would be to synthesise non-fully-functionalised peptides (4A/BB). For this, amino acid PG-Dap(NH₂)-OH (PG = Boc or Bzl) can be introduced whenever the position is to be left free and in a very last step they can be deprotected. Both Boc and Bzl are labile under strong acidic conditions (TFA and HF are normally employed for Boc and Bzl, respectively) so their deprotection can happen at the same time as the cleavage of the peptide from the resin (Figure 89).

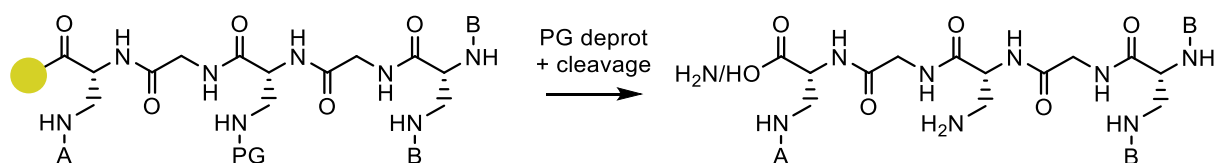
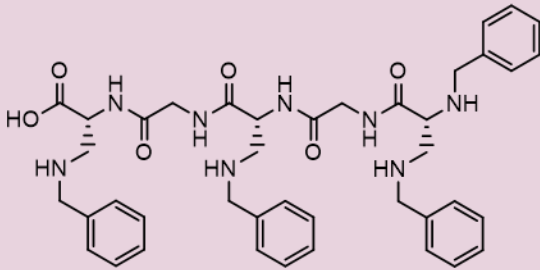
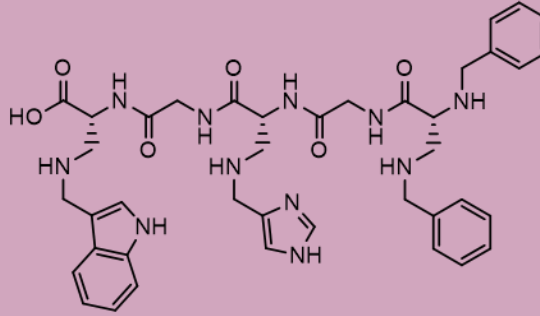
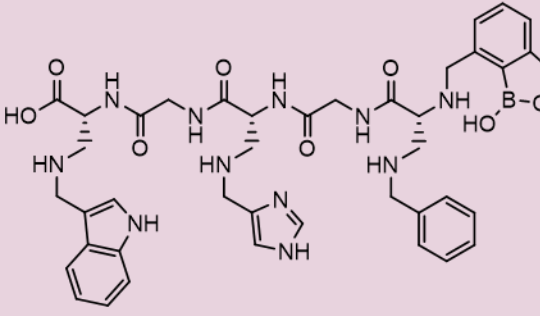
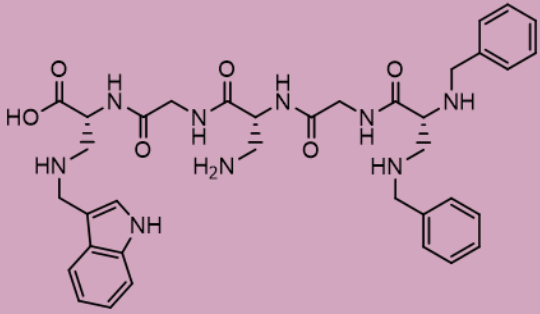


Figure 89: Proposed strategy to afford non-fully-functionalised peptides.

With these methodologies every library member arising from DCC 3.0 experiment could be synthesised (Table 10).

Table 10: Different methods to afford each type of peptide derivative in DCC 3.0.

Peptide type	Type code	example molecule	method
Tetra-substituted homopeptide	4AAAA		1
Tetra-substituted heteropeptide (3=4)	4ABCC		2
Tetra-substituted heteropeptide (3 ≠ 4)	4ABCD		3
Non-tetra-substituted peptide	4A/BB		4

6.4 DCC 3.0

The experimental protocol to carry out DCC 3.0 was similar to the one for DCC 2.0 with the difference that now the targets, as per the goals and objectives of Dr Ignacio Alfonso's group, were two polysaccharide glycosaminoglycans (GAGs): Heparin and Chondroitin Sulfate.

Peptide **4** (5 mM) and amines **Q**, **R**, **S**, and **T** (20 mM each) were dissolved in MeOH and left stirring O.N. Then, it was diluted 1/10 with a solution of 200 mM Phosphate buffer pH 7.5, containing 1.2 mg/mL of GAG template (Heparin or Chondroitin Sulfate), or no template for the blank experiment. NaCNBH₃ 5 mM in water was added quickly afterwards, and it was left stirring again ON before being analysed by LC-MS.

DCC 3.0 meant a considerable increment in library size from 26 members in DCC 2.0 to 625 members as per the formula n^r being ' n ' the number of items to choose from, **Q**, **R**, **S**, **T**, 0 (0 being free amine left unreacted) equals to five; and ' r ' the number of reactive positions, four.

This time, unfortunately, we were unable to confirm by LC-MS the presence of new products being formed in the DCL. The spectrum obtained corresponded merely to the sum of the starting BBs, and some tiny peaks that could correspond to low-substituted species (peptide with only one or two amino groups functionalised with **Q**, **R**, **S**, or **T**), or could just be noise. The peaks were comparable in size with the background noise and therefore we could not confirm their nature. In any case peaks with m/z that could be assigned to high-substituted products (peptide with 3 or 4 BBs attached) were found. We concluded that starting BBs are not reactive enough and highly substituted species may be too difficult to achieve in a DCC

experiment in aqueous environment, and therefore our next step was to attempt a simpler approach.

6.5 DCC 3.1 and 3.2

Back in Birmingham, we designed a much simpler peptide-based DCC 3.1 experiment. We used the same experimental conditions and recognition units **Q**, **R**, **S**, and **T** but with a shorter and easier to make peptide (Ac)Lys-Ala-Lys(CONH₂) (Lys-Ala-Lys, or **KAK** for short) as scaffold.

We wanted to explore as well the possibility of employing a longer spacer to reduce potential steric hindrance between the reactive positions, as a possible cause behind the low reactivity observed. To this aim, it was synthesised an analogue to **KAK**, with two consecutive units of Ala connecting the terminal Lys: **KAACK**. This peptide, tested in the same experimental conditions and with the same BBs as **KAK** (**Q**, **R**, **S**, and **T**) would constitute DCC 3.2 experiment (Figure 90).

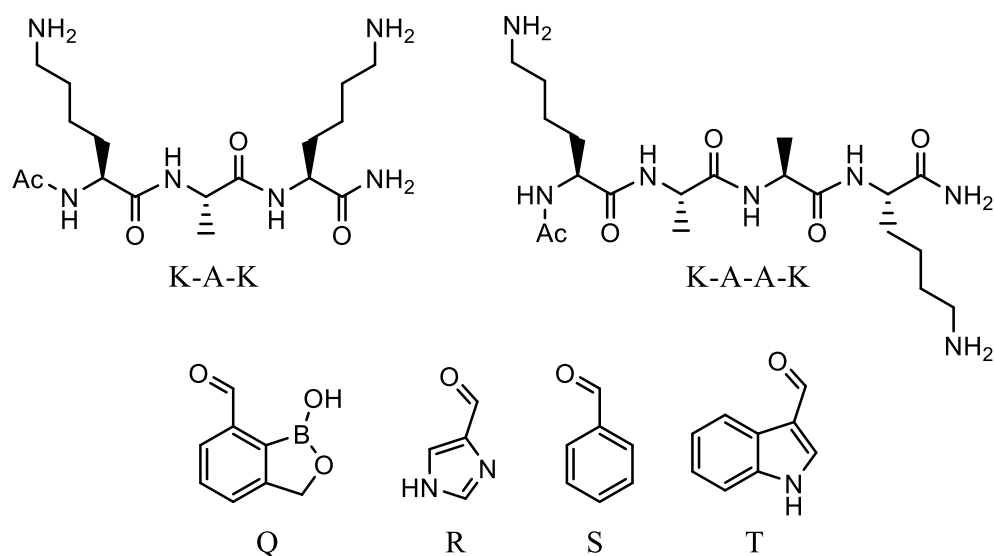


Figure 90: Starting BBs of DCC 3.1 and DCC 3.2 experiments.

A smaller peptide of three or four amino acids and with the N-terminal acetylated so it has only 2 reactive positions would afford a DCL of 25 members according to the n^r equation and would be perhaps be easier to fully functionalise.

Dap was swapped by Lys with the expectation that a longer length of the side chain (larger distance between the scaffold and the reactive amine) would benefit the IFR even though that could play against the binding properties of the potential receptors (less rigidity of the recognition unit).

Gly was swapped by Ala in an attempt to facilitate the synthesis of the peptide, as we found that the coupling reactions of Glycine onto the peptide were challenging and required a lot of excess of amino acid. Both the Dap or the Gly could be responsible for this low reactivity, but overall we found that when using Ala instead of Gly less equivalents of amino acid were required to accomplish a successful coupling reaction. Since Gly only functioned as a spacer, employing Ala instead would not affect the properties of the finished peptide while making its synthesis considerably easier.

These changes resulted in an LC-MS spectrum in which many new peaks could be observed in the UV chromatogram (i.e. peaks that do not correspond to the starting BBs). Nevertheless, the majority of the library members could not be identified in the MS chromatogram and therefore they could not be quantified. This is represented in Figure 91, where the bottom spectrum shows a fraction of the run, in which many new peaks are observed (i.e., peaks not corresponding to starting materials. Starting materials were eluted off the column in the first two minutes). However, none of the MS settings tried served to ionise the analytes and therefore the nature of such UV peaks could not be evaluated.

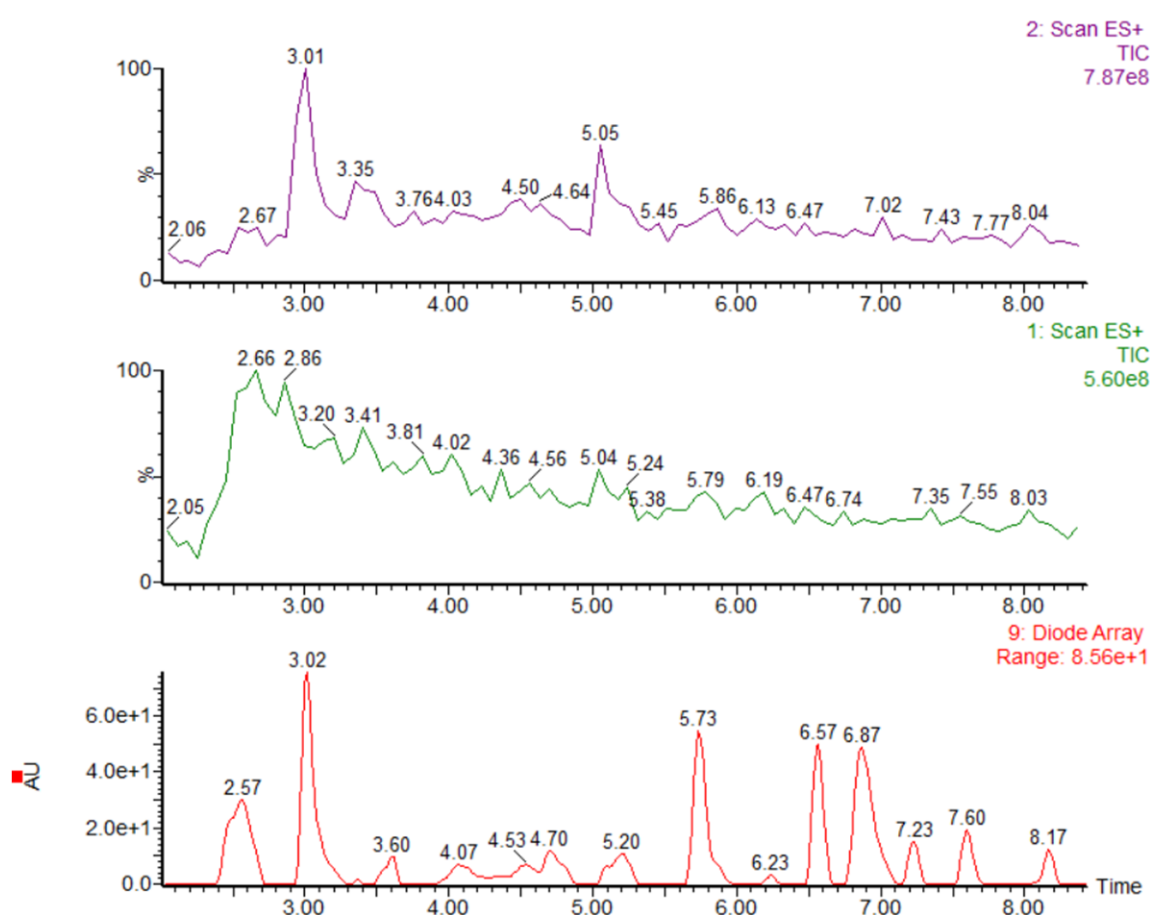


Figure 91: LC-MS spectra of DCC 3.1 experiment. UV response (bottom), and two different electrospray (ES) ionisation methods with positive ionisation (+) employed (middle and top).

The evidence suggests that there are new products being formed and therefore a DCL (as per the UV response), but the MS settings employed were just not fit for these sorts of molecules and could not ionise the analytes. We encountered a similar issue when optimising the analytical procedures for DCC 1.0 and DCC 2.0 experiments and we were able to solve it by testing different electrospray (ES) ionisation settings in positive mode (+) until finding the optimal ones.

This time, unfortunately, none of the ES modes in positive or negatives mode were able to ionise the analytes. Different ionisation methods (MALDI, APCI) should be tested in the future.

6.5.1 Synthetic protocol to afford KAK and KAAK by automated SPPS

In order to carry out DCC 3.1 and 3.2 experiments peptides **KAK** and **KAAK** were synthesised automatically using the peptide synthesiser CEM MultiPep1. This instrument allowed the complete synthesis in parallel of both products, which were carried out overnight. Only the last cleavage step was performed manually, under the same conditions and with the same apparatus detailed previously in this chapter for manual SPPS. Starting amino acid needed were Fmoc-D-Lys(Alloc)-OH and Fmoc-D-Ala-OH.

The protocol for the automated synthesis was the following:

- Piperidine 20% in DMF 1 x 3 min; 1 x 8 min
- DMF 4 x 1 min
- AA (0.5 M in DMF, 4 equiv.), Oxyma (1.0 M in DMF, 4.5 equiv.), DIC (1.0 M in DMF, 5 equiv.), 2 x 20 min
- Ac₂O (10 equiv.) + DIPEA (10 equiv.) in DCM, 1 x 10 min

- DCM 1 x 1 min
- DMF 5 x 1 min

These 6 steps were repeated 3 and 4 times for the syntheses of **KAK** and **KAAK**, respectively. Once the last amino acid was coupled, a last Fmoc deprotection and acetylation cycles were carried out in order to afford the desired N-terminal-acetylated peptides (Figure 92):

- Piperidine 20% in DMF 1 x 3 min; 1 x 8 min
- Ac₂O (10 equiv.) + DIPEA (10 equiv.) in DCM, 1 x 10 min

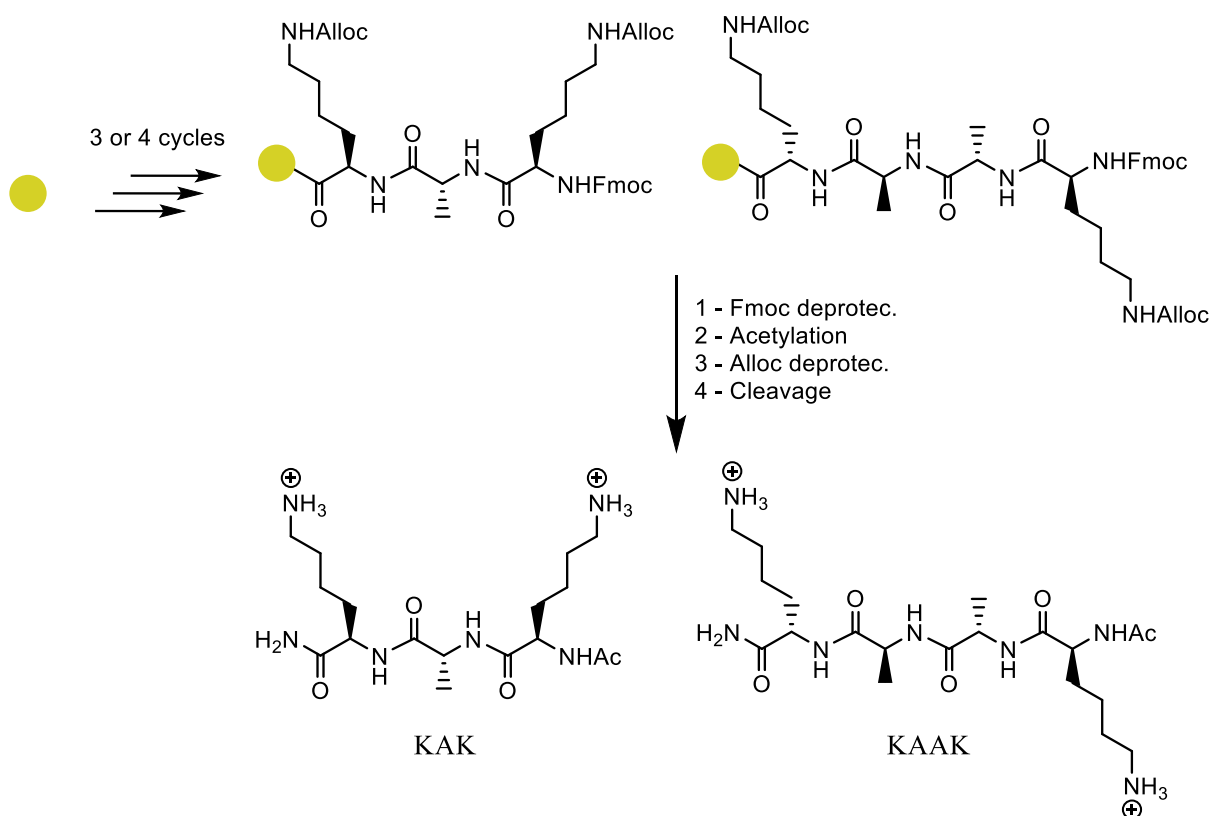


Figure 92: Diagram showing the synthetic route to afford peptides **KAK** and **KAAK** via SPPS.

And finally, in order to proceed with the cleavage of the peptide, the reaction cartridge was taken out of the peptide synthesiser and placed in the apparatus for manual SPPS where it

was treated with the same cleavage cocktail and with the same protocol employed in manual SPPS,

- TFA/TIPS/H₂O 1 x 30 min; 1 x 60 min; 1 x 1 min

6.6 Conclusions of Chapter 6

In this chapter we discussed the potential benefits of using peptides as scaffolds in DCC experiments. They are easy to make, they are customisable, they are flexible, and with a relative small number of amino acids, one can build a huge variety different peptides.

With DCC 3.0 experiment, even though we could not create a functional DCL, we laid the basis for peptide-based DCC. We learnt that the reactivity of the starting BBs is a key factor to consider and perhaps the bottleneck of the technique. In aqueous environments, when reactivity of the starting BBs is reduced (as in the case of the BBs needed for imine formation reaction), products that rely on the multiple functionalisation of a BB may not be formed.

With DCC 3.1 and 3.2 we overcame that issue by employing a shorter and simpler peptide as scaffold that contained, as in previous DCC 2.0 experiments, only two reactive positions. This time and according to the UV response of an LC-MS screening of the library, the creation of the DCL was successful. However, ionisation process must be sharpened to be able to analyse and quantify the analytes present in the library.

Peptide **4**, **4.1**, **KAK**, and **KAAC** were successfully synthesised and purified following SPPS procedures. The synthesis of the peptide derivatives that could arise from DCC 3.0 experiment was also investigated employing non-traditional on-resin IFR.

The methodology for the synthesis of two out of 4 types of products being formed in DCC 3.0 (4AAA and 4ABCC) was proven successful, and a methodology to afford the other 2 types (4ABCD and 4A/BB) was suggested.

4 and **4.1** and their derivatives were made manually in solid phase. Different polymeric supports and coupling methodologies were studied.

Peptide **4** was synthesised as well via manual SPPS but employing high temperature for the coupling reactions to try and enhance the efficiency of this reaction, hence improving the discrete yields obtained at room temperature. According to the colorimetric analysis carried out after each coupling reaction, it appeared like the coupling reactions indeed got better. Nevertheless, the overall yield did not improve significantly, probably due to partial premature cleavage induced by the higher temperatures.

Finally, peptides **KAK** and **KAAC** were automatedly synthesised with the help of a peptide synthesiser. Adding the resins onto the cartridges and cleaving the peptide in the end were the only steps that had to be carried out manually while everything else was done automated, in no longer than one night. Although the yields for these peptides were indeed better than those for peptides synthesised manually, the results cannot be compared as they were different target peptides. It is out of question however, the benefits that automated synthesis can offer in terms of simplicity and time consumption.

Chapter 7 - Experimental

This chapter will explain the experimental part of the thesis. Starting with the materials and methods employed, to then continue with the synthetic protocols to afford starting BBs, DCL members to be tested by ITC, and finally a detailed explanation on the SPPS methodology followed to synthesise the required peptides.

7.1 Materials & methods

If it is not stated otherwise, the chemicals were purchased from Sigma-Aldrich, Acros Organics, Merck, or Alfa Aesar and used without further purification. 1-Hydroxy-1,3-dihydro-2,1-benzoxaborole-7-carbaldehyde and 1-Hydroxy-1,3-dihydro-2,1-benzoxaborole-6-carbaldehyde were purchased from BLD Pharmatech Ltd. Fmoc-Asp(OAll)-OH was purchased from Fluorochem Ltd. Oxyma Pure was purchased from NovaBioChem. Fmoc-rink-amide-2CTC resin and Fmoc-L-Dap(Aloc)-OH were obtained from Iris Biotech GmbH.

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 permanganate or ninhydrin.

Silica gel column chromatography was carried out using Sigma-Aldrich 60 Å silica gel (35-70 µm).

Chemical structures were created in ChemDraw.

7.1.1 NMR

^1H NMR and ^{13}C NMR spectra were recorded at room temperature on the following spectrometers and analysed with MestreNova software: 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.

7.1.2 MS

Mass spectra were recorded with a Waters Xevo G2-XS spectrometer working with electrospray ionization (positive mode). LC-MS experiments were performed in a Waters SQD 2 (MS) instrument connected to a Waters Alliance e2695 (LC) and UV detector in the range 210 to 500 nm, working with electrospray ionization (positive mode). Different chromatography columns were employed, and these were explained in a dedicated chapter (LC-MS method development). LC-MS data was analysed with MassLynx software.

7.1.3 HPLC

HPLC was performed with an Agilent 1260 series quaternary pump, an automatic injector Agilent 1260 series, a dual variable wavelength detection Agilent 1260 series and a micro vacuum degasser device Agilent 1260. The column used was an AerisTM WIDEPORE (C4, 150 x 4.6 mm) with 3.6 μm particle size and 200 Å pore size.

7.1.4 ITC

ITC binding studies were conducted on a TA nano-LV instrument running in aqueous mode. The reference cell was filled with 300 μ L of deionised, filtered, and degassed water. The sample cell was filled with 300 μ L of solution of the receptors to be analysed and the syringe was filled with the solutions of the sugars of interest as titrants. The general procedure was 40 x 1.5 μ L injections for an ITC run. The heats of dilution for the sugar were obtained running sugar into buffer experiments and they were subtracted from the raw isotherm to provide the corrected heats, followed by their subsequent modelling. The model employed to analyse the data was the independent model. The corresponding isotherms were analysed utilising TA nano-analyse software. For this thesis the n value was fixed to 1, this has been previously reported in literature as viable method for studying weak interactions (i.e., $C < 10$).²⁹⁷

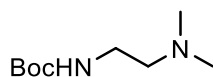
7.1.5 SPPS

Automated peptide synthesis was carried out with two different instruments as they were the available ones in the different labs where this research project took place. CEM Liberty Blue 2.0, that uses microwave radiation to improve the reaction effectiveness in coupling and deprotection steps was utilised to synthesise peptide **4**. Peptide was made at a scale of 1 mmol. Peptides **KAK** and **KAAC** were made with CEM MultiPep1 system at a scale of 1 mmol.

7.2 Synthesis of amine I for IFR optimisation experiments

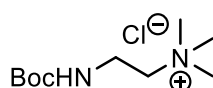
(2-Aminoethyl)trimethylammonium chloride, **I**, was synthesised from commercially available N,N-Dimethylethylenediamine. The synthesis was carried out in 3 steps:

1) Synthesis of N'-Boc-N,N-Dimethylethylenediamine.



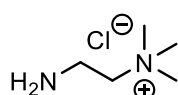
Triethylamine (0.75 mL, 5.4 mmol) and Boc_2O (1.28 g, 5.85 mmol) were dissolved in CHCl_3 (10 mL). Then, N,N-dimethylethylenediamine (0.5 mL, 4.95 mmol) was added and the reaction was stirred for 4 h. The solution was then washed with NaHCO_3 saturated solution and water. Organic phase was dried over MgSO_4 and solvent was removed under vacuum to afford N'-Boc-N,N-dimethylethylenediamine pure as a pale-yellow solid.

2) Synthesis of (N-Boc-Aminoethyl)trimethylammonium chloride.



To a solution of N'-Boc-N,N-dimethylethylenediamine (4.95 mmol) in MeOH, CH_3I (0.42 mL, 6.65 mmol) and KHCO_3 (0.56 mg, 5.65 mmol) were added. The reaction was stirred O.N. Solvent was removed under vacuum, crude redissolved in CHCl_3 , and the precipitate was filtrated out. The solution was dried under vacuum, and the solid washed with Et_2O , affording pure (N-Boc-Aminoethyl)trimethylammonium chloride as a white solid.

3) Synthesis of (2-Aminoethyl)trimethylammonium chloride.



(N-Boc-Aminoethyl)trimethylammonium chloride (0.62 g, 1.89 mmol) was dissolved in anhydrous MeOH (10 mL). It was cooled at 0°C and then acetyl chloride (2 mL, 28 mmol) was

added dropwise. After 3h, reaction was taken to dryness under vacuum, leaving a dark-red solid that was washed with Et₂O and MeOH, affording the product, **1**, as a pale yellow solid.

7.3 NMR Studies

7.3.1 Buffer preparation

The preparation of the deuterated phosphate buffer for NMR experiments to study the influence of pH in IFR was unconventional. As the chemicals required (KD₂PO₄, K₂DPO₄) are not commercially available in their deuterated forms, the analogues containing hydrogen instead (KH₂PO₄, K₂HPO₄) were employed. The right amount of them was weighed and dissolved in 10 mL of D₂O and then dried out under reduced pressure. This process was repeated 3 times to ensure that all the hydrogens present in the mixture are replaced by deuterium. Lastly, the solid is dissolved in the correct amount of D₂O to afford the deuterated buffered solution at the desired concentration (100 mM).

7.3.2 Preparation of solutions for NMR

For NMR experiments to study the influence of pH in NMR, aldehyde **2** and each one of the amines were dissolved in acetonitrile-d₃ and the buffer of choice, respectively. The solutions were combined to afford a final mixture with concentrations of aldehyde 1mM and amine 20 mM. The mixtures were left stirring O.N. and analysed by ¹H NMR.

7.4 Design of DCC experiment small-molecule approach

DCC 1.0

2 (1 mM) was mixed with **E**, **G**, and **H** (20/3 mM each) in 100 mM carbonate buffer pH 10 90%-10% MeCN. Then, the reaction mixture was split into four 1 mL Eppendorf tubes and to each one of them it was added a solution of glucose so its concentration in the reaction mixture would be 0.1 mM, 2 mM and 20 mM. To the fourth Eppendorf tube, the same volume of water was added instead so it can be used as a blank experiment. These mixtures were left stirring O.N. The next day, a solution of NaCNBH₃ (20 mM in the reaction mixture) was added to each one of the Eppendorf tubes and it was stirred for 30 min before analysing them by LC-MS.

Every one of the library members was found in the MS Chromatogram as their (M+H)⁺ adducts with m/z equal to their corresponding molecular weights plus one unit .

DCC 2.0

2 (1 mM) was mixed with **E**, **G**, **H**, **D**, and **P** (20/5 mM each) in 100 mM carbonate buffer pH 10 90%-10% MeCN. Then, the reaction mixture was split into five 1 mL Eppendorf tubes and to each one of them it was added a solution of glucose, mannose, galactose, and fructose 2 mM; as well as the same volume of distilled water for the fifth tube so it can be used as a blank experiment. These mixtures were left stirring O.N. The next day, a solution of NaCNBH₃ (20 mM in the reaction mixture) was added to each one of the Eppendorf tubes and it was stirred for 30 min before analysing them by LC-MS.

Every one of the library members was found in the MS Chromatogram as their (M+H)⁺ adducts with m/z equal to their corresponding molecular weights plus one unit .

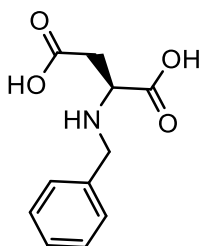
7.5 Synthesis of small-molecule receptors

The library members that gave the most interesting results in DCC 2.0 experiment were individually synthesised to evaluate their binding properties. The molecules to be tested were **D**, **P**, **1D**, **1P**, **2DD**, and **2PP**. Amino acids **D** and **P** were not synthesised by us as they are commercially available.

7.5.1 Synthesis of monosubstituted receptors **1D** and **1P**

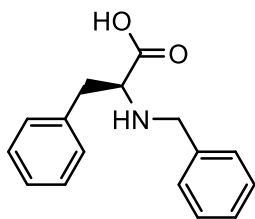
Mono-substituted molecules **2D** and **2P** were not synthesised exactly as they are present in the DCL. As it was already explained in a separate chapter, functionalising only one of the two aldehydes of starting isophthalaldehyde (**2**) would have been a rather complex procedure that would involve several reaction steps. Instead, we pursued a much simpler approach using benzaldehyde (**1**), the benzene ring posing only one aldehyde group, as starting material. This would afford the mono substituted products similar to **2D** and **2P**, without the extra CHO group in position 3 of the benzene ring. These molecules will be called **1D** and **1P**. The missing aldehyde group is far away from the ‘recognition positions’ of the receptor and therefore their absence would not affect the overall binding properties of the molecule. In other words, the affinity of **1D** and **1P** for the target saccharides would be the same as their analogues in the DCC experiment, **2D** and **2P**.

