THE UNIVERSITY
of EDINBURGH

NMR and *in silico* studies of fucosylated chondroitin sulfate (fCS) and its interactions with selectins

Natalia Brodaczewska

Thesis submitted for the degree of Doctor of Philosophy

School of Chemistry

University of Edinburgh

2017

Declaration

I hereby declare that this thesis was composed entirely by myself, in the School of Chemistry at the University of Edinburgh. The work reported here is my own and has not been submitted for any other degree or professional qualification.

Natalia Brodaczewska

August 2017

Abstract

This thesis describes structural studies on the interactions between the fucosylated chondroitin sulfate (fCS) oligosaccharides and human proteins known as selectins. fCS is a carbohydrate obtained from sea cucumbers, that can be classified as a branched glycosaminoglycan (GAG). It has attracted much attention due to its anti-coagulant, anti-inflammatory, antimetastatic and anti-HIV properties and its structure was previously determined by NMR. Selectins constitute a family of proteins involved in cell adhesion processes, such as inflammation, attachment of viral particles and migration of tumour cells. fCS oligosaccharides have been shown to bind to selectins, which is likely a reason behind their biological activity. However, the mechanism of this interaction is currently unknown.

The initial part of the thesis describes the experimental work on expression and purification of the recombinant L- and P-selectin constructs in *Pichia pastoris*, *Escherichia coli* and HEK 293 cells. The aim of these experiments was to produce two constructs for each selectin, a single domain construct, consisting of the C-type lectin domain only, and a double domain construct, consisting of both the C-type lectin and the EGF-like domains. The intention was that the recombinant proteins would be labelled with ^{13}C and ^{15}N to allow for the in-depth structural NMR studies on the fCS-selectin interaction. Various experimental approaches have been explored, including the use of different cell lines, modifications to construct design, as well as alterations to expression and purification conditions. Although it was not possible to produce soluble selectin constructs in either bacterial or yeast cells, protein expression tests in HEK293 cells, performed in collaboration with the Oxford Protein Production facility (OPPF), led to production of a soluble L-selectin construct, consisting of the L-selectin C-type lectin domain.

The produced L-selectin construct, as well as two commercially available constructs of the L- and P-selectin extracellular domains, were used in the Saturation Transfer Difference (STD) NMR experiments to provide new information about the nature of the fCS-selectin binding. The STD experiments allowed to identify the regions within the fCS oligosaccharides that are in direct contact with the protein and likely play an important role in this interaction. Experiments on different protein constructs allowed the comparison of fCS binding to P-selectin and to two different recombinant constructs of L-selectin. Results of these studies suggest that the binding occurs via a similar mechanism for both L- and P-selectins and that

the fCS oligosaccharides bind to one-domain L-selectin construct with similar affinity as to a larger construct, consisting of the entire extracellular region of the protein.

Alongside the experimental work, theoretical *in silico* studies on the fCS-selectin binding were undertaken as part of this project. The existing X-ray structures of selectin complexes were subjected to Molecular Dynamics (MD) simulations, which allowed to explore the dynamic behaviour of E-selectin upon binding to sialyl Lewis x (sLe^x). It was found that sLe^x forms a more favourable interaction with the extended conformation of E-selectin and that the protein in this conformation is characterised by a high degree of interdomain flexibility, with a new type of interdomain movement observed in the MD studies on this complex.

In further *in silico* studies, the fCS oligosaccharides were docked to the existing P-selectin structures. The docking tests were performed on the computationally produced fCS trisaccharides with fucose branches either 2,4 or 3,4-sulfated. Results were evaluated with MD simulations and analysed in the light of current knowledge of selectin-ligand binding and the STD NMR experimental results. The *in silico* studies allowed to identify a subset of P-selectin residues that are likely involved in the interaction with fCS oligosaccharides *in vivo*. The conformational behaviour of P-selectin upon binding to fCS was also explored and it was found that the interdomain hinge is flexible during this interaction and allows transition from bent to extended conformational state.

Finally, a new NMR method was developed to facilitate the studies of complex carbohydrates, incorporating the concepts of G-matrix Fourier Transform (GFT) NMR into 2D HSQC and 2D HSQC-TOCSY experiments. The method allows to separate peaks in the regions of high spectral overlap, providing information that can simplify the assignment process. The new experiments facilitated the structural evaluation of a sample containing a mixture of oligosaccharides resulting from the depolymerisation of fCS polysaccharide.

Lay summary

Marine organisms have adapted to life in harsh environments, surviving at great ocean depths with high salt concentrations and extreme pressure and temperature conditions. Because of that, they are able to produce bioactive compounds that are not found anywhere else on Earth. Some of these compounds have medical properties that could serve as an inspiration in design of new drugs. One example of such compound is a new type of carbohydrate found in the body walls of the sea cucumbers, named fucosylated chondroitin sulfate (fCS). It has attracted much interest as it was shown to reduce inflammation, HIV infection and migration of cancer cells to new tissues. The main aim of the work described in this thesis was to study the molecular mechanisms that underlie these biological properties of fCS and to obtain new information that would help in design of new drugs for treatment of these conditions.

We believe that the biological activity of the fCS is due to its interaction with a family of proteins called selectins. Selectins help cells to attach to each other, allowing certain cell types to migrate to new regions within the body. For example, the cells of our immune system need to travel to the site of injury during inflammation. Previous research has shown that fCS binds to two selectins more strongly than carbohydrates found on the surface of our cells. This means that it could outcompete the natural binders of selectins, interfering with their function and preventing migration of some of the cells, such as tumour cells.

The studies described in this thesis allowed to describe the interaction between the fCS and selectins in more detail, providing information about the structural regions within both molecules that are involved in the binding. We were able to compare the binding between the two types of selectins and predict the region within selectins which is responsible for the binding to fCS. This information was obtained by using a combination of experimental and computational methods.

Acknowledgements

I want to thank my supervisor Prof. Dušan Uhrín for giving me the opportunity to work on this project and for providing me with much needed support, as well as for his patience in explaining NMR, and for helping me to finish this thesis and not giving up on me!

I would also like to thank Glycomar – Dr Charlie Bavington, Dr Claire Moss and Dr Derek Thomson – for co-funding and collaborating with me on this project and for hosting me in beautiful Oban and teaching me all about the sea cucumbers.

Thank you to the amazing team at the Oxford Protein Production Facility who have put so much effort into helping me express my protein. Special thanks to Prof. Ray Owens for finding time and resources to host me in the OPPF labs and to Dr Nahid Rahman and Dr Joanne Nettleship for the kind support and hard work that they have contributed to this project.

A big thank you to the Barlow group – Dr Nicky McIntosh, Dr Eliza Makou, Dr Andy Herbert and Christopher Weir – for providing me with protocols, reagents and hours and hours of support in my protein expression experiments, for explaining everything with so much patience and for all the fun we had in the unforgettable lab 120.

I especially want to thank my wonderful research group and I would need way more pages to say all the thanks I need here. First of all, I would like to thank Haris who has been my guide from start to finish, kindly sharing his Greek wisdom with me every day, for believing in me from my interview day and never letting me give up. Ευχαριστώ!

Second of all, I would like to thank John who has been such an amazing friend and who has shared all stages of the PhD experience with me, providing me with clever advice, a manly arm to cry on and the most amazing hugs!

And of course I have to say a massive THANK YOU to Will, without whom half of this thesis would never have happened, for being so very patient with me and providing me with constant IT support, smart advice and ideas for every occasion. And for teaching me about whisky and lots of random things and for being an amazing friend too.

I would also like to thank Lorna and Juraj who have explained to me ten times or more (each) how to use the 600 and archive my data and didn't kill me at any point for not remembering

again. Lorna for all the coffee breaks and the drinking sessions she's added to my days and for being so kind and supportive about everything. Juraj (and Zlata) for being so friendly and helpful, for all the Slivovica and for the after-hours support with my STD experiments.

Special thanks go to Dr Zuzana Kostalova who has been my carbohydrate mentor and an unforgettable friend – for the kind support and wise advice about both science and life, and for making work so much more fun again and again. ☺

Thank you to my dear student Theodora who has put so much effort and hard work into this project, never complaining about the smelly yeast I made her grow every day, and for her amazingly detailed lab book which has helped so much in writing this thesis. And for being such an awesome and supportive friend and never forgetting me.

I would love to thank all the awesome friends that I have made on the way that I don't really have more space to thank in more detail but who have been super supportive. Adam and Hannah – I love you guys, you have made this PhD so much more fun and made me laugh all the time. If only Adam stopped speaking Russian to me, I would thank him more. Nicholle, Elaine, Alan – thank you for making the office so colourful and fun to be in and for always being so helpful and nice to me. And, of course, thank you to Ari for being Ari – one and only incredibly fun personality, with the gasps and snorts and dancing around the office and laughing with me for hours without any reason – but also, for all the encouragement and positive energy you've given me every time I thought I wasn't going to make it.

I would also like to thank my boyfriend Jonny for believing in me, for persevering through all the stress I've put him through and for taking such a great care of me. I don't think I would ever finish this thesis if not for you (even though you have no idea what it's about).

But most of all, I would like to thank my parents (and my cat Moja-Moja), without whom I would never be here in the first place. Thank you for never losing faith in me and for providing me with your unconditional support on every step of the way. Thank you for always being there for me, for turning my tears into laughter and for being so convinced that I am amazing despite all the stupid mistakes I make.

If I have forgotten to mention anyone by name, I apologise. It's really amazing to have such a huge number of people support me through my PhD experience. Thank you all!!!

Contents

Declaration ii

Abstract iii

Lay summary v

Acknowledgements vi

Chapter 1: Introduction 1

1.1. Glycosaminoglycans 2

1. 1. 1. Classification of GAGs 3

1. 1. 1. 1. Heparin and heparan sulfate 5

1. 1. 1. 2. Chondroitin sulfate and dermatan sulfate 7

1. 1. 2. GAG-protein interactions 9

1. 2. Fucosylated chondroitin sulfate 11

1. 2. 1. Marine organisms as a source of novel polysaccharides 11

1. 2. 2. Biological properties of the sea cucumber extracts 12

1. 2. 3. Structure of the fucosylated chondroitin sulfate (fCS) 14

1.3. Selectins and their ligands 17

1. 3. 1. Members of selectin family 17

1. 3. 2. Selectin structure 19

1. 3. 2. 1. C-type lectin domain 21

1. 3. 2. 2. EGF-like domain 23

1. 3. 2. 3. SCR domains 25

1. 3. 3. Selectin Ligands 26

1. 3. 4. Existing models for selectin-ligand binding 31

1. 3. 4. 1. Selectins and 'catch' bonds 31

1. 3. 4. 2. Lectin-EGF domain interface 32

1. 3. 4. 3. Sliding-rebinding and allosteric binding models 34

1. 3. 5. Selectins and cancer 36

1. 3. 6. Selectin-fCS binding 36

Chapter 2: Aims 39

Chapter 3: Selectin expression tests 40
3. 1. Introduction 40
3. 2. Selectin construct expression in *P. pastoris* 41
3. 2. 1. Expression system 41
3. 2. 2. Construct design 42
3. 2. 3. Materials and methods 44
3. 2. 3. 1. Plasmid preparation 44
3. 2. 3. 2. Plasmid transformation into E. coli 44
3. 2. 3. 3. Extraction of the plasmid DNA 45
3. 2. 3. 4. Plasmid transformation into P. pastoris 45
3. 2. 3. 5. Small-scale protein expression tests 46
3. 2. 3. 6. Large-scale protein expression tests 46
3. 2. 3. 7. Fermentation conditions 47
3. 2. 3. 9. Large-scale protein purification tests 48
3. 2. 3. 10. SDS-PAGE analysis 48
3. 2. 3. 11. Western blot analysis 48
3. 2. 4. Results and discussion 49
3. 2. 4. 1. Construct design and plasmid preparation 49
3. 2. 4. 2. Small-scale protein expression tests 52
3. 2. 4. 3. Large-scale expression tests 54
3. 2. 4. 4. Small-scale protein purification tests on the LS2 construct 58
3. 2. 4. 5. Small-scale protein purification tests on the LS1 construct 60
3. 2. 4. 6. Fermenter growth of the LS1 construct 61
3. 2. 5. Conclusions 64
3. 3. Selectin construct expression in *E. coli* 65
3. 3. 1. Expression system 65
3. 3. 2. Construct design 66
3. 3. 3. Materials and methods 68
3. 3. 3. 1. Plasmid preparation 68
3. 3. 3. 2. Plasmid transformation into E. coli 68
3. 3. 3. 3. Protein expression tests 69
3. 3. 3. 4. Preparation of cell extracts and SDS-PAGE analysis 69
3. 3. 3. 5. Preparation of the inclusion bodies 70
3. 3. 3. 6. Solubilisation and purification of inclusion bodies 71
3. 3. 3. 7. Protein refolding tests by infinite dilution 71

3. 3. 3. 8. On-column protein refolding test ... 72
3. 3. 3. 9. TCA precipitation ... 72
3. 3. 4. Results and discussion ... 73
3. 3. 4. 1. Construct design and plasmid preparation ... 73
3. 3. 4. 2. Soluble protein expression tests ... 74
3. 3. 4. 3. Protein refolding tests ... 79
3. 3. 5. Conclusions ... 86
3. 4 Selectin construct expression in HEK293 cells ... 87
3. 4. 1. Expression system ... 88
3. 4. 2. Construct design ... 88
3. 4. 3. Materials and methods ... 90
3. 4. 3. 1. Fusion cloning of DNA constructs ... 90
3. 4. 3. 2. Transformation into E. coli and plasmid preparation ... 91
3. 4. 3. 3. Transfection into HEK293 cells ... 92
3. 4. 3. 4. Scale-up and purification ... 93
3. 4. 4. Results and discussion ... 94
3. 4. 4. 1. Construct design and plasmid preparation ... 94
3. 4. 4. 2. Small-scale expression tests in HEK293 cells ... 95
3. 4. 4. 2. Large-scale expression of the construct 20313 ... 97
3. 4. 5. Conclusions ... 99
Chapter 4: STD NMR studies of the fCS-selectin interactions ... 100
4.1. Introduction ... 100
4.2. Material and methods ... 103
4.3. Results and discussion ... 104
4. 3. 1. fCS binding to the L-selectin construct SELL-7019H ... 104
4. 3. 2. 2D STD-TOCSY experiment on fCS binding to L-selectin ... 109
4. 3. 3. fCS binding to the single-domain L-selectin construct ... 111
4. 3. 4. fCS binding to the P-selectin construct SELP-474H ... 112
4.4. Conclusions ... 113
Chapter 5: *In silico* studies of selectin-carbohydrate interactions ... 115
5.1. Introduction ... 115
5. 2 Materials and Methods ... 118
5. 2. 1. PDB file preparation for MD and docking studies ... 118

5. 2. 2. MD simulation set-up 118
5. 2. 3. Docking experiments 119
5. 2. 4. Free energy calculation 119
5. 3. Results and discussion 121
5. 3. 1. Sialyl Lewis X and its interaction with E-selectin 121
5. 3. 1. 1. Preparation of the sLe^x tetrasaccharide 121
5. 3. 1. 2. Docking experiments on sLe^x and E-selectin 125
5. 3. 1. 3. MD study of the X-ray structures of sLe^x-E-selectin complex 133
5. 3. 1. 3. Conclusion 138
5. 3. 2. Docking studies of fCS-selectin interactions 139
5. 3. 2. 1. Preparation of the fCS oligosaccharides 140
5. 3. 2. 2. Docking of the fCS 2,4S trisaccharide to P-selectin 141
5. 3. 2. 3. MD analysis of fCS 2,4S trisaccharide-P-selectin complexes 144
5. 3. 2. 4. Docking of the fCS 3,4S trisaccharide to P-selectin 149
5. 3. 2. 5. Docking of the fCS trisaccharide - conclusions 154
Chapter 6: Development of GFT NMR experiments for carbohydrate studies 157
6.1. Introduction 157
6.2. Material and methods 160
6. 2. 1. fCS sample preparation 160
6. 2. 2. NMR spectroscopy 160
6. 2. 3. Pulse sequence parameters 161
6.3. Results and discussion 162
6. 3. 1. Pulse sequence 162
6. 3. 2. (3,2)D $BIRD^{r,x}$-HSQC 163
6. 3. 2. (3,2)D $BIRD^{r,x}$-HSQC-TOCSY 165
6.3. Conclusions 171
Chapter 7: Summary and future directions 173
References 177
Appendix 1 188
Appendix 2 189
Appendix 3 190

Appendix 4 191
Appendix 5 192
Appendix 6 193
Appendix 7 194
Appendix 8 199
Appendix 9 205

Chapter 1: Introduction

Glycosylation is a language that cells use to communicate with each other. Glycans present on the cell surface allow cells to interact with the surrounding environment and to pass on important messages to their neighbours. Efficient communication is crucial for processes such as the inflammatory response, cell growth, differentiation, proliferation and homeostasis. Misregulation of these processes can in turn lead to development of disease states such as the chronic inflammatory disease or cancer.

Even though carbohydrates perform a multitude of substantial functions in the human organism, their importance is often overlooked. The field of structural biology is dominated by the studies of the structure and function of proteins and nucleic acids, yet our understanding of structural properties of proteoglycans and protein-carbohydrate complexes is still very limited. The main issue in glycobiology is the incredible complexity of sugar molecules, characterised by a huge number of possible combinations of monosaccharides and glycosidic linkages, as well as the limited knowledge about the processes involved in the regulation of carbohydrate synthesis. Studies of glycan structure and interactions are also hindered by the complexity of carbohydrate samples, heterogeneous and polydisperse in nature, which makes them difficult to study using spectroscopic techniques.

The major aim of this work is to gain a more in-depth understanding of the protein-carbohydrate interactions, focusing on the fucosylated chondroitin sulfate – a marine carbohydrate closely related to a group of mammalian glycans known as glycosaminoglycans – and its interactions with a group of cell adhesion proteins known as selectins. This chapter will introduce these biomolecules, describing their structure, function and binding properties.

1.1. Glycosaminoglycans

Glycosaminoglycans (GAGs) are complex carbohydrates found in the majority of animal species, both vertebrates and invertebrates. Interestingly, GAGs have never been found in plant species which indicates that they occurred on evolutionary timescale after the eukaryotes had divided into plant and animal kingdoms[1]. Carbohydrates structurally similar to GAGs are found in certain species of bacteria, however, it is believed that their evolutionary origin may be different to the animal GAGs. Structure of GAGs varies greatly amongst the different animal species, as well as among the different organ and tissues within a single organism, due to the extensive modifications present on the carbohydrate chains[1].

GAGs participate in a variety of physiological processes. Many of them are attached to the cell-surface proteins, forming proteoglycans which participate in recognition and signal transmission between different cell types. Each tissue produces its own set of GAGs on its surface that helps in regulation of tissue-specific contacts with proteins such as chemokines, cytokines, growth factors, morphogens, enzymes and adhesion molecules. Some types of GAGs are also secreted by the cells to the surrounding extracellular matrix (ECM) where they become part of the complex extracellular environment that regulates cell proliferation and differentiation. Regulation of the surface presentation and secretion of GAGs and proteoglycans moderates cell-cell communication by allowing the cells to receive signals and respond to the extracellular environment[2-4].

GAGs have been implicated in a variety of biological functions, such as the structural organisation of the ECM, control of homeostasis, cell adhesion, regulation of cell growth and differentiation[4]. Acting as co-receptors for growth factors, they play important role in cell proliferation and contribute to development of malignant cancer tissue and tumour progression[4]. By interacting with bacterial receptors, they can also take part in microbial invasion. Sulfated GAGs are associated with inflammatory response and are also a common constituent of amyloid plaques in neurodegenerative diseases[3]. These varied biological roles and the importance of GAGs in disease progression have attracted much interest from the research community and have served as an inspiration in the search of new GAG-based drug candidates.

1. 1. 1. Classification of GAGs

GAGs are linear polysaccharides, built of repeating disaccharide units consisting of a hexosamine – N-glucosamine (GlcN) or N-galactosamine (GalN) – and a hexuronic acid, which can be either glucuronic (GlcA) or iduronic acid (IdoA). The only exception to this rule is observed in keratan sulfate, where hexuronic acid is replaced by galactose (Gal). The individual subunits are connected to one another by glycosidic linkages which may involve atoms at different positions along the carbohydrate ring. The individual GAG chains also differ from each other in the pattern of substituents – hexuronic acid subunits can be decorated with sulfate groups and hexosamine subunits with either sulfate or acetyl groups. These features lead to a great structural diversity among the glycosaminoglycan family. Based on their monosaccharide composition, configuration of their glycosidic linkages and the amount and location of the sulfate substituents, GAGs are divided into six major groups: heparin, heparan sulfate (HS), chondroitin sulfate (CS), dermatan sulfate (DS), keratan sulfate (KS) and hyaluronan (HA). Table 1.1.1 summarises the main structural features of these GAG subclasses.

Table 1.1.1. Structural characteristics of different subclasses of GAGs[4].

GAG	Average MW (kDa)	Monosaccharide composition	Sulfates/ disaccharide
Heparin	5 – 30	L-IdoA, D-GlcA, D-GlcN	1.6 - 3.0
Heparan sulfate (HS)	30	L-IdoA, D-GlcA, D-GlcN	0.4 - 2.0
Chondroitin sulfate (CS)	5 – 50	D-GlcA, D-GalN	0.1 - 1.3
Dermatan sulfate (DS)	15 – 40	L-IdoA, D-GalN	1.0 - 3.0
Keratan sulfate	4 – 19	D-Gal, D-GlcN	0.9 - 1.8
Hyaluronan (HA)	5 – 20000	D-GlcA, D-GlcN	N/A

Even among the individual subclasses, GAGs exhibit high degree of variability as different regions of the polysaccharide chains can have different chemical composition. The distribution of sulfates along the chain is not uniform which leads to formation of negatively

charged 'patches' along the chain. In some GAGs, even the distribution of individual monosaccharide subunits can vary, for example in some regions GlcA is more prevalent than IdoA. This multitude of structural variants of GAG chains, their heterogeneity, as well as their different patterns of attachment to proteins in formation of proteoglycans, are the source of diverse structural motifs that can be utilised in interactions with other macromolecules. For example, the arrangement of sulfate groups along the GAG chains is believed to play an important role in their recognition by the cell surface proteins[5,6]. A brief overview of some of the main types of GAGs will be given in the following sections.

All GAGs, except for hyaluronan, are synthesised as O-linked proteoglycans, covalently attached to a specific serine residue of the core protein via a tetrasaccharide linkage GlcAβ1→3Galβ1→3 GlcAβ1→4Xylβ1→O-(Ser)[4]. Synthesis of the linkage region is catalysed by four glycosyltransferases: xylosyl transferase (XylT), β-4-galactosyltransferase (GalT-I), β-3-galactosyltransferase (GalT-II) and β-3-GlcA transferase (GlcAT-I). Once the synthesis of the linker is complete, the first hexosamine unit is added onto the chain by the action of another glycosyltransferase. Initially, all hexosamine units are N-acetylated. The first residue determines whether the chain matures into a HS-type (GlcNAc) or CS-type (GalNAc) GAG. The chain is then gradually extended by alternative addition of hexuronic acid and hexosamine residues. Glycosyltransferases involved in these processes vary, depending on the type of GAG chain being synthesised[6].

Once the synthesis of the sugar backbone is complete, modifications can be added by the action of deacetylases, sulfotransferases and epimerases. N-deacetylase-N-sulfotransferase is a large enzyme that cleaves off acetyl group from some of the hexosamine residues and attaches an N-sulfate groups instead[6]. The final chain is divided into sections with predominantly N-sulfated residues (NS domains), N-acetylated residues (NAc domains) or a mixture of both (NS/NAc mixed domains)[6]. Other sulfotransferases add O-sulfate groups onto specific positions along the carbohydrate chain. GlcA-C5 epimerase catalyses epimerisation of GlcA to IdoA in biosynthesis of heparin and dermatan sulfate[6]. Iduronic acid residues provide the carbohydrate chains with greater flexibility and are therefore important in facilitating interactions with proteins[2,7].

It is important to note here that GAG biosynthesis is a nontemplate-driven process, in contrast to synthesis of nucleic acids or proteins. For this reason, GAG biosynthesis is very challenging to study, control or recreate under laboratory conditions. Therefore, our knowledge of the processes involved in the regulation of GAG biosynthesis, including the production of different types of chains and the addition of specific modifications onto

carbohydrate backbone is still very limited. Several studies indicate that the type of GAG chain and the pattern of modifications can be influenced by the protein core to which the chain is attached[8]. However, it remains unclear how the regulatory information is transferred from the protein to the biosynthetic machinery.

1. 1. 1. 1. Heparin and heparan sulfate

Without a doubt, the most familiar GAG is heparin, which is a molecule well-recognised for its wide use in medicine as a potent anticoagulant agent. Heparin is built of repeating units of α-D-GlcN (1⟶4) linked to either one of the hexuronic acids, although in the majority of cases it is α-L-IdoA. Heparin is the most densely sulfated type of GAG, with sulfate substituents at various positions – most commonly the amine group or the O6 of GlcN or the O2 of IdoA. Occasionally, O3 of GlcN can also be sulfated. The amino group of GlcN can be either sulfated or acetylated[4] (Fig 1. 1. 1).

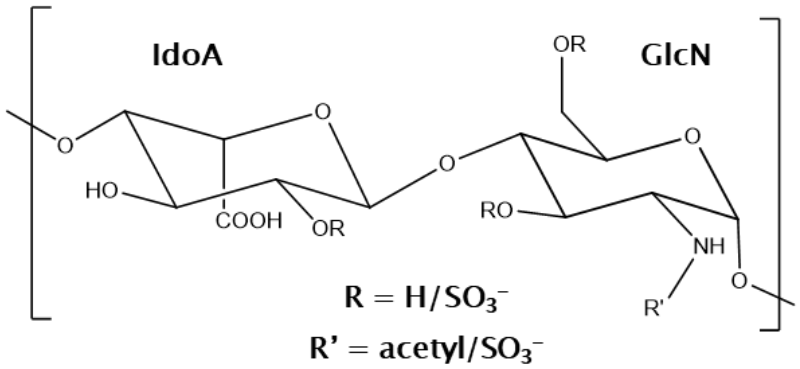


Fig. 1. 1. 1. Chemical structure of a typical heparin disaccharide unit. In over 90% of heparin units the hexuronic acid is IdoA. Heparan sulfate has higher proportion of GlcA subunits. R designates potential sites of sulfation, whilst R' – of sulfation or acetylation.

Pharmaceutically available heparin is derived from animal sources – it is found in the mast cells of certain organs, such as the liver, intestine or lungs. Its natural function is unclear but it is known to interact with histamine, proteases and inflammatory mediators, also located in the mast cell granules[9]. Heparin acts as an anticoagulant agent in several ways but its main mechanism of action is activation of antithrombin, a small protein which inhibits activity of thrombin – a coagulation cascade protease. Antithrombin binds with high affinity to a specific

pentasaccharide sequence within heparin via a set of basic amino acid residues[9]. Binding to heparin causes local conformational changes in antithrombin which in turn lead to higher affinity binding between antithrombin and thrombin. Following the heparin-induced conformational changes, antithrombin is able to covalently bind to thrombin, blocking its activity and stopping the coagulation cascade[9]. Several other clotting factors are also affected by the altered antithrombin which enhances the anticoagulatory effect of heparin even further[9].

Aside for its importance in preventing blood clotting, heparin also acts as an anti-inflammatory agent and has been shown to have beneficial effects for patients with conditions such as asthma, arthritis, inflammatory bowel disease, osteoporosis and cancer[5]. In treatment of such conditions, however, the anticoagulant activity of heparin becomes a problem as it may pose risks to patients. For this reason, efforts have been made to develop non-anticoagulant heparin derivatives and heparin-mimicking polyanions. Low molecular weight heparins (LMWHs), which are the heparin chains of size 2 to 10 kDa, display more predictable anticoagulant activity than the large heparin polysaccharides and can therefore be administered with less risk. Compared to heparin, they allow for more controlled treatment of thrombosis and have also been shown to inhibit angiogenesis better than the unfractionated heparin[5,10].

Both heparin and LMWHs have been reported to inhibit cell adhesion processes, blocking the recruitment of leukocytes, pathogens and migrating cancer cells to vascular endothelium. Studies show that heparin acts by binding to cell adhesion proteins, L- and P-selectins, competing with their natural ligands, and that the binding affinity can be regulated by introducing structural modifications onto heparin chains. For example, the elimination of sulfation at position O6 of GlcN dramatically reduced its inhibitory effect on tumour cell adhesion[11]. This inhibitory activity of heparins could have a potential use in treatment of inflammation and prevention of pathogenic invasion or cancer metastasis. Moreover, heparin binds to gp120 – an adhesion molecule present on the surface of HIV particles that plays an important role in the attachment and spread of the virus[5].

