# Visualization of Multivalent Binding Events using α-Hemolysin




A thesis submitted to the Board of the Faculty of Physical Sciences in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford

Pieter E Gregoir

Pembroke College, Oxford

Michaelmas 2020

## Author's declaration

The work presented in this thesis was conducted at the Chemistry Research Laboratory at the University of Oxford under the supervision of Professor Hagan Bayley and/or Professor Ben Davis. All the work is my own, except where otherwise stated, and has not been submitted for any other degree at this or any other university.

Pieter Gregoir

November 2020

**Short Abstract for Bodleian Library Indexing**

Visualization of multivalent binding events using α-Hemolysin

Pieter Gregoir

Pembroke College

University of Oxford

Submitted for the degree of Doctor of Philosophy in Chemical Biology

Michaelmas 2020

The work done in this thesis aims to visualize multivalent binding events of proteins and their ligands, more specifically lectins and sugar moieties. This visualization is done by single channel recording (SCR) via stochastic sensing. Herein, the binding is studied at a single molecular level. Two mannose moieties are covalently connected by linker molecules to an ion channel forming protein, forming bivalent mannopores. These bivalent mannopores are used in a SCR set-up, wherein different lectins are added to study their binding behaviour. The lectins differ in their valency, with monovalent, bivalent, trivalent and tetravalent lectins being used. For each of these lectins, a binding model was derived that shows the dynamic nature of a multivalent binding event. This work shows that lectin-sugar binding is much more dynamic than previous theories would suggest.

**Acknowledgement**

In first place I would like to thank Hagan Bayley, Nick Green, David Hodgson, Aga Borkowska, Rob Adligton, Mark Fricker and Alison Franklin in helping to resubmit this improved version of my thesis since the first submission of this thesis in October 2015.

I would like to thank the Marie Curie Actions and the Frame Work 7 program for the opportunity and the funding to achieve this work. I would like to thank all the members of the Carmusys network, and I felt fortunate to be part of this. I am sure that meeting all these people will benefit the rest of my career.

I would like to thank Hagan Bayley, Ben Davis, Samantha Shanley and Anne Hammerstein for their academic help.

Thanks to Frank Fieschi for the DC-SIGN and Langerin protein.Thanks to Mark Wormald for the molecular modelling.

Thanks to the members of the Ben Davis group and the Hagan Bayley group, the Chemistry department support staff, NMR staff and Mass spec staff.

Special thanks to Hua, Amy, Regis, Chris, Matt, Shirley, Bala, Alba, Macarena, Jonny and Kim for all the good times in and out of the lab. Hua, thank you for the help with the corrections.

Special thanks go out to my family and friends in Belgium, especially Ward Huybrechts, Kristel Van den Meerssche and Dennis Guilliams, who helped me through a difficult time.

I could not have finished this work without the help from Smita Gunnoo, thank you Smita for everything you have done over the years and all the help during writing.

I would like to thank Chris Schofield and Peter Cragg to be my Examiners. I appreciate the time it must have taken to correct the first version of this thesis.

## Abbreviations

| | |
|---|---|
| α | Degree of cooperativity |
| α-HL | α-Hemolysin |
| Ac | Acetate |
| ACN | Acrylonitrile |
| amp | Ampicillin |
| aq | Aqueous |
| Arg | Arginine |
| Asp | Aspartic acid |
| β | Enhancement factor |
| Bn | Benzyl |
| $C_{bulk}$ | Bulk concentration |
| $C_{EFF}$ | Effective concentration |
| Con A | Concanavalin A |
| CRD | Carbohydrate recognizing domain |
| δ | Chemical shift |
| d | Doublet |
| $D_8H_6$ | $(Aspartic\ acid)_8(histidine)_6$ tail |
| DBU | 1,8-Diazabicyclo[5.4.0]undec-7-ene |
| DC SIGN | Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin |
| DCM | Dichloromethane |
| dd | Doublet of doublets |
| DIBALH | Diisobutylaluminium hydride |
| DMAP | 4-Dimethylaminopyridine |
| DPhPC | 1,2-Diphytanol-sn-glycero-3-phosphocholine |
| DTT | Dithiothreitol |
| EDTA | Ethylenediaminetetraacetic acid |
| EG | Ethylene glycol |
| EGFR | Epidermal growth factor receptor |

| | |
|---|---|
| eq | Equivalents |
| Et | Ethyl |
| EtOAc | Ethyl acetate |
| Gal | Galactose |
| Gly | Glycine |
| h | Hour(s) |
| IgA | Immunoglobulin A |
| IgD | Immunoglobulin D |
| IgE | Immunoglobulin E |
| IgG | Immunoglobulin G |
| IgM | Immunoglobulin M |
| inter | Intermolecular |
| intra | Intramolecular |
| ITC | Isothermal titration calorimetry |
| ITG | Isopropyl-β-D-thiogalactoside |
| J | Coupling constant |
| K | Equilibrium constant |
| $K_8C$ | α-HL variant, lysine at position 8 is substituted for cysteine |
| $K_a$ | Affinity constant |
| $K_d$ | Dissociation constant |
| $k_{off}$ | Reverse rate constants |
| $k_{on}$ | Forward rate constant |
| L | Ligand |
| LB | Luria broth |
| Lev | Levulinate |
| m | Multiplet |
| m/z | Mass to charge ratio |
| MALDI-TOF-MS | Matrix assisted laser desorption ionisation time of flight mass spectrometry |
| man | Mannose |

| | |
|---|---|
| Me | Methyl |
| mono | Monovalent |
| multi | Multivalent |
| NIS | N-Iodosuccinimide |
| NMR | Nuclear magnetic resonance |
| OPSS | Orthopyridyl disulfide |
| PDB | Protein data base |
| PEG | Polyethylene glycol |
| Ph | Phenyl |
| PMA | Phosphomolybdic acid |
| PMB | p-Methoxybenzyl |
| poly | Polyvalent |
| PS | Pisum sativum |
| PTFE | Polytetrafluoroethylene |
| pTSA | para Toluene sulphonic acid |
| Pyr | Pyridine |
| q | Quartet |
| RGD | Arginine-glycine-aspartic acid tripeptide |
| rRBC | Rabbit red blood cell membranes |
| RT | Room temperature |
| RXR | Retinoid X receptor |
| s | Singlet |
| SCR | Single channel recording |
| SDS | Sodium dodecyl sulfate |
| SPR | Surface plasma resonance |
| t | Triplet |
| TBDMS | tert-Butyldimethylsilyl |
| TFA | Trifluoroacetic acid |
| THF | Tetrahydrofuran |

| | |
|---|---|
| TLC | Thin layer chromatography |
| TOF | Time of fight |
| $T_{off}$ | Dwell time of binding events |
| $T_{on}$ | Dwell time of non-binding events |
| Tris | Tris hydroxymethylaminoethane |
| Ts | Tosylate |

# TABLE OF CONTENTS

1 MULTIVALENCY 1

**1.1 Introduction....1**

**1.2 Terms and definitions ....2**

**1.3 Relevance of multivalent binding events in biology ....3**

1.3.1 Attachment of viruses to host cells .... 3

**1.4 Interaction between bacteria and cells....5**

**1.5 Binding of cells to other cells....7**

**1.6 Binding of cells to polyvalent molecules e.g. antibodies ....8**

**1.7 Binding of polyvalent molecules to polyvalent molecules ....9**

2 THERMODYNAMICS OF MULTIVALENT BINDING 11

**2.1 Cooperativity according to Whitesides....11**

**2.2 Hill equation ....12**

2.2.1 Positive cooperativity ($\alpha > 1$)..... 14

2.2.2 Negative cooperativity ($\alpha < 1$) ..... 15

**2.3 Gibbs free energy and the enhancement factor....16**

**2.4 Entropy and enthalpy....17**

**2.5 Multivalency according to Kitov: ....23**

**2.6 Effective concentration ....29**

**2.7 Kinetics of multivalent binding ....31**

**2.8 Cooperativity according to Anderson ....34**

**2.9 Multivalency according to Ercolani....37**

**2.10 Isothermal titration ....38**

3 STOCHASTIC SENSING 40

**3.1 Experimental design ..... 41**

**3.2 Single channel recording ..... 42**

**3.3 α-Hemolysin ..... 45**

**3.4 Preparation of a monovalent pore ..... 46**

**3.5 Preparation of a bivalent pore ..... 53**

**3.6 Current-voltage (IV) characteristic recordings ..... 56**

**3.7 FimH control ..... 57**
3.7.1 Production of $FimH_{1-156}$ ..... 58
3.7.2 Mass spectrometry ..... 61
3.7.3 FimH and monovalent mannopore ..... 61
3.7.4 FimH and bivalent mannopore ..... 67
3.7.5 Explanation of difference in $K_d$ as reported in literature ..... 71

**3.8 Con A ..... 76**
3.8.1 Introduction ..... 76
3.8.2 Con A and monovalent mannopore ..... 76
3.8.3 Con A controls ..... 77
3.8.4 Con A and bivalent mannopore ..... 79
3.8.5 Docking of Con A and a bivalent mannopore ..... 84
3.8.6 Conclusion ..... 90

**3.9 PS lectin ..... 91**
3.9.1 PS monovalent ..... 92
3.9.2 PS bivalent ..... 94

**3.10 Langerin ..... 99**
3.10.1 Langerin monovalent pore ..... 100
3.10.2 Langerin Bivalent ..... 102

**3.11 DC-SIGN ..... 105**
3.11.1 DC-SIGN and monovalent mannopore ..... 106
3.11.2 DC-SIGN and bivalent mannopore ..... 107

**3.12 Discussion and Conclusion ..... 111**

4 CARBOHYDRATE SYNTHESIS 115

**4.1 $PEG_{10}$ Synthesis ..... 115**

**4.2 Mannose D1 arm ..... 116**

**4.3 Convergent route ..... 116**
4.3.1 Donor Synthesis ..... 117
4.3.2 Acceptor Synthesis ..... 118
4.3.3 Glycosylation ..... 119

**4.4 Linear route ..... 120**

**4.5 *2-C* branched sugar analogues ..... 123**
4.5.1 Synthesis of 2-*C*-substituted branched D-mannose analogues .. 123

5 EXPERIMENTAL 126

6 REFERENCES 154

7 ADDENDUM 168

# 1 Multivalency

## 1.1 Introduction

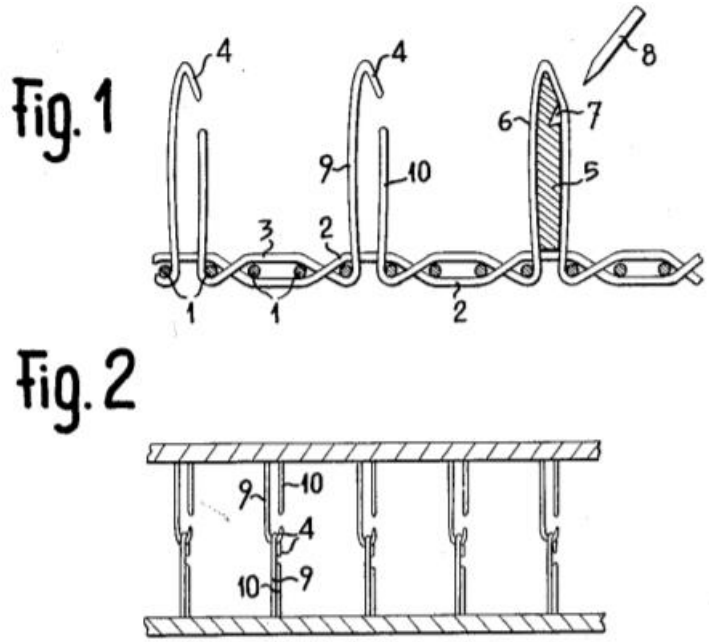



**Figure 1.1.** Drawing from the original Patent[100] of Velcro®, as a macroscopic exemplar of multivalency.

The adhesive nature of Velcro® provides a macroscopic scale example of multivalency. Velcro® contains nylon "hooks" that grab polyester "loops". Although each hook to loop interaction is very weak, when multiple hook to loop interactions are present, the overall binding is stronger and yet reversible (Figure 1.1). On a molecular scale, multivalency effects are observed when certain receptors (the hooks) bind to their cognate ligands (the loops). Understanding multivalency is important, and the principles governing multivalent effects can be used in the field of medicine, biochemistry, supramolecular chemistry and material sciences.[1] To date,

multivalency, although extensively researched, is not completely understood. To obtain a better insight into multivalent binding processes, multivalent interactions between carbohydrates and lectins have been explored. In nature, carbohydrate-lectin interactions are common. There are many examples of carbohydrate-lectin interactions following multivalency principles, where single carbohydrate-lectin interactions are mostly weak, and the increase in binding strength where multiple molecules are involved allows the required biological function to be fulfilled.[2]

## 1.2 Terms and definitions

Some terminology commonly used in the field of multivalency is described below.

**Cooperativity** arises from the interplay of two or more interactions, such that the behaviour of the system as a whole differs from expectations based on the properties of the individual interactions acting in isolation. Coupling of interactions can lead to positive ($\alpha > i$) or negative ($\alpha < i$) cooperativity, depending upon whether one interaction favours or disfavours another.[3]

**Multivalency or polyvalency** is used to describe binding interactions that are due to a combination of several simultaneous binding events. With multivalency, the overall binding is significantly greater than the simple

combination of the individual binding events, i.e. the "whole" of the interaction is greater than the sum of its parts.[4]

**Chelate cooperativity** arises from the presence of one or more intramolecular binding interactions, that is, as a consequence of the chelate effect (multivalency).[5]

The **chelate effect**, observed where two binding groups are covalently connected and the overall affinity is larger than the sum of the two individual affinity constants of each group.[5] This effect was first observed by Schwarzenbach in 1952.[6]

## 1.3 Relevance of multivalent binding events in biology

A range of examples demonstrating the importance of multivalency in biological systems is described in the section below. The examples range from two relatively small multivalent molecules interacting with each other to multivalency involved in cell – cell interactions.

### 1.3.1 Attachment of viruses to host cells

The first step of viral infection is the multivalent interaction of multiple molecules on the surface of the virus with multiple molecules on the surface

of the host cell allowing the host cell to perform endocytosis of the virus.[7] For example the influenza virus displays trimers of the lectin haemagglutinin on its surface, which is then densely packed on the virus surface with 600 to 1200 copies per virus particle.[7] Bronchial epithelial cells have multiple glycoproteins on their surface, where the terminal sugar is sialic acid. These sugar moieties are also densely arranged over the surface and bind strongly to the lectin haemagglutinin trimers as can be seen in Figure 1.2. This example demonstrates how the virus exploits multivalency to infect cells, and how the human body responds using multivalency itself to prevent infection. Mucins-containing fluid on the inside of the surface of the lungs is responsible for defence against infection by the influenza virus. Mucins are heavy glycosylated proteins with the oligosaccharides bearing terminal sialic acid moieties. When the influenza virus or other sialic acid binding viruses try to penetrate the fluid layer in the lungs, the sialic acid moieties of the mucins bind to the receptors on the virus surface, preventing binding to the epithelial cells.[8, 9] Current drugs preventing endocytosis are based on monovalent molecules, which mean that a high concentration of the drug is needed to be effective. Multivalent drugs should be effective at much lower concentrations due to stronger binding interactions as a result of multivalency.[10]

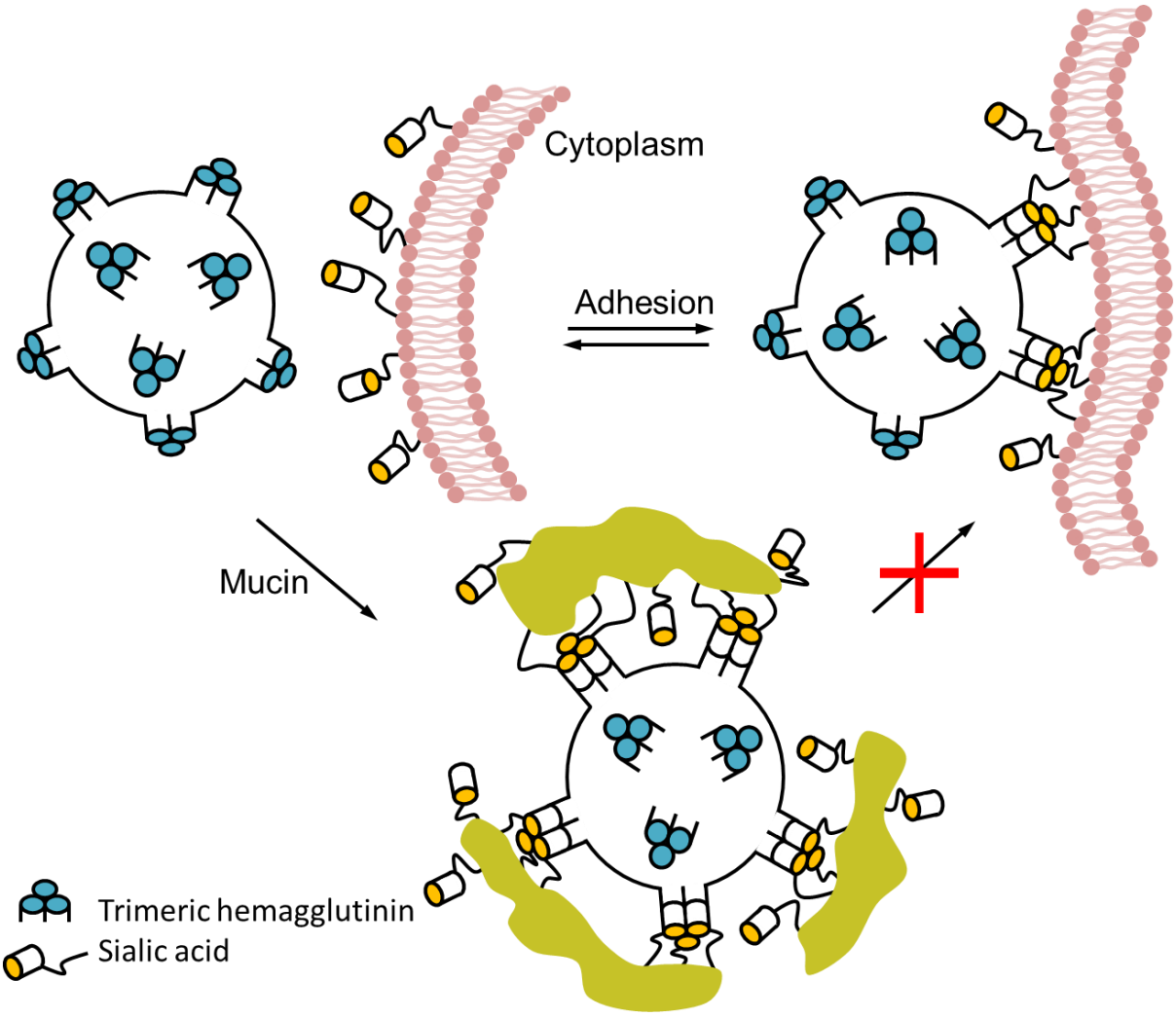



**Figure 1.2.** The first step in the uptake of influenza virus by host cells. The haemagglutinins on the surface of the virus bind in a multivalent manner to sialic acids on the surface of host cells. A natural defense is to produce mucin, a glycoprotein with a high surface density of sialic acids. The mucin prevents the virus from binding to the host cells.

## 1.4 Interaction between bacteria and cells

Bacteria adhere to the surface of host cells in a multivalent manner. The uropathogenic strains of *Escherichia coli* (*E. coli)* attach to the surface of epithelial cells in the bladder and in the urethra using lectins on the P-fimbrae (protein G) and on the type I fimbrae (FimH) (Figure 1.3).[11,12] Protein G binds strongly to Gal(α1,4)Gal (Gal – galactose) a part of a

glycolipid present in the cell membrane of the epithelial cells within the urinary tract. FimH binds specifically to mannose and uses the high-mannose glycoprotein, uroplakin-1a on the surface of fibronectins. [11,12] These fibronectins also display multiple copies of the tripeptide Arg-Gly-Asp (RGD) which strongly binds to the integrin-type receptors on the cell surface. As all these type of interactions are too weak to hold the whole complex together, multiple bindings is required resulting in an overall hetero-multivalency.[9-12]

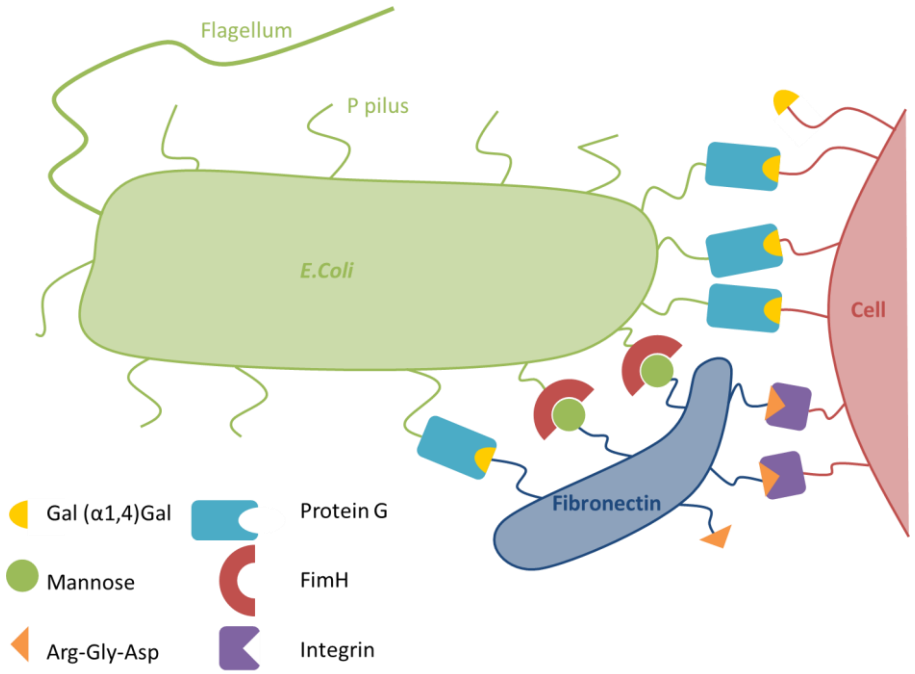



**Figure 1.3.** Attachment of uropathogenic *E. coli* to endothelial urethral cells. *E. coli* displays two proteins on the surface involved in the attachment: Fim H and protein G. Protein G adheres to Gal (α1,4) Gal which is displayed on the surface of the endothelial cells. Additionally, FimH adheres to mannose present on the surface of fibronectins. Fibronectin presents multiple copies of the tripeptide (RGD). These RGD peptides have a high affinity for integrin-type receptors present on the cell surface of the endothelial cells.

## 1.5 Binding of cells to other cells

Upon inflammation, neutrophils suspended in the blood attach to endothelial cells around the site of the inflammation in a multivalent fashion (Figure 1.4).[9] The endothelial cells lining the interior of the blood vessels near the site of inflammation start to display multiple copies of E- and P-selectins together with glycoproteins containing sialyl Lewis$^X$.[15] The neutrophils display L-selectin and also sialyl Lewis$^X$ molecules on their surface. The various selectins bind to the sialyl Lewis$^X$ sugars, this slows down the neutrophils and makes them stick to the surface of the blood vessel allowing them to migrate out of the blood vessel. [9, 13-15]

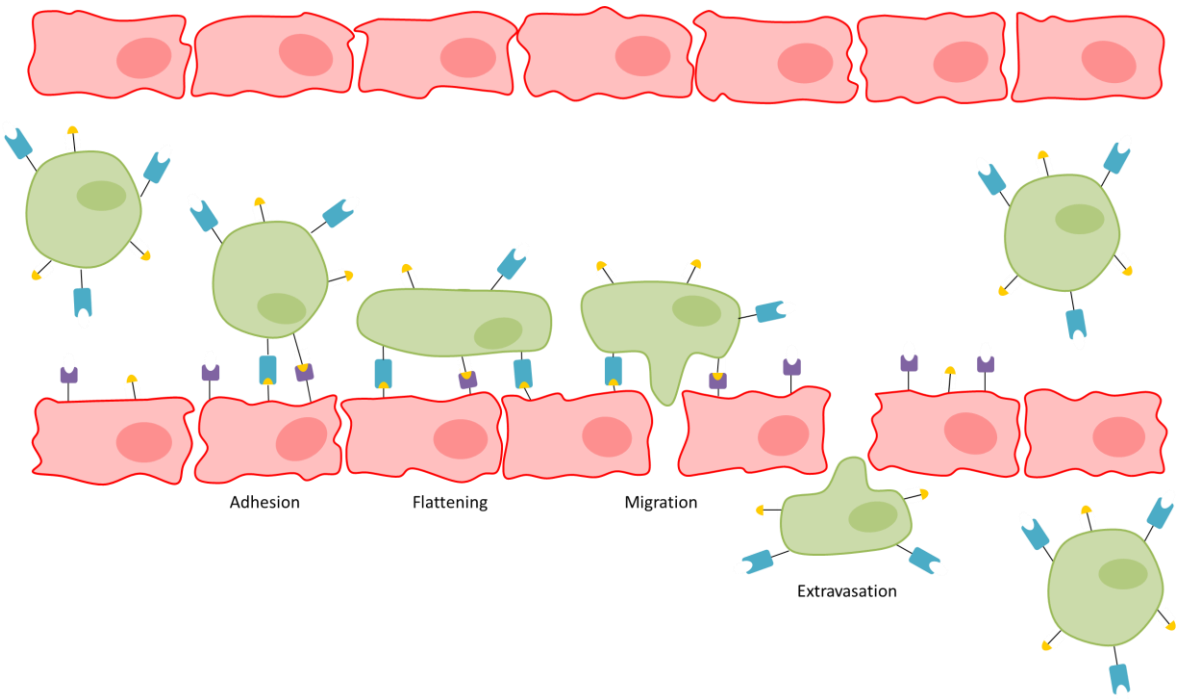



**Figure 1.4.** Inflammation causes display of E- and P- selectins on nearby endothelial cells lining the inner wall of blood vessels. Neutrophils interact in a multivalent fashion with the E-and P- selectins. In addition, L-selectin displayed on the surface of the neutrophils interact with sialyl Lewis$^X$ moieties are present on the endothelial cells and on the neutrophils. This causes the neutrophils to adhere, flatten, migrate and eventually extravagate out of the blood vessel.

## 1.6 Binding of cells to polyvalent molecules e.g. antibodies

Antibodies have multiple, equivalent antigen binding sites. They exist as monomers with two binding sites (IgD, IgE and IgG), dimers with four binding sites (IgA) and pentamers with ten binding sites (IgM)[16]. One single antibody bound to a pathogen is not sufficient to induce an immune response from macrophages. For macrophages to ingest a pathogen, multiple antibodies need to be bound to the surface of the invading particle. This results in a three-layered structure held together by multivalent interactions as depicted in Figure 1.5.[17] In this case, multivalency is involved in the mechanism of recognition as shown in Fig 1.5. [9, 18]

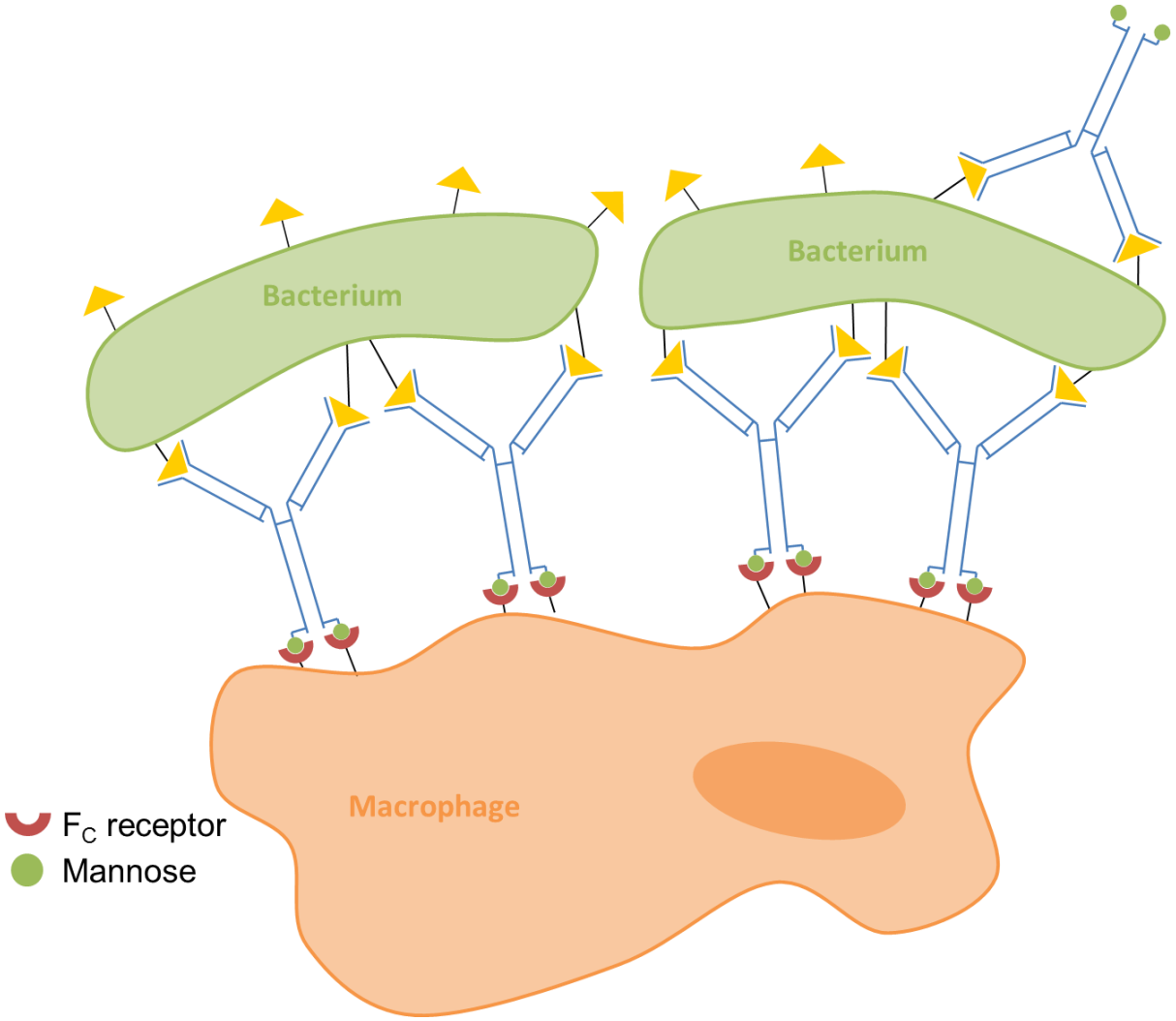



**Figure 1.5.** Three–layered structure of macrophage pathogen binding via antibodies. The antibodies recognise and bind bivalently to the bacterial surface. The Fc receptors on the macrophages bind bivalently to mannose residues on the constant region of the antibodies leading to this staged structure.

## 1.7 Binding of polyvalent molecules to polyvalent molecules

Multivalency is also used by nature for gene regulation. The complex of the retinoid X receptor (RXR) and its ligand (L) function as a transcription factor and is shown in Figure 1.6. When one RXR-L complex binds to single strand DNA the affinity is low, whereas when the RXR-L dimerises, the affinity for DNA increases but is still low. Further still, the affinity of the tetramer (RXR-

$L)_4$ is moderate and the affinity of the pentamer $(RXR\text{-}L)_5$ is higher as depicted in Fig 1.6. The DNA has multiple binding sites allowing these binding processes with increasing affinities to occur. The rate of gene expression is more dependent on what multimeric form the $(RXR\text{-}L)_n$ takes, followed by the concentration of RXR-L.[9]

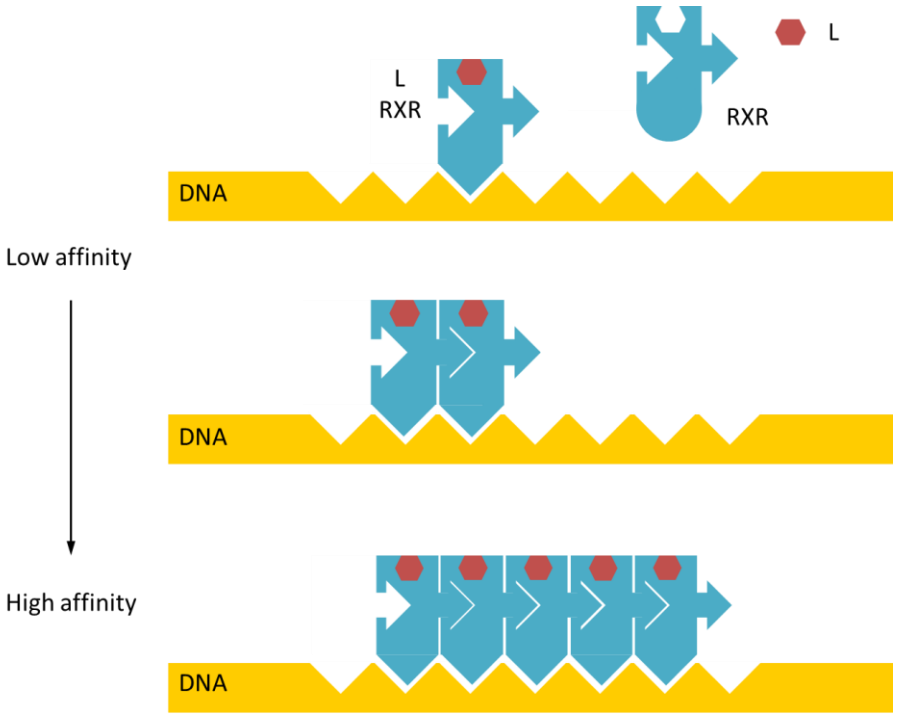



**Figure 1.6.** Binding of transcription factors to multiple sites on DNA.

# 2 Thermodynamics of multivalent binding

As described in Section 1, multivalent interactions resulting in strengthened binding interactions are paramount to many biological processes. This makes the study of multivalency interesting for rational drug design, where molecules bearing multiple ligands are designed to interfere with naturally occurring multivalent binding processes. Ideally, the molecules should be competitive inhibitors with affinities higher than that of the natural process. Theories about multivalency have been proposed with the aim of designing the best possible scaffolds to compete with naturally occurring multivalent interactions. A good understanding of the thermodynamic properties involved with multivalent binding combined with computational modelling should allow the prediction and design of multivalent molecules with increased binding affinities, reducing the degree of serendipity in the synthesis of targeted multivalent ligands.

## 2.1 Cooperativity according to Whitesides

The first major theory about multivalent binding was proposed by Whitesides[9] and co-workers where a clear differentiation between cooperativity and multivalency was described.

The outcome of cooperativity, the overall average free energy of a polyvalent interaction ( $\Delta G_{avg}^{poly}$ ) (Equation 2.1 - 2.3), can be greater to, equal

to or less than the individual free energy of a monovalent interaction ( $\Delta G^{mono}$ ).

$$\Delta G_{avg}^{poly} = \alpha \Delta G^{mono} \quad \text{(Eq. 2.1)}$$

$$N\Delta G_{avg}^{poly} = \alpha N \Delta G^{mono} \quad \text{(Eq. 2.2)}$$

$$K_N^{poly} = (K_{avg}^{poly})^N = (K^{mono})^{\alpha N} \quad \text{(Eq. 2.3)}$$

$$\alpha = \frac{\ln(K_N^{poly})}{\ln(K^{mono})^N} \quad \text{(Eq. 2.4)}$$

$\alpha > 1$: postitively cooperative interactions (synergistic)

$\alpha = 1$: non cooperative interactions (additive)

$\alpha < 1$: negative cooperative interactions (interfering)

The factor α (Equation 2.4) is the degree of cooperativity, and the unit of α depends on the order ($N$) of the polyvalent interaction, wherein $K_N^{poly}$ represents the affinity constant associated with the total polyvalent interaction, $K^{mono}$ represents the affinity constant associated with a monovalent interaction and $K_{avg}^{poly}$ represents the affinity constant associated with one single interaction in the total polyvalent interaction.[9]

## 2.2 Hill equation

In 1910, Archibald Vivian Hill proposed a model for the relation between the partial pressure of oxygen and the saturation of haemoglobin with oxygen.[19] He studied the influence of different salts on the oxygen saturation of haemoglobin, and proposed three equations in the original article, one of which is now known as the Hill equation (Equation 2.5).

$$y_1 = 100\frac{Kx^{\alpha}}{1+Kx^{\alpha}} \qquad \text{(Eq. 2.5)}$$

In Equation 2.5, $y_1$ represents the saturation of haemoglobin with oxygen, x is the partial pressure of oxygen and α (the degree of cooperativity) and K (the equilibrium constant) are the two variable parameters.

Since the original publication[19], the Hill equation (Equation 2.6) has been used in a slightly different form to study drug concentration – effect relations, or saturation of other enzymes and molecules[20-22].

$$y_{\alpha} = \frac{[M]^{\alpha}}{K_{\alpha}[M]^{\alpha}}, \quad \alpha > 0 \qquad \text{(Eq. 2.6)}$$

Equation 2.6 can be used to study cooperativity, with $K_{\alpha}$ being an equilibrium constant for the overall binding, [M] representing the concentration of the ligand, and α representing the degree of cooperativity. The effect of the degree of cooperativity factor can be seen easily in the shape of the concentration saturation plots (Figure 2.1). Positive cooperativity shows as a sigmoidal curve, hyperbolic curves relate to neutral or negative cooperativity depending on the slope at lower concentrations (Figure 2.1 a).[3]

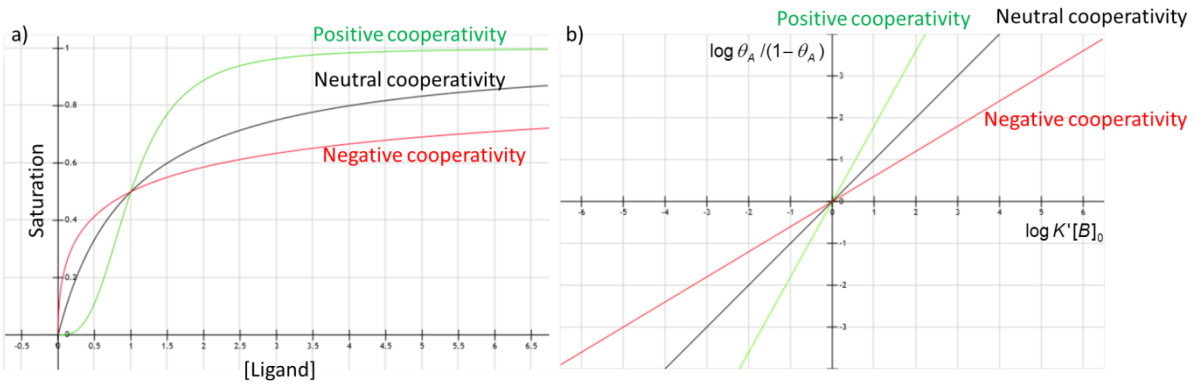



**Figure 2.1.** Concentration saturation plots, a) linear, b) logarithmic. Green lines is positive cooperativity, black lines is neutral cooperativity and red lines is negative cooperativity.

### 2.2.1 Positive cooperativity (α > 1)

Haemoglobin, a protein present in the blood of most vertebrates responsible for transporting oxygen from the lungs to the rest of the body and transporting carbon dioxide back to the lungs, is a tetravalent protein containing four iron (II) dependent haem groups where the oxygen coordinates.[23] The monomeric form is myoglobin. The affinity of deoxy-haemoglobin for oxygen is rather low. However, after the coordination of a first oxygen molecule to one of the four haem groups of the haemoglobin, the affinity for the second oxygen molecule increase considerably. A further increase in affinity is noticeable for the binging of the third oxygen molecule and consequently the binding of the fourth oxygen molecule. The affinity of 3-oxy-haemoglobin for oxygen is approximately 300 times stronger than the affinity of deoxy-haemoglobin for oxygen ( $-\Delta G^{poly}_{avg} > -\Delta G^{mono}_{(first)}$ ). Said increasing affinity of haemoglobin for oxygen is reflected in the shape of the oxygen-affinity curve, which is sigmoidal (Figure 2.2 blue), whereas the oxygen-affinity curve of monovalent myoglobin is hyperbolical (Figure 2.2

red). The difference in the shapes of the oxygen-affinity curves is an example of positive or synergistic cooperativity (α > 1) with no multivalency involved. The degree of cooperativity is therefore greater than 1 and is unitless.[23]

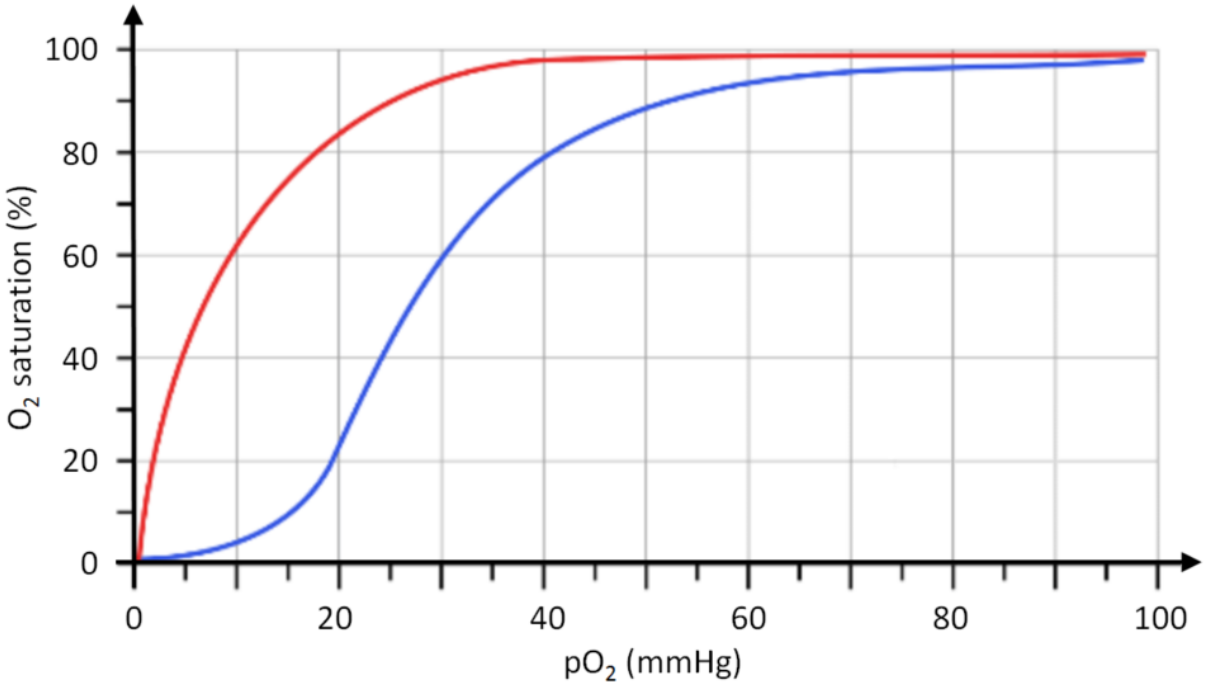



**Figure 2.2.** $O_2$-affinity curve of haemoglobin (blue) and myoglobin (red).

### 2.2.2 Negative cooperativity (α < 1)

An example of negative cooperativity is the transmembrane signalling by the epidermal growth factor receptor (EGFR).[25] Dimerization is induced by ligand binding, but negative cooperativity is observed when the second ligand binds. The mechanism for this behaviour has been revealed by studying crystal structures[25] of the *Drosophila melanogaster* EGFR extracellular region as this is the only EGFR that retains its negative cooperativity behaviour when its extracellular region is studied in isolation, allowing the growth of crystals of different binding states of the EGFR and

its growth factor ligand Splitz.[25] This complex is asymmetric and affects the free binding site, making it less favourable for a second ligand to bind to the complex ($-\Delta G_{avg}^{poly} < -\Delta G_{(first)}^{mono}$).[24, 25]

## 2.3 Gibbs free energy and the enhancement factor

Multivalency is a form of cooperativity. In more recent publications the term chelate cooperativity is used to indicate multivalency. The Gibbs free energy associated with multivalent binding ($\Delta G_N^{poly}$) is given by Equation 2.7. By analogy with degree of cooperativity factor α, Whitesides[9] introduced a parameter β in the case of multivalency. The parameter β (Equation 2.8) is the enhancement factor and can be calculated if the number of interactions N between the ligand and receptor is unknown, making it more practical in cases where polymers or surfaces are used with an unknown density of ligands. β is defined as the ratio of the avidity of the multivalent ligand to the avidity of its monovalent counterpart.[9]

$$\Delta G_N^{poly} = \Delta G^{mono} - RT\ln(\beta) \qquad \text{(Eq. 2.7)}$$

$$\beta = \frac{K_N^{poly}}{K_{mono}} \qquad \text{(Eq. 2.8)}$$

Since the parameter β has been introduced[9], β values reported in publications have been based on dissociation constants instead of association constants. By using dissociation constants, the β parameter becomes much more intuitive. In Whitesides latest publication[26] on the

subject of multivalency he defines β as in Equation 2.9 based on the dissociation constant of the monovalent interaction ( $K_d^{mono}$ ) and the dissociation constant of the multivalent interaction ( $K_d^{avidity}$ ).[26]

$$\beta = \frac{K_d^{mono}}{K_d^{avidity}} \qquad \text{(Eq. 2.9)}$$

## 2.4 Entropy and enthalpy

As in every thermodynamic study, effects influencing both the entropy and enthalpy of the system need to be taken into account (Equation 2.10).[9] In this section the effects multivalency has on these factors will be discussed.

$$\Delta G_N^{poly} = \Delta H_N^{poly} - T\Delta S_N^{poly} \qquad \text{(Eq. 2.10)}$$

The value for the enthalpy of the multivalent interaction ( $\Delta H_N^{poly}$ ) is not always equal to the sum of the enthalpy of *N* monovalent interactions ( $N\Delta H^{mono}$ ). Some ligand binding enhances the strength of binding of a second ligand. The example of oxygen binding to haemoglobin is an example of enthalpically enhanced binding.[9]

Binding of the first ligand can also decrease the strength of binding of a second ligand. This can be the result of conformational changes in the receptor, but also due to steric restrictions of the multivalent ligand. When the multivalent ligand does not have the correct spatial orientation of the individual ligands compared to the geometry of the receptors, enthalpy is

lost due to deformations in torsion angle and bond lengths in the molecule linking the ligands. The more rigid one of the multivalent entities is the more likely it is that small spatial mismatches between receptor and ligand will result in enthalpically diminished binding as shown in Figure 2. 3.[9]

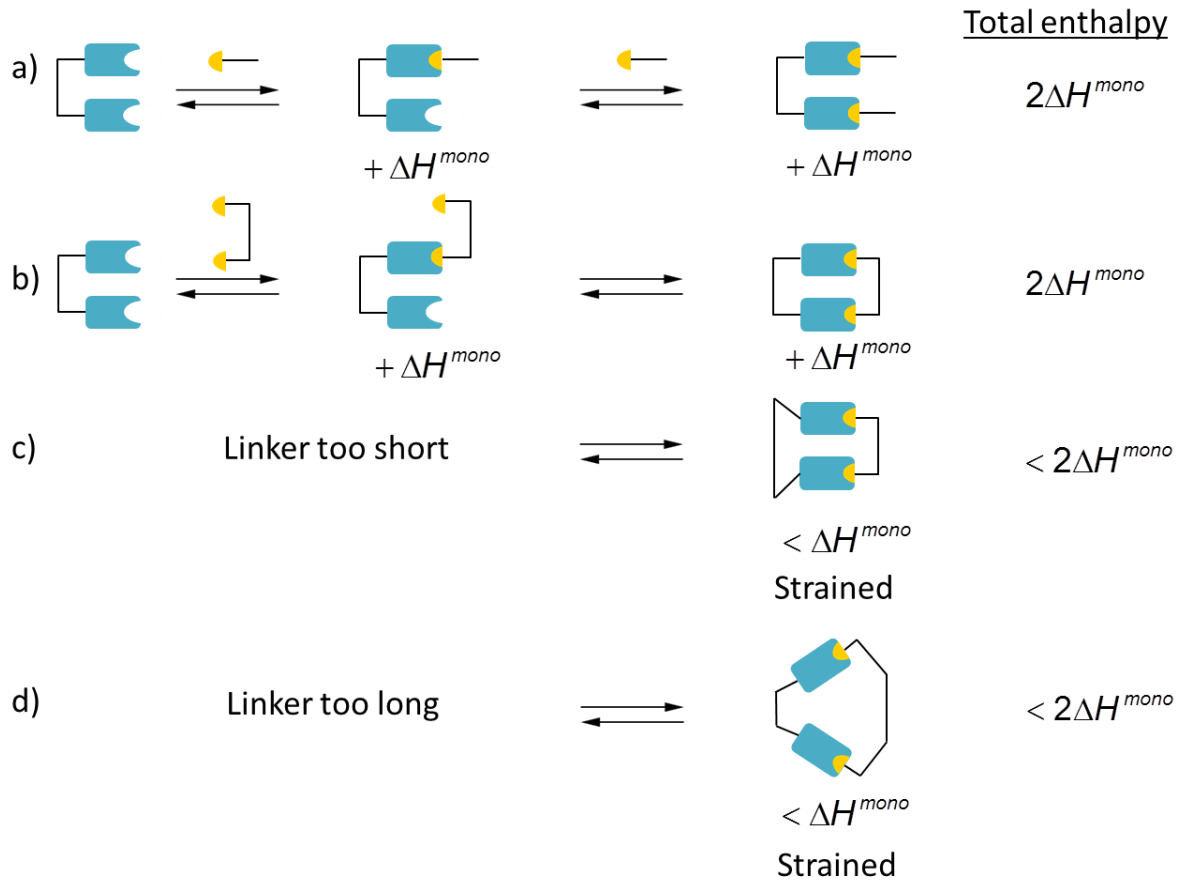



**Figure 2.3.** Effect of special orientation on the binding enthalpy.

To optimise the multivalent effect, ligands have been designed with a long flexible linker to overcome the negative effects of the enthalpically diminished binding. However the binding for these ligands was observed to be equal or slightly more favourable than the corresponding monovalent ligand. This is due to poor understanding of the entropy in the linker design.