Synthesis of **1D**



L-Aspartic acid (**D**, 652.0 mg, 4.9 mmol) was slowly dissolved in a solution of NaOH (0.4 g) in water (5 mL) and methanol (10 mL). Then, Benzaldehyde (**1**, 508 μ L, 5.0 mmol) was slowly added to the solution. The reaction mixture was stirred for 1h. NaBH₄ (226.0 mg, 6 mmol) was dissolved in methanol (1 mL) and slowly added to the reaction mixture. Reaction left stirring for 1h before acidifying it to pH 5-6 with acetic acid. The solvent was evaporated under reduced pressure. The resulted oil was purified by HPLC to afford **1D** as a white solid (907.2 mg, 83%). HPLC purification method was optimised as follows: A 95%-5% B isocratic for 20 minutes, with a retention time for **1D** of 9-12 minutes. Being A: water + 0.1% TFA and B: acetonitrile + 0.1% TFA. ¹H NMR (400 MHz, D₂O, 298 K) δ 7.48 (s, 5H), 4.38 (d, $J = 13.1$ Hz, 1H), 4.32 (d, $J = 13.1$ Hz, 1H), 4.18 (t, $J = 5.6$ Hz, 1H), 3.08 (dd, $J = 17.6, 3.9$ Hz, 2H). ¹³C NMR (100 MHz, D₂O, 298 K) δ 170.79, 130.02, 129.28, 55.84, 50.63, 33.49. ¹H and ¹³C NMR is in agreement with literature.³³⁸ HRMS (ESI⁺): calculated for C₁₁H₁₄NO₄ [M+H]⁺: 224.0923, found: 224.0929 (Annex 6).

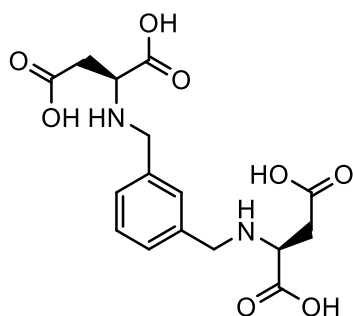
Synthesis of **1P**



L-Phenylalanine (**P**, 821.7 mg, 4.9 mmol) was slowly dissolved in a solution of NaOH (0.4 g) in water (5 mL) and methanol (10 mL). Then, Benzaldehyde (**1**, 508 μ L, 5.0 mmol) was slowly added to the solution. The reaction mixture was stirred for 3h. NaBH₄ (226.0 mg, 6 mmol) was dissolved in methanol (1 mL) and slowly added to the reaction mixture. Reaction left stirring for 1h before acidifying it to pH 5-6 with acetic acid. The solvent was evaporated under reduced pressure. The resulted oil was purified by HPLC to afford **1P** as a white solid (950.1 mg, 76%). HPLC purification method was optimised as follows: A 95%-5% B isocratic for 25 minutes, with a retention time for **1P** of 20 minutes. Being A: water + 0.1% TFA and B: acetonitrile + 0.1% TFA. ¹H NMR (400 MHz, DMSO-d₆, 298 K) δ 7.20 (m, 10H), 3.61 (d, $J = 13.1$ Hz, 1H), 3.40 (d, $J = 13.1$ Hz, 1H), 2.92 (m, 2H), 2.63 (m, 1H). ¹H NMR is in agreement with literature.³³⁹ HRMS (ESI+): calculated for C₁₆H₁₈NO₂ [M+H]⁺: 256.1338, found: 256.1342 (Annex 7).

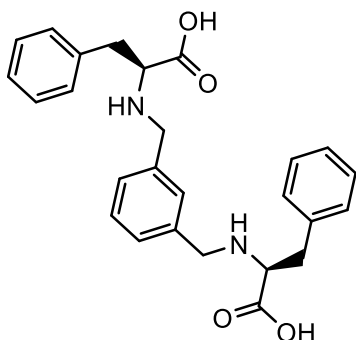
7.5.2 Synthesis of disubstituted receptors 2DD and 2PP

Synthesis of **2DD**



L-Aspartic acid (**D**, 665.3 mg, 5.0 mmol) was slowly dissolved in a solution of NaOH (0.4 g) in water (5 mL) and methanol (10 mL). Then, isophthalaldehyde (**2**, 335.0 mg, 2.5 mmol) was dissolved in methanol (10 mL) and slowly added to the amino acid solution. The reaction mixture was stirred for 6h. NaBH₄ (226.0 mg, 6 mmol) was dissolved in methanol (1 mL) and slowly added to the reaction mixture. Reaction left stirring for 1h before acidifying it to pH 5-6 with acetic acid. The solvent was evaporated under reduced pressure. The resulted oil was purified by HPLC to afford **2DD** as a white solid (552.1 mg, 60%). HPLC purification method was optimised as follows: A 95%-5% B isocratic for 10 minutes, with a retention time for **2DD** of 5.1 minutes. Being A: water + 0.1% TFA and B: acetonitrile + 0.1% TFA. ¹H NMR (400 MHz, D₂O, 298 K) δ 7.36 (t, J = 7.64 Hz, 1H), 7.28 (s, 1H), 7.27(d, J = 7.64 Hz, 2H), 3.72 (d, J = 13.1 Hz, 2H), 3.61 (d, J = 13.1 Hz, 2H), 3.43 (dd, J = 8.13, 5.63 Hz, 2H) 2.42 (m, 4H). HRMS (ESI⁺): calculated for C₁₆H₁₉N₂O₈ [M-H]⁺: 367.1141, found: 367.1139 (Annex 8).

Synthesis of **2PP**



L-Phenylalanine (**P**, 825.9 mg, 5.0 mmol) was slowly dissolved in a solution of NaOH (0.2 g) in water (5 mL) and methanol (10 mL). Then, isophthalaldehyde (**2**, 335.0 mg, 2.5 mmol) was dissolved in methanol (10 mL) and slowly added to the amino acid solution. The reaction mixture was stirred for 6h. NaBH₄ (226.0 mg, 6 mmol) was dissolved in methanol (1 mL) and slowly added to the reaction mixture. Reaction left stirring for 1h before acidifying it to pH 5-6 with acetic acid. The solvent was evaporated under reduced pressure. The resulted oil was purified by HPLC to afford **2PP** as a white solid (713.1 mg, 66%). HPLC purification method was optimised as follows: A 90%-10% B isocratic for 15 minutes, with a retention time for **2PP** of 10.1 minutes. Being A: water + 0.1% TFA and B: acetonitrile + 0.1% TFA. ¹H NMR (400 MHz, D₂O, 298 K) δ 7.37-7.19 (m, 13H), 7.10 (s, 1H), 3.73 (d, *J* = 13.1 Hz, 2H), 3.55 (d, *J* = 13.1 Hz, 2H), 3.36 (s, 2H N-H) 3.35 (t, *J* = 6.98 Hz, 2H) 2.90 (m, 4H). ¹³C NMR (100 MHz, D₂O, 298 K) δ 181.17, 139.09, 138.09, 129.27, 128.71, 128.53, 128.44, 127.42, 126.60, 117.78, 114.89, 64.48, 50.87, 39.04. HRMS (ESI⁺): calculated for C₂₆H₂₇N₂O₄ [M-H]⁺: 431.1971, found: 431.1983 (Annex 9).

7.6 Solid-phase peptide synthesis (SPPS)

7.6.1 Manual SPPS

As detailed in the previous chapter dedicated to the DCL peptide approach, the steps to follow for the synthesis of peptides **4** and **4.1** by manual SPPS are the following

7.6.1.1 *Conditioning of the resin and incorporation of the first amino acid*

Rink amide resin

- DCM 1 x 60 min ; 1 x 1 min
- DMF 2 x 1 min
- Piperidine 20% in DMF 2 x 4 min
- DMF 2 x 1 min
- AA, Oxyma, DMF, DIC 1 x 60 min
- DMF 2 x 1 min

2-CTC resin

- DCM 1 x 30 min ; 1 x 1 min
- DMF 2 x 1 min
- DCM 2 x 1 min
- AA, DCM, DIPEA 1 x 90 min
- MeOH
- DCM 3 x 1 min
- MeOH 1 x 1 min

7.6.1.2 Loading determination by Fmoc quantification

The loading determination of 2-CTC resin for the synthesis of **4.1** was done following the protocol detailed in the previous chapter and employing Equation 7.1 and Equation 7.2, the average corrected loading factor of the 2 values obtained was 0.32 mmol/g of resin.

7.6.1.3 Capping of unreacted positions

- A) MeOH (0.8 mL/g resin) + DIPEA (3 equiv.) in DCM, 1 x 30 min
- B) Ac₂O (10 equiv.) + DIPEA (10 equiv.) in DCM, 1 x 30 min

7.6.1.4 Fmoc removal

- Piperidine 20% in DMF 2 x 5 min
- DMF 2 x 1 min

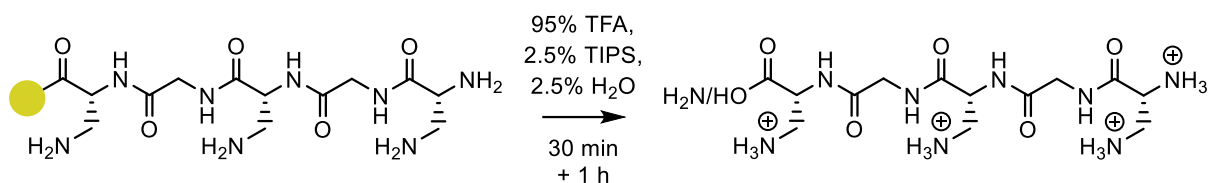
7.6.1.5 On-resin elongation of the peptide

- AA, (K) Oxyma, DIC, DMF 2 x 1 h // AA, PyBOP, DIPEA, DMF 2 x 1 h
- DMF 2 x 1 min
- Kaiser test

7.6.1.6 Alloc removal

- Pd(PPh₃)₄, PhSiH₃, DCM 3 x 10 min
- DCM 2 x 1 min
- DMF 2 x 1 min
- Piperidine 20% in DMF 2 x 4 min
- DMF 2 x 1 min

7.6.1.7 Peptide cleavage



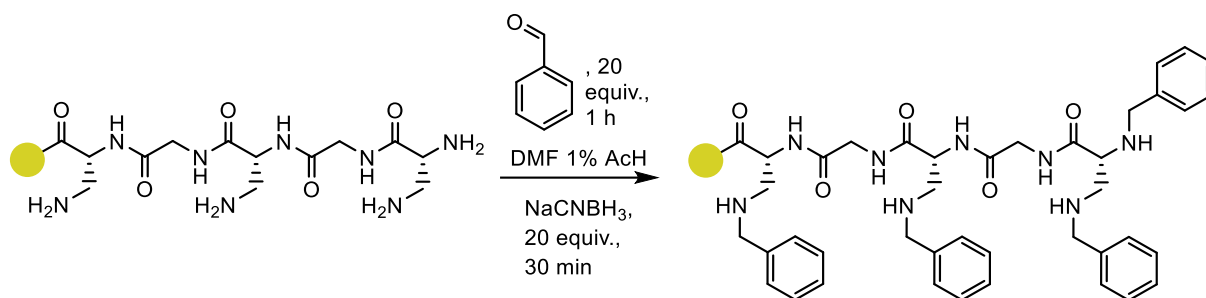
- TFA/TIPS/H₂O 1 x 30 min ; 1 x 60 min ; 1 x 1 min

The filtrates were collected in a round-bottomed flask containing cold Et₂O and the product peptides precipitated out. The peptides were vacuum-filtrated and dried. The peptides were purified by HPLC. HPLC purification method was optimised as follows: 5-100% B over 20 minutes, with a retention time for both **4** and **4.1** of 4.80 min and 4.66 min, respectively. Being A: water + 0.1% TFA and B: acetonitrile + 0.1% TFA. HRMS (ESI⁺): calculated for **4**: C₁₃H₂₇N₉O₅Na [M+Na]⁺: 412.2023, found: 412.2025 (Annex 10).

After lyophilisation, it was obtained 32.1 mg of **4** (63% yield, according to the 260 mg of resin employed and its loading factor of 0.5 mmol/g of resin) and 12 mg of **4.1** (32% yield, according to the 300 mg of resin employed and its corrected loading factor of 0.32 mmol/g of resin).

7.6.1.8 On-resin reductive amination

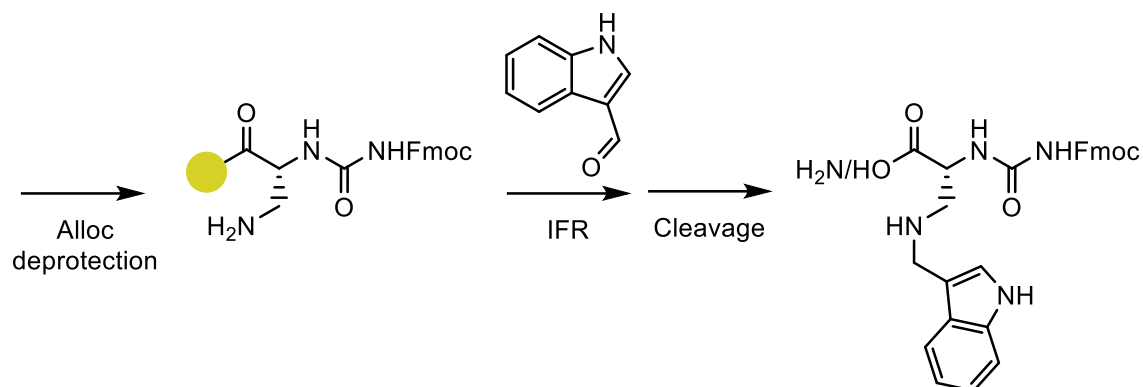
Method 1.



- AA, DMF+1%AcOH 1 x 1 h // AA, PyBOP, DIPEA, DMF 2 x 1 h
- NaCNBH₃, 1 x 30 min

The reaction gave negative result in a ninhydrin test so the peptide was cleaved from the resin and the formation of the tetra-substituted product **4SSSS** was confirmed by LC-MS. LC-MS running conditions were as follows: 5% B to 100% B over 5 minutes, with a retention time for **4SSS** of 3.47 minutes. Being A: water + 0.1% TFA and B: acetonitrile + 0.1% TFA. [M+H]⁺: 805 (Annex 11).

Method 2

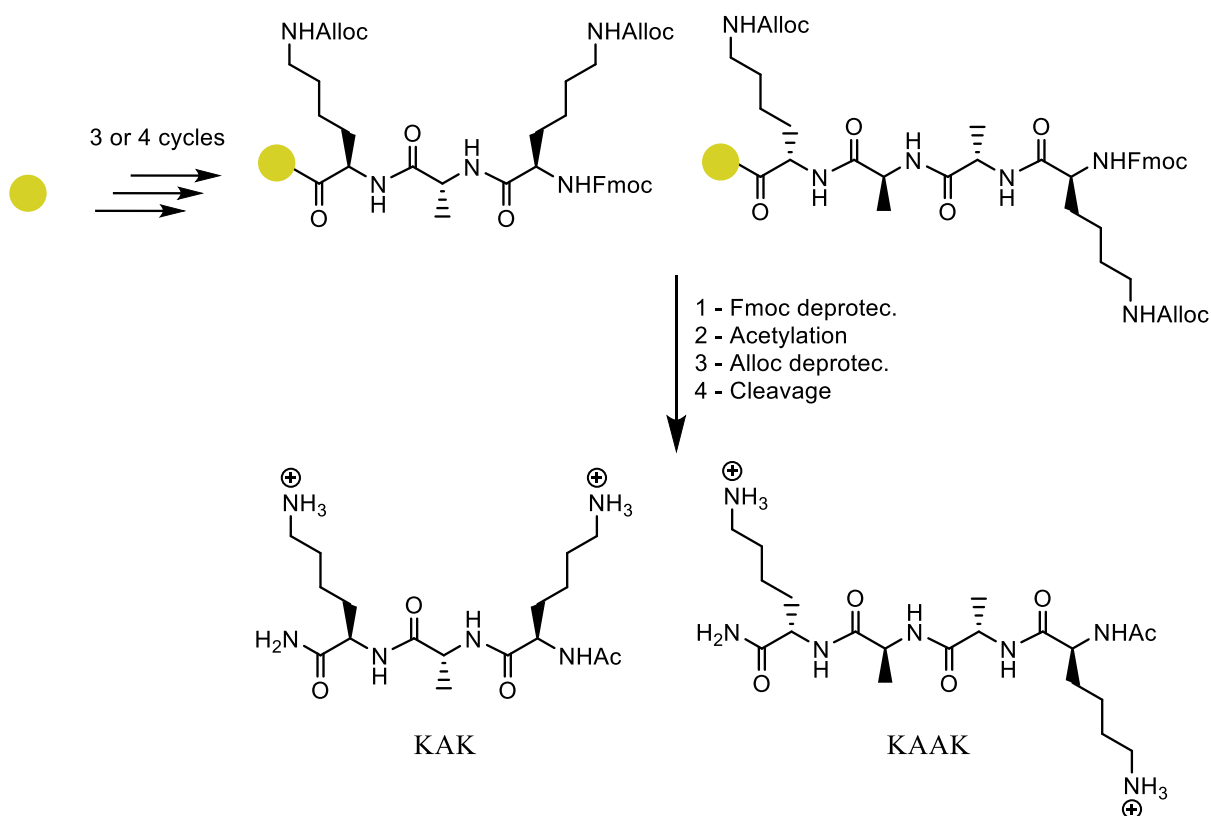


- AA, DMF+1%AcOH 1 x 1 h // AA, PyBOP, DIPEA, DMF 2 x 1 h
- NaCNBH_3 , 1 x 30 min

The reaction gave negative result in a ninhydrin test so the peptide was cleaved from the resin and the formation of the mono-substituted product **4T** was confirmed by LC-MS. LC-MS running conditions were as follows: 5% B to 100% B over 5 minutes, with a retention time for **4T** of 0.37 minutes. Being A: water + 0.1% TFA and B: acetonitrile + 0.1% TFA. $[\text{M}+\text{H}]^+$: 161 (Annex 12).

7.6.2 Automated SPPS

As detailed in the previous chapter dedicated to the DCL peptide approach, the steps to follow for the synthesis of peptides **KAK** and **KAAK** by automated SPPS are the following:



- Piperidine 20% in DMF 1 x 3 min; 1 x 8 min
- DMF 4 x 1 min
- AA (0.5 M in DMF, 4 equiv.), Oxyma (1.0 M in DMF, 4.5 equiv.), DIC (1.0 M in DMF, 5 equiv.), 2 x 20 min
- Ac₂O (10 equiv.) + DIPEA (10 equiv.) in DCM, 1 x 10 min
- DCM 1 x 1 min
- DMF 5 x 1 min

These 6 steps above were repeated 3 and 4 times for the syntheses of **KAK** and **KAAC**, respectively. Then:

- Piperidine 20% in DMF 1 x 3 min; 1 x 8 min
- Ac₂O (10 equiv.) + DIPEA (10 equiv.) in DCM, 1 x 10 min
- TFA/TIPS/H₂O 1 x 30 min; 1 x 60 min; 1 x 1 min

HPLC purification method was optimised as follows: 5-100% B over 10 minutes, with a retention time for both **KAK** and **KAAC** of 2.9 min and 3.2 min, respectively. Being A: water + 0.1% TFA and B: acetonitrile + 0.1% TFA. HRMS (ESI⁺): calculated for **KAK**: C₁₇H₃₄N₆O₄Na [M+Na]⁺: 409.2482, found: 409.2541 (Annex 13). HRMS (ESI⁺): calculated for **KAAC**: C₂₀H₃₈N₆O₆Na [M+Na]⁺: 480.2991, found: 480.2923 (Annex 14)

After lyophilisation, it was obtained 11.2 mg of **KAK** (72% yield, according to the 200 mg of resin employed and its loading factor of 0.2 mmol/g of resin) and 12.6 mg of **KAAC** (69% yield, according to the 200 mg of resin employed and its corrected loading factor of 0.2 mmol/g of resin).

Chapter 8 -Conclusions & Future work

Glycans are integral to a variety of biological processes including metabolism, immune system, cell adhesion, and cell-to-cell communication. Moreover, they are implicated in the development and progression of diseases, particularly neurodegenerative disorders and cancer. As certain pathological conditions can alter the normal expression of glycans in cells, glycans have become critical biomarkers in modern medicine for early disease diagnosis and risk stratification. However, detecting and quantifying these glycans is challenging due to their vast variety and diversity. The effective recognition of saccharide chains contained within these glycans is often the key to distinguishing between "normal" glycans present in healthy cells and "malicious" ones associated with disease (Chapter 1.1).

Hence, a valid strategy to follow in order to diagnose these diseases is the development of systems to tackle the detection of saccharides. However, molecular recognition of saccharides poses as well several challenges of their own: their simplicity (or in other words, their lack of functional groups susceptible of being bound by a receptor), their resemblance to one another (they have many chiral carbons and therefore the number of stereoisomers is large), and the fact that monosaccharides in solution are present in a number of interconvertible forms (tautomers), affording a number of regioisomers (Chapter 1.2).

35 years of research in the field of saccharide recognition yielded a huge number of receptors for saccharides reported. Often, they were based on natural compounds or directly extracted from natural sources, and in some other cases they were synthesised in the lab based on rational design approaches. Synthetic receptors have significantly improved over time, with

binding constants enhancing several orders of magnitude and overcoming major obstacles such as saccharide recognition in aqueous environments. However, there is still a need for a better understanding of the underlying mechanisms behind molecular recognition in order to be able to design more sophisticated saccharide receptors with even better binding properties. Chapters 1.3 to 1.5 give a detailed summary on the work carried out to date in the use of saccharide receptors, both extracted from natural sources and developed via rational design.

Several factors are thought to be crucial for the rational design of effective receptors, including multivalency and the right choice of intermolecular forces necessary for binding target ligands. H bonds, π -H interactions, and electrostatic forces are responsible for molecular recognition in nature. In the particular case of recognition of saccharides, boronic acids have been widely studied due to their ability to covalently bind sugars via boronate ester formation. Combining covalent and non-covalent interactions may hold the key to the creation of an outstanding new class of receptors.

Rather than attempting to create one potentially effective receptor for a particular saccharide, we propose a more ambitious approach involving rational design for the selection of a reduced number of compounds or building blocks in combination with a supramolecular technique called Dynamic Combinatorial Chemistry (DCC) that allows the self-assembly of a large library of potentially good receptors and the self-selection of the best receptor available. In Chapter 1.6 this technique is explained in detail.

An exhaustive method development is detailed in Chapter 3 - and Chapter 4 -to optimise both the Imine Formation Reaction (IFR), reaction in equilibrium responsible for the Dynamic Combinatorial Chemistry (DCC) technique, as well as to optimise the DCC methodology itself.

In DCC 1.0, isophthalaldehyde (**2**), which bears two reactive positions, is used as a scaffold with three chemically diverse molecules (**E**, **G**, and **H**). This system provided a library of 13 compounds that possessed multivalency and the tools to create covalent (through the BA group in **E**) and non-covalent interactions (H bonds and π -H) with saccharides. Molecule **2EE**, which possesses two boronic acid groups, emerged as the best receptor for glucose. We then carried out DCC 2.0, employing the same aldehydic scaffold to ensure multivalency, the same three building blocks (BBs) as in DCC 1.0, plus two new molecules, amino acids **D** and **P**. This doubled the library size to 26 members and enriched the system with more molecules capable of creating strong hydrogen bonds, π -H interactions, and electrostatic forces.

We tested this dynamic combinatorial library (DCL) against four different isomeric monosaccharides (glucose, mannose, galactose, and fructose). The first three sugars gave similar results with **2DD** standing out as the best receptor. The experiment with fructose yielded markedly different results, with **1D** and especially **1P** outperforming the other library members in a more competitive environment. This differential behaviour could be attributed to the unequal distribution of tautomers for those sugars, with fructose exhibiting a much larger fraction of furanose forms in solution than the others. It was interesting to see how molecules containing BA groups and therefore able to covalently bind the saccharides, **2E** and **2EE**, were outperformed by molecules capable of interacting with the templates merely by non-covalent forces.

The library members that gave the best results, as well as some other as negative controls, were synthesised and tested by ITC to obtain their binding data. The synthetic protocols are detailed in Chapter 7 - and the ITC studies can be found in Chapter 5 -. The tested molecules were **D**, **P**, **1D**, **1P**, **2DD**, and **2PP**. ITC validated DCC results, proving that the DCC methodology was successful in predicting the best receptors within the library of compounds. Molecule **1P**, with K_a for fructose = 1762 M^{-1} , is in the range of the best yet discovered small-molecule receptors for fructose that can operate in water media. The interaction between **1P** and fructose was also studied by NMR titration. While the poor solubility of **1P** limited the scope of the technique to obtain quantitative results, it was useful to corroborate the binding and obtain qualitative data regarding the regions of **1P** involved the most in the binding. The hydrogens directly adjacent to the secondary amine in **1P** were significantly more affected ($\Delta\delta = 0.02 \text{ ppm}$) than the benzylic protons in β to the carboxylic acid ($\Delta\delta = 0.008 \text{ ppm}$) and the aromatic ones ($\Delta\delta = 0.056 \text{ ppm}$). This effect was only visible in titration 3 and titration 5 experiments, where there is significant amount of complex **1P**-fructose.

In an attempt to exploit even more the benefits of multivalency, we designed DCC experiments employing peptides as scaffolds, in combination with BBs that would bring similar covalent and non-covalent intermolecular forces as the ones used in previous DCC experiments, to achieve molecular recognition. The synthetic protocols to afford such peptides are included as well in Chapter 7 -, and the attempts to employ such peptides in DCC experiments can be found in Chapter 6 -. DCC 3.0 experiment included pentapeptide, **4**, with four reactive positions, combined with four building blocks (**Q**, **R**, **S**, **T**). This DCL could potentially form up to 256 library members.

Unfortunately, highly substituted molecules (tri and tetra-functionalised peptides) were not detected by LC-MS. Even the mono and di-substituted ones were found in a very low amount. Evidence suggests that in the aqueous reaction conditions, starting BBs are not reactive enough to form multivalent products. To overcome this limitation, DCC 3.1 and 3.2 experiments were then designed as a simpler version of their predecessor, 3.0. Tri and tetrapeptides **KAK** and **KAAC** pose only 2 reactive positions so in combination with BBs **Q**, **R**, **S**, and **T** would afford a DCL of only 25 members. When screening these DCLs by LC-MS, many new peaks were found in the UV spectrum. This indicates that we indeed solved the reactivity issue, and there are new products being formed. However, we failed to find ionisation settings in the MS apparatus of the LC-MS system that could ionise the analytes. Hence, we could not identify and quantify the new products. Ionisation methods different to the ones tested herein (electrospray, ES) should be tested: MALDI, APCI.

We believe in the potential of DCC with multivalent peptidic scaffolds as a tool to create powerful receptors for saccharides. We found that the bottleneck of DCC technology employing IFR as the core of the equilibrium, is the limited reactivity of the starting BBs in aqueous media, and therefore we propose two different strategies to potentially overcome this limitation:

First, peptides **4** and **4.1** should be redesigned. We hypothesise that a plausible reason for the low reactivity exhibited by these peptides towards IFR could be the proximity between the reactive amines in Dap, and the peptidic scaffold, which could cause steric hindrance. Therefore, different natural and unnatural amino acids should be explored to replace Dap as

the unit responsible for IFR. L-2,4-diaminobutyric acid (Dab), L-ornithine, (Orn), and L-lysine (Lys) could for instance be tested. Their longer side chains will keep separated further away the reactive primary amine from the peptidic scaffold potentially improving their reactivity. Another source of steric hindrance that could be responsible for the low conversion to highly substituted products in DCC 3.0 experiment could be the proximity between the reactive units. In this context, new peptides should be designed with longer spacers (2, 3, or 4 units of Gly) between the Dap (or Dab, or Orn, or Lys) reactive units.

Secondly, as the aqueous environment seems to be a major obstacle for the formation of highly substituted library members in DCC experiments, a methodology should be tested that creates a library of multivalent peptides in organic anhydrous solvent, where IFR will be favoured. Then, after the equilibrium reaction is quenched with a reducing agent so there is no possibility of hydrolysis of the imines, these peptides can be exposed to the biological targets of interest (glycans) attached to solid supports and in aqueous buffers, so the target remains in its natural form. -A selection process could then be used to capture the most selective peptide.

List of references

- (1) Tommasone, S.; Allabush, F.; Tagger, Y. K.; Norman, J.; Köpf, 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.
- (2) Ohtsubo, K.; Marth, J. D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **2006**, *126* (5), 855–867.
- (3) Varki, A. Biological Roles of Glycans. *Glycobiology* **2017**, *27* (1), 3–49.
- (4) de Haas, P.; Hendriks, W. J. A. J.; Lefeber, D. J.; Cambi, A. Biological and Technical Challenges in Unraveling the Role of N-Glycans in Immune Receptor Regulation. *Front. Chem.* **2020**, *8* (February).
- (5) Perdivara I; Deterding LJ; Cozma C; Tomer KB; Przybylski M. Glycosylation Profiles of Epitope-Specific Anti- β -Amyloid Antibodies Revealed by Liquid Chromatography–Mass Spectrometry. *Glycobiology* **2009**, *19* (9), 958–970.
- (6) Haukedal, H.; Freude, K. K. Implications of Glycosylation in Alzheimer’s Disease. *Front. Neurosci.* **2021**, *14* (January), 1–18.
- (7) Campbell BJ; Yu LG; Rhodes JM. Altered Glycosylation in Inflammatory Bowel Disease: A Possible Role in Cancer Development. *Glycoconj J.* **2001**, *18*, 851–858.
- (8) 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.
- (9) 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.
- (10) 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.
- (11) 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.
- (12) 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.
- (13) Shiyue Zhou, Yunli Hu, Janie L. DeSantos-Garcia, and Y. M. Quantitation of Permethylated N-Glycans through Multiple Reaction Monitoring (MRM) LC-MS/MS. *J Am Soc Mass Spectrom* **2015**, 26 (4), 596–603.
- (14) Hart, G. Overview of Glycobiology. In *Encyclopedia of Cell Biology (Second Edition)*; 2023; pp 297–308.
- (15) Ho, W.; Hsu, W.; Huang, M.; Kadomatsu, K.; Nakagawara, A. Protein Glycosylation in Cancers and Its Potential Therapeutic Applications in Neuroblastoma. *J. Hematol. Oncol.* **2016**, 1–15.
- (16) 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. Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **2015**, *25* (12), 1323–1324.
- (17) Stowell, S. R.; Ju, T.; Cummings, R. D. Protein Glycosylation in Cancer. *Annu Rev Pathol.* **2015**, *10*, 473–510.
- (18) Cummings, R. D.; Pierce, J. M. The Challenge and Promise of Glycomics. *Chem. Biol.* **2014**, *21* (1), 1–15.
- (19) Dube, D.; Bertozzi, C. Glycans in Cancer and Inflammation — Potential for Therapeutics and Diagnostics. *Nat Rev Drug Discov* **2005**, *4*, 477–488.
- (20) Joachim, H.; Jürgen, G. An Introduction to the Sugar Code. *Histochem. Cell Biol.* **2017**, *147* (2), 111–117.
- (21) Stanley, P.; Taniguchi, N.; Aebi, M. N-Glycans. In *Essentials of Glycobiology (3rd Ed)*; 2017.
- (22) Trombetta, E. S. The Contribution of N-Glycans and Their Processing in the Endoplasmic Reticulum to Glycoprotein Biosynthesis. **2003**, *13* (9), 77–91.
- (23) Fuster, M. M.; Esko, J. D. THE SWEET AND SOUR OF CANCER : GLYCANS AS NOVEL THERAPEUTIC TARGETS. **2005**, *5* (July), 526–542.
- (24) Dhanisha, S. S.; Guruvayoorappan, C.; Abeesh, P. Mucins : Structural Diversity , Biosynthesis , Its Role in Pathogenesis and as Possible Therapeutic Targets. *Crit. Rev. Oncol. / Hematol.* **2018**, *122* (October 2017), 98–122.
- (25) Yu, R. K.; Tsai, Y.; Ariga, T.; Yanagisawa, M. Structures , Biosynthesis , and Functions of Gangliosides-an Overview. *J. Oleo Sci.* **2011**, *544* (10), 537–544.
- (26) Rabu, C.; McIntosh, R.; Jurasova, Z.; Durrant, L. Glycans as Targets for Therapeutic

Antitumor Antibodies. *Futur. Oncol.* **2012**, *8* (8), 943–960.