Heparan sulfate (HS) is closely related to heparin, the main differences being the density of the sulfate substituents along the carbohydrate chain, usually higher in the case of heparin, and a higher percentage of GlcA subunits in HS chains. HS is expressed in all types of cells, as well as in the ECM proteoglycans. Similarly to heparin, it interacts with a wide variety of proteins via its negatively charged sulfate groups. Acting like an antenna on the cell surface, HS mediates intercellular signals to proteins such as the growth factor receptors, bone

morphogenetic proteins, chemokines, interleukins, enzymes and apolipoproteins. Heparan sulfate interacts with complement regulators, helping to distinguish between the self and non-self cells. It is also believed to be an important participant in the pathogenic invasion, where it acts as a receptors for viral adhesion proteins. In addition to these roles, HS chains found in the ECM enhance structural integrity of tissues[5,12].

1. 1. 1. 2. Chondroitin sulfate and dermatan sulfate

Chondroitin sulfate (CS) is the most abundant GAG subclass in the human body. It is found in cartilage, tendons, ligaments and in the lining of blood vessels. It consists of β-D-galactosamine (GalNAc) and β-D-GlcA subunits, linked through alternating β (1⟶3) and β (1⟶4) linkages[4] (Fig. 1. 1. 2). There are four variants of CS, classified based on the position of sulfate groups on its subunits. CS-A and CS-C are monosulfated at positions C4 and C6 of GalNAc, respectively. CS-D is disulfated at position C6 of GalNAc, as well as at position C2 of GlcA. CS-E is disulfated at positions C4 and C6 of the GalNAc residue[12].

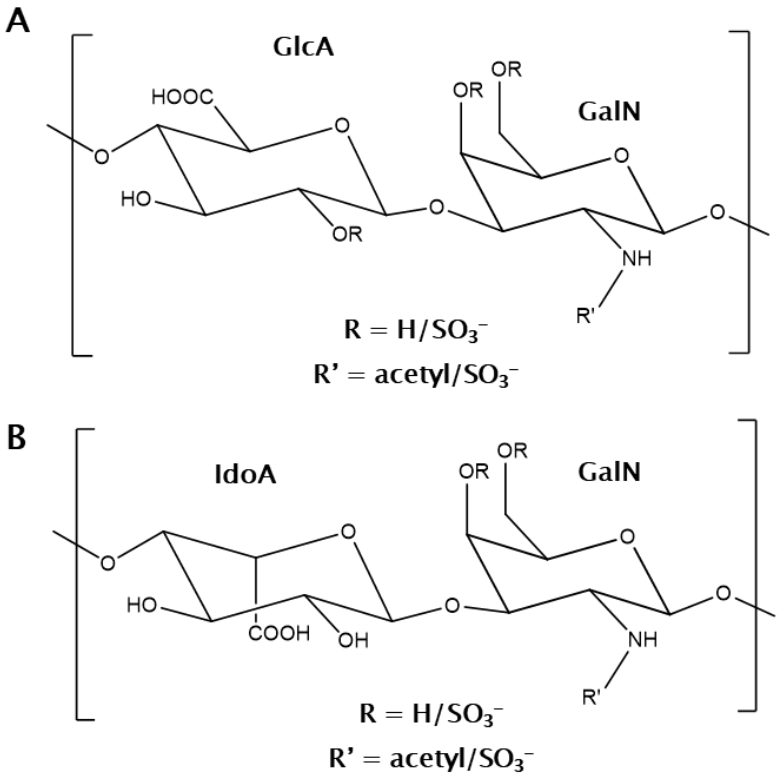


Fig. 1. 1. 2. Chemical structure of CS and DS disaccharides. A. Disaccharide subunit of chondroitin sulfate. **B**. Disaccharide subunit of dermatan sulfate. R designates potential sites of sulfation, whilst R' – of sulfation or acetylation.

Less abundant dermatan sulfate (DS) is found in the skin and also in the blood vessels and heart valves[3]. The main difference between CS and DS polysaccharides is the presence of GlcA instead of IdoA in the DS subunits (Fig. 1. 1. 2). Because of this high level of similarity, DS used to be classified as a subtype of CS (CS-B). In fact, it is not uncommon that these two types of GAGs are found within the same polysaccharide chain – such a chain would be referred to as a CS/DS chain[7]. Both sugars are characterised by the sulfation density lower than in the case of heparin and HS. As they are closely related and often occur together on the cell surface, it can be difficult to distinguish between the individual contributions of CS and DS to some of the cell functions.

Proteoglycans that contain CS have been shown to regulate many cellular processes, including proliferation, apoptosis, cell adhesion and migration. CS is a major component of the ECM of many connective tissues, including cartilage, bone, skin, ligaments and tendons. CS is responsible for many of the important biochemical properties of these tissues, providing them with high resistance and elasticity and is therefore considered as a potential therapeutic agent in treatment of osteoarthritis and joint-related pathologies[13]. CS-rich proteoglycans are also present in the ECM of the central nervous system, where they interact with the hyaluronan and the ECM proteins, contributing to the development and maintenance of the nervous tissue. There is evidence that chondroitin sulfates contribute to the plasticity of the nervous system and take part in the processes that underlie learning and memory[14]. CS proteoglycans have also been implicated in wound repair, by regulating cell adhesion, cell proliferation and cell migration. It is possible that different CS subtypes show different contributions to these processes and a specific combination of CS chains is required for correct wound healing[15]. It has also been shown that the CS sulfation pattern plays an important role in cancer progression – CS-E chains are overexpressed in many types of cancer and may be involved in the mediation of metastasis as they interact with the cell adhesion proteins such as selectins. Structural alterations in the CS chains have been identified in pancreatic, rectal and gastric carcinomas and certain subtypes of CS have been explored as pharmacological agents in the treatment of cancer[12,13].

Similarly to CS, DS is involved in the interactions with proteins of the ECM, the growth factors, chemokines and the pathogenic adhesion proteins. Dermatan sulfates are major constituents of the skin where they provide structural support to the tissue and assist in wound repair. In addition to that, DS-rich chains exhibit anticoagulant activity, although close to 50-fold less potent than their cousin heparin. However, its different mode of binding to proteins involved

in the coagulation cascade suggests that DS may be even more effective than heparin in preventing postoperative thrombosis[16].

The above examples highlight only some of the biological functions that GAGs are involved in. Even these few instances clearly illustrate how crucial these sugars are for the correct functioning of the human organism. It is not surprising that these compounds have attracted much interest as potential therapeutic agents. The following sections will introduce the novel type of GAG-like molecule obtained from the sea cucumbers and its interesting biological properties that could be useful in design of new drugs.

1. 1. 2. GAG–protein interactions

As described above, GAGs are involved in a variety of physiological processes thanks to their interactions with multiple types of proteins. Many of these interactions originate in the negative charge of the highly sulfated GAG chains which are able to form multiple electrostatic interactions with the positively charged residues on the protein surface. However, due to the fact that the nature of this interactions is electrostatic, as well as the flexibility of carbohydrate ligands and the fact that GAG-binding proteins often lack a clearly defined binding site, GAG-protein interactions tend to be fairly weak, with milimolar binding affinities.

Initial studies of GAG-protein interactions focused on the protein sequence analysis and aimed to find common traits between the heparin-binding domains of different proteins. It was proposed that such domains are characterised by an array of basic residues, often following a very similar pattern. Modelling studies suggested that these residues would be located on the same face of any secondary structure, forming a positively charged cluster that could interact with the sulfated sugars[17]. Since then, many X-ray structures of GAG-protein complexes have been obtained, largely confirming the initial hypothesis. Patches of positively charged residues – lysine, arginine or histidine – are a common feature of the sugar-binding sites, typically accompanied by other polar residues that participate in hydrogen bonding interactions with the hydroxyl and sulfate groups. Tyrosine, glutamine and asparagine are examples of residues commonly involved in such interactions[2,18].

Many previous studies have focused on the heparin binding proteins such as the fibroblast growth factor (FGF), heparinases or antithrombin. As these proteins vary greatly in structure and function, there are some major differences between the heparin-binding domains. However, all of them share the same feature – a positively charged cluster of basic residues that forms the basis for protein's interaction with GAGs. Such cluster may not necessarily be formed by residues that are sequentially close to one another. Often, basic residues are brought together from different regions of the sequence and it is the protein's tertiary structure that determines its affinity for GAGs. One such example would be the binding site of antithrombin III, where the residues located far apart from each other in sequence are brought together in the folded protein to form a heparin binding site (Fig 1. 1. 3.)[2].

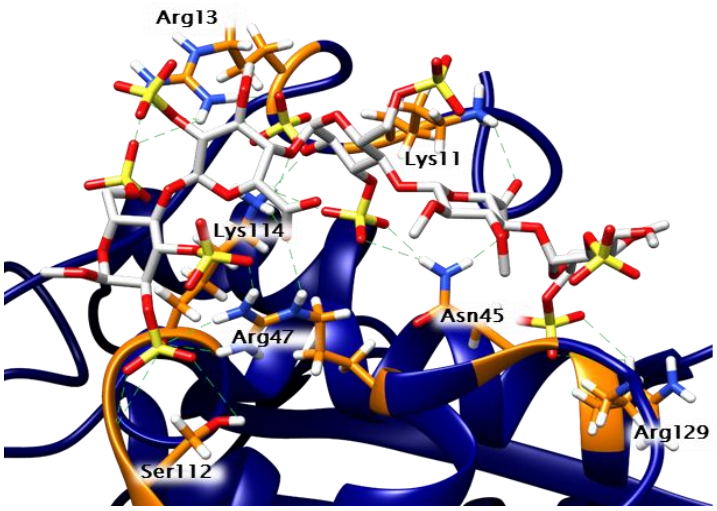


Fig. 1. 1. 3. Synthetic heparin pentasaccharide bound to antithrombin III[2]. Ribbon representation of the antithrombin III heparin binding site, highlighting the basic residues that support electrostatic interactions with the negatively charged oligosaccharide. The protein is shown in blue, the heparin pentasaccharide is shown in light grey and the interacting residues are shown in orange.

The protein-binding behaviour of GAGs is similar to that of nucleic acids and is driven by the so-called polyelectrolyte effect[18]. In physiological conditions, these long negatively charged chains are surrounded by the positively charged cations. Upon binding to protein, positively charged side chains of the protein residues displace these cations and the GAG-protein interaction is therefore entropically favoured[18].

Some studies have also proposed that there are specific sequences within GAGs that may contribute to protein recognition. Disulfated disaccharide sequences found in DS chain have been shown to bind to heparin cofactor II, whilst highly sulfated sequences in heparan sulfate

and heparin – to growth factor proteins such as fibroblast growth factor (FGF). All of the specific protein-binding sequences within GAGs have been shown to contain at least one iduronic acid residue and it is believed that the conformational flexibility of this residue plays an important role in GAG-protein interaction[17,18].

1. 2. Fucosylated chondroitin sulfate

1. 2. 1. Marine organisms as a source of novel polysaccharides

In the recent years there has been much interest in the analysis of the bioactive compounds from marine organisms. Because of the necessity to adjust to the aquatic environment, the underwater creatures had developed specific features different or even completely absent from those found in the land organisms. The necessity to adapt to very harsh living conditions such as the low sunlight and high pressure levels, low temperatures, high salt concentration and low nutrient content, has led these organisms to development of new structural and chemical mechanisms that provide their bodies with ability to adjust to the difficult environment that they live in.

The main sources for heparin at the moment are meat by-products, such as bovine lungs or porcine intestines. Production of heparin from such sources is expensive and not very efficient, and there have been incidents of heparin contamination with other polysaccharides. For these reasons, finding an alternative source of heparin and other GAG polysaccharides is highly desirable. GAG-like polysaccharides are often found in the extracts from tissues of marine organisms. For example, heparin-like polysaccharides have been prepared from certain species of marine bacteria and algae. As these organisms are easy and cheap to culture, they attract an interest as a means for production of GAG-like polysaccharide on an industrial scale. Indeed, such approach has been already used in production of food stabilisers such as xanthan and gellan[19]. Moreover, marine polysaccharides are often suitable as starting materials for production of heparin-like drugs. Structural derivatives of marine carbohydrates can allow to design compounds with increased activity and specificity. Marine

organisms have therefore attracted much commercial interest as a potential cheap alternative source of heparin and heparin mimetics[19]. Similarly, many marine organisms are rich in chondroitin sulfate and dermatan sulfate as they use these polysaccharides to built their cartilaginous skeletons and body walls[20]. Since many marine tissues are currently discarded as waste-products of the fishing industry, some marine GAG-like carbohydrates could be obtained more cheaply than similar mammalian products.

Although structurally similar to mammalian GAGs, polysaccharides from sea organisms can have distinct structural features and these differences are a source of the unique properties of these sugars[20,21]. The CS and DS chains found in marine organisms often differ from their mammalian equivalents in sulfation patterns[22–24]. Unlike the mammalian GAGs which are always linear, marine polysaccharides can have branches, as in the case of the fucosylated chondroitin sulfate (fCS), a sea cucumber polysaccharide which has been the main interest of this work[25,26]. The following sections will introduce the fCS, its structure, biological properties and its potential to be used as a therapeutic agent.

1. 2. 2. Biological properties of the sea cucumber extracts

Sea cucumbers (*Holothuroidea*) are animals belonging to a phylum of echinoderms, alongside the organisms such as star fish and sea urchins. When translated from Latin, *Echinodermata* means 'spiny skin' – this name refers to the endoskeleton of these animals, composed of hard, calcium-rich plates. Sea cucumbers are extraordinary members of this family, unique in both their body structure and behaviour. Although their body has a five-fold radial symmetry, it is not as obvious as in other echinoderms because of their elongated shape. As their endoskeletal plates are microscopic in size, the sea cucumber body is soft and gelatinous, protected only by a thick skin, sometimes decorated with spikes that give it the familiar look of a well-known vegetable. (Fig. 1. 2. 1). They are characterised by an unusual defence mechanism – they are able to eject their internal organs in the direction of approaching predators and entangle them, blocking the attack, as well as to regrow the organs lost or damaged during the fight.

Fig. 1. 2. 1. Sea cucumbers are a culinary delicacy in some regions of Asia. Although at first glance they may look like ordinary cucumbers, these are the sea organisms that keep attracting scientific interest.

There are over 1500 species of sea cucumbers world-wide and many of them are able to survive in the extreme conditions of the deepest ocean trenches. Their curious properties have caused them to be a sought-for delicacy in some regions of the world – they are a popular food source and a traditional remedy in some Asian countries, such as China or Malaysia. In these countries sea cucumber products are a home remedy for many medical problems, ranging from cuts and burns to more serious conditions like rheumatism, hypertension or impotence[27]. In recent years, pharmacological properties of sea cucumber extracts have been studied and it has been confirmed that the body wall of these organisms contains compounds that can stimulate wound healing[28], inhibit inflammatory processes[29] and act as anticoagulants or antithrombotics[30–39]. They have also been shown to have anticancer properties, by reducing angiogenesis and metastasis[40–44], as well as anti-HIV properties by interfering with viral adhesion processes[45–48].

The body wall of the sea cucumbers consists of collagen fibres, surrounded by a matrix of polysaccharides. A large proportion of these polysaccharides consists of GAG-like compounds closely related to chondroitin sulfate. As mammalian CS is involved in a range of biological functions, it is likely that its derivatives, present in large quantities in the sea cucumber extracts, are the source of unique biological properties of these extracts. As an example, sulfated glycosaminoglycans extracted from the body wall of the sea cucumber species *Stichopus hermani* has been demonstrated to accelerate wound healing in rats and to

stimulate growth of mesenchymal stem cells which are self-renewing cells that differentiate into cells of connective tissues[49]. Therefore, it is likely that the regenerative and wound-healing properties of the sea cucumber extracts are brought about by the presence of the sulfated GAG-like carbohydrates such as fCS.

Low molecular weight fCS chains from the sea cucumber species *Thelenota ananas* was shown to inhibit replication of HIV, reducing infection by clinical isolates of the virus[49]. fCS was found to bind to viral envelope glycoprotein gp120 with milimolar affinity and further studies suggested that the anti-HIV activity of fCS is dependent on the presence of sulfated fucose branches, similarly to other fCS activities[50]. Development of fCS derivatives allows to fine-tune some of the carbohydrate properties to upregulate the anti-HIV activity and downregulate the anticoagulant properties. Therefore, it is speculated that fCS could be developed into a novel HIV inhibitor[49].

Of special interest for this project are the anti-metastatic properties of the sea cucumber extracts, also attributed to fCS polysaccharides. fCS samples from the sea cucumber species *Ludwigothurea grisea* demonstrated inhibitory activity towards adenocarcinoma MC-38 cells in experimental model for lung colonisation in mice[50]. The same samples were also able to inhibit neutrophil recruitment in two models of inflammation. These biological activities were associated with high binding affinity for L- and P-selectins[50], a topic which will be explored in more detail in further sections of this thesis.

1. 2. 3. Structure of the fucosylated chondroitin sulfate (fCS)

Fucosylated chondroitin sulfate (fCS) was first described in 1988 by Vieira and Mourão in their study of chemical composition of polysaccharide fractions extracted from the sea cucumber *Ludwigothurea grisea*[51]. The study reported that the largest polysaccharide fraction was composed of equimolar quantities of α-L-fucose, β-D-glucuronic acid and N-acetyl β-D-galactosamine. Further examination revealed that this new type of polysaccharide is a chondroitin sulfate with fucose branches attached to glucuronic acid and that the presence of such branches blocks the access of chondroitinases normally used to depolymerise CS chains[51]. NMR analysis of this compound allowed to establish that the fucose branches are sulfated and are attached to position O3 of glucuronic acid[25].

These findings were soon followed by the reports of similar compounds found in other sea cucumber species. Analysis of fucose chemical shifts from *Stichopus japonicus* helped to determine that there are three types of fucose present in the fCS sample, differing from each other in the sulfation pattern. In this species, the most dominant form of fucose residues (>55%) was 2,4-sulfated, slightly less (>33%) was 3,4-sulfated and the smallest fraction (>11%) was 4-sulfated[26]. Recent studies, comparing structures and properties of fCS from different sea cucumber species, have found that the distribution of these different fucose forms can vary between the different species and that it is correlated to the anticoagulant potential of different fCS samples[36,50,52–54]. It was generally observed that the fCS samples characterised by a higher proportion of 2,4-sulfated fucose showed higher anticoagulant activity[50].

Based on the above findings, a single trisaccharide building block of the fCS can be described as: ⟶3)GalNAcβ(1⟶4)[Fucα(1⟶3)]GlcAβ(1⟶, where [Fuc(1→3)] denotes fucose branch from position O3 of glucuronic acid, with potential sulfation sites at positions O4 and O6 of galactosamine and O2, O3 and O4 of fucose (Fig. 1. 2. 2.).

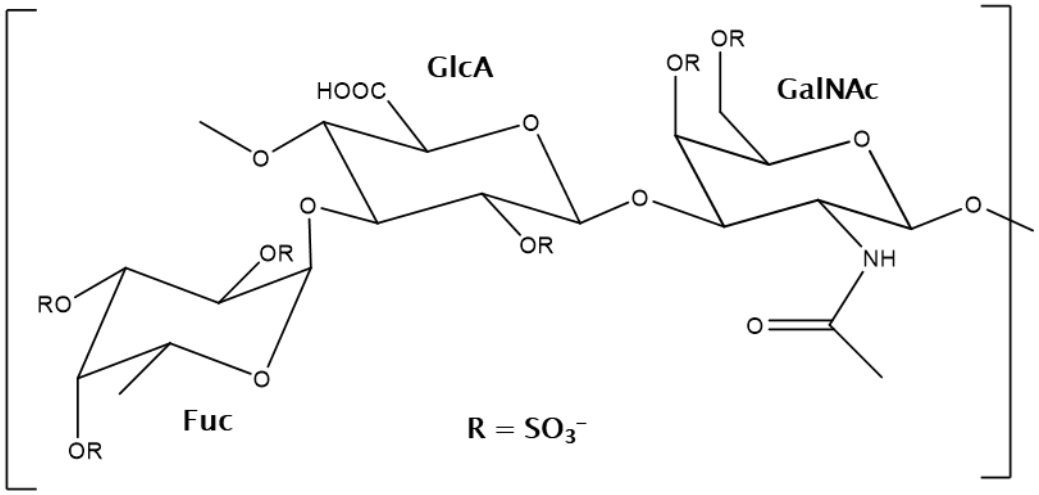


Fig. 1. 2. 2. Chemical structure of a repeating trisaccharide unit of the fCS. Fucosylated chondroitin sulfate (fCS) is built of a CS disaccharide with a sulfated fucose branch attached at position O3 of glucuronic acid. R designates potential sites of sulfation.

Previous studies in the Uhrín group have allowed to elucidate the structure of the fCS from a sea cucumber species *Holothuria forskali*. The structure of this GAG resembles that of the fCS polysaccharides from other sea cucumber species, presented in Fig. 1. 2. 2. Analysis of the

NMR chemical shifts indicated that galactosamine units are always disulfated at positions O4 and O6. The sulfation pattern on fucose residues is more variable – approximately 46% of fucose residues are disulfated at positions O3 and O4, 39% at positions O2 and O4 and 15% are monosulfated at position O4[55]. In addition to these results, some interesting information about the fCS conformation was obtained: fucose H-5 was found to have an unusually high chemical shift (4.79-4.89 ppm), compared to that of free fucose (~4.2 ppm). The 2D ^{1}H, ^{13}C NOESY-HSQC spectrum showed strong NOEs between the protons of fucose and galactosamine residues, which indicates that these monosaccharides are in close proximity to one another. These observations indicate that fucose residue may be stack on top of galactosamine and that this stacking interaction is independent of the fucose sulfation pattern[55].

As a result of the presence of sulfated fucose branches, fCS has more charged groups per unit than a standard CS chain. This property could have implications for the protein-binding behaviour of fCS and its resulting biological activity. Previous studies have shown that fCS has approximately five-fold higher affinity for Ca^{2+} ions than mammalian CS chains[56]. As some of the carbohydrate binding proteins rely on Ca^{2+} ions in their interactions with sugars, this could suggest that fCS binds such proteins with higher affinity than human GAGs.

fCS also shares structural similarities with some of the carbohydrate determinants recognised by the proteins involved in cell adhesion processes. As cell adhesion is a crucial step in the processes such as inflammation, blood coagulation, pathogenic adhesion and cancer metastasis, there is a strong indication that the fCS could act as a competitive inhibitor to some of the natural ligands of these proteins and, by doing so, disrupt their function. This topic will be discussed in more detail in further sections, focusing on a family of cell adhesion proteins known as selectins.

1.3. Selectins and their ligands

This section will describe a family of cell adhesion glycoproteins known as selectins. These proteins are known for their ability to mediate cell rolling – the earliest response to inflammatory stimuli or tissue damage at cellular level. Cell rolling is likely also an important step in attachment of viral particles and tumour cells as they spread to new tissues. Importance of selectins in the above processes makes them an exciting target in development of novel anti-inflammatory, antiviral or antimetastatic drugs.

1. 3. 1. Members of selectin family

Selectin family consists of three members, L-, P- and E-selectin. All selectins mediate cell rolling, however, the three proteins are expressed in different cell types, have different ligand preferences and perform different functions during the leukocyte recruitment process.

L-selectin is constitutively expressed on the surface of leukocytes, playing central role in formation of initial tethers between leukocytes and vascular endothelium at sites of inflammation. It also mediates the attachment of naïve lymphocytes to high endothelial venules (HEV) in secondary lymphoid tissues, where they are able to encounter antigens and undergo activation[57,58]. L-selectin is rapidly lost from the surface of leukocytes following their activation, forming a soluble protein, in the process known as shedding. It is believed that soluble form of L-selectin can compete with the cell-surface protein for ligand binding which is likely to cause dampening of inflammatory response[58,59].

E-selectin is expressed on the surface of vascular endothelium. However, its expression is inducible at the transcription level, meaning that the protein will be expressed only in presence of inflammatory stimuli, such as interleukin-1 (IL-1), tumour necrosis factor-α (TNF-α) or bacterial lipopolysaccharide (LPS), and that it will be transported to cell membrane after several hours. Therefore, this protein does not participate in immediate inflammatory response[60,61]. E-selectin was long thought to merely enhance P-selectin's function. However, it is now known that E-selectin has its own unique properties and

independent role. This selectin is capable of mediating slow rolling of neutrophils which facilitates cell arrest and transmigration to the sites of injury – processes crucial for correct immune response[62]. Recent studies have shown that it is also involved in trafficking of hematopoietic stem cells to the bone marrow[63].

P-selectin is expressed in platelets, as well as in endothelium. Similarly to E-selectin, its expression on the cell surface is also inducible by a variety of chemical stimuli, such as thrombin, histamine, complement fragments or adenine diphosphate (ADP). However, in contrast to E-selectin, P-selectin is stored in the granules inside the cells, ready to be released to the membrane within minutes from activation, which allows for a much quicker response than in the case of E-selectin[61]. P-selectin is mostly involved in the initial stages of leukocyte rolling. However, P-selectin and E-selectin are often coexpressed in endothelial cells and complement each other's function. A single gene knockout of either E- or P-selectin does not impair leukocyte recruitment – only a double knockout of those two genes will cause serious defects in cell rolling[57,61].

The summary of selectin expression sites and their individual roles in leukocyte recruitment cascade is presented in Table 1. 3. 1.. As P- and E-selectins are located on the endothelium, it is likely that they are also involved in adhesion of other cell types, including platelets, pathogens and tumour cells.

Table 1. 3. 1. Comparison of selectin expression sites and their roles in the leukocyte recruitment cascade.

	Expression site	Function
L–selectin	Lymphocytes	Lymphocyte entry into secondary lymphoid organs.
	Neutrophils	Initial attachment and cell rolling on endothelium at sites of inflammation or tissue injury.
E–selectin	Endothelium	Slow rolling of leukocytes at later stages of inflammatory response.
P–selectin	Platelets	Leukocyte-platelet interactions in wound healing and homeostasis
	Endothelium	Early leukocyte rolling in inflammatory response.

1. 3. 2. Selectin structure

Despite the differences in their individual roles, the three proteins are highly similar in both their sequence and structure. Selectins are type I transmembrane glycoproteins composed of five domain types. Starting at the N-terminus, selectins are built of a Ca^{2+}-dependent (C-type) lectin domain, an epidermal growth factor (EGF)-like domain, from two to nine short consensus repeat (SCR) domains, a membrane spanning region and a C-terminal cytoplasmic tail[61] (Fig. 1. 3. 1.). As the number of SCR domains is different for each selectin – there are two SCR domains in human L-selectin, six in E-selectin and nine in P-selectin[64] – the three proteins extend to different heights above the cell surface which is likely to contribute to recognition of specific ligands and help in regulation of cell rolling.