The total entropy of a multivalent system ($\Delta S_N^{\text{Poly}}$) (Equation 2.11) consists of changes in the translational ($\Delta S_{\text{Trans},N}^{\text{Poly}}$), rotational ($\Delta S_{\text{rot},N}^{\text{Poly}}$), and conformational entropies ($\Delta S_{\text{conf},N}^{\text{Poly}}$) of receptor and ligands, plus the entropy changes of the surrounding water ($\Delta S_{\text{H}_2\text{O},N}^{\text{Poly}}$).[9]

$$\Delta S_N^{\text{Poly}} = \Delta S_{\text{Trans},N}^{\text{Poly}} + \Delta S_{\text{rot},N}^{\text{Poly}} + \Delta S_{\text{conf},N}^{\text{Poly}} + \Delta S_{\text{H}_2\text{O},N}^{\text{Poly}} \qquad \text{(Eq. 2.11)}$$

The translational contribution to the entropy originates from the ability of a molecule to travel freely through space. The value of $\Delta S_{\text{Trans}}$ is related to the logarithm *M*, ($\Delta S_{\text{Trans}} \propto \ln(M)$) and inversely related to the logarithm of the ligand concentration ($\Delta S_{\text{Trans}} \propto \ln([\text{L}])^{-1}$). The rotational entropy results from the freedom of a molecule to rotate around its three principle axes and is related logarithmically to the product of the three principle moments of inertia, $I_{\text{x}}$, $I_{\text{y}}$ and $I_{\text{z}}$ ($\Delta S_{rot} \propto \ln(I_x I_y I_z)$). This logarithmic (and weak) dependence on the rotational and translational entropy values for its mass and dimensions results in the approximation that the values can be treated as constants in the multivalent binding process.[9]

To understand the conformational entropy involved with multivalent binding, one starts from the case where the conformational entropy is equal to zero ($\Delta S_{conf} = 0$). The starting point is the comparison of two monovalent interactions and one bivalent interaction where the ligands and receptors are linked via a rigid linker orientating the binding sites and ligands in the optimal orientation for bivalent binding. The entropy for the two monovalent interactions (not taking into account the hydration entropy) is $2\Delta S_{\text{Trans}}^{\text{mono}} + 2\Delta S_{\text{rot}}^{\text{mono}}$ (Figure 2.4). The total entropy in the case with the rigid linkers is equal to $\Delta S_{\text{Trans}}^{\text{mono}} + \Delta S_{\text{rot}}^{\text{mono}}$ using the approximations mentioned above for the translational and rotational entropy. This results in the entropy change of a bivalent binding process being half of that of the monovalent equivalent. Once the first binding event has occurred, the second event arises without any entropic cost. This second event is entropically enhanced and $\Delta G_{\text{second}} \approx \Delta H_{\text{mono}}$ as $\Delta H_{\text{second}} \approx \Delta H_{\text{mono}}$ and $\Delta S_{\text{second}} \approx 0$.[9]

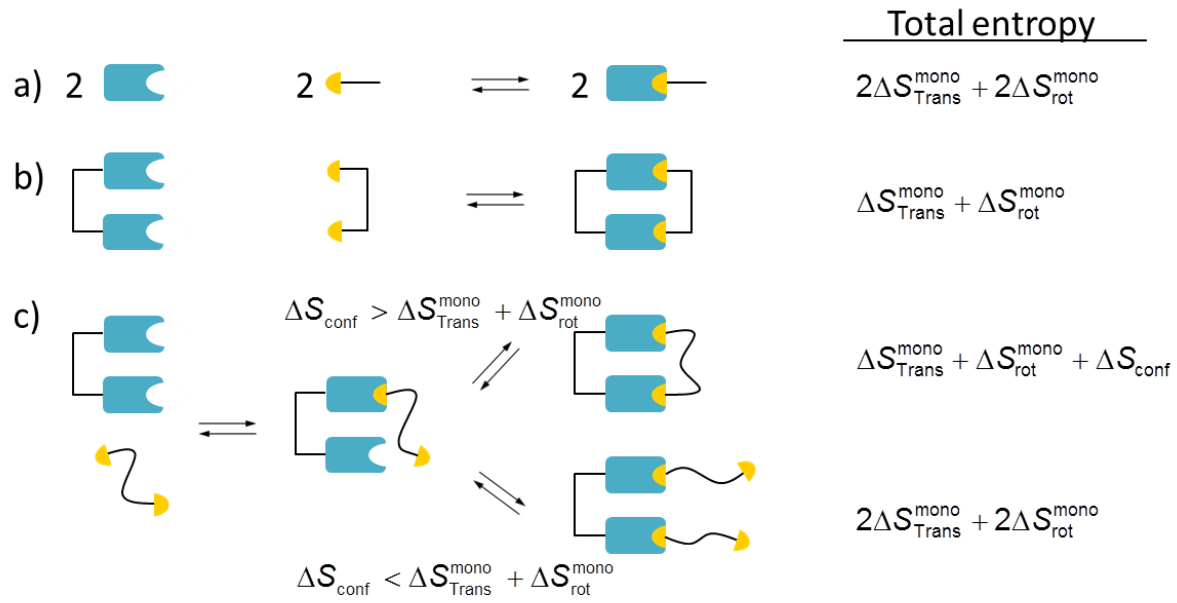



**Figure 2.4.** Influence of a linker on the entropy of multivalent systems.

In nature, the linker molecule connecting the receptors and ligand is often slightly flexible, resulting in the conformational entropy being less than zero ($\Delta S_{conf} < 0$). When complexation between receptor and ligand occurs, the number of available conformations for the bivalent ligands and the bivalent receptors reduces drastically, resulting in an unfavourable entropy contribution to the binding event. If the conformational cost is less than the sum of the rotational and the translational entropy cost ($\Delta S_{conf} > \Delta S_{Trans}^{mono} + \Delta S_{rot}^{mono}$), then the binding is still entropically enhanced (Figure 2.4.b). If the conformational entropy increases and becomes equal to the sum of the rotational and translational entropy, the binding becomes entropically neutral ($\Delta S_{conf} = \Delta S_{Trans}^{mono} + \Delta S_{rot}^{mono}$) where the complex formation of the second ligand from the same dimer is equal in entropic cost as the complex formation of a second bivalent molecule (Figure 2.4.c). When the

conformational entropy is even bigger than the sum of the rotational and translational entropy, it is more favourable to complex two bivalent molecules instead of forming a complex with one bivalent ligand (Figure 2.4.c). The bivalent binding becomes entropically diminished ( $\Delta S_{conf} < \Delta S_{Trans}^{mono} + \Delta S_{rot}^{mono}$ ).[9]

Enthalpy and conformational entropy can partially compensate each other, when the linker connecting two ligands gets longer and more flexible, the more likely negative effects of strain in the linker molecule decreases the enthalpy loss for the binding of the second ligand, making it more favourable. However, the longer the linker molecule gets, the higher the loss of conformational entropy upon binding. This makes the design of a multivalent ligand somewhat problematic to optimise. Values for the entropy loss for converting a single rotational carbon – carbon bond into a rigid non-rotatable bond have been estimated at 0.5 kcal $mol^{-1}$, whilst other single bonds estimates are in the range of 0.1 to 1.0 kcal $mol^{-1}$. The highest estimate for the loss in conformational entropy will arise when every bond loses its rotational freedom upon binding and can be valued at 0.5 *N* kcal $mol^{-1}$, where *N* is the number of rotatable bonds in the linker. For example, for a tri-ethylene glycol linker, where every bond is rotatable, the conformational entropy loss can be as high as 10 kcal $mol^{-}1$.[9,27, 28]

The final contribution to the entropy is the entropy associated with the water molecules surrounding the ligands and the binding sites ($\Delta S_{H_2O}$). In multivalent systems, the value of the solvation entropy is almost the same as the solvation entropy of the corresponding monovalent binding events. This results in a negative value of $\Delta S_{H_2O}$ for every receptor – ligand pair added to the total entropy, which is irrelevant whether it is part of a multivalent interaction or not.[9]

## 2.5 Multivalency according to Kitov:

In 2003, Kitov[29] and co-workers presented a new approach on the thermodynamics of multivalency. The starting point is based on the work done by Jencks in 1981 on additivity of free energies.[30] Jencks started from a molecule with two different binding pairs, with every pair having its own intrinsic binding energy ($\Delta G^{i}_{A}$ and $\Delta G^{i}_{B}$ for binding pairs A and B). From experimentation, Jencks observed that the total free energy of binding ($\Delta G^{O}_{AB}$) was different to the sum of the two single intrinsic binding energies, with the "connection Gibbs energy" ($\Delta G^{s}$) compensating for this difference. According to Jencks, this term represents the changes involved with the translational and rotational entropy. Now it is known that there is also conformational entropy and an enthalpy contribution involved in this term, ($\Delta G^{s}$).[30]

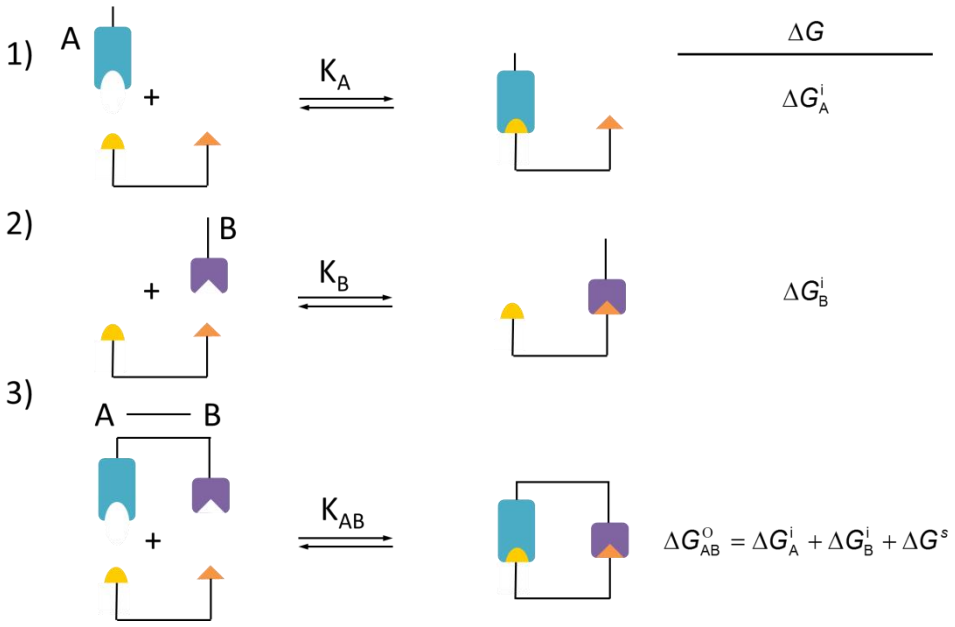



**Figure 2.5.** Gibbs free energy for multivalent systems.

Kitov and co-workers began from the following Equation 2.12 for the free energy ($\Delta g^{O}_{multi}$) of formation of the complex rl(*i*), which he adapted for *i* multiplied by the same ligand – receptor rl(*i*) pair resulting in:

$$\Delta g^{O}_{multi} = i\Delta G^{O}_{mono} + \Delta G^{O}_{interaction} \qquad \text{(Eq. 2.12)}$$

Where $\Delta G^{O}_{mono}$ is the binding energy of a monovalent interaction between a binding site and a monovalent ligand, *i* is the valency of the system and $\Delta G^{O}_{interaction}$ allows for the favourable and the unfavourable effects of linking the ligands together. For further thermodynamic analysis of the multivalent binding, Kitov rearranged the different contributions to the total free energy according to the origin of the interaction, either being the initial intermolecular free energy ($\Delta G^{O}_{inter}$) or the subsequent intramolecular free

energy( $\Delta G^{O}_{intra}$ ). The initial binding event between the ligand and the receptor is intermolecular and monovalent free energy ( $\Delta G^{O}_{mono}$ ) and is equal to the intermolecular free energy ( $\Delta G^{O}_{inter} = \Delta G^{O}_{mono}$ ). The sum of all the free energies contributions due to the intramolecular binding events is given by Equation 2.13.

$$(i-1)\Delta G^{O}_{intra} = (i-1)\Delta G^{O}_{mono} + \Delta G^{O}_{interaction} \qquad \text{(Eq. 2.13)}$$

The sum of all the free energy ( $\Delta g^{O}_{i}$ )associated with a microscopic complex of one multivalent receptor molecule with one multivalent ligand molecule rl(*i*) is described by Equation 2.14:

$$\Delta g^{O}_{i} = \Delta G^{O}_{inter} + (i-1)\Delta G^{O}_{intra} \qquad \text{(Eq. 2.14)}$$

Every binding unit of a ligand or a receptor has the same binding properties and all microscopic complexes rl(*i*) with *i* binding sites engaged have the same free energy. All these microstates together can be represented by a macroscopic complex RL(i). In order to translate Equation 2.14 from the microscopic world to a macroscopic world, a degeneracy coefficient *Ωi* is required. This represents that complex RL(*i*) is not an individual molecule but an ensemble of $\Omega_i$ microscopically distinguishable complexes rl(*i*). The free energy ( $\Delta G^{O}_{i}$ )of the macroscopic complex RL(*i*) becomes Equation 2.15:

$$\Delta G_{\mathrm{i}}^{\mathrm{O}} = \Delta G_{\mathrm{inter}}^{\mathrm{O}} + (i-1)\Delta G_{\mathrm{intra}}^{\mathrm{O}} - RT \ln \Omega_i \qquad \text{(Eq. 2.15)}$$

The free energy of every microstate follows a Boltzmann like distribution law. Since the macroscopic complex RL(*i*) is represented by a degenerate free energy level ($\Delta G_{\mathrm{i}}^{\mathrm{O}}$), the probability of an *i*th bound level to be found over any of all the other bound states is given by Equation 2.16:

$$w_i = \frac{e^{-\Delta G_i^{\mathrm{O}}/RT}}{\sum_{i=1}^{i_{\max}} e^{-\Delta G_i^{\mathrm{O}}/RT}} \qquad \text{(Eq. 2.16)}$$

The weight coefficient $w_i$ for every *i*th state is proportional to the contribution of every *i*th state to the overall free energy of the complex RL.[29]

Definitions for affinity constants for multivalent systems are not well defined and vary greatly from publication to publication.[1, 4, 9, 29] Often there are different names used for the affinity constant, ranging from "effective", "functional", "pseudo", "observed" and "apparent" depending on how the affinity constant has been recorded. Kitov and co-workers defines the affinity constant (Equation 2.18) according to the unambiguous thermodynamic definition as the ratio between reactants and products. Here, the products are the sum of the concentrations of all possible complexes.[29] This thermodynamically defined parameter is called the

avidity binding constant and this definition[29] is different from the affinity constant proposed (or used) by Whitesides and co-workers where $K_{avidity}$ is defined as $(K_{avidity})^N$.[9]

$$R + L \rightleftarrows RL(1) + RL(2) + RL(3) + \cdots + RL(i) \qquad \text{(Eq. 2.17)}$$

$$K_{avidity} = \frac{\sum[RL(i)]}{[R][L]} = \frac{[R_{bound}]}{[R][L]} = \frac{[R]_O - [R]}{[R][L]} \qquad \text{(Eq. 2.18)}$$

where $[R]$, $[R_{bound}]$ and $[R_O]$ are the unbound, bound and total receptor concentrations respectively.

Equation 2.18 can be rewritten as a sum of all *i* affinity constants ($K_i$)for every complex possible and becomes Equation 2.19.

$$K_{avidity} = \sum_{i=1}^{i_{max}} K_i \qquad \text{(Eq. 2.19)}$$

Conversion of Equation 2.19 into free energy terms ( $\Delta G^{O}_{avidity}$ ) gives following Equation 2.20:

$$\Delta G^{\circ}_{avidity} = -RT \ln \sum_{i=1}^{i_{max}} K_i = -RT \ln \sum_{i=1}^{i_{max}} e^{-\Delta G_i^{O}/RT} \qquad \text{(Eq. 2.20)}$$

Using the definition for the weight coefficient $w_i$, Kitov managed to separate Equation 2.20 into three terms, two of them corresponding to the inter- and intramolecular free energy, and one representing the degeneracy of multivalency.

$$\Delta G^{\circ}_{\text{avidity}} = \Delta G^{\circ}_{\text{inter}} + \Delta G^{\circ}_{\text{intra}} \sum_{i=1}^{i_{\max}} w_i (i-1) + RT \sum_{i=1}^{i_{\max}} w_i \ln(w_i / \Omega_i) \qquad \text{(Eq. 2.21)}$$

Equation 2.21 summarises all the thermodynamic components of multivalent binding. All enthalpic effects are represented by the first two terms, the enthalpy contribution due to inter- and intramolecular binding energies. The first term of Equation 2.21, $\Delta G^{\circ}_{\text{inter}}$, is similar to the free energy of monovalent binding, whereas the second and third term represent the multivalent effect, the additional specific interaction and the effect of a statistical factor. The maximum value for second term of Equation 2.21 depends on the number of additional interactions but will only be approached asymptotically as the number of branches of the ligand increases.[29]

The third term of Equation 2.21, $-RT \sum_{i=1}^{i_{\max}} w_i \ln(w_i / \Omega_i)$, agrees with a generalized form of the Boltzmann-Gibbs definition of entropy. This entropy definition is unique for multivalency, and is called the avidity entropy,

$\Delta S^{\circ}_{avidity}$, and is a measurement of disorder in the distribution of microscopic distinct complexes. The avidity entropy is always positive and favours association of multivalent ligands and receptors. The third term becomes significantly more important as the number of branches in the ligand and binding sites on the receptor molecule increases.[29]

## 2.6 Effective concentration

The concept of effective concentration ($C_{EFF}$) was introduced by Kramer and Karpen[31] and is depicted in Figure 2.6. Mulder and co-workers[32] further developed the concept by devising a theory around it. Again, the multivalent binding event must been seen as a two-step process. The first step is an intermolecular interaction between a ligand molecule and a receptor molecule. The second step is an intramolecular interaction. Following the first interaction, the fact that the ligands moieties are linked together forces the second ligand moiety to be proximate to a second receptor moiety. This gives rise to a local high concentration, called the effective concentration. [31, 32]

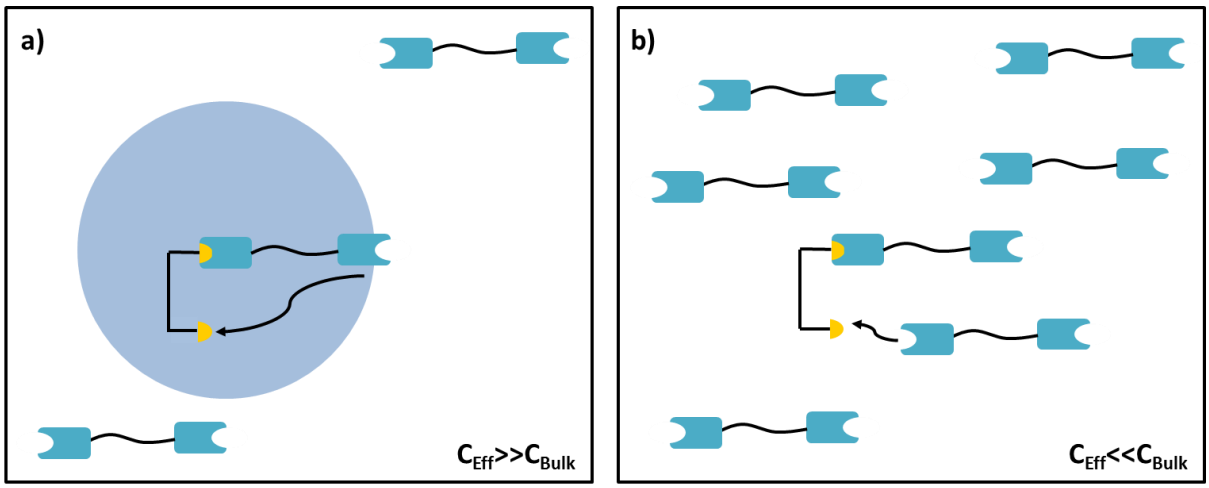



**Figure 2.6.** Effect of the effective concentration upon multivalent binding.

If $C_{EFF}$ is higher than the bulk concentration ($C_{Bulk}$), intramolecular interaction is more likely to occur and multivalency is favoured (Figure 2.6.a).[32] The effective concentration represents the concentration of the ligand as well as that of the receptor. When $C_{Bulk}$ is higher than the effective concentration (Figure 2.6.b), the intermolecular interactions will be favoured resulting in a complex with two different ligand molecules and one receptor instead of multivalent binding.[32]

The effective concentration depends on the length of the linker between the two ligands. The longer the linker, the lower the effective concentration, but if the linker is too short, such that the second ligand cannot reach the receptor, the effective concentration can be reduced to zero. Also, the flexibility of the linker has an influence on the size of the sphere, and rigid linkers can result in a different complex shape instead of a sphere. The rigidity of the linker can significantly reduce the volume of the "sphere",

which results in an increase in effective concentration. Here again, if the linker becomes too rigid such that the ligand cannot reach the receptor, the effective concentration is reduced to zero. This makes it extremely difficult to calculate or predict effective concentration for a designed ligand. This theory hides also the enthalpic and entropic contributions for multivalent binding discussed earlier, not allowing design and prediction of affinities for new ligands. [2, 32]

## 2.7 Kinetics of multivalent binding

Idealising all previous concerns, an estimate of reaction kinetics can be made. For a bivalent binding process, the first binding event (Figure 2.7b) would be four times more likely than its monovalent counterpart (Figure 2.7a). Hence, the affinity constant of the bivalent event ($K_a^{multi}$), is four times the affinity constant of the monovalent event ($K_a^{mono}$), Equation 2.22.

$$(K_a^{multi} = 4K_a^{mono}) \qquad \text{(Eq. 2.22)}$$

The association constant ($K_a^{multi,2}$) associated with the second binding event is half of the product of the effective concentration and the monovalent affinity ($K_a^{mono}$), Equation 2.23.

$$K_a^{multi,2} = \tfrac{1}{2} C_{Eff} K_a^{mono} \qquad \text{(Eq. 2.23)}$$

Overall the affinity constant for a bivalent system becomes Equation 2.24:

$$K_a^{multi} = 2C_{Eff}(K_a^{mono})^2 \qquad \text{(Eq. 2.24)}$$

Or in general, an *n*-valent binding event is given by Equation 2.25:

$$K_a^{multi} = b(C_{Eff})^{n-1}(K_a^{mono})^n \qquad \text{(Eq. 2.25)}$$

Where *b* is a statistical value, representing the ratio of permutations leading to product over the permutations of the reverse reactions.

An estimate for the effective concentration was proposed by Mulder and co-workers[82] using the formula for cyclisation probability (Equation 2.26)

$$C_{\text{Eff}} = \frac{1}{N_{\text{AV}}}\left(\frac{3}{2\pi\bar{r}_0^2}\right)^{3/2} \qquad \text{(Eq. 2.26)}$$

where $N_{\text{AV}}$ is Avogadro's number and $\bar{r}_0$ is the root mean squared distance between the free ligand and the free receptor moiety.

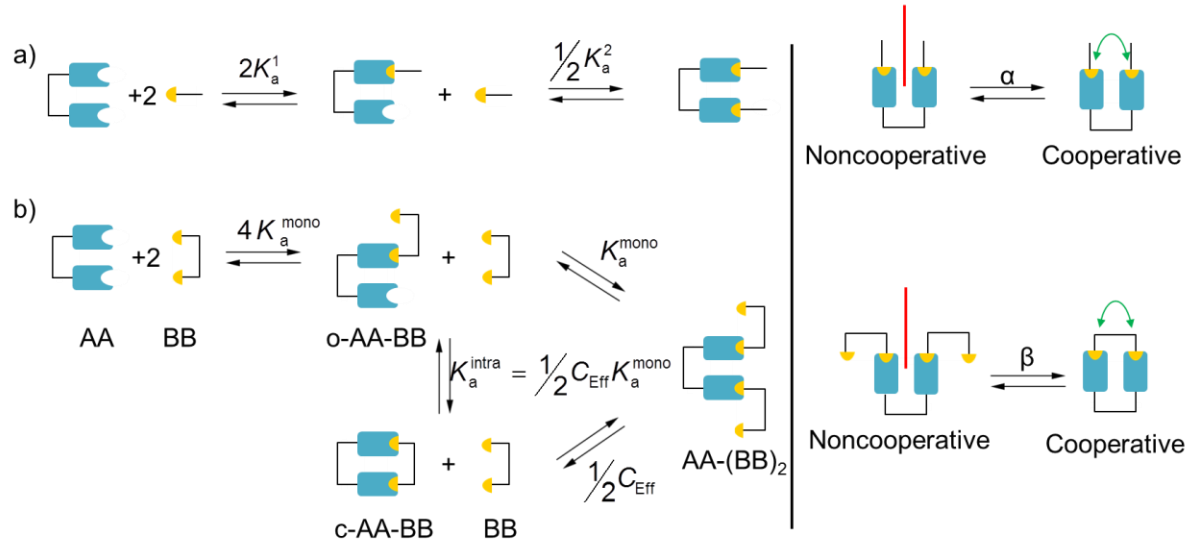



**Figure 2.7.** Kinetic approach to multivalency.

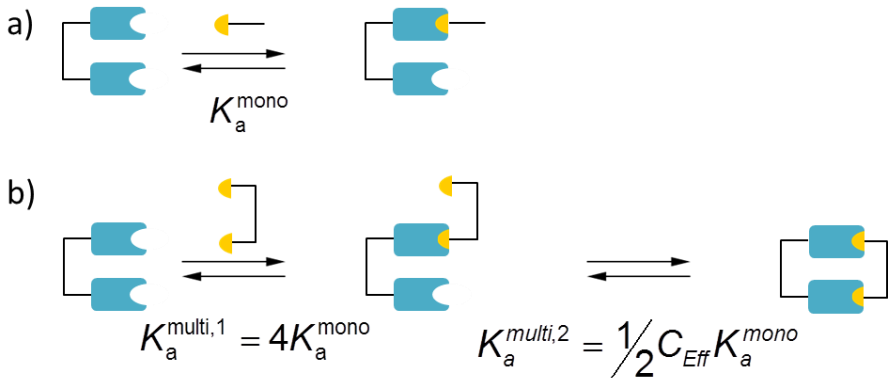



**Figure 2.8.** Multivalency only appears when the halftime of the monovalent binding is long enough.

Multivalent binding can only been achieved if the half-life of the monomerically bound complex is large enough. This intermediate could be seen as a "steady state", and its life time depends on the dissociation rate constant ($K_d^{mul,1}$) and the association constant ($K_a^{mul,2}$). If the dissociation of

the monomerically bound complex is too fast, the complex will dissociate before the influence of the higher effective concentration can result in a multivalent complex. Where $K_a^{mul,2}$ is very large, the steady state will be too short lived to experience any benefits of the higher effective concentration (Figure 2.8).[2]

## 2.8 Cooperativity according to Anderson

Anderson[3] postulates that positive cooperativity leads to an "all-or-nothing" behaviour. With strong cooperativity, only the extreme states are significantly populated as can be seen in speciation profiles (Figure 2.9). One species dominates under certain conditions, until a point where the conditions change and a sharp transition occurs from one species to another with very little species found in any intermediate form. All-or-nothing behaviour caused by positive cooperativity is very common in nature[3] and results in "on-off" behaviour by a small change in conditions.

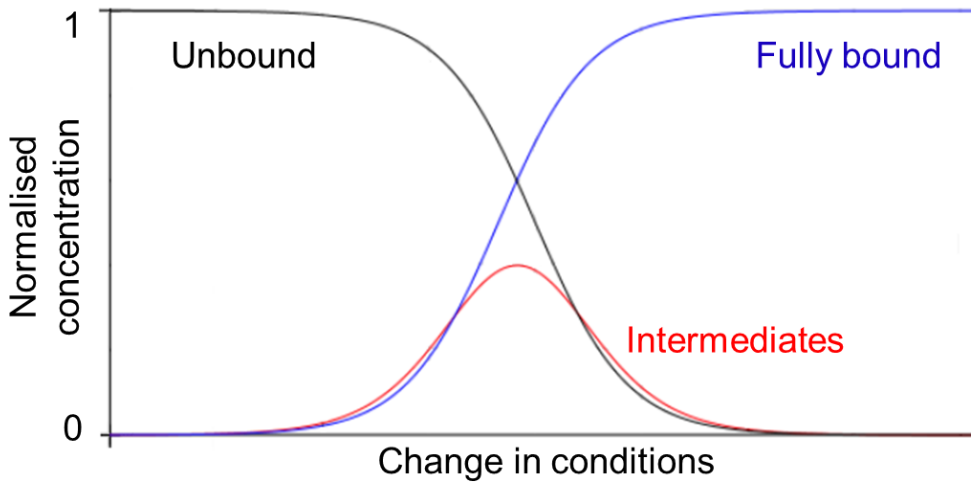



**Figure 2.9.** "All or nothing" behaviour.

Positive chelate cooperativity has a different outcome, it has a nothing-all-nothing outcome as can be seen in the speciation profiles in Figure 2.10. Assume a situation where the concentration of the receptor AA is very low compared to the concentration of ligand BB, so no complex can be formed with more than one molecule of AA. There will be four states the receptor molecule AA can adopt, the free receptor AA, two 1:1 complexes - the open state o-AA-BB and the cyclic state c-AA-BB - and one 1:2 complex, state AA-$(BB)_2$ (Figure 2.7.b). Under the assumption that there is no allosteric cooperativity (α=1), the key equilibrium between the open and the cyclic 1:1 complex defines the effective concentration as Equation 2.27:

$$\frac{1}{2} K\, C_{Eff} = \frac{[c-AA-BB]}{[o-AA-BB]} \qquad \text{(Eq. 2.27)}$$

$C_{Eff}$ is independent of the ligand concentration and the product $K\ C_{Eff}$ determines the population of the cyclic complex c-AA-BB, wherein $K$ is the association constant associated with the binding of a monovalent receptor A with a monovalent ligand B. Depending on the value of $K\ C_{Eff}$, varying behaviours of the system can be observed.[3]

Where $K\ C_{Eff} \ll 1$, the open form o-AA-BB (red) is more stable than the closed form c-AA-BB (green), and at higher concentrations of ligand BB the system goes into an "all" AA-$(BB)_2$ (blue) (Figure 2.10c). The reference system with one monovalent ligand B bound to receptor AA is plotted in

black. This speciation profile approaches the profile of two monovalent ligands B binding to one receptor molecule BB (Figure not shown).

Where K $C_{Eff}$ ≫1, the cyclic complex c-AA-BB (green) is more stable than the open state o-AA-BB (red) (Figure 2.10a). The cyclic state is the dominating species ("all" c-AA-BB) and covers a wide concentration range while the open form is hardly populated. When the concentration of the ligand BB increases, a transition is made to the "all" 1:2 complex AA-$(BB)_2$ state (blue). The cyclic complex c-AA-BB opens to form the open 1:2 complex AA-$(BB)_2$ only when 2 $[BB]_0 > C_{Eff}$ , where $[BB]_0$ = [BB] + [c-AA-BB] + [o-AA-BB] + 2[AA-$(BB)_2$]. Therefore, $C_{Eff}$ defines the concentration at which simple monovalent intermolecular interactions compete with cooperative intramolecular interactions. By comparing the speciation profile of the closed complex (green line in fig 2.10.a) to the profile of the monovalent reference system (black line), it can be shown that intramolecular interactions leads to an increased stability at lower concentrations of ligand BB.[3, 33]

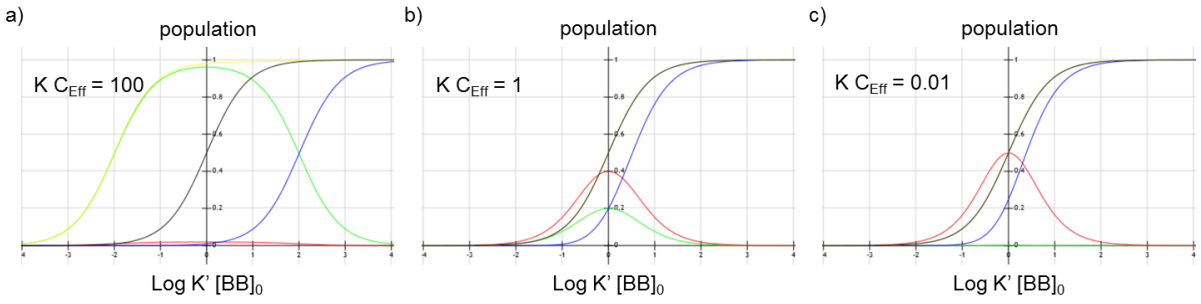



**Figure 2.10.** Speciation profiles at different effective concentrations.

## 2.9 Multivalency according to Ercolani

The last major publication on the theory about multivalency is the essay written by Ercolani[5] where he acknowledged the two types of cooperativity discussed in the literature so far, the allosteric and chelate cooperativity. This last form of cooperativity is the same as multivalency. According to Ercolani, there are still some major pitfalls in the understanding of allosteric cooperativity, and states that a new kind of cooperativity must be considered, interannular cooperativity arising from the interplay of two or more intramolecular binding interactions.

Allosteric cooperativity is relatively well understood, and the earlier definition of alpha applies. This cooperativity factor can be seen as an equilibrium constant for the conversion of the hypothetical complex with non-interacting sites (non-cooperativity) into the complex with interacting sites (cooperativity) See Fig 2.10.c.

The understanding of chelate cooperativity or multivalency is still open for discussion. According to Ercolani[5] and Anderson[3,33], the concentration of the species c-AA-BB is linear with the concentration of the ligand BB, whereas the species with two ligand molecules being part of the complex AA-$(BB)_2$ depends on the square of the concentration of the ligand. This leads to the conclusion that the chelate cooperativity is concentration dependent, in contrast with allosteric cooperativity. This has been overlooked in previous theories.

Ercolani[5] explains multivalency by saying that the Gibbs free energy of binding of a species AB to its corresponding receptor is larger than the sum of the species A and B to its receptor. However, this would assume additivity of the Gibbs free energy which was disapproved by Jencks. Jencks showed that the sum of two Gibbs free energies ($\Delta G^{O}_{AB} = \Delta G^{O}_{A} + \Delta G^{O}_{B}$) is equivalent to the product of two binding constants $K_A$ and $K_B$. The units for $K_A$, $K_B$ and $K_{AB}$, are $mol^{-1}L$, so the overall equation $K_{AB}$ ($mol^{-1}L$) = $K_A K_B$ ($mol^{-1}L$)$^2$ is meaningless.[5]

## 2.10 Isothermal titration

Brewer[34-39] and co-workers studied the interaction of a set of dendrimers presenting different numbers of tri-mannose moieties for the binding to Con A.[35-38] They started by mapping out which hydroxyl groups were important in binding to Con A by making deoxy-tri-mannose analogues and determining the affinity using isothermal titration calorimetry (ITC). The results of this study are in agreement with X-ray crystal structures, which Brewer used to justified the use of ITC in studying study lectin – carbohydrate interactions.[34]

Brewer further presented the tri-mannose to Con A in a multivalent way by using different dendrimers. Curvilinear Hill plots were observed, implying negative cooperativity. The authors explained that when more epitopes are in the bound state, the effective valency of the ligands decreases with every step in the multivalent interaction. This is directly opposite to the theory of

effective concentration. Brewer and co-workers identified some pitfalls as aggregates were observed due to crosslinking when concentrations of the ligands get high. ITC cannot distinguish between heat generated by binding events, and heat generated due to a transition in the protein allowing multivalent binding. By looking at the set of dendrimers, none of the linker molecules seemed long enough to allow bivalent binding on the same Con A molecule as the results of the molecular modelling show (see later).[36, 38] In later publications, Brewer offers further proof of the negative cooperativity, until 2005, where a similar set of dendrimers were used for crosslinking Con A molecules and multivalency was no longer mentioned.[35, 37]

# 3 Stochastic sensing

About 40 years ago an experiment was performed whereby a current was detected flowing through a single ion-conducting channel formed by the peptide antibiotic gramicidin on a planar lipid bilayer.[83] This was the first observation of a single molecular event. This experiment was the inspiration for the development of similar ion channels where "channel blockers" can reversibly bind in the lumen of said ion channels causing fluctuations in the current detected going through said ion channels.[84] This work formed the basis for stochastic sensing with engineered pores. Stochastic sensing is a process involving detecting a randomly determined sequence of observations each of which is considered as a sample of one element from a probability distribution. Stochastic sensing by the detection of single molecular events can therefore be used as an analytical technique. For this reason, different transmembrane pores have been engineered by genetic mutation and chemical modification to contain a binding site for analytes. These analytes reversibly occupy the engineered binding site causing characteristic fluctuations in the current going through the ion channel and these characteristic fluctuations can be used to determine the nature and quantity of the analyte present. Such engineered ion channels have been employed as rapid and sensitive biosensors for the detection of processes such as ligand gating, channel blocking, ion permutations, small cations or anions, organic molecules, proteins and DNA, but also in pharmaceutical screening. [40-45]

## 3.1 Experimental design

As described in the previous section, many theories exist about multivalent binding events. These theories are separate from each other and are difficult to combine. Since these theories have emerged, very few experiments gave been designed to challenge them or to illustrates their shortcomings. It is a common habit in the field of experimental physics to design experiments in an effort to validate a theory, or to make observations that cannot be explained by known known theories.

The work presented in this thesis is set out to advance the understanding of multivalent binding events. More specifically, it relates to multivalent binding events between proteins and their ligands, where the binding is specific and the species have to interact in a precise manner. The experiments look at multivalent binding events at a single molecular level and not at an ensemble of binding events. Therefore, properties of binding can be observed for one single binding event and not for an average of multiple binding events, this really helps to understand what happens during multivalent binding.[46]

The first steps in this field have been made by Hammerstein[43, 47], who looked at multivalent binding of ions at a single molecular level by using an ion channel-forming protein modified with two ligands. Shanley[57] further modified the system of Hammerstein to bind proteins in a specific way. Shanley prepared systems with single ligands to optimize variable such as linker length and the place of the modification.[57] Shanley finally presented

a bivalently modified ion channel-forming protein that was the starting point of the work done in this thesis.[57]

The aim of this thesis was to test the system presented by Shanley[57], and to obtain reproducible data for monovalent proteins, bivalent proteins, trivalent proteins and tetravalent proteins. From these data binding models can be derived, comprising every transition in the binding process.

A further aim is to provide a system with more complex ligands, and more specific more complex carbohydrates as ligands. Therefore, the work includes carbohydrate synthesis.

## 3.2 Single channel recording

For the following experiments, a technique known as single channel recording (SCR) was used[48]. These experiments are carried out in an artificial recording cell, comprising two recording chambers representing the intra and extra cellular compartments. These two chambers are separated by a polytetrafluoroethylene (PTFE) or Teflon® membrane. In this Teflon® membrane, an aperture is made with an approximate diameter of 100 µm. The aperture is where the artificial lipid bilayer is formed by folding two lipid monolayers together. The lipid monolayers form naturally on an air-water interface. By manipulating these lipid monolayers the lipid bilayer can be formed in the aperture. To stabilise the lipid bilayer in the aperture, the aperture is pre-treated with a hexadecane solution in pentane (10% v/v). After the pentane evaporates, the Teflon® membrane is covered

in a hydrophobic coating which causes a better addition of the lipids to the Teflon® membrane.[44]

To each chamber an aqueous recording buffer is added having a high enough salt concentration so that a current can be detected during the recording. To each amount of recording buffer in each chamber, an amount of lipid dissolved in pentane is added. After the evaporation of the pentane, the lipid monolayers form at the air-water interfaces. Sequential lowering of the buffer level below the aperture, and elevating the buffer level higher than the aperture, folds the artificial lipid bilayer in the aperture. The buffer solution in each chamber is then provided with an Ag/AgCl electrode connected to an ammeter.[44]

The artificial bilayer is an electrical insulator, meaning no current can be detected when a voltage is applied over the two chambers, Figure 3.1.a. In order to measure a current an ion channel-forming protein needs to be inserted in the lipid bilayer. An amount of detergent-solubilised ion channel-forming protein is added to one chamber, and the recording buffer is stirred until insertion occurs and a current can be detected. In all experiments in this thesis the ion channel-forming protein was added to the cis-chamber, meaning the chamber that is connected to the electrical ground. The other chamber is called the trans-chamber and is connected to the positive or negative potential of the power supply. A positive current as applied in the experiments means the movement of cations from the trans-chamber to the cis-chamber (see Figure 3.1.b).[44]

The final step in single channel recording is the addition of the analyte to the chamber where the binding site of the ion channel-forming protein is positioned (Figure 3.1.c). In all the experiments in this thesis the cis-chamber is the chamber where the analyte is added as the ion channel-forming protein mostly orientates itself in the bilayer with the sugar moieties pointing inwards towards the cis-chamber. [44, 49]

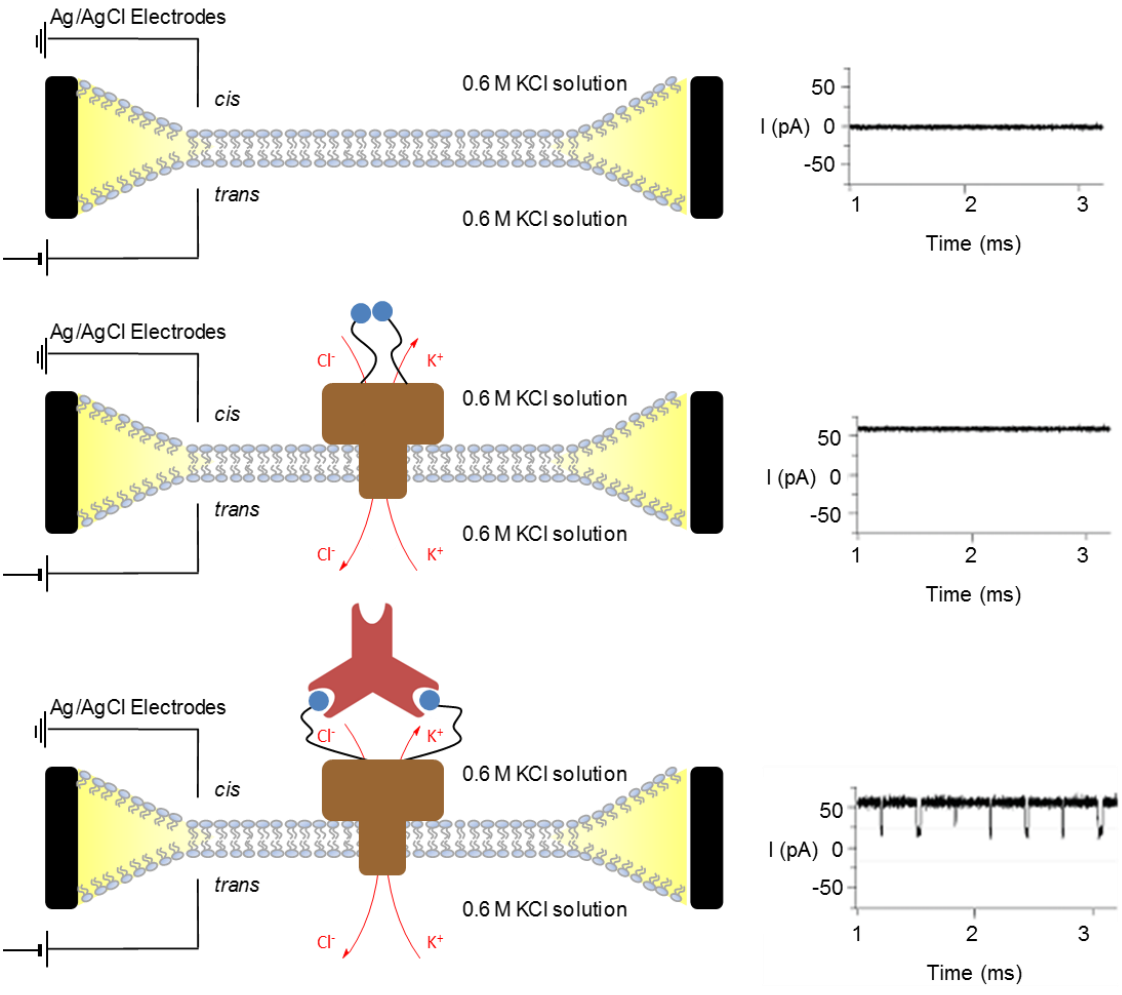



**Figure 3.1.** Working principle of single channel recording as a method for stochastic sensing.

## 3.3 α-Hemolysin

α-Hemolysin (α-HL) is a pore-forming toxin secreted by *Staphylococcus aureus*.[50] It is able to disturb the cell membrane allowing low molecular weight molecules to leak out the cell, damaging the osmotic balance and finally lysing the cells releasing nutrients from within the cell. These nutrients are available for the *Staphylococcus aureus* to grow and multiply. The α-HL ion channel itself is a heptamer, but is secreted as a 293 amino acid monomeric protein by the bacterium.[50] The 33.2 kDa water soluble monomer forms spontaneous heptameric homomers (240 kDa) in the presence of cell membranes. The shape of the heptamer resembles a mushroom, the cap is spacious and is hydrophilic, the hydrophobic stem is formed as a 14-stranded β-barrel and this is the part that can insert in a natural or artificial lipid bilayer. [50]

To study multivalent carbohydrate interactions, it was proposed to modify a pore with one or two carbohydrate molecules. These carbohydrate molecules could be linked to the pore via a water soluble linker molecule. It was proposed to use a polyethylene glycol linker. Three different linker lengths, with three, five and ten ethylene glycol moieties, were used to optimise the pore. It was proposed to use a disulphide bridge to covalently connect the sugar molecule to the ion channel allows later reduction of the disulphide bridge. This is useful as the carbohydrate molecules can cleave off by simple reduction, allowing be used as a control in later experiments.

## 3.4 Preparation of a monovalent pore

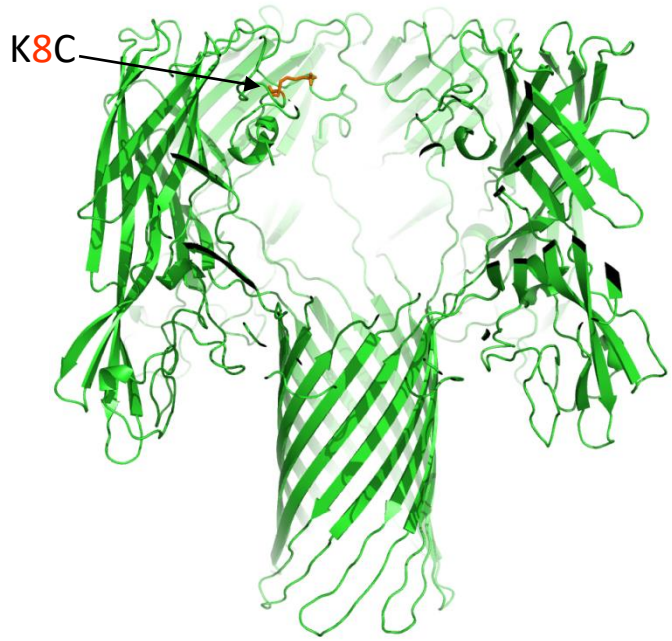



**Figure 3.2.** Site of modification in the α-HL heptamer.

The monovalent pore consists of six wildtype α-hemolysin (α-HL) monomer units and another monomer unit in which the lysine at position 8 is substituted for a cysteine (K8C) (Figure 3.2). This cysteine residue, containing a uniquely nucleophilic thiol group can be used in the 'tag and modify' strategy for functionalisation purposes. The thiol of this residue is used to form a disulphide bridge with another thiol group positioned at the terminus of a 10 unit ethylene glycol molecule linking a mannose moiety to the protein pore.[51-56] This last molecule, 1-(3, 6, 9, 12, 15, 18, 21, 24, 27-nonaoxa-29-methylthiosulphonate-29yl)-α-D-mannopyranoside (abbreviated to Man-EG-$SSO_2Me$), was prepared by Shanley.[57]

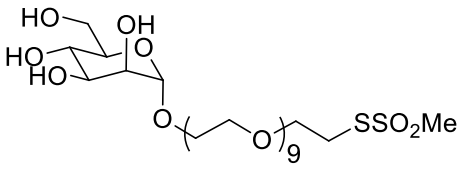



**Figure 3.3.** Man-$EG_{10}$-$SSO_2Me$.

To prepare the modified pores, wildtype monomers and K8C monomeric variants were produced by *in vitro* transcription translation (IVTT). An *E. coli* T7 S30 Extract System for circular DNA was used to produce $^{35}S$ methionine radiolabelled protein. These proteins were analysed by SDS-PAGE and photographic paper was used to visualize the bands. The protein was isolated by gel extraction of the desired band.

Once production of the K8C monomeric variant was complete, modification via the cysteine residue was carried out. The protein was incubated with Man-$EG_{10}$-$SSO_2Me$ at 25 °C for one hour. A control reaction was also carried out whereby a 25 mM solution of $mPEG_{5k}$-OPSS was added after this first reaction and incubated at 25 °C for 30 minutes.

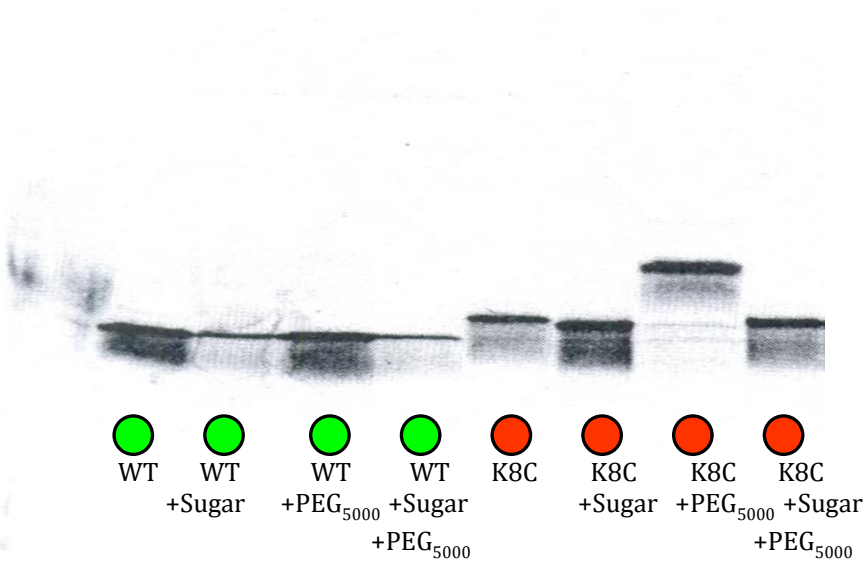



**Figure 3.4.** Native gel after chemical modification of the monomer and control reactions. WT stands for wildtype monomer, K8C stand for the α-HL monomeric variant with a cysteine in position 8, sugar stands for Man-$EG_{10}$-$SSO_2Me$ and $PEG_{5000}$ stand for $mPEG_{5k}$-OPSS.

The first four lanes in the gel in Figure 3.4 were control experiments carried out with the wildtype monomer (lacking a cysteine residue for modification purposes). The first lane contained only the wildtype monomer and in the second lane Man-$EG_{10}$-$SSO_2Me$ was added to the wildtype monomer and incubated for 1 hour at 25 ºC. In the third lane, m$PEG_{5k}$-OPSS was added to the wildtype monomer and incubated for 1 hour at 25ºC. If the $PEG_{5k}$ had reacted with the monomers, a considerable gel shift would be expected. In the fourth lane, Man-$EG_{10}$-$SSO_2Me$ was added to the wildtype monomer and after 30 minutes incubation at 25 ºC, m$PEG_{5k}$-OPSS was added for a further 30 minutes. As no gel shift was observed in any of the lanes, it can be concluded that in none of these cases the wildtype monomer was PEGylated, and also that no mannosylation occurred. This result was expected as the wildtype monomer contains no cysteine residues for modification.