- (27) Hsu, K.; Mahal, L. A Lectin Microarray Approach for the Rapid Analysis of Bacterial Glycans. *Nat Protoc* **2006**, *1*, 543–549.
- (28) Stevens, J.; Blixt, O.; Glaser, L.; Taubenberger, J. K.; Palese, P.; Paulson, J. C.; Wilson, I. A. Glycan Microarray Analysis of the Hemagglutinins from Modern and Pandemic Influenza Viruses Reveals Different Receptor Specificities. *J. Mol. Biol.* **2006**, *355*, 1143–1155.
- (29) Wang, J.; Gao, W.; Grimm, R.; Jiang, S.; Liang, Y.; Ye, H.; Li, Z.; Yau, L.; Huang, H.; Liu, J.; Jiang, M.; Meng, Q.; Tong, T.; Huang, H.; Lee, S.; Zeng, X.; Liu, L.; Jiang, Z. A Method to Identify Trace Sulfated IgG N-Glycans as Biomarkers for Rheumatoid Arthritis. *Nat. Commun.* **2017**, *8*, 631.
- (30) Klasi, M.; Markulin, D.; Vojta, A.; Samar, I.; Biru, I.; Dobrini, P.; Ventham, N. T.; Trbojevi, I.; Mirna, Š.; Jerko, Š.; Razdorov, G.; Kennedy, N. A.; Satsangi, J.; Dias, A. M.; Pinho, S.; Annese, V.; Latiano, A.; Inca, R. D.; Lauc, G.; Zoldo, V. Promoter Methylation of the MGAT3 and BACH2 Genes Correlates with the Composition of the Immunoglobulin G Glycome in Inflammatory Bowel Disease. **2018**, 1–14.
- (31) Keir, M. E.; Yi, T.; Lu, T. T.; Ghilardi, N. The Role of IL-22 in Intestinal Health and Disease. **2020**, *217* (3), 1–9.
- (32) Meany, D. L.; Zhang, Z.; Sokoll, L. J.; Zhang, H.; Chan, D. W. Glycoproteomics for Prostate Cancer Detection: Changes in Serum PSA Glycosylation Patterns. *J. Proteome Res.* **2009**, *8* (2), 613–619.
- (33) Meany, D. L.; Chan, D. W. Aberrant Glycosylation Associated with Enzymes as Cancer

Biomarkers. *Clin Proteom* **2011**, *8* (7).

- (34) Mereiter, S.; Campos, D.; Gomes, J.; Reis, C. A. Glycosylation in the Era of Cancer-Targeted Therapy : Where Are We Heading ? *Cancer Cell* **2019**, *36*.
- (35) Tabarés, G.; Radcliffe, C. M.; Barrabés, S.; Ramírez, M.; Aleixandre, N.; Hoesel, W.; Dwek, R. A.; Rudd, P. M.; Peracaula, R.; de Llorens, R. Different Glycan Structures in Prostate-Specific Antigen from Prostate Cancer Sera in Relation to Seminal Plasma PSA. *Glycobiology* **2006**, *16* (2), 132–145.
- (36) Ferrer-Batallé, M.; Llop, E.; Ramírez, M.; Aleixandre, R. N.; Saez, M.; Comet, J.; de Llorens, R.; Peracaula, R. Comparative Study of Blood-Based Biomarkers, A2,3-Sialic Acid PSA and PHI, for High-Risk Prostate Cancer Detection. *Int. J. Mol. Sci.* **2017**, *18* (4), 1–12.
- (37) Ishikawa, T.; Yoneyama, T.; Tobisawa, Y.; Hatakeyama, S.; Kurosawa, T.; Nakamura, K.; Narita, S.; Mitsuzuka, K.; Duivenvoorden, W.; Pinthus, J. H.; Hashimoto, Y.; Koie, T.; Habuchi, T.; Arai, Y.; Ohyama, C. An Automated Micro-Total Immunoassay System for Measuring Cancer-Associated A2,3-Linked Sialyl N-Glycan-Carrying Prostate-Specific Antigen May Improve the Accuracy of Prostate Cancer Diagnosis. *Int. J. Mol. Sci.* **2017**, *18* (2).
- (38) Yoneyama, T.; Ohyama, C.; Hatakeyama, S.; Narita, S.; Habuchi, T.; Koie, T.; Mori, K.; Hidari, K. I. P. J.; Yamaguchi, M.; Suzuki, T.; Tobisawa, Y. Measurement of Aberrant Glycosylation of Prostate Specific Antigen Can Improve Specificity in Early Detection of Prostate Cancer. *Biochem. Biophys. Res. Commun.* **2014**, *448* (4), 390–396.
- (39) Sarrats, A.; Comet, J.; Tabarés, G.; Ramírez, M.; Aleixandre, R. N.; De Llorens, R.; Peracaula, R. Differential Percentage of Serum Prostate-Specific Antigen Subforms

- Suggests a New Way to Improve Prostate Cancer Diagnosis. *Prostate* **2010**, *70* (1), 1–9.
- (40) Sarrats, A.; Saldova, R.; Comet, J.; O'Donoghue, N.; De Llorens, R.; Rudd, P. M.; Peracaula, R. Glycan Characterization of PSA 2-DE Subforms from Serum and Seminal Plasma. *Omi. A J. Integr. Biol.* **2010**, *14* (4), 465–474.
- (41) Peracaula, R.; Tabarés, G.; Royle, L.; Harvey, D. J.; Dwek, R. A.; Rudd, P. M.; de Llorens, R. Altered Glycosylation Pattern Allows the Distinction between Prostate-Specific Antigen (PSA) from Normal and Tumor Origins. *Glycobiology* **2003**, *13* (6), 457–470.
- (42) Hatakeyama, S.; Yoneyama, T.; Tobisawa, Y.; Ohyama, C. Recent Progress and Perspectives on Prostate Cancer Biomarkers. *Int. J. Clin. Oncol.* **2017**, *22* (2), 214–221.
- (43) Oliveira-ferrer, L.; Legler, K.; Milde-langosch, K. Role of Protein Glycosylation in Cancer Metastasis. *Semin. Cancer Biol.* **2017**, *44*, 141–152.
- (44) Saldova, R.; Royle, L.; Catherine, M.; Hamid, U. M. A.; Evans, R.; James, N.; Banks, R. E.; Hutson, R.; Harvey, J.; Antrobus, R.; Petrescu, S. M.; Dwek, R. A.; Rudd, P. M. Ovarian Cancer Is Associated with Changes in Glycosylation in Both Acute-Phase Proteins and IgG. *Glycobiology* **2007**, *17* (12), 1344–1356.
- (45) Ohyama, C.; Tsuboi, S.; Fukuda, M. Dual Roles of Sialyl Lewis X Oligosaccharides in Tumor Metastasis and Rejection by Natural Killer Cells. *EMBO J.* **1999**, *18* (6), 1516–1525.
- (46) Williams, G. T.; Kedge, J. L.; Fossey, J. S. Molecular Boronic Acid-Based Saccharide Sensors. *ACS Sens.* **2021**, *6*, 1508–1528.
- (47) Wang, S. S.; Cheng, C.-M. Glycan-Based Diagnostic Devices: Current Progress, Challenges and Perspectives. *Chem. Commun* **2015**, *51*, 16750.

- (48) Varki, A. Biological Roles of Oligosaccharides: All of the Theories Are Correct. *Glycobiology* **1993**, 3 (2), 97–130.
- (49) Szarek, W.; Hay, G. W.; Vyas, D. M.; Ison, E. R.; Hronowski, L. J. J. L-Glucose. A Convenient Synthesis from D-Glucose. *Can. J. Chem* **1984**, 67, 671.
- (50) James, T. D.; Phillips, M. D.; Shink, S. *The Molecular Recognition of Saccharides*; 2006.
- (51) Aoyama, Y.; Yasutaka, T.; Hiroo, T.; Ogoshi, H. Polar Host-Guest Interaction. Binding of Nonionic Polar Compounds with a Resorcinol-Aldehyde Cyclooligomer as a Lipophilic Polar Host. *J. Am. Chem. Soc.* **1988**, 110, 635–637.
- (52) Imberty, A. 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) Tromans, R. A.; Carter, T. S.; Chabanne, L.; Crump, M. P.; Li, H.; Matlock, J. V.; Orchard, M. G.; Davis, A. P. A Biomimetic Receptor for Glucose. *Nat. Chem.* **2019**, 11 (1), 52–56.
- (54) McNichols, R. J.; Cote, G. L. Optical Glucose Sensing in Biological Fluids : An Overview. *J. Biomed. Opt.* **2000**, 5 (1), 5–16.
- (55) Wang, B.; Boons, G.-J. *Carbohydrate Recognition : Biological Problems, Methods, and Applications*, 1st ed.; John Wiley & Sons, Incorporated: Hoboken, NJ, USA, 2011.
- (56) James, T. D.; Sandanayake, R. A. S. Saccharide Sensing with Molecular Receptors Based on Boronic Acid. *Angew. Chem. Int. Ed. Engl.* **1996**.
- (57) Li, Z.; Lazaridis, T. The Effect of Water Displacement on Binding Thermodynamics : Concanavalin A. *J. Phys. Chem.* **2005**, 109, 662–670.
- (58) Michel, J.; Tirado-rives, J.; Jorgensen, W. L. Energetics of Displacing Water Molecules

- from Protein Binding Sites : Consequences for Ligand Optimization. *J. Am. Chem. Soc.* **2009**, *131* (42), 15403–15411.
- (59) 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.
- (60) Schiebel, J.; Gaspari, R.; Wulsdorf, T.; Ngo, K.; Sohn, C.; Schrader, T. E.; Cavalli, A.; Ostermann, A.; Heine, A.; Klebe, G. Intriguing Role of Water in Protein-Ligand Binding Studied by Neutron Crystallography on Trypsin Complexes. *Nat Commun* **2018**, *9*, 3359.
- (61) Mandal, C. Sialic-Acid Binding Lectins. *Experientia* **1990**, *46* (5), 433–441.
- (62) Lehmann, F.; Tiralongo, E.; Tiralongo, J. Sialic Acid-Specific Lectins: Occurrence, Specificity and Function. *Cell. Mol. Life Sci.* **2006**, *63* (12), 1331–1354.
- (63) Hendrickson, O. D.; Zherdev, A. V. Analytical Application of Lectins. *Crit. Rev. Anal. Chem* **2018**, *48* (4), 279–292.
- (64) Quiocho, F. A. Carbohydrate-Binding Proteins: Tertiary Structures and Protein-Sugar Interactions. *Annu. Rev. Biochem.* **1986**, *5*, 287–315.
- (65) Weis, W. I.; Drickamer, K. Structural Basis of Lectin-Carbohydrate Recognition. *Annu. Rev. Biochem.* **1996**, *65*, 441–473.
- (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) 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* (15152–15160), 15152–15160.
- (68) Crocker, P.; Vinson, M. Molecular Analysis of Sialoside Binding to Sialoadhesin by NMR and Site-Directed Mutagenesis. *Biochem J* **1999**, *341* (2), 355–361.
- (69) Varki, A.; Angata, T. Siglecs—the Major Subfamily of I-Type Lectins. *Glycobiology* **2006**, *16* (1), 1R-27R.
- (70) Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. Glycosylation in Health and Disease. *Nat. Rev. Nephrol.* **2019**, *15* (6), 346–366.
- (71) 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.
- (72) 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.
- (73) Wright, C. S. A-Resolution Structure-Analysis of 2 Refined N-Acetylneuraminyllactose - Wheat-Germ-Agglutinin Isolectin Complexes. *J. Mol. Biol.* **1990**, *215* (4), 635–651.
- (74) Yamamoto, K.; Konami, Y.; Irimura, T. Sialic Acid-Binding Motif of *Maackia Amurensis* Lectins. *J. Biochem.* **1997**, *121* (4), 756–761.
- (75) 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 Nterminal Octapeptides from Human Glycophorin A. *FEBS Lett.* **1994**, *342* (3), 334–338.
- (76) Yuan, Q. Q.; He, J. L.; Niu, Y. Z.; Chen, J.; Zhao, Y. L.; Zhang, Y. C.; Yu, C. Sandwichtype

- Biosensor for the Detection of Alpha 2,3-Sialylated Glycans Based on Fullerene-Palladiumplatinum Alloy and 4-Mercaptophenylboronic Acid Nanoparticle Hybrids Coupled with Aumethylene Blue-MAL Signal Amplification. *Biosens. Bioelectron.* **2018**, *102*, 321–327.
- (77) 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.
- (78) Syed, P.; Gidwani, K.; Kekki, H.; Leivo, J.; Pettersson, K.; Lamminmäki, U. Role of Lectin Microarrays in Cancer Diagnosis. *Proteomics* **2016**, *16* (8), 1257–1265.
- (79) Silva, M. L. S. Lectin-Based Biosensors as Analytical Tools for Clinical Oncology. *Cancer Lett.* **2018**, *436*, 63–74.
- (80) Shibuya, N.; Goldstein, I. J.; Broekaert, W. F.; Nsimba-Lubaki, M. Peeters, B.; Peumans, W. The Elderberry (*Sambucus Nigra* L.) Bark Lectin Recognizes the Neu5Ac(Alpha 2-6)Gal/GalNAc Sequence. *J. Biol. Chem.* **1987**, *262* (4), 1596–1601.
- (81) Shibuya, N. Goldstein, I. J. Broekaert, W. F. Nsimbalubaki, M. Peeters, B. Peumans, W. Fractionation of Sialylated Oligosaccharides, Glycopeptides, and Glycoproteins on Immobilized Elderberry (*Sambucus Nigra* L.) Bark Lectin. *Arch. Biochem. Biophys.* **1987**, *254* (1), 1–8.
- (82) 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**, *214* (10), 3398–3404.

- (83) 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).
- (84) 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.
- (85) 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.
- (86) 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.
- (87) 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.
- (88) 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.
- (89) Dotan N; Altstock RT; Schwarz M; Dukler A. Anti-Glycan Antibodies as Biomarkers for Diagnosis and Prognosis. *Lupus* **2006**, *15* (7), 442–450.
- (90) Wandall, H. H.; Blixt, O.; Tarp, M. A.; Pedersen, J. W.; Bennett, E. P.; Mandel, U.; Ragupathi, G.; Livingston, P. O.; Hollingsworth, M. A.; Taylor-Papadimitriou, J.;

- Burchell, J.; Clausen, H. Cancer Biomarkers Defined by Autoantibody Signatures to Aberrant O-Glycopeptide Epitopes. *Cancer Res.* **2010**, *70* (4), 1306–1313.
- (91) Pifferi, C.; Thomas, B.; Goyard, D.; Berthet, N.; Renaudet, O. Heterovalent Glycodendrimers as Epitope Carriers for Antitumor Synthetic Vaccines. *Chem. – A Eur. J.* **2017**, *23* (64), 16283–16296.
- (92) Qin, Q.; Yin, Z.; Wu, X.; Haas, K. M.; Huang, X. Valency and Density Matter: Deciphering Impacts of Immunogen Structures on Immune Responses against a Tumor Associated Carbohydrate Antigen Using Synthetic Glycopolymers. *Biomaterials* **2016**, *101*, 189–198.
- (93) Astronomo, R. D.; Burton, D. R. Carbohydrate Vaccines: Developing Sweet Solutions to Sticky Situations? *Nat. Rev. Drug Discov.* **2010**, *9* (4), 308–324.
- (94) Xu, H.; et. al. The in Vitro and in Vivo Effects of Anti-Galactose Antibodies on Endothelial Cell Activation and Xenograft Rejection. *J. Immunol.* **2003**, *170* (3), 1531–1539.
- (95) Estrada, J. L.; Al., E. Evaluation of Human and Non-Human Primate Antibody Binding to Pig Cells Lacking GGTA1/CMAH/B4GalNT2 Genes. *Xenotransplantation* **2015**, *22* (3), 194–202.
- (96) Polonskaya, Z.; et. al. T Cells Control the Generation of Nanomolar-Affinity Antiglycan Antibodies. *J. Clin. Invest.* **2017**, *127* (4), 1491–1504.
- (97) Polonskaya, Z.; Savage, P. B.; Finn, M. G.; Teyton, L. High-Affinity Anti-Glycan Antibodies: Challenges and Strategies. *Curr. Opin. Immunol.* **2019**, *59*, 65–71.
- (98) Haji-Ghassemi, O.; Blackler, R. J.; Young, N. M.; Evans, S. V. Antibody Recognition of

- Carbohydrate Epitopes. *Glycobiology* **2015**, *25* (9), 920–952.
- (99) 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.
- (100) Bovin, N.; et al. Repertoire of Human Natural Anti-Glycan Immunoglobulins. Do We Have Auto-Antibodies? *Biochim Biophys Acta* **2012**, *1820* (9), 1373–1382.
- (101) Makeneni, S.; et al. Predicting the Origins of Anti-Blood Group Antibody Specificity: A Case Study of the ABO A- and B-Antigens. *Front. Immunol.* **2014**, *5*, 397.
- (102) 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), 17–23.
- (103) Badjic, J. D.; Nelson, A.; Cantrill, S. J.; Turnbull, W.; Stoddart, J. F. Multivalency and Cooperativity in Supramolecular Chemistry. *Acc. Chem. Res* **2005**, *38* (9), 723–732.
- (104) Davis, A. P. Biomimetic Carbohydrate Recognition. *Chem. Soc. Rev.* **2020**, *49*, 2531–2545.
- (105) Davis, A. P.; Wareham, R. . Carbohydrate Recognition through Noncovalent Interactions: A Challenge for Biomimetic and Supramolecular Chemistry. *Angew Chem Int Ed Engl.* **1999**, *38*, 2978–2996.
- (106) James, T. D.; Davis, A. P. *Carbohydrate Receptors, in Functional Synthetic Receptors*; Schrader, T., Hamilton, A. D., Eds.; Wiley–VCH, 2005.
- (107) 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.
- (108) Tuerk, C.; Gold, L. Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science (80-.)*. **1990**, *249* (4968), 505–510.
- (109) Aptamer discovery. Selex <https://2018.igem.org/Team:Madrid-OLM/AptDiscovery> (accessed Apr 12, 2023).
- (110) Ciesiolka, J.; Yarus, M. Small RNA-Divalent Domains. *RNA* **1996**, *2*, 785–793.
- (111) Hofmann, H. P.; Limmer, S.; Hornung, V.; Sprinzl, M. Ni²⁺ Binding RNA Motifs with an Asymmetric Purine-Rich Internal Loop and a G-A Base Pair. *RNA* **1997**, *3*, 1289–1300.
- (112) Rajendran, M.; Ellington, A. D. Selection of Fluorescent Aptamer Beacons That Light up in the Presence of Zinc. *Anal. Bioanal. Chem.* **2008**, *390*, 1067–1075.
- (113) Geiger, A.; Burgstaller, P.; Von der Eltz, H.; Roeder, A.; Famulok, M. RNA Aptamers That Bind L-Arginine with Sub-Micromolar Dissociation Constants and High Enantioselectivity. *Nucleic Acids Res.* **1996**, *24*, 1029–1036.
- (114) Connell, G. J.; Illangesekare, M.; Yarus, M. Three Small Ribooligonucleotides with Specific Arginine Sites. *Biochemistry* **1993**, *32*, 5497–5502.
- (115) Nieuwlandt, D.; Wecker, M.; Gold, L. In Vitro Selection of RNA Ligands to Substance P. *Biochemistry* **1995**, *34*, 5651–5659.
- (116) Williams, K. P. NBioactive and Nuclease-Resistant I-DNA Ligand of Vasopressin. *Proc. Natl Acad. Sci. USA* **1997**, *94*, 11285–11920.
- (117) Ellington, A. D.; Szostak, J. W. In Vitro Selection of RNA Molecules That Bind Specific

Ligands. *Nature* **1990**, 346, 818–822.

- (118) Chen, H.; McBroom, D. G.; Zhu, Y.-Q.; Gold, L.; North, T. W. Inhibitory RNA Ligand to Reverse Transcriptase from Feline Immunodeficiency Virus. *Biochemistry* **1996**, 35, 6923–6930.
- (119) Morris, K. N.; Jensen, K. B.; Julin, C. M.; Weil, M.; Gold, L. High Affinity Ligands from in Vitro Selection: Complex Targets. *Proc. Natl Acad. Sci. USA* **1998**, 95, 2902–2907.
- (120) Blank, M.; Weinschenk, T.; Priemer, M.; Schluesener, H. Systematic Evolution of a DNA Aptamer Binding to Rat Brain Tumor Microvessels. Selective Targeting of Endothelial Regulatory Protein Pigpen. *J. Biol. Chem.* **2001**, 276, 16464.
- (121) Kwame, S.; Shangguan, D.; Xiong, X.; Donoghue, M. B. O.; Tan, W. Development of DNA Aptamers Using Cell-SELEX. *Nat. Protoc.* **2010**, 5 (6).
- (122) 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.
- (123) 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.
- (124) Yang, K.; Barbu, M.; Halim, M.; Pallavi, P.; Kim, B.; Kolpashchikov, D. M.; Pecic, S.; Taylor, S.; Worgall, T. S.; Stojanovic, M. N. Recognition and Sensing of Low-Epitope Targets via Ternary Complexes with Oligonucleotides and Synthetic Receptors. *Nat. Chem.* **2014**, 6 (11), 1003–1008.
- (125) 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.
- (126) 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.
- (127) R. P. Bonar-Law; A. P. Davis; B. A. Murray. Artificial Receptors for Carbohydrate Derivatives. *Angew. Chem. Int. Ed.* **1990**, *29*, 1407–1008.
- (128) S. Anderson, U. Neidlein, V. Gramlich, F. D. A New Family of Chiral Binaphthyl-Derived Cyclophane Receptors: Complexation of Pyranosides. *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 1596–1600.
- (129) A. P. Davis; R. S. Wareham. A Tricyclic Polyamide Receptor for Carbohydrates in Organic Media. *Angew. Chem. Int. Ed.* **1998**, *37*, 2270–2273.
- (130) H. Abe; T. Yoneda; Y. Ohishi; M. Inouye. D-3h-Symmetrical Shape-Persistent Macrocycles Consisting of Pyridine-Acetylenephenol Conjugates as an Efficient Host Architecture for Saccharide Recognition. *Chem. – Eur. J.* **2016**, *22*, 18944–18952.
- (131) Francesconi, O.; Ienco, A.; Moneti, G.; Nativi, C.; Roelens, S. Self-Assembled Pyrrolic Cage Receptor Specifically Recognizes Beta-Glucopyranosides. *Angew Chem Int Ed Engl.* **2006**, *45*, 6693–6696.
- (132) D. W. Zhang; A. Martinez; J. P. Dutasta. Emergence of Hemicryptophanes: From Synthesis to Applications for Recognition, Molecular Machines, and Supramolecular Catalysis. *Chem. Rev.* **2017**, *117*, 4900–4942.
- (133) E. Klein, M. P. Crump; A. P. Davis. Carbohydrate Recognition in Water by a Tricyclic

Polyamide Receptor. *Angew. Chem. Int. Ed.* **2005**, *44*, 298–302.

- (134) C. Ke; H. Destecroix; M. P. Crump; A. P. Davis. A Simple and Accessible Synthetic Lectin for Glucose Recognition and Sensing. *Nat. Chem.* **2012**, *4*, 718–723.
- (135) P. K. Mandal; B. Kauffmann; H. Destecroix; Y. Ferrand; A. P. Davis; I. Huc. Crystal Structure of a Complex between Beta-Glucopyranose and a Macrocyclic Receptor with Dendritic Multicharged Water Solubilizing Chains. *Chem. Commun.* **2016**, *52*, 9355–9358.
- (136) H. Destecroix; C. M. Renney; T. J. Mooibroek; T. S. Carter; P. F. N. Stewart; M. P. Crump; A. P. Davis. Affinity Enhancement by Dendritic Side Chains in Synthetic Carbohydrate Receptors. *Angew. Chem. Int. Ed.* **2015**, *54*, 2057–2061.
- (137) T. J. Mooibroek, J. M. Casas-Solvas, R. L. H.; C. M. Renney; T. S. Carter; M. P. Crump; A. P. Davis. A Threading Receptor for Polysaccharides. *Nat. Chem.* **2016**, *8*, 69–74.
- (138) Y. Jang; R. Natarajan; Y. H. Ko; K. Kim. Cucurbit[7]Uril: A High-Affinity Host for Encapsulation of Amino Saccharides and Supramolecular Stabilization of Their Alpha-Anomers in Water. *Angew. Chem. Int. Ed.* **2014**, *53*, 1003–1007.
- (139) O. Francesconi; M. Martinucci; L. Badii; C. Nativi; S. Roelens. A Biomimetic Synthetic Receptor Selectively Recognising Fucose in Water. *Chem. – Eur. J.* **2018**, *24*, 6828–6836.
- (140) Y. Ferrand; M. P. Crump; A. P. Davis. A Synthetic Lectin Analog for Biomimetic Disaccharide Recognition. *Science (80-.)*. **2007**, *318*, 619–622.
- (141) B. Sookcharoenpinyo, E. Klein, Y. Ferrand, D. B. W.; P. R. Brotherhood; C. Ke; M. P. Crump; A. P. Davis. Highaffinity Disaccharide Binding by Tricyclic Synthetic Lectins.

Angew. Chem. Int. Ed. **2012**, *51*, 4586–4590.

- (142) Toone, E. J. Structure and Energetics of Protein-Carbohydrate Complexes. *Curr. Opin. Struct. Biol.* **1994**, *4*, 719–728.
- (143) Das, G.; Hamilton, A. D. Molecular Recognition of Carbohydrates: Strong Binding of Alkyl Glycosides by Phosphonate Derivatives. *J. Am. Chem. Soc.* **1994**, *116*, 11139.
- (144) Mazik, M. Recent Developments in the Molecular Recognition of Carbohydrates by Artificial Receptors. *RSC Adv.* **2012**, *2*, 2630–2642.
- (145) O. Francesconi, C. Nativi, G. Gabrielli, I. D. S.; S. Noppen, J. Balzarini, S. Liekens, S. R. Antiviral Activity of Synthetic Aminopyrrolic Carbohydrate Binding Agents: Targeting the Glycans of Viral Gp120 to Inhibit HIV Entry. *Chem. – Eur. J.* **2015**, *21*, 10089–10093.
- (146) J. Billing, H. Grundberg, U. J. N. Amphiphilic Anthracene-Amino Acid Conjugates as Simple Carbohydrate Receptors in Water. *Supramol. Chem.* **2002**, *14*, 367–372.
- (147) M. Mazik; H. Cavga. Carboxylate-Based Receptors for the Recognition of Carbohydrates in Organic and Aqueous Media. *J. Org. Chem.* **2006**, *71*, 2957–2963.
- (148) Corredor, M.; Carbajo, D.; Domingo, C.; Yolanda, P. Ø.; Bujons, J.; Messeguer, A.; Alfonso, I. Dynamic Covalent Identification of an Efficient Heparin Ligand. *Angew Chem Int Ed Engl.* **2018**, *57*, 11973–11977.
- (149) Carbajo, D.; Perez, Y.; Bujons, J.; Alfonso, I. Live-Cell-Templated Dynamic Combinatorial Chemistry. **2020**, *54*, 17202–17206.
- (150) T. S. Carter, T. J. Mooibroek, P. F. N. Stewart, M. P. C.; M. C. Galan; A. P. Davis. Platform Synthetic Lectins for Divalent Carbohydrate Recognition in Water. *Angew.*

Chem. Int. Ed. Int. Ed. **2016**, *55*, 9311–9315.

- (151) James, T. D.; Phillips, M. D.; Shinkai, S. *Boronic Acids in Saccharide Recognition*; Rowan, A. E., Rowan, S. J., Aida, T., Stoddart, J. F., Eds.; The Royal Society of Chemistry, 2006.
- (152) Lorand, J. P.; Edwards, J. O. Polyol Complexes and Structure of the Benzeneboronate Ion. *J. Am. Chem. Soc.* **1959**, *24* (6), 769–774.
- (153) 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.
- (154) Nakatani, H.; Hiromi, K. Binding of m -Nitrobenzeneboronic Acids into the Active Site of Subtilisin BPN'. *Biochim. Biophys. Acta* **1978**, *524*, 413–417.
- (155) Springsteen, G.; Wang, B. A Detailed Examination of Boronic Acid-Diol Complexation. *Tetrahedron* **2002**, *58* (26), 5291–5300.
- (156) 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.
- (157) 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*, 25–27.
- (158) 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.
- (159) 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.
- (160) 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.
- (161) Norrild, J. C.; Eggert, H. Evidence for Mono- and Bidentate 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.
- (162) 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.
- (163) 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.
- (164) 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.
- (165) Fossey, J. S.; James, T. D. Boronic Acid Based Modular Fluorescent Saccharide Sensors. In *Reviews in Fluorescence*; Geddes, C., Ed.; 2007; pp 103–118.
- (166) 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.

- (167) 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**, No. 17, 3237–3248.
- (168) 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**, No. 6, 792–793.
- (169) 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.
- (170) 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.
- (171) 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.
- (172) 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.
- (173) 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.

- (174) Duggan, P. J.; Offermann, D. A. Remarkably Selective Saccharide Recognition by Solidsupported Peptide Boronic Acids. *Tetrahedron* **2009**, *65* (1), 109–114.
- (175) 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.
- (176) Bruen, D.; Delaney, C.; Diamond, D.; Florea, L. Fluorescent Probes for Sugar Detection. *ACS Appl. Mater. Interfaces* **2018**, *10* (44), 38431–38437.
- (177) 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.
- (178) Stephenson-Brown, A.; Wang, H. C.; Iqbal, P.; Preece, J. A.; Long, Y.; 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.
- (179) 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.* **2019**, *131* (34), 12022.
- (180) 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).
- (181) 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.