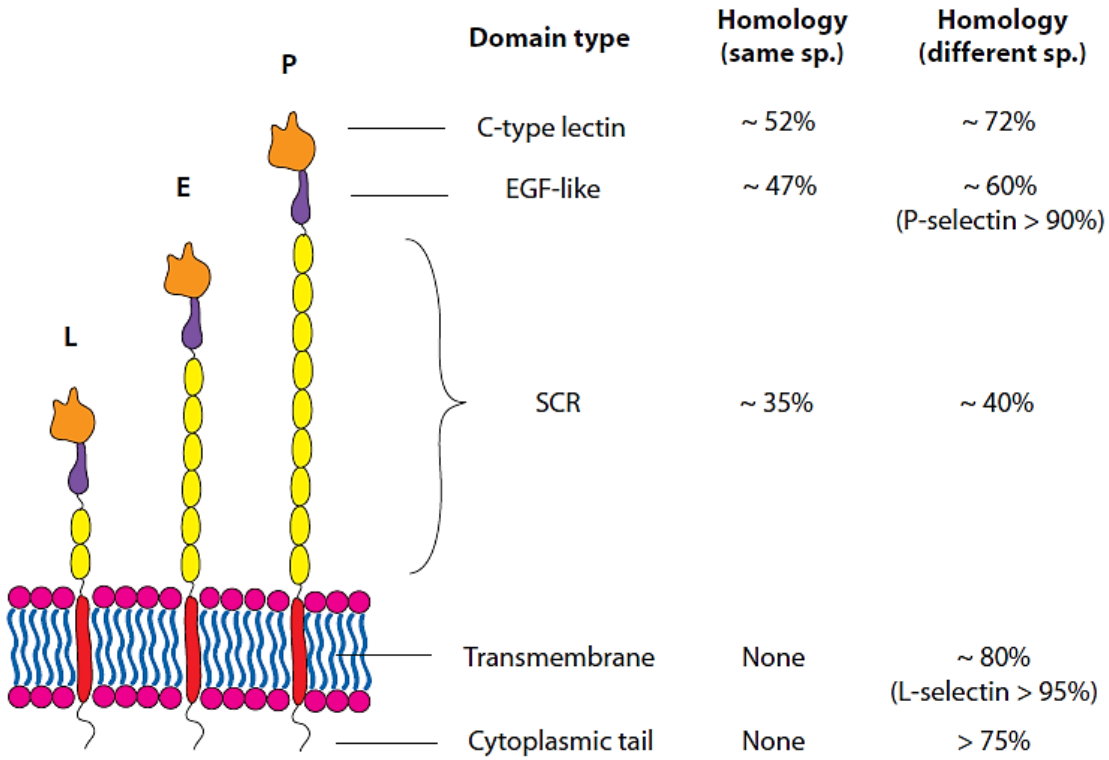


Fig. 1. 3. 1. Schematic representation of selectin structure and domain homologies. Individual domains are represented with different colours – orange for the C-type lectin domain, purple for the EGF-like domain, yellow for the SCR domains and red for the transmembrane domain. The cytoplasmic domain is a short peptide tail, consisting of less than 20 residues. The right hand side of the figure includes information on average percentage homology between the individual domains of different selectins from the same species, as well as between the domains of the same selectin from different species, based on the sequences for human, mouse and cow[61].

The extracellular domains display high degree of sequence similarity among selectins of the same species, as well as when comparing human selectins to selectins from different species, such as a mouse or a cow[61] (Fig. 1. 3. 1.). High sequence conservation observed for the C-type lectin and the EGF-like domains suggests their importance for correct protein function. Interestingly, although the C-type lectin, EGF and SCR domains occur in many proteins, no known proteins other than selectins combine these three domains types together. This could indicate that this particular structural arrangement is key to selectin function[61].

Because of the large size and transmembrane nature of these proteins, their full-size structure has not been solved as yet. However, the individual domains and their combinations have been studied by X-ray crystallography, NMR and molecular dynamics. Summary of all selectin domain structures published to date is presented in Table 1. 3. 2..

Table 1. 3. 2. List of selectin domain structures published to date.

Selectin	Construct	Ligand	Technique	Reference	PDB code
E	Lectin + EGF	N/A	X-ray	Graves *et al.*[65] (1994)	1ESL
	Lectin + EGF	sLe^x	X-ray	Somers *et al.*[66] (2000)	1G1Q
	Lectin + EGF + 2xSCR	sLe^x	X-ray	Preston *et al.*[67] (2016)	4CSY
	Lectin + EGF + 2xSCR	glycomimetic	X-ray		4C16
P	EGF	N/A	NMR	Freedman *et al.*[68] (1996)	1FSB
	Lectin + EGF	N/A	X-ray	Somers *et al.*[66] (2000)	1G1T
	Lectin + EGF	sLe^x	X-ray		1G1R
	Lectin + EGF	PSGL-1	X-ray		1G1S
L	Lectin + EGF	glycomimetic	X-ray	Mehta *et al.*[69] (2017)	3CFW

1. 3. 2. 1. C-type lectin domain

Many proteins of the immune system belong to the family of lectins. There are at least twelve structural subtypes of lectins, one of them being the C-type lectin, where the term 'C-type' refers to the calcium-dependent activity manifested by this domain. The C-type lectin domain occurs frequently in virtually all species of the animal kingdom, has been identified in over a thousand of proteins and thoroughly reviewed in the literature[70–72].

A distinguishing feature of the C-type lectin domain is the presence of a Ca^{2+}-coordination site which participates in binding of carbohydrate ligands. The overall domain fold is well conserved among the different proteins, despite its high sequence variability. The C-type lectin domain is usually between 110 – 130 amino acids long, with the N- and C-terminal regions aligned in a two-stranded antiparallel β-sheet. These regions are connected by a series of loops, two α-helices and antiparallel β-sheets[72].

This general description fits well with the published structures of selectin domains. The domains are 121 amino acids long with highly conserved N- and C-termini that fold over each other into an antiparallel β-sheet. The sequence if the N-terminus is well conserved between P- and L-selectins, however, several changes in that region are observed in E-selectin. Tyrosine at position 2 is replaced by serine, histidine 4 – by asparagine and tyrosine 5 – by threonine. Nevertheless, the N-terminus maintains the same fold in E-selectin as in the other two selectins. There are two α-helices flanking two opposite sides of the domain, between the residues 17-27 and 32-41, and multiple antiparallel β-strands in the centre of the domain (Fig. 1. 3. 2. A). The Ca^{2+}-coordination sphere is well conserved among selectins, in all three proteins consisting of four amino acids, Glu80, Asn82, Asn105 and Asp106 (Fig. 1. 3. 2. B)[66]. There are two disulfide bridges within the C-type lectin domain, conserved among all selectins. One of them is the disulfide bridge between the residues 19 and 117 that pins the two domain termini together. The second disulfide bridge is located close to the Ca^{2+}-coordination site, between the residues 90 and 109.

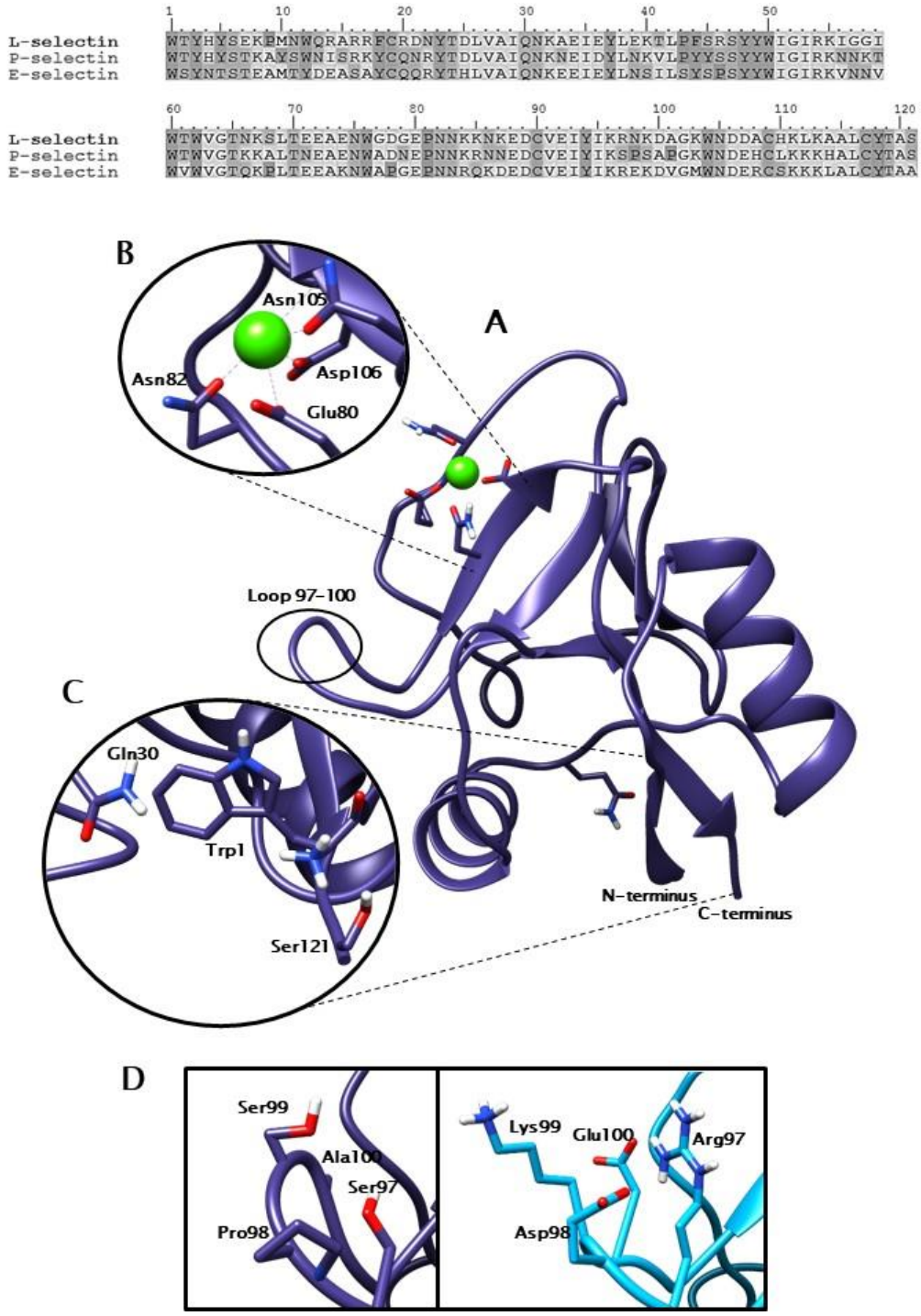


Fig. 1. 3. 2. X-ray structure of the C-type lectin domain from P-selectin[66]**.** Top of the figure shows a multiple sequence alignment of the C-type lectin domain amino acid sequences from human selectins. **A**. Ribbon representation of the overall structure of the C-type lectin domain. Calcium ion is shown in green. **B.** Ca^{2+}-coordination sphere, highlighting the contacts made by residues Glu80, Asn82, Asn105 and Asp106. These four residues are conserved amongst all human selectins. **C.** Close-up of the region where N- and C-termini meet – this region is the only site of interdomain interactions between the lectin and EGF-like domains. **D.** Close-up on the loop 97-100 (on the left) and the corresponding region in E-selectin. The differences in this region are likely to contribute to specific ligand binding properties of individual selectins.

Aside of the carbohydrate-binding region, the point of particular interest is the link between the C-type lectin and the EGF-like domains, closely associated with the protein's N-terminus (Fig. 1. 3. 2. C). There are very few noncovalent interactions between the lectin and the EGF-like domains and most of them are conserved among the three selectins.[73] The N-terminal Trp1 residue plays an important role in these interactions as a triple hydrogen bond donor and a hydrogen bond acceptor via its backbone oxygen atom. Another two residues worth noting here are Gln30, involved in the interdomain interactions via contacts with Glu135, and Ser121 that seems to play an important role in stabilising the interdomain hydrogen bond network. The latter residue is conserved only between L- and P-selectins – in E-selectin it is substituted with Ala121, suggesting that the interdomain interactions differ between the three proteins.

There are other differences between the lectin domains that may account for some of the specific properties of each individual protein. One region that is especially worth noting is the loop between the residues 97 – 100 (Fig. 1. 3. 2 D). Residues found in this loop in E- and L-selectins are polar amino acids with long side chains reaching towards the binding side. In P-selectin these residues are replaced with either non-polar alanine and proline, or polar but much shorter serine. Presence of additional charged residues in this loop in E- and L-selectins may facilitate interactions with some of the carbohydrate ligands[66].

There may be other regions of the C-type lectin domain that contribute to the protein's ligand binding affinity by mediating allosteric changes throughout the domain. Previous studies have proposed that one such region is a loop between the residues 83-89. This region has been shown to adopt an alternative conformation in some of the X-ray studies and has been suggested to enhance the ligand binding affinity of selectins via allosteric changes in domain conformation[66,67,69,74]. This topic will be discussed in more detail later in this work in the light of presented results of computational studies on selectin-ligand binding.

1. 3. 2. 2. EGF-like domain

The EGF-like domains are found in many extracellular proteins, involved in a variety of biological functions such as the blood coagulation, complement system function, embryonic

development and, as in case of selectins, cell adhesion. They occur in a variety of domain combinations, often found in multiple copies, although their biological role is still unsure[75].

The EGF-like domains are fairly small and compact, in general between 40 to 50 amino acid long. Although their sequence and conformation can vary between the different proteins, all of them feature three disulfide bridges with a signature connectivity between residues 122 and 123, 127 and 142, as well as 144 and 153 (Fig. 1. 3. 3.)[68]. Selectin EGF-like domains are characterised by a series of loops and turns in the N-terminal region, with a two-stranded antiparallel β-sheet in the centre (Fig. 1. 3. 3.). The region between residues 135 – 138 is the only region of the EGF-like domain involved in interactions with the lectin domain. This region is quite well conserved among selectins, with the exception of Thr136 which is replaced with Ile136 in L-selectin and Asn138, replaced with Gly138 in P-selectin.

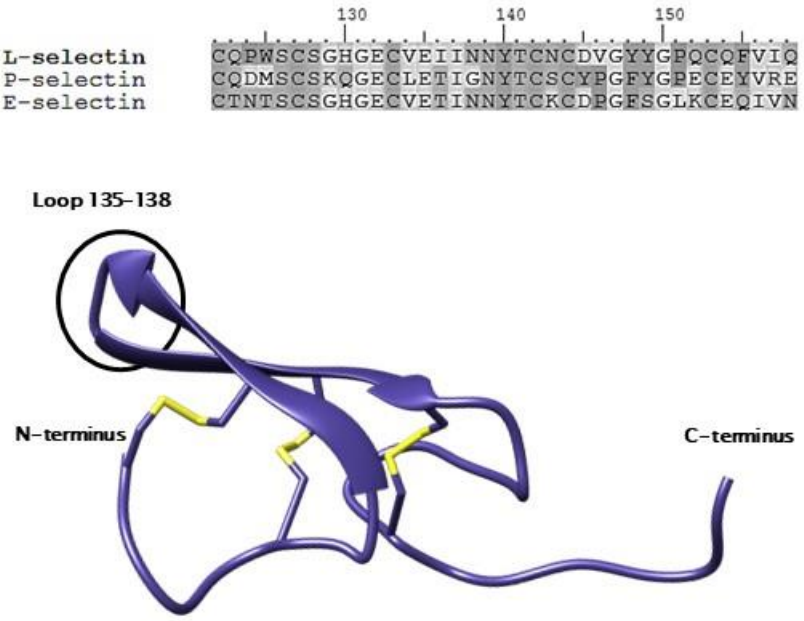


Fig. 1. 3. 3. X-ray structure of the EGF-like domain from P-selectin[66]. Top of the figure shows a multiple sequence alignment of the EGF-like domain amino acid sequences from human selectins. **A**. Ribbon representation of the overall structure of the EGF-like domain with highlighted N- and C-termini and the region between residues 135-138, involved in interdomain interactions with the C-type lectin domain. Disulfide bonds are marked in yellow.

1. 3. 2. 3. SCR domains

Selectins contain between two to nine short consensus repeat (SCR) domains, otherwise known as the Sushi domains. Only two of selectin crystal structures published so far contain the SCR domains – the two structures produced in the study by Preston *et al.* – 4CSY and 4C16[67]. Although the sequence conservation among selectin SCR domains is low, they all consist of 60 residues arranged into a compact β-sheet fold with conserved three disulfide bonds, one at each end of the domain and one in the centre[67] (Fig. 1. 3. 4). Each SCR domain contains a conserved tryptophan residue which allows for formation of a tightly packed rigid core in the centre of the domain. The neighbouring SCR domains are connected by a linker containing a pattern of four hydrophilic-hydrophobic-hydrophobic-hydrophilic residues.

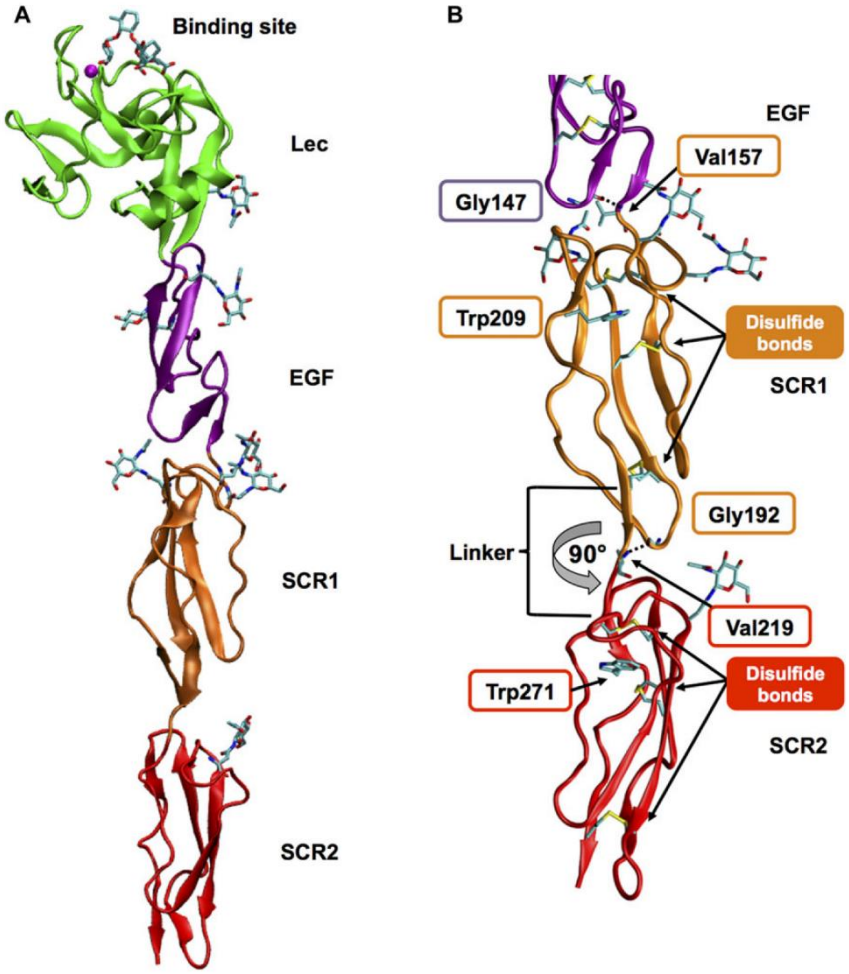


Fig. 1. 3. 4. X-ray structure of the lectin, EGF-like and two SCR domains of E-selectin[67]. **A.** In this structure (4C16), E-selectin was co-crystallised with a glycomimetic structurally resembling sLe^x. Two SCR domains align linearly with the EGF-like domain. **B**. Detailed view of the first two SCR domains, highlighting the disulfide bonds, conserved tryptophans and interdomain interactions. Figure taken from Preston *et al.*[67]

Not much is known about the behaviour of the two E-selectin SCR domains upon ligand binding. In the two published structures, both of which were obtained in the presence of a ligand, these domains align in a linear fashion with the EGF-like domain. 4CSY structure allowed to establish that, upon binding to sLe^x, two consecutive SCR domains are rotated by 90° relative to each other due to a stabilising interdomain interaction between the interdomain linker (residue Gly192) and SCR domain core (residue Val219) (Fig. 1. 3. 4. B). A similar stabilising interaction is also observed between the EGF-like domain (Gly147) and the interdomain linker to the first SCR domain (Val157) in E-selectin (Fig. 1. 3. 4. B)[67].

1. 3. 3. Selectin Ligands

Selectins bind to carbohydrate determinants on glycoproteins. Although the three proteins share some of their ligands, they differ from each other in their binding affinities and often exhibit different binding preferences. This section will introduce the major selectin ligands, focusing on the interactions that have been evaluated by X-ray crystallography.

All three selectins recognise a common carbohydrate determinant, sialyl $Lewis^X$ (sLe^x). Although it is a low-affinity interaction, sLe^x has been identified as a minimal recognition unit necessary to enable cell adhesion under flow for all selectins[76]. sLe^x is a tetrasaccharide of the following formula: NeuAcα(2→3)Galβ(1→4)[Fucα(1→3)]GlcNAcβ(1→R). Its structure is presented in Fig. 1. 3. 5..

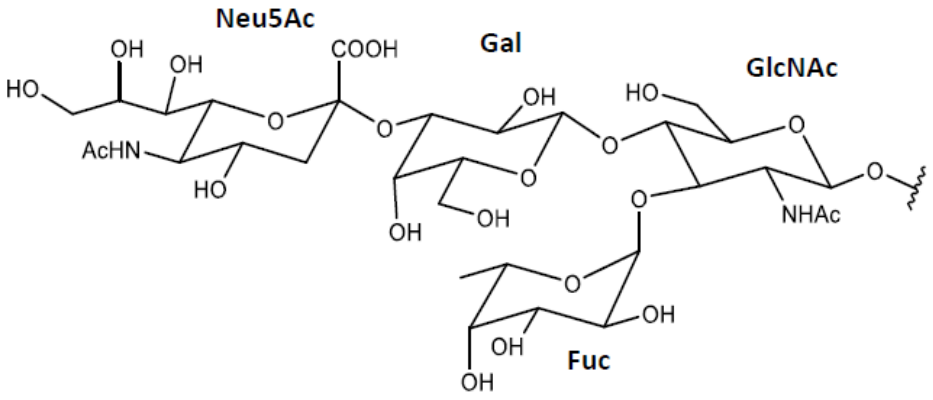


Fig. 1. 3. 5. Chemical structure of the sialyl Lewis x (sLe^x). sLe^x is a tetrasaccharide that consists of D-glucosamine (GlcNAc), D-galactose (Gal), L-fucose (Fuc) and N-acetylneuraminic acid (Neu5Ac).

Although all selectins recognise sLe^x, the individual proteins differ in their affinity for this ligand. The interaction is by far the strongest in the case of E-selectin, with K_D of 720 μM. The other two selectins exhibit binding affinities within the milimolar range, with K_D of 3.9 mM for L-selectin and 7.8 mM for P-selectin[77]. The exact cause for these differences is unclear, however, X-ray studies of selectin-sLe^x complexes have provided structural detail that sheds some light on possible reasons for the variation in binding affinities.

In crystal structures of P- and E-selectins, sLe^x is bound via the Ca^{2+}-coordination site in the lectin domain, adopting similar conformation in both cases[66,67]. Fucose 3- and 4-hydroxyl groups coordinate the Ca^{2+} ion and hydrogen bond to amino acids also involved in Ca^{2+} coordination. Additional interactions come from galactose which forms hydrogen bonds to Tyr94 and Glu92 (Fig. 1. 3. 6). The interactions observed for the two proteins diverge at the sialic acid-binding site – position of this sugar residue seems to be affected by the differences in the loop between residues 97-100. E-selectin forms tighter contacts with sLe^x in this region, with Arg97 forming hydrogen bonds to glycosidic oxygen and the carboxylate group of sialic acid. In P-selectin this position is occupied by a serine residue and, although there is a hydrogen bonding interaction between Ser99 and 4-hydroxyl group of sialic acid, there are less contacts overall between the carbohydrate and protein residues than in the case of E-selectin (Fig. 1. 3. 6)[66].

Many of selectin ligands bind, at least in part, via the sLe^x motif. Probably the best-studied selectin ligand is the P-selectin glycoprotein ligand-1 (PSGL-1) – a high-affinity ligand for L- and P-selectins, expressed on the surface of leukocytes and some of the activated endothelial cells. The binding affinity of P-selectin for PSGL-1 is 778 nm which is slightly stronger than its affinity for sLe^x determinant[66]. PSGL-1 is a 240-kDa transmembrane glycoprotein which contains multiple O-glycans and several N-glycans attached to two peptide chains extending into the extracellular space. The N-terminal region of the PSGL-1 peptide includes a sLe^x determinant and three sulfated tyrosine residues. The X-ray structure of P-selectin bound to PSGL-1 shows that the sLe^x motif binds to the lectin domain in the same manner as in the case of unconjugated sLe^x and that the sulfated tyrosines contribute to P-selectin binding by interacting with residues Arg85 and His114[66]. Indeed, multiple studies have shown that tyrosine sulfation plays an important role in the mediation of cell rolling and that at least one tyrosine needs to be sulfated to allow for the P-selectin-PSGL-1 binding[76,78–81]. However, there are some inconsistencies in the findings about the exact number of sulfated tyrosines required for binding, as well as their structural role, which has led to suggestions that there may be more than one possible conformations of PSGL-1 N-terminus capable of binding to P-selectin[66].

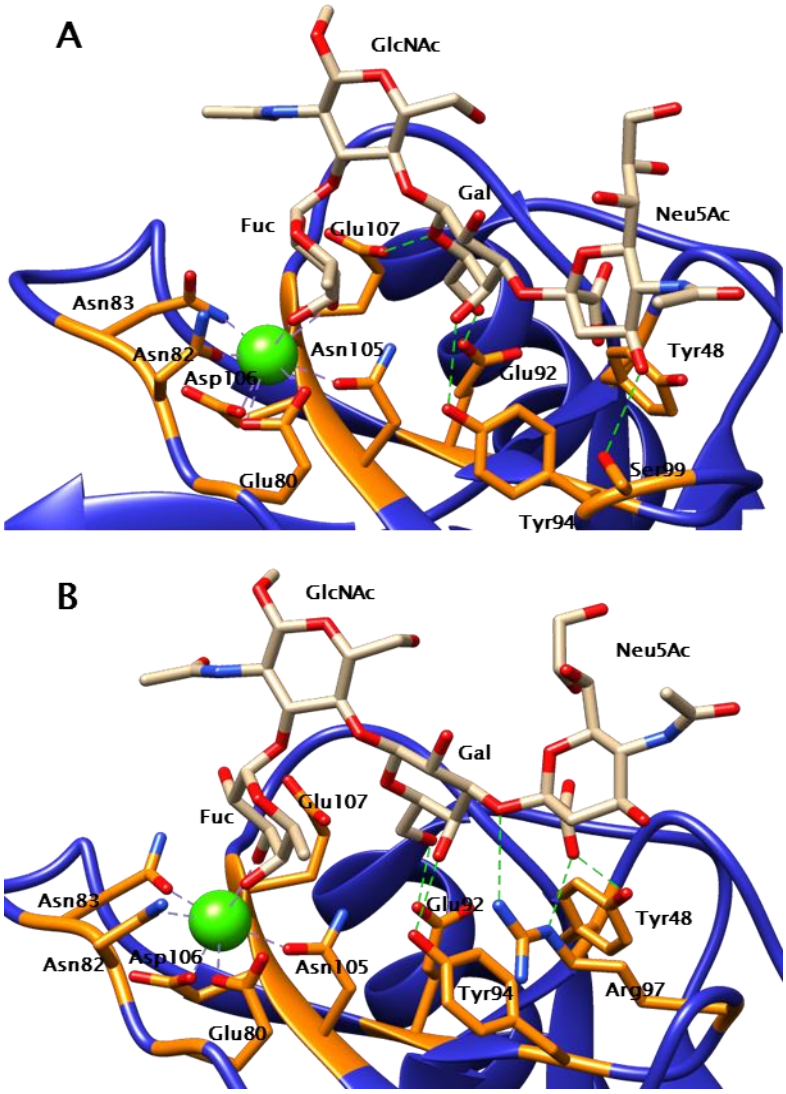


Fig. 1. 3. 6. X-ray structure of sLe^x binding site in P- and E-selectins. Ribbon representation of sLe^x binding site in P-selectin (**A**) and E-selectin (**B**), based on the X-ray structures from Somers *et al.*[66]. Proteins are shown in blue, with the interacting residues highlighted in orange. Carbohydrate residues are shown in light grey. Most differences between the two binding sites occur in residues surrounding the sialic acid (Neu5Ac).

In contrast to P- and L-selectins, E-selectin does not require sulfated tyrosines for binding to PSGL-1. It is possible that this behaviour is due to amino acid variants in its lectin domain that prevent it from forming stable interactions with the sulfates, however, the exact nature of these differences in binding is unclear[66,79,80]. E-selectins binds to PSGL-1 with much lower affinity than P- and L-selectins and the interaction between E-selectin and PSGL-1 is not required for cell attachment and rolling in vivo. This suggests that E-selectin binding to other ligands plays more important role in its function. In fact, E-selectin binds to the largest

spectrum of ligands, recognising a variety of fucosylated and sialylated glycolipids and glycoproteins. P- and L-selectins differ from E-selectin in that they prefer to bind to glycoproteins that are sulfated, as well as sialylated and fucosylated[82].