The last four lanes in Figure 3.4 were the K8C monomeric variants that have been reacted. This K8C monomeric variant contains one cysteine residue and an aspartic acid tail. The reaction in lane 5 served as a control and contained only the K8C monomeric variant, whereas lane 6 contained the K8C monomeric variant following incubation with Man-$EG_{10}$-$SSO_2Me$. In the sample in the seventh lane, m$PEG_{5k}$-OPSS was added to the K8C monomeric variant and incubated for 1 hour at 25 ºC. Lane 8 contained the K8C monomeric variant that was incubated with Man-$EG_{10}$-$SSO_2Me$ for 30 min at 25 ºC followed by addition of m$PEG_{5k}$-OPSS for a further 30 minutes. In lane 7 of Figure 3.4, a band shift corresponding to a change in mass was

observed. Only a very faint band was observed at the level of the control in lane 8. From this, it can be concluded that the majority of the K8C monomeric variants was PEGylated. No shift was observed in the 6$^{th}$ lane of Figure 3.4. No shift was expected by the mannosylation of the K8C monomeric variants as the addition of a molecule of this size would not be expected to cause a significant change in charge over mass of the protein. There was no shift observed in lane 8 of Figure 3.4. Hence, it can be concluded that the mannosylation with mPEG$_{5k}$-OPSS was completed after 30 min, and that the further addition of mPEG5k-OPSS had no effect on the position of the band. The lack of reaction in this extra control experiment implies that the modification with the mannose-containing compound occurred via the introduced cysteine, and that the reaction was completed.

From these experiments, it can be concluded that the modification of the K8C monomeric variants with mPEG$_{5k}$-OPSS was successful. In an attempt to further validate the modification of the K8C monomeric variants, MALDI-TOF analysis was performed on three different variants as shown in Figure 3.5. The first mass spectrum (a.) was of the wildtype $\alpha$-HL monomer which has an expected mass of 17485.9 Da. The peak at 17491.7 Da was believed to represent said wildtype monomer. The second mass spectrum (b.) showed a peak at 17471.6 Da, which was expected to be the K8C monomeric variant, as it was in reasonably close proximity to the expected mass of 17473.4 Da. The peak at 17790.5 Da in third spectrum (c.) corresponds to the chemically modified K8C monomeric variant (expected mass 17790.5 Da).

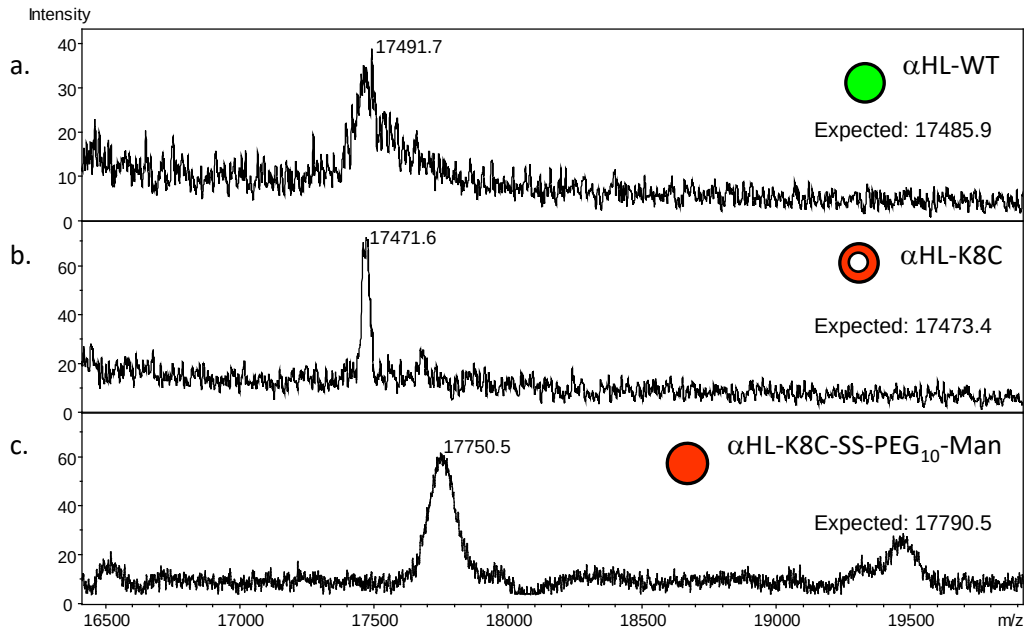



**Figure 3.5.** MALDI-TOF spectra for a.) the wildtype α-HL monomer, b.) the K8C monomeric α-HL variant, c.) the chemically modified K8C monomeric α-HL variant.

Figure 3.6 depicts the overall scheme followed to obtain the monovalent mannosylated α-HL pore or mannopore using the 'tag and modify' strategy. The 'tag' step involves the genetic mutation of lysine at position 8 to cysteine, such that the protein will be produced with a modifiable 'tag'. Also, an aspartic acid tail with 8 residues and 6 histidine residues was added for ease of purification later on in the process.

Both wildtype α-HL monomer and K8C monomeric variant were produced. The K8C monomeric variants were incubated for one hour with Man-$EG_{10}$-$SSO_2Me$ at 25 ºC to obtain the modified (mannosylated) monomer. The final step in the process was oligomerisation the form the mannopore. This was carried out by mixing 6 parts wildtype monomer with 1 part modified monomer (man-$EG_{10}$-K8C) in the presence of rabbit red blood cell membranes (rRBC) for 2 hours at 37 ºC. The presence of lipid membranes

was used to induce the oligomerisation of the α-HL monomers into a heptameric ion channel.

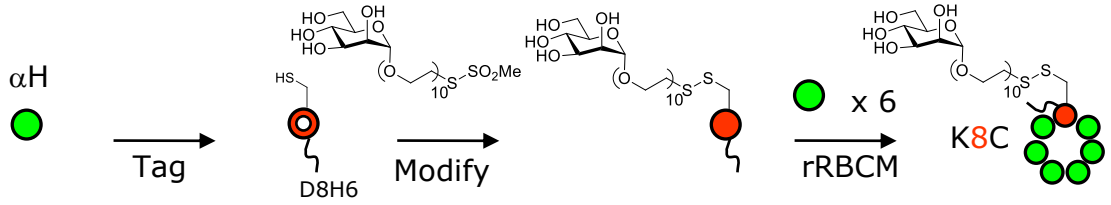



**Figure 3.6.** Overall reaction scheme to obtain a monovalent mannosylated pore, or monovalent mannopore.

The oligomerisation of the α-HL monomers is a stochastic event, and so different type of heptamers can be obtained. Oligomers comprising one, two, three or even four modified monomers may be obtained. The presence of the aspartic acid tail on the modified monomers allows us to separate these oligomers. The charged residues of a single modified monomer will cause a significant change in charge during oligomerisation depending on the number of modified monomers included in the heptamer. After oligomerisation, a native gel can be run, where oligomers with various numbers of modified monomers will separate, and the bands can be extracted, thus isolating pure oligomer. For the following experiments, only the oligomers with a single modification were isolated. Figure 3.7 shows the native gel in which the first lane is the wildtype oligomer, the second lane is result of the oligomerisation of 6 parts wildtype with 1 part of modified monomer, every different band is an oligomer with one more modified monomer in its structure. The third lane is the result of an oligomerisation of 5 parts wildtype α-HL monomer and 1 part of a modified

α-HL dimer, here again the different bands are different oligomers with a different number of modified monomers included.

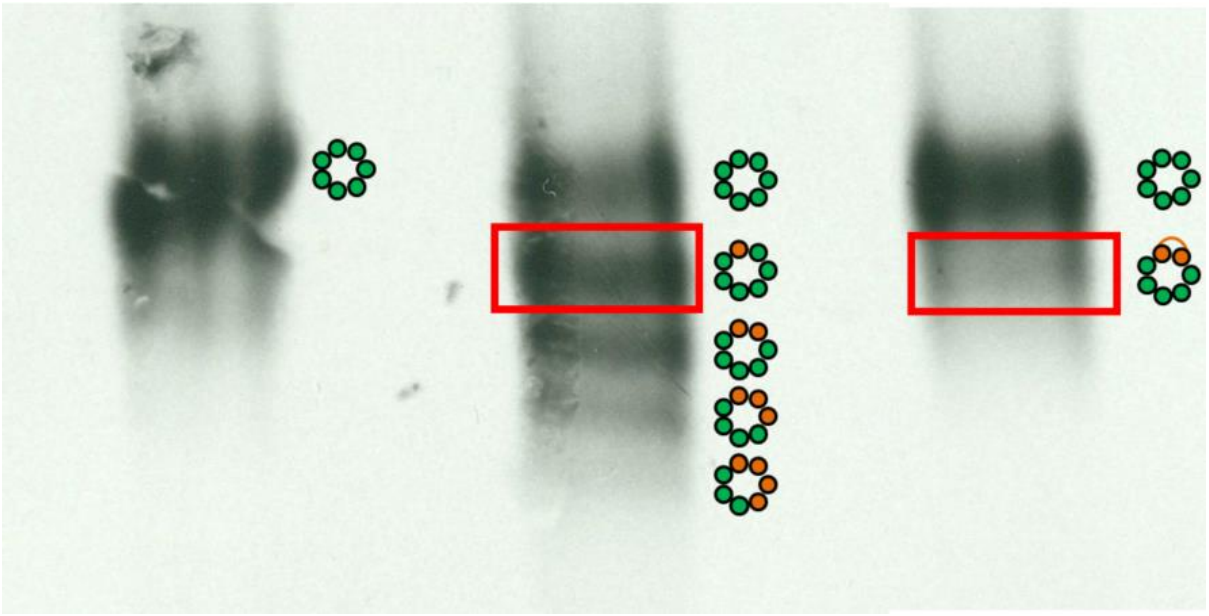


**Figure 3.7.** Native gel after oligomerisation of wildtype and modified monomers, first lane comprises wildtype heptamers, second lane comprises the monovalent mannopores and third lane comprises the bivalent mannopores.

## 3.5 Preparation of a bivalent pore

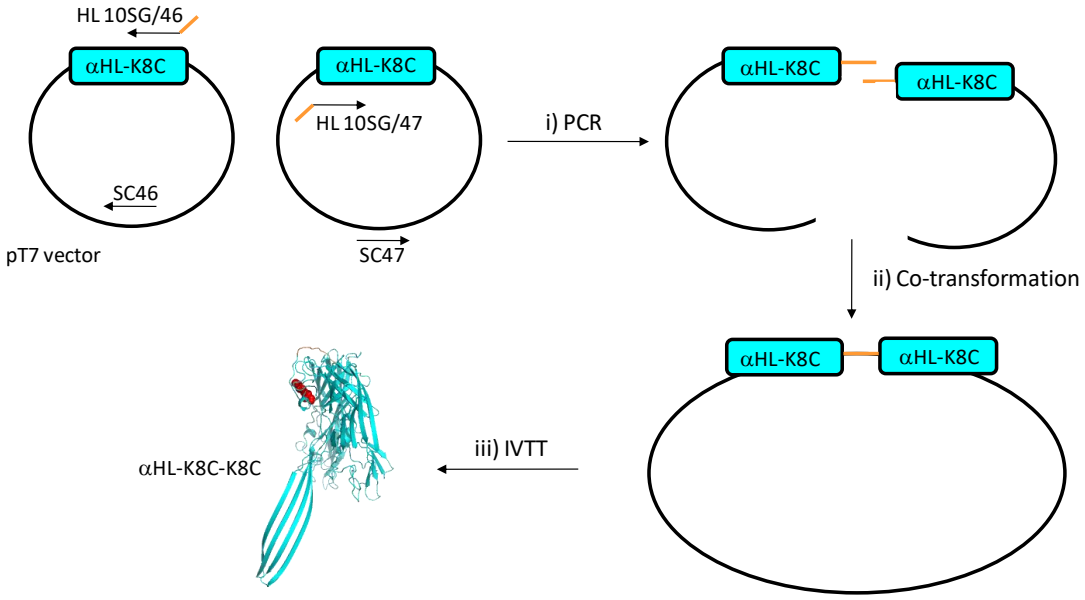



**Figure 3.8.** Preparation of the bivalent gene, linking two α-HL monomers together, by Hammerstein.

Hammerstein[47] prepared a gene that encodes for two α-HL monomer units linked together via a peptide linker as depicted in Figure 3.8. Because these two monomers are linked together via a short peptide linker to form a dimer, the two modified monomers will always be adjacent to each other upon oligomerisation. This gene encodes for the K8C variant of both of the linked monomers. α-HL dimer production was performed in a similar way to the production of α-HL monomers. The amount of dimer produced, is much lower than the amount of monomer produced under the same conditions.[47]

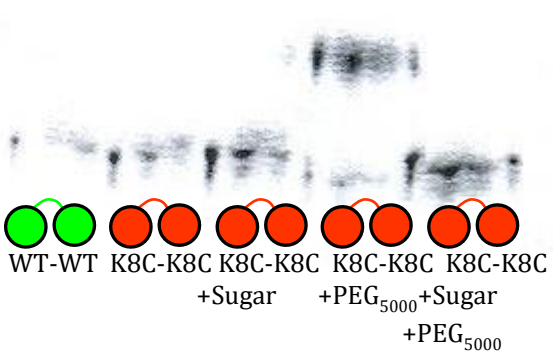



**Figure 3.9.** Native gel after chemical modification of the dimers and control reactions. WT-WT stands for wildtype dimers, K8C-K8C stand for the dimeric variant comprising cysteine at both positions 8, sugar stands for Man-$EG_{10}$-$SSO_2Me$ and $PEG_{5000}$ stand for $mPEG_{5k}$-OPSS.

The visualization technique for the native gel of the α-HL dimers and the control reactions is different from the visualization technique previously used for the α-HL dimers. Because the produced amounts of dimers are much lower than the produced amounts of monomers, it was difficult to visualise the gel under similar conditions. A gel loaded with a higher concentration of protein was ran and visualized by phosphor imaging (Figure 3.9). This technique is faster than using photographic paper as used for Figure 3.7, however the technique gives lower quality images. The first lane of the gel contains the wildtype-wildtype α-HL dimers, the last four lanes contain the K8C-K8C dimeric α-HL variants. Lane one contains the wildtype – wildtype control; in lane two the Man-$EG_{10}$-$SSO_2Me$ was added to the K8C-K8C dimeric α-HL variants and was incubated with for 1 hour at 25 ºC; in lane 3 the $mPEG_{5k}$-OPSS was added to the K8C-K8C dimeric α-HL variants and was incubated for 1 hour at 25 ºC. Lane four contains the reaction product of K8C-K8C dimeric α-HL variants first incubated for 30

minutes at 25 ºC with Man-$EG_{10}$-$SSO_2Me$ followed by a 30 minutes incubation at 25 ºC with $mPEG_{5k}$-OPSS. Here again the same conclusions can be drawn as compared to the modification of the K8C monomeric variants, only in lane 3 can a shift be observed due to the PEGylation. The absence of a shift upwards in lane 4 indicates full mannosylation after 30 minutes incubation. Hence, it therefore can be concluded that the dimer is fully mannosylated after the reaction. [51-58]

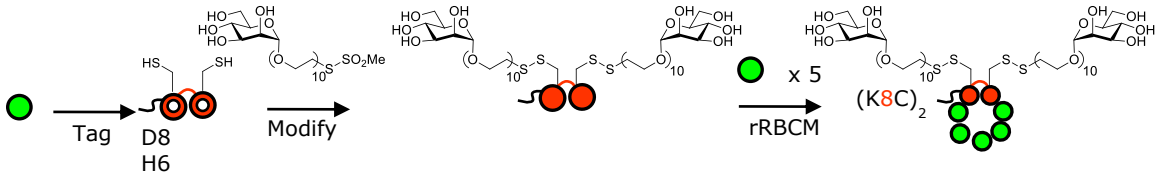



**Figure 3.10.** Overall reaction scheme to obtain a bivalent mannosylated pore, or bivalent mannopore.

The overall scheme followed to obtain a bivalent mannosylated pore or bivalent mannopore is shown in Figure 3.10. Starting from the wildtype plasmid, a K8C dimeric variant was produced including an aspartic acid tail and a histidine tail. After production of both the wildtype α-HL monomer and the K8C dimeric variant, the latter was incubated with Man-$EG_{10}$-$SSO_2Me$ for one hour at 25 ºC to obtain the mannosylated or modified α-HL dimer. In the last step five parts of wildtype α-HL monomer were mixed with one part of modified K8C α-HL dimer in the presence of rabbit red blood cell membrane for two hours at 37 ºC. This resulted in a mixture of different oligomers. Said different oligomers were separated by running a

native gel (lane 3 in Figure 3.7). For the purpose of the following experiments the heptamer comprising one modified dimer was isolated.

## 3.6 Current-voltage (IV) characteristic recordings

To find the optimal recording conditions, different buffers with a different pH were used and IV curves recorded. IV Curves in Figure 3.11 were measured for each different buffer condition to check the behaviour of the pore at the new conditions. From these measurements, it can be concluded that at a voltage of 75 mV and higher, spontaneous blockage of the pore occurs. Recordings at negative potential result in a significant increase in noise. Recordings were performed at a +50 mV potential, a condition which resulted in observing stable bilayer and pore, with a low level of noise.

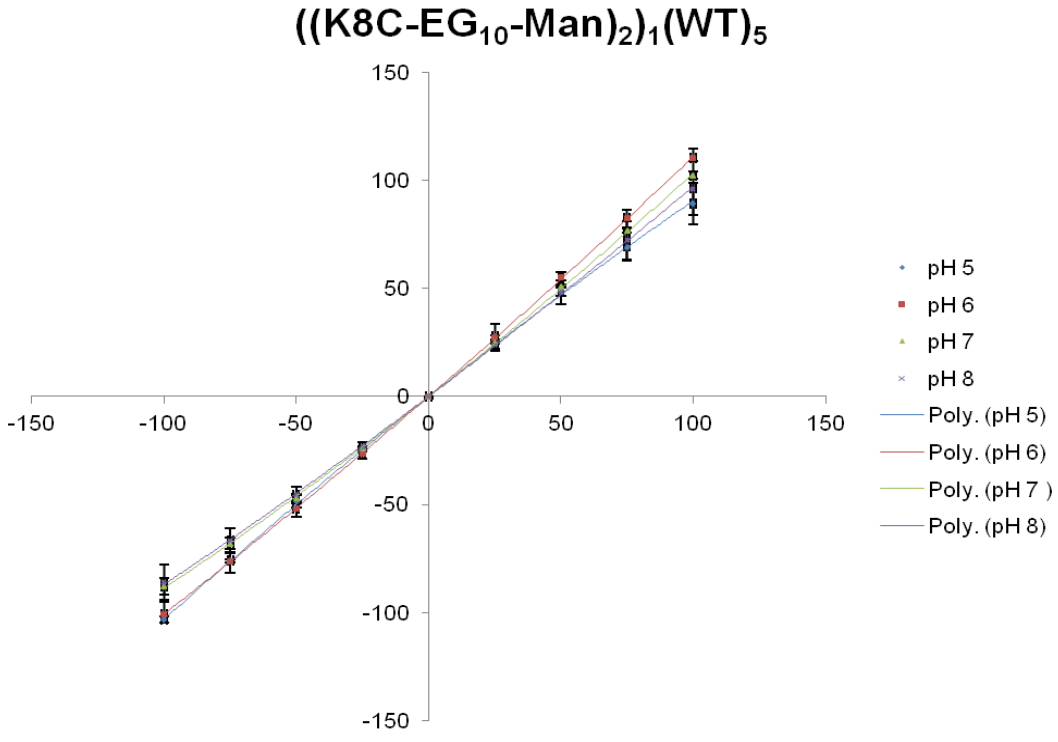



**Figure 3.11.** IV-Curves at different buffer conditions.

## 3.7 FimH control

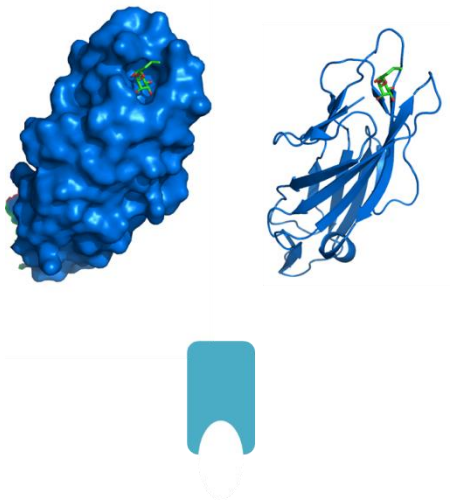


**Figure 3.12.** View of a crystal structure[59] of FimH and a cartoon representation of FimH as used herein.

Before measurements with the multivalent lectins, the behaviour of a monovalent lectin, FimH (Figure 3.12) was studied as a control to rule out effect of multiple proteins binding to one bivalent mannopore. FimH is a lectin, present on type I fimbrae of certain uropathgenic *E. coli* strains. FimH is specific for mannose and binds strongly to mannose-containing glycoproteins on the surface of epithelial cells in the bladder and in the urethra.[11, 59]

For this experiment a truncated form of FimH (containing only 156 amino acids out of 279) was produced in *E. coli*.[59] It has been previously shown that this truncated form of FimH is capable of binding to mannose, and further, unlike the full form is stable in buffered solution without any need for detergents. This truncated FimH protein was the smallest protein measured, but does have the highest affinity for mannose as reported in

literature. The reported values range from 2.3 μM for mannose to 5 nM for heptyl-α-D-mannose both measured via surface plasma resonance (SPR).[11, 59]

### 3.7.1 Production of $FimH_{1-156}$

A single colony of *E. coli* FimH1-156 pPKL241 (a pQE-30 vector) in *E. coli* HB101 (F'lacITN5) was picked from a LB-(amp) plate and used to inoculate 10 ml of LB media (100 μg $ml^{-1}$ ampicillin). This was incubated for 8 hours at 37 °C. This culture was used to inoculate LB (6 X 800 ml, 100 μg $ml^{-1}$ ampicillin). The cultures were incubated at 37 °C for 15 hours until A600= 0.8 a.u. Protein production was induced with ITG (isopropyl-β-D-thiogalactoside, 1mM) and the cultures were incubated for a further 90 min at 37 °C. The cultures were spinned down at 8000 X g for 10 min at 4 °C. Bacteria were harvested and the cell pellet was resuspended in 20 mM Tris-HCl, pH = 8.0, 0.5 mM EDTA, 30% sucrose (100ml) by stirring on ice for 10 min. Cells were harvested by centrifugation at 8000 X g for 15 min. Bacteria were lysed by osmotic shock with 100 ml ice-cold $MgSO_4$ (5 mM solution). Cells were centrifuged at 8000 X g for 15 min and the supernatant was loaded on 5 ml HisTrap HP(GE Healthcare) (equilibrated with 5 column volumes loading buffer: 20 mM Tris HCl, 200 mM KCl, 10 mM imidazole, pH = 7.5). The column was washed with 2 column volumes of loading buffer before the protein was eluted using a two-step gradient. The first step goes to 10 % eluting buffer (20 mM Tris HCl, 200 mM KCl, 300 mM imidazole, pH = 7.5), followed by 5 column volumes of 10% eluting buffer and a gradient to 100% eluting buffer over 10 column volumes. A flow rate of

5ml/min was used. Protein was detected by UV absorbance at 80 nm and the chromatogram is shown in Figure 3.13. Fractions of the flow through and eluted fractions were analysed by SDS-Page using a 12% Bis-Tris gel (Invotrogen) and MES running buffer (Invitrogen), 200V, 55 min (Figure 3.14). The relevant fractions were concentrated using a viva spin concentrator (3 kDa cut off, Sartorius Stedim) and buffer exchanged into a 60 mM Tris HCl pH 7.5 buffer. The protein concentration was measured using Nanodrop UV-Vis spectrophotometer, using extinction coefficient 37025 $cm^{-1}M^{-1}$, and a mass of 18 kDa. The protein was split into aliquots and snap frozen in liquid nitrogen and stored at -20 °C.[60]

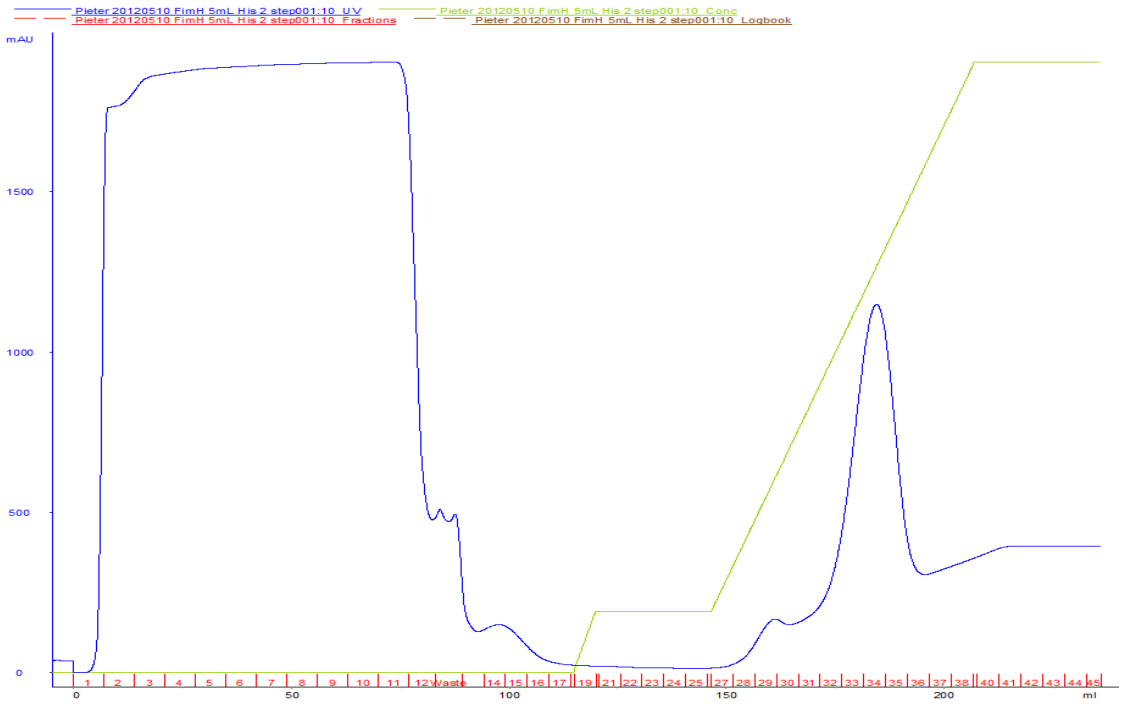



**Figure 3.13.** Chromatogram during purification of FimH.




**Figure 3.14.** SDS-PAGE from the crude and purified FimH protein.

Gene sequencing

The plasmid was submitted for gene sequencing using primers for the pQE-30 vector, the following protein sequence was obtained for the pre-protein:

AA sequence out of gene: 175 AA, Mw= 18864.001 Da

MGWSVNAWSF ACKTANGTAI PIGGGSANVY VNLAPVVNVG QNLVVDLSTQ
IFCHNDYPET ITDYVTLQRG SAYGGVLSNF SGTVKYSGSS YPFPTTSETP
RVVYNSRTDK PWPVALYLTP VSSAGGVAIK AGSLIAVLIL RQTNNYNSDD
FQFVWNIYAN NDVVVPTGGH HHHHH

The actual FimH protein sequence is the same sequence minus the first 9 amino acids.

AA sequence found: 166 AA, Mw= 17844.858 Da (-9 first AA).

F ACKTANGTAI PIGGGSANVY VNLAPVVNVG QNLVVDLSTQ
IFCHNDYPET ITDYVTLQRG SAYGGVLSNF SGTVKYSGSS YPFPTTSETP
RVVYNSRTDK PWPVALYLTP VSSAGGVAIK AGSLIAVLIL RQTNNYNSDD
FQFVWNIYAN NDVVVPTGGH HHHHH

### 3.7.2 Mass spectrometry

Figure 3.15 shows the mass spectrum of the purified protein. The observed mass is 17844.00 which is sufficiently close to the calculated mass of 17844.86 to assume that the correct protein was isolated.

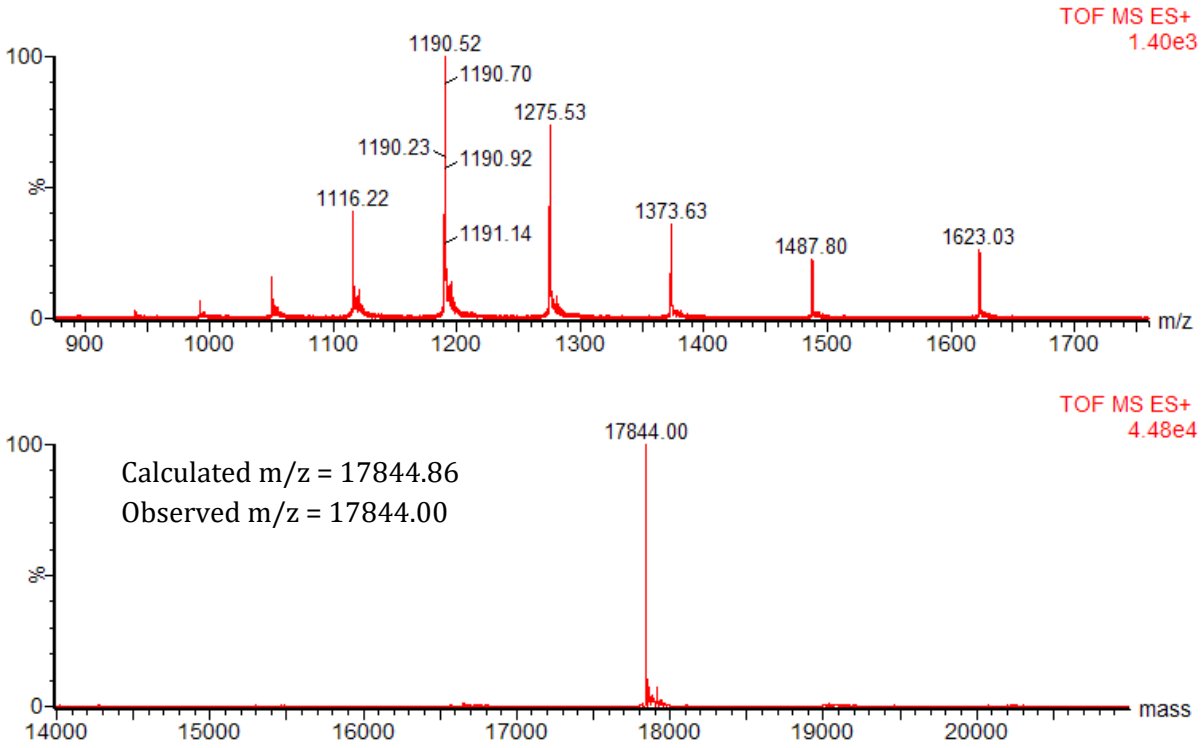



**Figure 3.15.** Mass spectrum of the produced and purified FimH protein.

### 3.7.3 FimH and monovalent mannopore

Figure 3.16 shows a section of a recording and an enlargement of some events. In this case, a mannopore carrying one mannose sugar was inserted into the artificial lipid bilayer and an amount of FimH protein was added to the buffer solution at the cis-side, meaning the side of the lipid bilayer connected to the cathode. The horizontal axis represents time and the vertical axis represents the current measured going through the pore.

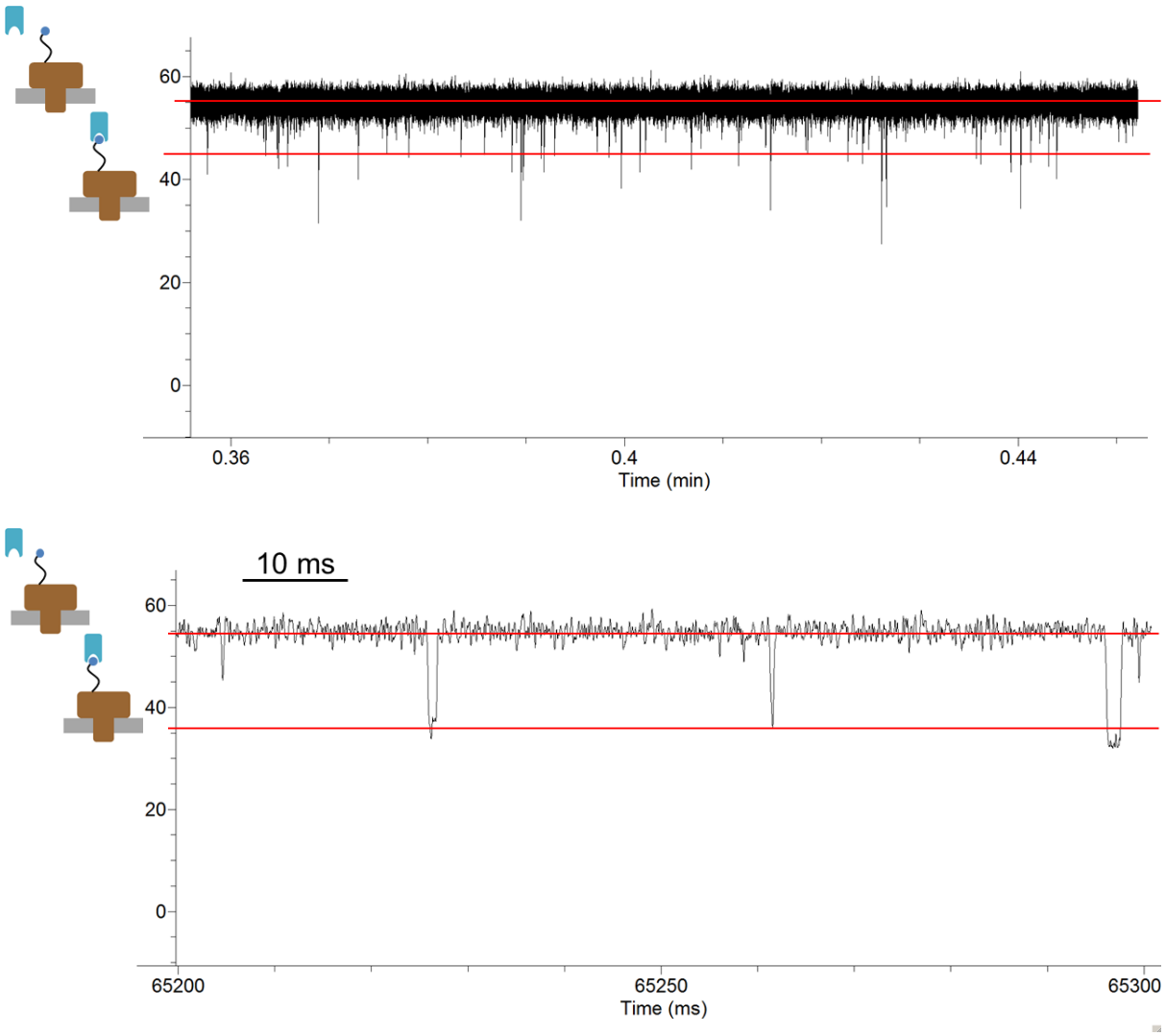



**Figure 3.16.** Detail of a trace recorded during interaction of FimH with a monovalent mannopore.

Typical for the mannopores and the buffer used, is the current around 57 pA when a voltage of 100 mV is applied. The thick black line around 57 pA is called the baseline. The baseline results from ions flowing through the channel as a consequence of a voltage applied across the artificial bilayer. Due to the detection of noise this line appears as a prominent black line but in fact it is a series of measuring points every 20 µs connected via a baseline. The black lines going down from the baseline are dips in the current caused by something blocking the ion flow through the pore. These

dips in the current are called the events.[44] If these events get more frequent by the addition of more protein, it can be concluded that it is the protein causing the events and that it must be the protein interacting with the mannopore. The trace in Figure 3.16 is a short detail from a much longer recording, and from this detail an enlarged view of some events is shown underneath. The appearance of events does not mean that the interactions between the protein and the mannopore are specific, meaning that the carbohydrate interacting domain is not necessarily interacting with the mannopore, and other parts of the protein can interact with the mannopore as well in a nonspecific way. By looking at the shape of the events, it can be seen that most events reach a specific level, or the amplitude is the same. The time spent at that level is called the dwell time of the event. This same level of blocking indicates that every interaction between FimH and mannopore is of the same nature, it blocks the ion flow in the same way. A statistical analysis of the amplitude of the events was carried out and a single distribution can be seen with an average value of 18 pA.

When FimH protein was added to the cis side of the monovalent α-HL pore (mannopore), events can be detected from a concentration of 10 μM FimH and upwards, whilst at lower concentrations, almost no events were detected. A concentration range from 10 μM to 25 μM FimH was used for single channel recording. No clear level can be identified and many of the events are triangular in shape. This can be the result of the duration of these events being the in the same order as the sampling rate. For these

events, not enough data points are recorded to fully see the real shape of these events. The baseline and events are subjected to noise, but the level of noise detected can be selected during the recording and can be filtered out afterwards during the analysis by applying a Gaussian-Bessel filter. Using a heavy filter will result in loss of detail or of the entire event, but the use of a too low filter results in noise being picked up as events and this will reflect the results of the analysis. The filter used will always be a compromise between these two consequences and some details will always be lost and some noise will be picked up as an event. To further minimise the amount of noise being picked up as events, in the analysis very short changes in level can be ignored. Filtering groups measuring point and average the difference between these measuring points out, this leads to less noise but also to fewer measuring points. For short events, the shape can change due to the decrease in points and it can look as if the event does not reach the same level of blockage of the other events. As a result of this the events have a triangular shape. For good analysis, these triangular events have to be included in the analysis, but they will give a slightly different dwell time as compared to the original event. The average event lasts around 0.27 ms. In Figure 3.16, a enlarged view of some events shows a level with an amplitude around -25 pA. Events that did not reach this level are in most cases triangular and are included in the analysis of the dwell time.

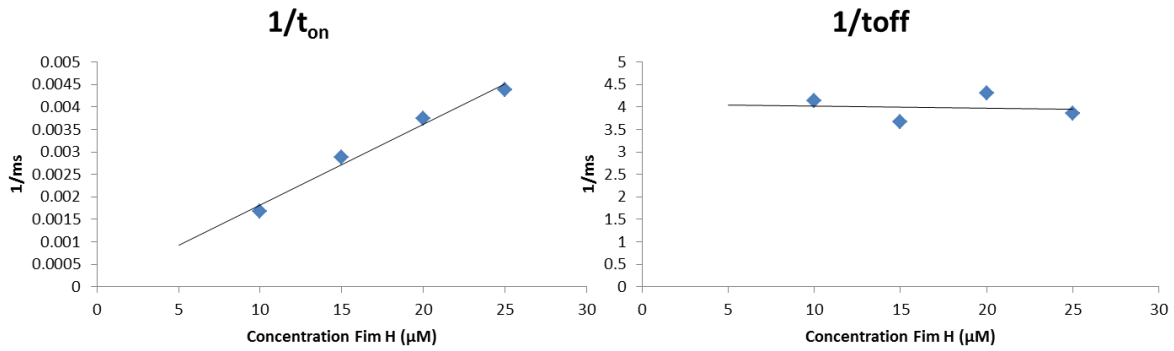



**Figure 3.17.** Concentration dependence of FimH in the interaction with a monovalent mannopore.

If FimH is responsible for the present events, the frequency of the events must be dependent on the concentration of FimH. Therefore the recordings were performed at four different concentrations. The dwell time spend at the base level, or the time between events is called $\tau_{on}$. The inverse of the average of these $\tau_{on}$ values should be linear to the concentration of FimH and should go through the origin. Slight deviations from this behaviour is due to the picking up of noise as events. In a similar way, the concentration of FimH should not affect the length of a single event. The dwell time spend at level one ($\tau_{off}$) should be independent of the concentration of FimH. This can be seen in Figure 3.17 where the reverse value of $\tau_{off}$ is plotted as a function of the concentration of FimH. All recordings have been repeated with at least three different pores to ensure the results are reproducible.

The number of events used for analysis is important, ideally at thousand events are needed for the conclusions to be significant. After the filter and minimum duration of events is tuned to detect all the events, software identifies all the events and the logarithm of the dwell time of each level

can be plotted in a histogram. If these histograms show a single distribution it indicates that the events in this level are the consequence of the same cause. To these histograms, exponential distributions can be fitted, and out of these curves the average of every level's dwell time can be calculated, in this case $\tau_{on}$ and $\tau_{off}$.[44] From these the forward rate constant $k_{on}$ = 1.10 ±0.01 $10^5$ $s^{-1}M^{-1}$ and reverse rate constants $k_{off}$ = 3703 ± 251 $s^{-1}$ for the binding of FimH to the mannopore can be calculated, and the ratio of the two give the dissociation constant $K_d$ = 34 ± 3 $10^{-7}$ mM for FimH and the mannopore, as shown in the following equations. Results are shown in Figure 3.18. The results are based on three independent experiments.

$$k_{\mathrm{Off}} = \frac{1}{\tau_{\mathrm{Off}}} \qquad \text{(Eq. 3.1)}$$

$$k_{\mathrm{On}} = \frac{1}{\tau_{\mathrm{On}} C} \qquad \text{(Eq. 3.2)}$$

$$K_{\mathrm{d}} = \frac{k_{\mathrm{Off}}}{k_{\mathrm{On}}} \qquad \text{(Eq. 3.3)}$$

A rough guidance for the quality of analysis is the quality of the histograms produced at the end and how the fitted curve corresponds to the histograms. With all this information and in this case, a simple model can be proposed for binding of FimH to the single mannopore, with two states, one where FimH is unbound and one where FimH is bound to the mannopore, see Figure 3.18.

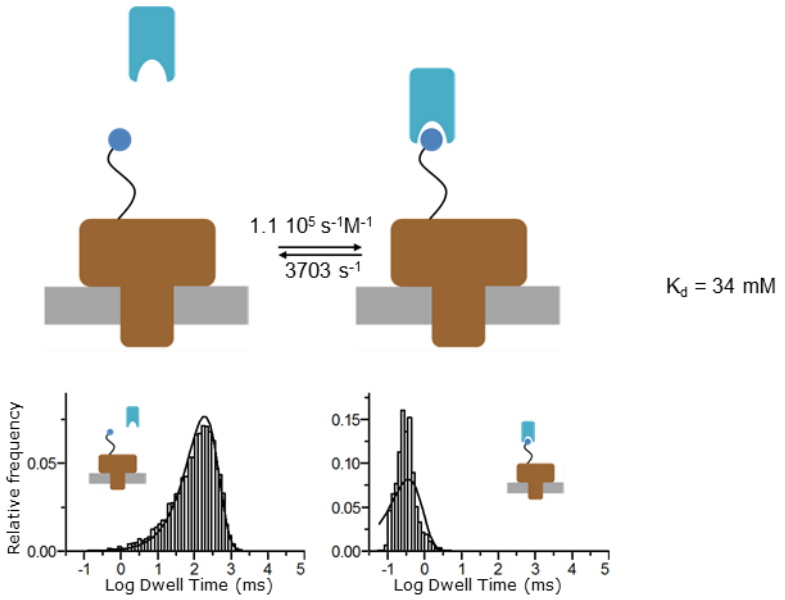



**Figure 3.18.** Model for the interaction between FimH and a monovalent mannopore.

### 3.7.4 FimH and bivalent mannopore

A set of new recordings was made, this time with a bivalent mannopore (containing two mannose molecules covalently attached to one single ion channel). The traces look almost identical to the traces of the monovalent case. The amplitude and dwell time of the events are similar and the time between events is almost the same.

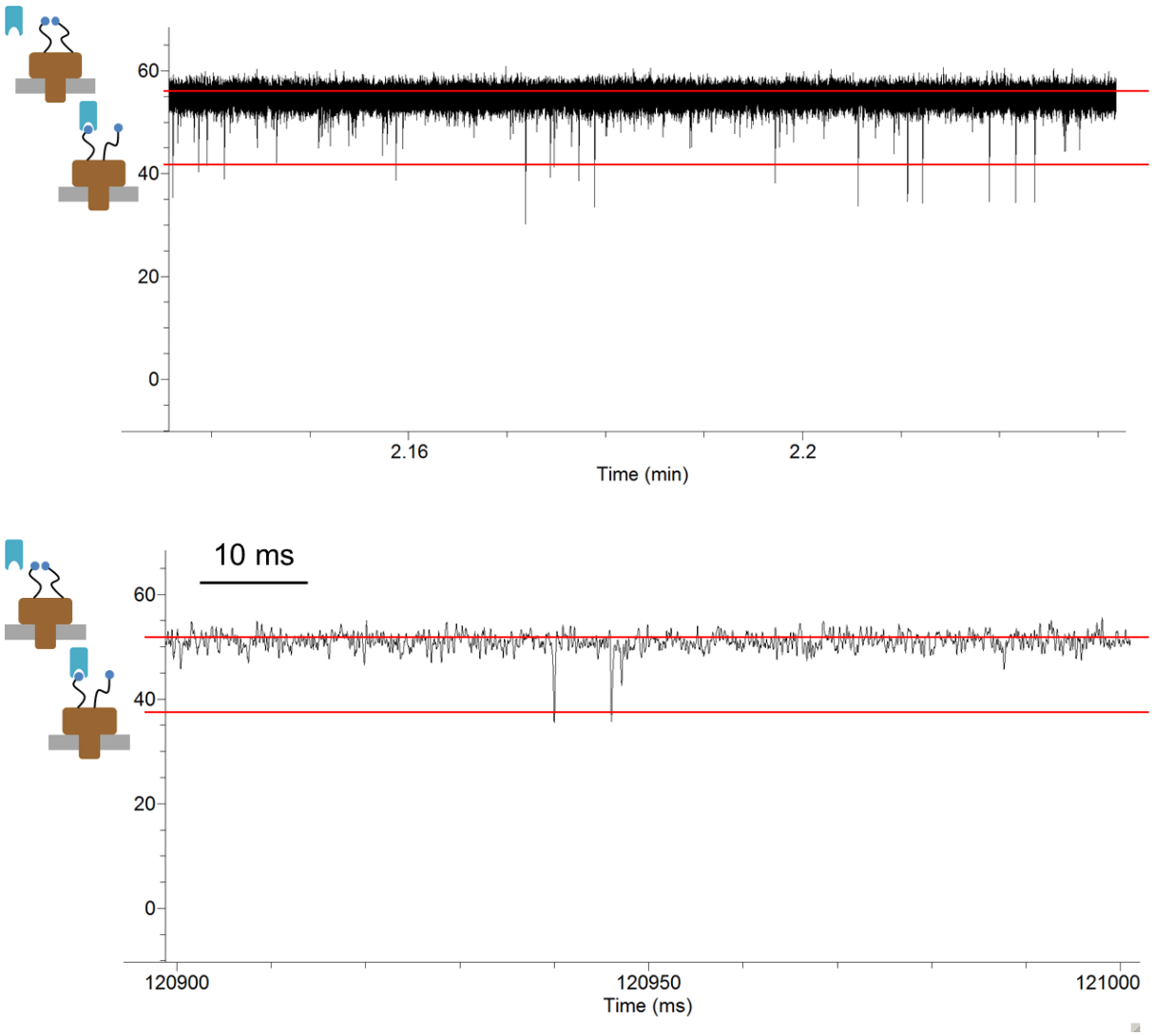



**Figure 3.19.** Detail of a trace recorded during interaction of Fim H with a bivalent mannopore.

Again the concentration dependence was looked at and similar concentration dependence curves were obtained (Figure 3.20), leading to a $k_{on}$ = 1.20 ± 0,02 $10^5$ $s^{-1}M^{-1}$, $k_{off}$ = 3722 ± 370 $s^{-1}$ and a $K_d$ = 31 ± 4 mM. These results are based on three independent experiments.

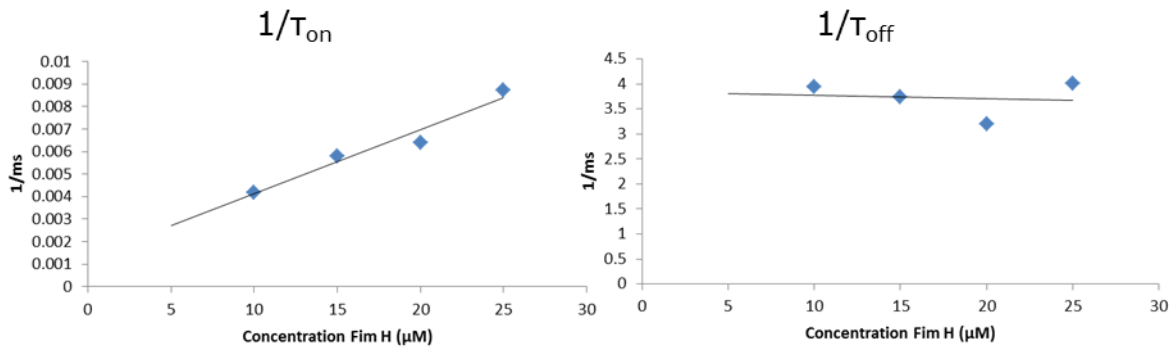



**Figure 3.20.** Concentration dependence of FimH in interaction with a bivalent mannopore.

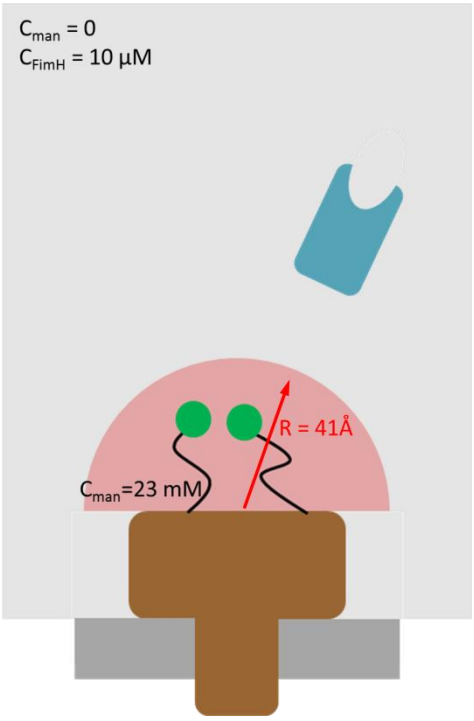



**Figure 3.21.** Effective concentration of FimH in binding to a mannopore.

An explanation for the similar results for the monovalent mannopore and the bivalent mannopore has been provided with the aid of the following simplified model as depicted in Figure 3.21. An estimate of the length of the decaethyleneglycol linker used to tether the mannose to the pores is 41 Å. If one assumes that the two mannose molecules can freely move in half a sphere with a radius of the linker-length, this results in a zone with a local concentration of mannose of 23 mM, outside this half sphere the concentration of mannose is equal to zero. The solution contains FimH at 10 μM or around 1000 times more dilute than the mannose in the half sphere. For binding to occur, FimH has to move into the sphere, and due to the low concentration this will be a slow step. Once a FimH molecule

has entered this zone, binding will occur relatively fast because of the high mannose concentration and the high $k_{On}$ value from FimH for mannose.

$$V_{1/2\text{Sphere}} = \frac{1}{2}\left(\frac{4}{3}\pi \cdot r^3\right) = \frac{4}{6}\pi\left(41\cdot 10^{-10}\,\text{m}\right)^3 = 1{,}44\cdot 10^{-25}\,\text{m}^3 = 1{,}44\cdot 10^{-22}\,\text{L} \quad \text{(Eq. 3.4)}$$

2 mannose molecules:

$$n = \frac{2}{N_A} = \frac{2}{6{,}02\cdot 10^{23}\,\text{mol}^{-1}} = 3{,}32\cdot 10^{-24}\,\text{mol} \quad \text{(Eq. 3.5)}$$

Concentration mannose in half sphere:

$$C_{\text{mannose}} = \frac{n}{V} = \frac{3{,}321\cdot 10^{-24}\,\text{mol}}{1{,}44\cdot 10^{-22}\,\text{L}} = 0.023\,\text{M} \quad \text{(Eq. 3.6)}$$

Volume for every FimH in solution:

$$V = \left(N_A \cdot C_{\text{FimH}}\right)^{-1} = \left(6{,}022\cdot 10^{23}\,\text{mol}^{-1} 10\cdot 10^{-6}\,\frac{\text{mol}}{\text{L}}\right)^{-1} = 1{,}66\cdot 10^{-19}\,\frac{\text{L}}{\text{Molecule}} \quad \text{(Eq. 3.7)}$$

The conclusions for binding of monovalent FimH to a bivalent mannopore are summarised in Figure 3.22; this figure also depicts the histograms for the base level and first level. Looking at the traces and at the histograms, no second level can be detected. This leads to the conclusion that no two molecules of FimH can bind to the same mannopore at the same time. As FimH is the smallest protein used in all of the following experiments, it can be expected that in all other cases no two protein molecules can bind to the bivalent mannopore at the same time. Also if this was the case the binding of the second protein would also be concentration dependent which can determined in the analysis of the binding events.