- (182) 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.
- (183) Tommasone, S.; Tagger, Y. K.; Mendes, P. M. Targeting Oligosaccharides and Glycoconjugates Using Superselective Binding Scaffolds. *Adv. Funct. Mater.* **2020**, 30 (31), 2002298.
- (184) 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.
- (185) Frei, P.; Hevey, R.; Ernst, B. Dynamic Combinatorial Chemistry: A New Methodology Comes of Age. *Chem. - A Eur. J.* **2019**, 25 (1), 60–73.
- (186) Nishinaga, T.; Tanatani, A.; Oh, K.; Moore, J. S. The Size-Selective Synthesis of Folded Oligomers by Dynamic Templatation. *J. Am. Chem. Soc.* **2002**, 124 (21), 5934–5935.
- (187) Sando, S.; Narita, A.; Aoyama, Y. A Facile Route to Dynamic Glycopeptide Libraries Based on Disulfide-Linked Sugar–Peptide Coupling. *Bioorg. Med. Chem. Lett.* **2004**, 14 (11), 2835–2838.
- (188) Rauschenberg, M.; Bomke, S.; Karst, U.; Ravoo, B. J. Dynamic Peptides as Biomimetic Carbohydrate Receptors. *Angew. Chemie Int. Ed.* **2010**, 49 (40), 7340–7345.
- (189) Hasenknopf, B.; Lehn, J.-M.; Kneisel, B. O.; Baum, G.; Fenske, D. Self-Assembly of a Circular Double Helicate. *Angew. Chemie Int. Ed. English* **1996**, 35 (16), 1838–1840.
- (190) Huc, I.; Krische, M. Dynamic Combinatorial Chemistry: Substrate H-Bonding Directed

- Assembly of Receptors Based on Bipyridine-Metal Complexes. *European Journal of Inorganic Chemistry*. pp 1415–1420.
- (191) Cai, M.; Shi, X.; Sidorov, V.; Fabris, D. Cation-Directed Self-Assembly of Lipophilic Nucleosides: The Cation's Central Role in the Structure and Dynamics of a Hydrogen-Bonded Assembly. *Tetrahedron* **2002**, *58*, 661–671.
- (192) 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. Engl.* **2021**, *60* (9), 4485–4490.
- (193) Nicolaou, K. C.; Hughes, R.; Pfefferkorn, J. A.; Barluenga, S.; Roecker, A. . Combinatorial Synthesis through Disulfide Exchange: Discovery of Potent Psammaplin A Type Antibacterial Agents Active against Methicillin-Resistant *Staphylococcus Aureus* (MRSA). *Chem. Eur. J.* **2001**, *7*, 4280–4295.
- (194) Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. Dynamic Combinatorial Chemistry. *Drug Discov. Today* **2002**, *7* (2), 117–125.
- (195) Ramström, O.; Lehn, J. M. Drug Discovery by Dynamic Combinatorial Libraries. *Nat. Rev. Drug Discov.* **2002**, *1* (1), 26–36.
- (196) Ramström, O.; Bunyapaiboonsri, T.; Lohmann, S.; Lehn, J. M. Chemical Biology of Dynamic Combinatorial Libraries. *Biochim. Biophys. Acta - Gen. Subj.* **2002**, *1572* (2–3), 178–186.
- (197) 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*, 2455.

- (198) Monjas, L.; Hirsch, A. K. Harnessing Dynamic Combinatorial Chemistry in the Search for New Ligands for Protein Targets. *Future Med. Chem.* **2015**, *7* (16), 2095–2098.
- (199) Maraković, N.; Šinko, G. The Lock Is the Key: Development of Novel Drugs through Receptor Based Combinatorial Chemistry. *Acta Chim. Slov.* **2017**, *64* (1), 15–39.
- (200) Huang, R.; Leung, I. K. H. Protein-Directed Dynamic Combinatorial Chemistry: A Guide to Protein Ligand and Inhibitor Discovery. *Molecules* **2016**, *21* (7).
- (201) 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.
- (202) Larsson, R.; Pei, Z.; Ramström, O. Catalytic Self-Screening of Cholinesterase Substrates from a Dynamic Combinatorial Thioester Library. *Angew. Chemie Int. Ed.* **2004**, *43* (28), 3716–3718.
- (203) Larsson, R.; Ramström, O. Dynamic Combinatorial Thiolester Libraries for Efficient Catalytic Self-Screening of Hydrolase Substrates. *European J. Org. Chem.* **2006**, *2006* (1), 285–291.
- (204) Zhang, Y.; Angelin, M.; Larsson, R.; Albers, A.; Simons, A.; Ramström, O. Tandem Driven Dynamic Self-Inhibition of Acetylcholinesterase. *Chem. Commun.* **2010**, *46* (44), 8457–8459.
- (205) Ramström, O.; Bunyapaiboonsri, T.; Lohmann, S.; Lehn, J. M. Chemical Biology of Dynamic Combinatorial Libraries. *Biochim. Biophys. Acta - Gen. Subj.* **2002**, *1572* (2–3), 178–186.
- (206) Cheeseman, J. D.; Corbett, A. D.; Gleason, J. L.; Kazlauskas, R. J. Receptor-Assisted Combinatorial Chemistry: Thermodynamics and Kinetics in Drug Discovery. *Chem.* – A

Eur. J. **2005**, *11* (6), 1708–1716.

- (207) 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.
- (208) Brady, P. A.; Sanders, J. K. M. Selection Approaches to Catalytic Systems. *Chem. Soc. Rev.* **1997**, *26* (5), 327–336.
- (209) Hunt, R. A. R.; Otto, S. Dynamic Combinatorial Libraries: New Opportunities in Systems Chemistry. *Chem. Commun.* **2011**, *47* (3), 847–858.
- (210) Belowich, M. E.; Stoddart, J. F. Dynamic Imine Chemistry. *Chem. Soc. Rev.* **2012**, *41* (6), 2003–2024.
- (211) Cougnon, F. B. L.; Sanders, J. K. M. Evolution of Dynamic Combinatorial Chemistry. *Acc. Chem. Res.* **2012**, *45* (12), 2211–2221.
- (212) 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.
- (213) Eichner, S.; Knobloch, T.; Floss, H. G.; Fohrer, J.; Harmrolfs, K.; Hermans, J.; Schulz, A.; Sasse, F.; Spiteller, P.; Taft, F.; Kirschning, A. No Title. *Angew. Chemie - Int. Ed.* **2012**, *51*, 752.
- (214) Matache, M.; Bogdan, E.; Hădăde, N. D. Selective Host Molecules Obtained by Dynamic Adaptive Chemistry. *Chem. – A Eur. J.* **2014**, *20* (8), 2106–2131.
- (215) Rasmussen, B.; Sørensen, A.; Beeren, S. R.; Pittelkow, M. *Dynamic Combinatorial Chemistry*; Wiley, Hoboken: New Jersey, 2014.
- (216) Zhang, Y.; Barboiu, M. Constitutional Dynamic Materials—Toward Natural Selection of

Function. *Chem. Rev.* **2016**, *116* (3), 809–834.

- (217) Klekota, B.; Miller, B. L. Dynamic Diversity and Small-Molecule Evolution: A New Paradigm for Ligand Identification. *Trends Biotechnol.* **1999**, *17* (5), 205–209.
- (218) Lehn, J. M. Dynamic Combinatorial Chemistry and Virtual Combinatorial Libraries. *Chem. – A Eur. J.* **1999**, *5*, 2455–2463.
- (219) Cousins, G. R. L.; Poulsen, S. A.; Sanders, J. K. M. Molecular Evolution: Dynamic Combinatorial Libraries, Autocatalytic Networks and the Quest for Molecular Function. *Curr. Opin. Chem. Biol.* **2000**, *4* (3), 270–279.
- (220) I, H.; R, N. Dynamic Combinatorial Chemistry. *Comb. Chem. High Throughput Screen.* **2001**, *4* (1), 186–193.
- (221) 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.
- (222) Rowan, S. J.; Cantrill, S. J.; Cousins, G. R. L.; Sanders, J. K. M.; Stoddart, J. . Dynamic Covalent Chemistry. *Angew. Chem. Int. Ed.* **2002**, *41*, 898–952.
- (223) Ladame, S. Dynamic Combinatorial Chemistry: On the Road to Fulfilling the Promise. *Org. Biomol. Chem.* **2008**, *6* (2), 219–226.
- (224) Miller, B. L. *Dynamic Combinatorial Chemistry: In Drug Discovery, Bioorganic Chemistry and Material Sciences*; Wiley: New Jersey, 2010.
- (225) Whitney, A. M.; Ladame, S.; Balasubramanian, S. Templated Ligand Assembly by Using G-Quadruplex DNA and Dynamic Covalent Chemistry. *Angew. Chemie - Int. Ed.* **2004**, *43* (9), 1143–1146.
- (226) Bugaut, A.; Toulmé, J. J.; Rayner, B. Use of Dynamic Combinatorial Chemistry for the

- Identification of Covalently Appended Residues That Stabilize Oligonucleotide Complexes. *Angew. Chemie - Int. Ed.* **2004**, *43* (24), 3144–3147.
- (227) Ladame, S.; Whitney, A. M.; Balasubramanian, S. Targeting Nucleic Acid Secondary Structures with Polyamides Using an Optimized Dynamic Combinatorial Approach. *Angew. Chemie - Int. Ed.* **2005**, *44* (35), 5736–5739.
- (228) Bugaut, A.; Toulmé, J. J.; Rayner, B. SELEX and Dynamic Combinatorial Chemistry Interplay for the Selection of Conjugated RNA Aptamers. *Org. Biomol. Chem.* **2006**, *4* (22), 4082–4088.
- (229) McNaughton, B. R.; Miller, B. L. Resin-Bound Dynamic Combinatorial Chemistry. **2006**, No. 5, 7–10.
- (230) McNaughton, B. R.; Gareiss, P. C.; Miller, B. L. Identification of a Selective Small-Molecule Ligand for HIV-1 Frameshift-Inducing Stem-Loop RNA from an 11,325 Member Resin Bound Dynamic Combinatorial Library. *J. Am. Chem. Soc.* **2007**, *129* (37), 11306–11307.
- (231) Bugaut, A.; Jantos, K.; Wietor, J. L.; Rodriguez, R.; Sanders, J. K. M.; Balasubramanian, S. Exploring the Differential Recognition of DNA G-Quadruplex Targets by Small Molecules Using Dynamic Combinatorial Chemistry. *Angew. Chemie - Int. Ed.* **2008**, *47* (14), 2677–2680.
- (232) Gareiss, P. C.; Sobczak, K.; McNaughton, B. R.; Palde, P. B.; Thornton, C. A.; Miller, B. L. Dynamic Combinatorial Selection of Molecules Capable of Inhibiting the (CUG) Repeat RNA-MBNL1 Interaction in Vitro: Discovery of Lead Compounds Targeting Myotonic Dystrophy (DM1). *J. Am. Chem. Soc.* **2008**, *130* (48), 16254–16261.

- (233) Nicolaou, K. C.; Hughes, R.; Cho, S. Y.; Winssinger, N.; Labischinski, H.; Endermann, R. Synthesis and Biological Evaluation of Vancomycin Dimers with Potent Activity against Vancomycin-Resistant Bacteria: Target-Accelerated Combinatorial Synthesis. *Chem. - A Eur. J.* **2001**, *7* (17), 3824–3843.
- (234) Baxter, P. N. W.; Khoury, R. G.; Lehn, J. M.; Baum, G.; Fenske, D. Adaptive Self-Assembly: Environment-Induced Formation and Reversible Switching of Polynuclear Metallocyclophanes. *Chemistry* **2000**, *6* (22), 4140–4148.
- (235) Cousins, G. R. L.; Furlan, R. L. E.; Ng, Y.-F.; Redman, J. E.; Sanders, J. K. M. Identification and Isolation of a Receptor for N-Methyl Alkylammonium Salts: Molecular Amplification in a Pseudo-Peptide Dynamic Combinatorial Library. *Angew. Chem. Int. Ed. Engl.* **2001**, *40* (2), 423–428.
- (236) Poulsen, S.-A. Direct Screening of a Dynamic Combinatorial Library Using Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2006**, *17* (8), 1074–1080.
- (237) Pei, Z.; Larsson, R.; Aastrup, T.; Anderson, H.; Lehn, J. M.; Ramström, O. Quartz Crystal Microbalance Bioaffinity Sensor for Rapid Identification of Glycosyldisulfide Lectin Inhibitors from a Dynamic Combinatorial Library. *Biosens. Bioelectron.* **2006**, *22* (1), 42–48.
- (238) André, S.; Pei, Z.; Siebert, H. C.; Ramström, O.; Gabius, H. J. Glycosyldisulfides from Dynamic Combinatorial Libraries as O-Glycoside Mimetics for Plant and Endogenous Lectins: Their Reactivities in Solid-Phase and Cell Assays and Conformational Analysis by Molecular Dynamics Simulations. *Bioorg. Med. Chem.* **2006**, *14* (18), 6314–6326.
- (239) Sindelar, M.; Lutz, T. A.; Petrera, M.; Wanner, K. T. Focused Pseudostatic Hydrazone Libraries Screened by Mass Spectrometry Binding Assay: Optimizing Affinities toward

- γ -Aminobutyric Acid Transporter 1. *J. Med. Chem.* **2013**, 56 (3), 1323–1340.
- (240) Mondal, M.; Groothuis, D. E.; Hirsch, A. K. H. Fragment Growing Exploiting Dynamic Combinatorial Chemistry of Inhibitors of the Aspartic Protease Endothiapepsin. *Med. Chem. Commun.* **2015**, 6 (7), 1267–1271.
- (241) Soubhye, J.; Gelbcke, M.; Van Antwerpen, P.; Dufrasne, F.; Boufadi, M. Y.; Nève, J.; Furtmüller, P. G.; Obinger, C.; Zouaoui Boudjeltia, K.; Meyer, F. From Dynamic Combinatorial Chemistry to in Vivo Evaluation of Reversible and Irreversible Myeloperoxidase Inhibitors. *ACS Med. Chem. Lett.* **2017**, 8 (2), 206–210.
- (242) Monjas, L.; Swier, L. J. Y. M.; Setyawati, I.; Slotboom, D. J.; Hirsch, A. K. H. Dynamic Combinatorial Chemistry to Identify Binders of ThiT, an S-Component of the Energy-Coupling Factor Transporter for Thiamine. *ChemMedChem* **2017**, 12 (20), 1693–1696.
- (243) Bhat, V. T.; Caniard, A. M.; Luksch, T.; Brenk, R.; Campopiano, D. J.; Greaney, M. F. Nucleophilic Catalysis of Acylhydrazone Equilibration for Protein-Directed Dynamic Covalent Chemistry. *Nat. Chem.* 2010 26 **2010**, 2 (6), 490–497.
- (244) Fang, Z.; He, W.; Li, X.; Li, Z.; Chen, B.; Ouyang, P.; Guo, K. A Novel Protocol to Accelerate Dynamic Combinatorial Chemistry via Isolation of Ligand–Target Adducts from Dynamic Combinatorial Libraries: A Case Study Identifying Competitive Inhibitors of Lysozyme. *Bioorg. Med. Chem. Lett.* **2013**, 23 (18), 5174–5177.
- (245) Fu, J.; Fu, H.; Dieu, M.; Halloum, I.; Kremer, L.; Xia, Y.; Pan, W.; Vincent, S. P. Identification of Inhibitors Targeting Mycobacterium Tuberculosis Cell Wall Biosynthesis via Dynamic Combinatorial Chemistry. *Chem. Commun.* **2017**, 53 (77), 10632–10635.

- (246) Bunyapaiboonsri, T.; Ramström, O.; Lohmann, S.; Lehn, J. M.; Peng, L.; Goeldner, M. Dynamic Deconvolution of a Pre-Equilibrated Dynamic Combinatorial Library of Acetylcholinesterase Inhibitors. *ChemBioChem* **2001**, 2 (6), 438–444.
- (247) Kehoe, J. W.; Maly, D. J.; Verdugo, D. E.; Armstrong, J. I.; Cook, B. N.; Ouyang, Y. Bin; Moore, K. L.; Ellman, J. A.; Bertozzi, C. R. Tyrosylprotein Sulfotransferase Inhibitors Generated by Combinatorial Target-Guided Ligand Assembly. *Bioorg. Med. Chem. Lett.* **2002**, 12 (3), 329–332.
- (248) Gerber-Lemaire, S. Popowycz, F.; Rodríguez-García, E. Asenjo, A. T. C.; Robina, I.; Vogel, P. An Efficient Combinatorial Method for the Discovery of Glycosidase Inhibitors. *ChemBioChem* **2002**, 3, 466–470.
- (249) 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.
- (250) Ramström, O.; Lohmann, S.; Bunyapaiboonsri, T.; Lehn, J.-M. Dynamic Combinatorial Carbohydrate Libraries: Probing the Binding Site of the Concanavalin A Lectin. *Chem. – A Eur. J.* **2004**, 10 (7), 1711–1715.
- (251) Eliseev, A. V; Nelen, M. I. Use of Molecular Recognition To Drive Chemical Evolution . 1 . Controlling the Composition of an Equilibrating Mixture of Simple Arginine Receptors. *J. Am. Chem. Soc.* **1997**, 119 (10), 1147–1148.
- (252) 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.

- (253) Zameo, S.; Vauzeilles, B.; Beau, J.-M. Dynamic Combinatorial Chemistry: Lysozyme Selects an Aromatic Motif That Mimics a Carbohydrate Residue. *Angew. Chemie Int. Ed.* **2005**, *44* (6), 965–969.
- (254) Valade, A.; Urban, D.; Beau, J.-M. Target-Assisted Selection of Galactosyltransferase Binders from Dynamic Combinatorial Libraries. An Unexpected Solution with Restricted Amounts of the Enzyme. *ChemBioChem* **2006**, *7*, 1023–1027.
- (255) Valade, A.; Urban, D.; Beau, J.-M. Two Galactosyltransferases' Selection of Different Binders from the Same Uridine-Based Dynamic Combinatorial Library. *J. Comb. Chem.* **2007**, *9* (1), 1–4.
- (256) Hochgürtel, M.; Kroth, H.; Piecha, D.; Hofmann, M. W.; Nicolau, C.; Krause, S.; Schaaf, O.; Sonnenmoser, G.; Eliseev, A. V. Target-Induced Formation of Neuraminidase Inhibitors from in Vitro Virtual Combinatorial Libraries. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (6), 3382–3387.
- (257) Zameo, S.; Vauzeilles, B.; Beau, J.-M. Direct Composition Analysis of a Dynamic Library of Imines in an Aqueous Medium. *European J. Org. Chem.* **2006**, *2006* (24), 5441–5444.
- (258) Nasr, G.; Petit, E.; Supuran, C. T.; Winum, J. Y.; Barboiu, M. Carbonic Anhydrase II-Induced Selection of Inhibitors from a Dynamic Combinatorial Library of Schiff's Bases. *Bioorg. Med. Chem. Lett.* **2009**, *19* (21), 6014–6017.
- (259) Hochgürtel, M.; Biesinger, R.; Kroth, H.; Piecha, D.; Hofmann, M. W.; Krause, S.; Schaaf, O.; Nicolau, C.; Eliseev, A. V. Ketones as Building Blocks for Dynamic Combinatorial Libraries: Highly Active Neuraminidase Inhibitors Generated via Selection Pressure of the Biological Target. *J. Med. Chem.* **2003**, *46* (3), 356–358.

- (260) Yang, Z.; Fang, Z.; He, W.; Wang, Z.; Gan, H.; Tian, Q.; Guo, K. Identification of Inhibitors for Vascular Endothelial Growth Factor Receptor by Using Dynamic Combinatorial Chemistry. *Bioorg. Med. Chem. Lett.* **2016**, *26* (7), 1671–1674.
- (261) Clipson, A. J.; Bhat, V. T.; Mcnae, I.; Caniard, A. M.; Campopiano, D. J.; Greaney, M. F. Bivalent Enzyme Inhibitors Discovered Using Dynamic Covalent Chemistry. *Chem. - A Eur. J.* **2012**, *18* (34), 10562–10570.
- (262) Mondal, M.; Radeva, N.; Fanlo-Virgós, H.; Otto, S.; Klebe, G.; Hirsch, A. K. H. Fragment Linking and Optimization of Inhibitors of the Aspartic Protease Endothiapepsin: Fragment-Based Drug Design Facilitated by Dynamic Combinatorial Chemistry. *Angew. Chemie Int. Ed.* **2016**, *55* (32), 9422–9426.
- (263) Frei, P.; Pang, L.; Silbermann, M.; Eriş, D.; Mühlethaler, T.; Schwardt, O.; Ernst, B. Target-Directed Dynamic Combinatorial Chemistry: A Study on Potentials and Pitfalls as Exemplified on a Bacterial Target. *Chem. – A Eur. J.* **2017**, *23* (48), 11570–11577.
- (264) Mondal, M.; Radeva, N.; Köster, H.; Park, A.; Potamitis, C.; Zervou, M.; Klebe, G.; Hirsch, A. K. H. Structure-Based Design of Inhibitors of the Aspartic Protease Endothiapepsin by Exploiting Dynamic Combinatorial Chemistry. *Angew. Chemie Int. Ed.* **2014**, *53* (12), 3259–3263.
- (265) Woon, E. C. Y.; Demetriades, M.; Bagg, E. A. L.; Aik, W.; Krylova, S. M.; Ma, J. H. Y.; Chan, M.; Walport, L. J.; Wegman, D. W.; Dack, K. N.; McDonough, M. A.; Krylov, S. N.; Schofield, C. J. Dynamic Combinatorial Mass Spectrometry Leads to Inhibitors of a 2-Oxoglutarate-Dependent Nucleic Acid Demethylase. *J. Med. Chem.* **2012**, *55* (5), 2173–2184.
- (266) Ramström, O.; Lehn, J.-M. In Situ Generation and Screening of a Dynamic

- Combinatorial Carbohydrate Library against Concanavalin A. *ChemBioChem* **2000**, *1*, 41–48.
- (267) Hotchkiss, T.; Kramer, H. B.; Doores, K. J.; Gamblin, D. P.; Oldham, N. J.; Davis, B. G. Ligand Amplification in a Dynamic Combinatorial Glycopeptide Library. *Chem. Commun.* **2005**, No. 34, 4264–4266.
- (268) Milanesi, L.; Hunter, C. A.; Sedelnikova, S. E.; Waltho, J. P. Amplification of Bifunctional Ligands for Calmodulin from a Dynamic Combinatorial Library. *Chem. – A Eur. J.* **2006**, *12* (4), 1081–1087.
- (269) Danieli, B.; Giardini, A.; Lesma, G.; Passarella, D.; Peretto, B.; Sacchetti, A.; Silvani, A.; Pratesi, G.; Zunino, F. Thiocolchicine–Podophyllotoxin Conjugates: Dynamic Libraries Based on Disulfide Exchange Reaction. *J. Org. Chem.* **2006**, *71* (7), 2848–2853.
- (270) Scott, D. E.; Dawes, G. J.; Ando, M.; Abell, C.; Ciulli, A. A Fragment-Based Approach to Probing Adenosine Recognition Sites by Using Dynamic Combinatorial Chemistry. *ChemBioChem* **2009**, *10* (17), 2772–2779.
- (271) Saiz, C.; Castillo, V.; Fontán, P.; Bonilla, M.; Salinas, G.; Rodríguez-Haralambides, A.; Mahler, S. G. Discovering Echinococcus Granulosus Thioredoxin Glutathione Reductase Inhibitors through Site-Specific Dynamic Combinatorial Chemistry. *Mol. Divers.* **2014**, *18* (1), 1–12.
- (272) Rose, N. R.; Woon, E. C. Y.; Kingham, G. L.; King, O. N. F.; Mecinovi, J.; Clifton, I. J.; Ng, S. S.; Talib-Hardy, J.; Oppermann, U.; McDonough, M. A.; Schofield, C. J. Selective Inhibitors of the JMJD2 Histone Demethylases: Combined Nondenaturing Mass Spectrometric Screening and Crystallographic Approaches †. *J. Med. Chem* **2010**, *53*, 1810–1818.

- (273) 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. Chemie Int. Ed.* **2012**, *51* (27), 6672–6675.
- (274) Leung, I. K. H.; Brown Jr, T.; Schofield, C. J.; Claridge, T. D. W. An Approach to Enzyme Inhibition Employing Reversible Boronate Ester Formation. *Med. Chem. Commun.* **2011**, *2* (5), 390–395.
- (275) Shi, B.; Stevenson, R.; Campopiano, D. J.; Greaney, M. F. Discovery of Glutathione S-Transferase Inhibitors Using Dynamic Combinatorial Chemistry. *J. Am. Chem. Soc.* **2006**, *128* (26), 8459–8467.
- (276) Shi, B.; Greaney, M. F. Reversible Michael Addition of Thiols as a New Tool for Dynamic Combinatorial Chemistry. *Chem. Commun.* **2005**, No. 7, 886–888.
- (277) Caraballo, R.; Dong, H.; Ribeiro, J.; Jiménez-Barbero, J Ramström, O. Direct STD NMR Identification of β -Galactosidase Inhibitors from a Virtual Dynamic Hemithioacetal System. *Angew. Chem. Int. Ed.* **2010**, *49*, 589–593.
- (278) Sakai, S.; Shigemasa, Y.; Sasaki, T. A Self-Adjusting Carbohydrate Ligand for GalNAc Specific Lectins. *Tetrahedron Lett.* **1997**, *38* (47), 8145–8148.
- (279) Lins, R. J.; Flitsch, S. L.; Turner, N. J.; Irving, E.; Brown, S. . Enzymatic Generation and In Situ Screening of a Dynamic Combinatorial Library of Sialic Acid Analogues. *Angew. Chem. Int. Ed.* **2002**, *41*, 3405–3407.
- (280) Lins, R. J.; Flitsch, S. L.; Turner, N. J.; Irving, E.; Brown, S. A. Generation of a Dynamic Combinatorial Library Using Sialic Acid Aldolase and in Situ Screening against Wheat

Germ Agglutinin. *Tetrahedron* **2004**, 60 (3), 771–780.

- (281) Elyashberg, M. Identification and Structure Elucidation by NMR Spectroscopy. *Trends Anal. Chem.* **2015**, 69, 88–97.
- (282) 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.
- (283) Hesse, Manfred Bernd Zeeh, H. M. *Spectroscopic Methods in Organic Chemistry*; 2008.
- (284) 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.
- (285) Wang, L.; Markley, J. L. Empirical Correlation between Protein Backbone ¹⁵N and ¹³C Secondary Chemical Shifts and Its Application to Nitrogen Chemical Shift Re-Referencing. *J. Biomol. NMR* **2009**, 44 (2), 95–99.
- (286) Malviya R, B. V.; Pal O.P.; Sharma P.K. High Performance Liquid Chromatography: A Short Review. *J. Glob. Pharma Technol.* **2010**, 2 (5), 22–26.
- (287) Hartman, A. M.; Gierse, R. M.; Hirsch, A. K. H. Protein-Templated Dynamic Combinatorial Chemistry: Brief Overview and Experimental Protocol. *European J. Org. Chem.* **2019**, 2019 (22), 3581–3590.
- (288) Pitt, J. J. Principles and Applications of Liquid Chromatography-Mass Spectrometry in Clinical Biochemistry. *Clin Biochem Rev* **2009**, 30 (1), 19–34.
- (289) 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.
- (290) Karas, M. Time-of-Flight Mass Spectrometer with Improved Resolution. *J. Mass Spectrom.* **1997**, *32* (1), 1–3.
- (291) Sanogo, S.; Silimbani, P.; Gaggeri, R.; Masini, C. Development and Validation of an HPLC-DAD Method for the Simultaneous Identification and Quantification of Topotecan, Irinotecan, Etoposide, Doxorubicin and Epirubicin. *Arab. J. Chem.* **2021**, *14* (1).
- (292) Wiseman, T.; Williston, S.; Brandts, J. F. Rapid Measurement of Binding Constants and Heats of Binding Using a New Titration Calorimeter. *Anal. Biochem.* **1989**, *179*, 131–137.
- (293) Freire, E.; Mayorga, O. L.; Straume, M. Isothermal Titration Calorimetry. *Anal. Chem.* **1990**, *62* (18), 950A-959A.
- (294) Doyle, M. L. Characterization of Binding Interactions by Isothermal Titration Calorimetry. *Curr. Opin. Biotechnol.* **1997**, *8* (1), 31–35.
- (295) Linton, B.; Hamilton, A. D. Calorimetric Investigation of Guanidinium-Carboxylate Interactions. *Tetrahedron* **1999**, *55* (19), 6027–6038.
- (296) Prozeller, D.; Morsbach, S.; Landfester, K. Isothermal Titration Calorimetry as a Complementary Method for Investigating Nanoparticle–Protein Interactions. *Nanoscale* **2019**, *11* (41), 19265–19273.
- (297) Turnbull, W. B.; Daranas, A. H. On the Value of c : Can Low Affinity Systems Be Studied by Isothermal Titration Calorimetry? *W. J. Am. Chem. Soc.* **2003**, *125* (12), 14859–

14866.

- (298) MicroCal, L. VP-ITC MicroCalorimeter User's Manual, MAU130030 REV. E. MicroCal, LLC: Northampton, MA.
- (299) Bhatnagar, R. S.; Gordon, J. I. Thermodynamic Studies of Myristoyl-CoA: Protein Nmyristoyltransferase Using Isothermal Titration Calorimetry. *Methods Enzym.* **1995**, *250*, 467–486.
- (300) Bundle, D. R.; Sigurskjold, B. W. Determination of Accurate Thermodynamics of Binding by Titration Microcalorimetry. *Neoglycoconjugates, Pt B Biomed. Appl.* **1994**, *247*, 288–305.
- (301) 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.
- (302) Quick Start: Isothermal Titration Calorimetry (ITC). *TA instruments*. 2016.
- (303) Fisher, H. F.; Singh, N. Calorimetric Methods for Interpreting Protein—Ligand Interactions. *Methods Enzym.* **1995**, *259*, 194–221.
- (304) Sigurskjold, B. W. Exact Analysis of Competition Ligand Binding by Displacement Isothermal Titration Calorimetry. *Anal. Biochem.* **2000**, *277* (2), 260–266.
- (305) Holdgate, G. Isothermal Titration Calorimetry and Differential Scanning Calorimetry. In *In Ligand-Macromolecular Interactions in Drug Discovery: Methods and Protocols*; Inc, E. H. P., Ed.; Totowa, 2010; pp 101–133.
- (306) Tellinghuisen, J. Isothermal Titration Calorimetry at Very Low C. *Anal. Biochem.* **2008**, *373*, 395–397.