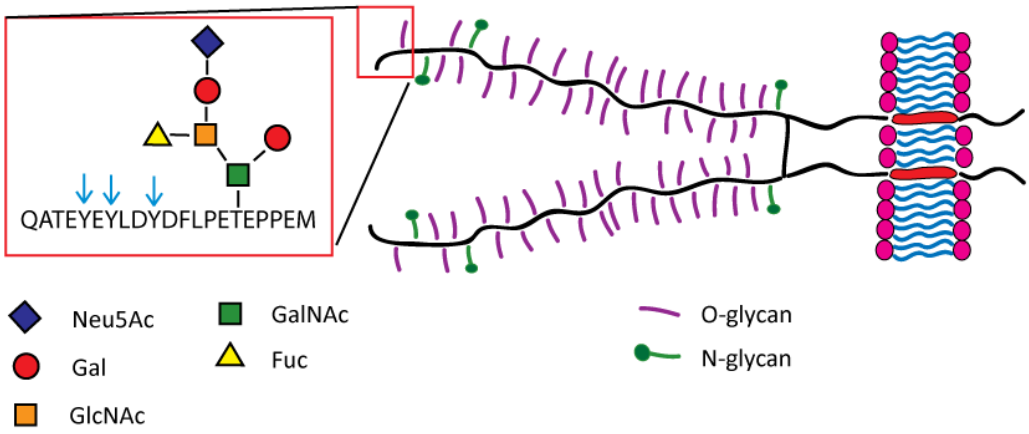


Fig. 1. 3. 7. Structure of P-selectin glycoprotein ligand-1 (PSGL-1). Schematic representation of the structure of PSGL-1, zooming in at the position of sLe^x motif (inset). Sulfated tyrosines are highlighted with blue arrows.

Whilst PSGL-1 is the only high-affinity ligand for P-selectin, both L- and E-selectins interact with other proteoglycans (Fig. 1. 3. 7.). Two major ligands for L-selectin are glycosylation-dependent cell adhesion molecule-1 (GlyCAM-1) and CD34, belonging to a family of peripheral node addressin (PNAd). Both of them are mucin-like glycoproteins decorated with multiple O-linked carbohydrate chains, expressed in HEV cells in the peripheral lymph nodes and at some sites of inflammation. CD34 is a transmembrane protein that plays part in the initial attachment of leukocytes to endothelium, whilst GlyCAM-1 is a soluble molecule secreted into blood that likely acts as a signal that initiates cell adhesion. These two ligands are capped with sLe^x motif, although modified with a sulfate ester at C6 position of GlcNAc. In combination with appropriate O-linked glycan chains this modification enhances L-selectin binding[82,83].

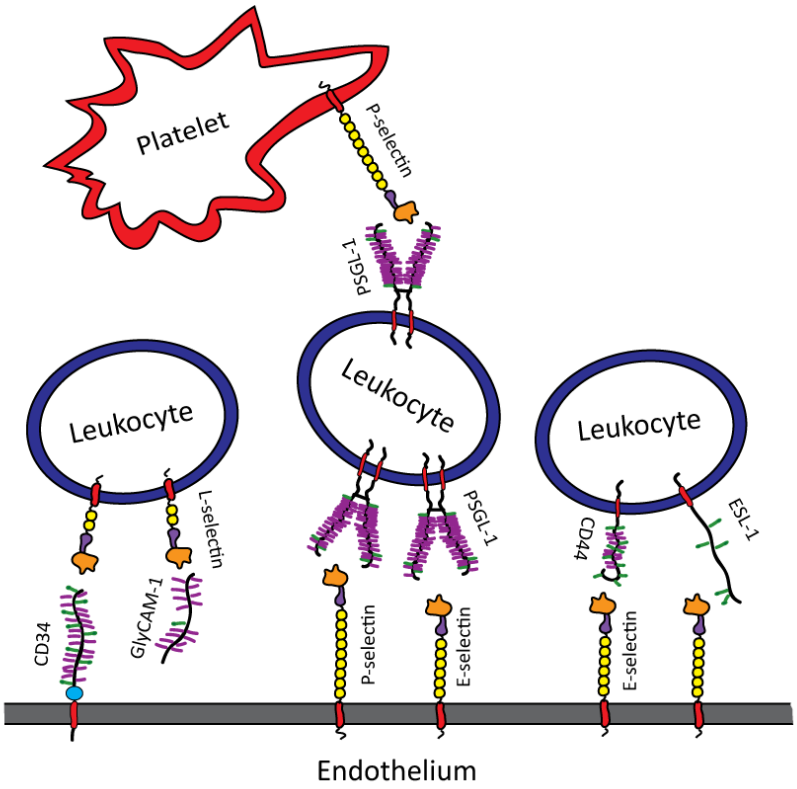


Fig. 1. 3. 8. Major selectin ligands and their expression sites. Schematic representation of major interactions that selectins are involved in, including the glycoproteins expressed on endothelium, platelets and leukocytes.

As mentioned above, E-selectin does not require sulfation for ligand recognition – the sLe^x motif is sufficient to support E-selectin-ligand binding[84]. Because of this lack of specificity, E-selectin has been proposed to recognise a variety of sialylated ligands, including PSGL-1, E-selectin ligand-1 (ESL-1), CD43, CD44 and glycolipids. It is also known to bind to L-selectin, depending on its glycosylation pattern[85]. Multiple E-selectin ligands are believed to collaborate with each other in mediation leukocyte rolling, facilitating the transition into slow rolling and controlling the velocity and arrest of the rolling cells on the endothelium[84]. Recent studies have shown that some of the E-selectin ligands may play different roles in adhesion of specific cell types. For example, PSGL-1 is involved in the attachment of mature neutrophils to endothelium, whilst ESL-1 supports rolling of hematopoietic progenitor cells[86].

1. 3. 4. Existing models for selectin–ligand binding

1. 3. 4. 1. Selectins and 'catch' bonds

In order to be able to fully understand the nature of selectin-ligand interactions, it is necessary to consider the natural environment selectins are found in. Located on the surface of leukocytes or vascular endothelium, these proteins are constantly exposed to the forces of the flowing blood. Their role is to recognise the carbohydrates on the interacting cell and to form a stable bond under such hydrodynamically challenging circumstances.

Multiple reports have shown that selectin bonds are quite unique in their response to shear force. Intuitively, one would think that increasing the external shear force would gradually weaken the intermolecular bonds, eventually reaching the point when ligand dissociates from the protein. This is the behaviour most commonly found in nature and is the definition of a 'slip' bond. However, what many studies have reported is that, as force is increased, the initial response of selectin-ligand complexes is that their interaction lifetimes also increase and that there is an optimal shear force threshold at which cell tethering and rolling is the most efficient. Therefore, high shear force can actually stimulate selectin function, rather than inhibit it[87–89]. This finding was quite revolutionary and caused some scientific disputes, especially that bonds of such nature had never been proven experimentally. As a result, a new type of bond was defined – a 'catch' bond – that is a bond whose lifetime is prolonged by shear force[88,89].

Selectins were not the only proteins that displayed this highly unusual behaviour. Further experiments have shown that shear-induced mechanical force enhances the interactions between bacterial protein subunit FimH and target host cells, as well as between Ibα glycoprotein and platelets during their adhesion to the sites of injury[90,91]. The existence of catch bonds was confirmed in 2003 by Marshall *et al.*, who used atomic force microscopy and flow-chamber experiments to show that increasing force first prolonged and then shortened the lifetimes of P-selectin complexes with PSGL-1[92].

Since their discovery, several models have been developed to explain the nature of selectin-ligand catch bonds in the light of selectin structure and interdomain interactions. The next section of this chapter will introduce the most important findings on the relationship between the lectin and EGF-like domains and how they led to development of such models.

1. 3. 4. 2. Lectin–EGF domain interface

Ever since the first reported X-ray structure of E-selectin, scientists have been trying to answer the question about the relationship between the lectin and EGF-like domains and how it affects selectin function. At first it may seem that the contacts between the two domains are very limited and that the domain interface is located so far from the binding site that it should not have a strong influence over the ligand binding. However, it is now well-established that the EGF-like domain does play a significant role in ligand binding and the adhesive processes.

Several studies have been undertaken, aiming to shed the light at the structure-function relationship of selectins and to answer the question of involvement of the EGF-like domain in ligand binding. A study published in year 2000 by Dwir *et al.*. was the first one to prove that the EGF-like domain affects the kinetics of selectin-ligand binding[93]. In this study, L-selectin mutant was expressed on the surface of leukocytes, with L-selectin EGF-like domain replaced with the same domain from P-selectin. In cells expressing the so-called LPL mutant cell adhesion was enhanced under shear stress and the shear threshold required for adhesion was lower than in the cells expressing the wild-type protein. In addition to these results, the study also observes that the EGF-like domain does not affect binding affinity in the absence of flow. Although the exact role of the EGF-like domain remained unexplained, the results clearly indicated that this domain is important for selectin-ligand binding under shear flow conditions, despite the fact that it does not interact directly with the ligand[93].

Another important finding came from the X-ray structure of the P-selectin lectin and EGF-like domains, published by Somers *et al.* in the same year[66]. In this study, the P-selectin was crystallised in two different conformations – in more compact 'bent' conformation when not bound to any ligands or when bound to sLe^x, and in 'extended' conformation when bound to PSGL-1 (Fig. 1. 3. 9.). The switch from bent to extended conformation occurs via movement of the interdomain hinge, displacing the EGF-like domain with respect to the lectin domain. These X-ray studies of P-selectin provided the first clue that the protein can exist in either bent or extended conformation and that this conformational change may be related to ligand-binding affinity under shear flow conditions. It was hypothesised that P-selectin has two states that support binding of different ligands: sLe^x in low-affinity bent conformation and PSGL-1 in high-affinity extended conformation. However, no experiments were undertaken at this stage to confirm this hypothesis. It remained unclear whether the conformational

changes are induced by the ligand binding or whether it is the external shear force that induces changes leading to higher binding affinity of the protein[66].

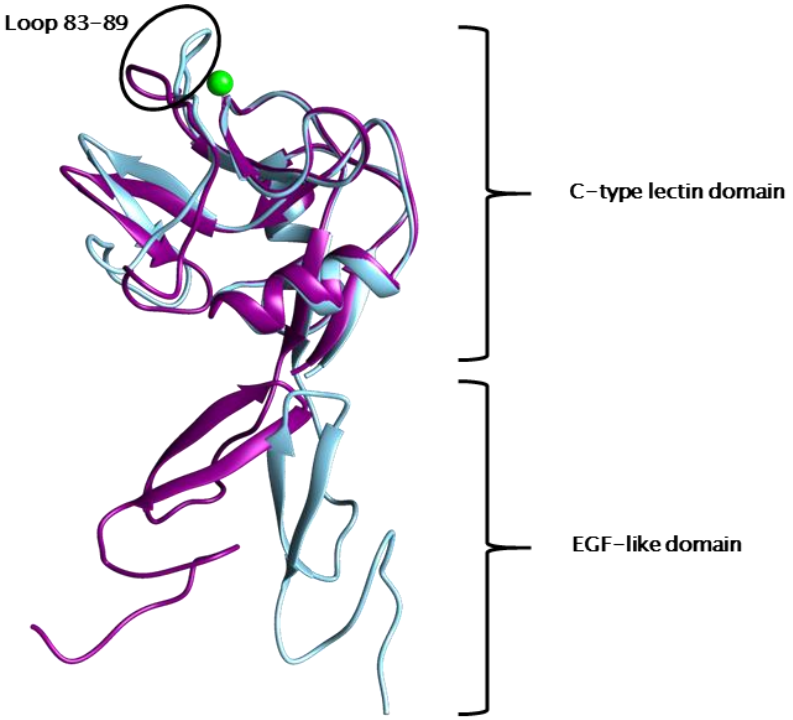


Fig. 1. 3. 9. P-selectin lectin and EGF-like domains in bent and extended conformations, as reported in the X-ray structure by Somers *et al*.[66]. Bent conformation, observed upon binding to sLe^x, is represented in purple and the extended conformation, observed upon binding to PSGL-1, is represented in light blue.

Six year later, two publications revealed important new information on the role of interdomain hinge in mediating the cell rolling[73,94]. By introducing modifications to the interdomain region of P- and L-selectins, it was possible to determine that flexibility of this region regulates protein's ability to tether and roll the cells. Both groups, independently of each other, mutated residue 138 of L-selectin, by substituting glycine for asparagine, to see how it will influence the ligand binding and cell adhesion. This residue attracts special interest as it is not conserved in all three selectins. P-selectin differs from L- and E-selectins in that it has glycine rather than asparagine at position 138 and is missing an interdomain hydrogen bond between Asn138 and Tyr37. The hypothesis was that such a change would influence flexibility of the protein and therefore its binding properties[73,94].

In Phan *et al.*, cells expressing the N138G mutant rolled more slowly over PSGL-1-expressing surfaces than cells expressing the wild-type protein[94]. In Lou *et al.*, cells expressing the mutant protein tethered to PSGL-1 coated surface at higher rates than those expressing the wild-type protein[73]. Another observation was that less shear force was required to activate bond formation, suggesting that hinge flexibility contributes to the 'catch' behaviour of selectin-ligand bonds. Additionally, in the study by Phan *et al.*, a new glycosylation site was introduced at the domain interface of P-selectin, acting as a wedge that locks the protein in extended conformation. Cells expressing the modified protein displayed fivefold increase in the binding affinity for PSGL-1, with reduced rates of the bond formation and detachment. As a result of these changes, cells expressing the mutant protein were characterised by reduced rolling speeds and tethering frequencies[94].

Taken together, these studies show that the interdomain region does play an important role in selectin function. It can be concluded that higher flexibility of the hinge lowers the shear threshold required for adhesion, which is likely due to the higher number of possible arrangements of the lectin domain with respect to the EGF-like domain and, therefore, higher chance for carbohydrate binding. It was also shown that locking the protein in a single conformation can increase its binding affinity but by the same means also lower its ability to mediate cell rolling. Based on these findings, a new model for transitions between the conformational states of selectins was developed, known as the 'sliding-rebinding' model[95].

1. 3. 4. 3. Sliding-rebinding and allosteric binding models

There are currently two leading models for the mechanism behind selectin-ligand catch bonds. In the sliding-rebinding model hinge extension does not lead to changes in the binding site conformation but it acts by tilting the ligand-binding interface in the direction of applied force (Fig. 1. 3. 10)[95]. Based on this model, the ligand does not bind to a single specific binding site but slides across the lectin domain surface, binding and dissociating from multiple overlapping binding sites[73,95].

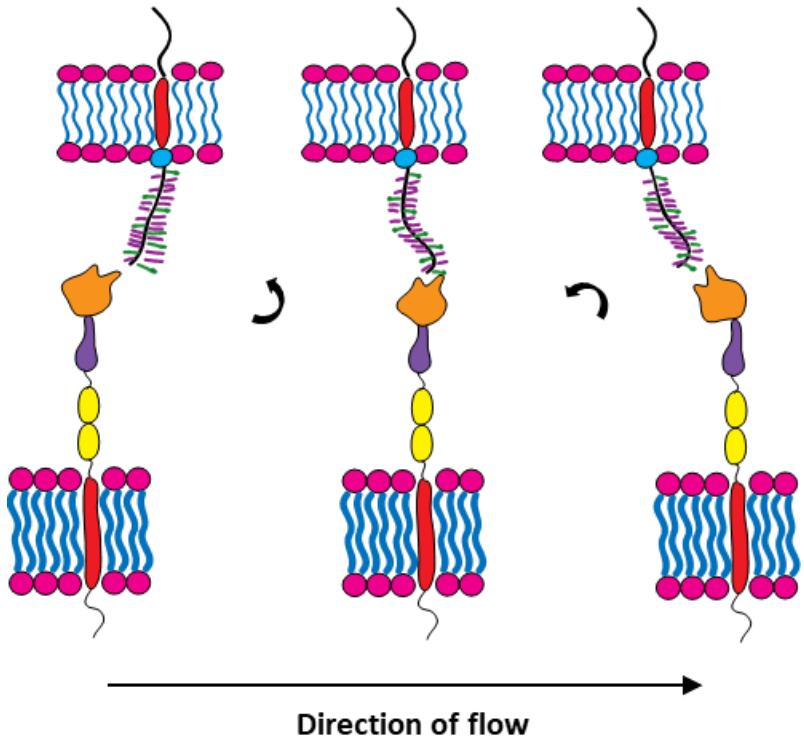


Fig. 1. 3. 10. Schematic representation of the sliding-rebinding model. In this model, flexibility in the interdomain hinge region allows the C-type lectin domain to adjust its position in the flow condition. The ligand binding is a dynamic process in which the carbohydrate ligand slides across the domain surface, continuously breaking and forming new interactions.

An alternative view is presented in the allosteric model. In this more traditional model the protein exists in two distinct states that are in equilibrium with each other – a lower affinity 'bent' state and higher affinity 'extended' state. Transition to extended state causes allosteric changes in the conformation of the binding site by closing the loop between residues 83-89, which is thought to increase protein's binding affinity (Fig. 1. 3. 9)[67,69,74,96].

There are unanswered questions about both models and neither of them is an ideal fit for the experimental results. It is possible that the actual mechanism for selectin binding is a mixture of these two models, where, although the extended state is a higher affinity state, the ligand binding is enhanced not by the allosteric changes within the binding site but by gradual mechanistic extension that reorients the lectin domain to allow for more efficient interaction with ligands as the cell rolls along the surface. The binding could then follow the sliding-rebinding model, allowing the protein to gradually increase its interaction with ligands as the cell rolls over the selectin-expressing surface.

Indeed, a study by Dwir *et al.*. has shown that higher shear rate causes formation of more tethers on leukocytes[97]. This result has led to suggestions that, after the formation of initial tether, cell rotates over the adhesive surface, gradually finding more tethers, and where the enhanced shear rate increases probability of finding the ligand[97]. It is possible that the extended conformation is easier to access and allows for formation of more stable tethers which could complement the sliding-rebinding model.

1. 3. 5. Selectins and cancer

Although the primary function of selectins is mediation of cell rolling in inflammatory response, other cell types can bind to selectins when migrating to new tissues via circulatory system. It is well-known that cell-surface glycosylation pattern is altered in metastatic cancer cells, improving their adhesion ability. Often the changes in glycosylation include increased presence of sialic acids, suggesting involvement of selectins in metastatic cell adhesion processes[98]. Indeed, multiple studies have shown that selectins have the ability to bind to cancer cells and that they are involved in metastatic processes[99–103].

For the above reasons, selectins are attractive targets in design of drugs acting to inhibit metastasis by preventing adhesion of tumour cells to blood vessels. In fact, it has been shown that the inhibition of P-selectin leads to reduction in metastasis *in vivo* in different types of tumour[104]. Heparin-related compounds have been shown to decrease metastasis in mice, likely through interaction with P-selectin. One of the recent studies used a holothurian GAG similar to heparin to decrease melanoma lung metastasis in mice[44]. This result is in agreement with the previous studies showing that sulfated GAG-like carbohydrates reduce metastasis by the inhibition of P-selectin[24,42,105].

1. 3. 6. Selectin–fCS binding

As fCS shares structural similarity to the compounds described above, it has a potential to act as a selectin inhibitor which could account for its anti-inflammatory and antimetastatic

properties. Indeed, fCS from *L. grisea* has been found to bind strongly to the L- and P-selectins, inhibiting their binding to carbohydrate determinant sialyl Lewis x (sLe^x) as well as to the LS180 carcinoma cells[42]. The inhibition was four- to eightfold more potent than in case of heparin for P- and L-selectins and, as in case of heparin, no inhibition was observed for E-selectin[42]. In addition to these results, mouse models have been used to demonstrate that fCS is able to inhibit lung colonisation by adenocarcinoma cells *in vivo*[42].

High affinity of fCS oligosaccharides for selectins may be a result of an interesting structural arrangement of the monosaccharide subunits, discovered in the NMR studies of the fCS from *H. forskali*, described earlier in the chapter (Section 1. 2. 3). The results of these studies indicate that fucose residue may be stack on top of galactosamine, in an arrangement strikingly similar to the one found in sLex tetrasaccharide, where fucose has been found to stack on top of galactose ring by X-ray crystallography[106] and NMR[107]. In both cases, fucose has been found to remain in close proximity to galactose (~2.5 Å) in a rigid arrangement, stabilised by non-polar van der Waals interactions[107]. In the analogous arrangement found in the fCS oligosaccharides, several sulfate groups are brought to close spatial proximity and a large negative patch is created. This accumulation of negative charge, along with the sLe^x-like conformation of the repeating trisaccharide unit are likely the key properties that enhance the fCS-selectin interaction.

Following these observations, a study was performed in collaboration with Prof. Feizi (Imperial College of London), using a microarray of fCS fragments to test their binding to L-, P- and E-selectins (Fig. 1. 3. 11)[55]. As the fCS samples extracted from the sea cucumber body wall are long polysaccharides (estimated MW ~120-140 kDa), they were cut into smaller oligosaccharide fragments by Cu^{2+}-catalysed Fenton type (symbolised by letter F) and photochemical (P) depolymerisation[55]. Fragments of different lengths were produced, ranging from three (tri) to ten (deca) fCS units in each chain. Although each sample displayed a level of heterogeneity, the majority of fragments in each sample should be of the estimated molecular weight (MW). It must be noted that high-MW fragments were also heterogeneous in sequence as it is not currently possible to produce fragments of exactly the same composition by depolymerisation. To separate the fragments even further, each oligosaccharide type was divided into two subfractions – upper (U) and lower (L) by preparative thin layer chromatography (TLC)[55]. Fig 1. 3. 11 shows the results of the microarray study using these fragments. Each fraction was arrayed at two levels, 2 and 5 fmol/spot in duplicate and the response was dose-dependent. The L, P and E-selectins were prepared as culture supernatants of IgM chimeras and their binding was detected with biotinylated anti-human IgM followed by Alexa Fluor-647-labelled streptavidin.

The results clearly show that all oligosaccharide fractions bind to L- and P-selectin but not to E-selectin. The response was higher for L-selectin than P-selectin. The lower affinities of some of the fragments, eg. tetra- and higher MW fractions, could be due to the higher heterogeneity of these fractions in monosaccharide sequence or degree of sulfation. Overall, the study confirmed that the fCS oligosaccharides from *H. forskali* bind to L- and P-selectins more strongly than sLe^x and that the binding affinity differs between oligosaccharides of different lengths[55].

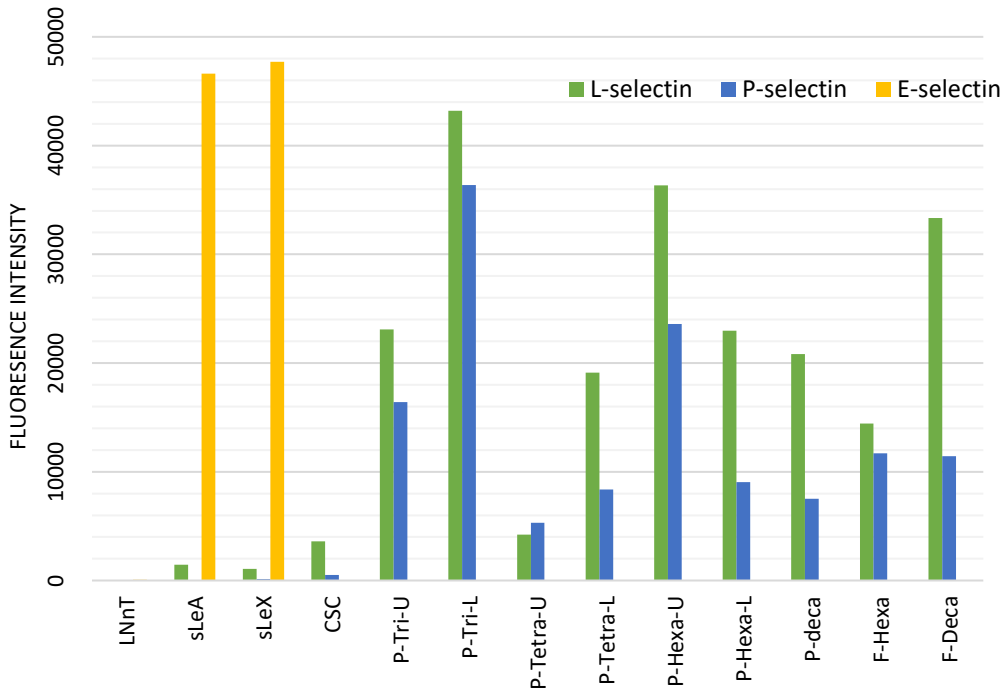


Fig. 1. 3. 11. Results of microarray analysis of binding between the different fCS oligosaccharides and selectins. Fluorescence intensity levels for different fCS oligosaccharides produced by depolymerisation (columns 5-13) upon binding to L-, P- and E-selectins. For comparison, fluorescent intensities for sLe^x and structurally closely related sLe^A are also shown (columns 2 and 3 respectively). Columns 1 and 4 are control results for lacto-N-neotetraose (LNnT) and chondroitin sulfate C hexasaccharide (CSC)[55].

Results of this microarray study confirm the literature findings on the strong affinity of L- and P-selectins for fCS. It is unclear why E-selectin presents such a dramatically different response to the other two proteins. Detailed structural studies would be needed to account for this difference, as well as to fully understand the nature of the binding between selectins and fCS oligosaccharides. In this work, the results of molecular dynamics and NMR studies of fCS-selectin binding will be presented and discussed in the light of the current knowledge on selectin-ligand interactions.

Chapter 2: Aims

Due to its diverse biological activities, fCS is a promising drug candidate that could be used in treatment of severe conditions such as inflammatory disease, cancer or HIV. Understanding the reason behind the strong affinity of L- and P-selectins for fCS oligosaccharides at the molecular level would provide a valuable contribution to the process of design and development of new selectin inhibitors. It could also be of great value to wider efforts of scientific community, providing data and tools to gain insight into other carbohydrate-protein complexes.

The major aim of this work was to obtain structural information that would shed more light on the nature of fCS-selectin interaction. The initial goal was to express ^{13}C, ^{15}N – labelled recombinant selectin constructs and to use them in Nuclear Magnetic Resonance (NMR) studies. As will be explained in the following chapters, due to problems with protein expression, a different approach had to be taken and in later stages of the project Saturation Transfer Difference (STD) NMR was used, along with computational methods, to study the fCS-selectin interactions.

The secondary aim of this work was to develop new NMR tools to facilitate the analysis of complex carbohydrate samples. Heterogeneous nature of these samples makes them challenging to study by spectroscopic methods. Even with a large library of one- and two-dimensional NMR experiments, structural elucidation of carbohydrate samples is hindered by the high amount of peak overlap in the NMR spectra. To target this problem, we developed a variation of the HSQC and HSQC-TOCSY experiments which uses principles of the G-matrix Fourier Transform (GFT) NMR to separate the overlapping signals.

The experimental work undertaken to achieve the above aims will be described in the following chapters, starting with the protein expression experiments, followed by the STD experiments and computational studies of fCS-selectin interactions, and, finally, by the development of new NMR methods for studies of carbohydrate samples.

Chapter 3: Selectin expression tests

3. 1. Introduction

This chapter will describe the work on expression of the ^{15}N, ^{13}C-labelled recombinant constructs of human selectins for structural studies of selectin-fCS interactions by NMR. Because the E-selectin has previously shown to have no affinity for the fCS oligosaccharides, all experiments described here were focused on the expression of L- and P-selectin constructs only. As selectins are large, transmembrane proteins, expression of the full-length protein was not considered feasible. The region of main interest was the extracellular part of the protein, in particular the C-type lectin domain, directly involved in the interactions with carbohydrate ligands, and the neighbouring EGF-like domain, believed to play a regulatory role in these interactions. The aim was to express two constructs for each selectin: a single-domain construct, consisting only of the C-type lectin domain, and a double-domain construct, consisting of both the lectin and the EGF-like domains. Having these two constructs would not only allow us to determine which residues within the lectin domain mediate the interactions with the fCS oligosaccharides, but also to perform comparative studies on the binding between the single- and double-domain constructs and to assess the role of the EGF-like domain in these interactions.

Double-domain selectin constructs have been previously expressed in the Chinese Hamster Ovary (CHO) cells for X-ray crystallography studies[65,66]. In the later studies of L-selectin glycosylation the protein was expressed in Human Embryonic Kidney (HEK293) cells[108]. Use of mammalian cells is a sensible option, providing the most reliable machinery for transcription and translation of human proteins. However, as the main purpose of the planned experiments was to produce the ^{15}N, ^{13}C-labelled constructs, a simpler expression system was desirable, for which well-established protocols for the isotopic labelling of produced proteins exist.

The School of Chemistry at the University of Edinburgh has expertise in protein production in two unicellular expression systems, the *Escherichia coli* and *Pichia pastoris*. The first of the two is the well-known bacterial expression system, used for fast, efficient and cost-effective production of many recombinant proteins. However, bacterial cells are often not able to

produce folded and functional human proteins. Formation of disulfide bonds, often found in human proteins, requires the presence of oxidative environment within the cell. In eukaryotic cells, such environment is found in the endoplasmic reticulum (ER) – a system of membrane compartments which allow for synthesis and chemical modification of different biomolecules. In prokaryotic cells the ER is missing, which means that the majority of protein folding processes will take place in the cytoplasm – an environment in which the formation of disulfide bridges is not favourable[109].