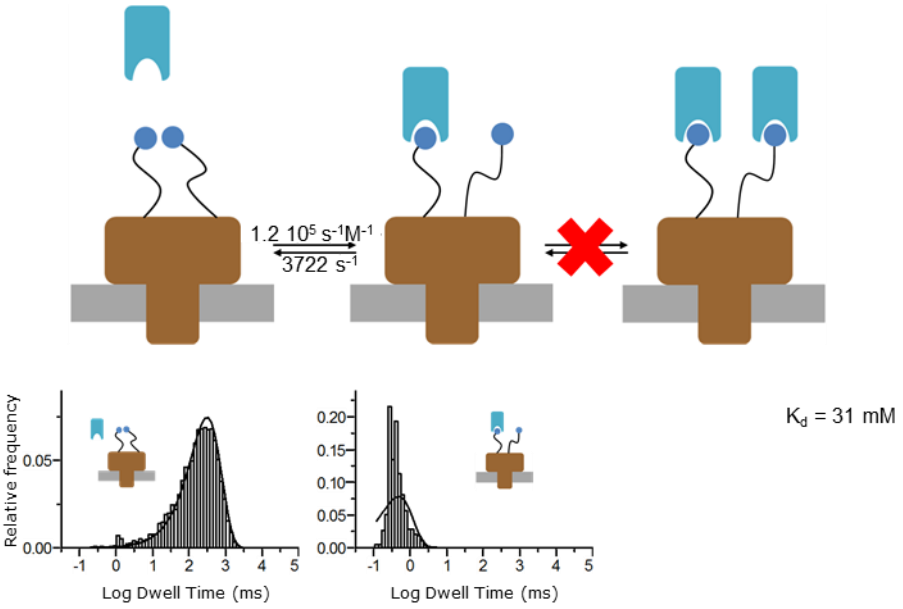



**Figure 3.22.** Model for the interaction between FimH and a bivalent mannopore.

### 3.7.5 Explanation of difference in $K_d$ as reported in literature

The determined affinity of FimH for the mannose connected to the pore is higher than the $K_d$ values reported in literature[11, 59] for mannose in solution. At the concentration the measurements are performed, the rate of the "on" reaction is determined by the rate of FimH reaching the zone where they can bind to the mannose (see fig 3.23). This is also demonstrated by the little difference in the $K_d$ value when one mannose molecule is present on the pore or two mannose molecules.

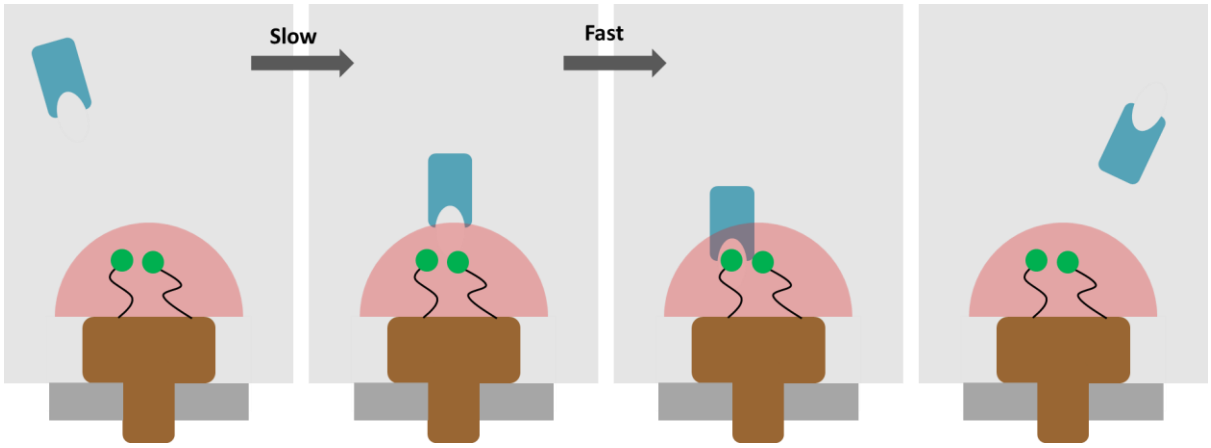



**Figure 3.23.** Kinetic model for binding of a FimH molecule to a bivalent mannopore.

Movement of molecules in a set up like this can be explained by three phenomena:

Convection:

Convection is the result of a force on a solution, this can be a pump or stirring, called forced convection. But also small thermal and density differences can cause convection, this is called natural convection. This cannot be controlled but due to the fast nature of the occurring binding events, this contribution will not affect the rate of binding of FimH to the mannopore. As there is no stirring during the measurements, there is also no forced convection to be accounted for.

Migration:

Migration is due to the electric potential over a liquid. Charged particles will move towards or be repelled from a charged surface. This effect is the strongest in close proximity to the electrodes and in close proximity to the ion channel. The high salt concentration in the measurement chambers acts

as a carrier of charges, which results in no effect of the potential on the bulky proteins with little charge.

Diffusion:

This only leaves diffusion as a contribution to the rate of the binding event. Diffusion is the consequence of local uneven concentrations of materials. A schematic representation is given in Figure 3.24. The main driving forces are entropic effects acting on molecules to even out those concentration differences.[86]

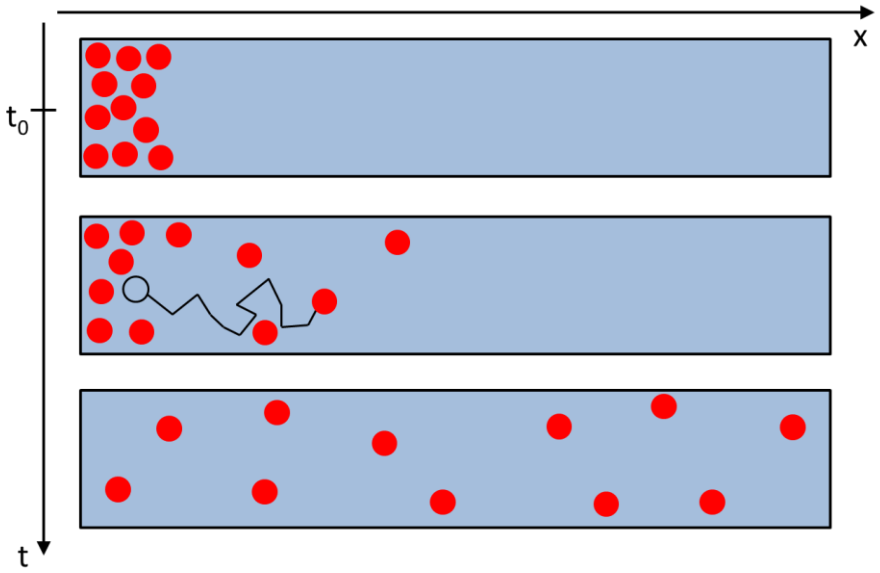



**Figure 3.24.** Schematic representation of diffusion.

Fick's law[85] provides a mathematical model to predict the movement of material due to diffusion.

$$J = -D\frac{\partial C_0}{\partial x} \qquad \text{(Eq. 3.8)}$$

Where J = diffusion flux (mol/m$^2$s), the amount of substance that flows through a surface area during a time

D = diffusion coefficient or diffusivity (m$^2$/s)

$C_0$ = concentration (mol/m$^3$)

X = position (m)

Fick's law in 2 or 3 dimensions becomes Equation 3.9:

$$J = -D\nabla C_0 \qquad \text{(Eq. 3.9)}$$

Stokes – Einstein equation, Equation 3.10

$$D = \frac{k_B T}{6\pi\eta r} \qquad \text{(Eq. 3.10)}$$

Where, D = diffusion coefficient or diffusivity (m$^2$/s)

η = viscosity

r = radius of spherical particle

$$(MW) = N_A \rho V = N_A \rho \left( \frac{4}{3} \pi r^3 \right) \qquad \text{(Eq. 3.11)}$$

$$\Rightarrow r = \left( \frac{3(MW)}{4\pi N_A \rho} \right)^{1/3} \qquad \text{(Eq. 3.12)}$$

Where $N_A$ = Avogadro's number

ρ = density

V = molar volume

Therefore, the expression for the diffusion coefficient becomes Equation 3.12:

$$D = \frac{k_B T}{6\pi\eta\left(\frac{3(MW)}{4\pi N\rho}\right)^{1/3}} \qquad \text{(Eq. 3.12)}$$

In the case of determining the binding affinity for FimH and the mannopore, the mannose molecule is bound covalently to the pore, resulting is a limited movement for the mannose moiety. The volume in which the mannose moiety can move is considerably smaller than the average volume in which a FimH protein molecule can move. For binding to happen, the FimH molecule must diffuse into the zone where it can bind to mannose. This is probably the opposite of what happens in solution, where the much smaller mannose molecule will diffuse towards the FimH protein. Although the diffusion coefficient is not a strong function of the molecular weight[86], the difference between the mass of a small molecule such as mannose and a high mass protein FimH will affect the diffusion drastically. And therefore, the determined dissociation constant will be much lower for the mannopore than the one reported in literature for solution.

# 3.8 Con A

## 3.8.1 Introduction

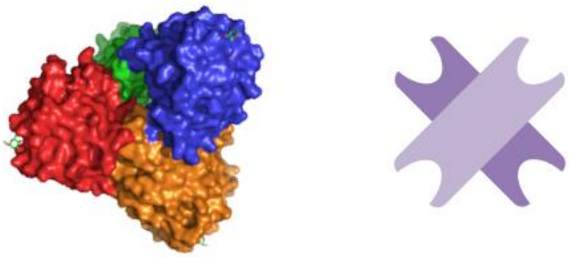


**Figure 3.25.** View of a crystal structure[61] of Con A and a cartoon representation of Con A as used herein.

The Con A[61] used in these experiments was purchased from Sigma Aldrich.

## 3.8.2 Con A and monovalent mannopore

A monovalent mannopore has been used in the following experiment. This pore was inserted into an artificial lipid bilayer and a voltage was applied over the bilayer. Due to the buffer used (see experimental section), the current through the mannopore was 57 pA when 100 mV was applied.

The concentration dependence was measured by the addition of Con A to the measuring chamber at the cis-side, in steps of 62.5 μM until a concentration of 312.5 μM was reached. Filtration and analysis of the data were performed in an identical way to the filtration and analysis of FimH. The measurement with the monovalent mannopore revealed next to the base level a single level with an amplitude of 5.5 pA down from the baseline as shown in Figure 3.26.

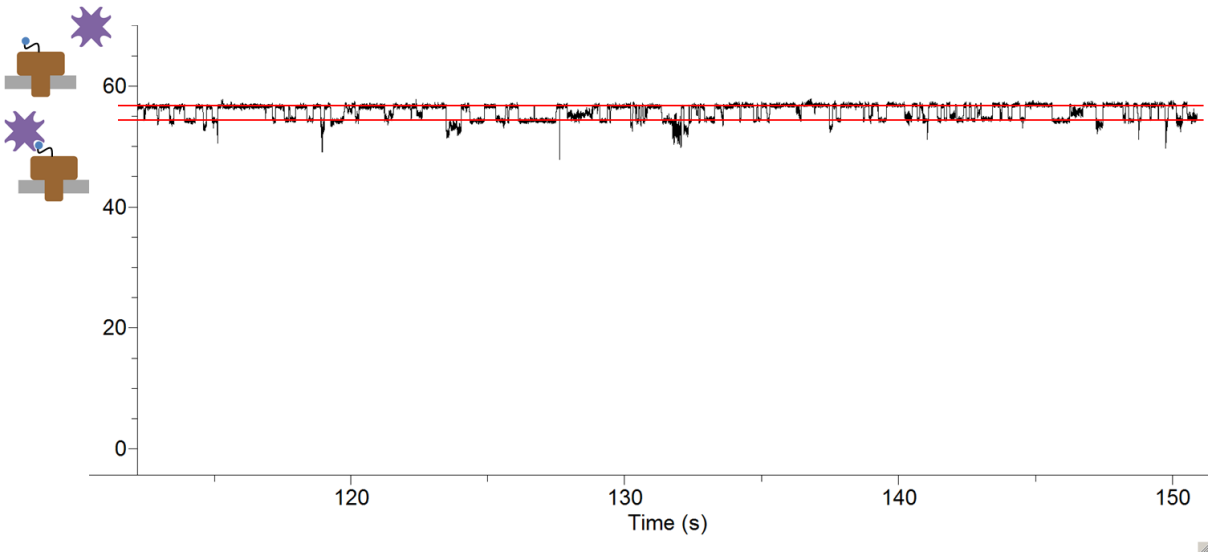



**Figure 3.26.** Detail of a trace recorded during interaction of Con A with a monovalent mannopore.

### 3.8.3 Con A controls

A number of controls were performed to show that the detected events are specific lectin – carbohydrate interactions. The first trace in Figure 3.27 is the baseline after insertion of a monovalent mannopore in the lipid bilayer. No events were observed. The next trace shows the same monovalent mannopore but after the addition of 125 μM Con A. Here, events were detected. The bottom trace in Figure 3.27 shows the current through the monovalent mannopore after addition of DTT. The DDT was added to reduce the disulphide bond between the α-HL and the PEG-linker. This way the mannose is detached from the α-HL and no binding of Con A to the α-HL should occur. Indeed, no events were detected.

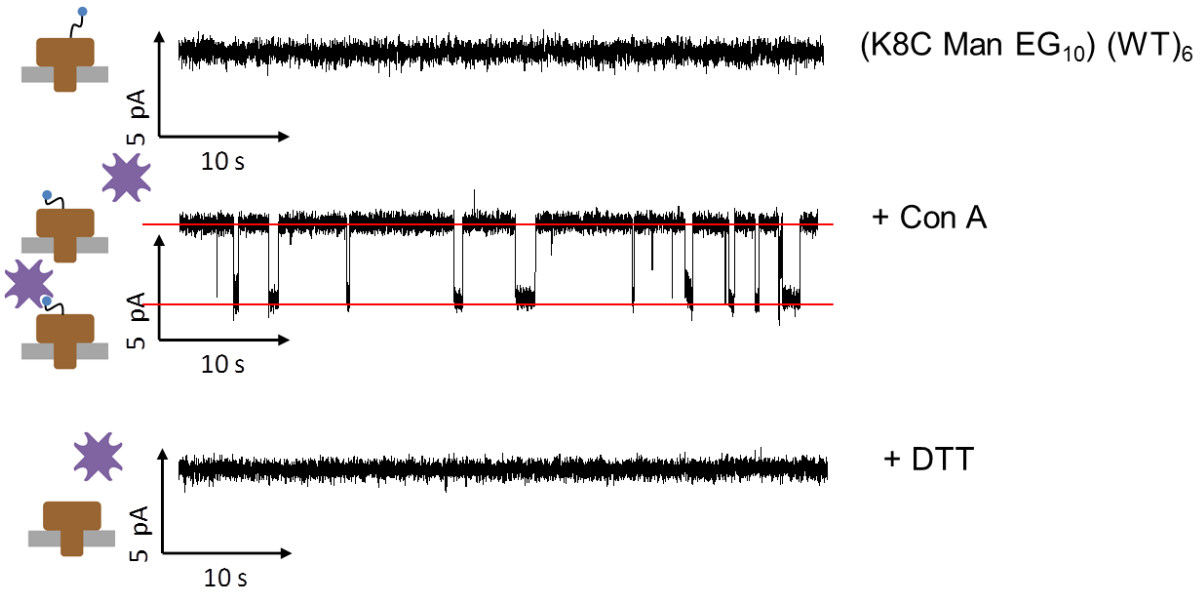



**Figure 3.27.** Control recordings, recordings with a monovalent mannopore and Con A show events, after the addition of DTT the mannose moiety is cleaved off and the events disappear.

Figure 3.28 summarises the binding kinetics, $k_{on}$ = 6500 ± 100 $s^{-1}M^{-1}$ and $k_{off}$ = 5.69 ± 0.26 $s^{-1}$, and also the $K_d$ = 0.875 ± 0.053 mM value of the single binding of Con A. These results are based on three independent experiments.

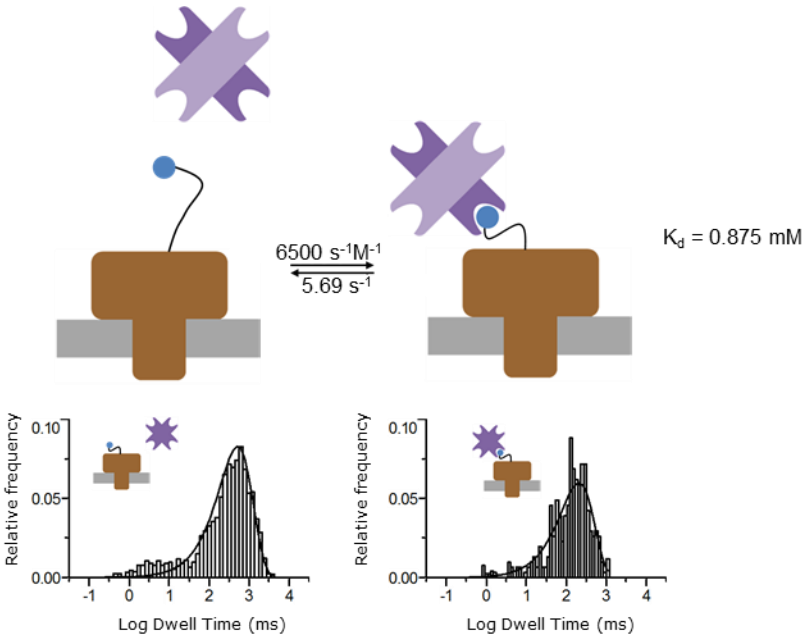



**Figure 3.28.** Model for the interaction between Con A and a monovalent mannopore.

### 3.8.4 Con A and bivalent mannopore

A bivalent mannopore was used in the following experiments described herein. The bivalent mannopore was inserted into an artificial lipid bilayer and a voltage was applied over said bilayer. Here again, due to the buffer used, the current through the pore was 57 pA when 100 mV was applied.[44]

Con A was added to the measuring chamber at the cis-side as before. The concentration dependence was measured by addition of Con A in steps of 62.5 μM until a concentration of 312.5 μM was reached. The same filters and analysis strategy was used here, as was used in the monovalent Con A measurements.

Addition of Con A to a bivalent mannopore resulted in complex events as shown in Figure 3.29. There was not just one level apart from the base level, but there were three levels that were apparently connected to each other. Looking at the details, the first step was a small blockage of the pore with the same amplitude as the monovalent binding of Con A. From this level there was another step down to a level around 10 pA down from the baseline, there were also plenty of switches between these two levels. Starting from this last level, some events occurred that almost completely block the ion flow. In most cases, the complex event will end by a sequence of switches going from the bottom level through all the intermediate levels and finally returning to the base level from the level providing the smallest blockage.

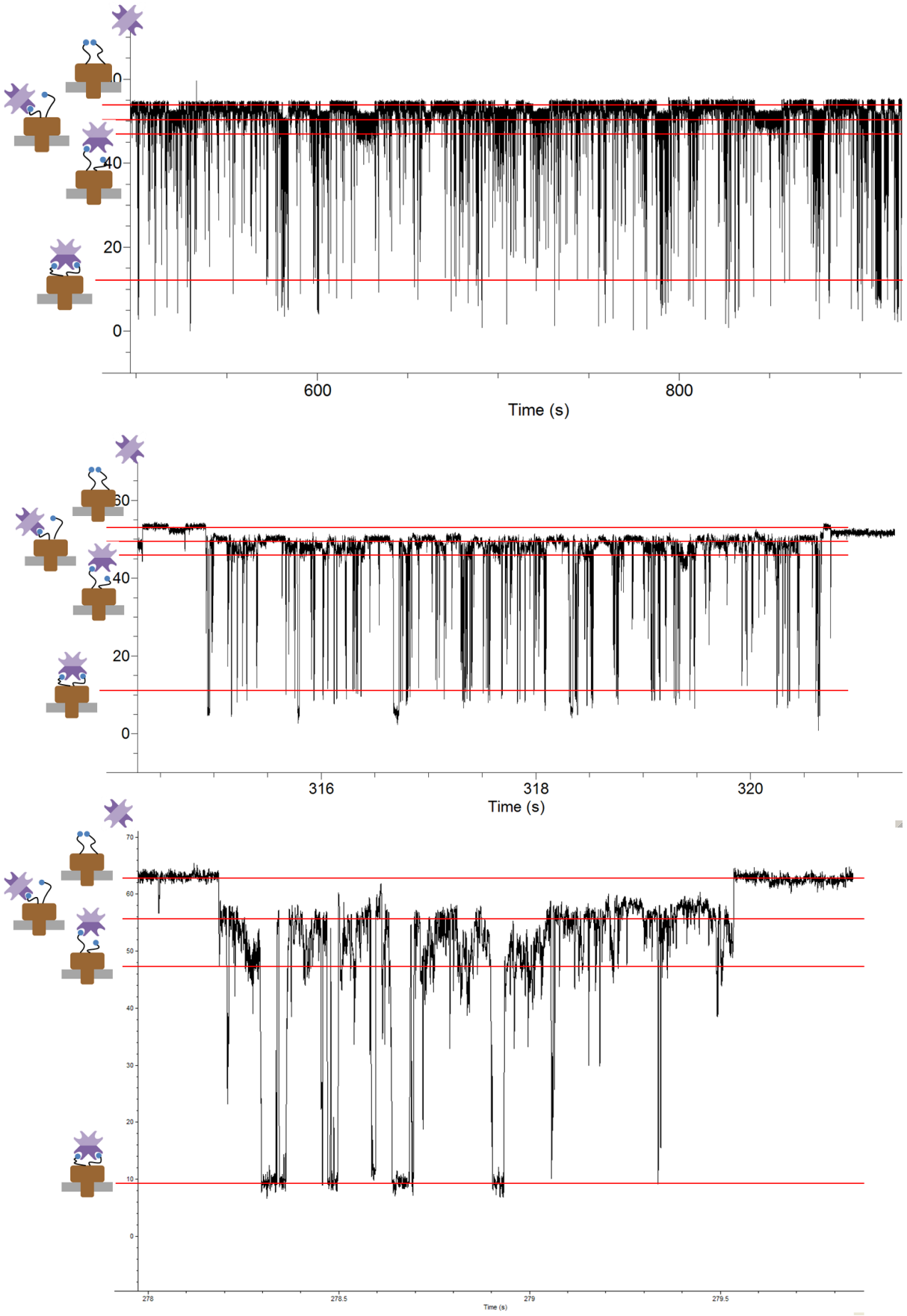



**Figure 3.29.** Detail of a trace recorded during interaction of Con A with a bivalent mannopore.

A model for the Con A binding to the bivalent mannopore is depicted in Figure 3.30 together with the kinetics between all the different levels, the histograms of the dwell times for each level and the fitted curves showing the validity of the model. The $k_{12}$ value between the base level (1) and the first level (2) is 6600 $s^{-1}M^{-1}$, the $k_{21}$ value for the switch from the first intermediate level (2) to the base level (1) is 14.89 $s^{-1}$. The transitions between the first intermediate level (2) and the second intermediate level (3) has a $k_{23}$ of 48.88 $s^{-1}$ and a $k_{32}$ of 238.10 $s^{-1}$, meaning that the forward transition is slower than the reverse reaction. The last transition between the second intermediate level (3) and the bottom level (4) is characterised by a $k_{34}$ of 54.61 $s^{-1}$ and a $k_{43}$ of 442.26 $s^{-1}$, meaning the reverse transition is much faster than the forward transition. These values suggest the following, once one binding site of Con A binds to one mannose moiety of the bivalent mannopore, the transition towards the second intermediate level (3) occurs faster than the transition back towards the base level (1) or the unbound state. Once the system is in the second intermediate level (3), the transition back towards the first intermediate level is faster than the transition towards the bottom level (4).

Transitions "skipping" a level were not observed in the analysis, transitions between the bottom level and the base level or first intermediate level were only observed in amounts that can be regarded as experimental error. A binding model suggesting these transitions was fitted to the experimental data, but no convolution towards a good fit could be obtained, meaning the data do not support such a binding model.

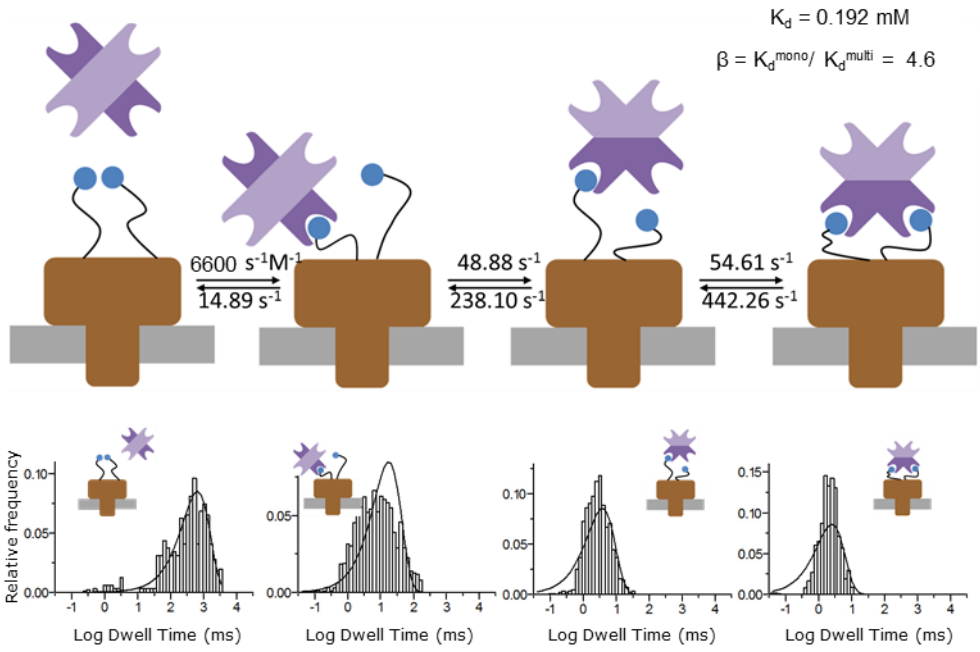



**Figure 3.30.** Model for the interaction between Con A and a bivalent mannopore.

Experiments with different concentrations of Con A added to the measuring chamber were conducted to study the concentration dependency of each transition. The results of these experiments are shown in Figure 3.31. The results reveal that only the transition between the base level (1) and the first intermediate level (2) are concentration dependent. The concentration of Con A does not affect any of the other rate constants associated with transitions between intermediate or bottom levels. These results also support the binding model suggested in Figure 3.30, where only the first step is intermolecular and as a consequence concentration dependent. The other steps in binding events are intramolecular, and therefore concentration independent.

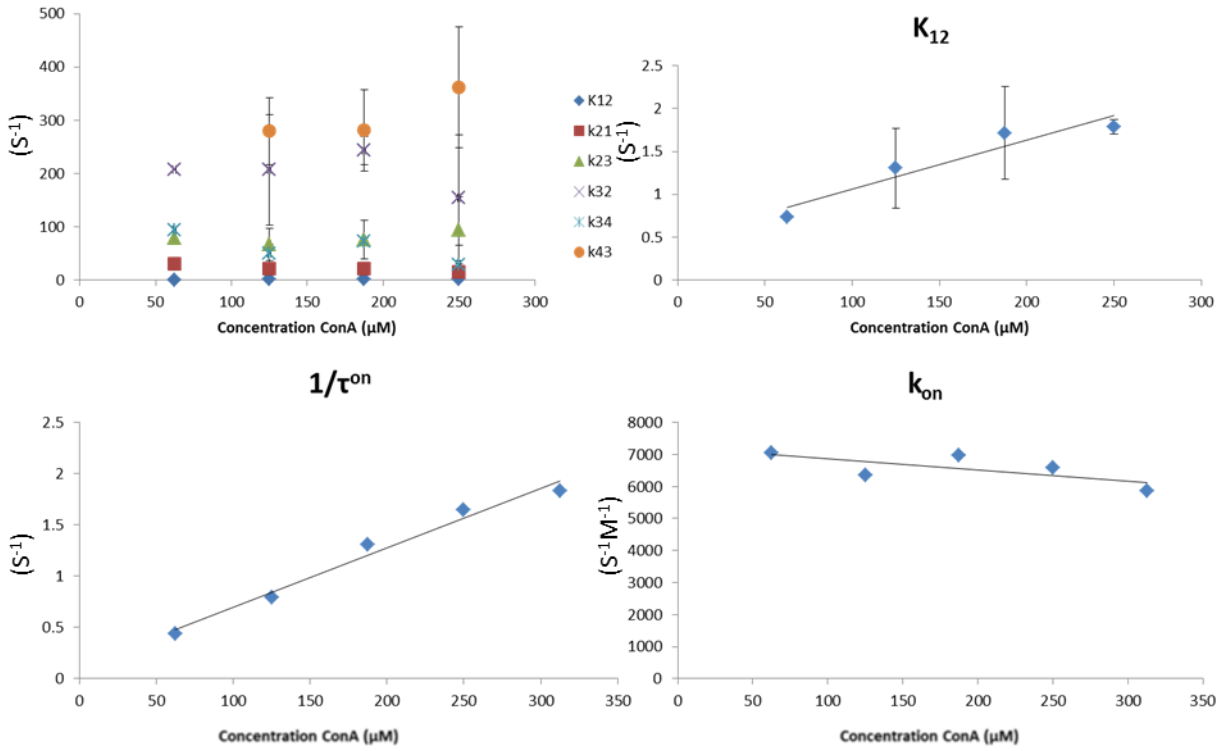



**Figure 3.31.** Concentration dependency of the different transitions in multivalent interaction of ConA with a bivalent mannopore.

In an effort to calculate a β-value to quantify the multivalent effect occurring during Con A binding to the bivalent mannopore the overall $k_{on}$ and $k_{off}$ were determined, thereby regarding only the base level and one binding level, which combines all three levels identified above. The ratio of $k_{off}$ value over the $k_{on}$ value results in a $K_d$ value for the multivalent binding event of 0.192 ± 0.033 µM, which is 4.6 times lower than the $K_d$ value of a monovalent binding event. Therefore the β-value for Con A is 4.6. Looking at the length of the events detected it can be seen that the binding events of Con A to a bivalent mannopore are much longer in length than those binding events to a monovalent mannopore.

$k_{on,\ multivalent} = 6600 \pm 500\ s^{-1}\mu M^{-1}$ (Eq. 3.13)

$k_{off,\ multivalent} = 1.26 \pm 0.12\ s^{-1}$ (Eq. 3.14)

$K_d^{multivalent} (= k_{off}/k_{on}) = 0.192 \pm 0.033\ mM$ (Eq. 3.15)

$K_d^{\ monovalent} = 0.875 \pm 0.053\ mM$ (Eq. 3.16)

$\beta = K_d^{mono}/K_d^{multi} = 4.6$ (Eq. 3.17)

The results for the multivalent experiments are based on three independent experiments.

### 3.8.5 Docking of Con A and a bivalent mannopore

To explain the observed results for the binding of Con A and the bivalent pore, docking studies were performed by Dr. Mark Wormald. The Con A tetramer was docked onto the bivalent α-HL pore with the two $EG_{10}$ linkers ending in a mannose moiety.

In Figure 3.32 Con A is depicted in 3 of the 6 orientations possible (purple). Con A is a dimer of dimers and these dimers are organised so that the 4 different binding sites are orientated in a pseudo tetrahedron (where binding sites A and B, and binding sites C and D are part of the same dimer). This leads to a symmetry which simplifies the system to 3 different orientations where A and B can be interchanged with C and D (values in brackets in Figure 3.32). Although distances between the two anomeric positions of the two mannose molecules (yellow) docked in the binding sites are similar in the three different orientations, the binding capabilities are notably different. In the case of A and B, the two mannose molecules are

directed towards the corners of the tetrahedron. Binding sites A and C, and binding sites A and D are positioned higher up in the protein and are not accessible for binding. The linkers need to bend around the protein more to reach the binding site upon bivalent binding.

The difference in these orientations results in Con A only being able to be docked in a bivalent way when binding sites A and B are used. The $EG_{10}$ linkers are too short to bind to another combination of binding sites simultaneously. This is mainly because Con A cannot come close enough to the α-HL without causing steric clashes.

The bottom half of Figure 3.32 shows a map of the charges on the surface of the protein, where blue are residues with a positive charge, red residues have a negative charge and the green residues are hydrophobic. From these images, it can be concluded that the overall charge over each surface is more or less neutral and that there are no hydrophobic patches. This is an indication that there is no significant protein - protein interaction involved between Con A and α-HL, and that only the interaction between the mannose moieties and Con A is responsible for the bivalent binding.

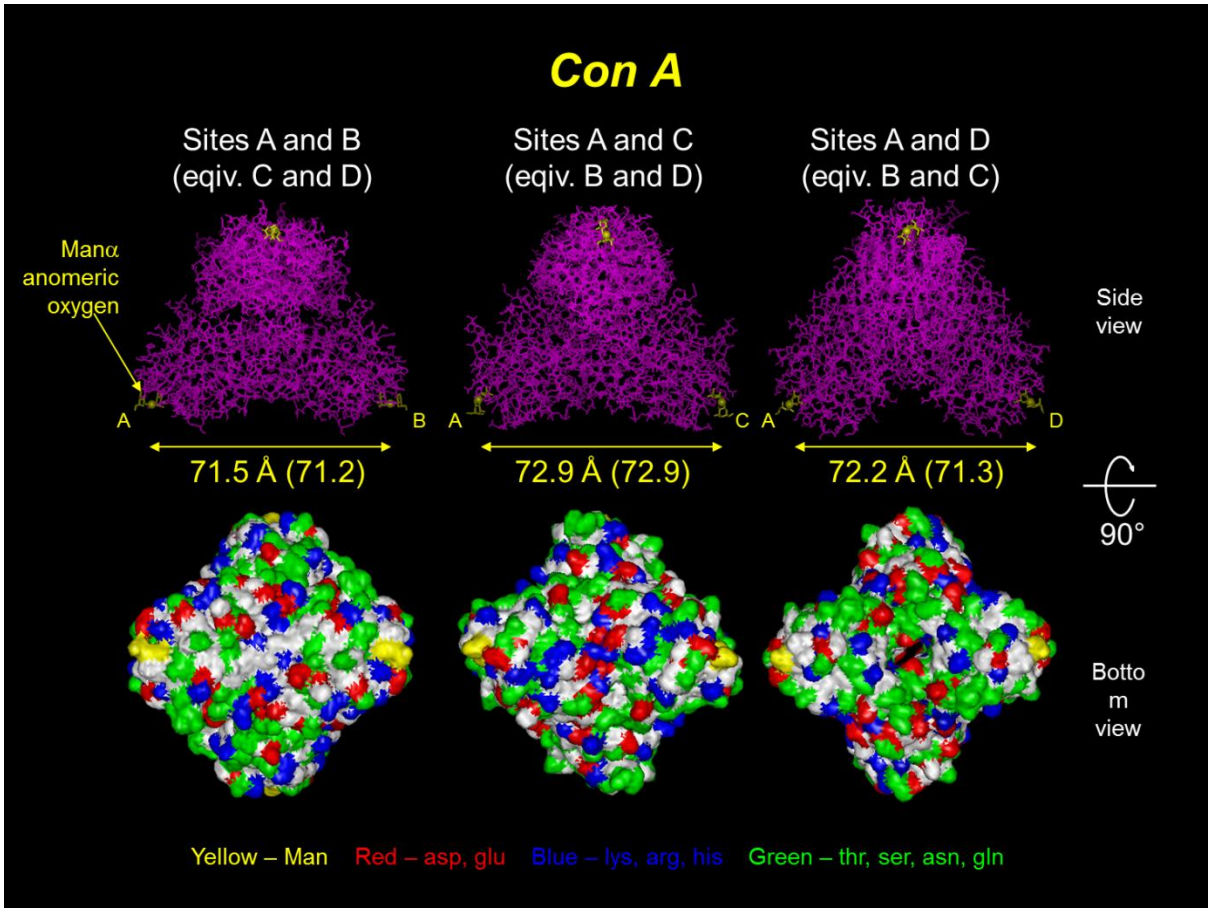



**Figure 3.32.** Different orientations of Con A crystal structure comprising mannose in the binding site and surface mapping of the different orientations of said crystal structures.

Figure 3.33 shows the maximum distance (79 Å) between the last oxygen atoms of the $EG_{10}$ linkers connected to the cysteine at position 8 of two α-HL monomers adjacent to each another in the heptamer. Because these modifications are not in the centre of the protein, binding of A-B and B-A are different, and will block the ion flow through the pore to a different extent. But in the case of Con A, the pseudo tetrahedral organisation the difference is so small that no difference in SCR signal is expected.

When Con A is docked onto the α-HL pore so that binding site A and B can be occupied (Figure 3.35), Con A almost completely blocks the entrance to the ion channel from three sides, leaving only a small gap where ions can flow through at one side of the channel. This might explain the observation of a blockage of more than 80% when Con A binds bivalent to the pore.

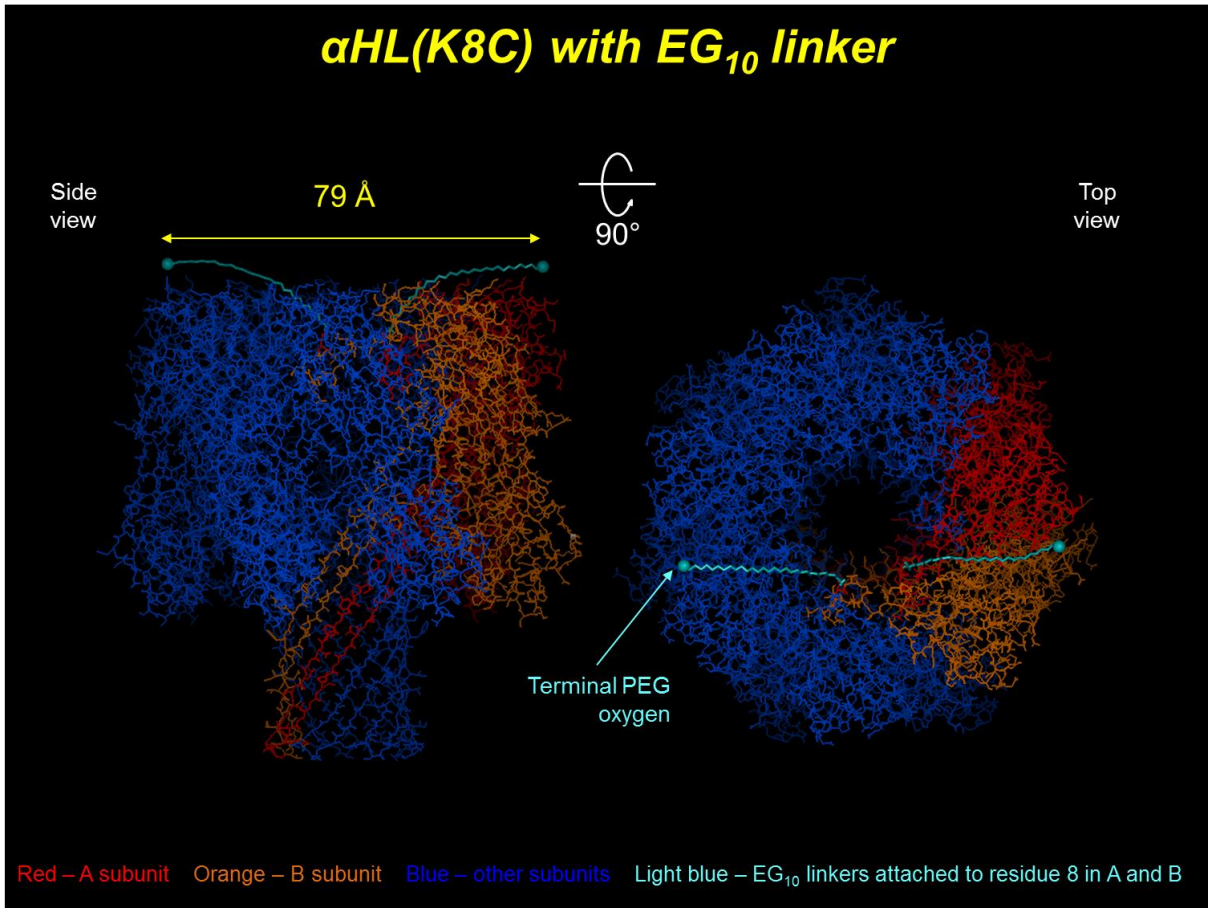



**Figure 3.33.** Side and top view of a bivalent mannopore with the 2 mannose moieties placed as far apart from each other as physically possible.

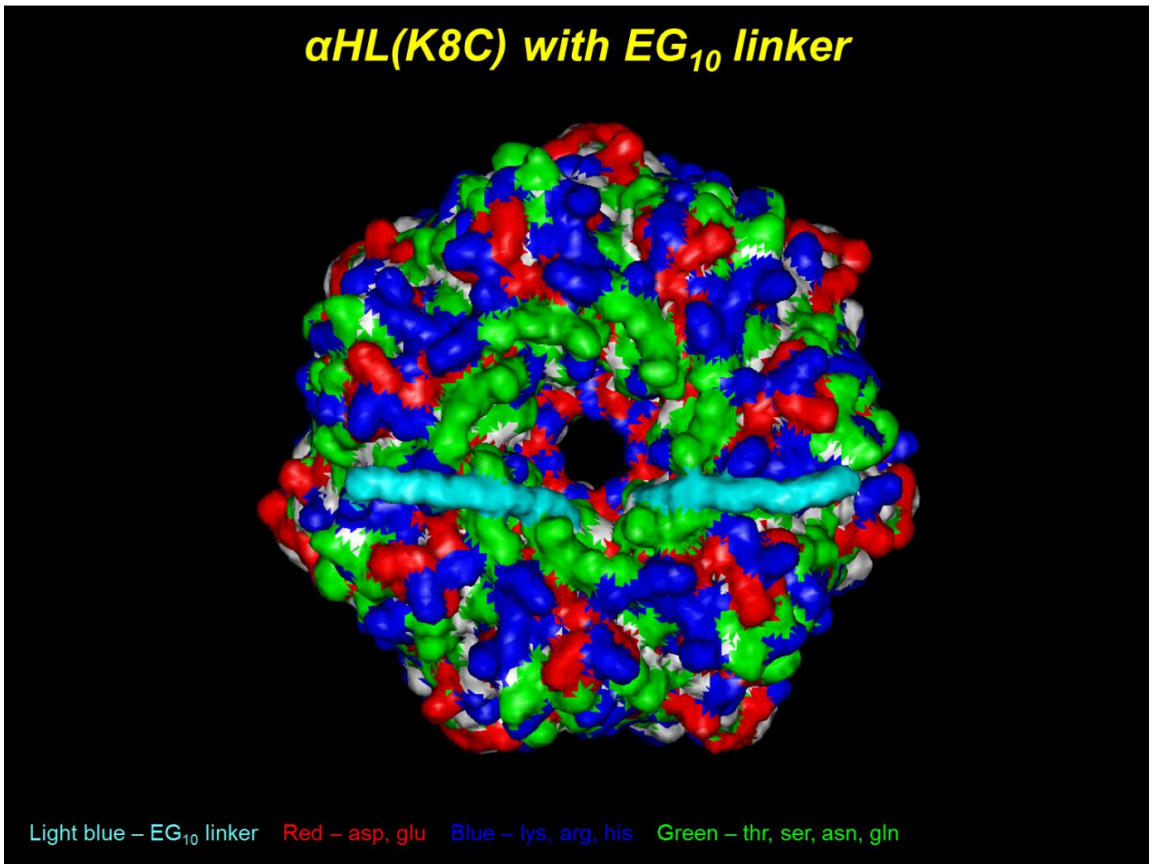



**Figure 3.34.** Surface map of the binding interface of the bivalent mannopore.

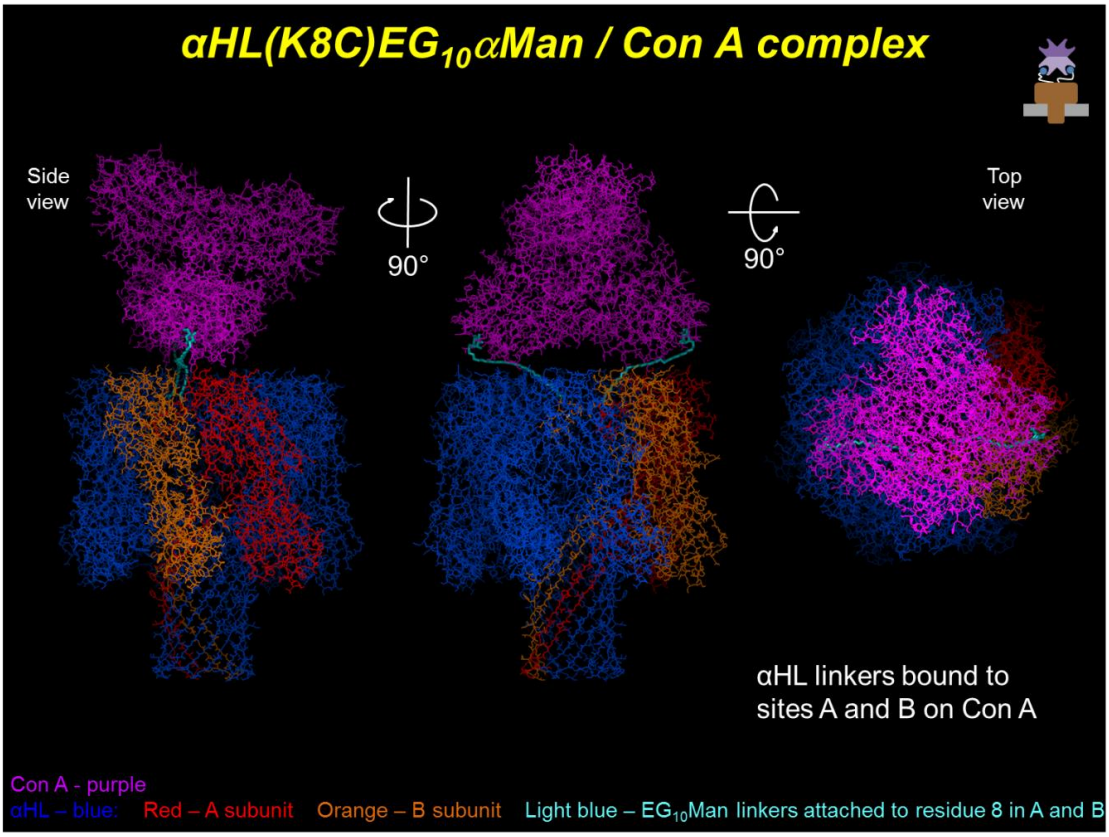



**Figure 3.35.** Side view and top view of Con A docked on a bivalent mannopore in the A and B orientation.

In the enlarged view in Figure 3.36, of the docking of Con A onto α-HL it can be seen that the linkers need to bend around a loop at the bottom of the Con A molecule, for the mannose moiety to occupy the same space as the mannose that was present in the crystal structure. If this loop would be considered flexible, more flexible binding would be allowed. The docking has been performed such that the mannose has the same orientation in the binding site as it has in the crystal structure. If the mannose would have had a beta-linkage to the $EG_{10}$ linker, reaching the correct orientation would have been easier resulting in more flexibility in the linkers.

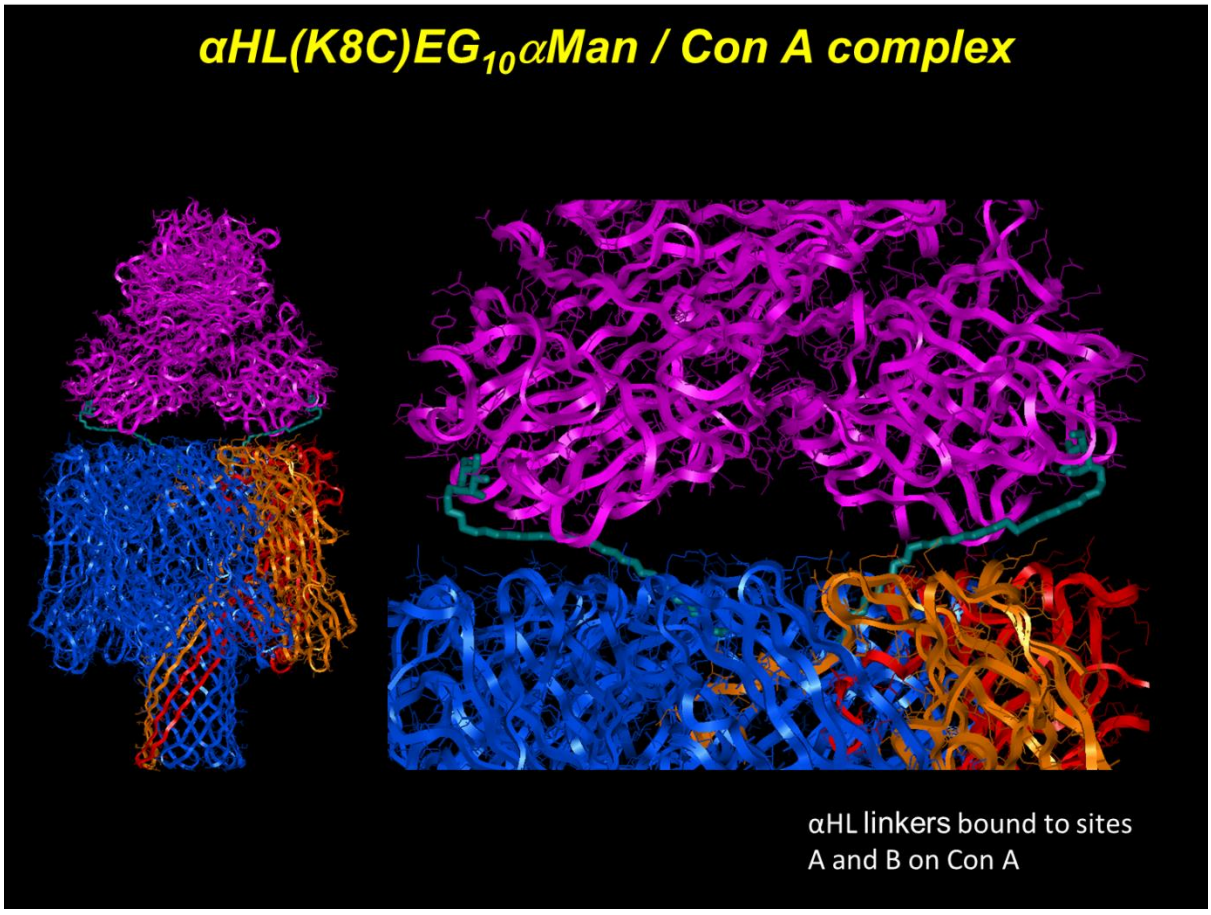



**Figure 3.36.** Detail of the bivalent binding interaction between Con A and a bivalent mannopore.

Figure 3.37 shows the different orientations docked on the bivalent mannopore. Although Con A is pseudo tetrahedral, only binding site A and B can be occupied by a mannose moiety of the bivalent mannopore at the same time. Other orientations, (A and C) and (A and D), require longer linker lengths to be occupied at the same time by the ligands of the bivalent mannopore.