- (307) Wadso, I. Isothermal Microcalorimetry for the Characterization of Interactions between Drugs and Biological Materials. *Thermochim. Acta* **1995**, *267*, 45–49.
- (308) Canal-Martin, A.; Perez-Fernandez, R. Protein-Directed Dynamic Combinatorial Chemistry: An Efficient Strategy in Drug Design. *ACS Omega* **2020**, *5* (41), 26307–26315.
- (309) Glasoe, P.; Lonf, F. Use of Glass Electrodes to Measure Acidities in Deuterium Oxide. *J. Phys. Chem.* **1960**, *64* (1), 188–190.
- (310) Lorand, J.; Edwards, J. Polyol Complexes and Structure of the Benzenboronate Ion. *J. Org. Chem.* **1959**, *24*, 769–774.
- (311) Godoy-Alcántar, 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.
- (312) Hamilton, A. Hydrogen Bonding in Biological and Artificial Molecular Recognition. *Adv. Supramol. Chem.* **1990**, *1*, 1.
- (313) Persch, E.; Dumele, O.; Diederich, F. Molecular Recognition in Chemical and Biological Systems. *Angew. Chemie - Int. Ed.* **2015**, No. 150, 3290–3327.
- (314) Canal-martín, A.; Sastre, J.; Sánchez-barrena, M. J.; Canales, A.; Baldominos, S.; Pascual, N.; Martínez-gonzález, L.; Molero, D.; Fernández-valle, M. E.; Sáez, E.; Blanco-gabella, P.; Gómez-rubio, E.; Martín-santamaría, S.; Sáiz, A.; Mansilla, A.; Cañada, F. J.; Jiménez-barbero, J.; Martínez, A.; Pérez-fernández, R. Insights into Real-Time Chemical Processes in a Calcium Sensor Protein-Directed Dynamic Library. *Nat. Commun.* **2019**, *10*, 2798.
- (315) Chen, Z.; Lv, Z.; Wang, X.; Yang, H.; Qing, G.; Sun, T. A Biomimetic Design for a

- Sialylated , Glycan-Specific Smart Polymer. *Nat. Publ. Gr.* **2018**, *10* (3), e472-13.
- (316) Barclay, T.; Ginic-markovic, M.; Johnston, M. R.; Cooper, P.; Petrovsky, N. Observation of the Keto Tautomer of D-Fructose in D2O Using H NMR Spectroscopy. *Carbohydr. Res.* **2012**, *347* (1), 136–141.
- (317) Inoue, K.; Kitahara, K.; Aikawa, Y.; Arai, S.; Masuda-hanada, T. HPLC Separation of All Aldopentoses and Aldohexoses on an Anion-Exchange Stationary Phase Prepared from Polystyrene-Based Copolymer and Diamine: The Effect of NaOH Eluent Concentration. *Molecules* **2011**, *16*, 5905–5915.
- (318) Doyle, M. Characterization Calorimetry of Binding Interactions by Isothermal Titration. *Curr. opiniion Biotechnol.* **1987**, *8*, 31–35.
- (319) Schumacher, S.; Katterle, M.; Hettrich, C.; Reiner, B.; Hall, D. G.; Scheller, F. W. Label - Free Detection of Enhanced Saccharide Binding at PH 7 . 4 to Nanoparticulate Benzoboroxole Based Receptor Units. *J. Mol. Recognit* **2011**, *24*, 953–959.
- (320) Chandramouli, N.; Ferrand, Y.; Lautrette, G.; Kauffmann, B.; Mackereth, C. D.; Laguerre, M.; Dubreuil, D.; Huc, I. Iterative Design of a Helically Folded Aromatic Oligoamide Sequence for the Selective Encapsulation of Fructose. *Nat. Chem.* **2015**, *7* (4), 334–341.
- (321) Zhai, W.; Male, L.; Fossey, J. S. Glucose Selective Bis-Boronic Acid Click-Fluor. *Chem. Commun.* **2017**, *53* (14), 2218.
- (322) Thordarson, P. Determining Association Constants from Titration Experiments in Supramolecular Chemistry. *Chem. Soc. Rev.* **2011**, *40*, 1305–1323.
- (323) Sijbesma, R. .; Nolte, R. J. . Binding of Dihydroxybenzenes in a Synthetic Molecular

- Clip : Effect of Hydrogen Bonding and Pi-Stacking. *J. Org. Chem.* **1991**, 56 (10), 3199–3201.
- (324) Macomber, R. S. An Introduction to NMR Titration for Studying Rapid Reversible Complexation. *J. Chem. Educ.* **1992**, 69 (5), 375–378.
- (325) Lopez-Macia, A.; Jime, J. C.; Royo, M.; Giralt, E.; Albericio, F. Synthesis and Structure Determination of Kahalalide F. **2001**, 123 (8), 11398–11401.
- (326) Crespo, L.; Montaner, B.; Pe, R.; Royo, M.; Pons, M.; Albericio, F.; Giralt, E. Peptide Dendrimers Based on Polyproline Helices. *J. Am. Chem. Soc.* **2002**, 124 (28), 8876–8883.
- (327) Cort, A.; Soriano, A.; Ventura, R.; Molero, A.; Hoen, R.; Casad, V.; Fanelli, F.; Albericio, F.; Lluí, C.; Franco, R.; Royo, M. Adenosine A_{2A} Receptor-Antagonist / Dopamine D₂ Receptor-Agonist Bivalent Ligands as Pharmacological Tools to Detect A_{2A}-D₂ Receptor Heteromers. *J. Med. Chem.* **2009**, 52, 5590–5602.
- (328) Systems, H. G.; Lagona, J.; Mukhopadhyay, P.; Chakrabarti, S.; Isaacs, L. The Cucurbit [n] Uril Family Angewandte. *Angew Chem Int Ed Engl.* **2005**, 44 (31), 4844–4870.
- (329) Masson, E.; Ling, X.; Joseph, R.; Kyeremeh-mensah, L.; Lu, X.; Masson, E. Cucurbituril Chemistry : A Tale of Supramolecular Success. *RSC Adv.* **2012**, 2, 1213–1247.
- (330) Tromans, R. A.; Carter, T. S.; Chabanne, L.; Crump, M. P.; Li, H.; Matlock, J. V; Orchard, M. G.; Davis, A. P. A Biomimetic Receptor for Glucose. *Nat. Chem.* **2019**, 11, 52–56.
- (331) Pochorovski, I.; Diederich, F. Fluorophore-Functionalized and Top-Covered Resorcin [4] Arene Cavitands. *Isr. J. Chem.* **2012**, 52, 20–29.
- (332) Pochorovski, I.; Diederich, F. Development of Redox-Switchable Resorcin [4] Arene

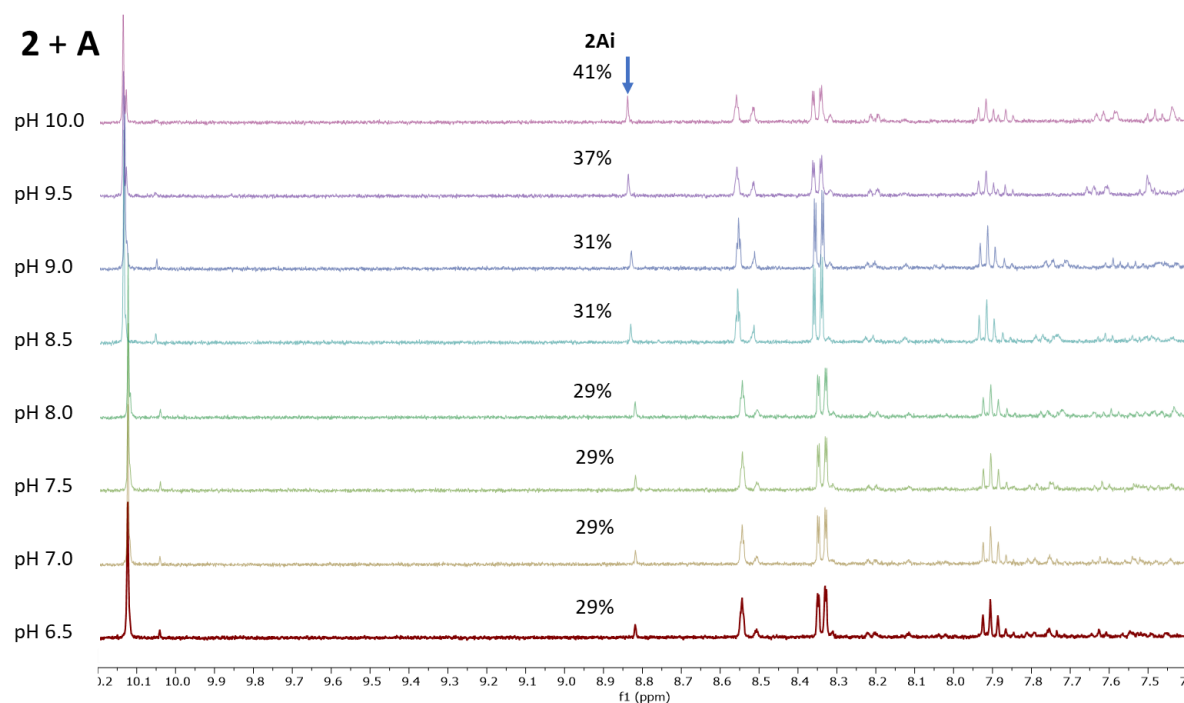
Cavitands. *Acc. Chem. Res.* **2014**, *47* (7), 2096–2105.

- (333) Gottschalk, T.; Jaun, B.; Diederich, F. Container Molecules with Portals : Reversibly Switchable Cycloalkane. *Angew Chem Int Ed Engl.* **2007**, *46*, 260–264.
- (334) Gottschalk, T.; Jarowski, P. D.; Diederich, F. Reversibly Controllable Guest Binding in Precisely Defined Cavities : Selectivity , Induced Fit , and Switching in Novel Resorcin [4] Arene-Based Container Molecules. *Tetrahedron* **2008**, *64*, 8307–8317.
- (335) Forrey, C.; Douglasb, J. F.; Gilsonc, M. K. The Fundamental Role of Flexibility on the Strength of Molecular Binding. *Soft Matter* **2012**, *8* (23), 6385–6392.
- (336) Dessolin, M.; Loffet, A. New Allyl Group Acceptors for Palladium Catalyzed Removal of Allylic Protections and Transacylation of Allyl Carbamates. *Tetrahedron Lett.* **1995**, *36* (32), 5741–5744.
- (337) Bacsá, B.; Horváti, K.; Bosze, S.; Andreae, F.; Kappe, C. O. Solid-Phase Synthesis of Difficult Peptide Sequences at Elevated Temperatures : A Critical Comparison of Microwave and Conventional Heating Technologies. *J. Org. Chem.* **2008**, *73*, 7532–7542.
- (338) Zhang, J.; Grandi, E.; Fu, H.; Saravanan, T.; Bothof, L.; Tepper, P. G.; Thunnissen, A. W. H.; Poelarends, G. J. Engineered C – N Lyase : Enantioselective Synthesis of Chiral Synthons for Artificial Dipeptide Sweeteners. *Angew. Chem. Int. Ed.* **2020**, *59*, 429–435.
- (339) Bermejo-Lopez, A.; Martín-Matute, B. Selective and Quantitative Functionalization of Unprotected α -Amino Acids Using a Recyclable Homogeneous Catalyst. *Chem. Commun.* **2022**, *8*, 3302–3323.

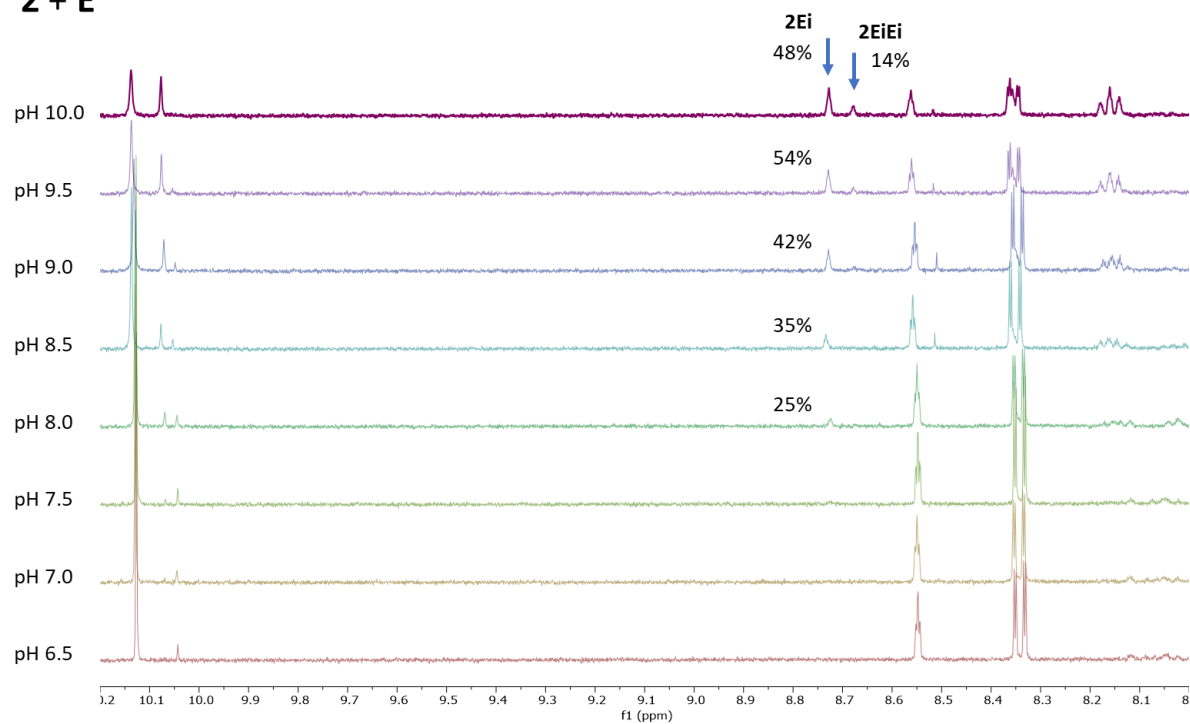
APPENDIX

Annex 1

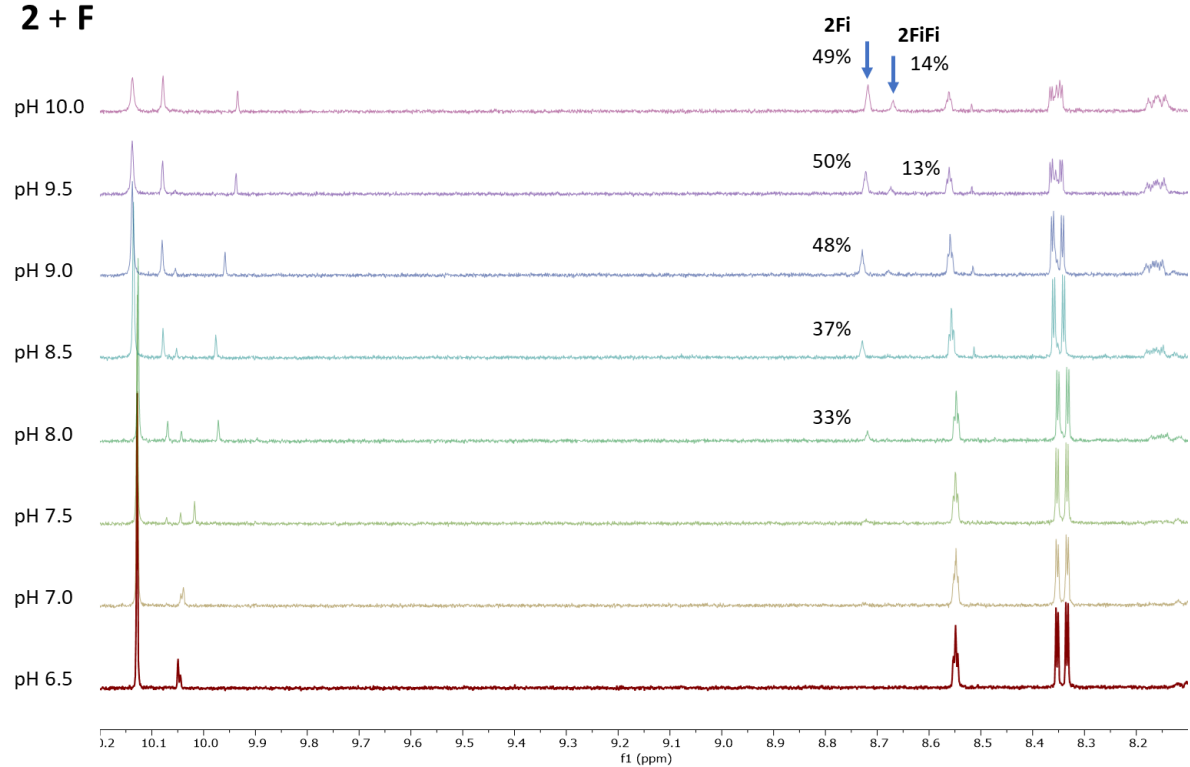
*Plot of ^1H NMR spectrums corresponding to reaction between aldehyde **2** and the amines tested in the section dedicated to the study of the influence of pH in IFR, at different pH from 6.5 to 10. Only the combinations that afforded imines are represented. Iminic proton (8-9 ppm) integrated to calculate the percentage of the monosubstituted (2Xi) and disubstituted (2XiXi) formed.*



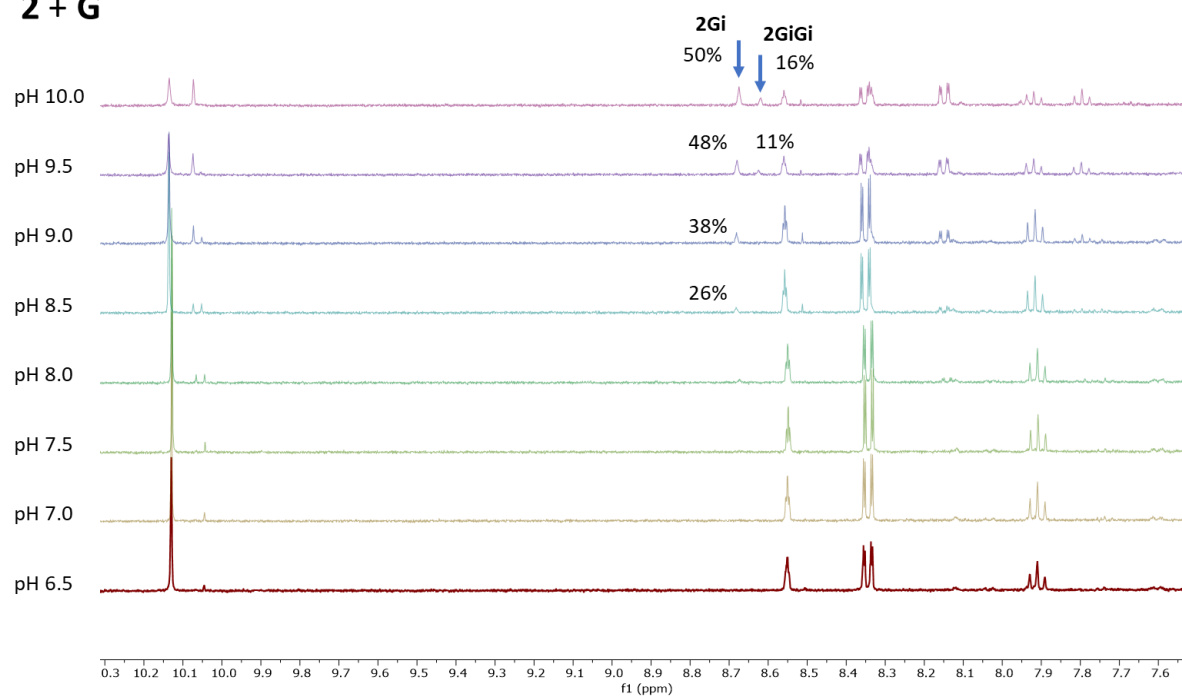
2 + E



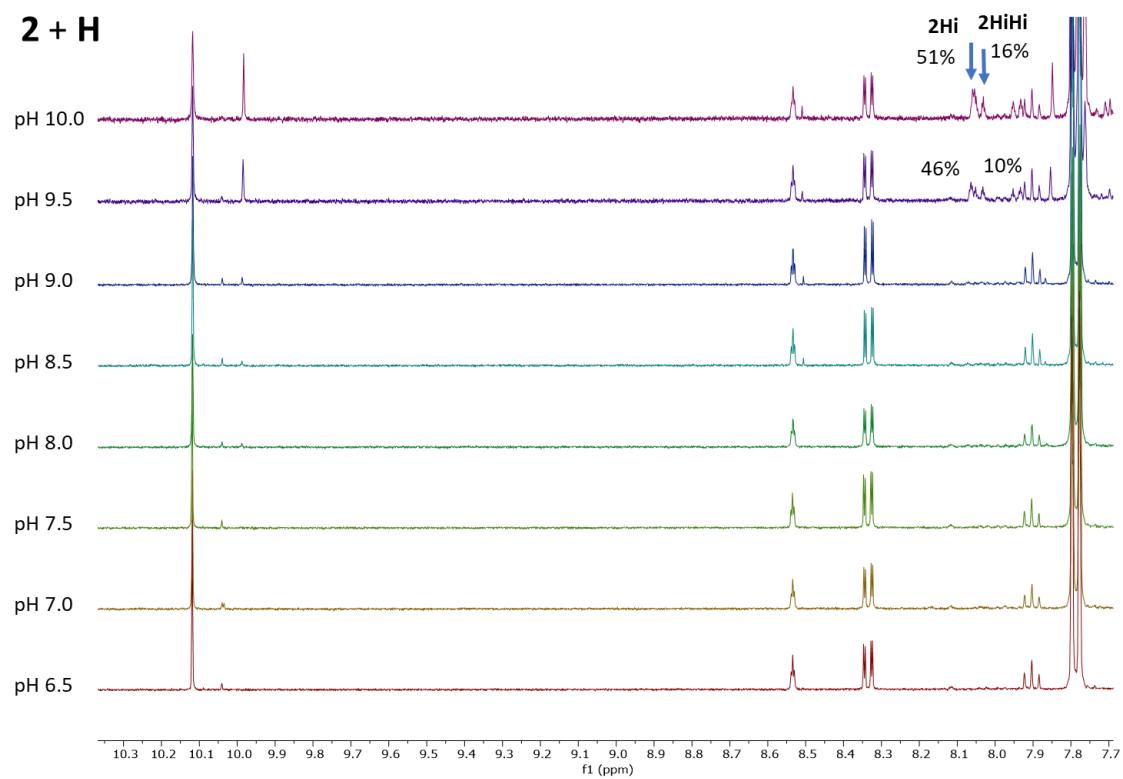
2 + F

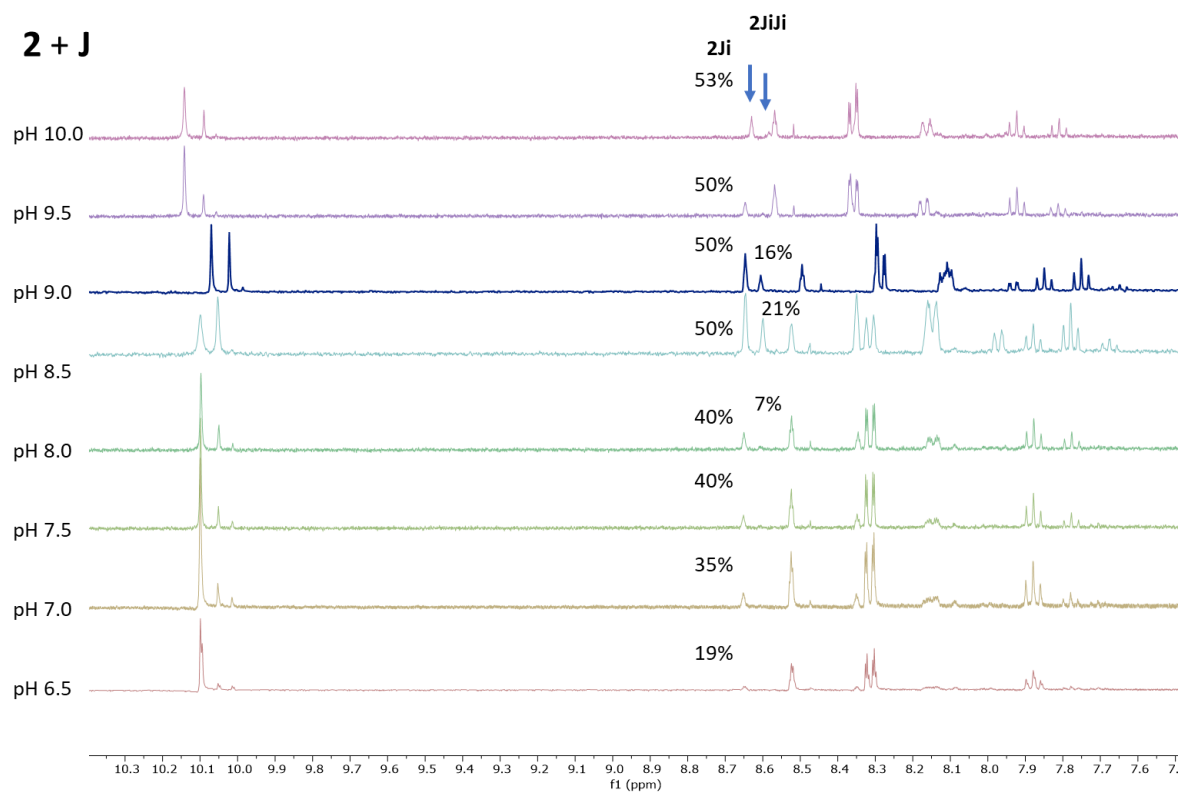
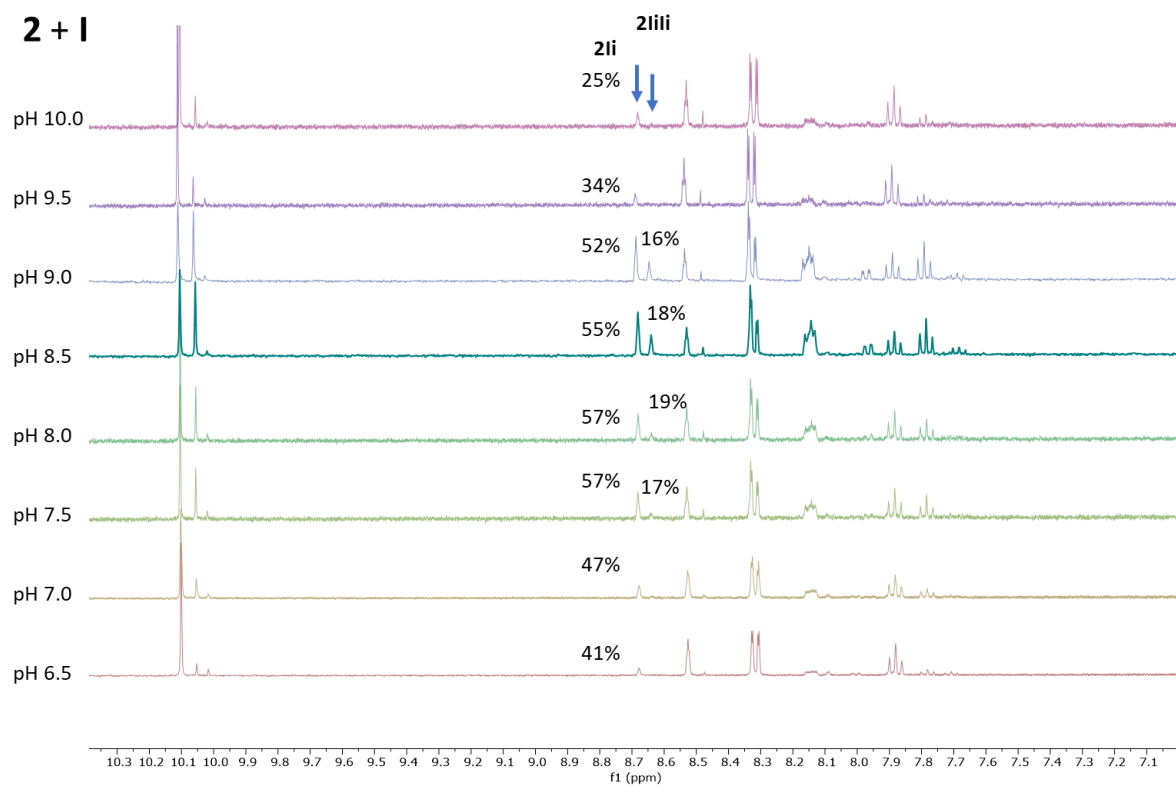


2 + G



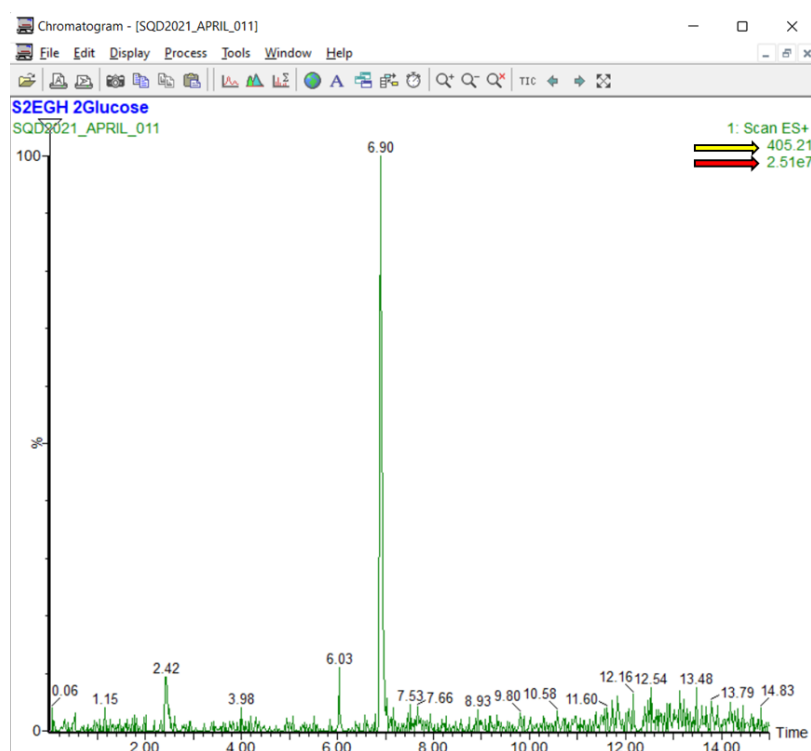
2 + H





Annex 2

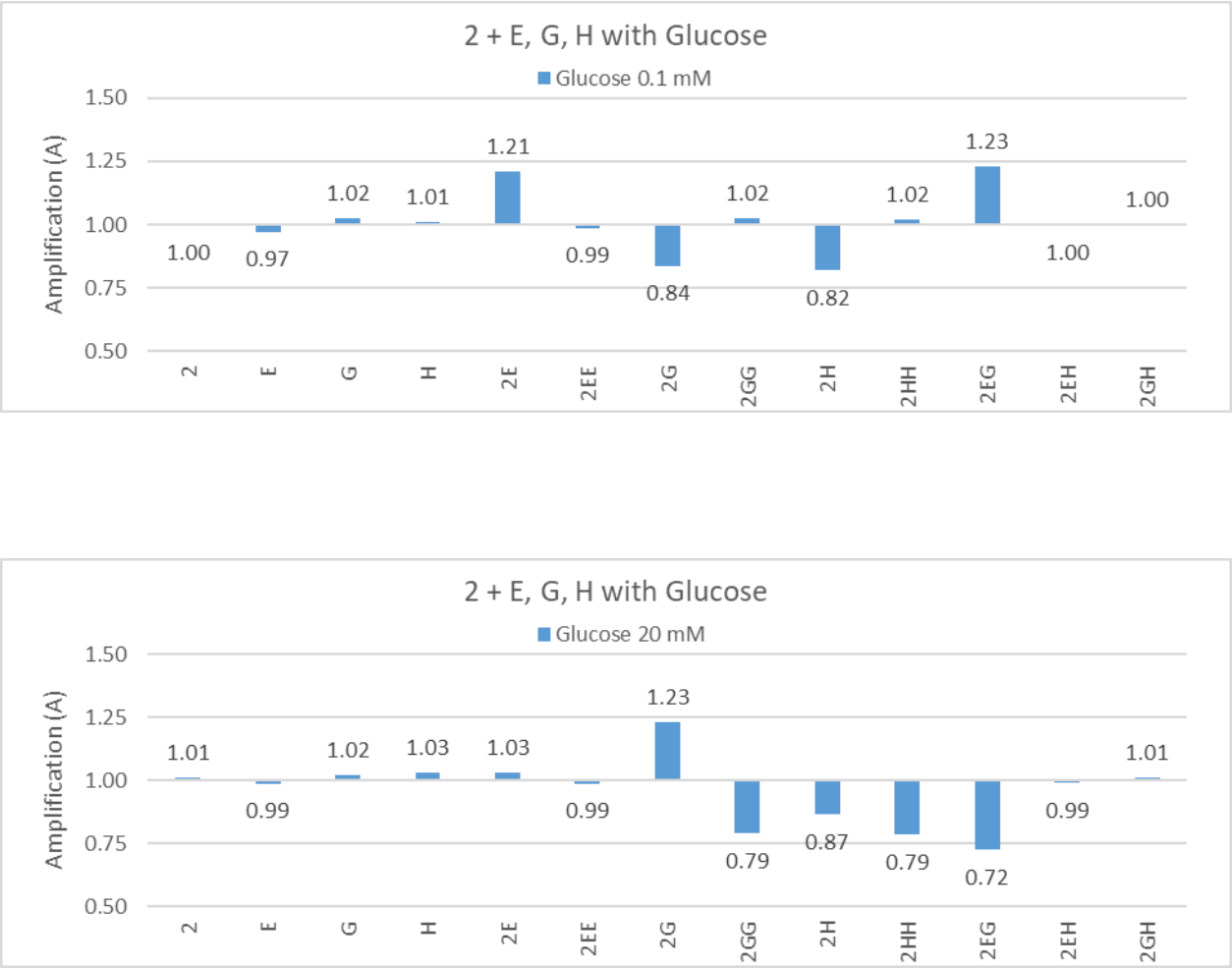
The m/z of each one of the library members (405.21 in the case of S2EE) was searched for in the MS chromatogram, and the area behind the peak, or number of ions with that m/z ($2.51e7$ in this example) was plotted in a chart. The division between the peak areas in both the Glucose and the blank experiments resulted in the Amplification values (A, 2.54 in the example) that was plotted and represented in a bar graph (Figure 44).