P. pastoris is also a simple, unicellular organism but it has the eukaryotic organelles and machinery that supports synthesis and folding of complex proteins. Therefore, it is a more promising candidate for expression of recombinant selectin constructs. Like *E. coli*, the cells are easy to transform and are able to express recombinant proteins in high yields[110,111]. Additional benefit of using this expression system comes from the fact that the proteins can be secreted into the extracellular media for easy purification. A disadvantage of using this expression system is that *P. pastoris* cells replicate at slightly lower rates and need longer times for culture growth and protein expression than *E. coli* cells. However, *P. pastoris* expression system has been successfully used for production of a variety of recombinant, isotopically labelled human proteins for NMR studies[110–113]. It was therefore decided that the first set of expression tests will be performed in the *P. pastoris* expression system.

3. 2. Selectin construct expression in *P. pastoris*

3. 2. 1. Expression system

Pichia pastoris is a methylotrophic species of yeast – it can utilise methanol as a sole source of carbon and energy. This unique ability to metabolise methanol allowed for development of an expression system for recombinant proteins, coupled to expression of alcohol oxidase (*AOX1/AOX2*) gene. This gene encodes for an enzyme that catalyses the first step in the methanol utilisation pathway – oxidation of methanol to formaldehyde. Expression of the *AOX1* gene is controlled at the transcription level – alcohol oxidase will be produced only in the presence of methanol which allows for regulated induction of protein expression. The

cells can be cultured to high densities and are easy to transform, with protocols similar to those used for other commonly used species of yeast[110].

Recombinant proteins can be secreted into the medium in presence of the α-factor signalling sequence from *S. cerevisiae*. As only low levels of endogenous proteins are secreted, the recombinant proteins should constitute the majority of the total protein in the culture media and can therefore be easily purified. However, the efficiency of secretion and purification vary for different proteins and may be difficult in certain cases[114].

Several strains of *P. pastoris* have been developed, with slight differences in the genotype and phenotype characteristics. In the experiments described here, strain KM71H was used – the only *P. pastoris* strain that relies on *AOX2* rather than *AOX1* gene for methanol metabolism, also known as Mut^S strain. *AOX1* gene is a much stronger gene and its deletion in KM71H strain significantly slows down methanol metabolism, reducing the cell growth rates. Lack of the *AOX1* gene, however, has its benefits: in other strains, high rate of methanol metabolism leads to accumulation of toxic by-products such as hydrogen peroxide and formaldehyde. Such substances increase oxidative stress, inhibit cell growth and ultimately lead to cell lysis and release of proteases into the media[115]. In addition to that, strong expression of AOX1 gene product may compete with the production of recombinant proteins[116]. Although slower-growing, KM71H strain allows to avoid the above problems and has been successfully used in production of several protein constructs at the University of Edinburgh. As AOX1 gene is only introduced on the vector containing the recombinant protein, it triggers high-levels of expression of the recombinant protein, in some cases offering higher yields than the Mut+ strains. Due to these benefits, as well as the long-standing experience of the use of KM71H strain for the recombinant protein expression at the University of Edinburgh (Barlow group), this strain was chosen as a suitable candidate for expression of selectin constructs.

3. 2. 2. Construct design

Two types of constructs were designed for expression of L- and P-selectins in *P. pastoris*: the single-domain (LS1, PS1) and the double-domain construct (LS2, PS2). The double-domain constructs were based on the example of constructs produced previously for the X-ray studies[65,66]. The N-terminus was conserved as in the full-length protein, with Trp1 being the N-terminal residue. The C-terminal residue was Gln158 in the case of L-selectin and Glu158

in the case of P-selectin double-domain construct. The single-domain constructs were designed to be 35-residue shorter, terminating at residue Ser121. This decision was based on the direct involvement of Ser121 in the stabilising interactions with the protein's N-terminus, observed in the reported crystal structures. There is no defined border between the C-type lectin and the EGF-like domains and residue Ser121 was chosen as the terminal residue based on the fact that it is the last residue which is actively involved in stabilisation of the lectin domain fold. The amino acid sequences of all four protein constructs are included in Appendix 1.

Selectin constructs designed here were inserted into the pPICZαB vector. This vector contains several elements designed to help in efficient expression of recombinant proteins. pUC bacterial origin of replication facilitates production of high number of copies of the plasmid. *AOX1* promoter and terminator allow for controlled expression of the gene of interest, whilst a sequence encoding for resistance to zeocin allows to selectively culture only the transformed strains (Fig 3. 2. 1). The constructs were designed to be inserted into the vector with the use of *PstI* and *XbaI* restriction enzymes, at the site following the α-factor signal sequence to allow for protein secretion into the media.

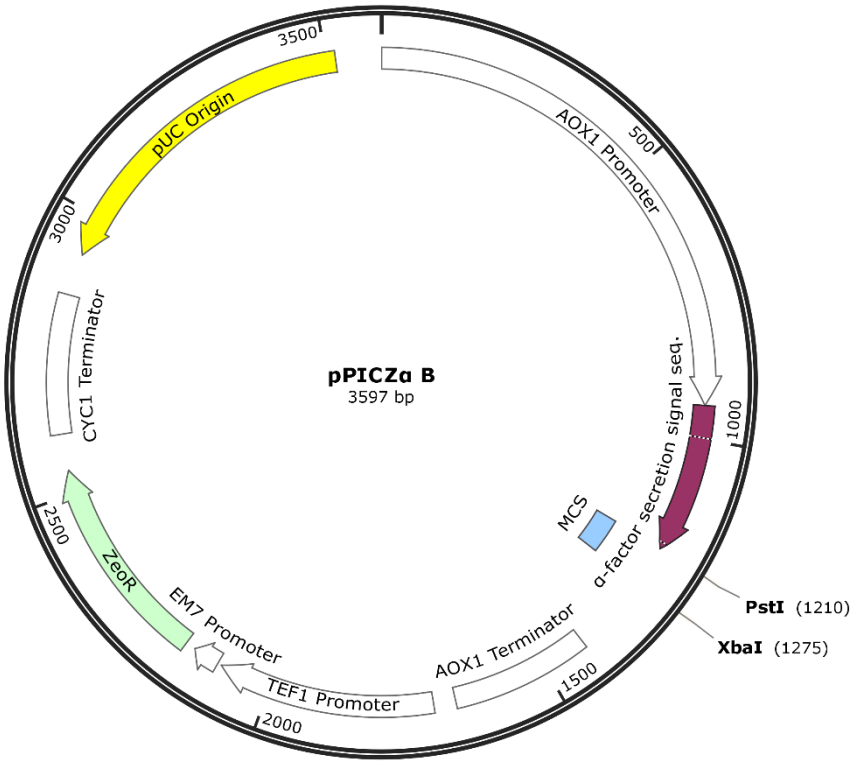


Fig. 3. 2. 1. Map of pPICZαB vector. The gene of interest is inserted after the AOX1 promoter and includes the N-terminal α-factor signalling sequence for protein secretion. Vector also encodes for resistance to zeocin. Vector image was produced in SnapGene® Viewer.

3. 2. 3. Materials and methods

3. 2. 3. 1. Plasmid preparation

The recombinant DNA fragments were based on the sequences deposited in the Uniprot database (# 14151, # 16109), codon optimised for expression in *P. pastoris*, synthesised and subcloned into the pPICZαB vector by GenArt™ (Thermofischer Gene Synthesis Suite), with the use of the PstI and XbaI restriction sites. The received vial was centrifuged at 6000 × g for 1 min at 4°C and the plasmids (5 µg) were resuspended in 50 µL of ultrapure water prior to transformation into *E. coli* cells.

3. 2. 3. 2. Plasmid transformation into E. coli

1 µL of the resuspended plasmid DNA was added into an aliquot (50 µL) of chemically competent Top10 cells. Cells were incubated on ice for 20 min, then heat-shocked for 30s at 42 °C, returned on ice and diluted with 80 µL of S. O. C. medium, incubated for 1 h at 37 °C and plated at four different concentrations (5 µL, 10 µL, 20 µL and 40µL) on 15 mL LB agar plates with 25 µg/mL of zeocin (InvivoGen). Plates were allowed to dry and incubated overnight at 37 °C.

The next day, transformed colonies (5 colonies per construct) were used to inoculate 5 mL of LSLB with 25 µg/mL zeocin. Cultures were grown for 8 h at 37 °C, with shaking at 200 rpm, transferred to 250 mL LSLB, incubated overnight at 37 °C with shaking at 200 rpm and harvested by centrifugation at 5000 × g for 10 min at 4 °C. Pelleted cells were then used for plasmid preparation. Plasmids were prepared from the *E. coli* cultures using the QIAgen Plasmid Maxi Kit, according to manufacturer's instructions. The final DNA samples were quantified (A_{260}), linearised with SacI-HF® enzyme (New England BioLabs) for 3 h at 37°C and separated by electrophoresis on a 1 % TBE agarose gel, visualised with SybrSafe™ (Invitrogen), with 5 µL of Hyperladder™ 1kb DNA GeneRuler (Thermo Scientific) as a reference and sequenced at the Edinburgh Genomics Facility.

3. 2. 3. 3. Extraction of the plasmid DNA

To obtain high-purity DNA for the transformation of *P. pastoris* cells, the linearised plasmid DNA was extracted using phenol-chloroform method. Phenol:chloroform:isoamyl alcohol (25:24:1, Sigma) was added to equal volume of linearised DNA. The extraction mixture was centrifuged at 20 000 × g for 1 min, the top layer transferred to a new tube and the bottom layer discarded. This procedure was repeated once. After the top layer was transferred again to a fresh tube, equal volume of 100 % chloroform was added to it. After another centrifugation step, the top layer was transferred again to a new tube and 0.1 volume of 3 M sodium acetate and 2 volumes of 100 % cold ethanol were added to it. The mixture was incubated at -20 °C overnight and centrifuged at 20 000 × g for 30 min. Supernatant was discarded and pellet was washed with 400 µL of 70% cold ethanol. Sample was centrifuged at 20 000 × g for 10 min and ethanol was removed by pipetting and air-drying. Pellet was resuspended in 12 µL of ultrapure water and DNA concentration was quantified (A_{260}).

3. 2. 3. 4. Plasmid transformation into P. pastoris

The KM71H strain of *P. pastoris* was kindly provided by Dr McIntosh. A single colony was picked and used to inoculate 5 mL of YPD media. Cells were incubated for 24 h at 30 °C, with shaking at 220 rpm, used to inoculate 100 mL of YPD media at three different concentrations and incubated again overnight in the same conditions. On the next day, OD_{600} was measured and the cell culture that reached the value of 1.3 was used for further experiments.

Cell culture was centrifuged at 1500 × g for 5 min at 4 °C, supernatant was discarded and pellet was resuspended in 100 mL of ice-cold sterile water. This procedure was repeated one more time. Sample was centrifuged for the third time, supernatant discarded and pellet resuspended in 8 mL of ice-cold 1M sorbitol. This step was repeated but this time the cells were resuspended in 200 µL of ice-cold 1M sorbitol. 80 µL of the cells were added to 10 µL of linearised plasmid DNA and transferred to a pre-chilled electroporation cuvette. Cells were pulsed with electric current in the electroporation tray (Gene Pulser Cuvette, Bio-Rad) and immediately resuspended in 1 mL YPDS with 1 M sorbitol, transferred to a new tube and incubated for 3 h at 30 °C. After the incubation, cells were streaked onto YPDS agar plates

containing zeocin at three different concentrations (100 μg/mL, 200 μg/mL, 300 μg/mL) and incubated at 30 °C for 72 h. The colonies that grew at the highest zeocin concentration were used in the protein expression tests.

3. 2. 3. 5. Small-scale protein expression tests

A single colony of transformed *P. pastoris* strain was used to inoculate 5 mL of buffered glycerol complex medium (BMGY). Cells were incubated overnight at 30 °C with shaking at 220 rpm. On the next day, the cell culture was transferred into 100 mL of BMGY and incubated for another 24 h at 30 °C with shaking at 220 rpm. The culture was then centrifuged at 3000 × g for 5 min at 4 °C, supernatant discarded and pellet resuspended in 25 mL of buffered methanol complex medium (BMMY). Methanol was added in concentrations ranging from 0.5 % to 1 % of the culture volume, at 24, 48, 66 and 72 h after the transfer to BMMY. Cells were harvested at 90 h from induction by centrifugation at 5000 × g for 20 min at 4 °C. Supernatant was collected and used in protein purification tests.

3. 2. 3. 6. Large-scale protein expression tests

A single colony of transformed *P. pastoris* strain was used to inoculate 8 samples of 5 mL of buffered glycerol complex medium (BMGY). Cells were incubated overnight at 30 °C with shaking at 220 rpm. On the next day, the cell culture was transferred into 8 samples of 1 L of BMGY and incubated for another 24 h at 30 °C with shaking at 220 rpm. After that time, the cultures were centrifuged at 3000 × g for 5 min at 4 °C, supernatant discarded and pellet resuspended in 8 samples of 125 mL of buffered methanol complex medium (BMMY). Methanol was added in concentrations ranging from 0.5 % to 1 % of the culture volume, at 24, 48, 66 and 72 h after the transfer to BMMY. Cells were harvested at 90 h from induction by centrifugation at 5000 × g for 30 min at 4 °C. Supernatant was collected and used in protein purification tests.

3. 2. 3. 7. Fermentation conditions

Test fermentation was performed in the Bioflow 3000 fermentor at the University of Edinburgh. A single colony of the LS2 construct was used to inoculate 10 mL of BMGY. Cell culture was incubated for 2 days at 30 °C with shaking at 220 rpm. This culture was used to inoculate 200 of BMGY and incubated for 24 h at 30 °C with shaking at 220 rpm. 100 mL of the cell culture was added to the fermentor, containing 600 mL of autoclaved basal salts (27 mL of 85 % H_2PO_4, 0.95 g $CaSO_4$, 18.2 g K_2SO_4, 15 g $MgSO_4$, 4.2 g KOH and 25 mL glycerol per 1 L solution) and 2.6 mL of filter-sterilised PTM salts (BioWorld). The fermentation was run at 30 °C and pH was maintained at 5.0, using undiluted ammonium hydroxide. Dissolved oxygen (DO) was maintained above 20 % by adjusting agitation rate and pure oxygen supply.

When the initial glycerol was depleted, as indicated by an abrupt increase in DO level, a 50 % solution of glycerol with 6 mL/L PTM salts was added to the fermentor. On day 2 from the start of fermentation growth, temperature was lowered to 20 °C and cells were induced with 6 mL of a solution of methanol with 12 mL/L PTM salts. 12 mL of the methanol solution was added again after 8, 19, 28, 39, 47, 54 and 66 h from the start of induction. Cells were harvested by centrifugation at 5000 × g for 30 min at 4 °C.

3. 2. 3. 8. Small-scale protein purification tests

10 – 20 mL disposable columns were used for the small-scale protein purification tests. The tested resins included SP Sepharose, SP-650M Toyopearl, Heparin Sepharose, Concavalin A (GE Healthcare). Supernatant from the *P. pastoris* cultures was diluted 1 in 6 and its pH was adjusted to be equal to pH of the cation exchange buffer, which ranged from 4.0 to 7.0. The buffer used was either 20 mM potassium phosphate or, in later experiments, 20 mM sodium acetate. In the cation exchange chromatography, column was equilibrated in five column volumes of the buffer and 30 – 60 mL of the supernatant was loaded onto the resin. The elution buffer contained sodium chloride in concentrations ranging from 0.1 M to 1 M. Fractions of 0.5 mL were collected and analysed by UV (A_{260}, A_{280}) and SDS-PAGE.

3. 2. 3. 9. Large-scale protein purification tests

The purification system used for the large-scale protein purification tests was ÄKTA FPLC P-920. The columns tested included SP Sepharose Fast Flow and HiLoad 16/600 Superdex 75 (GE Healthcare). Supernatant from the *P. pastoris* cultures was diluted 1 in 6, 10 mM $CaCl_2$ and PMSF protease inhibitor (Sigma) were added and the pH was adjusted to be equal to pH of the purification buffer, which ranged from 4.0 to 7.0 for cation exchange chromatography and 7.4 for gel filtration. The buffer used was 20 mM sodium acetate or, in case of the gel filtration experiments, 20 mM Tris with 500 mM NaCl. In the cation exchange experiments, column was equilibrated in five column volumes of the buffer and 1 L of the supernatant was loaded onto the resin. The elution buffer contained sodium chloride in concentrations ranging from 0 M to 1 M. Fractions of 0.2 mL were collected and analysed by UV (A_{260}, A_{280}) and SDS-PAGE.

3. 2. 3. 10. SDS-PAGE analysis

Protein MW and purity were assessed by SDS-PAGE. Selected fractions were deglycosylated, using the Endo-HF enzyme (New England BioLabs) according to manufacturer's instruction. For the SDS-PAGE analysis, 15 – 20 µL of the sample was combined with 3 µL of the NuPAGE® LDS Sample Buffer (4X, Invitrogen) and 2 µL of the NuPAGE Sample Reducing Agent (10X, Invitrogen), heated for 3 min at 100°C and stored on ice before loading on the gel.

The SDS-PAGE was run using the precast NuPAGE Novex 4 – 12 % Bis-Tris Protein Gels (Invitrogen) with the Precision Plus Protein™ (Bio-Rad) or PageRuler Plus (ThermoFisher) molecular weight marker.

3. 2. 3. 11. Western blot analysis

Western blot was performed in order to selectively detect the L-selectin protein constructs. Jurkat Whole Cell Lysate (Santa-Cruz Biotechnology), containing the full-size L-selectin, was

used as a positive control. SDS-PAGE was run as described in Section 3. 2. 3. 10. XCell II Blot Module (Invitrogen) was used to transfer proteins from the SDS-PAGE gel onto the PVDF membrane (Bio-Rad). Transfer module was packed with 3 blotting pads and a filter paper (Bio-Rad) pre-soaked in transfer buffer (1x NuPage Transfer buffer, 10 % methanol). The membrane was soaked for 10 min in methanol, 5 min in transfer buffer and then placed on top of the SDS-PAGE gel and in the transfer module. Another filter paper was placed on top of the membrane, followed by another two pads, also pre-soaked in the transfer buffer. Transfer was performed at 30 V for 1 h. After the transfer, the membrane was placed in the Falcon tube containing 40 mL of 5 % (w/v) skim milk (Fluka Analytical) in PBS and incubated on the rollers for 2 h at 4 °C. The membrane was transferred to a new Falcon tube, containing 10 mL of milk in PBS with a 1:200 dilution of primary antibodies (N-18, Santa-Cruz Biotechnology) and placed on the rollers at 4°C for overnight incubation. On the following day, the membrane was washed in 10 mL of PBS for 20 min, twice in 10 mL of 0.05% (v/v) Tween20 in PBS for 20 min and twice in 10 mL of PBS for 20 min. After the washes the membrane was placed in 10 mL of milk in PBS with a 1:5000 dilution of secondary antibodies (donkey anti-goat IgG-HRP, Santa-Cruz Biotechnology) and incubated on the rollers for 2 h at 4 °C. Following the incubation, the membrane was washed twice in 10 mL of PBS for 5 min, once in 20 mL of 0.05 % Tween20 in PBS for 20 min, 10 mL of PBS for 10 min and 10 mL of PBS for 5 min. The membrane was stained for 30 min in 3,3'-diaminobenzidine (Sigma).

3. 2. 4. Results and discussion

3. 2. 4. 1. Construct design and plasmid preparation

Experiments described in this section aimed to express the recombinant L- and P-selectin constructs in *P. pastoris* strain KM71H. The initial tests attempted to find an optimal protocol for expression of these constructs using standard growth media. Had these attempts been successful, the next step would be to find suitable conditions for expression of the ^{15}N, ^{13}C – labelled protein in minimal media. The protocols described here were kindly provided by the Barlow group (University of Edinburgh).

The double-domain constructs were based on the constructs expressed previously by other groups for X-ray studies[65,66]. The preferred option was to express the protein constructs without any added tags as these can be problematic to cleave off and can interfere with NMR studies. For this reason, the only addition to the native protein sequence was the N-terminal α-factor secretion signal sequence. This sequence is necessary for protein secretion into the extracellular media and should be cleaved off by *P. pastoris* proteases prior to translocation through the cell membrane.

The α-factor sequence adds sixty-six amino acid residues to protein's N-terminus. There are two proteases that are responsible for the processing of this sequence and their recognition sites are also encoded within the pPICZαB vector. The first protease is a dibasic endopeptidase encoded by the *Kex2* gene that cuts the peptide between arginine and glutamic acid within the terminal region of the α-factor signal sequence (Fig. 3. 2. 2). The second protease, a product of the *Ste13* gene, is a dipeptidyl amino peptidase that removes the remaining glutamic acid and alanine residues (Fig. 3. 2. 2). Successful post-translational processing of the α-factor sequence is dependent on correct function of both of these proteases. However, in some cases the *Ste13* protease displays low efficiency and can leave a permanent two-residue tail at the protein's N-terminus[117].

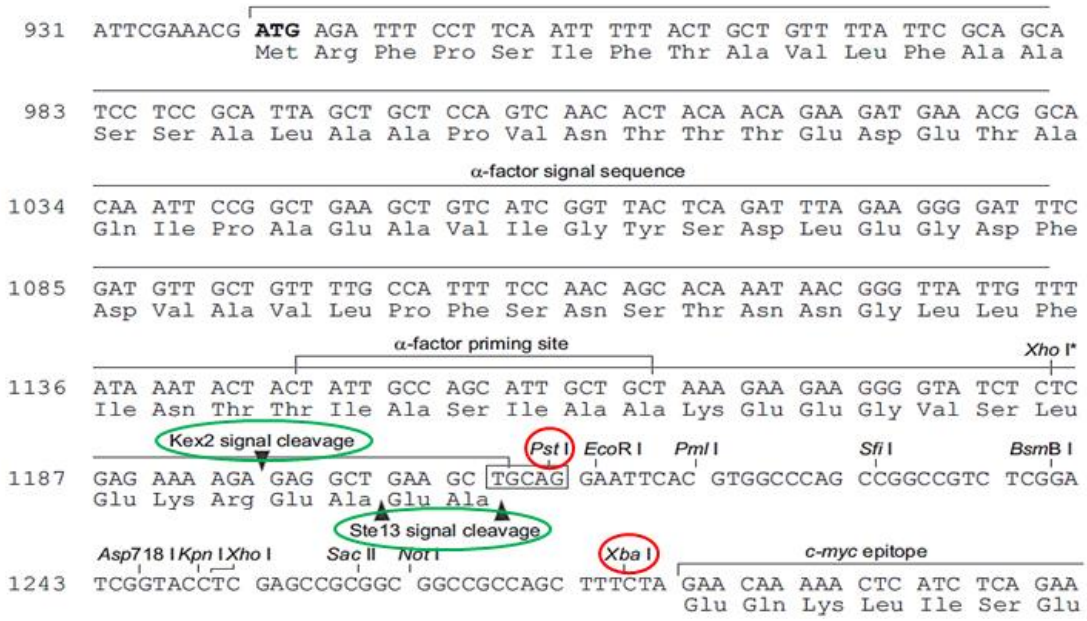


Fig. 3. 2. 2. Fragment of the pPICZαB vector genetic sequence. The red circles highlight the *PstI* and *XbaI* restriction sites used to for the cloning of selectin constructs into the vector. The green circles highlight the protease cleavage sites, necessary for the correct cleavage of the α-factor signalling sequence.

Selectin constructs were designed to be cloned into the pPICZαB vector with the use of two restriction sites: *PstI* at the 5' (N-terminal) site and *XbaI* at the 3' (C-terminal) site of the protein sequence (Fig. 3. 2. 2). Although these restriction sites are required to insert the desired gene into the vector, the introduction of six extra bases onto the 5' and 3' ends of the gene sequence leads to addition of two amino acids onto protein's N- and C-termini. These amino acids cannot be cleaved off and in some proteins can interfere with the folding processes and affect protein's stability. However, similar approach has been used for successful expression of many proteins and this design was regarded as the most promising option for selectin construct expression NMR studies at the time.

The genes were synthesised and cloned into pPICZαB vector externally by GenArt™. The resulting plasmids were transformed into *E. coli* cells to generate high amounts of plasmid DNA for expression tests. The extracted DNA was sequenced and no mutations were found within any of the constructs. For efficient uptake by the *P. pastoris* cells, the plasmids needed to be linearised, using the SacI-HF restriction enzyme. The difference between the circular and linearised plasmid can be observed by agarose gel electrophoresis, where the linear plasmids will behave as larger molecules and will migrate through gel matrix more slowly than the circular plasmids. The majority of circular plasmids adopt a twisted coil conformation, making them appear smaller on the agarose gel. As presented on Fig 3. 2. 3, all of the plasmids were linearised successfully and migrated to the expected region on the gel (ca. 4 kb) after linearization.

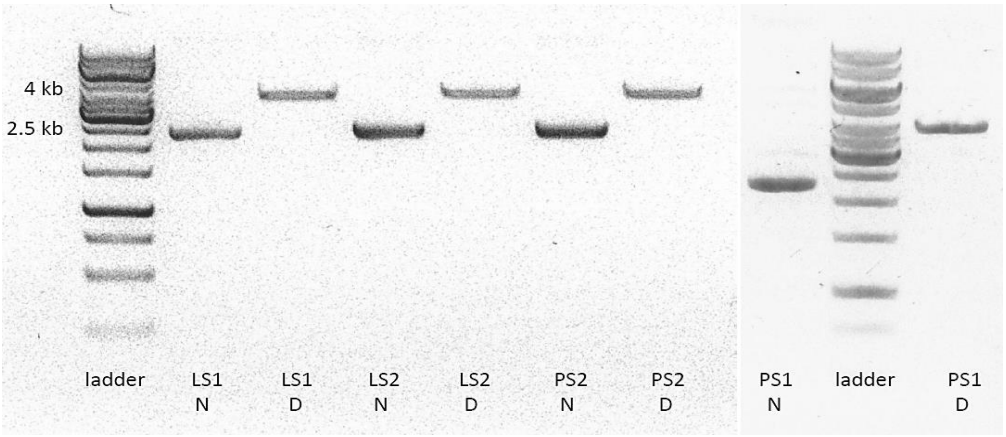


Fig. 3. 2. 3. Linearisation of the plasmid DNA. Agarose gels of the LS1, LS2, PS2 and PS1 plasmid DNA before and after the linearisation reaction. Letter N indicates the DNA sample prior to incubation with the SacI-HF enzyme, letter D indicates the sample following the digestion reaction.

In the next step, the linearised DNA was transformed into *P. pastoris* by electroporation. The transformed cells formed colonies on agar plates containing up to 300 μg/mL of zeocin, proving that the transformation was successful.

3. 2. 4. 2. Small-scale protein expression tests

In the initial experiments, small-scale expression tests were performed on all of the constructs, aiming to identify the colonies that express the protein in the highest amounts. For this purpose, 100 mL cultures of individual colonies were grown, induced with 0.5% of methanol for 3 – 4 days and harvested by centrifugation. As the protein should be secreted by the cells, the supernatant was analysed by the SDS-PAGE. The expected molecular weight (MW) of the LS1 and PS1 constructs was approximately 14.5 kDa and of the LS2 and PS2 constructs – 18.6 kDa, without taking glycosylation into account. Summary of the calculated MW and isoelectric point (pI) values for all constructs expressed in *P. pastoris* can be found in Appendix 1.

Analysis of protein expression levels was hindered by the large amount of smearing observed on the SDS-PAGE (Fig. 3. 2. 4). The smearing appeared to be correlated to the amount of recombinant protein as it increased at longer induction times and was the most intense in the regions where selectin band was expected. Reasons for the smearing remain unclear but could be explained by the fact that the constructs may be forming complexes with other molecules such as nucleic acids or carbohydrates. Indeed, absorbance at 260 nm was measured and proved to be higher than the absorbance at 280 nm for most samples, indicating presence of nucleic acids. In order to remove any N-linked glycans which could affect protein's migration on the gel and contribute to the smearing, the supernatant samples were digested with the Endo-HF glycosidase. As the C-type lectin domain contains two N-glycosylation sites, it is possible that there could be two glycan chains attached to this domain when produced in yeast cells. However, treatment with Endo-HF did not improve the clarity of SDS-PAGE results (Fig 3. 2. 4. B).