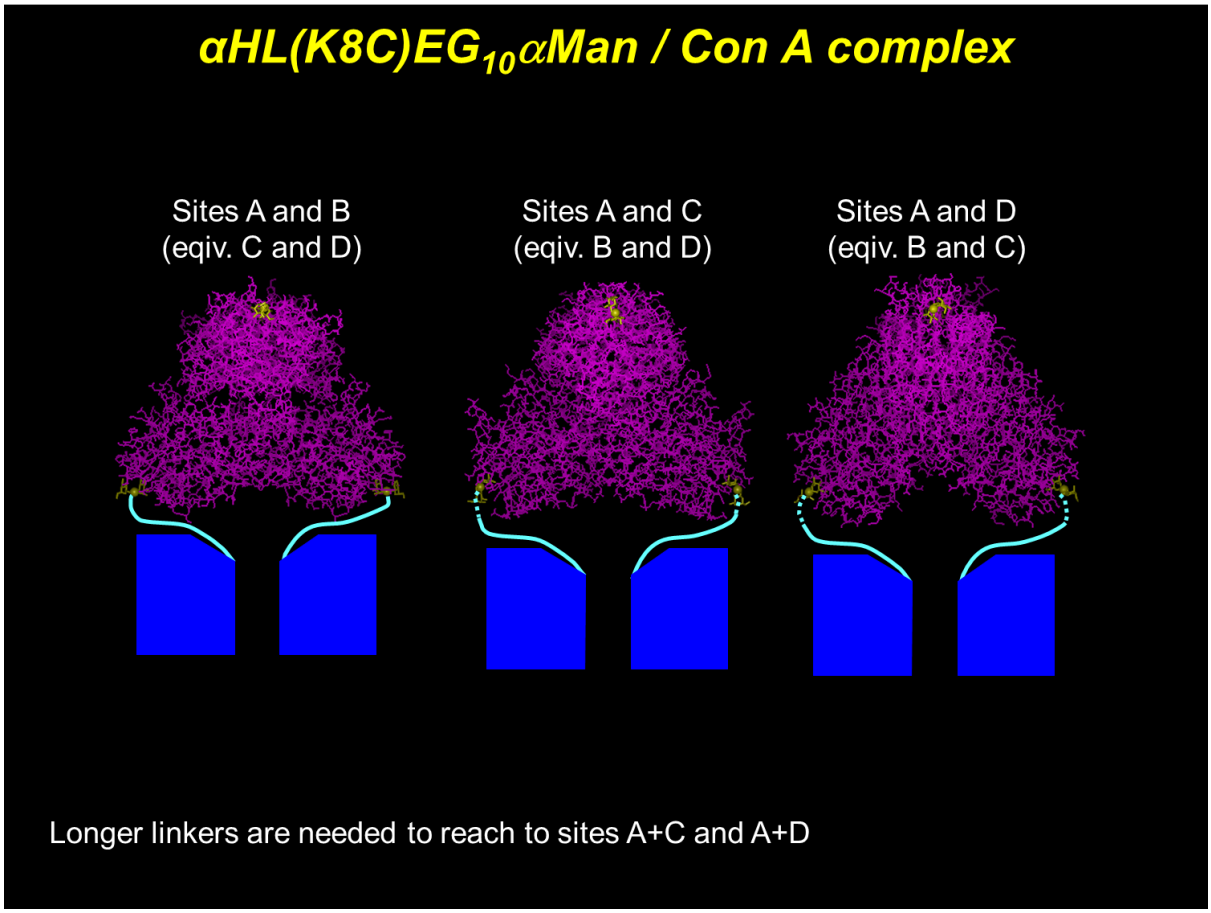



**Figure 3.37.** Con A binding modes, only the AB orientation can actually bind to the bivalent mannopore, the linkers are to short to bind, A-C or A-D at the same time.

### 3.8.6 Conclusion

Binding experiments of Con A to a monovalent and bivalent mannopore revealed the binding interaction of a bivalent binding event. From the results a model was suggested for bivalent binding. The model comprises

four states, one state with no binding, a state where only one binding domain is bound to the bivalent mannopore, an intermediate binding state and one state where Con A is bivalently bound to the mannopore. The experimental data support this model and molecular modelling suggests that the model is possible. For the intermediate binding state we suggest that a conformational change occurs in the Con A tetramer, and that from this conformation bivalent binding is possible to the mannopore.

## 3.9 PS lectin

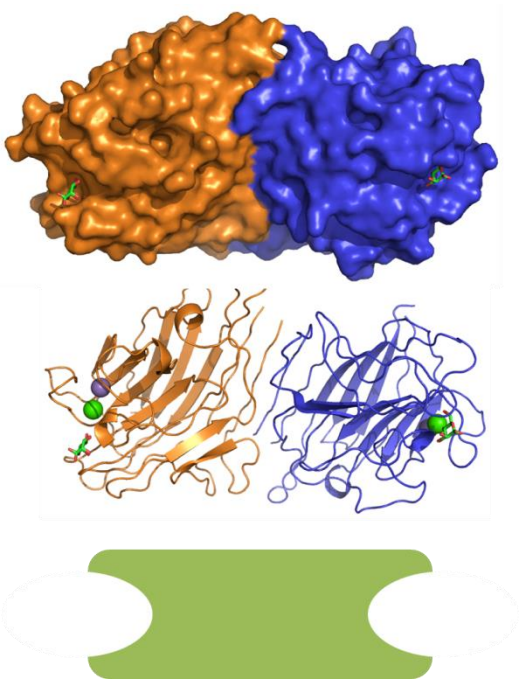


**Figure 3.38.** View of a crystal structure of PS lectin and a cartoon representation of PS lectin as used herein.

The PS lectin[62,63] in these experiments was purchased from Sigma Aldrich.

### 3.9.1 PS monovalent

For this set of experiments a monovalent mannopore was inserted in an artificial lipid bilayer. The measuring chambers were filled with recording buffer and a voltage of 100 mV was applied over the bilayer. Due to this voltage a current from around 57 pA was measured through the mannopore. To the cis side of the mannopore PS lectin was added starting at a concentration of 1 µM. The concentration dependence of binding events was measured and therefore more experiments were conducted at 2 µM, 3 µM and 4 µM PS lectin.

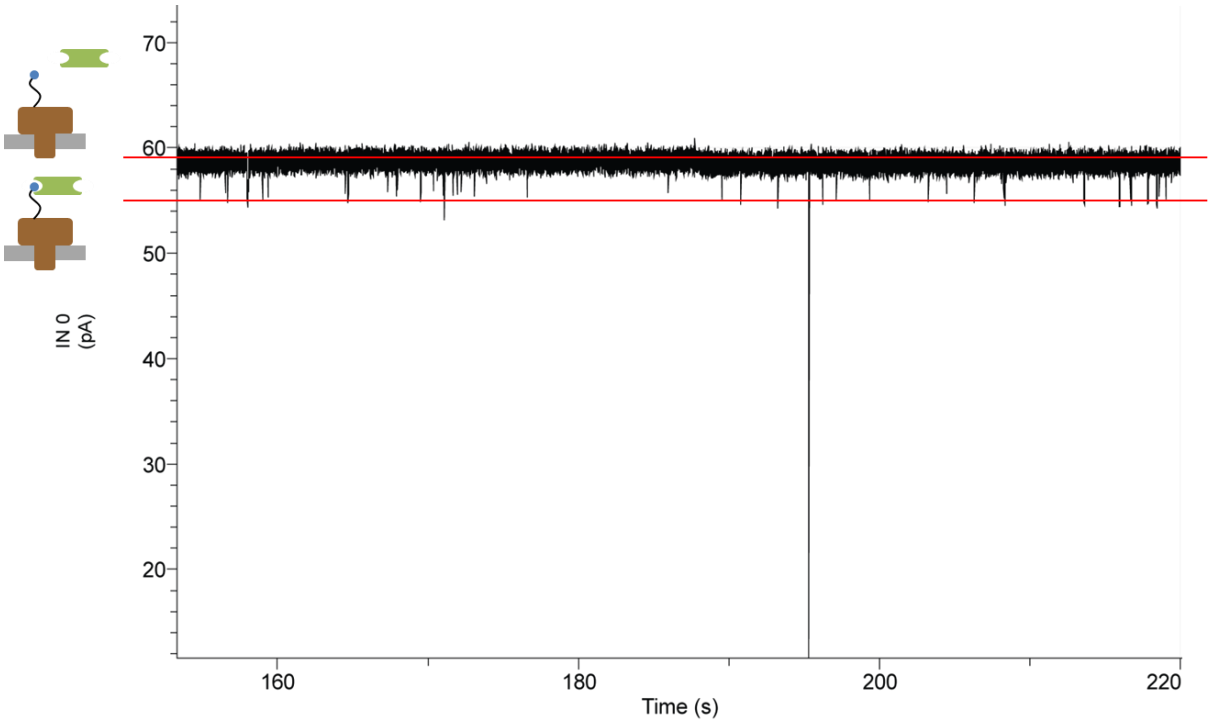



**Figure 3.39.** Detail of a trace recorded during interaction of PS lectin with a monovalent mannopore.

The traces recorded in these experiments show a single level in Figure 3.39 with an amplitude of 4.5 pA. The dwell time of these events is 14 ms. A model with two states is suggested in Figure 3.40, one state with no

binding, and a state where one binding domain of PS is bound to the monovalent mannopore. The fitted curves match the histograms suggesting a good model. The $k_{on}$ of this binding event is 32362 ± 647 $s^{-1}M^{-1}$ and the $k_{off}$ is 71.4 ± 4.9 $s^{-1}$. This results in a $K_d$ for the monovalent binding of 2.2 ± 0.2 mM. These results are based on three independent experiments. As shown in the trace in Figure 3.39, some events occurred that completely blocked the ion flow through the pore. These events are very short and triangular in shape. These events are too rare to include in the analysis and were ignored as they are probably the effect of some impurity or non-specific interaction between protein and mannopore.

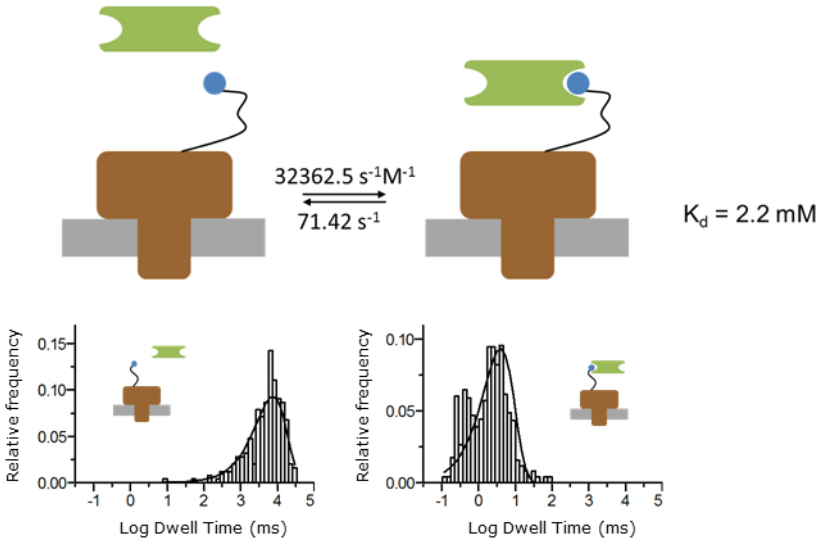



**Figure 3.40.** Model for the interaction between PS lectin and a monovalent mannopore.

### 3.9.2 PS bivalent

To study the bivalent binding behaviour of PS lectin, a bivalent mannopore was inserted into an artificial lipid bilayer as described herein before. The same buffer and recording conditions were used as in the monovalent experiments.

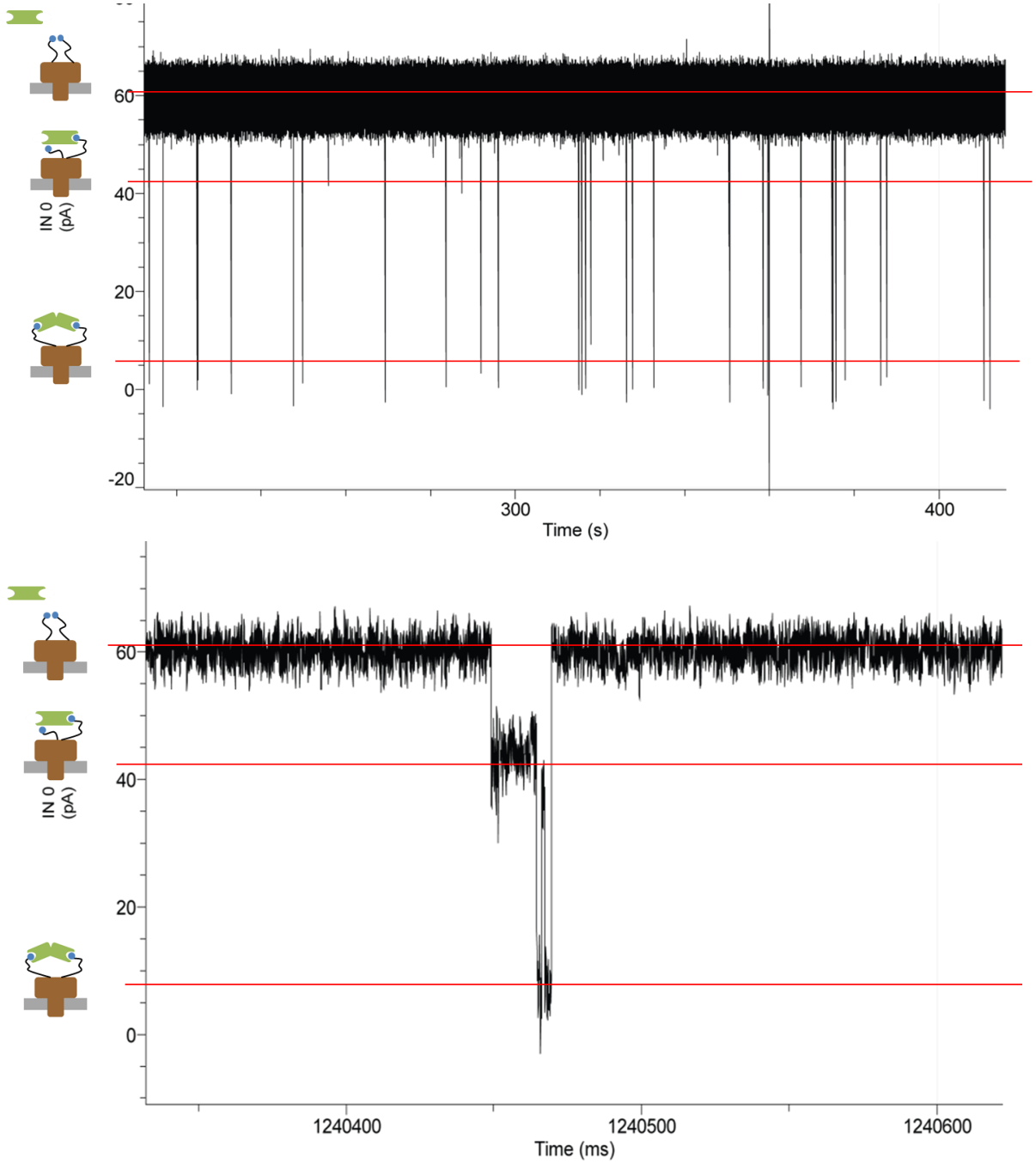



**Figure 3.41.** Detail of a trace recorded during interaction of PS lectin with a bivalent mannopore.

Figure 3.41 shows a part of a recording of PS lectin and a bivalent mannopore and an enlarged view of one of the events is shown. In this Figure two levels can be seen. The first level has a similar amplitude than the monovalent binding and the second level has a much larger amplitude of 17 pA. Few transitions between these levels have been observed during one binding event. The most remarkable observation is that when the system returns to the base level it always starts at the lower level without going through the first level. The $K_d$ value for the multivalent binding event was determined to be 1.03 ± 0.06 mM, as a $k_{on}$ was observed of 1.17 $10^4$ $s^{-1}M^{-1}$ and a $k_{off}$ of 12.1 $s^{-1}$. The observed β-value for the PS lectin was 2.13 ± 0.32. These results are based on three independent experiments.

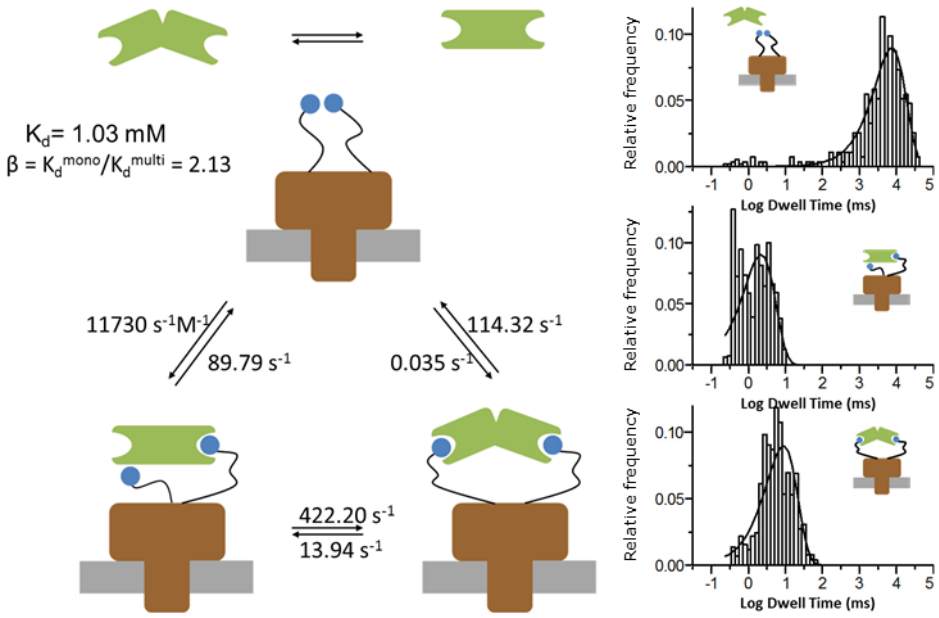



**Figure 3.42.** Model for the interaction between PS lectin and a bivalent mannopore.

A model as shown in Figure 3.42 has been fitted to these results comprising three binding states, a first state where no PS lectin is bound to the mannopore (1), a second binding state (2) where the PS lectin only binds

via one binding site to the mannopore and a third binding site where PS lectin binds bivalently to the mannopore. The rate constant $k_{12}$ of the transition between the first and the second binding state is 11730 $s^{-1}M^{-1}$ the reverse rate $k_{21}$ is 89.79 $s^{-1}$. The rate constant $k_{23}$ is 422.20 $s^{-1}$ and the reverse $k_{32}$ is 13.94 $s^{-1}$ between the second and third binding states. The transition between the first and the third binding states are characterised by a $k_{13}$ of 0.035 $s^{-1}$ and a $k_{31}$ of 114.32 $s^{-1}$. This means that once a PS lectin molecule is monovalently bound to the mannopore, the binding with second mannose moiety is much faster than returning to the non-bound state. Once the system is in the second binding state the transition towards monovalent binding is slower than the transition to the state where PS lectin is not bound to the mannopore. This can be observed by looking at the details of the traces as well. Almost all events stop with a direct transition from the bivalently bound state towards the non-bound state. No extra step can be seen at the second level when an event stops. Analysis of the crystal structure[63] of the PS lectin itself can explain this observation. Here, docking experiments were carried out by Dr Mark Wormald.

The PS lectin exists as two globular parts with a flexible region in the middle linking these globular parts. In the crystal structure[63], the two binding states are 69 Å apart as shown in Figure 3.43. The binding sites are indicated by the presence of the mannose in the crystal structure in yellow. The lower face will be oriented towards the ion channel upon binding shows no areas with an overall cationic (blue), anionic (red) or hydrophobic (white) nature. This indicates that the chance of non-specific interaction

between PS lectin and the ion channel is low. The overall appearance of the PS lectin in the crystal structure is that the lectin is bent around the flexible region so that the two binding sites are closer to each other.

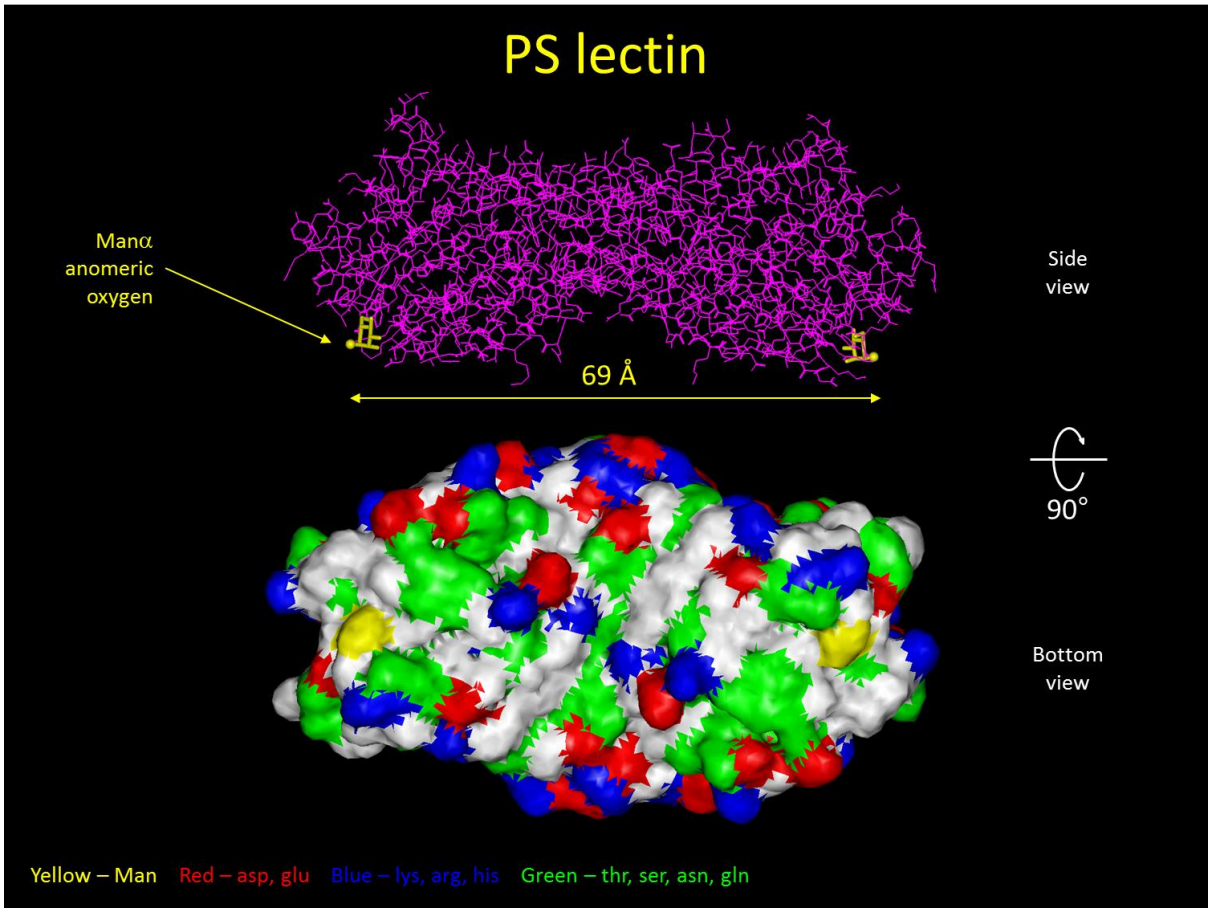



**Figure 3.43.** Crystal structure of PS lectin and surface mapping.

Figure 3.44 depicts the docked PS-lectin binding bivalently to the bivalent mannopore. The $PEG_{10}$ linkers depicted in light blue are quite extended, and have to make a sharp bend to accommodate the mannose via an α-linkage in the binding site. But bivalent binding is possible. Figure 3.45 provides an enlarged view of the $PEG_{10}$ linker configuration upon binding. This Figure also shows the bent conformation of the PS lectin. A change in this conformation could lead to the linkers being insufficiently long for

binding which could explain the behaviour observed in the recordings. Bivalent binding can exist as long as the PS lectin is in a bent configuration, as soon as in solution the PS lectin changes to a more elongated configuration, tension forms on the $PEG_{10}$ linkers and the whole complex falls apart rapidly.

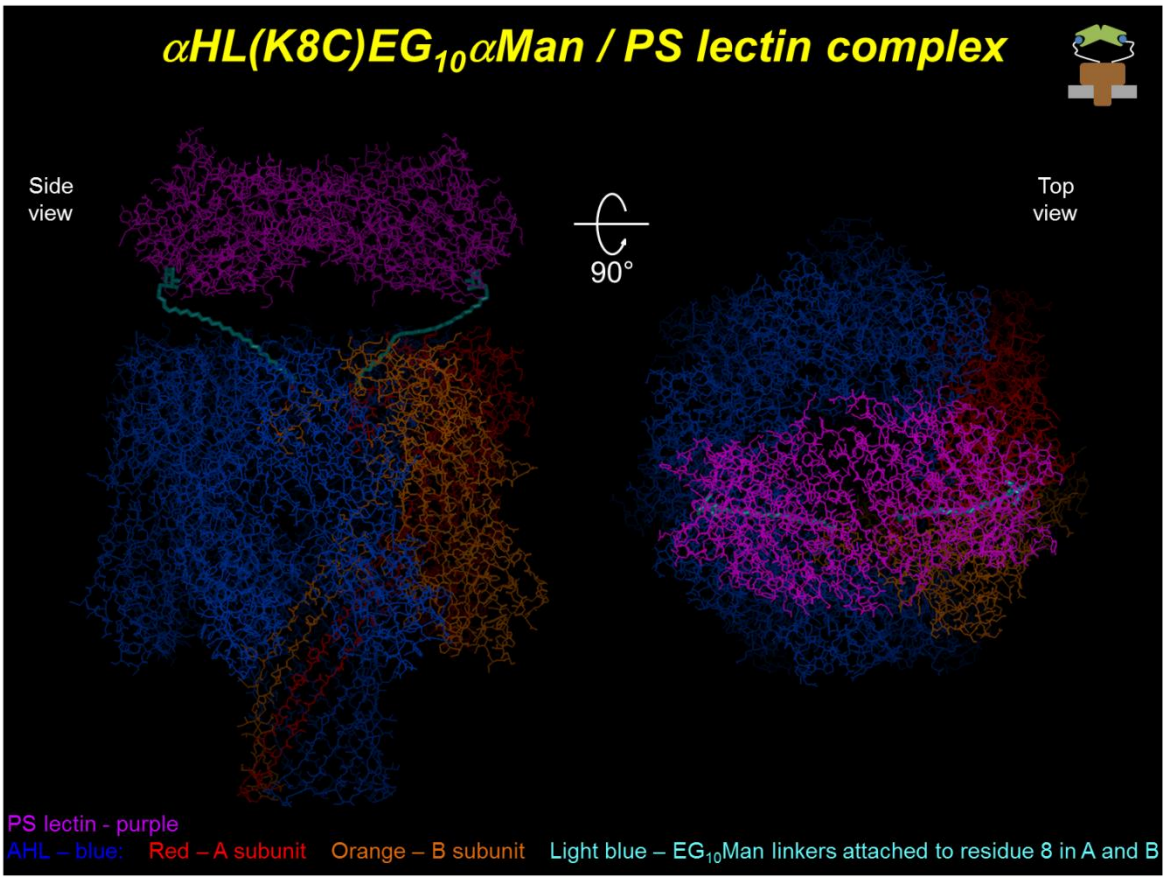



**Figure 3.44.** PS lectin docked on a bivalent mannopore.

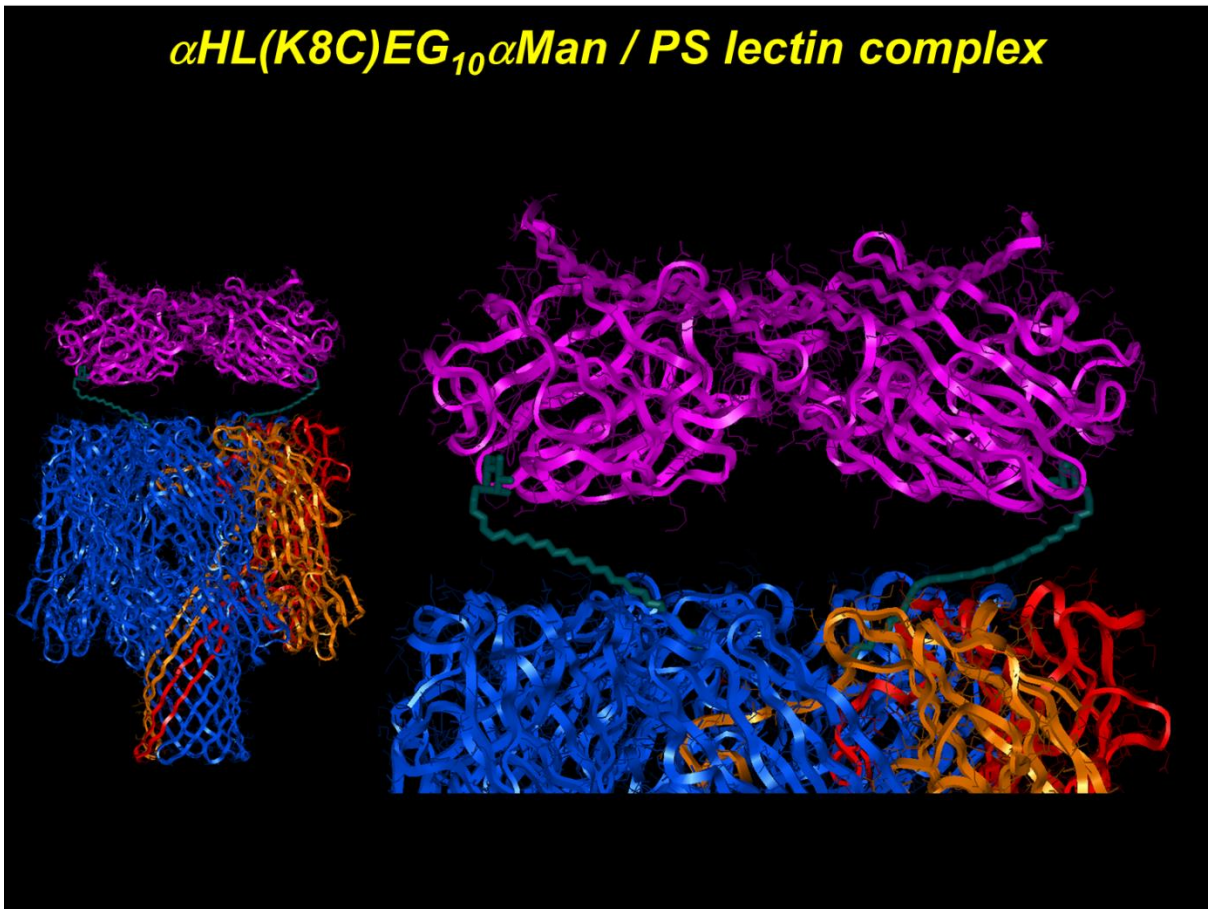



**Figure 3.45.** Detail of the docked binding between PS lectin and a bivalent mannopore.

## 3.10 Langerin



**Figure 3.46.** View of a crystal structure[65] of Langerin and a cartoon representation of Langerin as used herein.

Langerin as depicted in Figure 3.46 is found on the surface of Langerhans cells, a subset of dermal dendritic cells.[64] The protein is a trans - membrane protein, and it forms spontaneous trimers via the outer cellular neck domain via coiled coil interactions between the α-helices.[65] The carbohydrate recognizing domain (CRD) of Langerin is positioned in the head of the protein and is dependent on the presence of a calcium ion, thus Langerin is a c-type lectin.[65] Langerin is specific for mannose and its function is to recognise and internalise pathogens. The lectin is also involved in the formation of birbeck granules in the Langerhans cells, the function of these granules is still not fully understood.[64] Franc Fieschi has kindly provided the Langerin protein needed in the following experiments.[64,65]

### 3.10.1 Langerin monovalent pore

Experiments were carried out with Langerin and the monovalent mannose pore. Different amounts of Langerin were added to the cis-chamber so that the concentration on langerin in the cell ranged from 1 μM to 4 μM in increments of 1μM. Again short events can be detected as shown in Figure 3.47 with an average amplitude of 8 pA from the base line which is smaller than the amplitude measured when using FimH. Dwell time analysis leads to a $k_{on}$= 231 ± 21 $s^{-1}M^{-1}$ and a $k_{off}$= 1013 ± 103 $s^{-1}$. The ratio of the two k values gives a $K_d$= 4.38 ± 0.80 mM. These results are based on three independent experiments. Histograms of the dwell time in each state show one population of events, giving an indication that all the events are caused by the same phenomena.

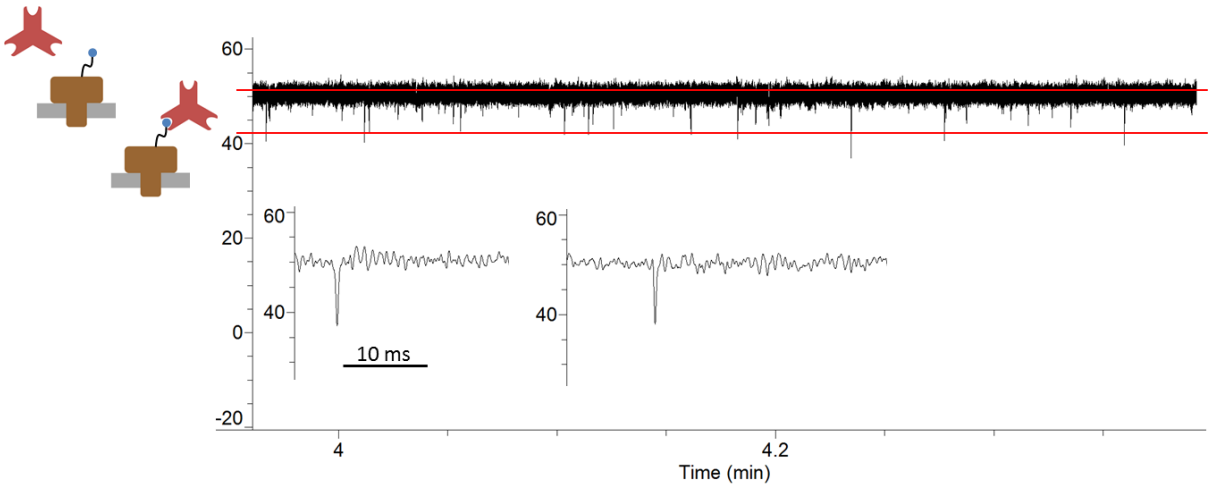



**Figure 3.47.** Detail of a trace recorded during interaction of Langerin with a monovalent mannopore.

The concentration dependence of the events was plotted (Figure 3.48) and it was determined that the inverse of the $\tau_{on}$ was concentration dependent, which is expected for the transition comprising the binding of free Langerin to the monovalent mannopore. The inverse value of $\tau_{off}$ is concentration independent which is in line with the transition of Langerin bound to the mannopore to the unbound state.

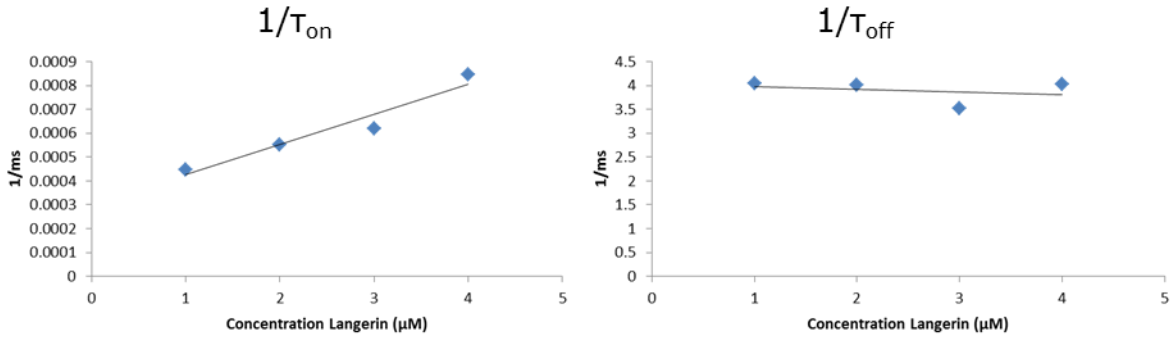



**Figure 3.48.** Concentration dependence for the monovalent binding interactions for langerin.

A binding model was fitted to the observed data as shown in Figure 3.49 comprising two states, i.e. the monovalently bound state (1) and the unbound state (2). The fit between the Gaussian distributions and the histograms indicate the quality of the model presented.

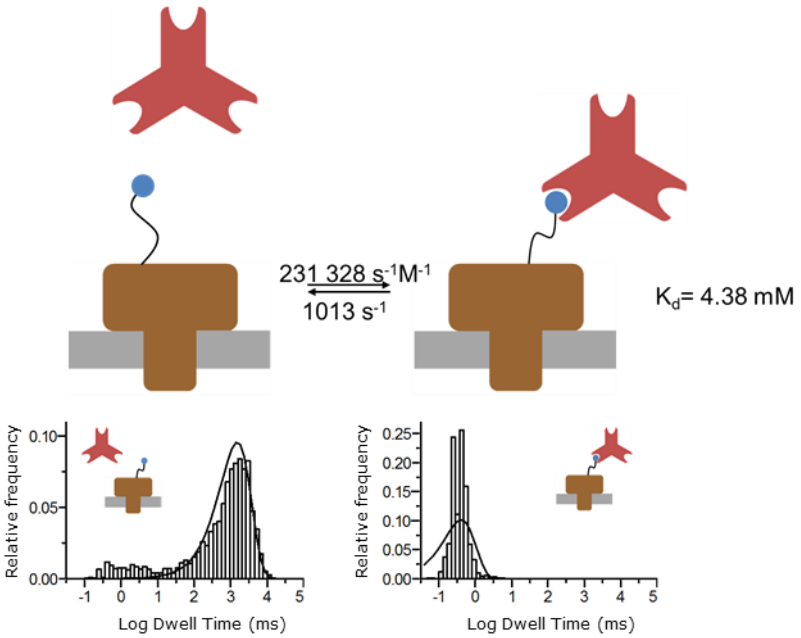



**Figure 3.49.** Model for the interaction between Langerin and a monovalent mannopore.

### 3.10.2 Langerin Bivalent

In the same recording buffer, traces were collected for Langerin and the bivalent mannose pore, using a concentration range from 1 µM to 4 µM. Complex events were observed, with two different levels as can be seen in Figure 3.50, a first level with an amplitude similar to the monovalent binding. Out of this level, events with a larger blockage of the pore start and in most cases the dwell time of these events is short. This observation is similar to the other multivalent ligands but the levels are different for

each ligand. Some examples are shown in the Figure 3.50, note that the time scale is different for the last example.

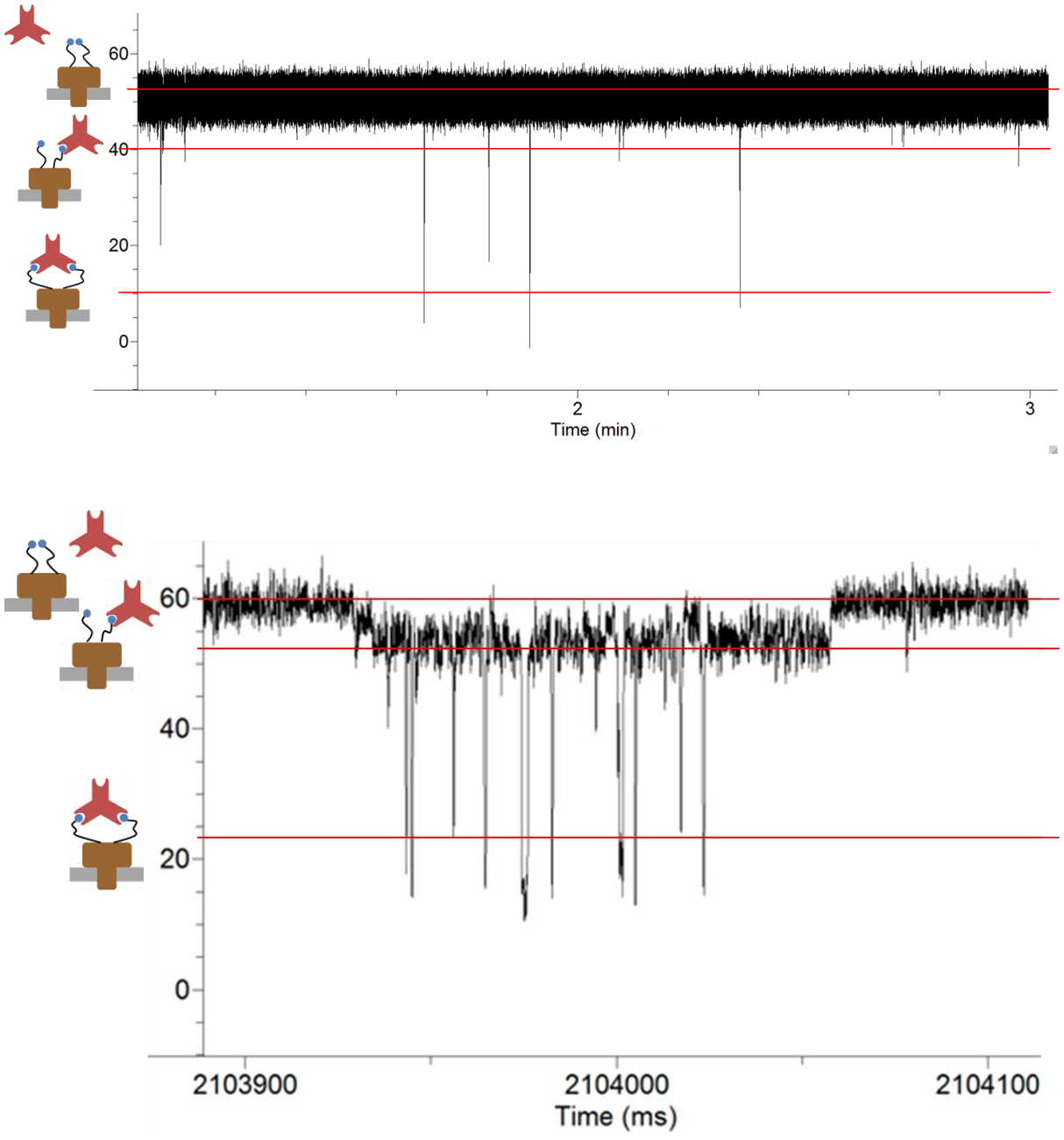



**Figure 3.50.** Detail of a trace recorded during interaction of Langerin with a bivalent mannopore.

A linear binding model was fitted to the experimental data as shown in Figure 3.51. The model comprises three states, one corresponding to the unbound state (1), a state corresponding to the monovalent bound state (2) and a state (3) corresponding to the bivalently bound state. The transition between the unbound state and the monovalent state is characterised by a $k_{12}$ of 18666 $s^{-1}M^{-1}$ and a $k_{21}$ of 231.12 $s^{-1}$, the transition between the monovalent bound state and the bivalent state is characterised by a $k_{23}$ of 28.72 $s^{-1}$ and a $k_{32}$ of 428.19 $s^{-1}$. Transitions between the bivalently bound state and the unbound state have not been observed and are excluded to obtain the best fitted model. This means that once a Langerin molecule is bound monovalently towards the mannopore, the transition to the unbound state is faster than the transition to the bivalently bound state. According to the observations, very few monovalent binding events result in a bivalent binding event. Once a Langerin molecule is bivalently bound to the mannopore, the complex will not fall apart directly to the unbound state as in the case with PS-lectin, but will first return to the monovalent binding state before returning to the unbound state completely.

In an effort to determine a β-value to quantify the multivalent effect an overall $K_d$ for the multivalent binding was determined of 7.98 ± 1.07 mM, as a $k_{on}$ was observed of 1.87 $10^4$ $s^{-1}M^{-1}$ and a $k_{off}$ of 149 $s^{-1}$. This results in a β-value of 0.549 ± 0.179, meaning that this multivalent binding interaction is an example of negative cooperatively. These results are based on three independent experiments.

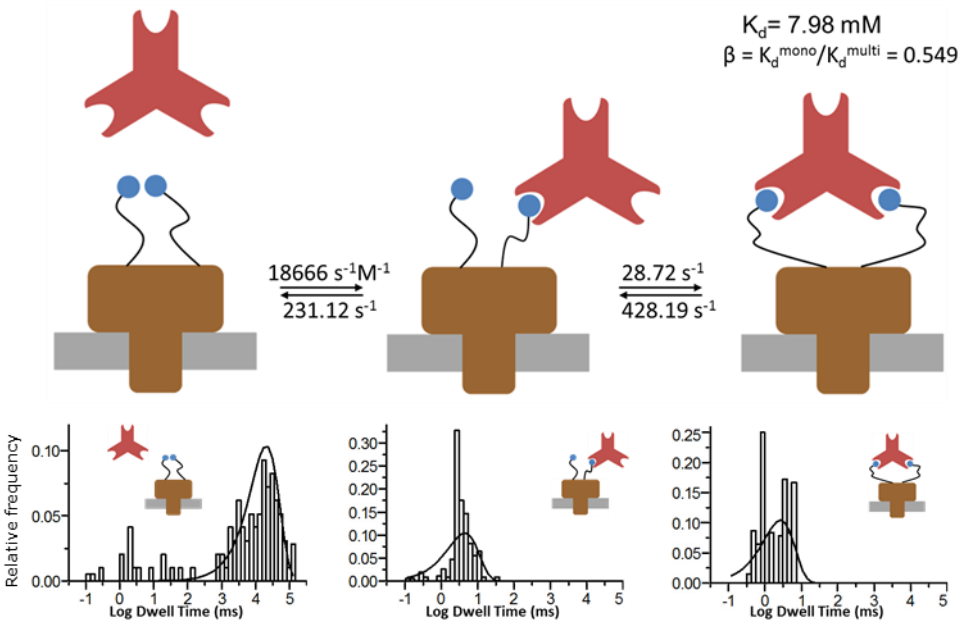



**Figure 3.51.** Model for the interaction between Langerin and a bivalent mannopore.

## 3.11 DC-SIGN

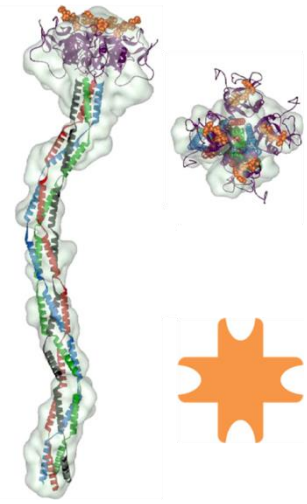


**Figure 3.52.** View of a crystal structure of DC-SIGN and a cartoon representation of DC-SIGN as used herein.

DC-SIGN[66-71] was provided by Dr. Franc Fieschi, from the University of Grenoble.

### 3.11.1 DC-SIGN and monovalent mannopore

Experiments were carried out where DC-SIGN was added to the SCR measuring cell comprising a monovalent mannopore in a recording buffer. Different amounts of DC-SIGN were added to the cis-chamber to obtain a concentration of DC-SIGN in said chamber of 1 to 4 µM in steps of 1 µM. Figure 3.53 shows a trace that has been recorded showing very few and very short events. A detail of a typical binding event is shown in the same Figure 3.53.

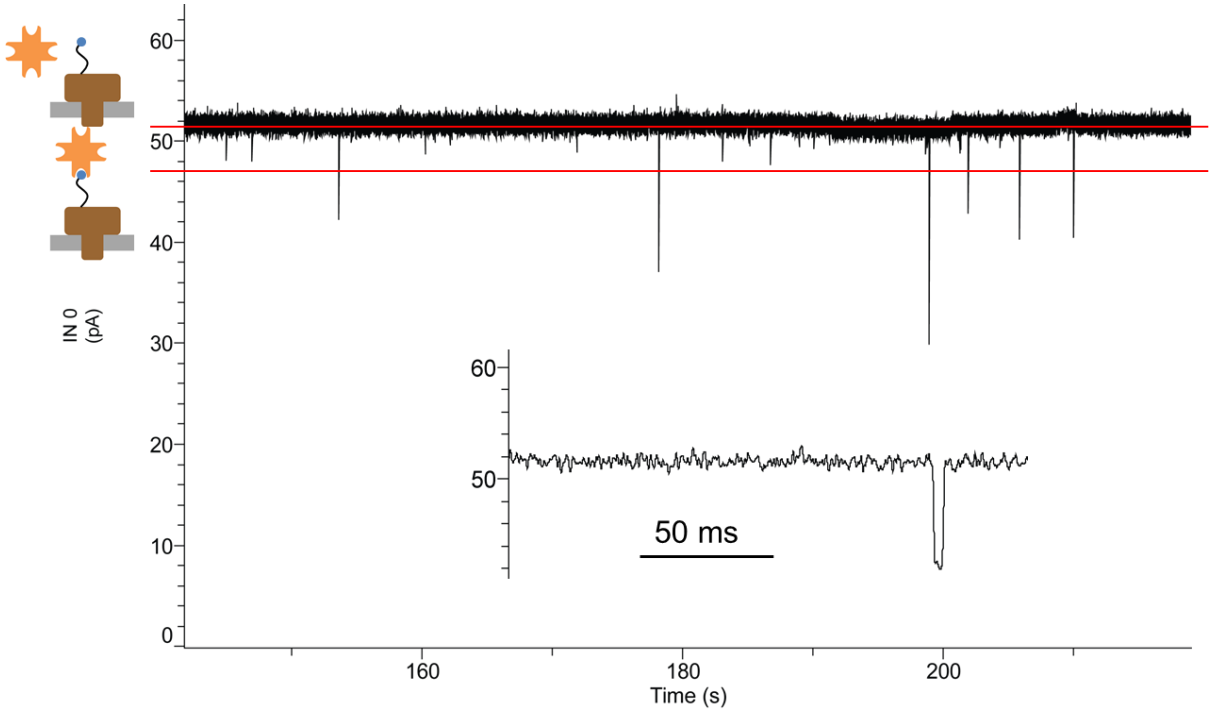



**Figure 3.53.** Detail of a trace recorded during interaction of DC-SIGN with a monovalent mannopore.

Although the events were very short a model could be fitted for the experimental results and is shown in Figure 3.54. The model comprises two states and is characterised by a $k_{on}$ of 41538 ± 1412 $s^{-1}M^{-1}$ and a $k_{off}$ of 702

± 86 $s^{-1}$. This resulted in a $K_d$ for the monovalent binding of 16.9 ± 2.7 mM. These results are based on three independent experiments.