DCL member	Mw	m/z	number of ions detected		Amplification (A)
			Blank Experiment	2 mM Glucose Experiment	
S2	134.04	135.04	1.83E+07	1.81E+07	0.99
SE	151.08	152.08	3.43E+08	2.84E+08	0.83
SG	123.07	124.07	2.85E+08	3.01E+08	1.06
SH	204.09	205.09	2.31E+08	2.46E+08	1.06
S2E	269.12	270.12	1.62E+07	2.62E+07	1.61
S2EE	404.21	405.21	9.90E+06	2.51E+07	2.54
S2G	241.11	242.11	2.82E+07	4.02E+07	1.42
S2GG	348.18	349.18	2.28E+06	2.84E+06	1.24
S2H	322.13	323.13	1.01E+07	6.56E+06	0.65
S2HH	510.23	511.23	8.13E+06	8.05E+06	0.99
S2EG	376.2	377.2	1.90E+06	2.96E+06	1.56
S2EH	457.22	458.22	6.01E+05	5.93E+05	0.99
S2GH	429.21	430.21	1.67E+06	1.90E+06	1.14

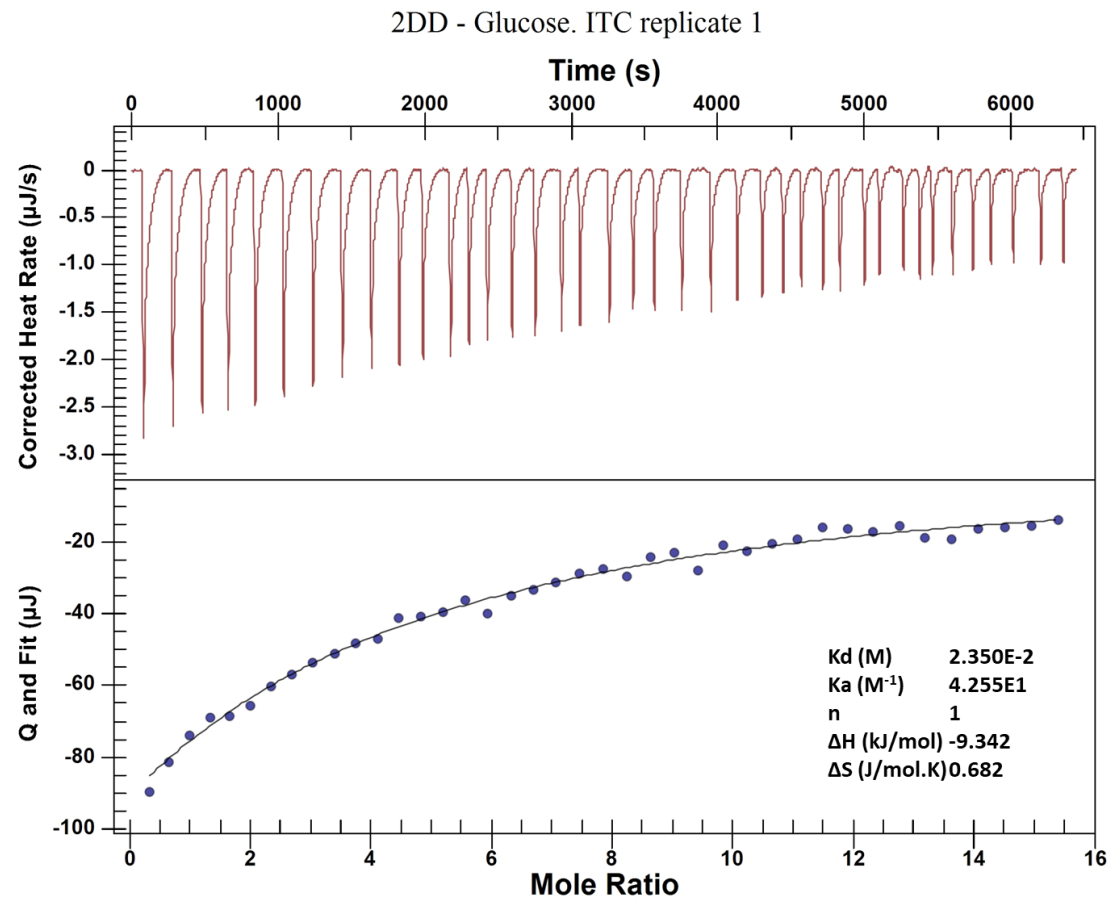
Annex 3

Amplification values for the DCL formed with 2, E, G, H, and Glucose 0.1 mM (top) and Glucose 20 mM (bottom) as template.

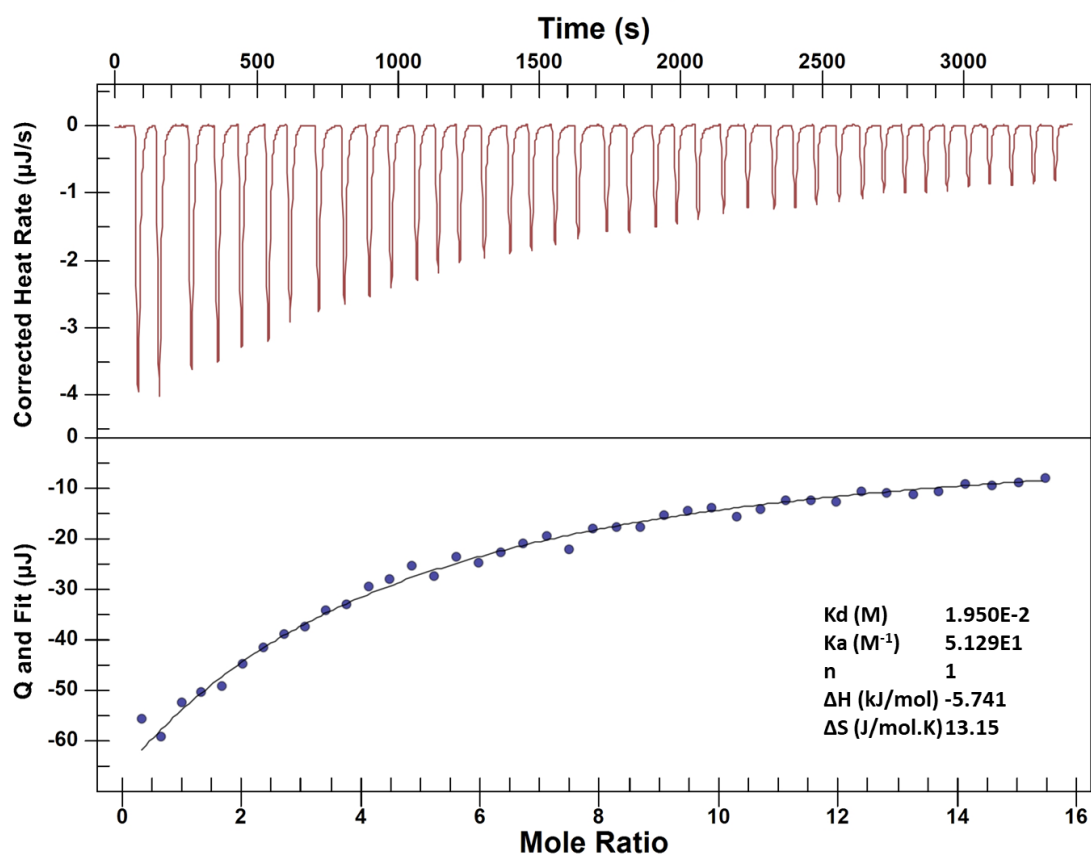


Annex 4

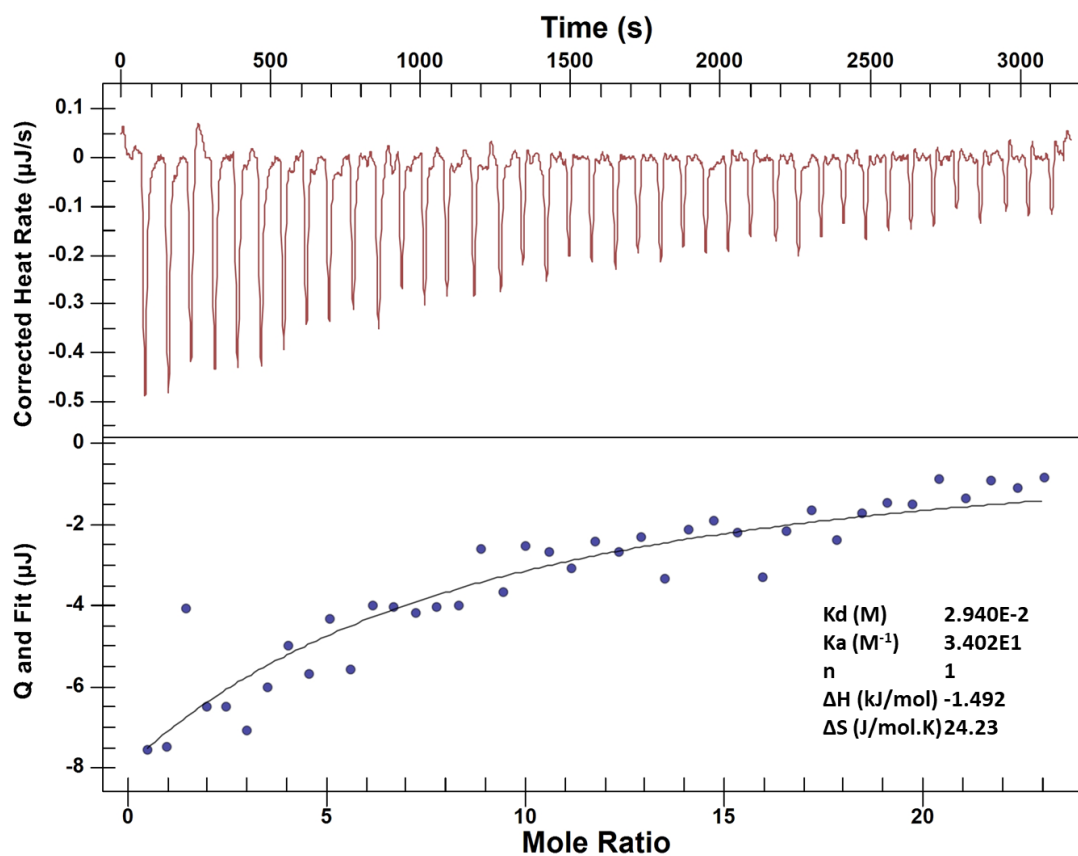
ITC graphs for all the combinations receptor-sugar tested that afforded a readable binding constant.