In experiments undertaken later in the project, attempts were made to reduce the smearing on the SDS-PAGE by modifying the methods for sample preparation and conditions for the gel electrophoresis (Fig. 3. 2. 4. C). These tests included alterations to the amounts of the SDS buffer, the reducing agent and the Endo-HF enzyme, purchasing new reagents, changing

electrophoresis voltage, incubation temperature for the SDS-PAGE samples, diluting or concentrating the samples. Despite many efforts, it was not possible to lessen the smearing or to determine its cause. Further experiments explored different methods for purification of the expressed constructs to obtain a sample that is easier to analyse.

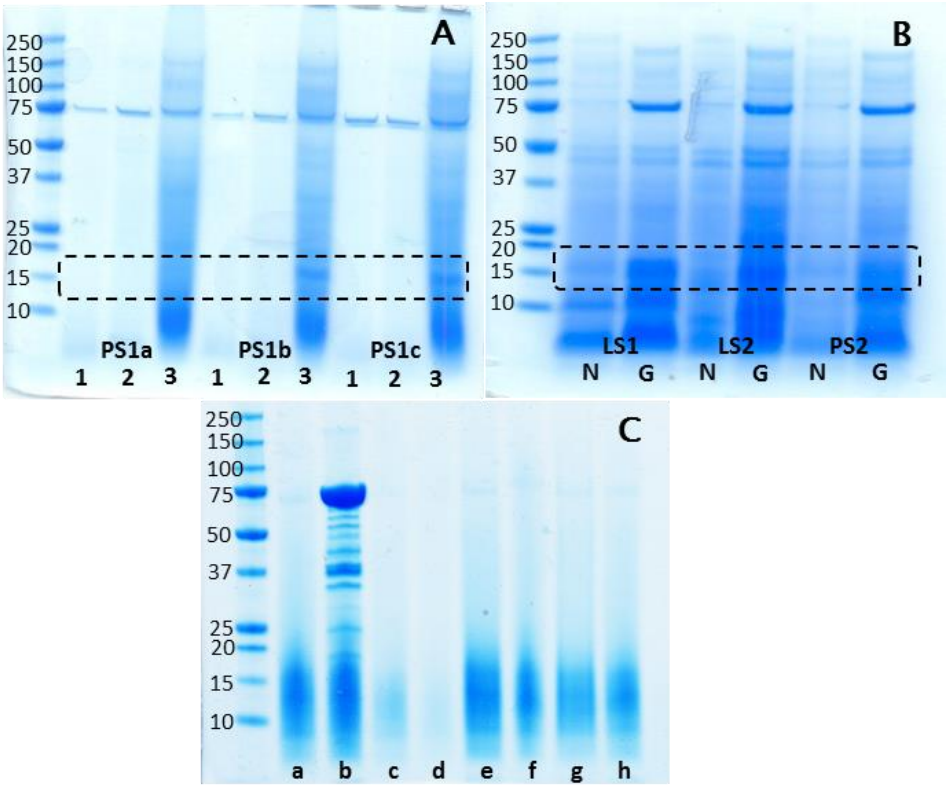


Fig. 3. 2. 4. SDS-PAGE of the cell culture supernatant in small-scale expression tests. **A** SDS-PAGE analysis of crude supernatant for three of the PS1-expressing colonies (a, b and c) in 1, 2 and 3 days after induction. All of the samples were treated with Endo-HF glycosidase and the samples from day 3 were concentrated 5-fold. **B** SDS-PAGE of crude supernatant for LS1, LS2 and PS2 transformants, prior to (N) and after (G) the treatment with Endo-HF glycosidase. All of the samples were concentrated 5-fold. **C** SDS-PAGE of a sample containing the LS2 protein, testing different methods of sample preparation in attempts to reduce the smearing. Sample prepared using standard methods (a) was compared with samples prepared with excess of the Endo-HF enzyme (b), by varying the sample concentration (c), ratios of the reducing agent and the SDS buffer (d – g) and by centrifugation, loading only the supernatant on the gel (h). The first lane is the MW marker in kDa.

By analysing the SDS-PAGE results of the small-scale expression tests, it was not possible to clearly determine whether the protein constructs were expressed in *P. pastoris*, how high the yields were and whether the proteins were folded or have undergone degradation. A sensible

next step was to express one of the protein constructs on a larger scale (4 L), in order to obtain higher amounts of the supernatant sample for more detailed analysis. As all of the test expression cultures were similar by SDS-PAGE, no priority was given to any of the constructs.

3. 2. 4. 3. Large-scale expression tests

The construct that was expressed in larger quantities was the double-domain L-selectin construct (LS2). Cells were cultured and induced as described previously and the collected supernatant was separated on the SP Sepharose resin in potassium phosphate buffer, pH 6.0. The calculated pI of the LS2 construct is close to 7.0 which meant that it should be possible to bind the protein to a cation exchange resin at pH 6.0. However, no protein was detected to bind to the column at this pH. Following this result, pH of the buffer and the supernatant was lowered to 4.0 to ensure that the protein binds this time. Indeed, at the lower pH, a large peak was detected in the fractions eluted from the column at approximately 0.3 – 0.7 M concentration of sodium chloride (Fig. 3. 2. 5. A).

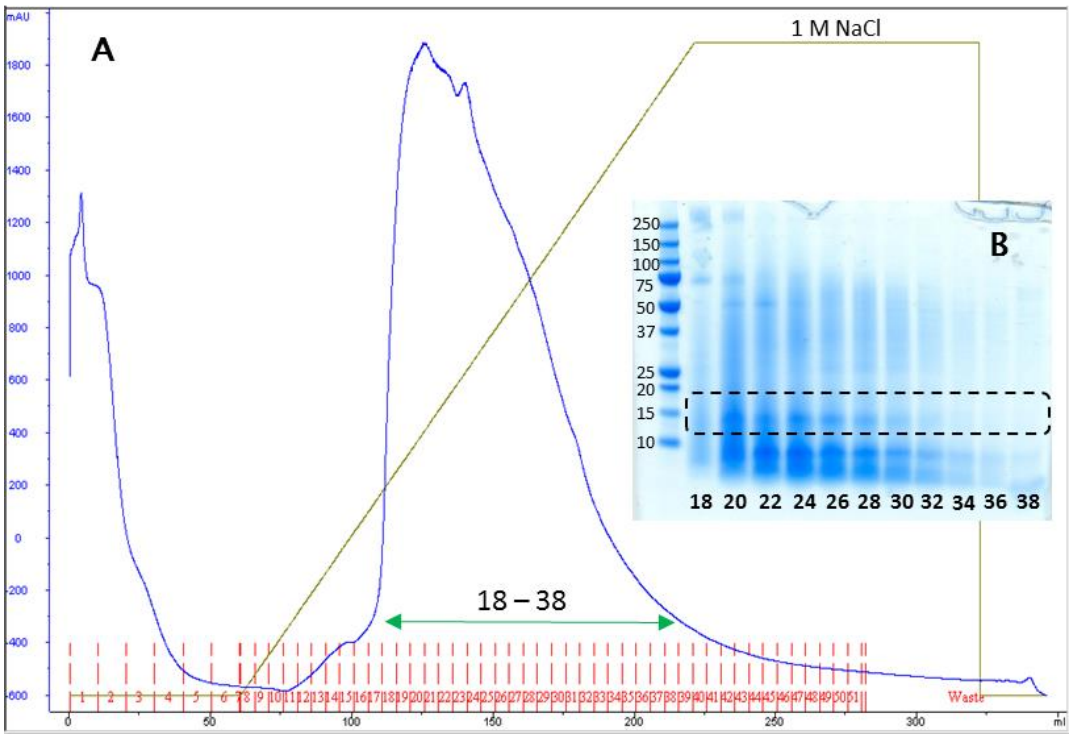


Fig. 3. 2. 5. SP Sepharose cation exchange chromatography of the LS2 supernatant from the large-scale growth. A Chromatogram of the fractions eluted with the salt gradient from 0 M to 1 M NaCl. Fractions 18 – 38, analysed by the SDS-PAGE, are indicated by the green arrow. **B** SDS-PAGE of selected fractions with fraction numbers underneath. Dotted line indicates the region where the LS2 band was expected. The first lane is the MW marker in kDa.

Fractions 18 – 38 were analysed by SDS-PAGE (Fig. 3. 2. 5. B). A band appearing at MW close to 15 kDa was considered to be the one that most likely corresponded to the LS2 construct. One of the fractions containing this band (fraction 24) was used in a test gel filtration chromatography experiment. The other fractions containing this band (20 – 31) were combined and stored for future analysis at -20 °C.

Gel filtration chromatography was performed in phosphate buffered saline (PBS) at pH 7.4. Several broad and merging peaks were observed (Fig. 3. 2. 6. A). A range of fractions from different regions of the chromatogram was analysed by SDS-PAGE. Fractions 40 – 90 contained no bands that could correspond to the full-size LS2 construct – only smaller size bands were observed, which could indicate that protein was degraded prior to or during this purification step. Bands of ca. 15 kDa and 20 kDa were observed in fractions 32 – 38 (Fig. 3. 2. 6. B). These fractions were combined and stored for further analysis at -20 °C.

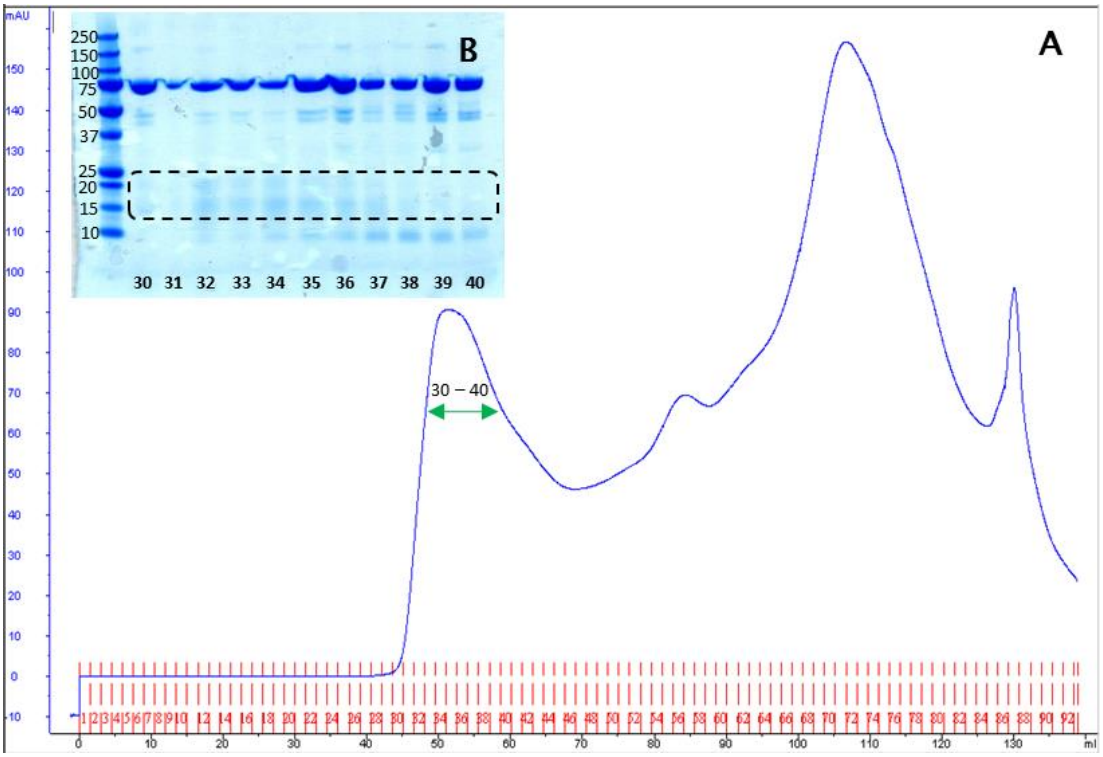


Fig. 3. 2. 6. Gel filtration chromatography of the fraction 24 eluted from the cation exchange resin. A Chromatogram of the fractions eluted from the gel filtration column. Fractions 30-40, analysed by SDS-PAGE, are indicated with the green arrow. **B** SDS-PAGE of selected fractions with fraction numbers underneath. Dotted line indicates the region where the LS2 band was expected. The first lane is the MW marker in kDa. The strong band at ca. 75 kDa, as well as two weaker bands between 37 – 50 kDa resulted from the excessive amount of Endo-HF enzyme used in the preparation of samples for SDS-PAGE.

Bands observed at lower MW in both experiments were considered to be LS2 fragments, produced by proteolytic cleavage. No protease inhibitors were used during this large scale protein preparation and purification, which may have caused protein degradation during and after the culture growth. In order to minimise this problem in the future experiments, PMSF was added to the supernatant after the harvest. EDTA was not used as a protease inhibitor because of its chelating affinity for divalent cations. It was expected that presence of Ca^{2+} ions may be important for stability and binding properties of the C-type lectin domain. For this reason, it was also decided, that in the further experiments, use of phosphate buffers, which also sequester calcium ions, would be avoided. Sodium acetate buffers were used instead from that point onwards.

Due to the problems with identification of the LS2 bands on SDS-PAGE, western blotting experiments were performed on the fractions from both the cation exchange and the gel filtration chromatography experiments. A single band was detected only in the combined fractions from the SP Sepharose chromatography and not in the fractions collected from the gel filtration. The band appeared at ca. 20 kDa, which was close to the estimated MW of 18.6 kDa (Fig 3. 2. 7). It should be noted that such band was not observed on any of the SDS-PAGE gels, suggesting that the protein may be present in very small quantities and its detection may only be possible by the western blot. This, in turn, would suggest that there is a problem with expression of this construct in *P. pastoris* as it is not being produced in high enough yields, and that modifications to culture growth conditions should be explored.

Fig. 3. 2. 7. Western blot of the fractions from chromatography experiments. The first two lanes correspond to the fractions collected after the gel filtration (GF). The next two lanes correspond to the fractions collected after the cation exchange chromatography (CE). The positive control (PC) was a cell extract containing the full-size L-selectin – the proteins did not transfer well onto the membrane but a strong band corresponding to the full-size L-selectin can be observed on top of the MW marker in the final lane.

As the combined fractions from cation chromatography were shown to contain the LS2 construct by western blot, another gel filtration experiment was performed, using a 10-fold concentrated sample. The conditions of this experiment were the same as previously, except for the higher amount of the loaded protein. However, since in the previous experiment the collected fractions did not prove to contain the desired construct, other regions of the gel filtration chromatogram were going to be explored.

The results were similar to those from the previous experiment –broad overlapping peaks were present on the chromatogram and the SDS-PAGE analysis, although cleaner than before, did not show any presence of the LS2 construct. Only small protein fragments were observed, strengthening the hypothesis that the protein was degraded prior to or during the chromatography experiment (Fig. 3. 2. 8). After 5-fold concentration of some of the fractions, the protein could be observed on the gel but only as a weak band that disappeared from the gel after a two hours of staining, indicating that the background noise was too high for detection of the very small amounts of the LS2 construct in the sample.

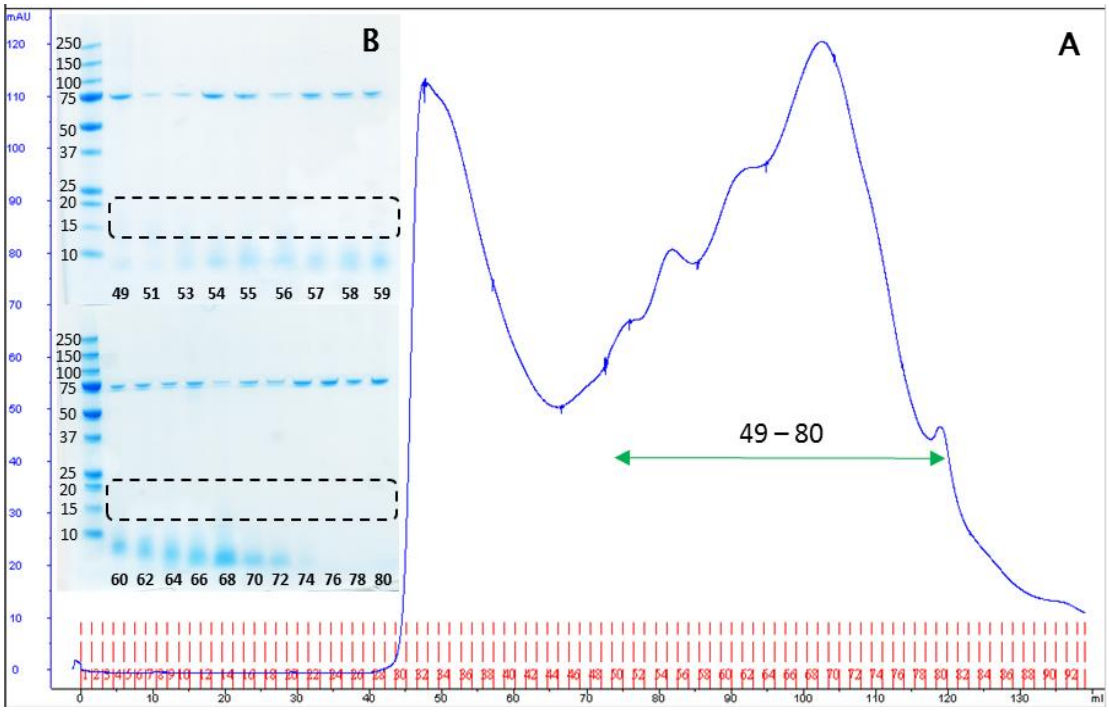


Fig. 3. 2. 8. Gel filtration chromatography on combined fractions 20 – 31 from cation exchange chromatography. A Chromatogram of the fractions eluted from the gel filtration column. Fractions 49-80, analysed by SDS-PAGE, are indicated with green arrow. **B** SDS-PAGE of selected fractions with fraction numbers underneath. Dotted line indicates the region where the LS2 band was expected. The first lane is the MW marker in kDa.

Interestingly, the small protein fragments (< 10 kDa) were present throughout the majority of collected fractions. This should not be the case in a correctly calibrated and well-functioning gel filtration column – all low-MW fragments should be eluted in the final stages of the experiment, after the larger proteins were eluted. This raised suspicions that there may be technical issues with the column that interfere with the result of this experiment. The exact reason for these problems was never confirmed.

The three most prominent low-MW contaminants were analysed by mass spectrometry and one of them was identified to be a fragment of the degraded construct. The identified peptide corresponded to residues 98-111 of the human L-selectin. The other fragments could not be identified but it is likely that they were also fragments of the LS2 construct.

3. 2. 4. 4. Small-scale protein purification tests on the LS2 construct

Further small-scale expression experiments were undertaken in attempt to establish a more suitable protocol for LS2 construct expression and purification. The aim of these tests was to increase the yields of produced protein and to minimise its degradation. In these experiments, cell cultures were fed with higher than previously amounts of methanol (1%) and PMSF was added to the supernatant immediately after harvest to prevent proteolysis.

Unfortunately, the LS2 construct, when expressed on a small scale, presented high amount of smearing on all SDS-PAGE gels, which hindered protein identification in the chromatography tests. Smearing in this case was so severe, that no individual bands could be observed. Although protein purification was attempted on SP Sepharose column pH 4.0 and pH 5.0, as well as on the Heparin column at pH 5.0, no clear bands could be observed below 20 kDa. Resins from both columns were also tested by SDS-PAGE for presence of any proteins that may still be bound after the elution with 1M NaCl. No bands were observed.

An interesting observation came from the comparison of the fractions eluted from the SP Sepharose resin under the Ca^{2+}-rich (addition of $CaCl_2$) and Ca^{2+}-deficient (presence of EDTA) conditions. Peaks in the UV absorbance for those two tests occurred at similar time points: in fractions 3 – 6 and 7 – 12. However, the levels of absorbance for the two observed peaks were different in the two cases. In the Ca^{2+}-rich environment, A_{280} absorbance for the peak observed in fractions 3 – 6 was higher than in the case of the Ca^{2+}-deficient environment.

Opposite trend was observed for the peak observed in fractions 7 – 12 (Fig. 3. 2. 9). In addition to these findings, analysis by SDS-PAGE showed much reduced smearing for the Ca^{2+}-rich samples, when compared to the Ca^{2+}-deficient fractions. Unfortunately, no clear bands were observed even after concentrating the samples 5-fold, which might be attributed to the small amount of protein present in the fractions. However, these tests clearly show that the concentration of Ca^{2+} ions affects the protein's behaviour in the purification tests and is an important factor to consider in future experiments.

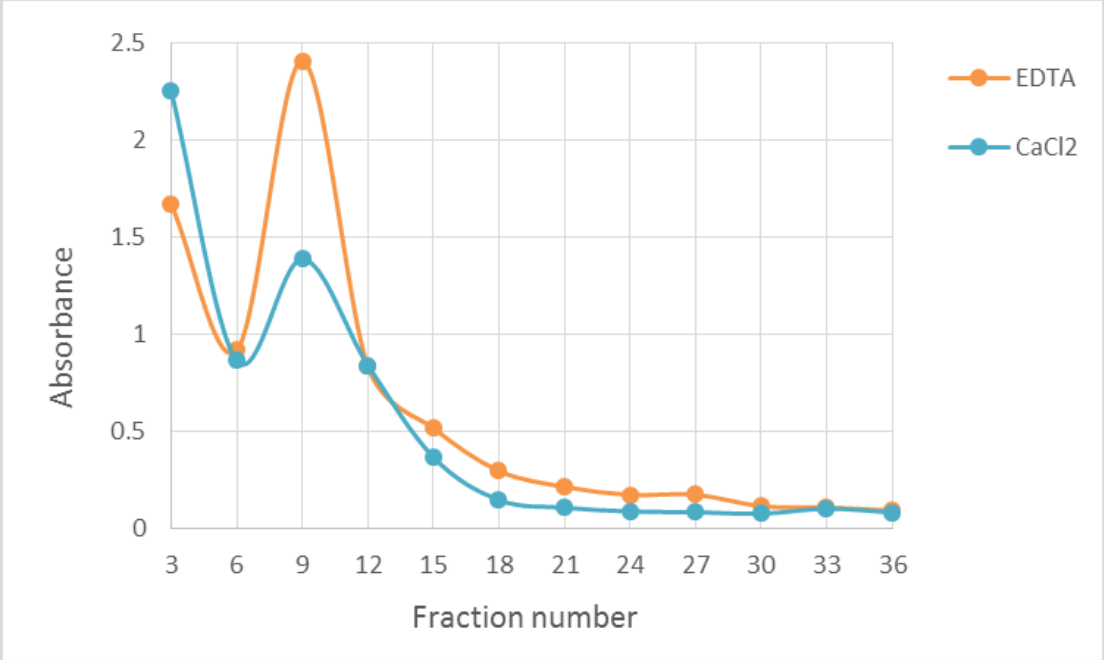


Fig. 3. 2. 9. UV absorbance (A_{280}) of the fractions eluted from the SP Sepharose resin under Ca^{2+}-rich and Ca^{2+}-deficient conditions. Sample of the supernatant from *P. pastoris* LS2 growth was passed through SP Sepharose resin after the addition of $CaCl_2$ (blue line) and EDTA (yellow line). A_{280} of individual fractions was recorded using a bench-top UV spectrophotometer.

Despite the efforts to modify the conditions for cell growth and protein purification, it was not possible to establish a protocol that would allow to obtain the LS2 constructs pure from any contaminants and in high enough yields to allow for further studies.

In the next set of experiments, small-scale expression and purification tests were performed on the single-domain L-selectin construct (LS1).

3. 2. 4. 5. Small-scale protein purification tests on the LS1 construct

The LS1 construct was prepared on a small-scale (300 mL) using the same protocols as described previously, with 1% methanol used for the induction of protein expression and PMSF added to the resulting supernatant to inhibit protease activity. 10 mM calcium chloride was also added to the supernatant to stabilise the protein.

In the first experiment, SP Sepharose resin was tested in sodium acetate pH 4.0. As in the previous tests, the UV absorbance (A_{280}) of the fractions was monitored and selected fractions were analysed by SDS-PAGE. Multiple protein bands were observed in the region between 15 and 20 kDa, indicating that there were many proteins within the *P. pastoris* supernatant binding to the resin at this pH. As calculated pI of the LS1 construct was 8.6, which is higher than the pI of the majority of proteins normally found in the *P. pastoris* supernatant, it should be possible to find experimental conditions in which the LS1 construct binds to a cation exchange resin more selectively.

Three different cation exchange resins were tested at pH 7.0 (SP Sepharose, SP Toyopearl 650M and Heparin Sepharose). No protein was detected in the eluted fractions at this pH. It was therefore decided that a lower pH value should be used at which LS1 construct binds to one of these resins more selectively than at pH 4.0. The arbitrary choice was pH 5.6, lying half-way between the pH 4.0 and 7.0. At this pH, purification on the SP Sepharose resin showed more promising results, with a strong band at ca. 17 kDa, close to the calculated molecular weight of the LS1 construct (14.5 kDa), and with fewer bands eluting at the same time (Fig. 3. 2. 10). However, a strong band was also observed in the column flow-through at 14 kDa and it is possible that this band, rather than the band in eluted fractions, could correspond to the LS1 construct. Due to the unavailability of L-selectin-specific antibodies at the time, it was not possible to confirm which of the two bands corresponded to the LS1 construct. Nevertheless, this purification method was found to be the most suitable from all of the cation exchange methods tested on the small-scale and it was used in the large expression tests.

It is worth noting that the smearing on the SDS-PAGE was greatly reduced in the LS1 samples, showing that this protein behaves differently to the LS2 construct. However, as in the case of the LS2 construct, multiple high intensity bands were observed at low MW values (Fig. 3. 2. 10. B), suggesting that high levels of degradation are observed also for the LS1 construct despite the use of PMSF in all experiments. Culture growth in more controlled conditions

could increase the expression levels and reduce the proteolysis. Therefore, in the next experiment, a larger scale growth in a fermenter (0.7 L) was attempted.

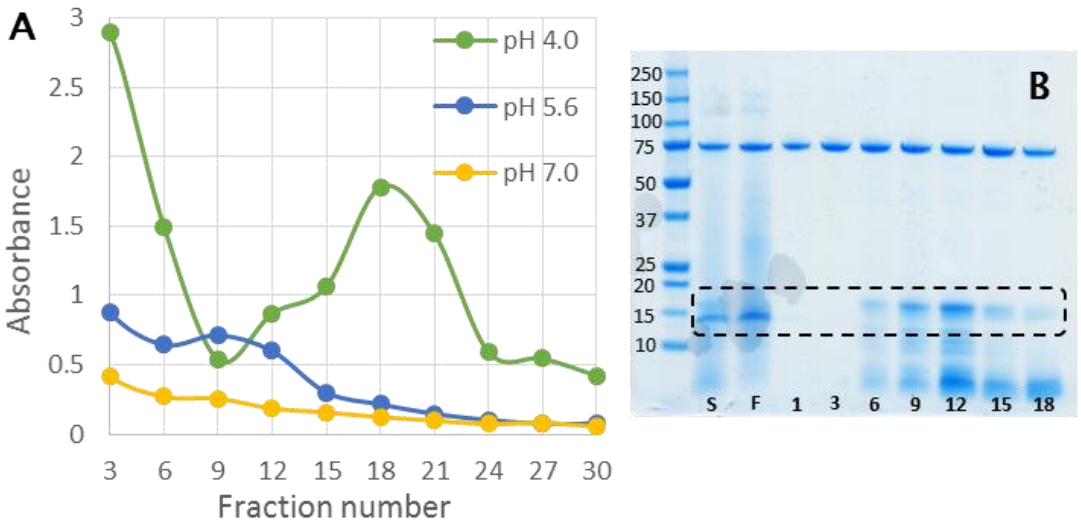


Fig. 3. 2. 10. UV absorbance (A_{280}) of the fractions eluted from SP Sepharose resin at pH 4.0, 5.6 and 7.0 with SDS-PAGE of selected fractions. A LS1 supernatant was applied to the SP Sepharose resin at three different pH values: 4.0 (green line), 5.6 (blue line) and 7.0 (yellow line) and change in A_{280} was recorded using a bench-top UV spectrophotometer. **B** SDS-PAGE of selected fractions eluted at pH 5.6 with fraction numbers underneath. Dotted line indicates the region where the LS1 band was expected. The first lane is the MW marker in kDa. The second lane is raw supernatant after the harvest (S) and the third lane is the column flow-through (F).