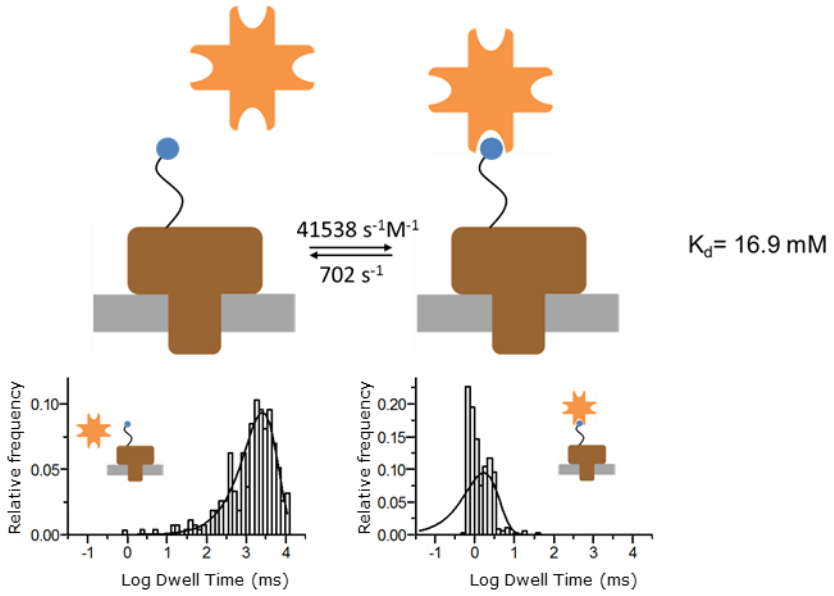



**Figure 3.54.** Model for the interaction between DC-SIGN and a monovalent mannopore.

### 3.11.2 DC-SIGN and bivalent mannopore

Using similar recording conditions and concentrations of DC-SIGN, experiments with a bivalent mannopore were carried out. A detail of a recorded trace can be seen in Figure 3.55. Very few complex events can be seen, but the length of the complex events is much longer than monovalent events. A detail is shown in the same Figure 3.55.

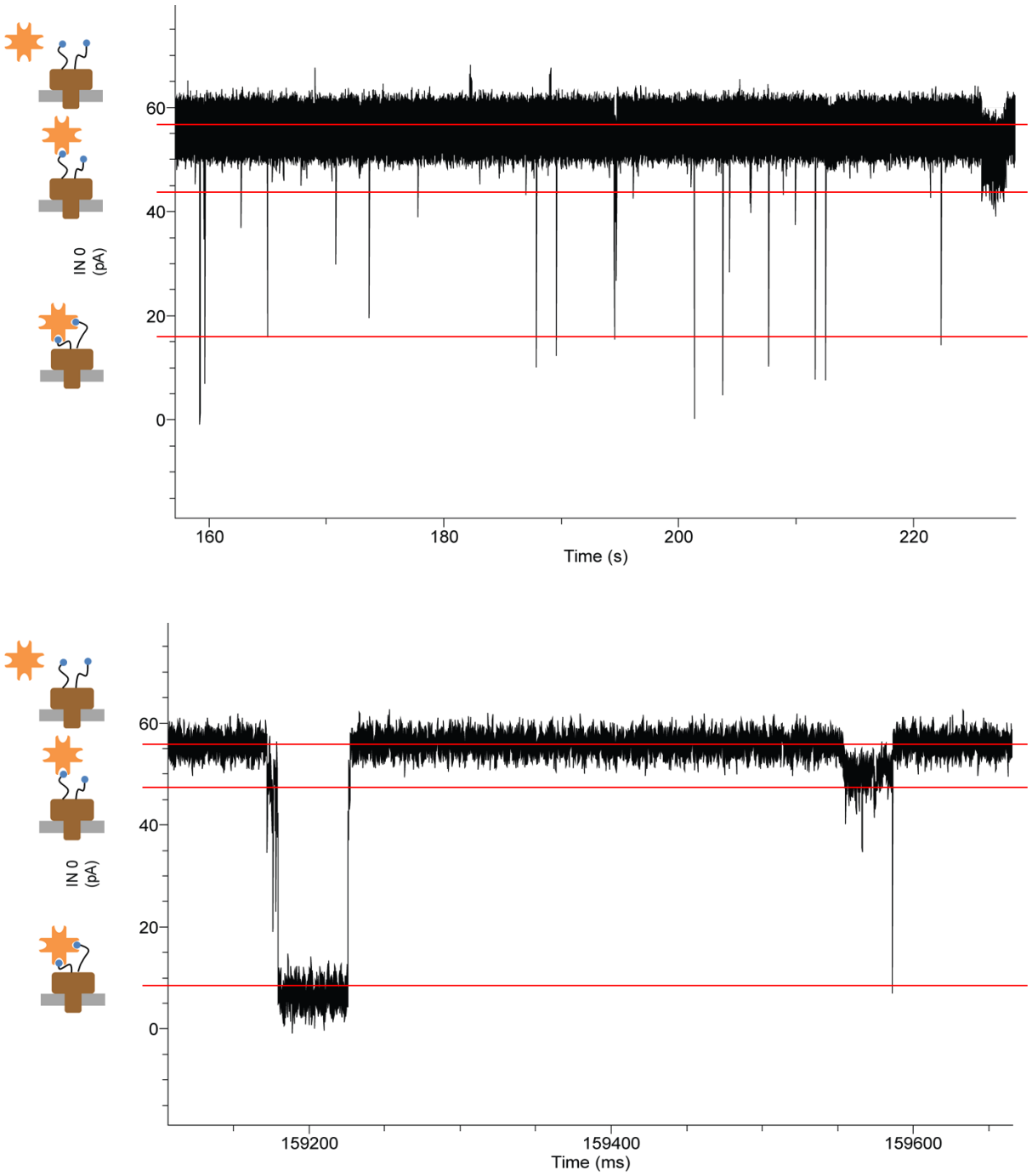



**Figure 3.55.** Detail of a trace recorded during interaction of DC-SIGN with a bivalent mannopore.

A linear binding model was fitted to the experimental data as shown in Figure 3.56. Here again, the model comprises three states, one corresponding to the unbound state (1), a state corresponding to the monovalent bound state (2) and a state (3) corresponding to the bivalently

bound state. The transition between the unbound state and the monovalent state is characterised by a $k_{12}$ of 48937 $s^{-1}M^{-1}$ and a $k_{21}$ of 2418.71 $s^{-1}$, the transition between the monovalent bound state and the bivalent state is characterised by a $k_{23}$ of 73.66 $s^{-1}$ and a $k_{32}$ of 2540.79 $s^{-1}$. Transitions between the bivalently bound state and the unbound state have not been observed and are excluded to obtain the best fitted model. This means that once a DC-SIGN molecule is bound monovalently towards the mannopore, the transition to the unbound state is very fast and much faster than the transition to the bivalently bound state. This is according to the observations, events are very short and very few monovalent binding events result in a bivalent binding event. Once a DC-SIGN molecule is bivalently bound to the mannopore, the complex will not fall apart directly to the unbound state as in the case with PS-lectin, but will first return to the monovalent binding state before returning to the unbound state completely like the Langerin does.

In an effort to determine a β-value to quantify the multivalent effect an overall $K_d$ for the multivalent binding has been determined of 2.54 ± 0.30 mM, as a $k_{on}$ was observed of 4.89 $10^4$ $s^{-1}M^{-1}$ and a $k_{off}$ was observed of 124.3 $s^{-1}$. This results in a β-value of 6.66 ± 1.8, meaning the multivalent binding is 6.66 time stronger than the monovalent binding. These results are based on three independent experiments.

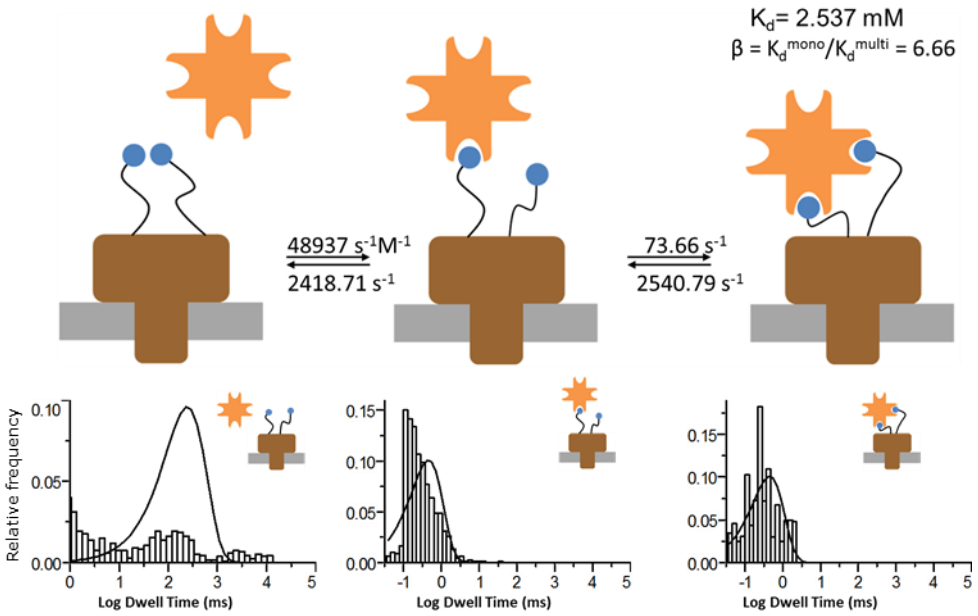



**Figure 3.56.** Model for the interaction between DC-SIGN and a bivalent mannopore.

## 3.12 Discussion and Conclusion

The experiments described above show a significant difference in binding behaviour of multivalent lectins, to a monovalent mannopore compared to a bivalent mannopore. The binding events to a monovalent mannopore are stochastic for each lectin, but a single level of blockage can be defined. The dwell time of the binding events forms a single population, which has a defining $\tau_{on}$ and $\tau_{off}$ values for each lectin. As expected, the inverse of the $\tau_{on}$ value was concentration-dependent. From these $\tau_{on}$ and $\tau_{off}$ values, dissociation constants $K_d$ were derived for each lectin. The $K_d$ values differ from those in literature for the relevant lectin with mannose. I postulate that for the mannopores, the movement of mannose is restricted due to its attachment to the α-HL protein, which may influence the observed $K_d$ values.

Binding events of the multivalent lectins to the bivalent mannopore were observed to be more complicated. A typical example of a complex binding event for each lectin is shown in Figure 3.57. For all four tested lectins, the first level of blockage can be observed in a similar region as in the experiments using monovalent mannopores. It has been observed that the transition between the first level and the base level is lectin-concentration dependent. Hence, it can be assumed that this first level corresponds to monovalent binding of the lectin to the bivalent mannopore.

Besides the first level of blockage, all four tested lectins display a lower

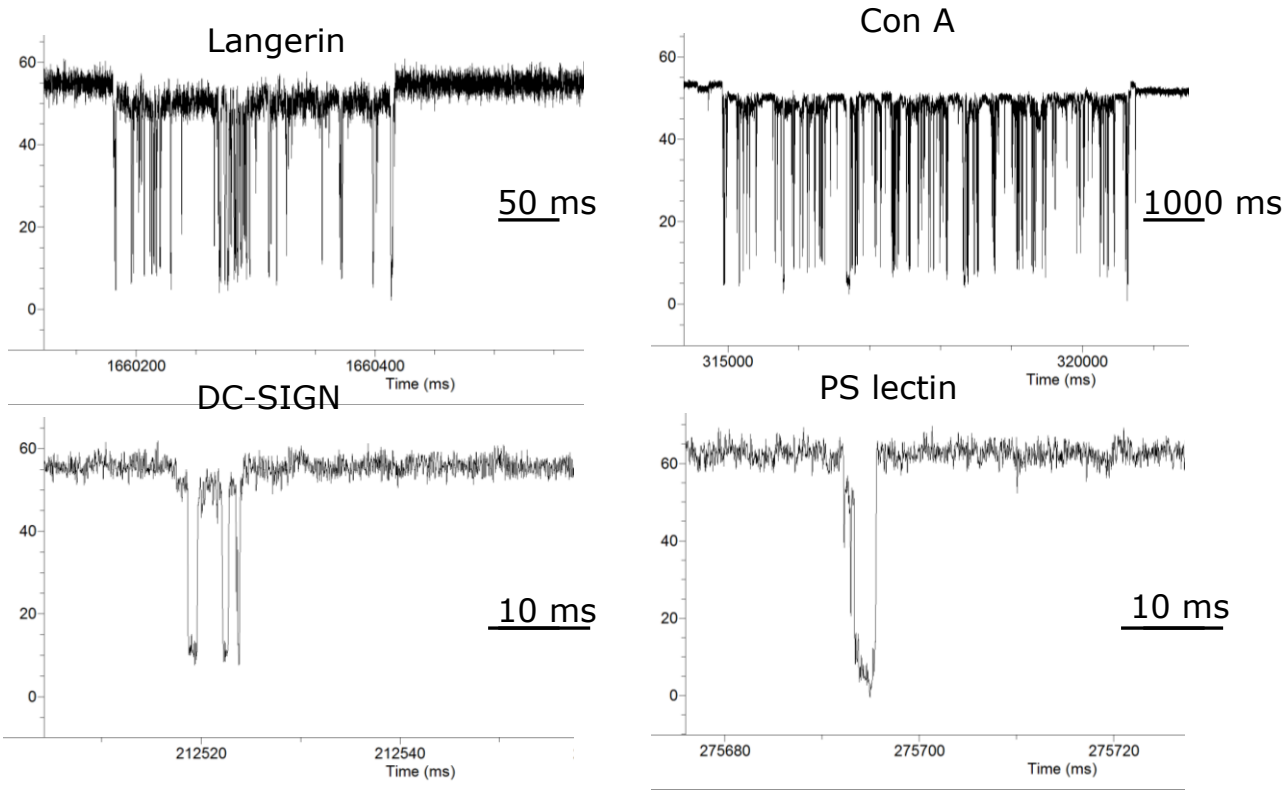



**Figure 3.57.** Comparison of the different bivalent binding events.

level. It appears that the events cause this lower level block the ion flow through the bivalent mannopore much more than the events associated with monovalent binding to the bivalent mannopore. No significant amount of events can be found that start from the base level and reach straight to the lower level. It appears that events reaching the lower level start from a level which corresponds to a level associated with monovalent binding. It has been observed that the transition between the level corresponding to monovalent binding and the lower level is lectin-concentration independent. Hence, this transition can be associated with an intra-molecular event. Therefore, it can be assumed that the lower level corresponds to the bivalent binding of the lectin to the bivalent mannopore.

Besides the observations mentioned above, there are some significant differences observed in the multivalent binding behaviour for the different lectins. For Con A, two levels could be identified in the region corresponding to a monovalent binding. Multiple transitions between these two levels were observed. However, it was observed that events reaching the lower level associated with bivalent binding originated from one single level associated with monovalent binding. It was therefore postulated that only one of the intermediate levels is associated with a monovalent binding event which is not accessible for the other mannose ligand attached to the bivalent mannopore. This postulation may be supported by the modelling of the system, as the four binding sites of Con A are asymmetrically distributed over the surface, the two mannose ligands on the bivalent mannopore may not be long enough to reach each possible combination of the four different binding sites.

In the experiments with the PS lectin, it could be observed that the complex events end with a transition from the lower level straight to the baseline. For the other three lectins, the complex events end with a transition from the lower level to the level associated with monovalent binding, followed by a further transition to the baseline. Hence, it was postulated that the bivalent binding complex with PS and the bivalent mannopore disintegrates faster than can be observed in the recordings.

For each lectin, a binding model was proposed taking into account the postulates mentioned above, see Figures 3.30, 3.42, 3.51 and 3.56. Experimental data were matched against these proposed binding models and a fit between the data and the models was observed.

| Lectin | $\tau_{off}$ monovalent (ms) | $\tau_{off}$ multivalent (ms) | $K_d$ monovalent (mM) | $K_d$ multivalent (mM) | β |
|---|---|---|---|---|---|
| Fim H | 0.27 | - | 34±3 | - | - |
| ConA | 176 | 794 | 0.875±0.053 | 0.192±0.033 | 4.56±1.06 |
| PS | 14 | 82 | 2.2±0.2 | 1.03±0.06 | 2.13±0.32 |
| Langerin | 0.99 | 6.71 | 4.38±0.80 | 7.98±1.07 | 0.549±0.179 |
| DC SIGN | 1.42 | 8.05 | 16.9±2.7 | 2.54±0.30 | 6.66±1.8 |

**Table 3.1** Overview of binding properties for the Fim H, ConA, PS, Langerin and DC SIGN to the monovalent mannopore and the bivalent mannopore.

It was further observed that the average length or dwell time of the complex events varies significantly for each lectin, see Table 3.1. For Con A, PS-lectin and DC SIGN, positive cooperativity ($\beta > 1$) could be observed. However, for Langerin negative cooperativity ($\beta < 1$) was observed. Unfortunately, no explanation for this negative cooperativity could be given.

# 4 Carbohydrate synthesis

In an effort to improve the dwell time of the binding events of DC-SIGN with the mannopore, several new types of ligands were designed and syntheses attempted. All ligands were designed with the same decaethylene glycol ($PEG_{10}$) linker as in previous experiments. Three ligands were selected and efforts to synthesise the ligands are discussed below.

## 4.1 $PEG_{10}$ Synthesis

Monodisperse ethylene glycol oligomers are commercially available up to six ethylene glycol units, but higher oligomerisation states of ethylene glycol need to be synthesised. The synthesis of $PEG_{10}$ is depicted in Figure 4.1.[72] For the synthesis, two diethylene glycol molecules **(1)** were added to both ends of a hexaetheylene glycol molecule in the following way: firstly, the diethylene glycol **(1)** was singly protected by using one equivalent of tritylchloride to give **(2)**. Both hydroxyl groups of the hexaethylene glycol **(4)** were tosylated to give **(3)**. These compounds **(2)** and **(3)** were mixed together with sodium hydride overnight to yield a protected decamer **(5)**[72]. Three methods were used for the removal of the trityl protecting groups on the decaethylene glycol **(5)**; hydrogenation in the presence of palladium on carbon was very slow, whilst the use of TFA or HCl with water led to faster reactions. Unfortunately, degradation of the ethyleneglycol chain was also observed to a small extent.[72]

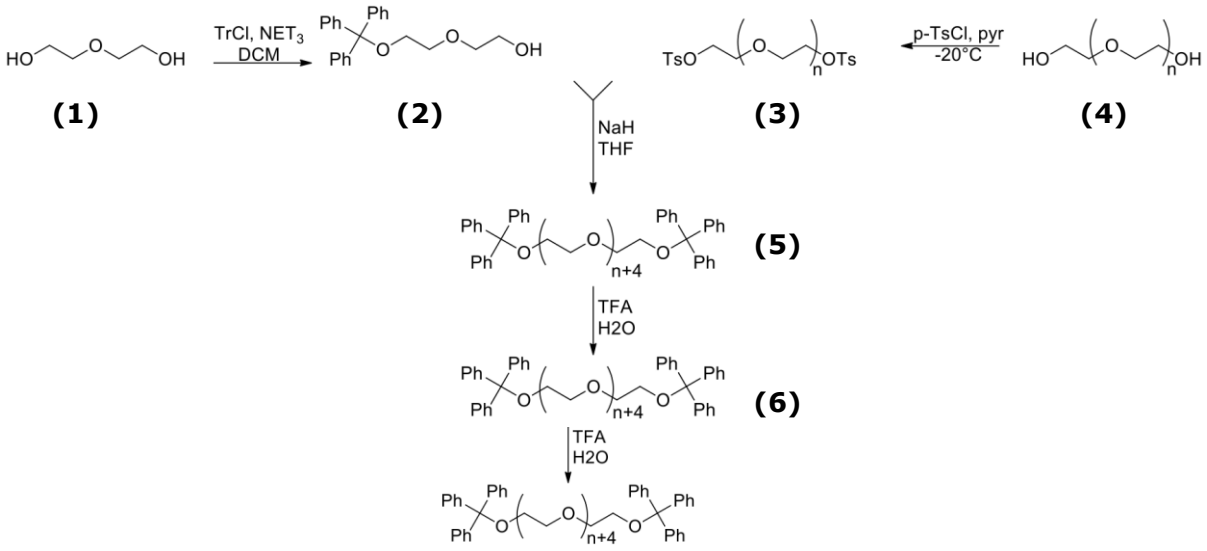



**Figure 4.1.** Synthesis of $PEG_{10}$.

## 4.2 Mannose D1 arm

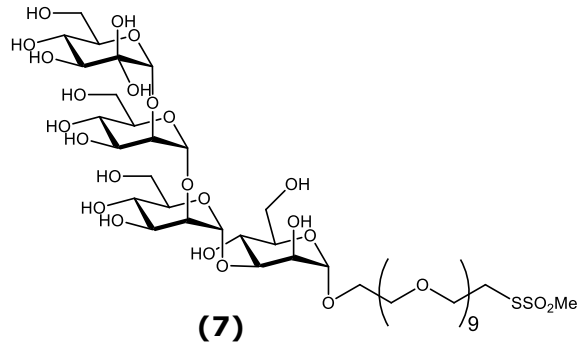



**Figure 4.2.** D1 arm to $PEG_{10}$ conjugate.

The mannose D1 arm is part of a mannose structure with 9 mannose molecules that is commonly found in the glycocalyx.[74,87]

One of the strongest carbohydrate ligands for DC-SIGN is $Man_9GlcNAc_2$ with a $K_d$ = 180 µM.[74] The D1 arm of this oligosaccharide -a tetrasaccharide- acts as a good mimic and has a $K_d$ of 350 µM.4

## 4.3 Convergent route

To synthesise the D1 arm to $PEG_{10}$ conjugate **(7)**, see Figure 4.2, a convergent synthesis route was proposed, see Figure 4.5. This would enable flexibility in the nature of sugar molecules at the non-reducing end. The route would allow the

synthesis of a different number of carbohydrate moieties and further, the incorporation of non-natural carbohydrates. The route described led to a mannose 1-3 disaccharide **(20)** from donor molecule **(13)** and acceptor molecule **(19)**.This disaccharide **(20)** functions as a donor in the glycosylation reaction of the decaethylene glycol with an orthogonal donor group to the monovalent donor **(13)**.

### 4.3.1 Donor Synthesis

As shown in Figure 4.3, the synthesis of the donor started from penta-acetyl mannose **(8)** which was converted in the first step to the mannose bromide (structure not shown) which was converted to the methylortho-ester **(9),** using lutidine. The following step involved protecting group manipulation in which the acetyl groups were exchanged for benzyl groups yielding **(10)**. Opening of the ortho-ester with ethanethiol resulted in the thiolate **(11)** in the alpha configuration. The last step was the deprotection of the acetyl group **(12)**, and protection of the free hydroxyl function with a levunilate protecting group. This led to donor **(13)**, which was later used in glycosylation reactions. To make the mannose 1-3 disaccharide this thioglycoside donor was converted into a TCA donor **(15)**. This allowed the use of orthogonal glycosylation conditions to prevent the thiolate on acceptor molecule **(19)** from reacting in the formation of the disaccharide. Donor **(15)** was prepared from donor **(13)** by hydrolysis of the anomeric ethane thiol group **(14)** followed by reaction trichloroacetonitrile and DBU.[74-77]

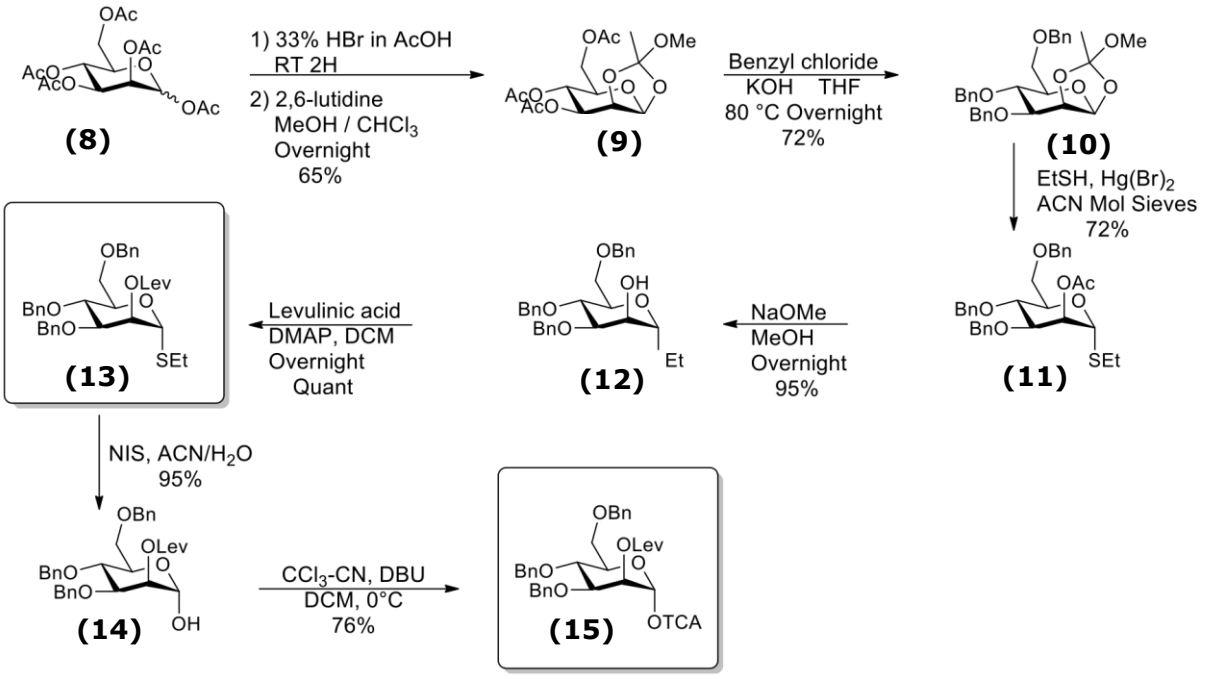



**Figure 4.3.** Synthesis of donor molecules.

## 4.3.2 Acceptor Synthesis

As shown in Figure 4.4, the synthesis of the acceptor **(19)** started from the penta-acetyl mannose **(8)**, which was first converted to the alpha-phenoltiolate **(16)**. The selective benzylidene protection of the hydroxyl functions in the C-4 and the C-6 position failed and the major isolated product was compound **(17)** with double benzylidene protection. Compound **(17)** was converted to compound **(18)** by exploiting the difference in reactivity between the less stable five membered ring over the six membered ring. Reaction of compound **(17)** with 10 % w/w phosphomolybdic acid (PMA)/silica resulted in the target compound **(18)**.[78-80] The last step in the synthesis of the acceptor **(19)** was the selective protection of the hydroxyl function on position two. This is done by forming the ortho-ester and selective opening resulted in the target compound **(19)**.

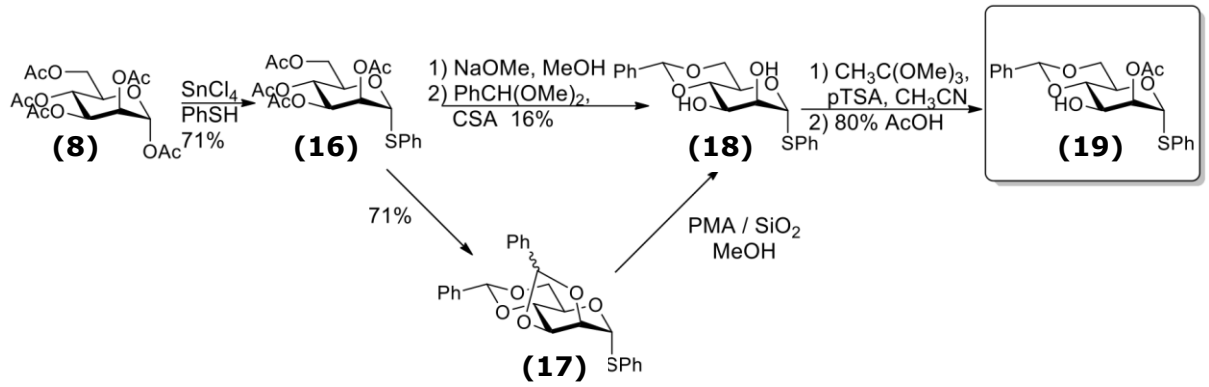



**Figure 4.4.** Synthesis of acceptor molecules.

### 4.3.3 Glycosylation

As shown in Figure 4.5, activation of donor **(15)** with boron trifluoride diethyl etherate in the presence of acceptor **(19)** yielded the disaccharide **(20)** in a 67% yield. Unfortunately, the glycosylation of this disaccharide to the $PEG_{10}$ molecule was problematic. No reaction was observed and $PEG_{10}$ was recovered from the reaction mixture. The donor **(15)** was observed as hydrolyzed product after work up or was seen as the methoxylated compound by mass spectrometry during the reaction. This shows that the activation was successful and the problem is the reactivity of the acceptor ($PEG_{10}$). Hence, the convergent route was abandoned.

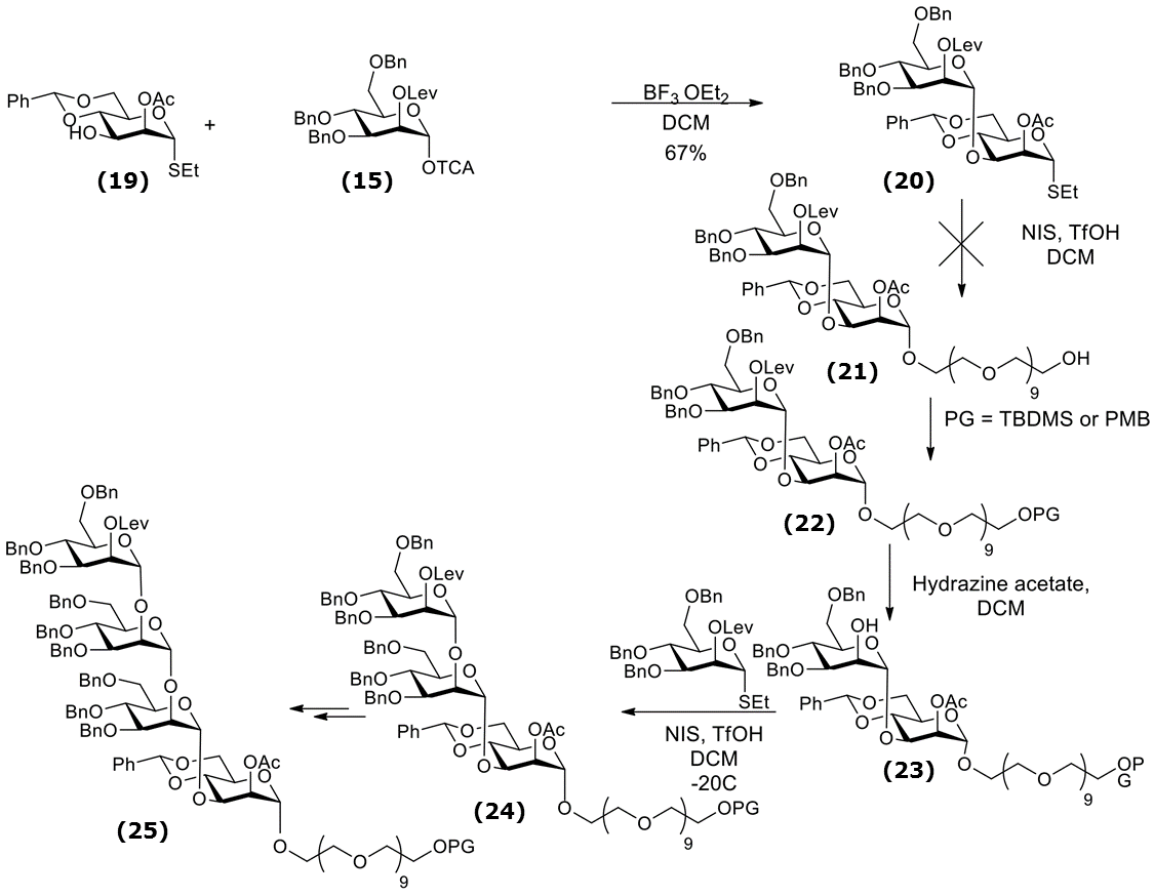



**Figure 4.5.** Convergent route.

## 4.4 Linear route

A linear route to compound **(25)** was attempted as an alternative to the convergent route. The first reaction was the glycosylation of the donor **(19)** with two equivalents $PEG_{10}$. This reaction was very low yielding. A series of test reactions, see Figure 4.6, was performed with different degrees of oligomerization of ethylene glycol. The yield of the reactions decreased to 27 % for the $PEG_{10}$ when the polymerization numbers increased. In the case of higher degrees of polymerization side reactions were observed. Acetylated $PEG_{10}$ **(32)** could be isolated. The NMR of the compounds showed three different acetyl peaks with almost the same intensity.

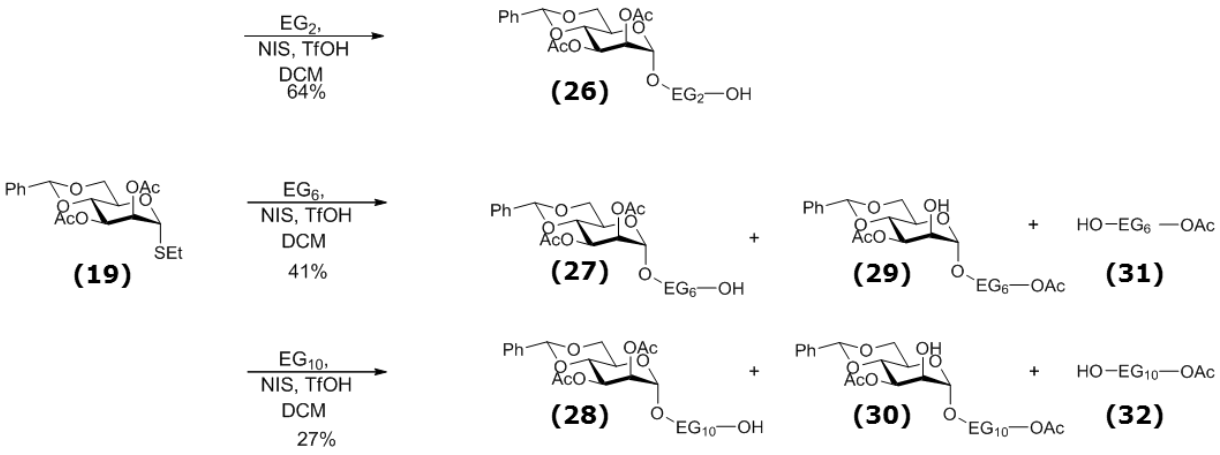



**Figure 4.6.** Donor - $PEG_{10}$ glycosylation.

This last observation can be explained by a different undesired opening of the ortho-ester during the acid-catalysed glycosylation reaction. This would result in the acetylated PEG molecules and this PEG molecule could undergo a glycosylation reaction with the donor that had lost the acetyl protecting group on the C-2 position. Separation of the compounds was very problematic due to similar $R_f$ values, and for this reason this route was abandoned.

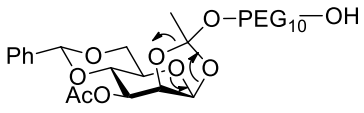


**Figure 4.7.** Ortho-ester opening.

A new linear synthesis, see Figure 4.8, was proposed where the low yielding PEGylation of compound **(8)** was the first reaction in the sequence. Again here a mixture of different compounds **(36)** was obtained and separation was almost impossible. Only the dimannosylated compound could be separated from the rest. The mixture was treated with sodium methoxide in order to remove all the acetyl protecting groups. This led to the single compound **(37)** in a 54% yield over two steps. The hydroxyl groups of the sugar were benzylidene protected to give **(38)**,

leaving only the primary hydroxyl group of the $PEG_{10}$ linker free. This hydroxyl group was TBDPS-protected to give **(39)**. Selective deprotection of one benzylidene group of compound **(39)** gave **(40)**. The ortho-ester was formed between the hydroxyl groups at position two and three of compound **(40)** before selective opening, yielding the acceptor **(41)** needed to start the D1-arm synthesis.[59, 81]

This synthetic route was also abandoned due to low yields when scaling up the reactions. The route would be inefficient for obtaining enough material to complete the synthesis, and carry out single channel recording experiments using the obtained compounds.

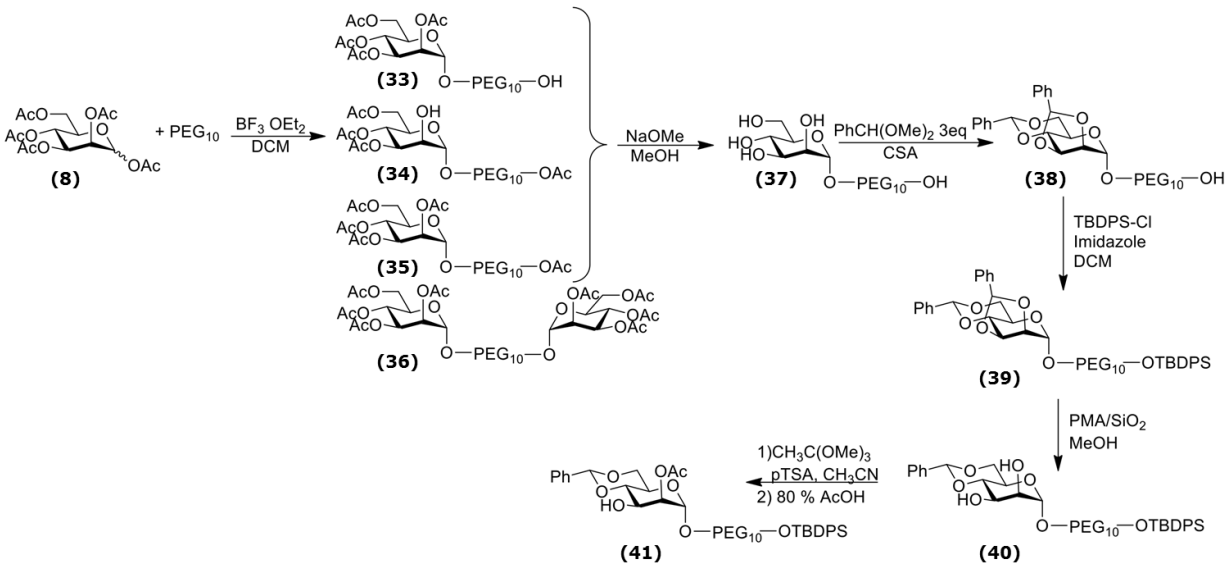



**Figure 4.8.** Linear route.

## 4.5 *2-C* branched sugar analogues

### 4.5.1 Synthesis of 2-*C*-substituted branched D-mannose analogues

It has been shown by Fleet that 2-*C*-substituted branched D-mannose have a higher affinity for DC SIGN compared to mannose.[88] Hence, efforts were made to functionalise the mannopores with such branched mannose ligands. The starting material for the synthesis of the branched sugars, as shown in Figure 4.9, is **D**-fructose **(42)**. After a Kiliani ascension with sodium cyanide followed by hydrolysis, a 2-C branched sugar (structure not shown) was obtained.[89] Treating this sugar with acetone and an acid gave the protected branched mannose **(43)** in a 25% yield. Esterification of the free hydroxyl group of **(43)** with trifluoromethanesulphonic (triflic) anhydride resulted in triflate **(44)**. Nucleophilic substitution with sodium azide yielded azide **(45)**. The following step was the conversion of the lactone to the lactol **(46)** by treatment with DIBAL-H. To obtain 2-*C*-azidomethyl-**D**-mannose **(47)**, lactol **(46)** was deprotected by stirring for 48h in the presence of DOWEX $H^+$ resin. Hydrogenation of **(47)** with palladium black as a catalyst led to 2-*C*-aminomethyl-**D**-mannose **(48)**.

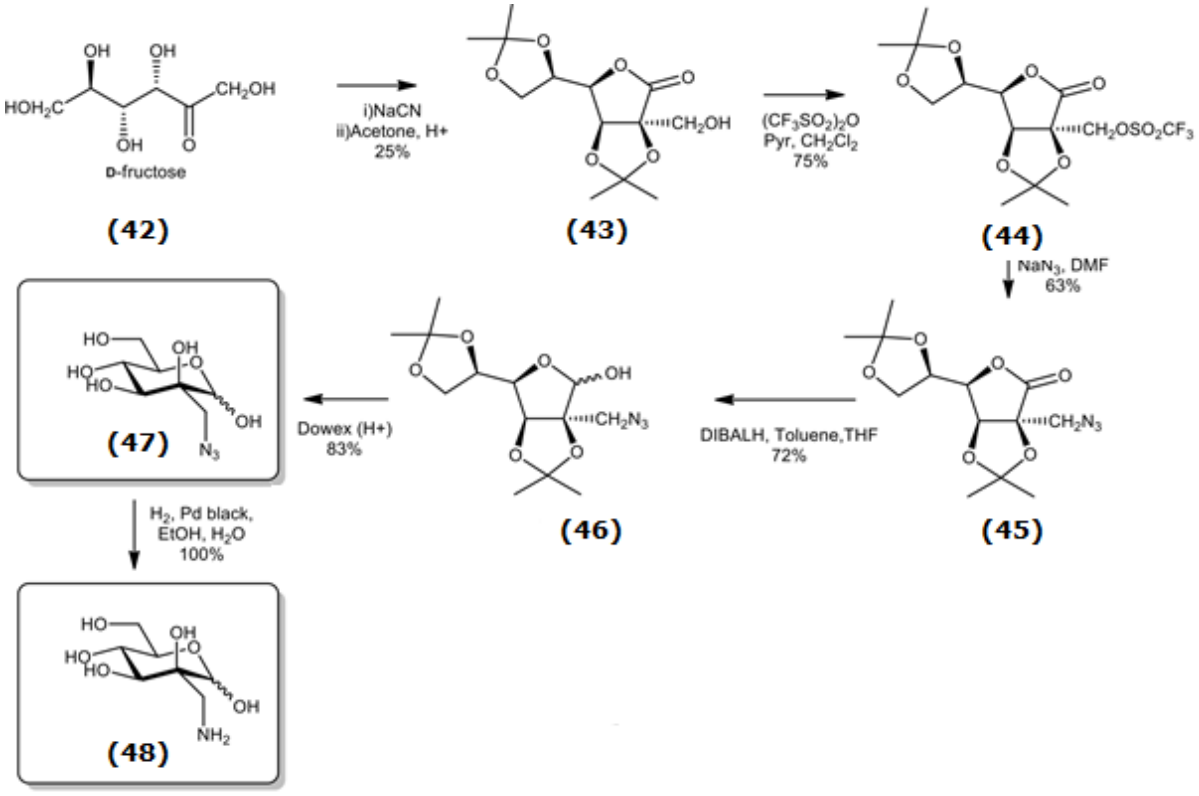



**Figure 4.9.** Synthesis of 2-*C*-substituted branched D-mannose.

To attach compound **(48)** to the pore for single channel recording, a $PEG_{10}$ linker with a terminal -$SSO_2Me$ group needed to be installed in the anomeric position, as shown in Figure 4.10. The synthesis started with the preparation of compound **(54)** and the peracetylated derivative **(49)** as starting material. Treating compound **(54)** with two equivalents of $PEG_{10}$ led to compound **(35)** in a very low yield. Any efforts to mesylate the free hydroxyl group on the PEG linker failed. Hence, also this synthesis route was abandoned. Due to time constraints and after considerable efforts, the idea of synthesising a pore with higher affinity for DC-SIGN was abandoned and the focus shifted to other lectins with higher affinities for the already existing pores.

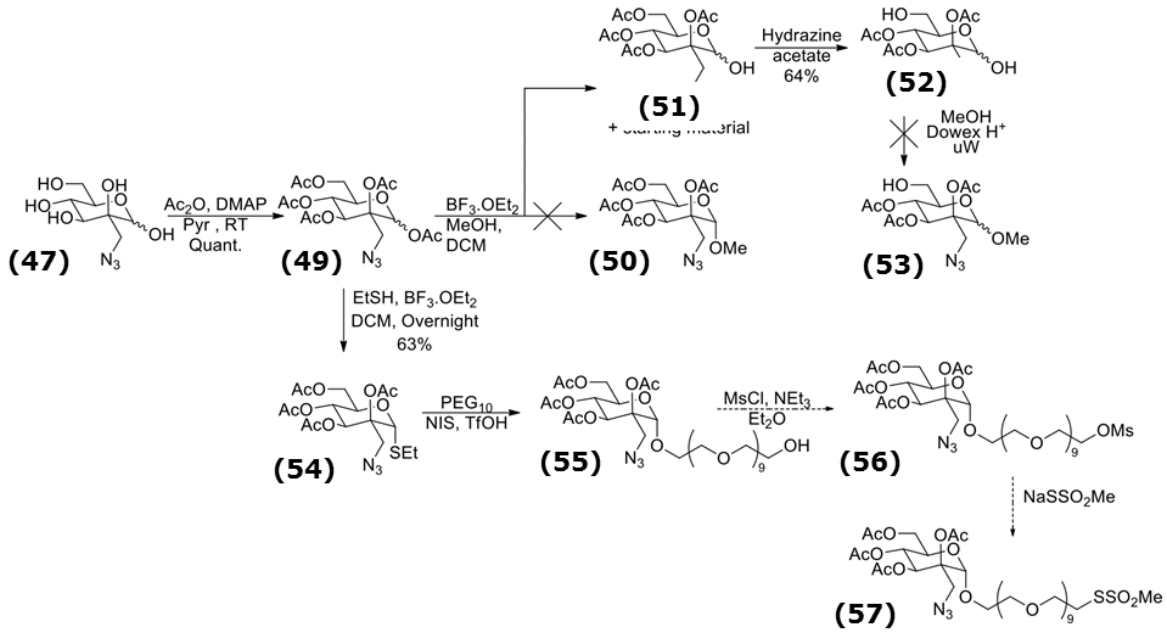



**Figure 4.10.** Glycosylation of 2-*C*-substituted branched D-mannose and $PEG_{10}$.

# 5 Experimental

## General

Proton nuclear magnetic resonance (δH) and Carbon nuclear magnetic resonance (δC) were recorded by Dr. T. Claridge or Dr. B. Odell on Bruker DPX 400 (400 MHz) or on a Bruker AMX 500 (500 MHz) instrument. The assignment of spectra was performed using COSY, DEPT and HMQC experiments. All chemical shifts are quoted on a δ-scale in parts per million (ppm) and are referenced to residual solvent frequencies. For the numbering of carbon atoms and protons attached to said carbon atoms, is based on conventional numbering used for carbohydrates.
Mass Spectra were recorded on a Micromass Platform 1 spectrometer or a Bruker Daltonics Micro-TOF spectrometer. Mass Spectra recordings were done in Daltons. Thin layer chromatography (TLC) was performed on Merck aluminium backed plates, precoated with silica (0.2 mm, 60 F254). Visualization of the plates was done by using a U.V. lamp (λmax=254 or 365 nm) and/or ammonium molybdate (5% in 2M H2SO4) or sulphuric acid (5% in EtOH) or potassium permanganate (0.5% in 1 M NaOH). Optical rotations were measured on a Perkin-Elmer 241 polarimeter with a path length of 1.0 dm; concentrations are given in g/100 mL. Flash column chromatography was carried out using silica gel (Sigma-Aldrich, Merck Grade 9385, 60 Å, 230-400 mesh). The used petroleum ether has a boiling in the range 40-60 °C. Dichloromethane, tetrahyrdofuran and acetonitrile were dried over alumina. All solvents used are anhydrous unless otherwise indicated.

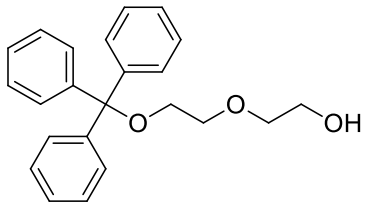
## Single channel recordings

Single channel recordings were carried out in a planar lipid bilayer apparatus.[44] The apparatus comprises a 25 µm polytetrafluoroethylene (PTFE) septum (Goodfellow Corporation, Malvern, PA), separating the two chambers in the apparatus. In the PTFE septum, a 100 µm is present. Before recordings this PTFE septum is pre-treated with a 10 vol% hexadecane solution in pentane. In the aperture, a lipid bilayer is formed from 1,2-Diphytanol-sn-glycero-3-phosphocholine lipids (DPhPC, Avanti Polar Lipids, Alabaster, AL), 25 mg lipids dissolved in 2.5 ml pentane. The chambers were filled with 0.6 M KCl, 0.6 mM $CaCl_2$, 0.3 mM $MnCl_2$, 6 mM Tris-HCl, pH 7.2 or 0.6 M KCl, 25 mM $CaCl_2$, 25 mM Tris-HCl, pH 7.2 for DC-SIGN. To the cis-chamber, the mannopores (0.05 to 0.5 ng/ml) were added and the buffer was stirred until a single pore inserted into the bilayer. The recordings where carried out according to the voltage clamp procedure, at a holding potential of +100 mV, with the cis-chamber connected to the ground. Currents were low-pass filtered with a 4-pole Bessel filter at 2 kHz. The sample rate of the recordings was 10 kHz, and samples were collected with a Digidata 1200 A/D converted (Axon Instruments, Union City, CA). Data analysis was performed as described.[44]

## Modelling

The modelling has been carried out using Modeller, using crystal structure of PDB 7AHL for α-HL, PDB 5CNA for Con A (with α-methyl-D-mannopyranoside), PDB 1RIN for PS Lectin. The models were checked for feasibility and Ramachandran plots were checked using RAMPAGE Ramachandran Plot Assessment.[99]

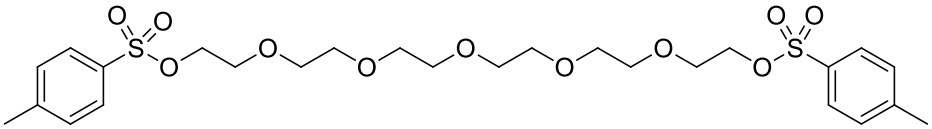
### 7,7,7-Triphenyl-3,6-dioxaheptanol (2)[90]




A solution of diethylene glycol (10.6 g; 100 mmol) and triethylamine (28 mL; 200 mmol) was dissolved in 150ml $CH_2Cl_2$ and cooled to 0 °C. Trityl chloride (27.9 g; 100 mmol) was added. The mixture was allowed to reach room temperature and left overnight. When the reaction was finished as judged by TLC analysis, the mixture was washed with a saturated $NaHCO_3$ solution and the aqueous layer was extracted 3 times with $CH_2Cl_2$. The organic phases where collected and concentrated *in vacuo*. The product **(2)** (14.3 g; 41%) was purified using flash chromatography, using $CH_2Cl_2$ as eluent.