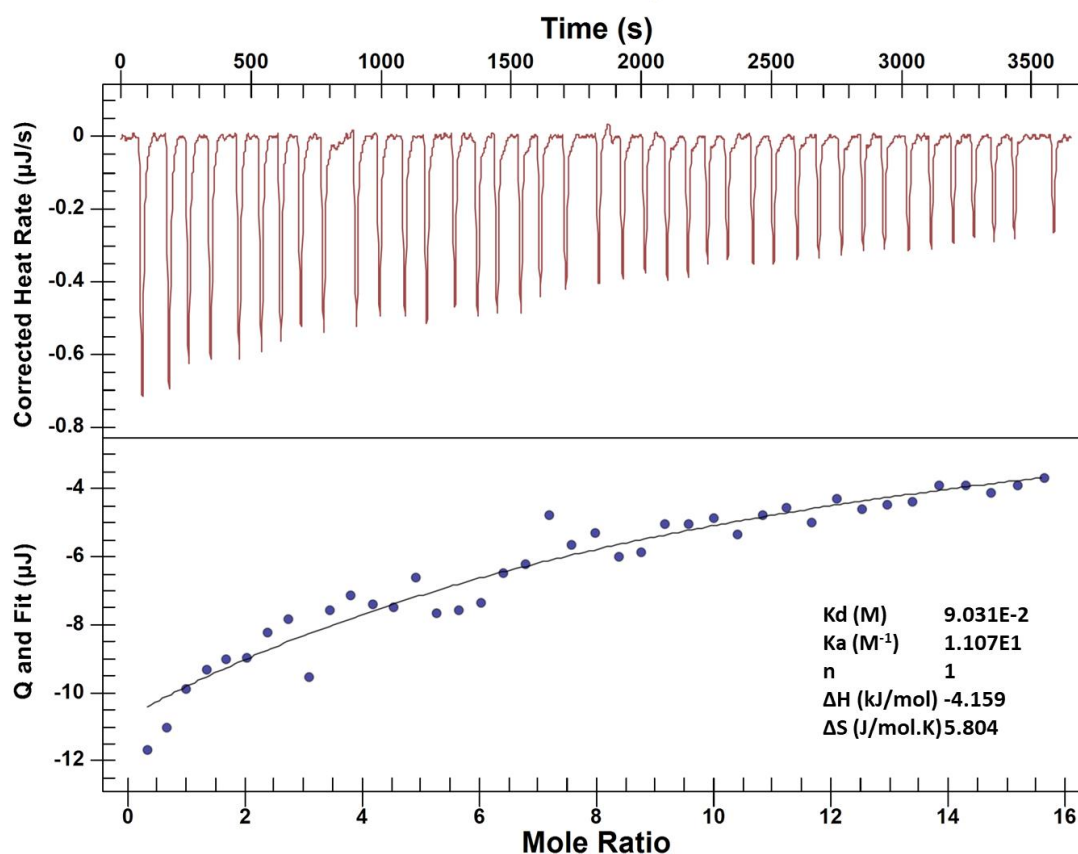
2DD - Mannose. ITC replicate 1



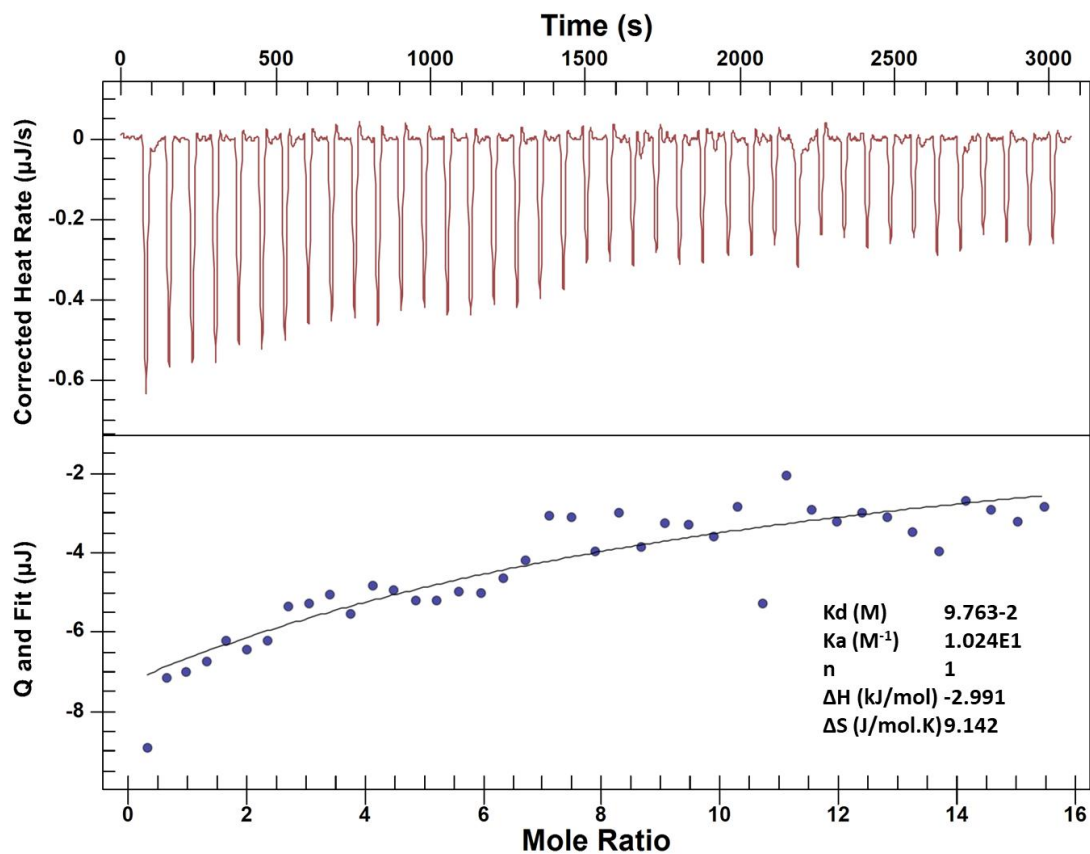
1P - Mannose. ITC replicate 1



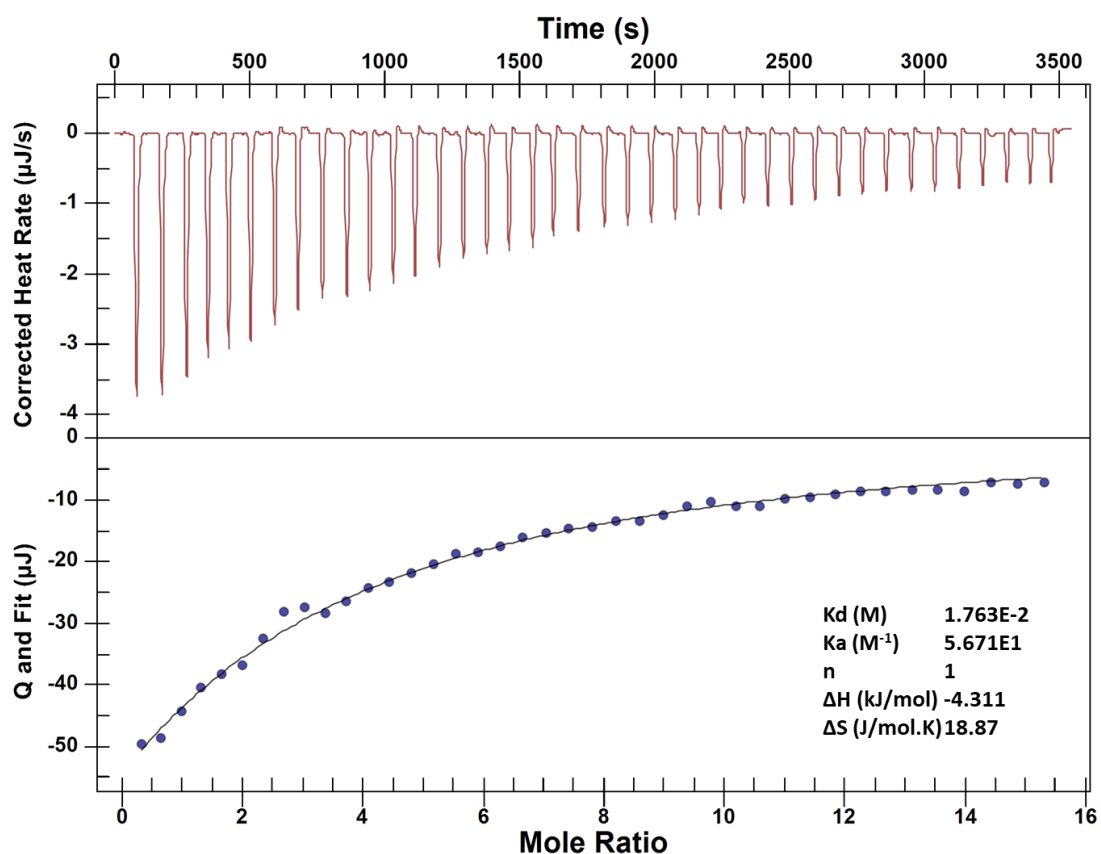
1D - Mannose. ITC replicate 1



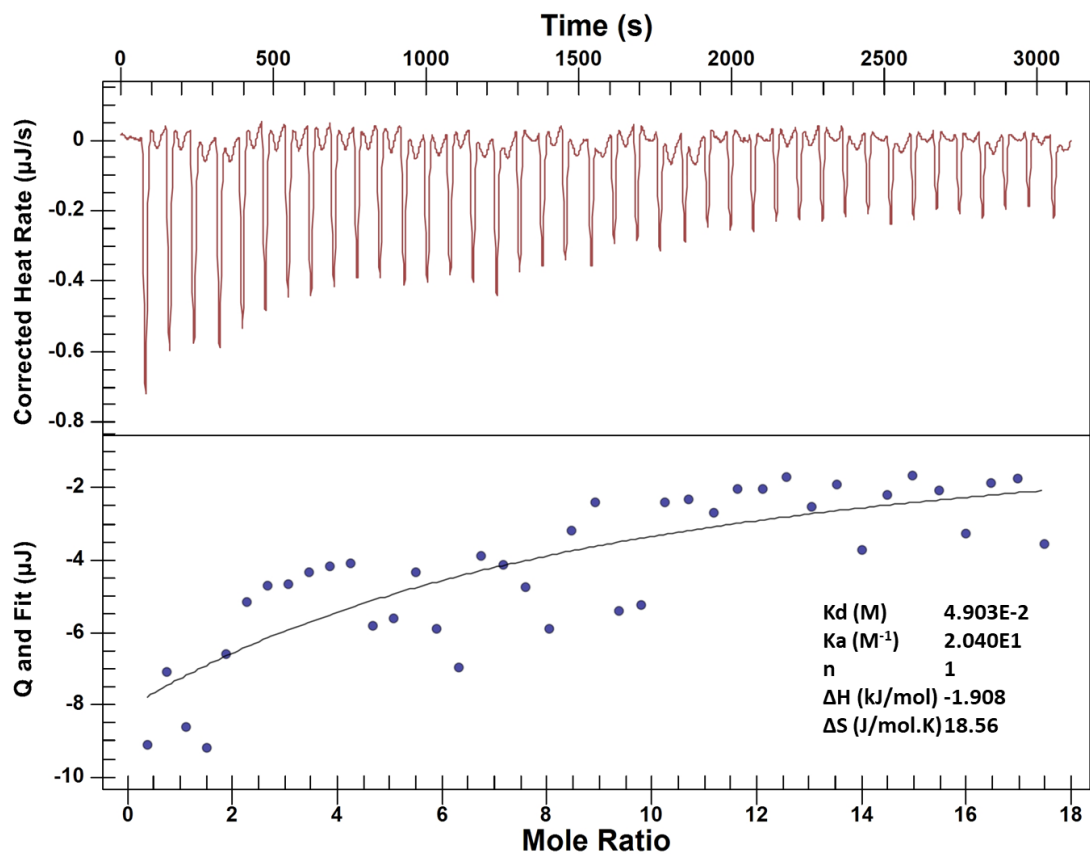
D - Mannose. ITC replicate 1



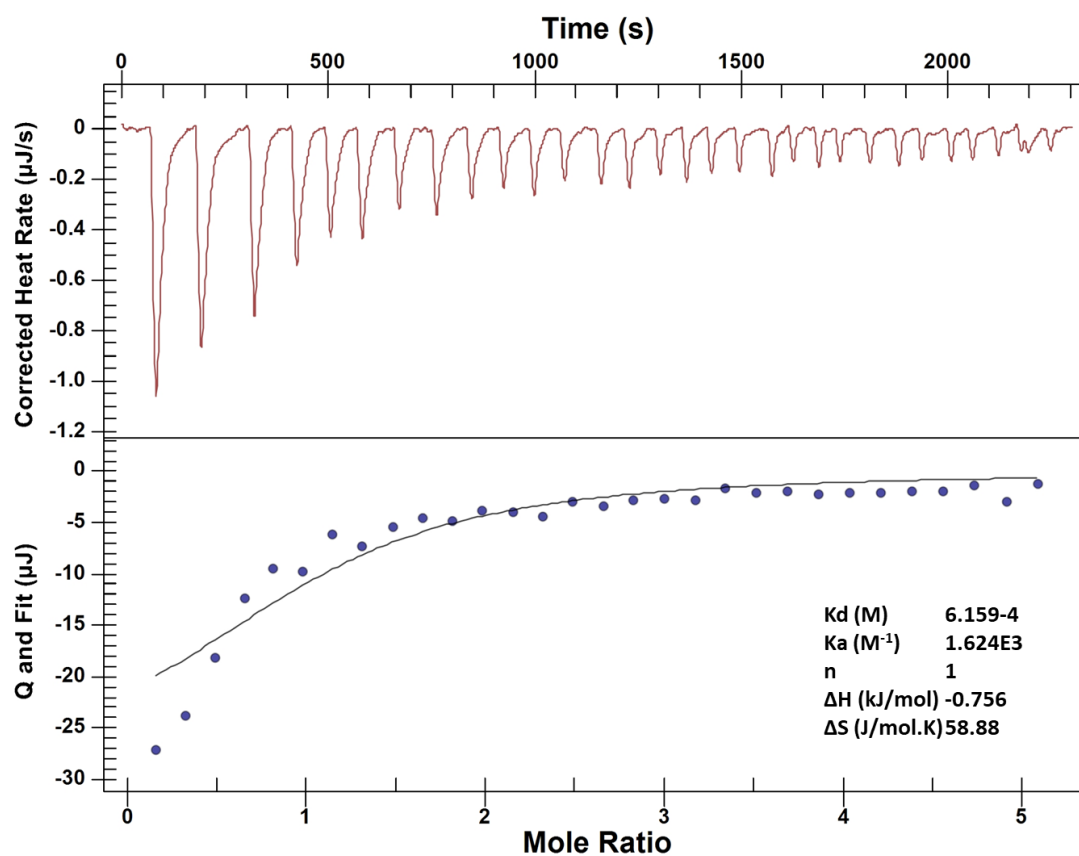
2DD - Galactose. ITC replicate 1



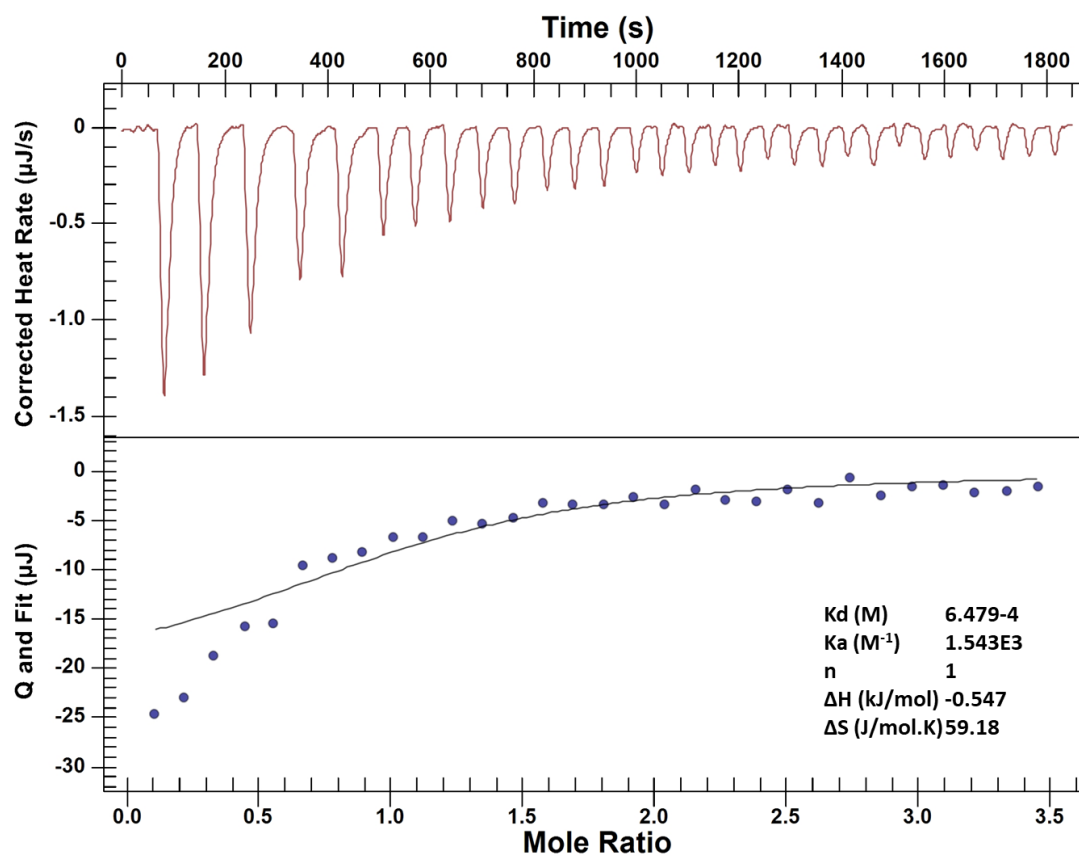
1P - Galactose. ITC replicate 1



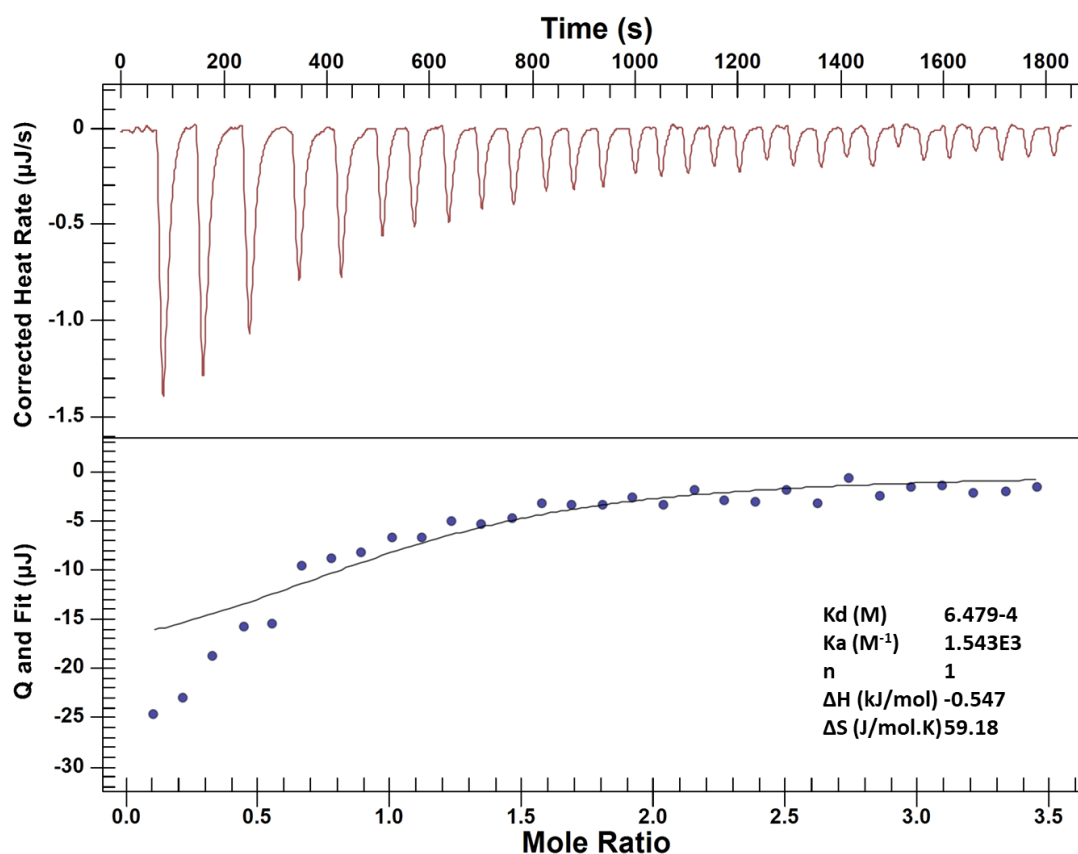
1P - Fructose. ITC replicate 1



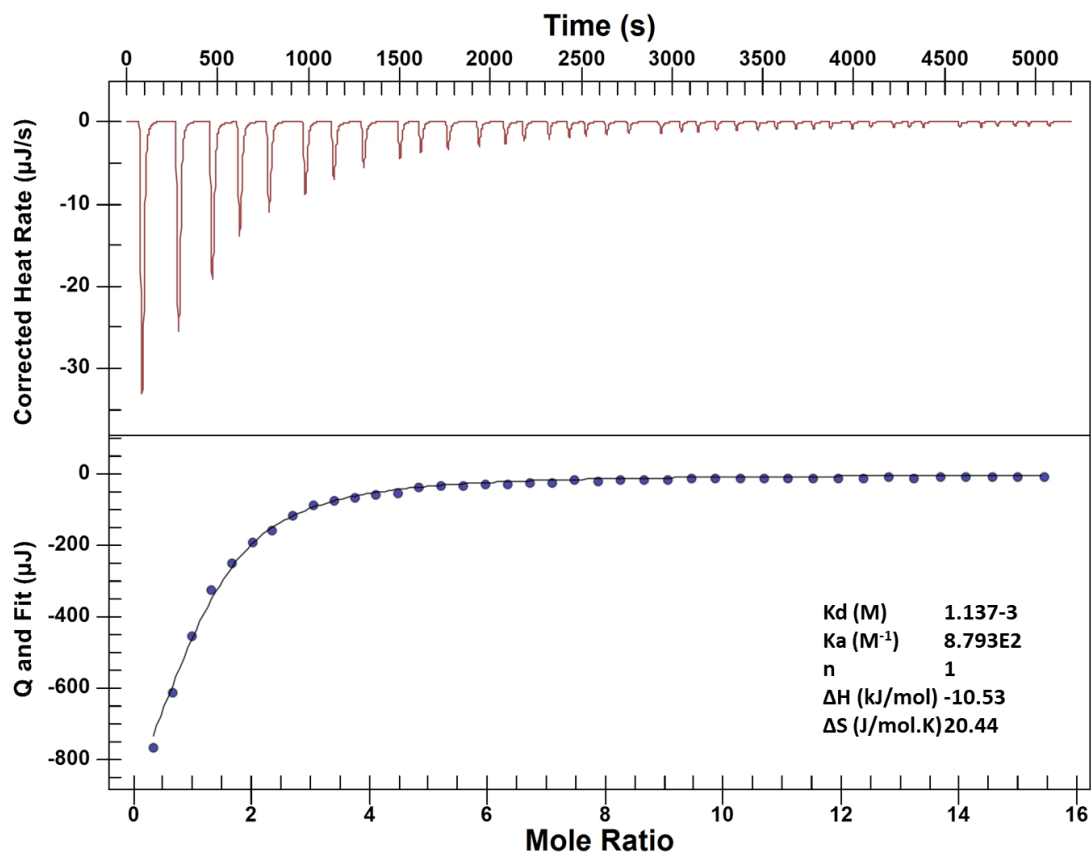
1D - Fructose. ITC replicate 1



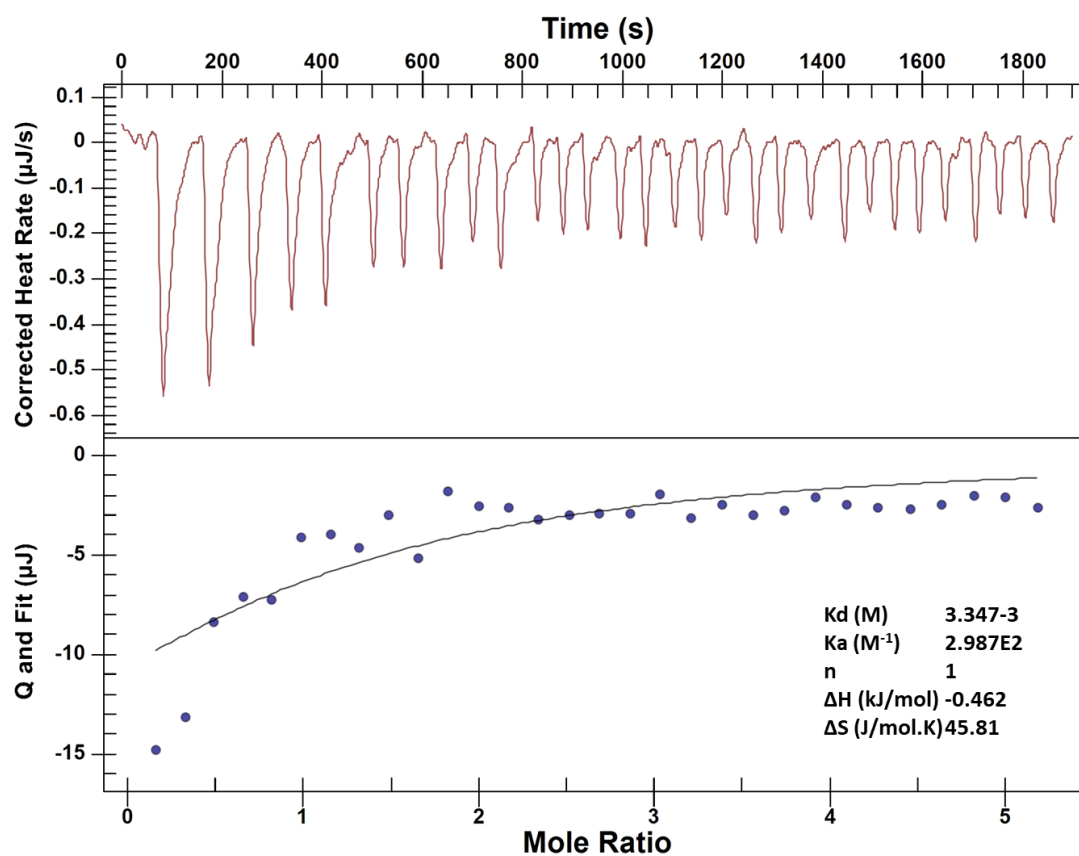
1D - Fructose. ITC replicate 1



2DD - Fructose. ITC replicate 1



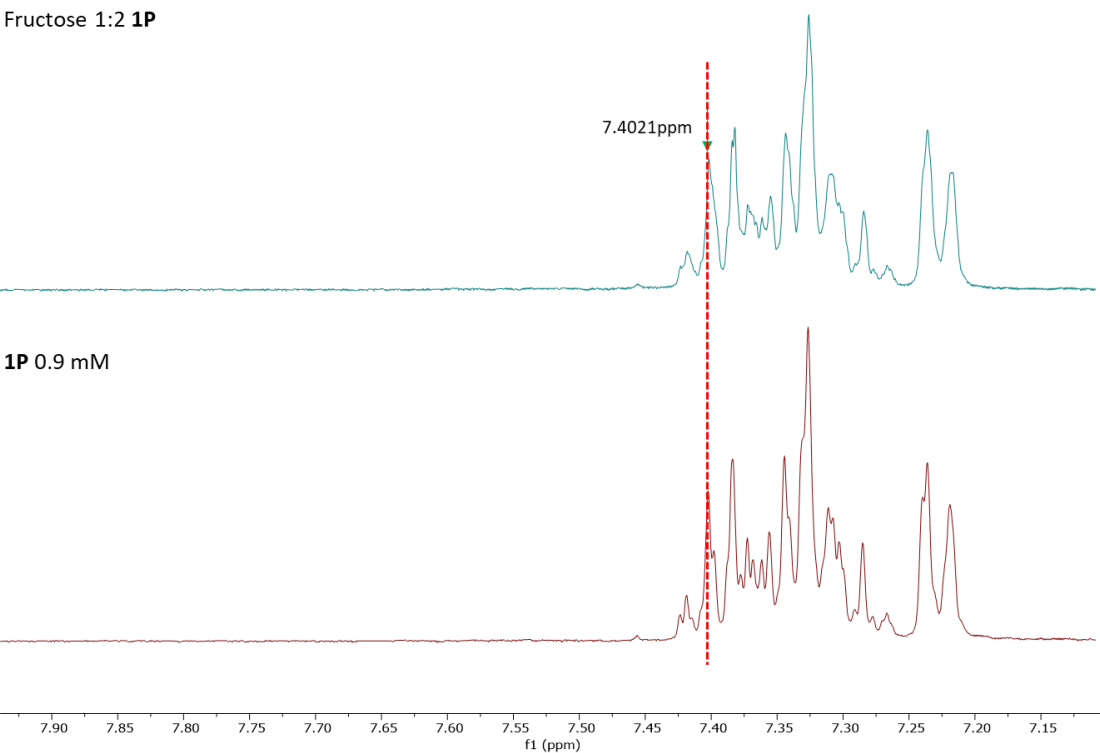
2PP - Fructose. ITC replicate 1



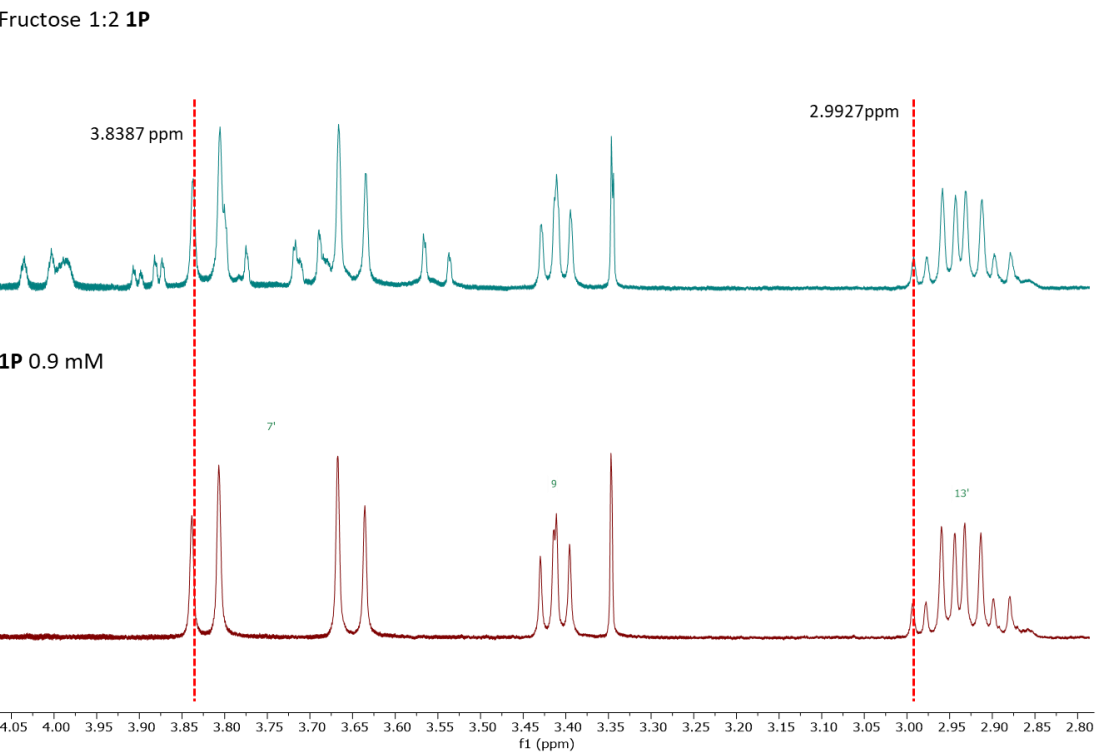
Annex 5

Titration 4 experiment ^1H NMR spectrum.

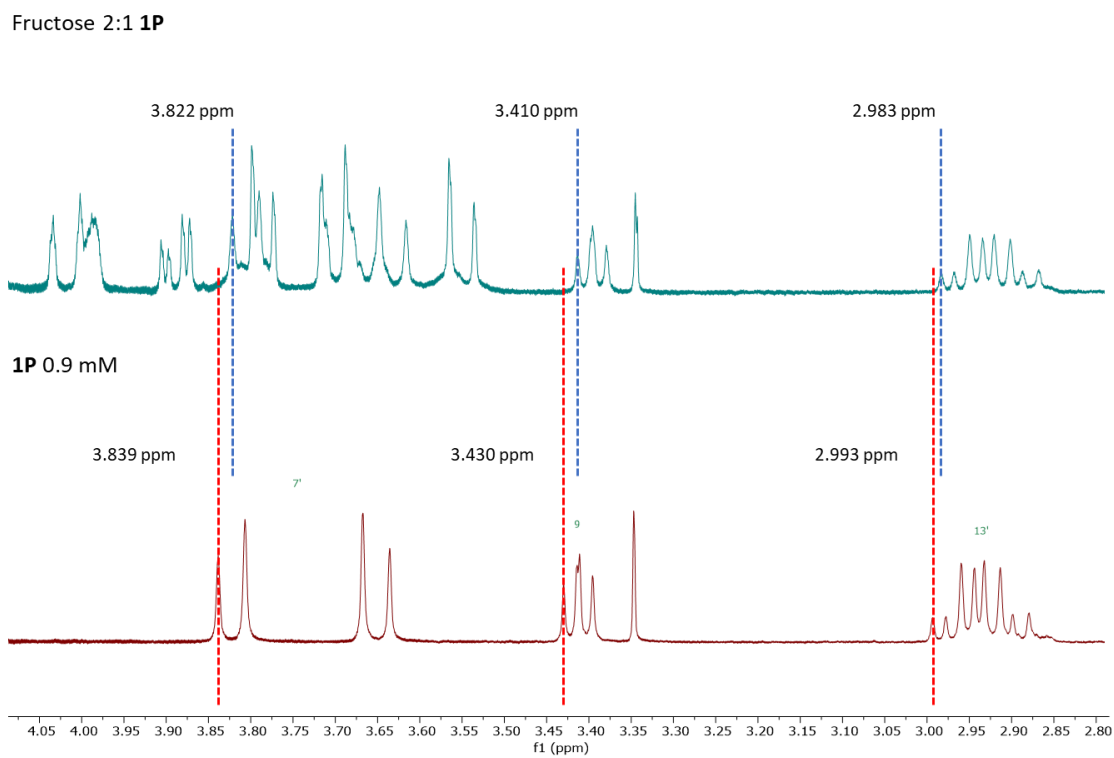
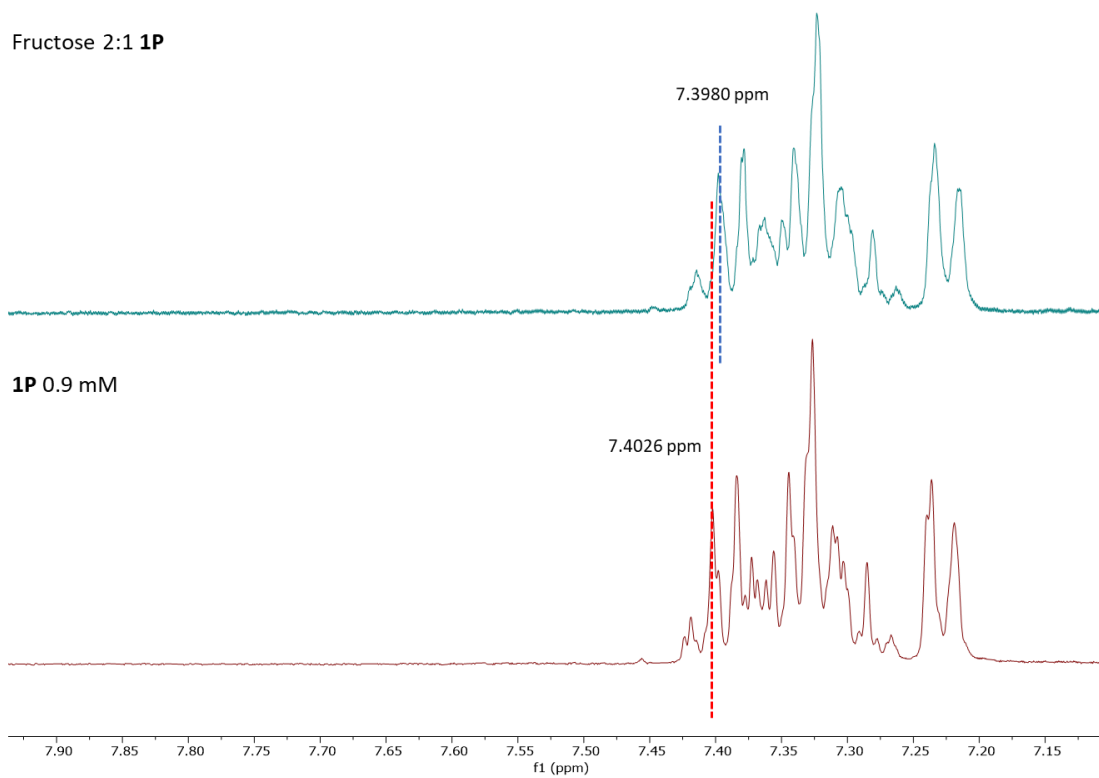
Fructose 1:2 **1P**



Fructose 1:2 **1P**

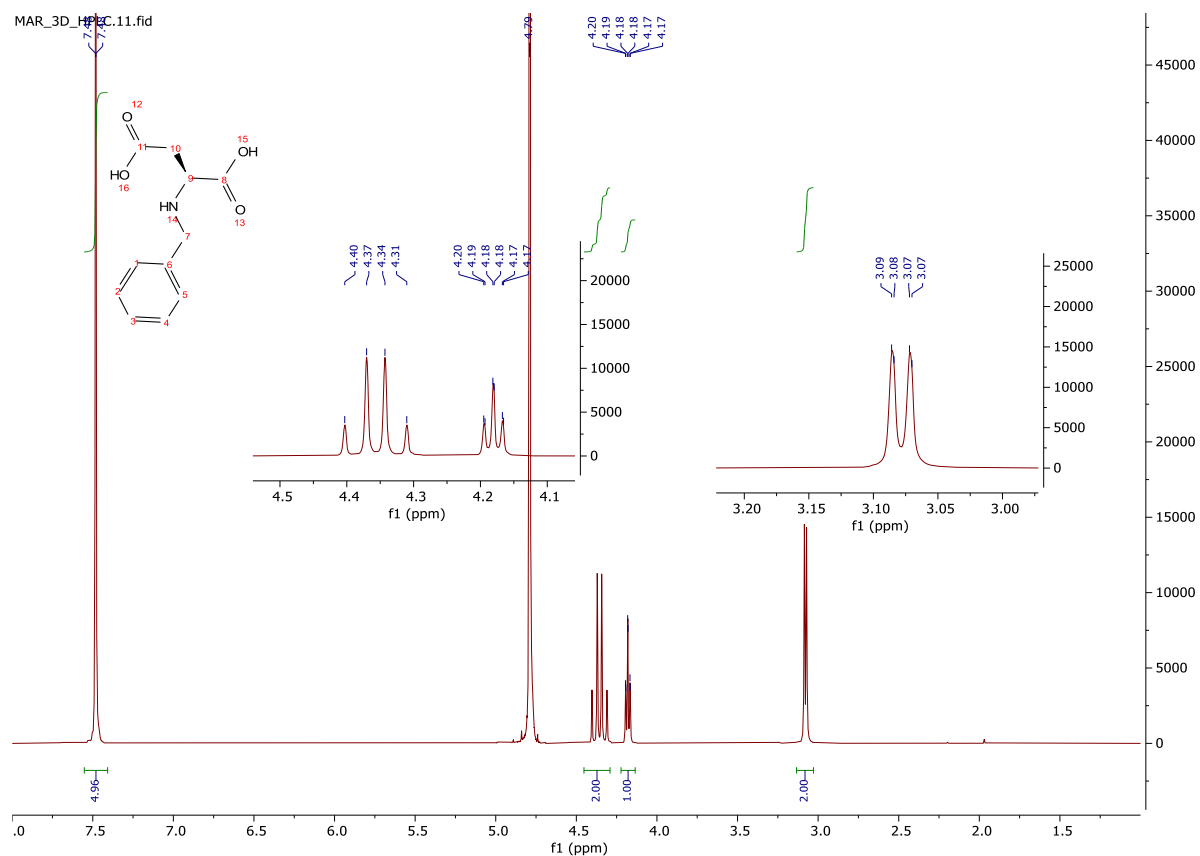


Titration 5 experiment ^1H NMR spectrum.

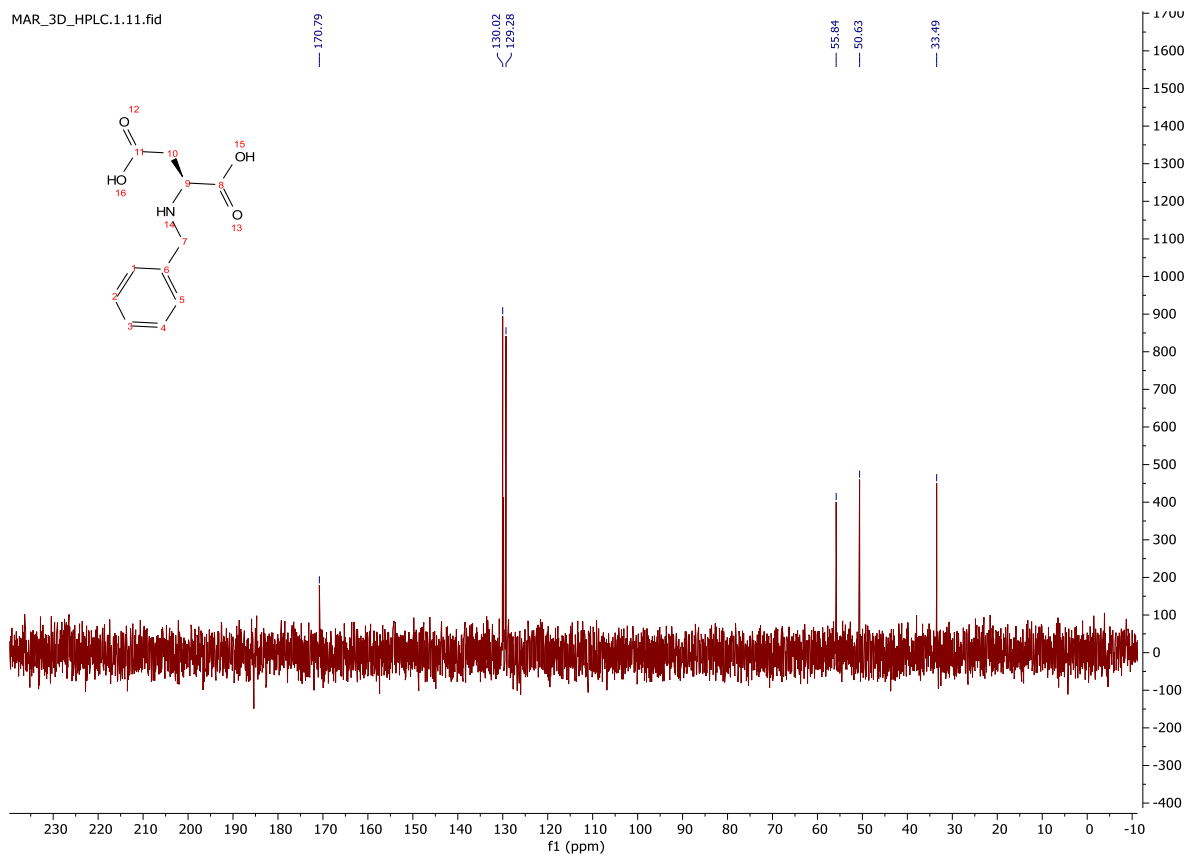


Annex 6

Characterisation of molecule **1D**: first ^1H NMR spectrum (400 MHz, D_2O , 298 K), then ^{13}C NMR spectrum (400 MHz, D_2O , 298 K) (next page, top), and MS (next page, bottom).



MAR_3D_HPLC.1.11.fid



MAR_3D C₁₁H₁₃NO₄ MW=223

(Water)/CH₃CN:H₂O:0.1% Formic Acid

PMM-MXA-9ELUF-nESI-Pos-1 40 (1.468) AM2 (Ar,18000.0,0.00,0.00); ABS

University of Birmingham, School of Chemistry

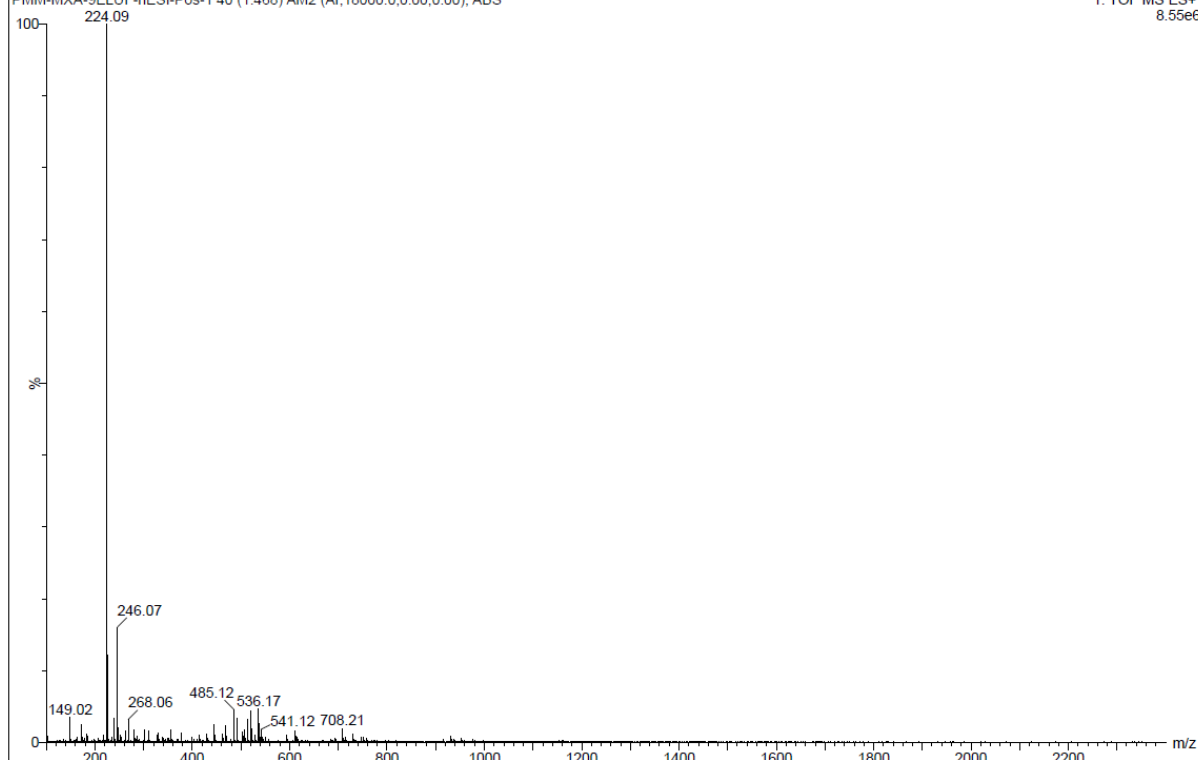
Waters Synapt G2-S

Miguel Alena-Rodriguez

22-Sep-2022

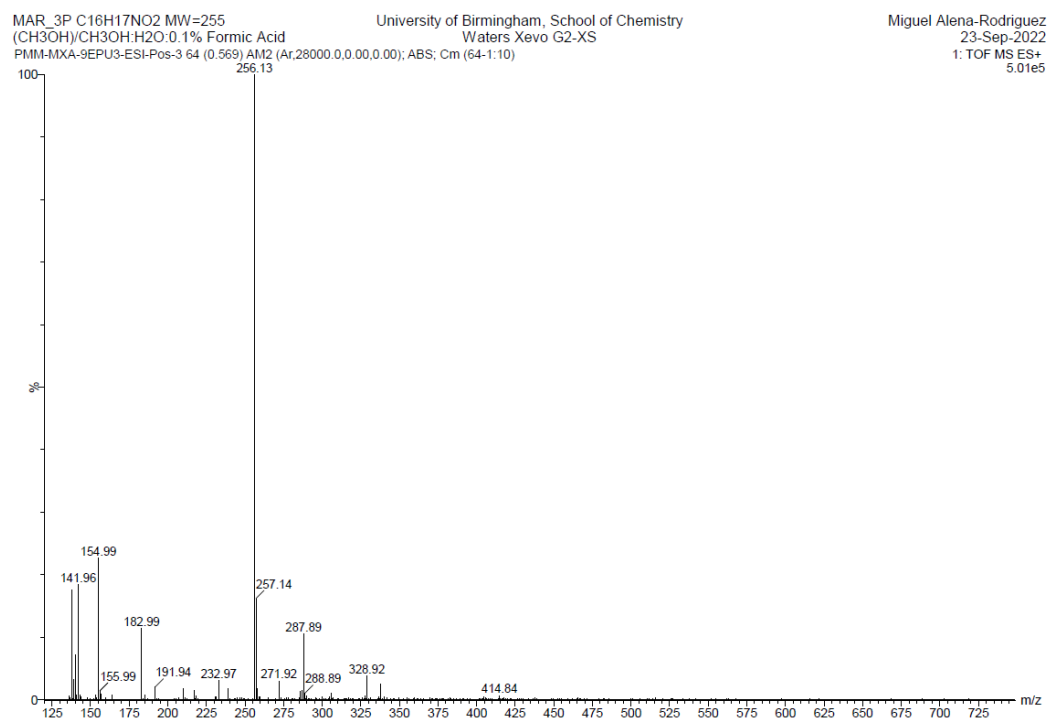
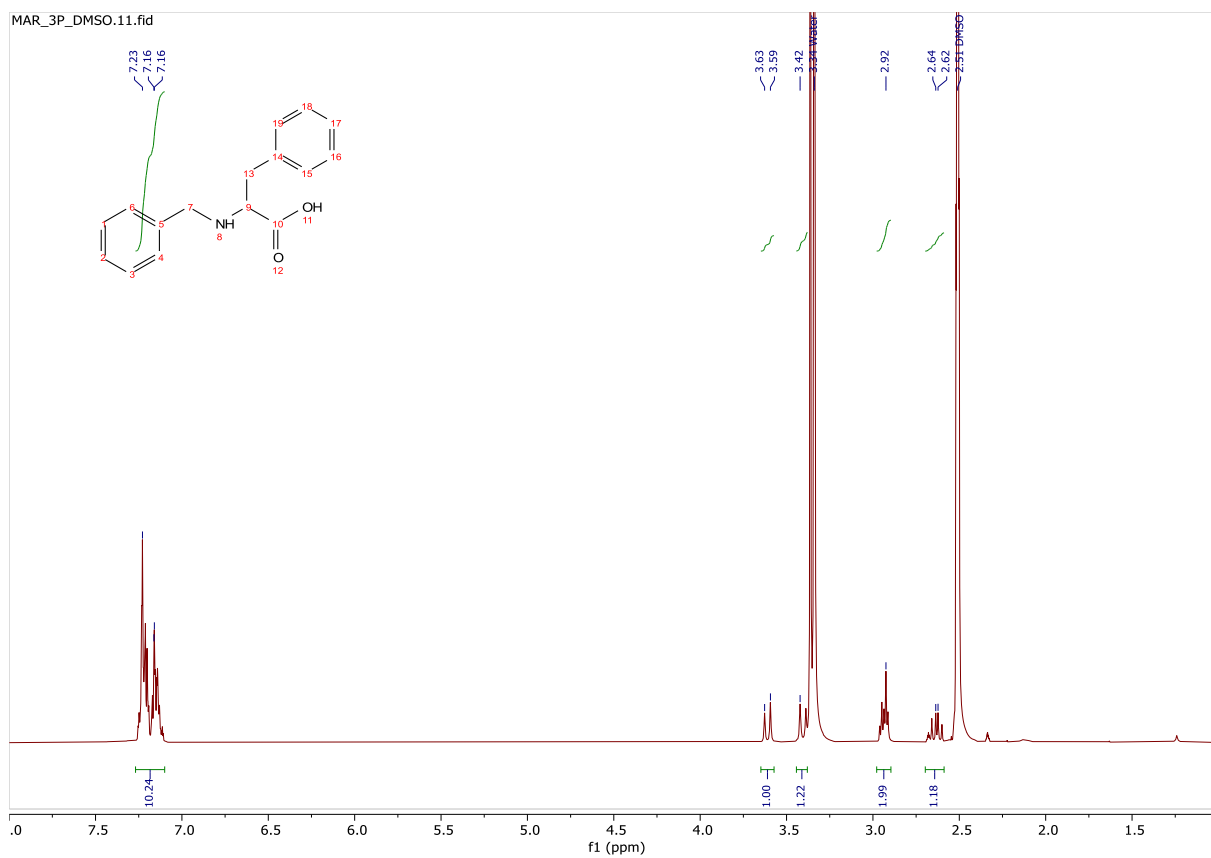
1: TOF MS ES+