3. 2. 4. 6. Fermenter growth of the LS1 construct

In the fermenter growth yeast cells are cultured under controlled conditions, which means that they are able to reach densities higher than in the flask growth, leading to increased protein production levels. During the LS1 fermentative growth, temperature was maintained at 30 °C initially and at 20 °C after the induction with 1% methanol. pH was maintained at 5.0 and dissolved oxygen levels were kept above 20% by constant agitation. Protein expression level in the cell culture media was monitored on each day of the induction. A gradual increase in band intensity was observed for a band at close to 15 kDa which could correspond to LS1, as well as two low-MW bands that likely correspond to protein degradation products (Fig. 3. 2. 11). This result suggests that LS1 construct may be degraded during the culture growth and use of protease inhibitors during the growth may be necessary.

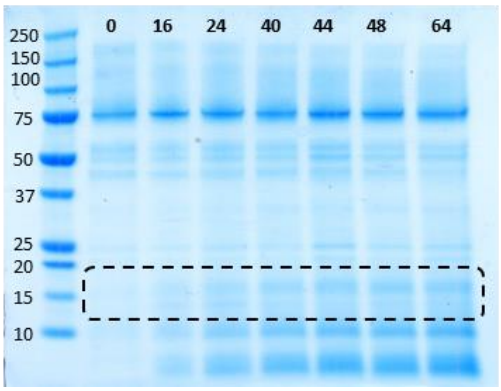


Fig. 3. 2. 11. SDS-PAGE of the *P. pastoris* fermentation media at different time points of the methanol induction. Numbers above the lanes indicate the number of hours after the start of induction. Dotted line indicates the region where the LS1 band was expected. The first lane is the MW marker in kDa.

Culture supernatant was harvested as before and PMSF and 10 mM calcium chloride were added. During the purification, protocols from the small-scale tests were followed. In the first step, the supernatant was loaded onto the SP Sepharose column at pH 5.6. The eluted fractions were monitored by UV and SDS-PAGE (Fig. 3. 2. 12. A). Fractions 17 – 20, 21 – 23 and 24 – 29 were collected into three separate samples that were analysed by the western blot (Fig. 3. 2. 12. B). The highest intensity response was detected in the sample containing fractions 17 – 20. This sample was purified further by gel filtration chromatography.

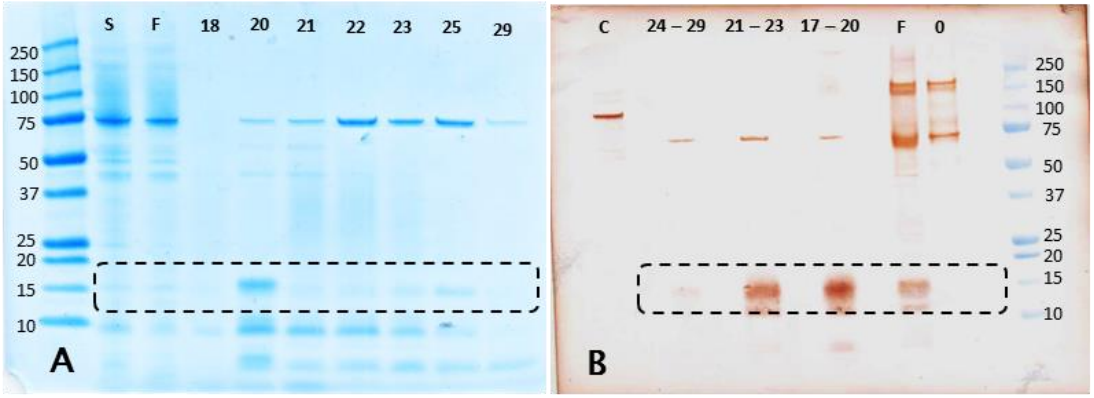


Fig. 3. 2. 12. SDS-PAGE and western blot of the fractions eluted from the SP Sepharose column at pH 5.6. A SDS-PAGE of selected fractions. Numbers above the lanes indicate the fraction number. Dotted line indicates the region where the LS1 band was expected. The first lane is the MW marker in kDa. The second lane is the raw supernatant after the harvest (S) and the third lane is the column flow-through (F). **B** Western blot of the combined fractions 17 – 20, 21 – 23 and 24 – 29. The last three lanes correspond to the column flow-through (F), the sample before induction (0) and the MW marker in kDa.

Gel filtration experiment on fractions 17 – 20 was performed at the Edinburgh Protein Production Facility (EPPF) to avoid technical problems experienced during the small expression tests. The column was equilibrated in 20 mM Tris pH 7.4 with 500 mM NaCl and 10 mM $CaCl_2$ and the sample was concentrated down to 250 µL. Two high-intensity overlapping peaks were observed in the initial part of the chromatogram, between the fractions B3 and B6, and a low-intensity peak later in the chromatogram, between fractions C5 and C6 (Fig. 3. 2. 13. A). These fractions were analysed by SDS-PAGE, however, the bands in the initial analysis were very weak and the samples needed to be concentrated approximately 5-fold. Under the reducing SDS-PAGE conditions all of the fractions contained low-MW proteins. As fractions B3 – B6 were eluted at retention times that would normally correspond to elution of larger proteins, this result suggested that the expressed proteins may be aggregating . For this reason, the same fractions were compared under the reducing and non-reducing SDS-PAGE conditions (Fig. 3. 2. 13. B).

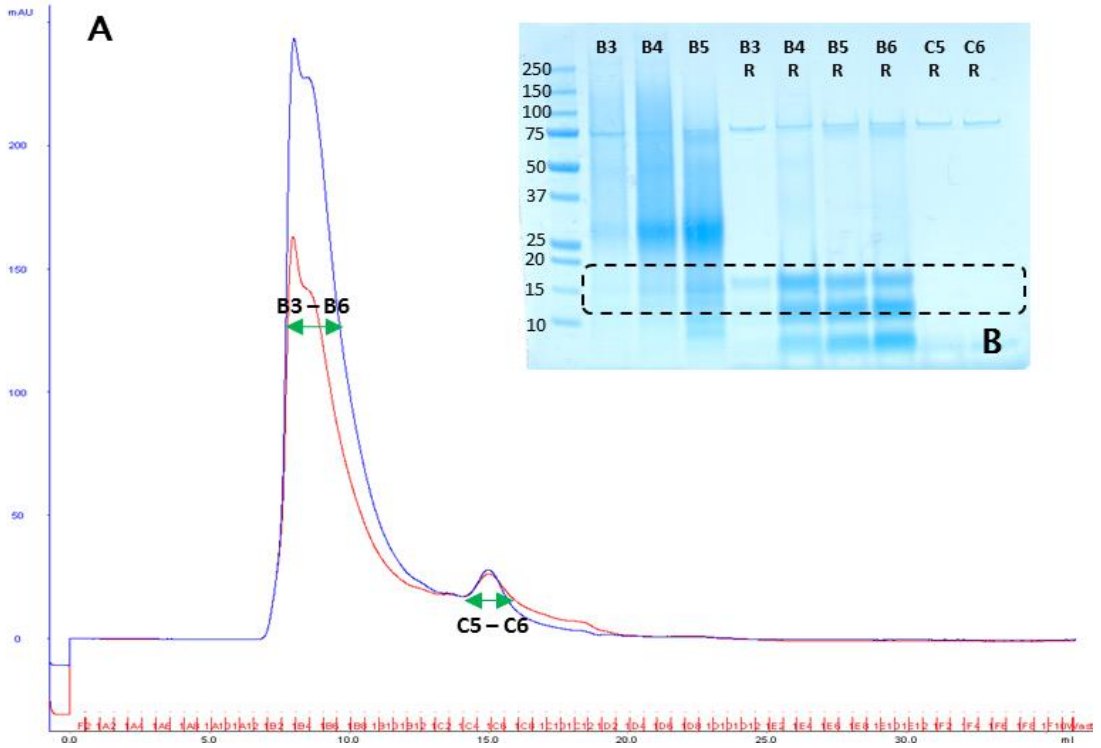


Fig. 3. 2. 13. Gel filtration chromatography on combined fractions 17 – 20 from the cation exchange chromatography. A Chromatogram of the fractions eluted from the gel filtration column (A_{280} in blue, A_{260} in red). Fractions B5 – B6, analysed by SDS-PAGE, are indicated with green arrows. **B** SDS-PAGE of selected fractions with fraction numbers above the lanes. Letter R indicates samples prepared in the reducing conditions. Dotted line indicates the region where the LS2 band was expected. The first lane is the MW marker in kDa.

In the SDS-PAGE experiment it can be clearly seen that the proteins are aggregating and behaving as a larger protein of approximately 28 kDa. Such MW value would suggest that the LS1 construct may form dimers in the buffer used for the gel filtration experiment. In addition to that, high-intensity low-MW bands were observed in all of the analysed fractions, indicating high-levels of protein degradation. It is possible that these low-MW fragments are also aggregating with the full-size protein which would explain why their separation by the gel filtration is not possible.

A few more small-scale experiments were undertaken in order to try and reduce the protein degradation levels. Nine more freshly transformed *P. pastoris* colonies were tested for higher LS1 expression levels with less degradation. However, none of the colonies expressed the full-size construct at a level significantly higher than the colony used in the earlier tests. In another experiment, induction in the presence of 0.1 mmol AEBSF protease inhibitor was attempted. Unfortunately, this experiment also was not successful. It is likely that the high degradation levels occurred due to the construct not being folded correctly and being treated as a misfolded protein by the cells. The reason for this could lie in the construct design – the N-terminal region of the C-type lectin domain may be crucial to the protein's stability. Presence of the α-factor signalling sequence and the additional two amino acids at this site could disturb the folding processes and destabilise the protein. It is therefore possible that *P. pastoris* is not a suitable system for selectin domain expression, at least not with the construct design used in the experiments described here.

3. 2. 5. Conclusions

After numerous test expression experiments of the L-selectin domain constructs in *P. pastoris*, conclusions were made that due to high levels of protein degradation observed in all of the samples, low yields of produced protein and problems with the purification of both the single- and double-domain constructs, expression tests in *P. pastoris* would not be continued. Change of strategy was required and, for this reason, a new expression system was explored in further experiments.

3. 3. Selectin construct expression in *E. coli*

3. 3. 1. Expression system

Escherichia coli has many advantages over other expression systems and is usually the first-choice method for protein production. Bacterial cells are characterised by the fast growth kinetics and high cell density cultures can be achieved within a short time from inoculation. The *E. coli* cells can be easily transformed with the recombinant DNA within just a few minutes. This expression system is also relatively inexpensive, as the culture media can be prepared from a few readily available ingredients.

Despite the differences between the eukaryotic and prokaryotic cells, it is possible to express the eukaryotic proteins in the *E. coli* cells. However, expression of such proteins may require some additional steps, for example redirection of the recombinant protein into periplasm post-translation or co-expression of fusion proteins or chaperones that aid in the protein folding processes. Periplasm is a region between the inner and outer cell membranes that provides oxidising environment necessary for formation of disulfide bridges. However, translocation of protein into periplasm requires a signalling sequence to be attached to protein's N-terminus and, in many cases, this step has a low success rate, leading to accumulation of misfolded intermediates in the cytoplasm[109].

To aid with the problems in disulfide bond formation, new strains of *E. coli* have been developed. For example, the Origami™ strains have mutations in two cytoplasmic enzymes, the thioredoxin reductase (trxB) and glutathione reductase (gor) which disrupt the function of these reductases, enhancing disulfide bond formation in the cytoplasm[109].

When the proteins expressed in bacterial cytoplasm are not folded correctly, they tend to form insoluble aggregates known as the inclusion bodies. Alternative experimental approach utilizes this fact and involves extraction of the inclusion bodies from bacterial cells followed by their solubilisation and refolding. As the inclusion bodies contain almost exclusively the protein of interest, refolding can yield high amounts of pure protein. However, finding the suitable refolding conditions for a specific protein can be a challenging, time and cost-consuming process as there are many variables that can be altered and as every protein construct behaves differently under a given set of conditions.

Acknowledging that there are multiple options for expressing eukaryotic proteins in *E. coli*, a set of experiments was undertaken as part of this project, testing different methods and experimental conditions in order to express correctly folded selectin constructs. These expression tests and their results are described in the following sections of this chapter.

3. 3. 2. Construct design

The vector used in the design of L-selectin constructs was pET26b (+) (Fig 3. 2. 1). This vector features the ColE1 origin of replication for high-copy number of the plasmid, the multiple cloning site (MCS), a promoter recognised by T7 RNA polymerase and the *lac* operon sequence. In this type of vector, the gene of interest is cloned behind the T7 promoter and protein expression can be induced by either lactose or its non-hydrolysable analogue, isopropyl β-D-1-thiogalactopyranoside (IPTG). The second option is usually preferred, as the presence of thiol bond in IPTG structure prevents this compound from being metabolised by the cell and allows to maintain constant concentration for induction of protein expression[122]. The pET26b (+) vector also features kanamycin resistance to selectively grow only the cells that contain the recombinant plasmid.

Two types of L-selectin constructs were designed for expression in *E. coli*, a one-domain construct (LS1), consisting of residues from Trp1 to Ser121, and a two-domain construct, consisting of residues from Trp1 to Glu158. Expression of the P-selectin constructs was not attempted at this stage. As in this expression system the protein constructs need to be purified from cell extracts containing many different proteins, His_6-tag was added to the C-terminus of both constructs.

One of the main issues with using the *E. coli* cells to express selectin domains is the presence of two disulfide bridges in the C-type lectin domain and three disulfide bridges in the EGF-like domain. As mentioned above, bacterial cytoplasm has a reducing environment, in which the formation of disulfide bonds is unfavourable. Along with the inability of bacterial cells to perform post-translational modifications, this issue leads to formation of unstable protein intermediates that aggregate inside the cells, forming the so-called inclusion bodies[112]. For this reason, a C-terminal thioredoxin (Trx) fusion tag was included in the construct design, aiming to enhance protein's solubility and facilitate protein folding processes. Trx is an intracellular protein, native to E. coli that is frequently used to improve the solubility of the

target protein. It is also believed to support formation of disulfide bridges via its intrinsic oxido-reductase activity[123]. The Trx-tag is relatively large in size (11.7 kDa) and the TEV-protease cleavage site was included in the construct to allow for the removal of both the Trx- and His_6-tags from the purified constructs. The NdeI and XhoI restriction sites were used for cloning of the gene constructs into the vector.

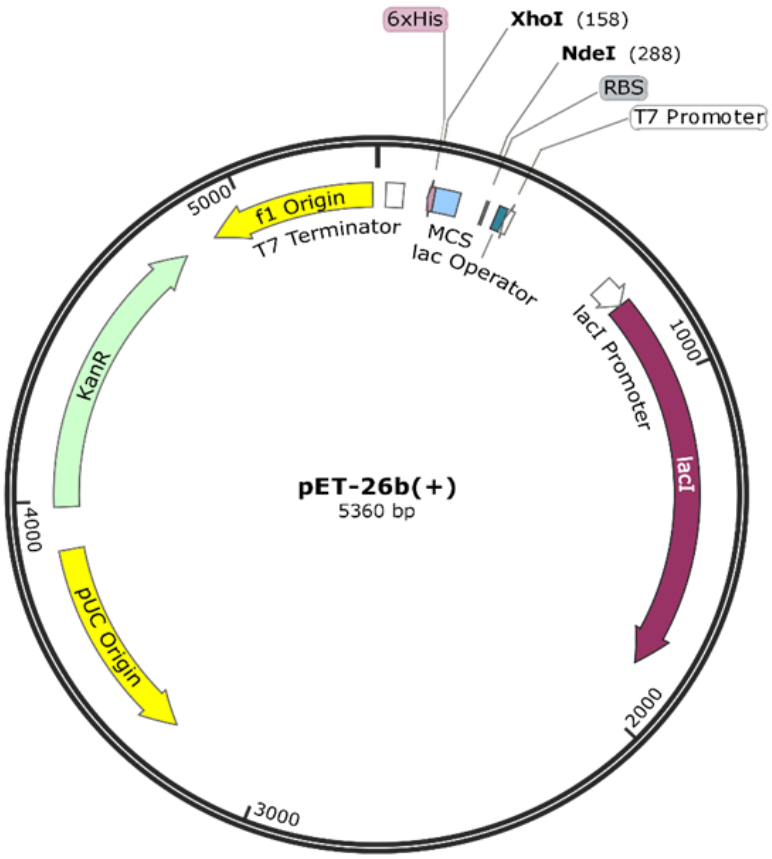


Fig. 3. 3. 1. Map of pET26b (+) vector. The gene of interest is inserted after the T7 promoter, lac operator and the ribosome binding site (RBS), between the NdeI and XbaI restriction sites. Vector also encodes for lacI gene and resistance to kanamycin. Vector image was produced in SnapGene® Viewer.

3. 3. 3. Materials and methods

3. 3. 3. 1. Plasmid preparation

The recombinant DNA fragments were based on the sequence deposited in the Uniprot database (# 14151), codon optimised for expression in *E. coli*, synthesised and subcloned into the pET-26b (+) vector externally by GenScript, with the use of the NdeI and XhoI restriction sites. The received vial was centrifuged at 6000 × g for 1 min at 4°C and the plasmids (4 μg) were resuspended in 20 μL of ultrapure water and vortexed for 1 min prior to transformation into *E. coli* cells.

3. 3. 3. 2. Plasmid transformation into E. coli

1 μL of resuspended plasmid DNA was added into an aliquot (50 μL) of chemically competent Top10 or DHα cells. Cells were incubated on ice for 20 min, heat-shocked for 30 s at 42 °C, returned on ice and diluted with 80 μL of S. O. C. medium. The cells were incubated at 37 °C for 1 h and plated at four different concentrations (5 μL, 10 μL, 20 μL and 40 μL) on 15 mL LB Agar plates with 30 μg/mL of kanamycin. The plates were allowed to dry and incubated overnight at 37 °C.

The next day, transformed colonies (5 colonies per construct) were used to inoculate 5 mL of LB with 30 μg/mL of kanamycin. The cultures were grown overnight at 37 °C, with shaking at 200 rpm, before harvesting by centrifugation at 5000 × g for 10 min at 4 °C. Pelleted cells were then used for plasmid preparation. Plasmids were prepared from the *E. coli* cultures using the QIAgen Plasmid Mini Kit, according to manufacturer's instructions. The final DNA samples were quantified (A_{260}) digested with ClaI enzyme for 30 min at 37°C and separated by electrophoresis on a 1% TBE agarose gel, visualised with GelRed™ (Biotum), with 5 μL of Hyperladder™ 1kb DNA GeneRuler (Thermo Scientific) as reference and sequenced at Edinburgh Genomics Facility.

3. 3. 3. 3. Protein expression tests

BL21 (DE3), BL21 Star (DE3), BL21 (DE3) pLys and LEMO21 (DE3) *E. coli* strains used for the protein expression tests were obtained from the Edinburgh Protein Production Facility. Origami™ (DE3) strain (Novagen) was purchased from Merck Millipore. All strains were transformed with the plasmid DNA using the procedure described in Section 3. 3. 3. 2.

Transformed colonies were used to inoculate 100 mL of LB with 30 μg/mL of kanamycin. In the Origami™ cell cultures 12.5 μg/mL of tetracycline was used in addition to kanamycin. The initial culture was incubated overnight with shaking (200 rpm) at 37°C. On the next day, the culture was diluted into four flasks containing 150 mL of fresh media to OD_{600} values between 0.1 – 0.2. The new cultures were incubated at varying temperatures (RT, 30 °C or 37 °C) with shaking (200 rpm) until the OD_{600} reached levels between 0.6 and 0.8. At this point, isopropyl β-1-thiogalactoparanoside (IPTG) was added to each flask at three different concentrations (0.1 mM, 0.5 mM and 1 mM). One flask was taken aside and incubated in the same conditions as the other flasks but without the addition of IPTG. After the induction was completed, cells were harvested by centrifugation at 5000 × g for 15 min.

3. 3. 3. 4. Preparation of cell extracts and SDS-PAGE analysis

Pellet from 50 mL cell culture was resuspended in 2 mL of BugBuster®, incubated with shaking (200 rpm) for 10 min at room temperature and centrifuged at 5000 × g for 20 min. Supernatant was kept for analysis of soluble protein expression. For the SDS-PAGE analysis, 50 μL of the supernatant was combined with 50 μL of the NuPAGE® LDS Sample Buffer (4X, Invitrogen), heated for 3 min at 100 °C and stored on ice before loaded on the gel.

Pellet was resuspended again in 2 mL of BugBuster® (Novagen). Lysozyme (Sigma) was added to the sample at 1 KU/mL. Solution was vortexed and incubated at room temperature for 5 min. BugBuster® was diluted 1 in 10 with ultrapure water and added to the sample in 6:1 ratio. The sample was vortexed for 1 min and spun at 5000 × g for 15 min at 4 °C. Supernatant was discarded and pellet was resuspended in 10 mL of 1:10 diluted BugBuster® solution and the centrifugation step was repeated. Supernatant was discarded and pellet was resuspended in 1.5 mL of phosphate-buffered saline (PBS). For the SDS-PAGE analysis, 50 μL

of the sample was combined with 50 µL of the NuPAGE® LDS Sample Buffer (4X, Invitrogen), heated for 3 min at 100 °C and stored on ice before loading on the gel.

The SDS-PAGE was run using the precast NuPAGE Novex 4 – 12% Bis-Tris Protein Gels (Invitrogen) with the Precision Plus Protein™ (Bio-Rad) or PageRuler Plus (ThermoFisher Scientific) molecular weight marker.

3. 3. 3. 5. Preparation of the inclusion bodies

In order to prepare the inclusion bodies for the *in vitro* refolding tests, BL21 pLys strain transformed with LS1 plasmid was used to inoculate 100 mL of LB media with 30 µg/mL of kanamycin. The starter culture was incubated overnight with shaking (200 rpm) at 37 °C. Next day, the starter culture was diluted into 2 - 8 flasks containing 500 mL LB with 30 µg/mL of kanamycin to OD_{600} values between 0.2 – 0.3 and incubated with shaking (200 rpm) at 37°C. Once the OD_{600} reached the levels between 0.6 – 0.8, cultures were induced with 1 mM IPTG and incubated for 4 h with shaking (200 rpm) at 37°C. Cells were harvested by centrifugation at 5000 g for 20 min and the weight of the pellet was recorded.

Each pellet from the 500 mL culture was resuspended in 25 mL of the lysis buffer (50 mM Tris, pH 8.0, 250 mM NaCl, 2 mM EDTA, 0.1% Triton™ X-100, 0.1 mM PMSF) and sonicated for 10 – 15 min with 30 s intervals. When the sonication was completed, samples were treated with DNAse (0.01 mg/mL, New England BioLabs) and lysozyme (0.1 mg/mL, Sigma) and incubated for 20 min at room temperature. The insoluble fraction was harvested by centrifugation at 5000 × g for 15 min, supernatant was discarded and pellet was resuspended in the lysis buffer and sonicated as before. Centrifugation step was repeated, but this time pellet was resuspended in the lysis buffer without Triton™ X-100. Samples were sonicated again and the insoluble fraction was harvested by centrifugation as before.

3. 3. 3. 6. Solubilisation and purification of inclusion bodies

Each pellet prepared from 500 mL culture was resuspended in 25 mL of solubilisation buffer (6 M guanidine-HCl, 250 mM NaCl, 20 mM Tris, pH 8.0, 10 mM $CaCl_2$) with 20 mM β-mercaptoethanol (β-ME, Sigma). Samples were incubated on a mixing rotor for 1 h at room temperature and centrifuged at 5000 × g for 10 min. Supernatant was transferred to a new sample tube and centrifuged again. The new supernatant was transferred to another sample tube, passed through a 0.45 μM filter and incubated overnight with 2 – 3 mL of HisPur™ Ni-NTA resin (Thermofisher Scientific) on the rollers at 4°C.

Next day, the resin was transferred to a 20 mL disposable column and flow-through was collected. Resin was washed with 20 mL of the solubilisation buffer and 20 mL of equilibration buffer (8 M urea, 250 mM NaCl, 20 mM Tris, pH 8.0, 10 mM $CaCl_2$). The first batch of proteins was eluted in 10 mL of the equilibration buffer containing 0.2 M imidazole, the second batch in 10 mL of the same buffer containing 0.5 M imidazole, and the third batch was eluted in 10 mL of the same buffer containing 1 M imidazole into fractions of 1.5 mL each. To elute any remaining proteins, resin was incubated in 10 mL of the buffer containing 1 M imidazole for 10 min. To ensure that all protein was bound and eluted from the resin, the resin was washed with the solubilisation buffer and incubated again overnight with flow-through collected after the first incubation. Next day, all purification steps were repeated.

Protein concentration in the individual fractions was estimated from A_{280}, using the corresponding buffer as a blank. Fractions were analysed on SDS-PAGE and fractions containing the recombinant protein were combined into one sample.

3. 3. 3. 7. Protein refolding tests by infinite dilution

For the refolding tests using the dilution method, collected protein fractions were diluted in the refolding buffer to less than 10 μM concentrations. None of the refolding solutions exceeded the volume of 250 mL. Dilution was performed slowly, adding the protein dropwise, with 15 – 20 s intervals between every drop and with fast, constant stirring. Different variants of the refolding buffer were tested, all containing 20 mM Tris (pH 8.0), 100 – 250 mM NaCl and 10 mM $CaCl_2$. One Pierce™ Protease Inhibitor Tablet (EDTA-free,

Thermofisher Scientific) was added to each refolding solution. The optional components of the buffer included urea (1 – 3M), arginine (500 mM), DTT (2 – 5 mM), reduced glutathione (1 – 35 mM), oxidised glutathione (0.1 – 1 mM), 5% (v/v) glycerol. The refolding solution was incubated at room temperature for up to 72 h, filtered and then transferred to a SnakeSkin™ Dialysis Tubing (Thermofisher Scientific) and dilalysed three times against 10:1 volume of the refolding buffer without the additives. At all stages, samples (1.6 mL) were collected for TCA precipitation to monitor the efficiency of the refolding reaction.

3. 3. 3. 8. On-column protein refolding test

Harvested cell pellets were resuspended in ultrapure water (5 mL per 1 g pellet) and sonicated for 30 min using 15 s intervals. Sample was centrifuged at 5000 × g for 30 min and the supernatant was removed. Pellet was resuspended in a 10 mL of solution of a few grams of methionine and 1mM EDTA, sonicated for 5 min using 15 s intervals and centrifuged again. Methionine wash was repeated two times. Five (5 – 10 mL) volumes of 1:1 mixture of acetone and water with 0.2% TFA were added to the pellet and the sample was incubated overnight. Next day, 2 mL of HisPur™ Ni-NTA resin (Thermofisher Scientific) was placed in a 20 mL disposable column and equilibrated with 20 mL of the binding buffer (100 mM Tris, pH 8.0, 250 mM NaCl, 10 mM $CaCl_2$). A range of buffers with varying urea concentrations (0 M, 1 M, 2 M, 3 M, 4 M, 5 M, 6 M, 7 M and 8 M) were prepared from the binding buffer. 10 mL of each of these buffers was passed slowly through the column, starting with the 8 M urea concentration and finishing with the 0 M urea concentration. Protein was eluted and analysed as described in Section 3. 3. 3. 7.

3. 3. 3. 9. TCA precipitation

In order to analyse the samples from protein refolding experiments, TCA precipitation was performed to remove the denaturing agents and concentrate the protein. For each sample, 20 µL of the sample was mixed with 980 µL of water and 100 µL pre-chilled 100% trichloroacetic acid (TCA) and incubated on ice for 1 h. Samples were centrifuged at 14 000 rpm for 15 min at 4 °C, supernatant was discarded and pellet was washed with 100% cold

acetone. Samples were centrifuged again for 5 min, supernatant was discarded and pellet was air dried. 15 µL of the NuPAGE® LDS Sample Buffer (4X, Invitrogen) was added to the pellet and mixed with the pipette. For the reduced samples, 2 µL of the NuPAGE® Sample Reducing Agent (Invitrogen) was added to the sample and the sample was heated at 100 °C for 3 min.

3. 3. 4. Results and discussion

3. 3. 4. 1. Construct design and plasmid preparation

In the experiments described here, expression of L-selectin constructs in E. coli cells was attempted. As in the case of the earlier expression tests in *P. pastoris*, the major aim of the tests described here was to find a method that could be optimised to allow for expression of a doubly labelled protein for NMR experiments.