$^1$H NMR (400 MHz, $CHCl_3$-*d*) $\delta$ ppm 2.16 (1H, br s, OH), 3.25 (2H, t, *J*=5.0 Hz, $CH_2OH$), 3.61 (2H, t, *J*=5.0 Hz, $C\underline{H}_2CH_2OTr$), 3.67 (2H, t, *J*=5.5 Hz, $C\underline{H}_2CH_2OH$), 3.73-3.77 (2H, t, *J*=5.0 Hz, $CH_2C\underline{H}_2OTr$), 7.32-7.20 (9H, m, 6 *o*-ArCH, 3 *p*-ArCH), 7.43-7.48 (6H, m, 6 *m*-ArCH) ; $^{13}$C NMR (400 MHz, $CHCl_3$-*d*) $\delta$ ppm 61.9 ($CH_2\underline{C}H_2OTr$), 63.4 ($CH_2OH$), 70.6 ($\underline{C}H_2CH_2OH$), 72.3 ($\underline{C}H_2CH_2OTr$), 86.7 ($O\underline{C}Ph_3$), 127.0 (*p*-Ar$\underline{C}$H), 127.8 (*o*-Ar$\underline{C}$H), 128.7 (*m*-Ar$\underline{C}$H), 144.0 (ArC). MS (ESI-) calculated (M) 348.43; found (2M-H): 695.35.

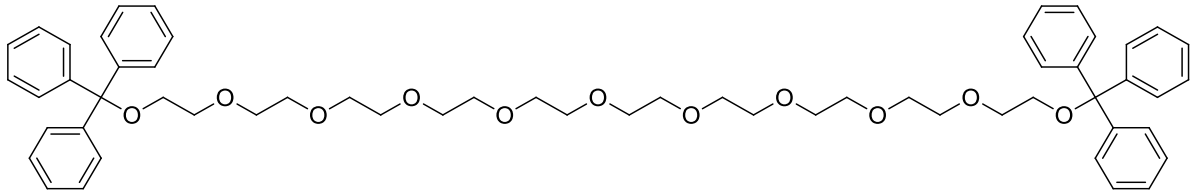
**1,17-Bis-(toluene-4-sulphonyloxy)-3,6,9,12,15,-pentaoxa-heptadecane (3)**[91]



Hexaethylene glycol (3 g; 10 mmol) was dissolved in pyridine (8 ml, 99 mmol) and cooled to -10°C. p-TsCl (5.91 g; 31 mmol) was dissolved in pyridine (8 ml, 99 mmol). These two pyridine solutions where mixed and stirred overnight. When reaction was finished as judged by TLC analysis, the solution was poured in to 50 ml ice water. This was then extracted with 2 x 60 ml $CH_2Cl_2$. The combined organic layers were washed with 2 M HCl (4 x 25 ml), saturated $NH_4Cl$ solution (4 x 25 ml) and dried with anhydrous $MgSO_4$. The product **(3)** (4.0 g; 64 %) was purified via flash chromatography, using EtOAc as eluent.

$^1$H NMR (400 MHz, $CHCl_3$-*d*) $\delta$ ppm 2.44 (6H, s, 2 $CH_3$), 3.58 (8H, s, 4 $CH_2$), 3.61 (4H, s, 2 $CH_2$), 3.68 (4H, t, *J*=5.0 Hz, 2 $CH_2$), 4.15 (4H, t, *J*=5.0 Hz, 2 $CH_2$), 7.34 (4H, d, *J*=8.0 Hz, 4 ArCH), 7.79 (4H, d, *J*=8.5 Hz, 4 ArCH); $^{13}$C NMR (400 MHz, $CHCl_3$-*d*) $\delta$ ppm 21.7 (2 $CH_3$), 68.7 (2 $CH_2$), 69.7 (2 $CH_2$), 70.5 (2 $CH_2$), 70.6 (2 $CH_2$), 70.6 (2 $CH_2$), 70.7 (2 $CH_2$), 128.0 (4 ArCH), 129.8 (4 ArCH), 133.0 (2 $ArCCH_3$), 144.8 ($ArCSO_2$). MS (ESI+) calculated (M) 590.18; found: ($M+NH_4^+$) 608.23, (M+ Na) 613.16, (M+K) 629.18.

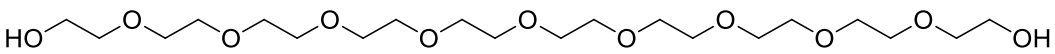
**1,1,1,33,33,33-Hexaphenyl-2,5,8,11,14,17,20,23,26,29,32-undecaoxa-tritriacontane (5)**[91]



The following three solution were prepared: NaH (1.4 g; 58.4 mmol) in 10 ml THF, **(2)** (2.7 g; 7.7 mmol) in 10 ml THF, and **(3)** (1 g; 1.7 mmol) in 10 ml THF. The first two solutions were mixed together at 0°C and stirred for 30 min, after which the third solution was added. This mixture was then stirred overnight at room temperature. Methanol (10 ml) was then added to quench the NaH. The solvent was removed, and the residue was redissolved in 100 ml water. The aqueous phase was 3 times extracted with EtOAc and the organic phase was concentrated *in vacuo*. Product **(5)** (1.0 g; 60%) was purified using flash chromatography, using EtOAc as eluent.

$^{1}$H NMR (400 MHz, $CHCl_3$-*d*) δ ppm 3.24 (4H, t, *J*=5.0 Hz, 2 C$\underline{H}_2$OTr), 3.63-3.69 (36H, m, 18 $CH_2$), 7.21-7.32 (18H, m, 12 *m*-ArCH, 6 *p*-ArCH), 7.47 (12H, d, *J*=7.5 Hz, *o*-ArCH); $^{13}$C NMR (400 MHz, $CHCl_3$-*d*) δ ppm 14.2 ($CH_2$OTr), 21.1 ($CH_2CH_2$OTr), 60.4 ($CH_2$), 70.6 ($CH_2$), 70.6 ($CH_2$), 70.6 ($CH_2$), 70.7 ($CH_2$), 70.7 ($CH_2$), 70.8 ($CH_2$), 86.5 (O$\underline{C}$Ph$_3$), 126.9 (*p*-ArCH), 127.8 (ArCH), 128.7 (ArCH), 144.1 (ArC). MS (ESI+) calculated (M) 973.17; found: (M+Na) 996.15.

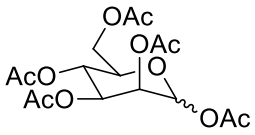
**3,6,9,12,15,18,21,24,27-Nonaoxanonacosane-1,29-diol (6)**[91]



**(5)** (544 mg; 0.56 mmol) was dissolved in 3ml TFA/TIS/$H_2O$ (95/2.5/2.5). After one hour stirring the liquid was evaporated, and the product **(6)** (207 mg; 81%) was purified using flash chromatography, using $CH_2Cl_2$ / MeOH (90:10) as eluent.

$^1$H NMR (400 MHz, $CHCl_3$-*d*) δ ppm 2.29 (2H, s, 2 x OH), 3.59-3.75 (40H, m, 10 x $OCH_2CH_2$); $^{13}$C NMR (400 MHz, $CHCl_3$-*d*) δ ppm 61.7 (2 x $CH_2OH$), 70.6 (16 x $CH_2$), 72.6 (2 x $\underline{C}H_2CH_2OH$) MS (ESI+) calculated (M) 458.27; found: (M+Na) 481.24, (M+K) 497.24.

**Penta-acetyl-D-mannose (8)**[92]

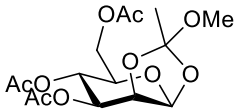


Mannose (1000 mg; 5.6 mmol) and NaOAc (457 mg; 5.6 mmol) were dissolved in acetic anhydride. The reaction mixture was then heated under refluxed for 2 hours. The mixture was then cooled to room temperature and the acetic anhydride was removed by co-evaporation with toluene. Ice was added to the residue, which was left to stir overnight at 0°C. This mixture was neutralized with $NaHCO_3$ and extracted 3 times with $CH_2Cl_2$, to give the product **(8)** (1704 mg; 78%). Apart form concentrating *in vacuo*, no further purification was needed.

$^1$H NMR (400 MHz, $CHCl_3$-*d*) δ ppm 2.01 (6H, s, $CH_3\alpha$), 2.01 (6H, s, $CH_3\beta$), 2.06 (6H, s, $CH_3\alpha$), 2.09 (6H, s, $CH_3\beta$), 2.10 (6H, s, $CH_3\alpha$), 2.10 (6H, s, $CH_3\beta$), 2.17 (6H, s, $CH_3\alpha$), 2.18 (6H, s, $CH_3\alpha$), 2.22 (6H, s, $CH_3\beta$), 3.81 (1H, ddd, *J* =9.5, 6.0, 2.0 Hz, H-5β), 4.03-4.06 (2H, m, H-5α), 4.10 (2H, dd, *J*=12.5, 2.5 Hz, H-6α),

4.14 (1H, dd, *J*=13.0, 2.4, H-6β), 4.29 (2H, dd, *J*=12.5, 5.0 Hz, H-6'α), 4.31 (1H, dd, *J*=12.5, 5.5 Hz, H-6'β), 5.13 (1H, dd, *J*=10.0, 3.3 Hz, H-3β), 5.26 (2H, apt t, *J*=2.0 Hz, H-2α), 5.29-5.36 (5H, m, H-3α, H-4α, H-4β), 5.49 (1H, dd, *J*=3.0, 1.0 Hz, H-2β), 5.86 (1H, d, *J*=1.0 Hz, H-1β), 6.09 (1H, d, *J*=2.0 Hz, H-1α). MS (ESI+) calculated (M) 390.09; found: (M+Na) 413.09, (2M+Na) 803.19.

**3,4,6-Tri-*O*-acetyl-β-D-mannose 1,2-(methyl orthoacetate) (9)**[93]

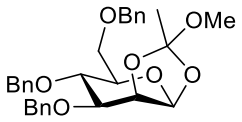



To a solution of penta-acetyl mannose (25.5 g, 65.33 mmol) in dry $CH_2Cl_2$ (10 ml) 33% HBr in acetic acid (68.72 g, 849.3 mmol) was added at room temperature; the mixture was stirred for 2 hours. 100 ml ice water was added and the organic layer separated and then back extracted with $CH_2Cl_2$ (2 x 100 ml). The combined organic layers were then washed with saturated $NaHCO_3$, brine, dried over $Na_2SO_4$ and concentrated to yield a residue. Without purification, the crude bromo derivative (27.0 g, 65.66 mmol) was solubilised in dry MeOH and dry $CHCl_3$ (200 ml, 50:50) followed by 2,6-lutidine (27.158 g, 253.46 mmol) and stirred at room temperature overnight. Solvents were removed using rotary evaporator, then co-evaporated with toluene (2 x 100 ml). The residue was then redissolved in $CHCl_3$ and washed with $H_2O$, brine, dried over $Na_2SO_4$, filtered and concentrated. The residue was purified by flash chromatography using EtOAc / petrol ether (20:80) as eluent, yielded the product **(9)** as a white solid (21.29 g, 65%) ($R_f$ = 0.45; EtOAc / petrol ether (50:50)).

$^1$H NMR (400 MHz, $CHCl_3$-*d*) δ ppm 5.50 (d, *J* = 2.5 Hz, 1H, H-1), 5.31 (t, *J* = 10.0 Hz, 1H, H-4), 5.15 (dd, *J* = 10.0, 4.00 Hz, 1H, H-3), 4.62 (dd, *J* = 4.0, 2.5 Hz,

1H, H-2), 4.24 (dd, *J* = 12.0, 5.0 Hz, 1H, H-6a), 4.15 (dd, *J* = 12.0, 2.5 Hz, 1H, H-6b), 3.69 (ddd, *J* = 9.5, 5.0, 2.5 Hz, 1H, H-5), 3.28 (s, 3H, OMe), 2.13 (s, 3H, $COCH_3$), 2.08 $COCH_3$), 2.06 $COCH_3$), 1.75 (s, 3H, $CH_3$); $^{13}C$ NMR (400 MHz, $CDCl_3$) δ ppm 170.6 (*C*$OCH_3$), 170.3 (*C*$OCH_3$), 169.4 (*C*$OCH_3$), 124.5 (C(OMe)$CH_3$), 97.3 (C-1), 76.5 (C-2), 71.2 (C-5), 70.6 (C-3), 65.4 (C-4), 62.2 (C-6), 49.9 (OMe), 24.3 ($CH_3$), 20.7 (CO*C*$H_3$), 20.7 (CO*C*$H_3$), 20.6 (CO*C*$H_3$). MS (ESI+) calculated (M) 362.12; found: (M+Na) 385.08.

**3,4,6-Tri-*O*-benzyl-β-D-mannose 1,2-(methyl orthoacetate) (10)**[94]

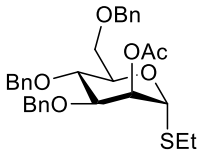



To a solution of **(9)** (5 g, 13.75 mmol) in 50 ml of dry THF followed by the addition of benzyl chloride (29.7 g, 243 mmol) and heated to 80 °C. After 15 min, the reaction mixture was cooled, and powdered KOH (11.6 g, 207 mmol) added and heating continued overnight. 50 ml $H_2O$ was added, and the reaction mixture was diluted with $CH_2Cl_2$. The organic layer separated and washed with saturated $NaHCO_3$, brine, dried over $Na_2SO_4$, filtered and concentrated. The residue was purified by flash chromatography, using EtOAc / petrol ether (20:80) as eluent, to yield the product **(10)** as a white solid (72%), ($R_f$ = 0.50;EtOAc / petrol ether (30:70)).

$^1H$ NMR (400 MHz, $CHCl_3$-*d*) δ ppm 7.42 – 7.24 (m, 15H, Ph), 5.36 (d, *J* = 2.5 Hz, 1H, H- 1), 4.91 (d, *J* = 10.0 Hz, 1H, CH*H*Ph), 4.80 (AB Quartet, *J* = 12.0 Hz, 2H, $CH_2Ph$), 4.65 -4.53 (m, 3H, $CH_2Ph$; C*H*HPh), 4.41 (dd, *J* = 4.0, 2.5 Hz, 1H, H-2), 3.93 (t, *J* = 9.5 Hz, 1H, H-4), 3.80 - 3.68 (m, 3H, H-3, H-6a,b), 3.43 (ddd, *J* = 9.5, 4.5, 2.5 Hz, 1H, H-5), 3.30 (s, 1H, OMe), 1.75 (s, 3H, $CH_3$); $^{13}C$ NMR (400

MHz, $CDCl_3$) δ ppm 138.2, 137.8, 128.5, 128.4, 128.3, 128.0, 128.0, 127.7, 127.5 (CH, C-Ph), 123.9 (*C*(OMe)$CH_3$), 97.5 (C-1), 79.0 (C-2), 77.1 (C-3), 75.2 ($CH_2$Ph), 74.1 (C-5), 74.1 (C-4), 73.3 ($CH_2$Ph), 72.3 ($CH_2$Ph), 68.9 (C-6), 49.7 (OMe), 24.4 ($CH_3$). MS (ESI+) calculated (M) 506.23; found: (M+Na) 529.18.

**S-Ethyl 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-1-thio-α-D-mannopyranoside (11)**[95]

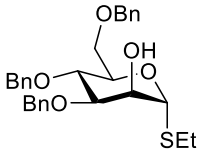



To a solution of **(10)** (10 g, 29.6 mmol), freshly activated 4 A molecular sieves in dry $CH_3CN$ (100 ml) was added ethane thiol (9.2. g, 148.14 mmol) and the resultant mixture was stirred for one hour. Then $Hg(Br)_2$ (2.16 g, 6 mmol) and heated to 60 °C for overnight stirring. Then the reaction mixture was filtered over the pad of Celite®, washed with EtOAc and the filtrate concentrated. The residue purified by flash silica gel column chromatography, using EtOAc / petrol ether (10:90) as eluent, yielded the product **(11)** as a white solid (11.3 g, 72%), ($R_f$ = 0.60; EtOAc / petrol ether (30:70)).

$^1$H NMR (400 MHz, $CHCl_3$-*d*) δ ppm 7.53 - 7.09 (m, 15H, Ph), 5.45 (dd, *J* = 2.5, 1.5 Hz, 1H, H-2), 5.34 (d, *J* = 1.5 Hz, 1H, H-1), 4.87 (d, *J* = 11.0 Hz, 1H, C*H*HPh), 4.70 ($AB_q$, *J* = 11.5 Hz, 2H, CH*2*Ph), 4.57 – 4.47 (m, 3H, $CH_2$Ph, CH*H*Ph), 4.18 – 4.21 (m, 1H, H-5), 3.98 – 3.90 (m, 2H, H-3, H-4), 3.85 (dd, *J* = 11.0, 4.0 Hz, 1H, H-6a), 3.70 (dd, *J* = 11.0, 2.0 Hz, 1H, H-6b), 2.75 - 2.53 (m, 2H, S-$CH_2$), 2.18 (s, 1H, $COCH_3$), 1.29 (t, *J* = 7.5 Hz, 3H, $CH_3$); $^{13}$C NMR (400 MHz, $CDCl_3$) δ ppm 170.4 (*C*$OCH_3$), 138.3, 138.1, 137.6, 128.4, 128.3, 128.1, 127.8, 127.7, 127.6, 127.6 (CH, C-Ph), 82.4 (C-1), 78.5 (C-4), 75.1 (C-3), 74.5 ($CH_2$Ph), 73.3, (C-5),

71.8 ($CH_2Ph$), 71.7 ($CH_2Ph$), 70.5 (C-2), 68.7 (C-6), 25.4 (S-$CH_2$), 21.1 ($COCH_3$), 14.8 ($CH_3$). MS (ESI+) calculated (M) 536.29; found: (M+Na) 559.17.

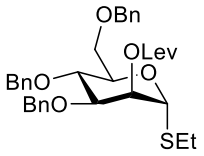
**S-Ethyl 3,4,6-tri-*O*-benzyl-1-thio-α-D-mannopyranoside (12)**[96]




Compound **(11)** (11 g, 20.5 mmol) was dissolved in 100 ml dry MeOH and then NaOMe (1.1 g, 20.5 mmol) was added. The solution was stirred at room temperature overnight before it was neutralized with Dowex 50wx8 (H+), filtered, concentrated to yield the crude, which was purified by flash chromatography, using EtOAc / petrol ether (10:90) as eluent, to yield the product **(12)** as a white solid (95 %), ($R_f$ = 0.45 EtOAc / petrol ether (30:70)).

$^1H$ NMR (400 MHz, $CHCl_3$-*d*) δ ppm 7.19 – 7.37 (m, 15H, Ph), 5.41 (d, *J* = 1.5 Hz, H-1), 4.84 (d, *J* = 11.0 Hz, C*H*HPh), 4.66 – 4.69 (m, 3H, $CH_2Ph$, CH*H*Ph), 4.51 – 4.53 (m, 2H, $CH_2Ph$), 4.18 (t, *J* = 9.5 Hz, 1H, H-5), 4.11 (dd, *J* = 1.0 Hz, H-2), 3.92 (t, *J* = 9.0 Hz, 1H, H-4), 3.85 (dd, *J* = 3.0, 9.0 Hz, 1H, H-3), 3.81 (dd, *J* = 4.5 Hz, 1H, H-6a), 3.70 (dd, *J* = 2.0, 11.0 Hz, 1H, H-6b), 2.54 – 2.72 (m, 2H, S-$CH_2$), 1.30 (t, *J* = 7.5 Hz, 3H, $CH_3$); HRMS (ESI+) calculated for: $C_{29}H_{34}O_5S$ (M + Na) 517.2019; found: 517.2019.

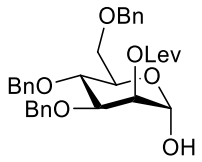
## S-Ethyl 3,4,6-tri-*O*-benzyl-2-*O*-levulinate-1-thio-α-D-mannopyranoside (13)




To a solution of **(12)** (10 g, 18.6 mmol) in dry $CH_2Cl_2$ (150 ml) were added levulinic acid (2.27 g, 19.53 mmol) and DMAP (2.499 g, 20.46 mmol). The solution was cooled to 0 °C. After 10 min DIC (2.33 g, 18.46 mmol) was added and stirred overnight while the temperature increased to room temperature. The reaction mixture was then concentrated and purified by flash chromatography, using EtOAc / petrol ether (25:75) as eluent, to yield the product **(13)** as a white solid (0.479 g, 87 %), ($R_f$ = 0.35; EtOAc / petrol ether (30:70)).

$^1$H NMR (400 MHz, $CHCl_3$-*d*) δ ppm 7.37 – 7.17 (m, 15H, Ph), 5.44 (s, 1H, H-2), 5.31 (s, 1H, H-1), 32 4.86 (d, *J* = 10.5 Hz, 1H, C*H*HPh), 4.73 - 4.63 (m, 2H, $CH_2$Ph), 4.55 - 4.46 (m, 3H, $CH_2$Ph, CH*H*Ph), 4.16 (s, 1H, H-5), 3.91 (d, *J* = 3.5 Hz, 2H, H-3, H-4), 3.85 – 3.83 (m, 1H, H-6a), 3.74 - 3.64 (m, 1H, H-6b), 2.83 - 2.53 (m, 6H, S-$CH_2$, 2 x $CH_2$ of Lev), 2.15 (s, 1H, $CH_3$ of Lev), 1.28 (t, *J* = 7.5 Hz, 3H, $CH_3$); $^{13}$C NMR (400 MHz, $CDCl_3$) δ ppm 206.3 (CO of Lev), 172.0 (COO of Lev), 138.3, 138.1, 137.7, 128.3, 128.3, 128.1, 127.8, 127.7, 127.6, 127.6 (CH, C- Ph), 82.3 (C-1), 78.5 (C-4) , 75.1 (C-3), 74.4 ($CH_2$Ph), 73.3 (C-5), 71.7 ($CH_2$Ph), 71.6 ($CH_2$Ph), 70.6 (C-2), 68.8 (C-6), 37.9 (OCO-$CH_2$ of Lev), 29.8 ($CH_2$ of Lev), 28.2 (S-$CH_2$), 25.4 ($CH_3$ of Lev), 14.9 ($CH_3$); $[\alpha]_D$ = 68.8; MS (ESI+) calculated (M) 592.24; found: (M+Na) 615.23.

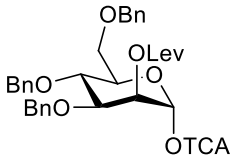
### 3,4,6-Tri-O-Benzyl-2-O-levulinoyl-α,β-D-mannopyranose (14)




To a stirred solution of **(13)** (5 g, 8.5 mmol) in $CH_3CN$ (50 ml) and $H_2O$ (5 ml) was added NIS (2.85 g, 12.6 mmol). The reaction mixture was stirred for 30 min at room temperature, then diluted with $CH_2Cl_2$. The solution was washed with saturated $Na_2S_2O_3$ and brine, and the combined organic layer was dried with anhydrous $MgSO_4$ and concentrated *in vacuo*. The residue was purified by flash column chromatography, using petrol ether / EtOAc (50:50) as eluent, to produce **(14)** (4.4g; 95%) as colourless syrup, ($R_f$ = 0.20; petrol ether / EtOAc (50:50)).

$^1$H NMR (400 MHz, $CHCl_3$-*d*, α-anomer as the major product) $\delta$ ppm 7.38-7.12 (m, 15 H), 5.35 (dd, *J* = 3.5, 1.5 Hz, 1 H), 5.17 (dd, *J* = 1.0 Hz, 1 H), 4.85 (d, *J* = 11.0 Hz, 1 H), 4.67 (d, *J* = 11.0 Hz, 1 H), 4.58 (d, *J* = 12.0 Hz, 1 H), 4.51-4.45 (m, 3 H), 4.08-4.02 (m, 1 H), 4.01 (dd, *J* = 9.5, 3.5 Hz, 1 H), 3.86 (d, *J* = 3.0 Hz, 1H), 3.76-3.64 (m, 3 H), 2.75-2.64 (m, 4 H), 2.10 (s, 3 H). $^{13}$C NMR (400 MHz, $CDCl_3$) 206.9, 172.3, 138.5, 138.2, 138.1, 128.7, 128.6, 128.5, 128.4, 128.3, 128.2, 128.1, 128.0, 127.9, 127.8, 92.6, 77.9, 75.3, 74.8, 73.6, 71.8, 71.2, 69.5, 69.4, 38.3, 30.0, 28.4. HRMS (m/e): calculated for $C_{32}H_{36}NaO_8$ (M + Na+) 571.2308, found 571.2303.

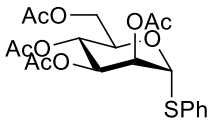
## 3,4,6-Tri-O-benzyl-2-O-levulinoyl-α-D-mannopyranosyl trichloroimidate (15)




Activated molecular sieves were added to a solution of **(14)** (4.66 g, 8.5 mmol) and $CCl_3$-CN (12.27 g, 85 mmol) in 60 ml $CH_2Cl_2$. The solution was stirred for two hours at 0°C, followed by the addition of DBU (260 mg, 1.7 mmol). The reaction was completed as judged by TLC analysis after 90 min stirring at 0°C. The reaction mixture was filters through Celite® and was concentrated *in vacuo*. The residue was purified by a Biotage automated system starting from a 12% EtOAc in petrol ether to produce **(15)** (3.8 g; 76%) ($R_f$ = 0.45; petrol ether / EtOAc (50:50))

$^1$H NMR (400 MHz, $CHCl_3$-*d*) δ ppm 8.73 (s, 1 H), 7.38-7.18 (m, 15 H), 6.28 (d, *J* = 2.0 Hz, 1 H), 5.48 (t, *J* = 2.0 Hz, 1 H), 4.87 (d, *J* = 11.0 Hz, 1 H), 4.71 (d, *J* = 11.0 Hz, 1 H), 4.66 (d, *J* = 12.0 Hz, 1 H), 4.56 (d, *J* = 11.0 Hz, 1 H), 4.54 (d, *J* = 11.0 Hz, 1 H), 4.51 (d, *J* = 12.0 Hz, 1 H), 4.03-3.94 (m, 3 H), 3.83 (dd, *J* = 11.0, 3.5 Hz, 1 H), 3.71 (dd, *J* = 11.5, 1.5 Hz, 1 H), 2.80-2.67 (m, 4 H), 2.15 (s, 3 H). $^{13}$C NMR (400 MHz, $CHCl_3$-*d*) δ ppm 206.5, 172.1, 160.2, 138.4, 138.3, 137.8, 128.7, 128.6, 128.6, 128.5, 128.5, 128.4, 128.3, 128.12, 128.06, 127.9, 95.5, 75.7, 74.5, 73.8, 73.7, 72.1, 68.6, 67.7, 53.7, 38.2, 30.0, 28.3. MS (ESI+) calculated (M) 592.23; found: (M+Na) 615.21, (M+K) 631.18.

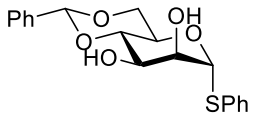
**S-Phenyl 2,3,4,6-Tetra-*O*-acetyl-1-thio-α-D-mannopyranoside (16)**[97]




To a solution of **(8)** (5 g, 12.8 mmol) in 100 ml dry $CH_2Cl_2$ phenyl thiol (2.82 g, 2.63 ml, 25.6 mmol) was added, followed by the addition of $BF_3{\cdot}OEt_2$ (4.54 g, 3.95 ml, 25.6 mmol). The mixture was stirred for 20 hours at room temperature before the reaction was completed as judged by TLC analysis. Triethylamine was added until the pH was neutral. the solution was then diluted with 100 ml $CH_2Cl_2$ and washed with saturated $NaHCO_3$ solution, brine and water. The organic layer was dried with anhydrous $MgSO_4$ and concentrated *in vacuo*. The residue was purified by a Biotage automated system, starting with a 12 % gradient of EtOAc in petrol ether to produce **(16)** (4.0 g; 71%) ($R_f$ = 0.50 petrol ether / EtOAc (50:50)).

$^1$H NMR (400 MHz, $CHCl_3$-*d*) δ ppm 7.50–5.30 (m, 5 H, ArH), 5.50 (dd,$J_{1,2}$ = 1.5 Hz, $J_{2,3}$ = 3.0 Hz, 1 H, H-2), 5.50 (d, 1 H, H-1), 5.33 (dd, $J_{4,5}$ = 10.0 Hz, $J_{3,4}$ = 10.0 Hz, 1 H, H-4), 5.32 (dd, 1 H, H-3), 4.55 (ddd, *J*5,6a = 2.0 Hz, $J_{5,6b}$ = 6.0 Hz, 1 H, H-5), 4.31 (dd, $J_{6a,6b}$ = 12.5 Hz, 1 H, H-6b), 4.11 (dd, 1 H, H-6a), 2.16, 2.08, 2.06, 2.02 (s, 12 H, 4 × OCCH3). $^{13}$C NMR (400 MHz, $CHCl_3$-*d*) δ ppm 170.6, 169.9, 169.8 (O*C*$CH_3$), 132.1–129.2 (Ar), 85.7 (C-1), 70.9 (C-2), 69.5 (C-5), 69.4 (C-3), 66.4 (C-4), 62.4 (C-6), 20.9, 20.7 (OC*C*$H_3$). MS (ESI+) calculated (M) 440.46; found: (M+Na) 463.46.

**S-Phenyl 4,6-O-benzylidene-1-thio-α-D-mannopyranoside (18)**[98]




To a solution of **(16)** (3.9 g, 8.86 mmol) in 75 ml dry methanol, NaOMe (270 mg, 4.44 mmol) was added and stirred for 1 hour at room temperature. A white precipitate was formed, which disappeared after addition of 50 ml MeOH. Dowex 50WX8 resin was added to the mixture until the pH was acidic. The resin was removed by filtration and the solution was concentrated *in vacuo* to yield a white solid. No purification was needed at this stage, apart from drying 3 times by co-evaporation with toluene, before the second step was performed.

The white solid obtained in the first step was dissolved in 100 ml dry acetonitrile. $PhCH(OMe)_2$ (1.7 g, 11.2 mmol) was added, followed by the addition of CSA (200 mg, 1 mmol). The mixture was stirred overnight and after completion as judged by TLS analysis, 250 µl triethylamine was added. The solution was concentrated *in vacuo* and the compound was purified by a Biotage automated system to produce **(18)** only in a 16 % yield, ($R_f$ = 0.45 petrol ether / EtOAc (50:50)), the major compound was **(17)** in a 71 % yield, ($R_f$ = 0.65; petrol ether / EtOAc (50:50)).

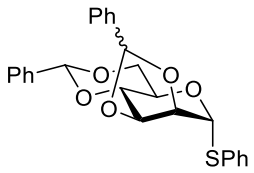
Conversion of **(17)** into **(18)**

A 10% w/w phosphomolybdic acid (PMA)/Silica was prepared by adding 4,5 g silica to 500 mg PMA solution in 25 ml MeOH. This suspension was stirred for 6 hours at room temperature and dried by evaporation of the MeOH (the silica turns from yellow to fluo-green during drying).

To a solution of **(17)** (2.0 g, 4.5 mmol) in 40 ml $CH_2Cl_2$ 500 mg of the 10% PMA/$SiO_2$ was added and stirred for two hours at room temperature. After the reaction is completed the silica was filtered off and the solution was concentrated *in vacuo*, to produce **(18)** (1.4g; 89 %).

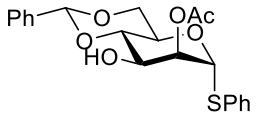
$^1$H NMR (400 MHz, $CD_3OD$) $\delta$ ppm 7.57–7.27 (m, 10H, arom.), 5.62 (s, 1H, benzylidene), 5.49 (d, 1H,$J_{1-2}$=1.5 Hz, H-1), 4.24 (m, 1H, H-5), 4.18 (d, 1H, $J_{2-3}$=3.5 Hz, H-2), 4.13 (dd, 1H, $J_{6b-6a}$=10.0 Hz and $J_{6b-5}$= 5.0 Hz, H-6b), 4.03 (t, 1H, $J_{4-3}$=$J_{4-5}$=10.0 Hz, H-4), 3.95 (dd, 1H, H-3), 3.84 (t, 1H, H-6a). MS (ESI+) calculated (M) 360.13; found: (M+Na) 383.13.

**S-Phenyl exo(phenyl)-2,3;4,6-di-O-benzylidene-1-thio-α-D-mannopyrano-side, S-Phenyl endo(phenyl)-2,3;4,6-di-O-benzylidene-1-thio-α-D-mannopyranoside (17)**[98]




$^1$H NMR (400 MHz, $CHCl_3$-*d*) $\delta$ ppm 7.60-7.28 (m, 15 H, aromatic), 6.31 (s, 1 H, PhCH, dioxolane), 5.87 (s, 1 H, PhCH, dioxane), 5.64 (s, 1 H, H-l), 4.69 (dd, 1 H, H-3), 4.41-3.73 (m 5 H, skeleton protons).

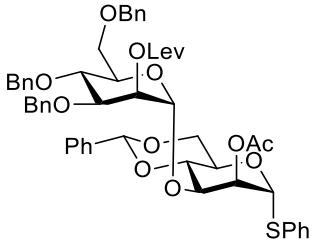
**S-Phenyl 2-*O*-acetyl-4,6-*O*-benzylidene-1-thio-α-D-mannopyranoside (19)**




To a solution of **(18)** (6.3 g, 17.5 mmol) in 200 ml acetonitrile, trimethylorthoacetate (3.14 g, 3.28 ml, 26.2 mmol) and para-toluene sulphonic acid (500 mg, 2.6 mmol) were added. The mixture was stirred for one hour at room temperature. Triethylamine was then added to neutralize, and 75 ml acetic acid was then added to the reaction mixture. After another hour stirring, the reaction mixture is concentrated and purified via flash chromatography, using EtOAc / petrol ether (30:70) as eluent, to produce **(19)** (5.4 g; 76 %).

$^{1}$H NMR (400 MHz, $CHCl_3$-*d*) $\delta$ ppm 7.56–7.30 (m, 10H, arom.), 5.66 (s, 1H, benzylidene), 5.52 (s, 1H,H-1), 5.51 (dd, 1H, $J_{2-1}$=1.5 Hz, $J_{2-3}$=3.3 Hz, H-2), 4.41 (ddd, 1H, $J_{5-4}$=9.5 Hz, $J_{5-6a}$=10.5 Hz, $J_{5-6b}$=5.0 Hz, H-5), 4.29 (dd, 1H, $J_{6b-6a}$=10.5 Hz, H-6b), 4.28-4.25 (m, 1H,H-3), 4.04 (t, 1H, $J_{4-3}$=9.5 Hz, H-4), 3.89 (t, 1H, H-6a), 2.65 (d, 1H, $J_{OH-3}$=3.5 Hz, OH), 2.21 (s, 3H, O– C=O–$CH_3$); $^{13}$C NMR(400 MHz): d 170.3 (–O–C=O–$CH_3$), 136.9, 133.0 (2 C arom.), 132.0–126.2 (10 CH arom.), 102.2 (benzylidene), 86.8 (C-1), 79.0 (C-4), 73.5 (C-2), 68.3 (C-6), 67.7 (C-3), 64.5 (C-5A), 20.9 (–O–C=O–CH3); MS (ESI+) calculated (M) 402.46; found: (M+H) 403.2.

**S-Phenyl (2-*O*-levulinoyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl)-(1→3)-2-*O*-acetyl-4,6-*O*-benzylidene-1-thio-α-D-mannopyranoside (20)**

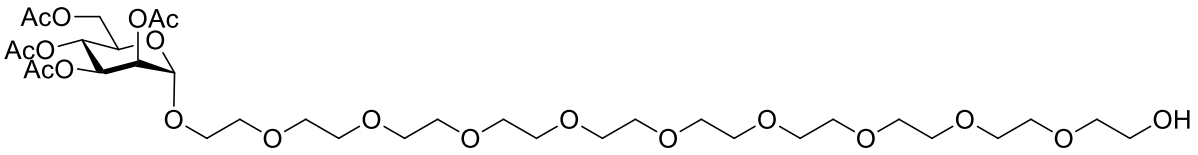



Compound **(15)** (4.0 g, 9.9 mmol) and **(19)** (2.6 g, 6.6 mmol) were dissolved in 120 ml dry $CH_2Cl_2$ with 3 g of freshly activated 4 Å molecular sieves and stirred for one hour. $BF_3{\cdot}OEt_2$ (1.41 g, 1.26 ml, 9.95 mmol) is added and the reaction is stirred overnight starting from 0°C and allowing to warm up to room temperature. Triethylamine is added until the reaction mixture is neutral before the solution is washed with saturated $NaHCO_3$ solution, brine and water. The organic layer is dried with anhydrous $MgSO_4$ and concentrated *in vacuo*. The compound is purified by a Biotage automated system starting with a gradient from 12 % EtOAc in petrol ether, to produce **(20)** (3.5 g, 67%) ($R_f$ = 0.25; petrol ether / EtOAc (50:50)).

$^1$H NMR (400 MHz, $CHCl_3$-*d*) δ ppm 8.15–7.25 (m, 25H, arom.), 5.69 (s, 1H, benzylidene), 5.55 (dd, 1H, $J_{2-1}$=2.0 Hz, $J_{2-3}$=2.5 Hz, H-2D), 5.52 (dd, 1H, $J_{2-1}$=1.5 Hz, $J_{2-3}$=3.5 Hz, H-2C), 5.49 (d, 1H, H-1C), 5.34 (d, 1H, H-1D), 4.89 (d, 1H, J=11.0 Hz, CHPh), 4.77 (d, 1H, J = 12.0 Hz, CHPh), 4.74 (d, 1H, J= 11.5 Hz, CHPh), 4.56 (d, 1H, J=12.0 Hz, CHPh), 4.56 (d, 1H, J = 11.5 Hz, CHPh), 4.54 (d, 1H, J = 11.0 Hz, CHPh), 4.42 (ddd, 1H, $J_{5-4}$=10.0 Hz, $J_{5-6b}$=5.0 Hz, $J_{5-6a}$= 10.5 Hz, H-5C), 4.40 (dd, 1H, $J_{3-4}$=10.0 Hz, H-3C), 4.30 (dd, 1H, $J_{6b-6a}$=10.5 Hz, H-6Cb), 4.20 (t, 1H, H-4C), 3.97– 3.87 (m, 4H, H-5D, H-4D, H-3D and H-6Ca), 3.86 (dd, 1H, $J_{6b-6a}$=12.0 Hz, $J_{6b-5}$=4.0 Hz, H-6Db), 3.78 (dd, 1H, $J_{6a-5}$=3.5 Hz, H-6Da), 2.21 and 2.16 (2 s, 6H, 2 O–C=O– CH3); $^{13}$C NMR (400 MHz, $CHCl_3$-*d*) δ ppm 170.1 and 169.7 (2 –O–C]O–CH3), 138.4, 138.2, 137.9, 136.9 and 133.0 (5 C

arom.), 132.0–125.9 (25 CH arom.), 101.3 (benzylidene), 98.8 (C-1D), 86.8 (C-1C), 78.9 (C-4C), 75.6, 74.1 and 72.1 (C-3D, C-4D and C-5D), 74.9 ($CH_2Ph$), 73.4 (CH2Ph), 73.1 (C-2C), 71.7 (CH2Ph), 70.8 (C-3C), 68.6 (C-6D), 68.4 (C-2D), 68.2 (C-6C), 64.6 (C-5C), 20.8 (2 times –O–C=O–$CH_3$). MS (ESI+) calculated (M) 796.90; found: (M+Na) 820.1.

**2,3,4,6 Tetra-*O*-acetyl-1-(3, 6, 9, 12, 15, 18, 21, 24, 27-nonaoxa-29-hydroxy-29yl)-α-D-mannopyranoside (33)**[94]

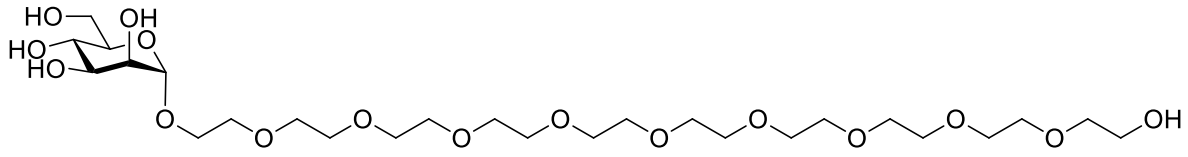



Boron trifluoride-diethyl etherate (4.40 mL, 35.0 mmol) was added dropwise to a stirring solution of decaethyleneglycol (3.51 g, 7.6 mmol) and **(8)** (2.00 g, 5.1 mmol) in dichloromethane (10 mL). The mixture was left to stir at room temperature. Over a period of six hours, further boron trifluoride-diethyl etherate (2.5 mL, 20.4 mmol) was added. After a further eighteen hours, TLC analysis (EtOAc) indicated complete consumption of starting material ($R_f$ = 0.60; EtOAc) to product ($R_f$ = 0.10; EtOAc). The reaction mixture was quenched with a saturated sodium bicarbonate solution (60 mL). The organic layer was washed with a saturated sodium bicarbonate solution (3 × 50 mL) and the aqueous layer was re-extracted with dichloromethane (50 mL). The combined organic layers were washed with brine (50 mL), dried with anhydrous $MgSO_4$, filtered, and concentrated *in vacuo*. The obtained residue was purified using flash column chromatography, using EtOAc / petrol ether (50:50) to EtOAc / MeOH (80:20), as eluent, to yield a light brown oil. However, only a small amount of pure compound

**(33)** could be isolated via flash column chromatography. the other fractions obtained in the flash column chromatography comprised large amounts of **(33)**. All the fractions were combined and used in the preparation of **(37)**.

$^{1}$H NMR (400 MHz, $CHCl_3$-*d*) δ ppm 2.01 (3H, s, $CH_3$), 2.05 (3H, s, $CH_3$), 2.11 (3H, s, $CH_3$), 2.16 (3H, s, $CH_3$), 3.63 - 3.71 (40H, m, 20 $CH_2$), 4.05 - 4.08 (1H, m, H-5), 4.10 - 4.14 (1H, m, H-6), 4.30 (1H, dd, *J*=12.5, 5.2 Hz, H-6'), 4.88 (1H, s, H-1), 5.27 (1H, m, H-2), 5.30 (1H, t, *J*=10.0 Hz, H-4), 5.36 (1H, dd, *J*=10.0, 3.5 Hz, H-3); $^{13}$C NMR (400 MHz, $CHCl_3$-*d*) δ ppm 20.7 ($CH_3$), 20.7 ($CH_3$), 20.8 ($CH_3$), 20.9 ($CH_3$), 60.4 ($CH_2$), 62.4 (C-6), 65.9 ($CH_2$), 66.2 (C-4), 67.4 ($CH_2$), 68.4 (C-5), 69.1 (C-3), 69.1 ($CH_2$), 69.6 (C-2), 70.0 ($CH_2$), 70.3 ($CH_2$), 70.6 (13 $CH_2$), 70.7 ($CH_2$), 97.7 (C-1), 169.7 (C=O), 169.8 (C=O), 170.0 (C=O), 170.7 (C=O); $[\alpha]_D$ = +21.7, (c. 1.3 in $CHCl_3$); HRMS *m/z* ($ES^+$): calculated for $C_{34}H_{61}O_{20}$ 788.3569, found: ($M+Na^+$) 811.3571.

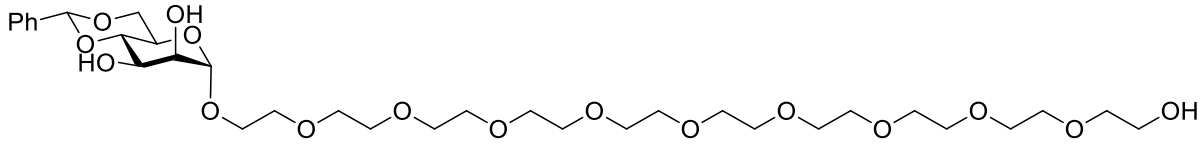
**1-(3, 6, 9, 12, 15, 18, 21, 24, 27-Nonaoxa-29-hydroxy-29yl)-α-D-mannopyranoside (37)**




Sodium methoxide (270 mg, 5 mmol) was added to a stirring solution of **(33), (34), (35)** (1.5 g, 2.0 mmol) in MeOH (10 mL). After thirty minutes, TLC analysis (MeOH / EtOAc, (20:80)) indicated complete consumption of starting material ($R_f$ = 0.40; MeOH / EtOAc, (20:80)) to product **(37)** ($R_f$ = 0.00; MeOH / EtOAc, (20:80)). Dowex 50WX8-200 ion exchange resin ($H^+$ form) was added until pH<7. The reaction mixture was filtered and concentrated *in vacuo*, to give a pale yellow oil (1.2 g; 100 %).

$^1$H NMR (400 MHz, $CHCl_3$-*d*) δ ppm 3.45 (2H, t, *J*=6.5 Hz, C*H*$_2$OH), 3.64 (36H, m, 18 x $CH_2$), 3.73-3.78 (3H, m, H-5, H-6, H-6'), 3.81 (2H, t, *J*=6.5 Hz, C*H*$_2$CH$_2$OH), 3.90-4.10 (3H, m, H-2, H-3, H-4), 4.85 (1H, s, H-1); $^{13}$C NMR (400 MHz, $CHCl_3$-*d*) δ ppm 30.4 ($CH_2$Br), 61.3 (C-6), 66.6 (C-4), 66.8 (*C*H$_2$CH$_2$OH), 70.1 ($CH_2$), 70.5 (13 x $CH_2$), 70.5 ($CH_2$), 70.6 ($CH_2$), 70.6 ($CH_2$), 71.3 ($CH_2$), 70.8 (C-5), 71.6 (C-3), 72.2 (C-2), 100.1 (C-1); $[\alpha]_D$ = +25.7; MS (ESI+) calculated (M) 620.68; found: (M+Na) 620.1.

**(3', 6', 9', 12', 15', 18', 21', 24', 27'-Nonaoxa-29'-hydroxy-29yl) 4,6-O-benzylidene-α-D-mannopyranoside (38')**

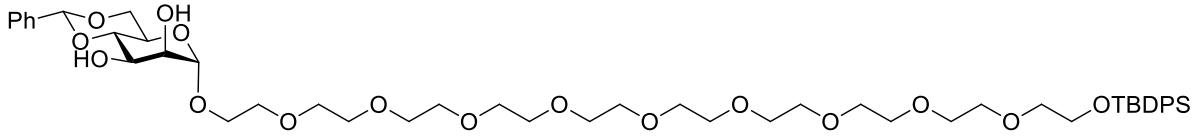



Benzyldimethyl acetal (797 mg ,3.9 mmol) was added to a solution of **(37)** (243 mg, 3.9 mmol) in 75 ml acetonitrile. Camphorsulphonic acid (455 mg; 1.95 mmol) was added to a stirring solution. After a couple of minutes, the solution turned clear. The reaction was left to stir overnight and was stopped by addition of triethylamine (197 mg; 1.95 mmol). Solvents were removed by rotary evaporation. The residue **(38')** ($R_f$ = 0.75; MeOH / $CH_2Cl_2$ (20:80)) was purified using flash chromatography, using MeOH / $CH_2Cl_2$ (3:97) as eluent. Product **(38')** (174 mg; 63%) was obtained as a yellow oil. Starting material was recovered and the reaction was repeated.

$^1$H NMR (400 MHz, $CHCl_3$-*d*) δ ppm 3.18 (2H, s, -OH), 3.50-3.65 (38H, m, H-6, H-6', 18 x $CH_2$), 3.73-3.84 (6H, m, H-3, H-5, 2 x $CH_2$), 3.95 (2H, broad s, H-2, CH*H*), 4.17 (1H, t, J = 6.0, H-4), 4.81 (1H, d, J = 0.5 Hz, H-1), 5.48 (1H,s,H-7), 7.26-

7.28 (3H, m, arom), 7.40-7.43 (2H, m, arom); $^{13}$C NMR (400 MHz, $CHCl_3$-*d*) δ ppm 61.5 (C-6), 63.3 (C-5), 66.7 (C-3), 68.3 (C-2), 68.8 (C-4), 70.1-72.5 (*C*$H_2$), 79.0 (*C*$H_2$), 100.7 (C-1), 102.0 (C-7), 126.3 (2 x *C*H), 128.1 (2 x *C*H), 129.0 (*C*H), 137.5 (C), IR (neat) ν $cm^{-1}$ 3410 (OH), 2873 (C-H) $[\alpha]_D$ = +21.3; HRMS *m/z* ($ES^+$): calculated for $C_{33}H_{56}O_{16:}$ 708.3568, found: ($M+Na^+$) 731.33.

**(29'-O-*Tert*-butyldiphenylsilane-3', 6', 9', 12', 15', 18', 21', 24', 27'-29'-decoxa-29yl) 4,6-O-benzylidene-α-D-mannopyranoside (40)**

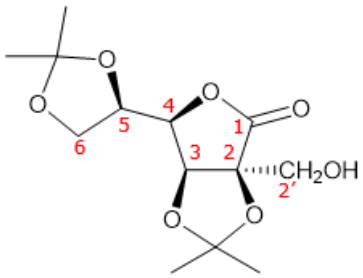



To a solution of **(38')** (463 mg, 0.65 mmol) in *tert*-butyl(chloro)diphenylsilane (255 µl, 0.981 mmol), 4-(dimethylamino)pyridine (0.8 mg, 0.0065 mmol) was added. The mixture was stirred overnight at room temperature before the solvent was removed by rotary evaporation. The residue was purified using flash chromatography using MeOH / $CH_2Cl_2$ (3:97), to obtain **(40)** (179 mg; 58.3%) ($R_f$ = 0.10; MeOH / $CH_2Cl_2$ (2:98)).

$^1$H NMR (400 MHz, $CHCl_3$-*d*) δ ppm 1.05 (9H,s, 3 x $CH_3$), 3.02 (2H, s, -OH), 3.57-3.69 (38H, m, H-6, H-6', 18 x $CH_2$), 3.78-3.94 (6H, m, H-3, H-5, 2 x $CH_2$), 4.06 (2H, m, H-2, CH*H*), 4.25 (1H, t, J = 5.5 Hz, H-4), 4.93 (1H, d, J = 0.5 Hz, H-1), 5.56 (1H,s,H-7), 7.34-7.43 (11H, m arom), 7.66-7.72 (4H, m, arom); $^{13}$C NMR (400 MHz, $CHCl_3$-*d*) δ ppm 19.2 (*C*-$CH_3$), 26.8 (3 X *C*$H_3$), 63.2 (C-6), 63.4 (C-5), 66.8 (C-3), 68.4 (C-2), 68.8 (C-4), 70.1-72.5 (*C*$H_2$), 79.0 (*C*$H_2$), 100.7 (C-1),

102.0 (C-7), 126.3 (*C*H), 127.6 (*C*H), 128.2 (*C*H), 129.0 (*C*H), 129.1 (*C*H), 129.6 (*C*H), 129.7 (*C*H), 133.7 (*C*H), 134.5 (*C*), 135.6 (*C*H) 137.4 (C) IR (neat) ν $cm^{-1}$ 3414 (OH), 2862 (C-H) $[\alpha]_D$ = +17.28, HRMS *m/z* ($ES^+$) calculated for $C_{49}H_{74}O_{16}Si$: 946.4746, found ($M+Na^+$) 969.48.