8.55e6



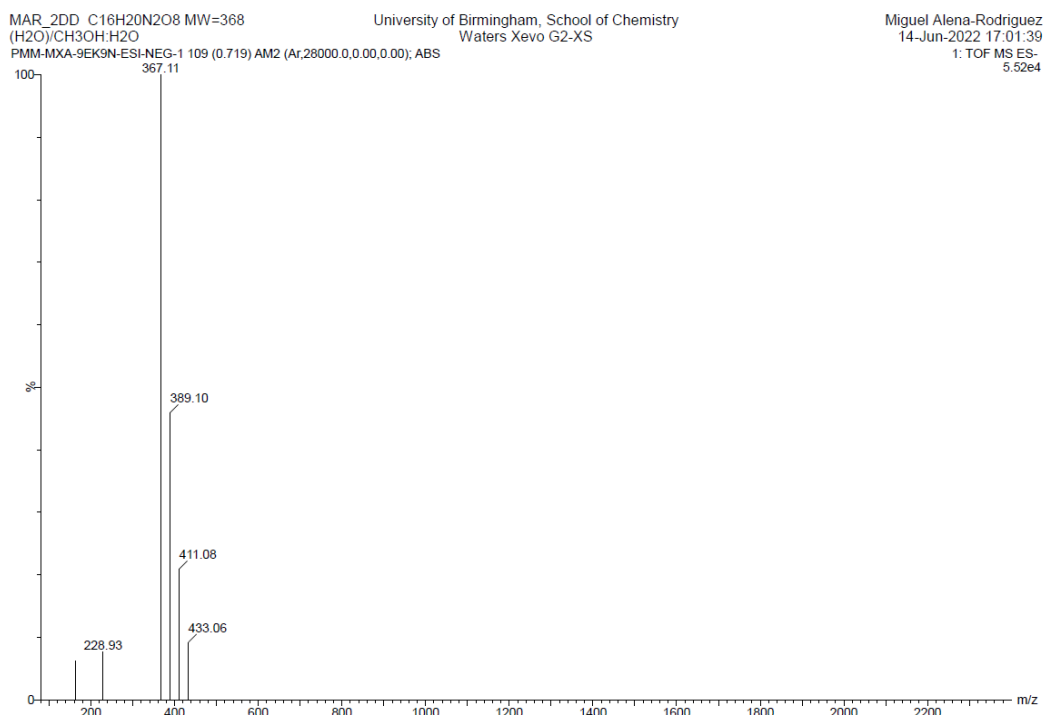
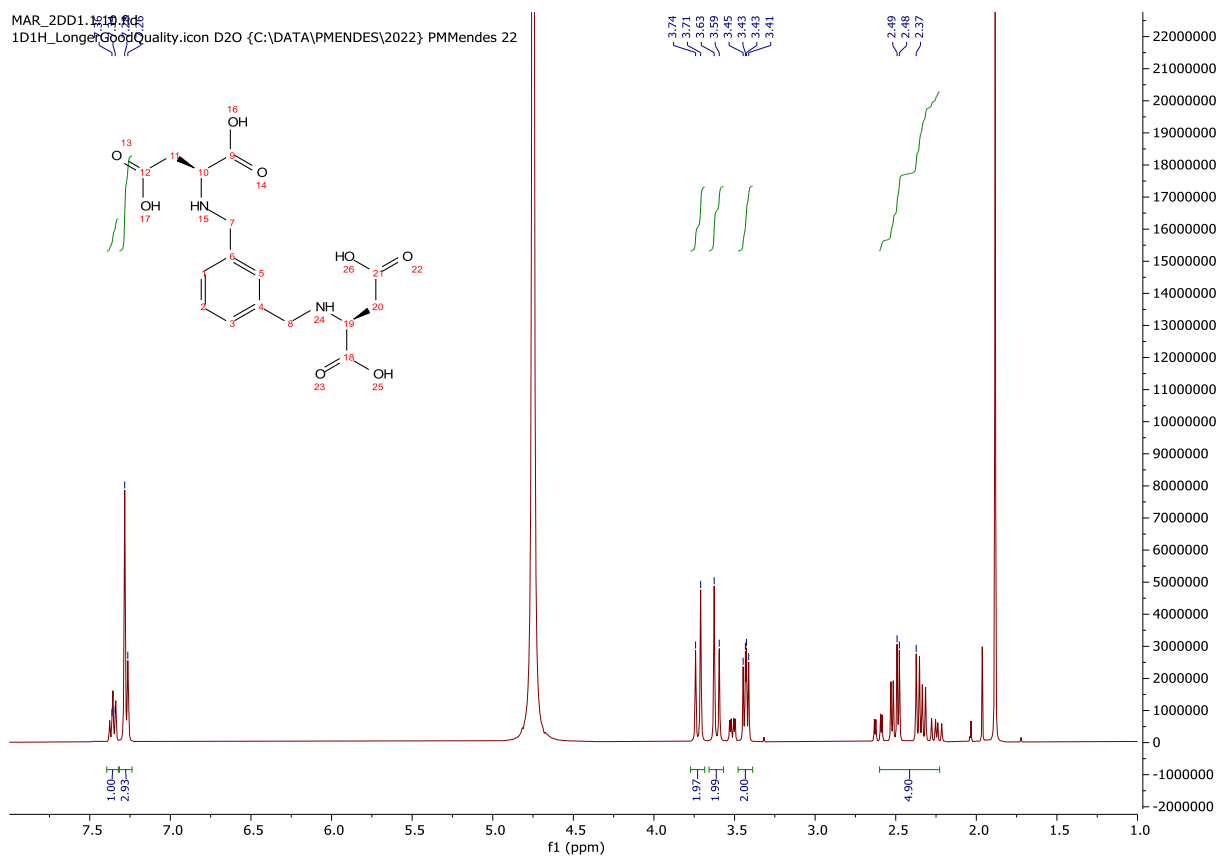
Annex 7

Characterisation of molecule **1P**: ^1H NMR spectrum (400 MHz, DMSO, 298 K) (top), and MS (bottom).



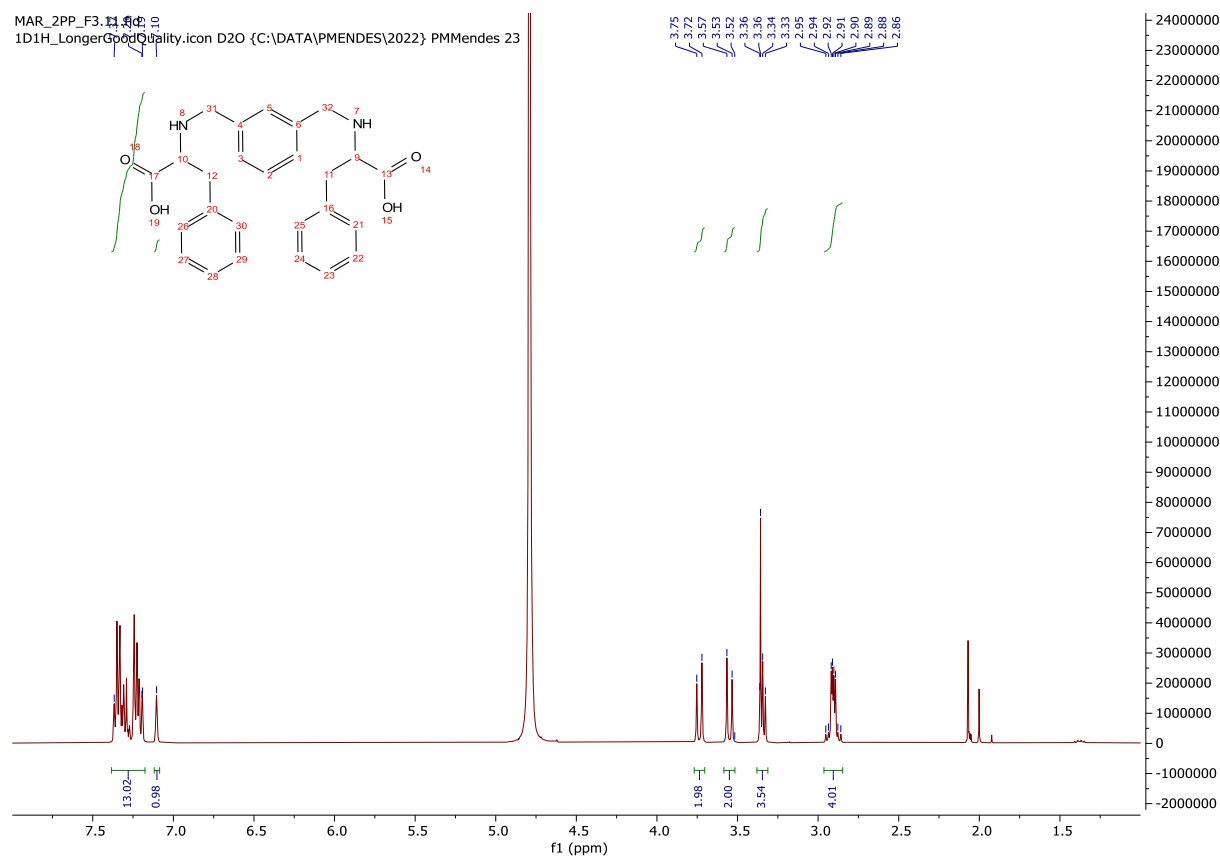
Annex 8

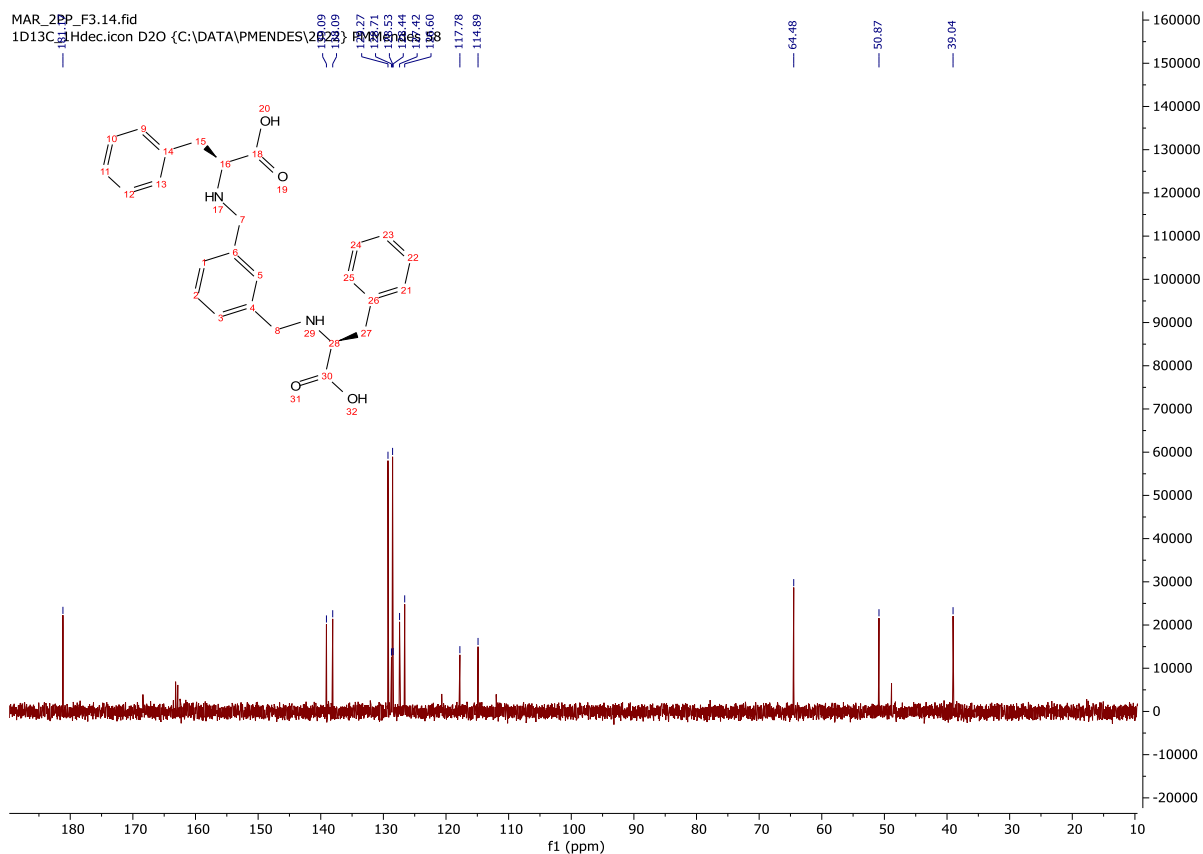
Characterisation of molecule **2DD**: ^1H NMR spectrum (400 MHz, D_2O , 298 K) (top), and MS (bottom).



Annex 9

Characterisation of molecule **2PP**: first ^1H NMR spectrum (400 MHz, D_2O , 298 K), then ^{13}C NMR spectrum (400 MHz, D_2O , 298 K) (next page, top), and MS (next page, bottom).





MAR_2PP C26H28N2O4 MW=433
Acetonitrile:Water
PMM-MXA-9EK7R-ESI-Neg-1 48 (0.326) AM2 (Ar,28000.0,0.00,0.00); ABS

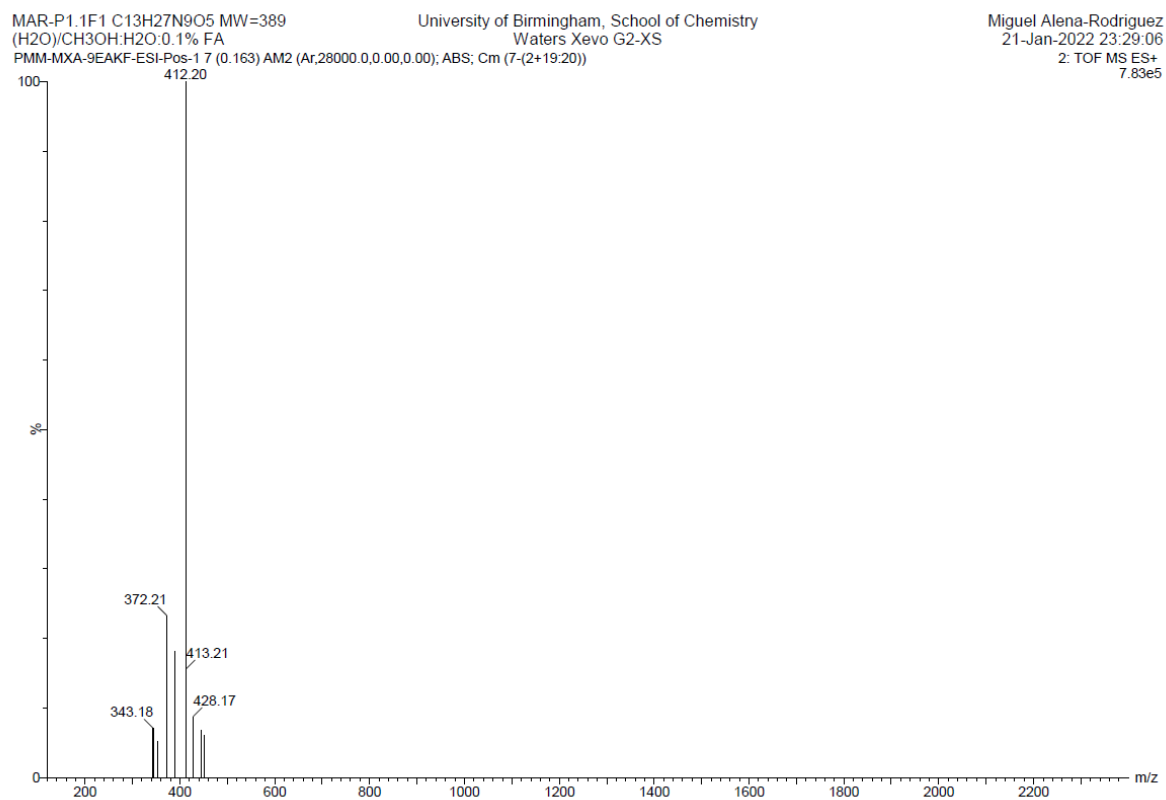
University of Birmingham, School of Chemistry
Waters Xevo G2-XS

Miguel Alena-Rodriguez
08-Jun-2022 23:48:45
1: TOF MS ES-
3.15e5



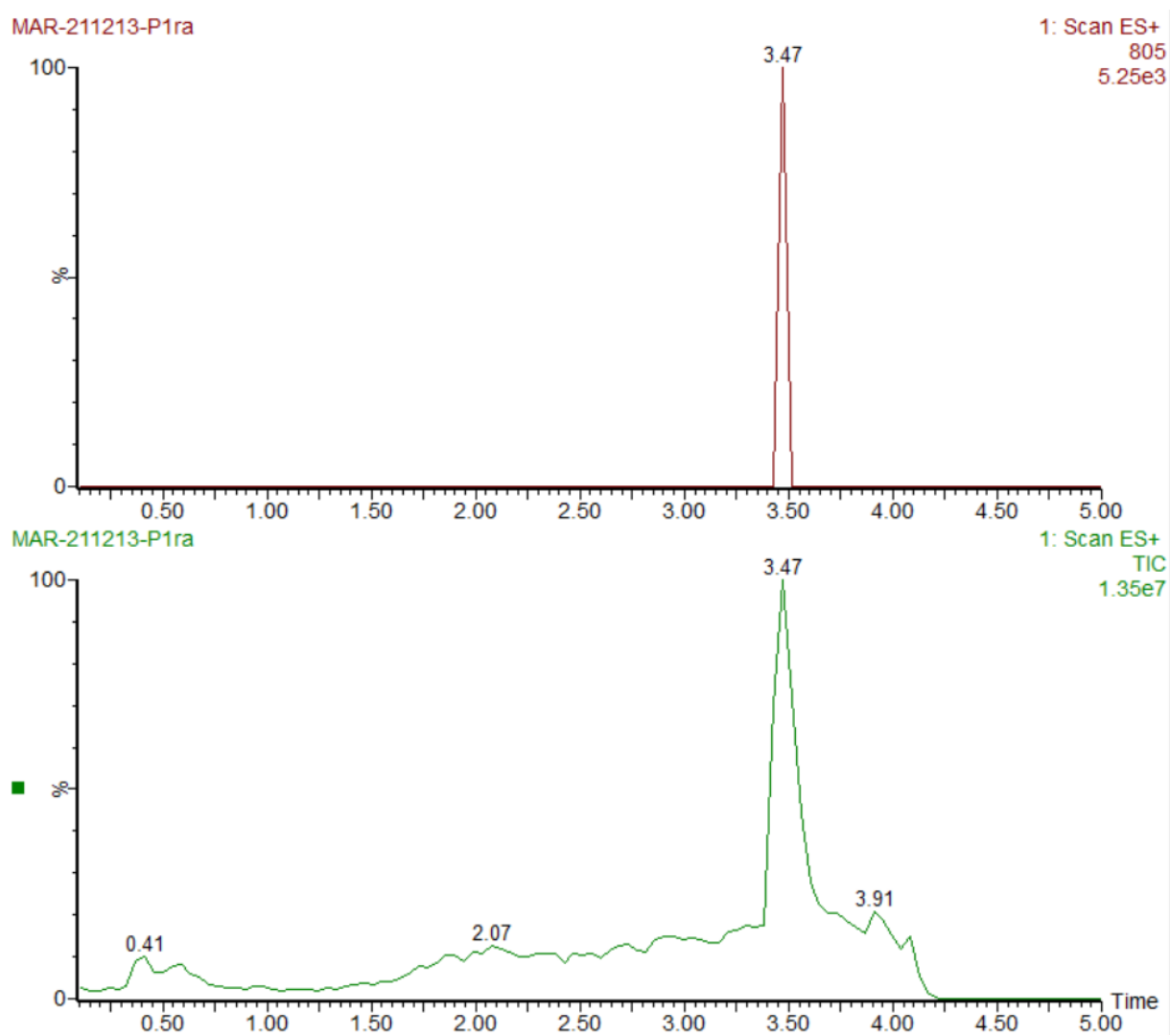
Annex 10

Characterisation of peptide **4** by MS.



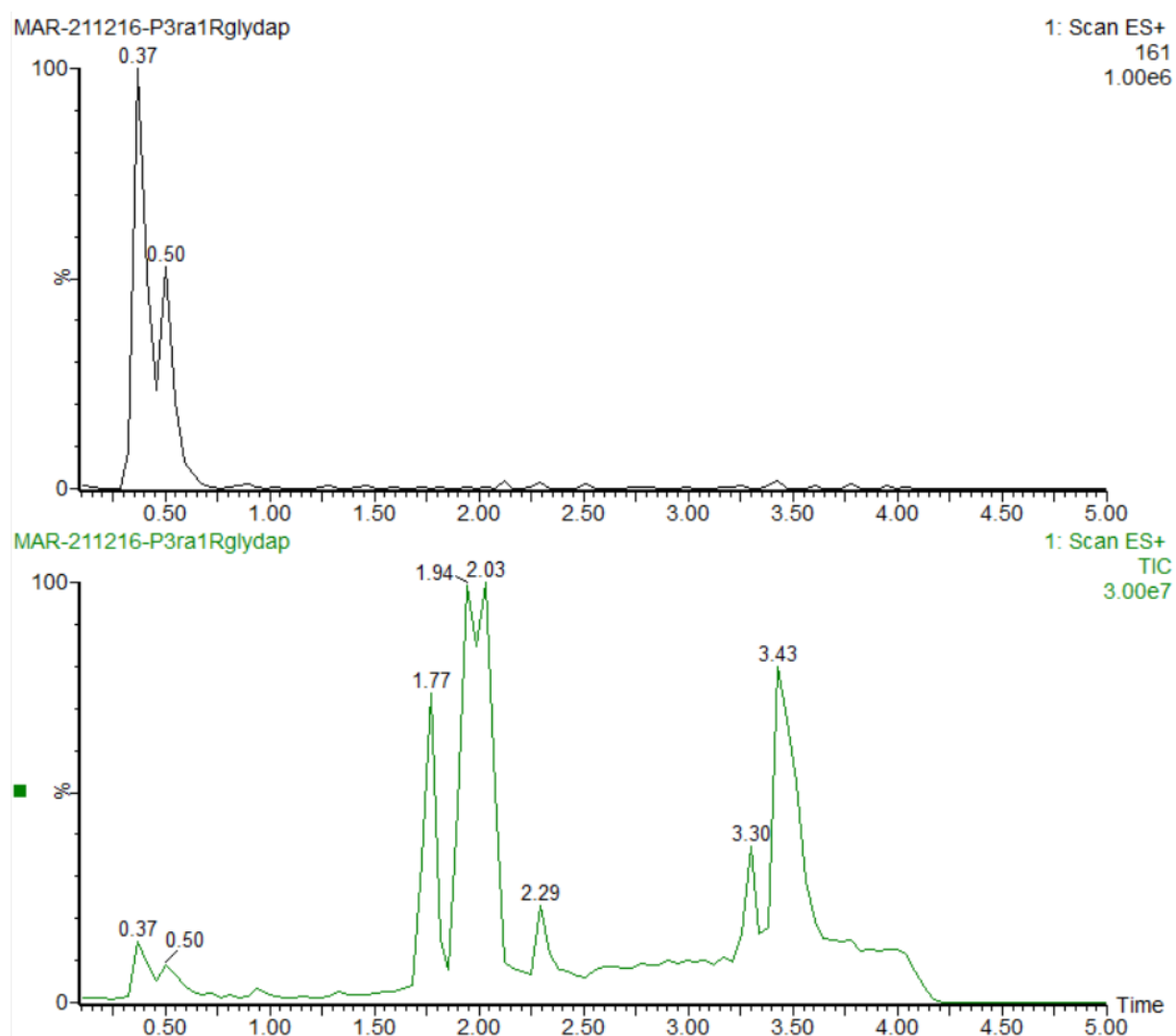
Annex 11

*LC-MS of the crude resulting from the cleavage of peptide **4SSS** resulting from the on-resin reductive amination reaction.*



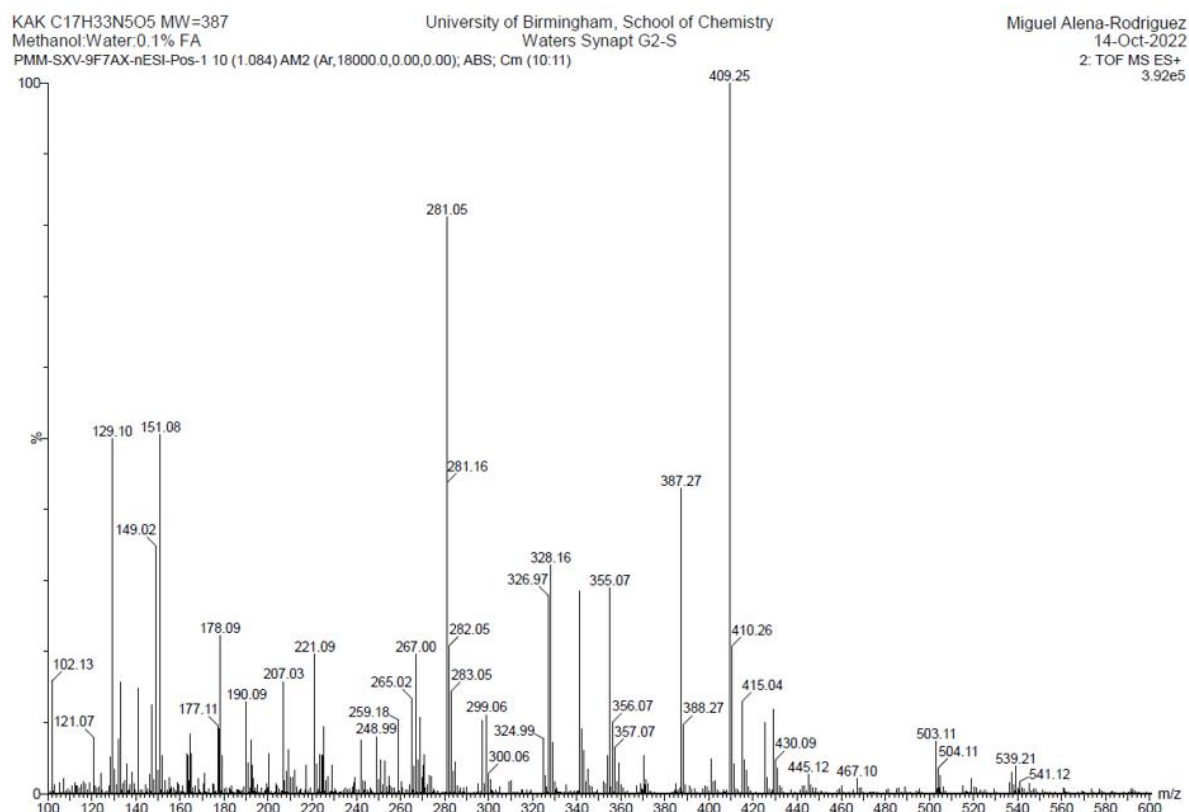
Annex 12

LC-MS of the crude resulting from the cleavage of peptide **4T** resulting from the on-resin reductive amination reaction.



Annex 13

Characterisation of peptide **KAK** by MS.



Annex 14

Characterisation of peptide **KAAK** by MS.