**2,3:5,6-Di-*O*-isopropylidene-D-mannono-1,4-lactone (43)**[88]



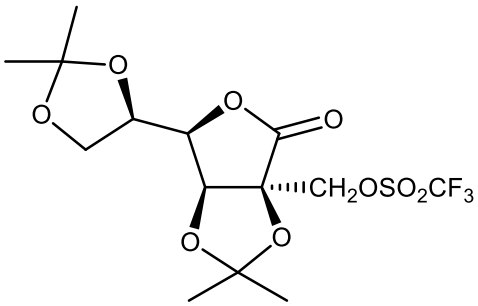

Sodium cyanide (1.06 g; 21.6 mmol) was added to a solution of D-fructose (3.00 g; 16.7 mmol) in 30 ml water. After 24 hours stirring at room temperature, the reaction mixture was refluxed for 12 hours to hydrolyse the cyanide. The mixture was cooled down and passed through an Amberlite IR-120$H^+$ column, before solvent was removed via lyophilisation.

In a second step, dry acetone 50 ml and concentrated sulphuric acid 0.75ml where added to the reaction product. Anhydrous copper sulphate was added as a drying agent. This mixture was stirred overnight and when reaction was completed as judged by TLC analysis, the mixture was neutralised with solid sodium carbonate. The solvent was then removed, and the residue was redissolved in 30 ml water. The aqueous layer then was extracted 3 times with $CH_2Cl_2$. The combined organic layers were concentrated by rotary evaporation. The residue was purified via flash

chromatography using EtOAc / petrol ether (50:50) as eluent, to yield **(43)** (1.2 g; 25%).

$^1$H NMR (400 MHz, $d_6$-acetone) $\delta$ ppm: 1.35, 1.41, 1.42 (4 x s, 12H, 4 x $-CH_3$), 3.89–4.03 (m, 3H, H-6, H-2'a, H-2'b), 4.15 (dd, 1H, $J_{6,6_0}$ = 9.0 Hz, $J_{6_0,5}$ = 6.0 Hz, H-$6_0$), 4.41–4.46 (m, 1H, $H_5$), 4.50 (dd, 1H, $J_{3,4}$ = 3.5 Hz, $J_{4,5}$ = 7.0 Hz, H-4), 4.69 (dd, 1H, $J_{2'a,\,OH}$ = 5.0 Hz, $J_{2'b,\,OH}$ = 6.0 Hz, –OH), 4.92 (d, 1H, $J_{3,4}$ = 3.5 Hz, H-3); $^{13}$C NMR (400 MHz, $CHCl_3$-*d*) $\delta$ ppm 26.99, 26.16, 26.76 (4 x $-CH_3$), 61.43, 66.24 (2 x $-CH_2-$), 72.42, 78.45, 78.70 (3 x –CH–), 86.00, 109.73, 113.79 (3 x –C–), 175.69 (–C=O). MS (ESI-) calculated (M) 288.30; found: (M+Cl) 323.09, MS (ESI+) calculated (M) 288.30; found: (M+Na) 311.11.

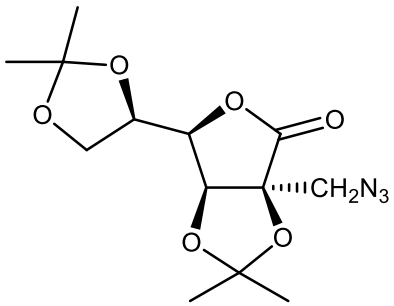
**2,3:5,6-Di-*O*-isopropylidene-2-*C*-trifluoromethanesulphonyloxymethyl-D-mannono-1,4-lactone (44)**[88]




To a solution of **(43)** (1.2 g, 4.185 mmol) in 1 ml pyridine and 6.25 ml $CH_2Cl_2$, 1 ml triflic anhydride was added, under nitrogen atmosphere. The reaction was finished after 1 hour, as judged by TLC analysis. The mixture was diluted with 50 ml $CH_2Cl_2$ and washed with 60ml 0.1 M HCl solution. The organic phase was dried with anhydrous $MgSO_4$ and concentrated by rotary evaporation. No further purification was performed to obtain **(44)** (1.3 g; 75%).

$^1$H NMR (400 MHz, $CHCl_3$-*d*) δ ppm 1.39, 1.42, 1.47, 1.48 (12H, 4 x s, 2 x $C(CH_3)_2$), 4.06–4.18 (2H, m, H-6a, H-6b), 4.35 (1H, dd, $J_{3,4}$ = 3.5 Hz, $J_{4,5}$ = 8.3 Hz, H-4), 4.45 (1H, ddd, $J_{4,5}$ = 8.5 Hz, $J_{5,6a}$ = 3.5 Hz, $J_{5,6b}$ = 6.0 Hz, H-5), 4.67 (1H, d, $J_{2'a;2'b}$ = 11.0 Hz, H-2'a), 4.76 (1H, d, $J_{2'a;2'b}$ = 11.0 Hz, H-2'b), 4.88 (1H, d, $J_{3,4}$ = 3.5 Hz, H-3); $^{13}$C NMR (400 MHz, $CHCl_3$-*d*) δ ppm 25.0, 25.9, 26.8, 26.9 ($C(CH_3)_2$), 66.3 (C-6), 70.7 (C-2$_0$), 72.2 (C-5), 77.6 (C-3), 78.1 (C-4), 83.5 (C-2), 110.1, 115.3 (2 $CMe_2$), 171.5 (C-1) MS (ESI+) calculated (M) 420,35; found: (M+Na) 443.04.

**2-*C*-Azidomethyl-2,3:5,6-di-*O*-isopropylidene-D-mannono-1,4-lactone (45)**[88]

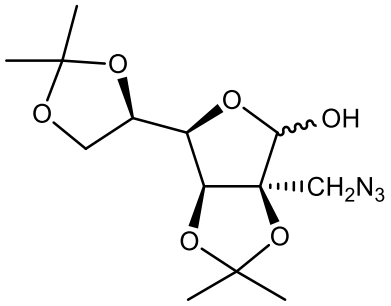



**(44)** (1.12g; 2.7 mmol) was dissolved in 3.5 ml DMF. To this solution, sodium azide (0.234 g; 3.6 mmol) was added. the mixture was stirred for 3 hours at room temperature under nitrogen atmosphere. After the reaction was completed as judged by TLC analysis, the solvent was removed by rotary evaporation and the residue was redissolved in 25 ml EtOAc. The solution was washed with brine, before the organic layer was dried with anhydrous $MgSO_4$ and concentrated by rotary evaporation, to obtain **(45)** (0.56 g; 63%).

$^1$H NMR (400 MHz, $CHCl_3$-*d*) δ ppm 1.38, 1.41, 1.47, 1.48 (12H, 4 x s, 2 x $C(CH_3)_2$), 3.58 (1H, d, $J_{2'a;2'b}$ = 13.0 Hz, H-2'a), 3.81 (1H, d, $J_{2'a;2'b}$ = 13.0 Hz, H-2'b), 4.06

(1H, dd, $J_{5,6a}$ = 3.5 Hz, $J_{6a,6b}$ = 8.5 Hz, H-6a), 4.14 (1H, dd, $J_{5,6b}$ = 6.0 Hz, $J_{6a,6b}$ = 8.5 Hz, H-6b), 4.31 (1H, dd, $J_{3,4}$ = 3.5 Hz, $J_{4,5}$ = 8.5 Hz, H-4), 4.41 (1H, ddd, $J_{4,5}$ = 8.5 Hz, $J_{5,6a}$ = 3.5 Hz, $J_{5,6b}$ = 6.0 Hz, H-5), 4.76 (1H, d, $J_{3,4}$ = 3.5 Hz, H-3); $^{13}$C NMR (400 MHz, $CHCl_3$-*d*) δ ppm 25.0, 26.1, 26.9, 27.0 ($C(CH_3)_2$), 50.5 (C-2'), 66.4 (C-6), 72.4 (C-5), 78.1 (C-4), 78.3 (C-3), 84.9 (C-2) 110.0, 114.5 (2 x $CMe_2$), 173.8 (C-1) MS (ESI+) calculated (M) 313.31; found: (M+Na) 336.10.

**2-*C*-Azidomethyl-2,3:5,6-di-*O*-isopropylidene-D-mannofuranose (46)**[88]

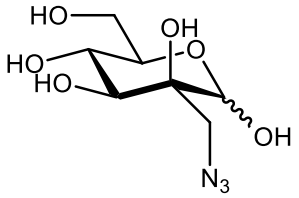



**(45)** (560 mg; 1.7 mmol) was dissolved in 6 ml $CH_2Cl_2$. This mixture was cooled to -78°C (acetone / $CO_2$ ice). To the cold mixture, DIBALH (305 mg; 2.1 mmol) solution in toluene was added. The reaction left to stir for 1,5 hours at this temperature. When reaction was finished as judged by TLC analysis, 2.5 ml of methanol was added to quench the reaction. The mixture was allowed to warm up to room temperature. Saturated potassium sodium tartrate solution (12.5 ml) was added to the mixture, and stirred for another 1.5 h. The mixture was then extracted with 3x 30 ml $CH_2Cl_2$, the combined organic layers were dried with anhydrous $MgSO_4$ and concentrated by rotary evaporation. The residue was purified via flash chromatography using petrol ether / EtOAc (50:10 to 20:10) as eluent, to obtain **(46)** (386 mg; 72%).

$^1$H NMR (400 MHz, $CHCl_3$-*d*) δ ppm 1.37, 1.45, 1.48, 1.49 (12H, 4 x s, 2 x $C(CH_3)^A_2$ ), 1.36, 1.44, 1.53, 1.56 (12H, 4 x s, 2 x $C(CH_3)^B_2$ ), 2.91 (1H, d, $J_{1,OH}$ =

1.0 Hz, $OH^A$), 3.47 (1H, dd, $J_{3,4}$ = 2.5 Hz, $J_{4,5}$ = 8.5 Hz, H-$4^B$), 3.54 (1H, d, $J_{2'a;2'b}$ = 13.0 Hz, H-2'$a^A$), 3.58 (1H, d, $J_{2'a;2'b}$ = 12.0 Hz, H-2'$a^B$), 3.68 (1H, d, $J_{2'a;2'b}$ = 13.0 Hz, H-2'$b^A$), 3.74 (1H, d, $J_{1,OH}$ = 11.5 Hz, $OH^B$), 3.87 (1H, d, $J_{2'a;2'b}$ = 12.0 Hz, H-2'$b^B$), 4.02–4.12 (5H, m, H-$4^A$, H-6$a^A$, H-6$b^A$, H-6$a^B$, H-6$b^B$), 4.35 (1H, ddd, $J_{4,5}$ = 8.5 Hz, $J_{5,6a}$ = 4.5 Hz, $J_{5,6b}$ = 6.0 Hz, H-$5^B$), 4.39 (1H, ddd, $J_{4,5}$ = 8.0 Hz, $J_{5,6a}$ = 4.5 Hz, $J_{5,6b}$ = 6.0 Hz, H-$5^A$), 4.51 (1H, d, $J_{3,4}$ = 3.0 Hz, H-$3^A$), 4.59 (1H, d, $J_{3,4}$ = 2.5 Hz, H-$3^B$), 4.83 (1H, d, $J_{1,OH}$ = 11.5 Hz, H-$1^B$), 5.34 (1H, d, $J_{1,OH}$ = 1.0 Hz, H-$1^A$); $^{13}$C NMR (400 MHz, $CHCl_3$-*d*) δ ppm 25.2, 26.6, 26.9, 27.0, 27.0, 27.1, 27.1, 27.5 ($C(CH_3)^A_2$ , $C(CH_3)^B_2$ ), 52.5 (C-2'$^A$), 52.9 (C-2'$^B$), 66.8 (C-$6^A$), 67.1 (C-$6^B$), 72.7 (C-$5^B$), 72.9 (C-$5^A$), 76.3 (C-$4^B$), 80.5 (C-$4^A$), 82.4 (C-$3^B$), 83.3 (C-$3^A$), 94.3 (C-$2^A$). 94.3 (C-$2^B$), 98.0 (C-$1^B$), 102.4 (C-$1^A$), 109.4, 114.7 (2 x $CMe^A_2$), 109.5, 114.9 (2 x $CMe^B_2$). MS (ESI+) calculated (M) 315.32; found: (M+Na) 338.12, (2M+Na) 653.27.

**2-*C*-Azidomethyl-D-mannopyranose (47)**[88]

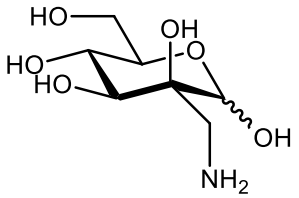



A mixture of **(46)** (454 mg; 1.4 mmol) and 300mg Dowex 50WX8-100 $H^+$ resin in 5 ml water / dioxane (50:50) was stirred for 2 days at 45°C mixture. The resin was then removed by filtration and the solvent was removed using lyophilisation, to yield **(47)** (273 mg; 83 %).

$^1$H NMR (400 MHz, $D_2O$) δ ppm 3.41 (1H, H-$5^B$), 3.47 (1H, H2'$^A$), 3.48 (1H, H2'$^B$), 3.58 (1H, H-4'$^B$), 3.60 (1H H-2'$^B$), 3.61 (1H, H-$3^B$), 3.65 (2H H-$3^a$, H-$4^A$), 3.68 (1H, H-2'$^A$), 3.73 (1H, H-$6^B$), 3.76 (1H, H-$6^A$), 3.83 (1H, H-$5^A$), 3.86 (1H, H$6^A$),

3.90 (1H, H-$6^B$), 4.85 (1H, H-$1^B$), 5,12 (1H, H-$1^A$). MS (ESI-) calculated (M) 235.08; found: (M-H) 234.09.

**2-*C*-Aminomethyl-D-mannopyranose (48)**[88]




**(47)** (20 mg; 0.8 mmol) was dissolved in 1 ml water / ethanol (50:50). To this solution, 2 mg palladium black was added. The flask was degassed and left overnight stirring under a hydrogen gas atmosphere. After the reaction was completed as judged by TLC, the mixture was passed through a Celite® filter and the aqueous phase was removed using lyophilisation, to yield **(48)** (16 mg; 100 %)

$^1$H NMR (400 MHz, $D_2O$) δ ppm 3.19 (2H, H-2'$^{A,B}$), 3.32 (1H, H-2'$^A$), 3.33 (1H, H-$6^A$), 3.38 (1H, H-2'$^B$), 3.46 (1H, H-$5^B$), 3.53 (1H, H-2'$^B$), 3.61 (1H, H-$3^B$), 3.67 (1H, H-$3^B$), 3.74 (1H, H-$6^B$), 3.78 (1H, H-$6^A$), 3.82 (1H, H-2'$^A$), 3.87 (2H H-$5^a$, H-$6^A$), 4,87 (1H H-$1^B$), 5,04 (1H, H-$1^A$). MS (ESI+) calculated (M) 209.09; found: (M+Na) 232.13, (2M+Na) 441.23.

# 6 References

1. Fasting, C.; Schalley, C. A.; Weber, M.; Seitz, O.; Hecht, S.; Koksch, B.; Dernedde, J.; Graf, C.; Knapp, E.-W.; Haag, R., Multivalency as a chemical organization and action principle. *Angew Chem Int Ed* **2012,** *51* (42), 10472-10498.

2. Reynolds, M.; Perez, S., Thermodynamics and chemical characterization of protein-carbohydrate interactions: The multivalency issue. *Comptes Rendus Chimie* **2011,** *14* (1), 74-95.

3. Hunter, C. A.; Anderson, H. L., What is cooperativity? *Angew Chem Int Ed* **2009,** *48* (41), 7488-7499.

4. Reynolds, M.; Pérez, S., Thermodynamics and chemical characterization of protein–carbohydrate interactions: The multivalency issue. *Cr Chim* **2011,** *14* (1), 74-95.

5. Ercolani, G.; Schiaffino, L., Allosteric, Chelate, and interannular cooperativity: a mise au point. *Angew Chem Int Ed* **2011,** *50* (8), 1762-1768.

6. Schwarzenbach, G., Der Chelateffekt. *Helv Chim Acta* **1952,** *35* (7), 2344-2363.

7. Mammen, M.; Dahmann, G.; Whitesides, G. M., Effective inhibitors of hemagglutination by influenza-virus synthesized from polymers having active ester groups - insight into mechanism of inhibition. *J Med Chem* **1995,** *38* (21), 4179-4190.

8. Hensley, S. E.; Das, S. R.; Bailey, A. L.; Schmidt, L. M.; Hickman, H. D.; Jayaraman, A.; Viswanathan, K.; Raman, R.; Sasisekharan, R.; Bennink, J. R.; Yewdell, J. W., Hemagglutinin receptor binding avidity drives influenza a virus antigenic drift. *Science* **2009,** *326* (5953), 734-736.

9. Mammen, M.; Choi, S. K.; Whitesides, G. M., Polyvalent interactions in biological systems: Implications for design and use of multivalent ligands and inhibitors. *Angew Chem Int Ed* **1998,** *37* (20), 2755-2794.

10. Fasting, C.; Schalley, C. A.; Weber, M.; Seitz, O.; Hecht, S.; Koksch, B.; Dernedde, J.; Graf, C.; Knapp, E. W.; Haag, R., Multivalency as a chemical organization and action principle. *Angew Chem Int Ed* **2012,** *51* (42), 10472-10498.

11. Connell, H.; Agace, W.; Klemm, P.; Schembri, M.; Marild, S.; Svanborg, C., Type 1 fimbrial expression enhances Escherichia coli virulence for the urinary tract. *Proc Natl Acad Sci U S A* **1996,** *93* (18), 9827-9832.

12. Westerlund, B.; Korhonen, T. K., Bacterial proteins binding to the mammalian extracellular-matrix. *Mol Microbiol* **1993,** *9* (4), 687-694.

13. Varki, A., Selectin ligands: Will the real ones please stand up? *J Clin Invest* **1997,** *99* (2), 158-162.

14. Lowe, J. B.; Ward, P. A., Therapeutic inhibition of carbohydrate-protein interactions in vivo. *J Clin Invest* **1997,** *99* (5), 822-826.

15. Gimbrone, M. A.; Nagel, T.; Topper, J. N., Biomechanical activation: An emerging paradigm in endothelial adhesion biology. *J Clin Invest* **1997,** *99* (8), 1809-1813.

16. Woof, J. M.; Burton, D. R., Human antibody - Fc receptor interactions illuminated by crystal structures. *Nat Rev Immunol* **2004,** *4* (2), 89-99.

17. Engelhardt, W.; Gorczytza, H.; Butterweck, A.; Monkemann, H.; Frey, J., Structural requirements of the cytoplasmic domains of the human macrophage Fc-gamma receptor-IIa and B-cell Fc-gamma receptor-IIb2 for the endocytosis of immune-complexes. *Eur J Immunol* **1991,** *21* (9), 2227-2238.

18. Ravetch, J. V.; Bolland, S., IgG Fc receptors. *Annu Rev Immunol* **2001,** *19*, 275-290.

19. Barcroft, J.; Hill, A. V., The nature of oxyhaemoglobin, with a note on its molecular weight. *J Physiol* **1910,** *39* (6), 411-28.

20. Byers, L. D., Probe-dependent cooperativity patterns in Hill-plots. *J Chem Educ* **1977,** *54* (6), 352-354.

21. Hill, A. V., The combinations of haemoglobin with oxygen and with carbon monoxide. *Biochem J* **1913,** *7* (5), 471-80.

22. Nissani, E.; Perlmutterhayman, B., Drug-binding to biological macromolecules - A kinetic-study of the system chlorodiazepoxide (librium) and bovine serum-albumin. *Int J Chem Kinet* **1986,** *18* (10), 1123-1132.

23. Koshland, D. E.; Neet, K. E., Catalytic and regulatory properties of enzymes. *Annu Rev Biochem* **1968,** *37*, 359-&.

24. Pike, L. J., Negative cooperativity in the EGF receptor. *Faseb J* **2009,** *23*.

25. Alvarado, D.; Klein, D. E.; Lemmon, M. A., Structural basis for negative cooperativity in growth factor binding to an EGF receptor. *Cell* **2010,** *142* (4), 568-579.

26. Whitesides, G. M.; Takayama, S.; Holmlin, E.; Rao, J. H., Multivalency. *Abstr Pap Am Chem Soc* **1999,** *217*, U223-U223.

27. Glick, G. D.; Knowles, J. R., Molecular recognition of bivalent sialosides by influenza-virus. *J Am Chem Soc* **1991,** *113* (12), 4701-4703.

28. Glick, G. D.; Toogood, P. L.; Wiley, D. C.; Skehel, J. J.; Knowles, J. R., Ligand recognition by influenza-virus - The binding of bivalent sialosides. *J Biol Chem* **1991,** *266* (35), 23660-23669.

29. Kitov, P. I.; Bundle, D. R., On the nature of the multivalency effect: A thermodynamic model. *J Am Chem Soc* **2003,** *125* (52), 16271-16284.

30. Jencks, W. P., On the attribution and additivity of binding-energies. *Proc Natl Acad Sci U S A,-Biological Sciences* **1981,** *78* (7), 4046-4050.

31. Kramer, R. H.; Karpen, J. W., Spanning binding sites on allosteric proteins with polymer-linked ligand dimers. *Nature* **1998,** *395* (6703), 710-713.

32. Huskens, J.; Mulder, A.; Auletta, T.; Nijhuis, C. A.; Ludden, M. J. W.; Reinhoudt, D. N., A model for describing the thermodynamics of multivalent host-guest interactions at interfaces. *J Am Chem Soc* **2004,** *126* (21), 6784-6797.

33. Hogben, H. J.; Sprafke, J. K.; Hoffmann, M.; Pawlicki, M.; Anderson, H. L., Stepwise effective molarities in porphyrin oligomer complexes: preorganization results in exceptionally strong chelate cooperativity. *J Am Chem Soc* **2011,** *133* (51), 20962-20969.

34. Gupta, D.; Dam, T. K.; Oscarson, S.; Brewer, C. F., Thermodynamics of lectin-carbohydrate interactions. *J Biol Chem* **1997,** *272* (10), 6388-6392.

35. Dam, T. K.; Oscarson, S.; Roy, R.; Das, S. K.; Page, D.; Macaluso, F.; Brewer, C. F., Thermodynamic, kinetic, and electron microscopy studies of concanavalin A and Dioclea grandiflora lectin cross-linked with synthetic divalent carbohydrates. *J Biol Chem* **2005,** *280* (10), 8640-8646.

36. Dam, T. K.; Roy, R.; Page, D.; Brewer, C. F., Thermodynamic binding parameters of individual epitopes of multivalent carbohydrates to concanavalin a as determined by "reverse" isothermal titration microcalorimetry. *Biochemistry* **2002,** *41* (4), 1359-1363.

37. Dam, T. K.; Roy, R.; Page, D.; Brewer, C. F., Negative cooperativity associated with binding of multivalent carbohydrates to lectins. thermodynamic analysis of the "multivalency effect". *Biochemistry* **2002,** *41* (4), 1351-1358.

38. Dam, T. K.; Roy, R.; Das, S. K.; Oscarson, S.; Brewer, C. F., Binding of multivalent carbohydrates to concanavalin A and Dioclea grandiflora lectin - Thermodynamic analysis of the "multivalency effect". *J Biol Chem* **2000,** *275* (19), 14223-14230.

39. Ahmad, N.; Gabius, H. J.; Sabesan, S.; Oscarson, S.; Brewer, C. F., Thermodynamic binding studies of bivalent oligosaccharides to galectin-1, galectin-3, and the carbohydrate recognition domain of galectin-3. *Glycobiology* **2004,** *14* (9), 817-825.

40. Bayley, H.; Cremer, P. S., Stochastic sensors inspired by biology. *Nature* **2001,** *413* (6852), 226-30.

41. Braha, O.; Walker, B.; Cheley, S.; Kasianowicz, J. J.; Song, L.; Gouaux, J. E.; Bayley, H., Designed protein pores as components for biosensors. *Chem Biol* **1997,** *4* (7), 497-505.

42. Clarke, J.; Wu, H. C.; Jayasinghe, L.; Patel, A.; Reid, S.; Bayley, H., Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol* **2009,** *4* (4), 265-70.

43. Hammerstein, A. F.; Shin, S. H.; Bayley, H., Single-molecule kinetics of two-step divalent cation chelation. *Angew Chem Int Ed Engl* **2010,** *49* (30), 5085-90.

44. Movileanu, L.; Howorka, S.; Braha, O.; Bayley, H., Detecting protein analytes that modulate transmembrane movement of a polymer chain within a single protein pore. *Nat Biotechnol* **2000,** *18* (10), 1091-5.

45. Herget, S.; Toukach, P. V.; Ranzinger, R.; Hull, W. E.; Knirel, Y. A.; von der Lieth, C. W., Statistical analysis of the Bacterial Carbohydrate Structure Data Base (BCSDB): characteristics and diversity of bacterial carbohydrates in comparison with mammalian glycans. *BMC Struct Biol* **2008,** *8*, 35.

46. Shin, S. H.; Luchian, T.; Cheley, S.; Braha, O.; Bayley, H., Kinetics of a reversible covalent-bond-forming reaction observed at the single-molecule level. *Angew Chem Int Ed Engl* **2002,** *41* (19), 3707-9; 3523.

47. Hammerstein, A. F.; Jayasinghe, L.; Bayley, H., Subunit dimers of alpha-hemolysin expand the engineering toolbox for protein nanopores. *J Biol Chem* **2011,** *286* (16), 14324-34.

48. Cheley, S.; Braha, O.; Lu, X.; Conlan, S.; Bayley, H., A functional protein pore with a "retro" transmembrane domain. *Protein Sci* **1999,** *8* (6), 1257-67.

49. Howorka, S.; Nam, J.; Bayley, H.; Kahne, D., Stochastic detection of monovalent and bivalent protein-ligand interactions. *Angew Chem Int Ed Engl* **2004,** *43* (7), 842-6.

50. Song, L.; Hobaugh, M. R.; Shustak, C.; Cheley, S.; Bayley, H.; Gouaux, J. E., Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* **1996,** *274* (5294), 1859-66.

51. Davis, B. G., Synthesis of glycoproteins. *Chem Rev* **2002,** *102* (2), 579-602.

52. Davis, B. G., Biochemistry. Mimicking posttranslational modifications of proteins. *Science* **2004,** *303* (5657), 480-2.

53. Gamblin, D. P.; Garnier, P.; Ward, S. J.; Oldham, N. J.; Fairbanks, A. J.; Davis, B. G., Glycosyl phenylthiosulfonates (glyco-PTS): novel reagents for glycoprotein synthesis. *Org Biomol Chem* **2003,** *1* (21), 3642-4.

54. Davis, B. G.; Lloyd, R. C.; Jones, J. B., Controlled site-selective protein glycosylation for precise glycan structure-catalytic activity relationships. *Bioorg Med Chem* **2000,** *8* (7), 1527-35.

55. Chalker, J. M.; Bernardes, G. J.; Lin, Y. A.; Davis, B. G., Chemical modification of proteins at cysteine: opportunities in chemistry and biology. *Chem Asian J* **2009,** *4* (5), 630-40.

56. van Kasteren, S. I.; Kramer, H. B.; Gamblin, D. P.; Davis, B. G., Site-selective glycosylation of proteins: creating synthetic glycoproteins. *Nat Protoc* **2007,** *2* (12), 3185-94.

57. Shanley, S.A., A Glycopore for Bacterial Sensing, *DPhil Thesis, University of Oxford* **2009.**

58. Chapman, T. M.; Courtney, S.; Hay, P.; Davis, B. G., Highly diastereoselective additions to polyhydroxylated pyrrolidine cyclic imines: Ready elaboration of aza-sugar scaffolds to create diverse carbohydrate-processing enzyme probes. *Chem Eur J* **2003,** *9* (14), 3397-3414.

59. Wellens, A.; Garofalo, C.; Nguyen, H.; Van Gerven, N.; Slättegård, R.; Hernalsteens, J. P.; Wyns, L.; Oscarson, S.; De Greve, H.; Hultgren, S.; Bouckaert, J., Intervening with urinary tract infections using anti-adhesives based on the crystal structure of the FimH-oligomannose-3 complex. *PLoS One* **2008,** *3* (4), e2040.

60. Schembri, M. A.; Hasman, H.; Klemm, P., Expression and purification of the mannose recognition domain of the FimH adhesin. *FEMS Microbiol Lett* **2000,** *188* (2), 147-51.

61. Mackiewicz, A.; Mackiewicz, S., Determination of lectin-sugar dissociation constants by agarose affinity electrophoresis. *Anal Biochem* **1986,** *156* (2), 481-8.

62. Prasthofer, T.; Phillips, S. R.; Suddath, F. L.; Engler, J. A., Design, expression, and crystallization of recombinant lectin from the garden pea (Pisum sativum). *J Biol Chem* **1989,** *264* (12), 6793-6.

63. Banerjee, R.; Mande, S. C.; Ganesh, V.; Das, K.; Dhanaraj, V.; Mahanta, S. K.; Suguna, K.; Surolia, A.; Vijayan, M., Crystal structure of peanut lectin, a protein with an unusual quaternary structure. *Proc Natl Acad Sci U S A* **1994,** *91* (1), 227-31.

64. Thépaut, M.; Valladeau, J.; Nurisso, A.; Kahn, R.; Arnou, B.; Vivès, C.; Saeland, S.; Ebel, C.; Monnier, C.; Dezutter-Dambuyant, C.; Imberty, A.; Fieschi, F., Structural studies of langerin and Birbeck granule: a macromolecular organization model. *Biochemistry* **2009,** *48* (12), 2684-98.

65. Thépaut, M.; Vivès, C.; Pompidor, G.; Kahn, R.; Fieschi, F., Overproduction, purification and preliminary crystallographic analysis of the carbohydrate-recognition domain of human langerin. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **2008,** *64* (Pt 2), 115-8.

66. Feinberg, H.; Mitchell, D. A.; Drickamer, K.; Weis, W. I., Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. *Science* **2001,** *294* (5549), 2163-6.

67. Guo, Y.; Feinberg, H.; Conroy, E.; Mitchell, D. A.; Alvarez, R.; Blixt, O.; Taylor, M. E.; Weis, W. I.; Drickamer, K., Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. *Nat Struct Mol Biol* **2004,** *11* (7), 591-8.

68. Mitchell, D. A.; Fadden, A. J.; Drickamer, K., A novel mechanism of carbohydrate recognition by the C-type lectins DC-SIGN and DC-SIGNR. Subunit organization and binding to multivalent ligands. *J Biol Chem* **2001,** *276* (31), 28939-45.

69. Feinberg, H.; Castelli, R.; Drickamer, K.; Seeberger, P. H.; Weis, W. I., Multiple modes of binding enhance the affinity of DC-SIGN for high mannose N-linked glycans found on viral glycoproteins. *J Biol Chem* **2007,** *282* (6), 4202-9.

70. Geijtenbeek, T. B.; van Kooyk, Y., Pathogens target DC-SIGN to influence their fate DC-SIGN functions as a pathogen receptor with broad specificity. *APMIS* **2003,** *111* (7-8), 698-714.

71. van Kooyk, Y.; Geijtenbeek, T. B., DC-SIGN: escape mechanism for pathogens. *Nat Rev Immunol* **2003,** *3* (9), 697-709.

72. Bandyopadhyay, P.; Bandyopadhyay, P.; Regen, S. L., An ion conductor that recognizes osmotically-stressed phospholipid bilayers. *J Am Chem Soc* **2002,** *124* (38), 11254-11255.

73. Ahmed, S. A.; Tanaka, M., Synthesis of Oligo(ethylene glycol) toward 44-mer. *J Org Chem.* **2006,** *71* (26), 9884-9886.

74. Wang, S. K.; Liang, P. H.; Astronomo, R. D.; Hsu, T. L.; Hsieh, S. L.; Burton, D. R.; Wong, C. H., Targeting the carbohydrates on HIV-1: Interaction of oligomannose dendrons with human monoclonal antibody 2G12 and DC-SIGN. *Proc Natl Acad Sci U S A* **2008,** *105* (10), 3690-3695.

75. Xue, J. L.; Cecioni, S.; He, L.; Vidal, S.; Praly, J. P., Variations on the $SnCl_4$ and $CF_3CO_2Ag$-promoted glycosidation of sugar acetates: a direct, versatile and apparently simple method with either alpha or beta stereocontrol. *Carbohyd Res* **2009,** *344* (13), 1646-1653.

76. Veeneman, G. H.; van Boom, J. H., An efficient thioglycoside-mediated formation of [alpha]-glycosidic linkages promoted by iodonium dicollidine perchlorate. *Tetrahedron Lett* **1990,** *31* (2), 275-278.

77. Konradsson, P.; Udodong, U. E.; Fraser-Reid, B., Iodonium promoted reactions of disarmed thioglycosides. *Tetrahedron Lett* **1990,** *31* (30), 4313-4316.

78. Kumar, G. D. K.; Baskaran, S., A facile, catalytic, and environmentally benign method for selective deprotection of tert-butyldimethylsilyl ether mediated by phosphomolybdic acid supported on silica gel. *J Org Chem* **2005,** *70* (11), 4520-4523.

79. Kumar, P. S.; Kumar, G. D. K.; Baskaran, S., Truly catalytic and chemoselective cleavage of benzylidene acetal with phosphomolybdic acid supported on silica gel. *Eur J Org Chem* **2008,** (36), 6063-6067.

80. Kumar, D.; Kumar, R.; Chakraborti, A. K., Tetrafluoroboric acid adsorbed on silica gel as a reusable heterogeneous dual-purpose catalyst for conversion of aldehydes/ketones into acetals/ketals and back again. *Synthesis-Stuttgart* **2008,** (8), 1249-1256.

81. Thompson, D. K.; Hubert, C. N.; Wightman, R. H., Hydroxylated pyrrolidines - Synthesis of 1,4-dideoxy-1,4-imino-l-lyxitol, 1,4,5-trideoxy-1,4-imino-d-glycero-l-altro-octitols and 1,4,5-trideoxy-1,4-imino-l-lyxo-hexitol, 2,3,6-trideoxy-3,6-imino-d-glycero-l-altro-octitols and 2,3,6-trideoxy-3,6-imino-d-glycero-l-galacto-octitols, and of a chiral potential precursor of carbapenem systems. *Tetrahedron* **1993,** *49* (18), 3827-3840.

82. Mulder, A.; Auletta, T; Sartori, A.; Del Ciotto, S.; Casnati, A.; Ungaro, R.; Huskens, J.; Reinhoudt, D.N., Divalent binding of a bis(adamantyl)-functionalized calix[4]arene to β-cyclodextrin-based hosts: an experimental and theoretical study on multivalent binding in solution and at self-assembled monolayers. *J Am Chem Soc* **2004**, *126* (21), 6627-6636**.**

83. Hladky, S., Haydon, D. Discreteness of conductance change in bimolecular lipid membranes in the presence of certain antibiotics. *Nature* **1970,** *225***,** 451–453.

84. Bayley, H., Braha, O. & Gu, L.Q. Stochastic sensing with protein pores. *Adv Mater* **2000,** *12*, 139-142.

85. Fick, A., On liquid diffusion. *Annalen der Physik und Chemie* **1855***, 94: 59.*

86. Atkins, P.; de Paula, J., Physical chemistry for the life science, *Oxford University Press* **2006**.

87. Morelli, L.; Poletti, L.; Lay, L., Carbohydrates and immunology: synthetic oligosaccharide antigens for vaccine formulation. *Eur J Org Chem* **2011,** (29), 5723-5777.

88. Mitchell, D.A.; Jones, N.; Hunter, S.; Cook, J.; Jenkinson, S.; Wormald, M.A.; Dwek, R.A.; Fleet, G.W.J., Synthesis of 2-C-branched derivatives of d-mannose: 2-C-aminomethyl-d-mannose binds to the human C-type lectin DC-SIGN with affinity greater than an order of magnitude compared to that of d-mannose, *Tetrahedron Asymmetry* **2007**, (18), 1502-1510.

89. Soengas, R.; Izumori, K.; Simone, M. I.; Watkin, D.J.; Skytte, U.P.; Soetaert, W.; Fleet, G.W.J., Kiliani reactions on ketoses: branched carbohydrate building blocks from D-tagatose and D-psicose, *Tetrahedron Lett* **2005**, *46*, 5755-5759.

90. Keegstra, E.M.D.; Zwikker, J.W.; Roest, M.R.; Jenneskens, L.W., A highly selective synthesis of monodisperse oligo(ethylene glycols). *J Org Chem* **1992***, 57*, 6678-6680.

91. Chen, Y.Y.; Baker, G.L. Synthesis and properties of ABA amphiphiles. *J Org Chem* **1999***, 64*, 6870-6873.

92 Lee, E., Bruzzi, A.; O'Brien, E., O'Colla, P.S. Selective acetylation of D -mannose and D -xylose. *Carbohyd Res* **1979**, *71*, 331-334.

93. Schröter, S., Male-specific modification of human CD52. *J Biol Chem* **1999**, 274(42): 29862-73.

94. Franks, N.E.; Montgomery, R., Stereoselective ring-opening of β-D-mannopyranose 1,2-(alkyl orthoacetates), *Carbohyd Res* **1968**, *6* (3), 286-298.

95. Jain, R.K.; Liu. X.; Oruganti, S.R.; Chandrasekaran E.V.; Matta, K.L., Synthesis of oligosaccharide substrates for N-linked glycoprotein processing enzymes, *Carbohyd Res* **1995**, *271* (2), 185-196.

96. Saksena, R.; Zhang, J.; Kováč, P., Synthesis of 2-(trimethylsilyl)ethyl α-D-mannopyranosides revisited, *J Carbohydr Chem* **2002**, *21* (6), 453-470.

97. Guo, X.; Cheng, C.; Liu, X.; Wang, R.; Xie, D.; Wang, L.; Zhao, Y., Crystallographic and theoretical study of phenyl 2,3,4,6-tetra-O-acetyl-1-thio-α-d-mannopyranoside, *J Mol Struc-THEOCHEM* **2004**, *710* (1–3), 201-205.

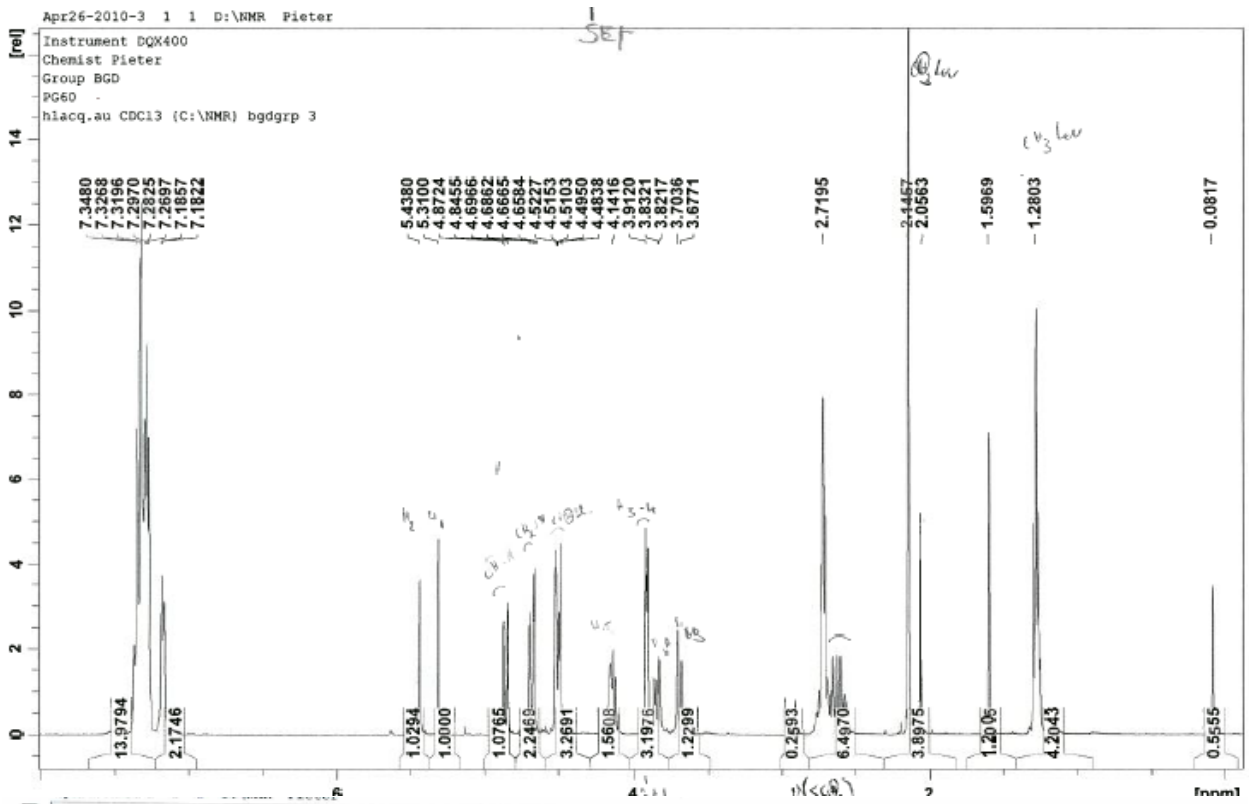
98. Van der Marel, G.; Codee, J., Carbohydrate chemistry: proven synthetic methods, Volume 2, *CRC Press* **2014**, 2, 177.

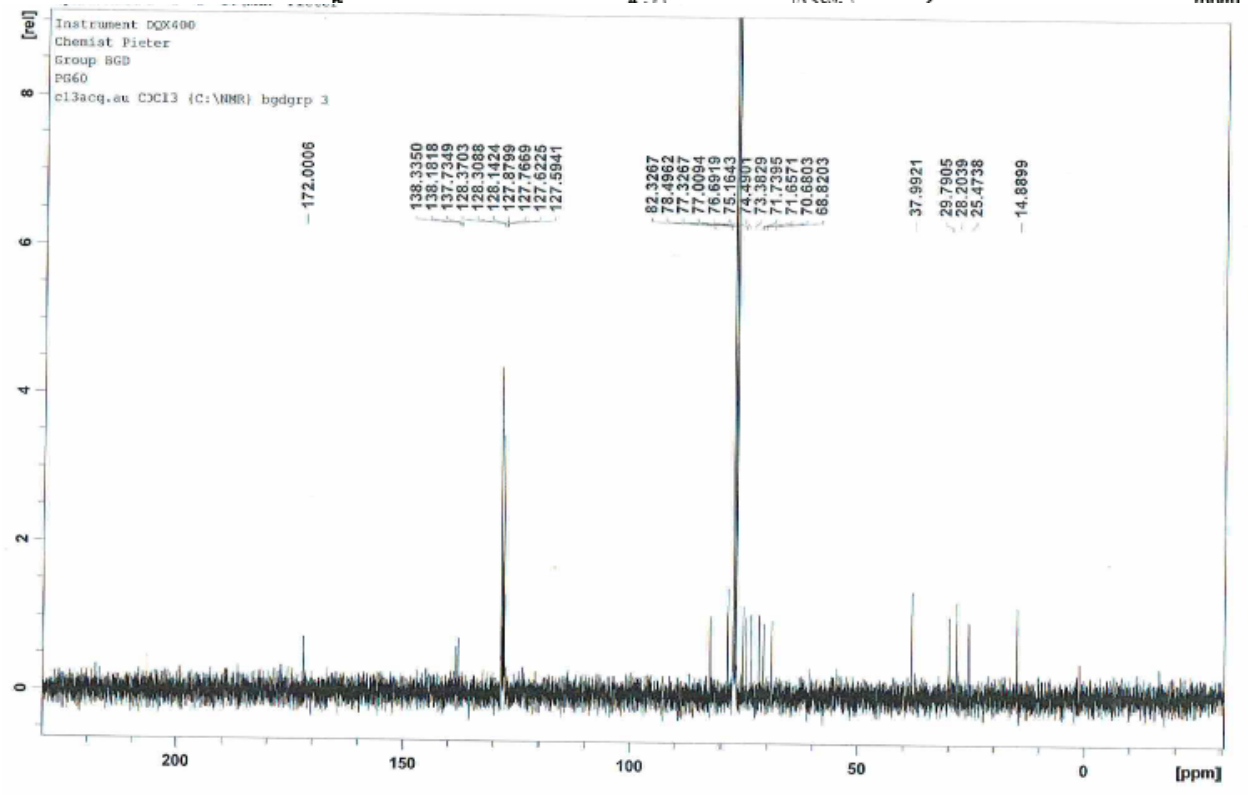
99. Lovell, S.C., Davis, I.W., Arendall III, W.B., de Bakker, P.I.W., Word, J. M., Prisant, M. G., Richardson, J. S., Richardson, D.C. Structure validation by C α geometry: φ, ψ and C β deviation, *Proteins: Struct, Funct, Bioinf* **2003**, *50*, 437-450.

100. De Mestral, G., Velvet type fabric and method of producing same, US Patent 2,717,437, **1955**.

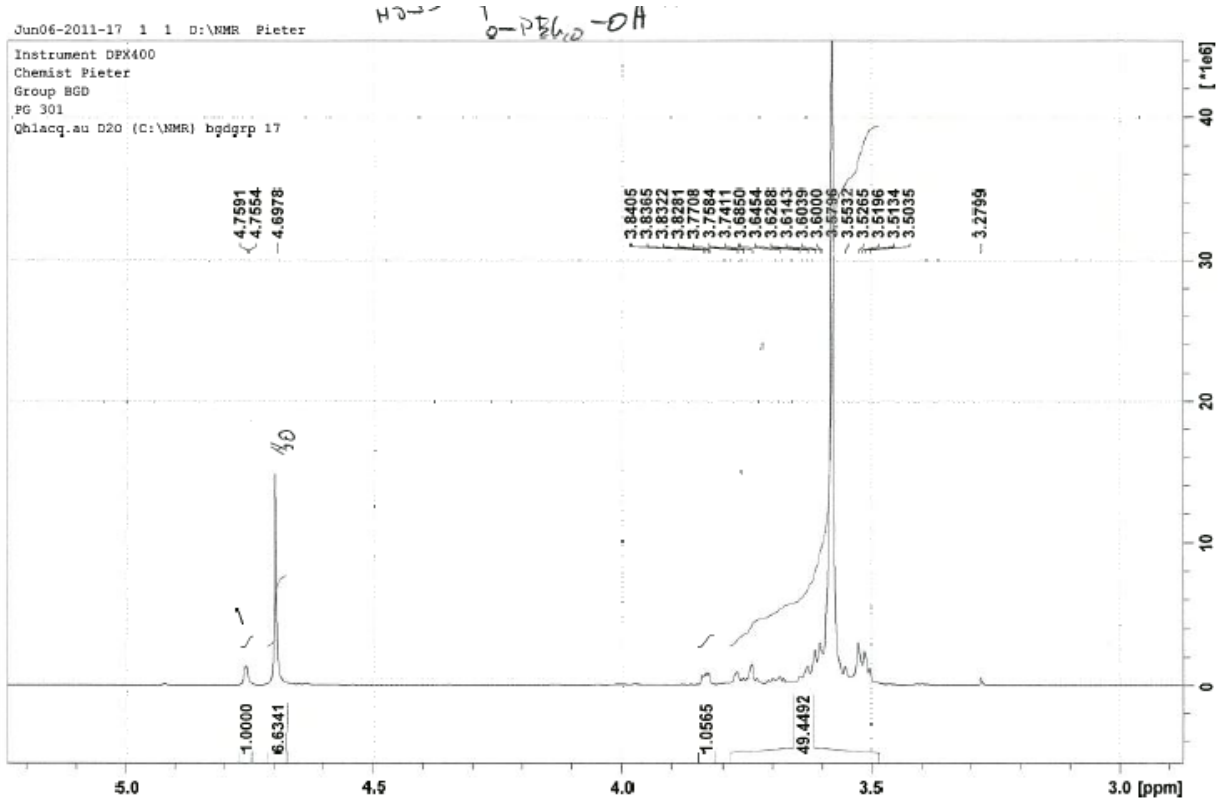
# 7 Addendum

## S-Ethyl 3,4,6-tri-*O*-benzyl-2-*O*-levulinate-1-thio-α-D-mannopyranoside (13)

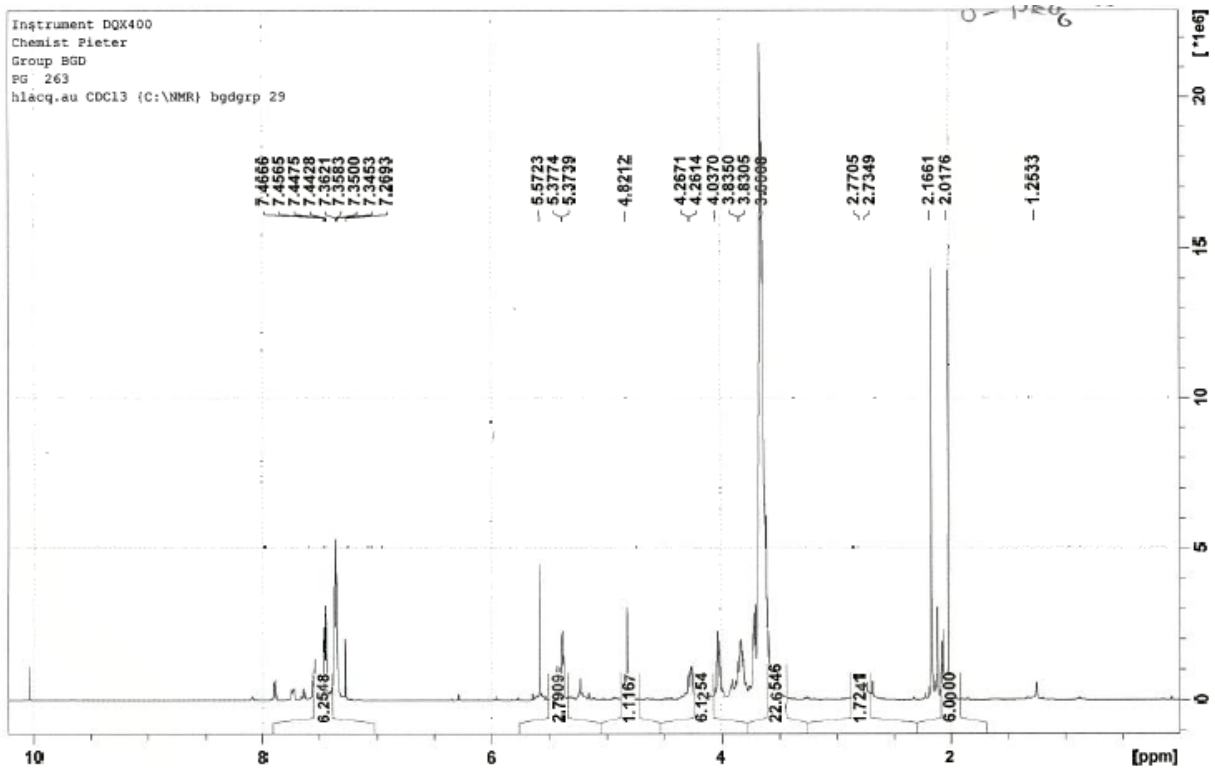


## 1-(3, 6, 9, 12, 15, 18, 21, 24, 27-Nonaoxa-29-hydroxy-29yl)-α-D-mannopyranoside (37)



## (3', 6', 9', 12', 15', 18', 21', 24', 27'-Nonaoxa-29'-hydroxy-29yl) 4,6-O-benzylidene-α-D-mannopyranoside (38')



**(29'-O-T*ert*-butyldiphenylsilane-3', 6', 9', 12', 15', 18', 21', 24', 27'-29'-decoxa-29yl) 4,6-O-benzylidene-α-D-mannopyranoside (40)**