The L-selectin construct design for expression in *E. coli* was similar to that used for expression in *P. pastoris* and to the constructs expressed previously for the X-ray studies[65,66]. However, in this study, addition of any tags onto the protein's N-terminus was avoided. Instead, a fusion protein was added to the C-terminus of the constructs, in order to target the challenges associated with expression of human proteins in the prokaryotic system. As described in section 3. 3. 2, the fusion protein chosen here was thioredoxin (Trx) – a protein tag widely used to enhance solubility and stimulate disulfide bond formation in the expressed protein. Because the fusion protein could interfere with the NMR studies, construct was designed in such a way that the Trx tag, along with the His_6-tag, could be removed by TEV protease, once all the expression and purification steps are completed.

Cloning reactions were performed externally (GenScript) and the resulting plasmids were transformed into *E. coli* cells to generate high amounts of plasmid DNA for expression tests. For each recombinant strain, five colonies were tested for the presence of the inserted plasmid, using ClaI digestion enzyme. Each plasmid has exactly one recognition site for this enzyme and digestion with ClaI should linearise the plasmids. The difference between the circular and linearised plasmid can be observed on agarose gel as the plasmids will migrate

differently through gel matrix. Circular plasmid will most likely occur in the supercoiled conformation, which is the native conformation of plasmids *in vivo*. In this conformation, the DNA is tightly twisted which causes it to behave like a smaller DNA fragment and to travel more quickly across the agarose gel. In some cases DNA can be nicked on one strand and the twists may be relieved, forming more open conformation. In such cases, the plasmid will travel through the gel matrix very slowly, appearing to be of size slightly higher than expected. As presented on Fig 3. 3. 2, both forms of plasmid DNA were observed in the different preparations from E. coli. When digested with ClaI enzyme, plasmids migrated on the agarose gel as expected for the linear DNA fragment of the length of between 5700 – 5800 bp. Because the supercoiled form of the plasmid is more efficient in transforming the bacterial cells than the nicked form, only samples containing mostly the supercoiled DNA were used for transformations of *E. coli* cells for the protein expression tests.

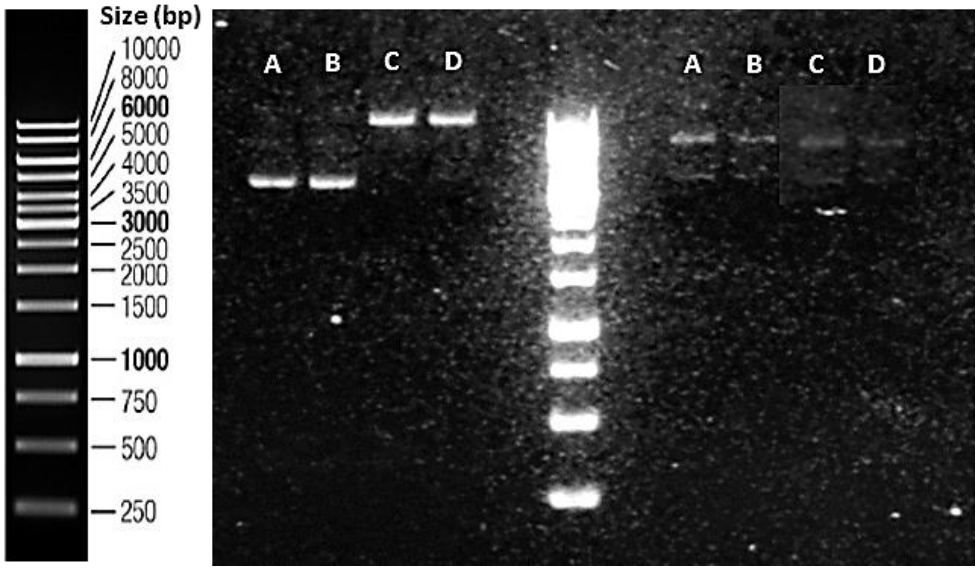


Fig. 3. 3. 2. Linearisation of the plasmid DNA. Agarose gels of four different DNA samples from the Mini prep for the LS1 construct, before (on the left from the ladder) and after (on the right from the ladder) the linearisation reaction. Two different forms of DNA were observed – the supercoiled form in samples A and B and the nicked form in samples C and D. All fragments migrated to the same region on the gel after linearisation.

3. 3. 4. 2. Soluble protein expression tests

In the initial set of expression tests, attempts were made to express well-folded, soluble L-selectin constructs in the *E. coli* cytoplasm. The majority of experiments focused on the single domain construct because the lectin domain has only two disulfide bridges and is found in

soluble form in the bloodstream. Therefore, it should be easier to express it as a soluble protein within the bacterial cytoplasm than the more complex two-domain construct which contains five disulfide bridges. The soluble protein could be extracted and purified from the cells using a His-Trap resin and, following the removal of the fusion tag, used for the structural NMR studies.

In the first experiment, three *E. coli* strains were transformed with the recombinant L-selectin DNA – BL21 (DE3) and its two variants, BL21 Star (DE3) and BL21 (DE3) pLys. BL21 (DE3) is one of the *E. coli* strains most commonly used for expression of recombinant proteins. BL21 Star (DE3) is a similar strain but with enhanced ability to express high levels of protein due to enhanced mRNA stability. BL21 (DE3) pLys is another variant that increases expression efficiency for proteins that are under control of the T7 promoter.

In order to test different induction conditions, three different IPTG concentrations (0.1 mM, 0.5 mM and 1 mM) were used for induction of protein expression in each strain. Expression levels of the LS1 construct in the soluble and insoluble fraction of the cell extract were analysed by SDS-PAGE after 3 h from induction at 37 °C (Fig. 3. 3. 3. A, B and C). Calculated molecular weight of the construct was 28 kDa (see Appendix 2) and in all of the strains a strong band was observed in the corresponding region on the SDS-PAGE. However, the band was only observed in the insoluble fraction of the cell extract which means that the protein was not folded correctly and aggregated in inclusion bodies. Construct was expressed at all IPTG concentrations, with no significant increase in expression levels at higher IPTG concentrations. The LS1 band was not observed in the extract from cells that were not induced with IPTG.

Following these results, expression in Origami™ strain was attempted using identical conditions. As the Origami™ strain enhances production of disulfide bridges, expression in this strain could produce higher amounts of fully-folded, soluble protein. After 3 h from induction, protein expression levels in soluble and insoluble fractions were analysed by SDS-PAGE (Fig. 3. 3. 3. D). However, the results were similar to those observed in other bacterial strains – the LS1 band was only observed in the insoluble fraction.

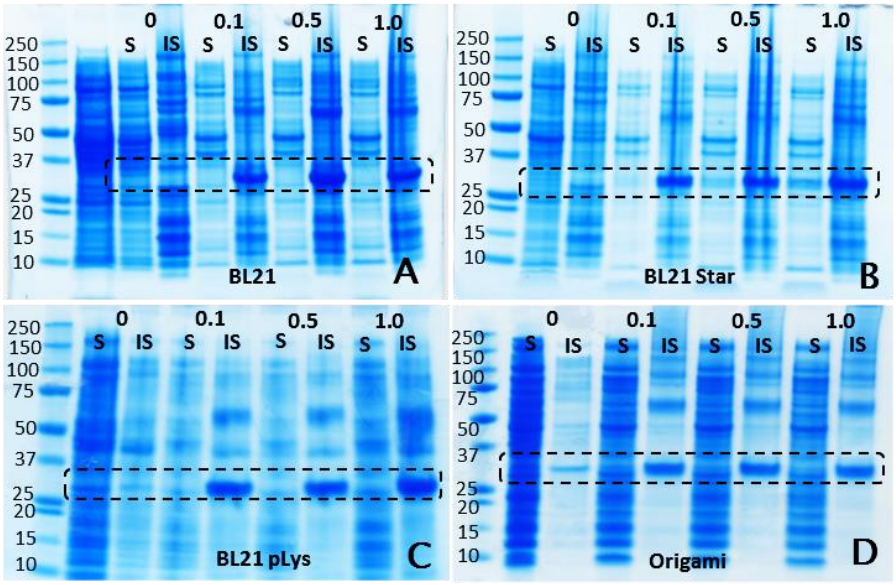


Fig. 3. 3. 3. SDS-PAGE of the soluble and insoluble fractions after small-scale expression tests of the LS1 construct in four different *E. coli* strains. Four strains were tested for expression of the LS1 construct in the soluble (S) and insoluble (IS) fractions: BL21 (DE3) (**A**), BL21 Star (DE3) (**B**), BL21 pLys (DE3) (**C**) and Origami™ (DE3) (**D**). Numbers above the lanes indicate the IPTG concentration used for the induction of protein expression. Dotted line indicates the region where the LS1 band was expected. The first lane is the MW marker in kDa.

Another *E. coli* strain transformed with the LS2 DNA was the LEMO21 (DE3) strain. This strain contains all the main features of the BL21 (DE3) but is designed to allow for tunable expression of difficult proteins. This is achieved by modulation of the levels of lysozyme *lysY* which varies at different L-rhamnose concentrations. Two different L-rhamnose concentrations were tested (0.5 mM and 1 mM) but yet again – no soluble protein was detected in the cell extracts.

Changes to induction conditions can sometimes shift the equilibrium towards production of soluble protein. For this reason, the next set of experiments tested how modified induction conditions influence the soluble protein content of the cell extract. Several different variants of induction conditions were tested. In the first few tests the length and temperature of induction were varied.

A common approach aiming to increase the solubility of expressed protein is to grow the cultures at reduced temperatures. At higher temperatures large amount of proteins are being produced which increases the likelihood of protein aggregation. In addition to that, hydrophobic interactions are temperature dependent which causes the aggregation

processes to be favoured at higher temperatures. There are also chaperones within the *E. coli* cytoplasm whose activity increases at temperatures close to 30°C. Maintaining high levels of activity of these chaperones and preventing protein aggregation could aid in the protein folding processes and increase the soluble fraction of the expressed protein[120].

Cells were induced again with three different concentrations of IPTG. This time, protein expression was induced at 30°C and at room temperature. Protein expression levels were measured as previously after 3h, 5h and 20 h from induction. Small changes in the levels of soluble protein were detected in the samples incubated at room temperature but the levels of insoluble protein were still very high (Fig. 3. 3. 4). A larger scale expression test (1 L) with overnight induction at room temperature was undertaken. Unfortunately, attempts to purify the protein from this expression test were not successful, likely due to the very low expression levels of soluble protein.

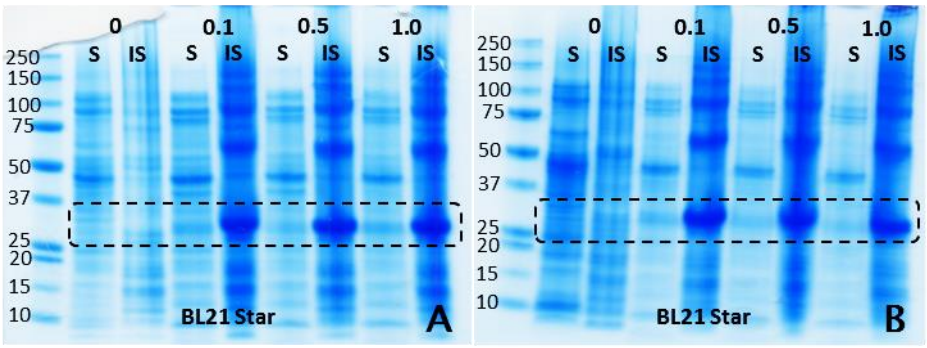


Fig. 3. 3. 4. SDS-PAGE of the soluble and insoluble fractions after small-scale expression tests of the LS1 construct at room temperature. Expression of the LS1 construct in the soluble (S) and insoluble (IS) fractions in BL21 Star (DE3) strain after 5 h (**A**) and 20 h (**B**) induction at room temperature. Numbers above the lanes indicate the IPTG concentration used for the induction. Dotted line indicates the region where the LS1 band was expected. The first lane is the MW marker in kDa.

The next set of experiments tested whether heat-shocking of the *E. coli* cells prior to induction could help to produce soluble protein. Exposing the cells to either hot or cold temperatures for a short time can lead to production of the heat shock proteins, including chaperones that assist in protein folding processes and stabilise unfolded intermediates, protecting the cells from the effects of extreme temperature conditions. It is possible that presence of such proteins could increase the production of soluble recombinant selectin constructs. In order to test this hypothesis, cells were subjected to very low (4°C) or very

high (42°C) temperatures for 30 min prior to induction with IPTG. Unfortunately, these tests have not increased the amounts of soluble protein produced by the cells (Fig 3. 3. 5. A and B).

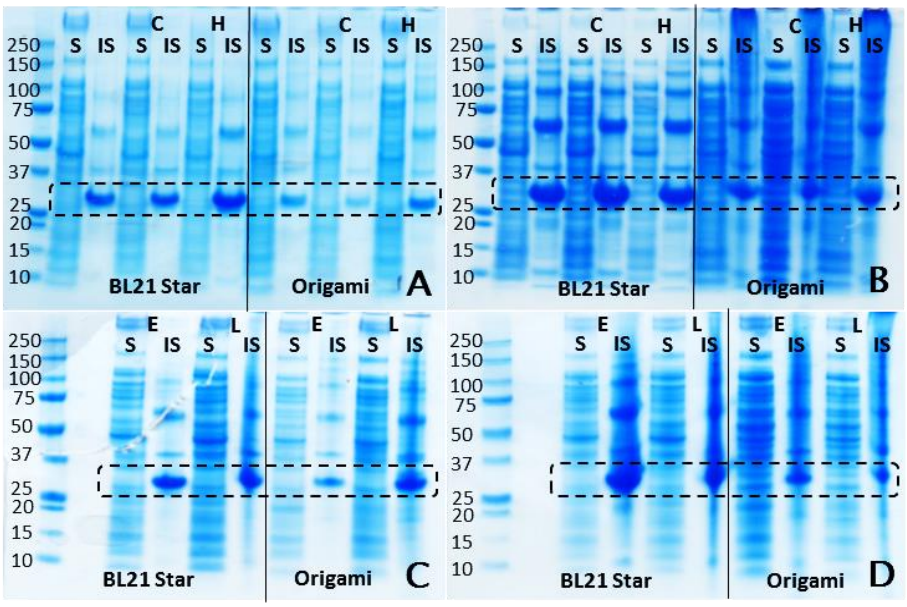


Fig. 3. 3. 5. SDS-PAGE of the soluble and insoluble fractions after small-scale expression tests of the LS1 construct with modified induction conditions. Expression of the LS1 construct in the soluble (S) and insoluble (IS) fractions in the BL21 Star (DE3) and Origami™ (DE3) strains after heat-shocking the cells at 4 °C (C) or 42 °C (H) prior to induction and after inducing at either early (E) or late (L) stages of the culture growth. Samples were taken at 5 h (**A, C**) and 20 h (**B, D**) from induction. Dotted line indicates the region where the LS1 band was expected. The first lane is the MW marker in kDa.

Finally, tests were performed to assess whether inducing the cells at different time points could affect the production of recombinant protein. At different stages of the culture growth, processes regulating protein expression and folding undergo changes. Typically, induction should be performed after the cells have entered the exponential growth phase but before the cell density grows to high levels and the culture starts to enter the stationary phase. Such optimal region corresponds to optical cell density (OD_{600}) levels of approximately 0.6 – 0.8 and this was the time point of induction for all previously described *E. coli* expression tests. Here, in one set of expression tests the cells were induced at OD_{600} of 0.2 – 0.3, corresponding to transition from the lag phase into exponential growth phase. In another set of expression tests the cells were induced at OD_{600} of values higher than 1.0, when the culture has high cell density and begins to enter the stationary phase. Sadly, no change was detected in the amount of soluble protein expressed by the cells in either of the tests (Fig. 3. 3. 5. C and D).

These experiments concluded the studies on expression of soluble LS1 construct in *E. coli* cells. A similar set of tests was performed on the two-domain L-selectin construct (LS2). Unfortunately, despite the high expression levels, this construct was expressed solely in the insoluble fraction of the cell extract.

It is worth noting that the presence of the thioredoxin fusion tag did not help with soluble expression of either of the constructs. There are several factors that could contribute to this outcome. First of all, the most optimal site for attachment of the fusion tags is the protein's N-terminus, where they can aid in the initiation of translation processes[119]. Due to the unusual fold of the selectin N-terminus and the problems observed in selectin expression in *P. pastoris* after the attachment of the α-factor secretion sequence at the protein's N-terminus, the constructs for *E. coli* expression were designed in the way where all tags were attached to the C-terminus. Success rate of the fusion partner is also dependent on the protein-fusion tag combination. Therefore, to find optimal conditions for expression of a given protein, a range of different fusion partners should be tested. Due to the time and financial limitations of this project, it was not possible to perform such an extensive research on the suitability of different fusion partners for selectin expression in *E. coli*.

Although the attempts to express the two L-selectin constructs in the soluble fraction were unsuccessful, both L-selectin constructs were expressed in high amounts in the *E. coli* cells. If a method was developed that would allow to extract the insoluble proteins and induce their transition into the correct conformation, it could be possible to obtain the constructs suitable for the NMR studies in high enough yields. Therefore, further experiments focused on the search for optimal conditions for refolding of the L-selectin constructs from inclusion bodies.

3. 3. 4. 3. Protein refolding tests

Protein refolding becomes increasingly more common as a tool that helps to express difficult proteins in the *E. coli* expression system. Many different approaches to protein refolding have been explored but the most common ones involve refolding by infinite dilution, dialysis or on-column refolding[121,122]. In all three methods, the inclusion bodies need to be isolated from the cells and solubilised with the use of denaturants such as urea or guanidinium-HCl. These strong chaotropes weaken the noncovalent interactions between the proteins, preventing aggregation of the unstable protein intermediates. Strong reducing agents such as

dithiotheitol (DTT) or β-mercaptoethanol (β-ME) are used to break disulfide bonds. After the solubilisation step, the proteins need to be refolded into their native conformation. This process involves several steps in which the concentration of denaturant is gradually decreased to allow for formation of a new set of noncovalent interactions. Buffers used for refolding experiments should provide oxidising environment to support the formation of disulfide bridges. In the infinite dilution and dialysis methods, protein concentration is maintained at very low levels throughout the experiments in order to prevent protein aggregation. Finding the optimal refolding environment to fulfil all of the above conditions for a given protein can be a challenging, trial-and-error process[121,122].

The dilution method is the most frequently used approach for protein refolding. After solubilisation and purification of inclusion bodies, the denatured protein is diluted in large volumes of the refolding buffer to concentrations below 10 μM and incubated with stirring for up to 72 h. The underlying principle of this method is that at very low concentrations, protein folding reactions will be favoured over the aggregation as collisions between individual protein molecules will be largely reduced. In the presence of oxidising reagents formation of disulfide bridges will be encouraged and the protein is likely to find a stable refolded conformation[121]. The experiments described here are based on the protocols kindly provided by Mr. Christopher Weir.

The LS1 construct was expressed in high yields in the *E. coli* cells. Therefore, it was possible to extract large amounts of unfolded protein from the inclusion bodies (ca. 25 mg per 1 L of bacterial culture). As the inclusion bodies consist mostly of the unfolded recombinant proteins, the extracted proteins are normally quite pure at this stage. To isolate the LS1 construct, the solubilised inclusion bodies were bound to the His-Trap resin, incubated overnight and eluted with imidazole buffer. The protein migrated to the expected MW region on the SDS-PAGE (28 kDa) under reducing conditions and appeared to be free from any major contaminants, although LS1 dimers and trimers were also observed (Fig. 3. 3. 6.).

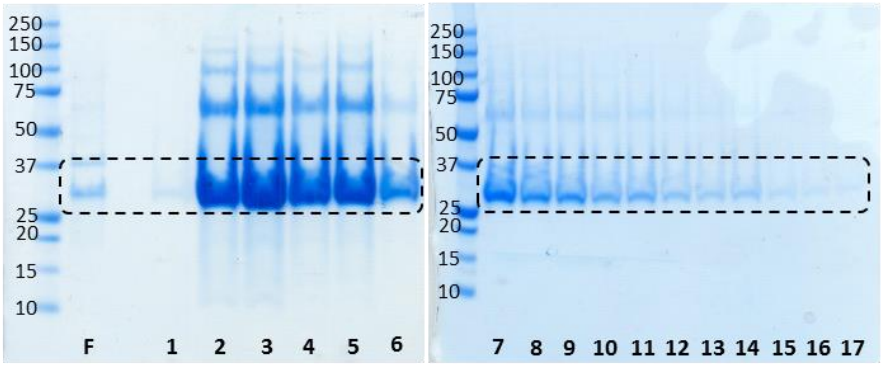


Fig. 3. 3. 6. SDS-PAGE of the fractions eluted from the His-Trap resin after the solubilisation of LS1 inclusion bodies. Fractions were eluted with 1 mM imidazole. Numbers below the lanes indicate the fraction numbers. Dotted line indicates the region where the LS1 band was expected. The first lane on each gel is the MW marker in kDa.

Because upon the transfer to refolding buffer the protein is suddenly exposed to an environment with reduced amounts of denaturing agents, it is very important that the protein is added in small portions and is quickly mixed into the buffer to prevent aggregation. During the experiments described here, care was taken to ensure that the protein additions were performed in very small portions (< 10 μL), with 15 – 20 s time interval between each added portion, and that efficient stirring was maintained throughout the whole duration of the experiment.

In the first instance, a series of small scale refolding tests was performed to assess the most suitable refolding conditions. Three different urea concentrations were tested (0 M, 1 M and 2 M), along with three different ratios of the reduced to oxidised glutathione (10:1, 2:1, 1:1). Glutathione is a thiol compounds found in the endoplasmic reticulum in eukaryotic cells that plays and important role in formation of disulfide bridges[123]. Oxidised glutathione (GSSG) acts as an oxidant facilitating the formation of disulfide bridges and reduced glutathione (GSH) acts as a reducing agent, cleaving the misplaced bridges in the process known as disulfide shuffling. Therefore, appropriate ratio of the two is important for correct protein folding[123].

Samples from all of the tests were subjected to SDS-PAGE analysis under non-reducing and reducing conditions. It is expected that, if the refolding is successful, a strong band will be observed under both conditions, with slight difference in migration between the reduced and non-reduced samples. Unfortunately, in this set of tests, strong bands in the reducing gel conditions were observed only at the start of refolding reaction. Very faint bands were

observed after 24 h and practically no bands could be seen after 72 h from the start of refolding reaction (Fig 3. 3. 7).

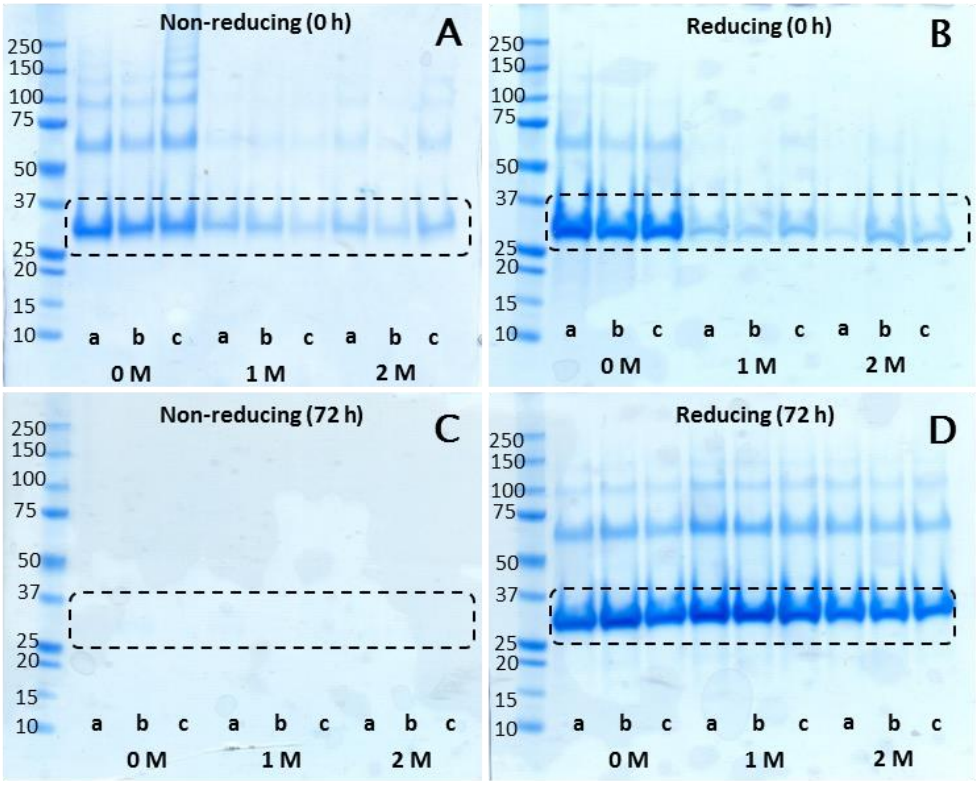


Fig. 3. 3. 7. SDS-PAGE of the samples taken during the small-scale refolding tests by infinite dilution on LS1 construct. Nine different buffers were tested with the concentrations of urea varying between 0 M, 1 M and 2 M (Numbers below the lanes) and the ratio of the reduced to oxidised glutathione varying between 10:1 (a), 2:1 (b) and 1:1 (c). SDS-PAGE was run under non-reducing (**A**, **C**) and reducing (**B**, **D**) conditions. Samples for the gels shown here were taken at 0 h (top) and 72 h (bottom gels) from the start of refolding reaction and prepared by TCA precipitation. Dotted line indicates the region where the LS1 band was expected. The first lane is the MW marker in kDa.

One of the refolding tests was performed on a larger scale (1 L), using the refolding buffer consisting of 20 mM Tris (pH 8.0), 150 mM NaCl, 10 mM $CaCl_2$, 1 M urea and a 10:1 ratio of reduced to oxidised glutathione. The aim of this experiment was to test whether some of the protein may be refolded, however, in too small amounts to be detected on a smaller scale. The protein was allowed to refold for 72 h and then dialysed into a buffer that did not contain any denaturing agents or other refolding additives. During the dialysis process the protein formed visible aggregates. Nevertheless, the dialysis was completed and the protein sample was filtered to remove the aggregates. Subsequent analysis by SDS-PAGE revealed that there

is still protein present in the filtered sample. The sample was passed through a HisTrap FF column to allow for its further purification and concentration from the large refolding volumes. However, it was not possible to elute the protein from the column – the protein has likely formed aggregates between the resin molecules and blocked the flow of the buffer through the column.

Additional small-scale tests were performed, using 2 – 5 mM DTT, 5% glycerol, increased salt concentration or a mixture of all these factors, still without any positive results. Following these tests, the refolding buffer composition was modified to include 5 mM DTT in addition to 150 mM NaCl and a 10:1 ratio of reduced to oxidised glutathione. With the use of this refolding buffer, more promising results were observed – presence of the LS1 construct was detected in the samples prepared under both the reducing and non-reducing conditions (Fig. 3. 3. 8). After 72 h, the refolding reaction was stopped and the sample was divided into eight parts to be used in a range of small-scale dialysis tests.

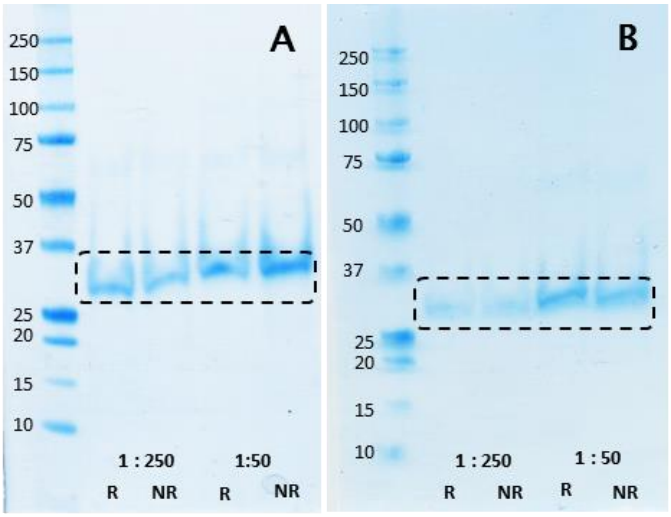


Fig. 3. 3. 8. SDS-PAGE of the samples taken during the small-scale refolding tests on LS1 construct with 5 mM DTT in the refolding buffer. The refolding reaction was performed in 20 mM Tris, pH 8.0 with 150 mM NaCl, 10 mM $CaCl_2$, 1 M urea, 10 mM reduced glutathione, 1 mM oxidised glutathione and 5 mM DTT. Samples for SDS-PAGE were prepared under reducing (R) and non-reducing (NR) conditions, after TCA precipitation. Two different dilutions of the protein sample in the refolding buffer were tested: 1 in 250 and 1 in 50. **A** Samples taken at the start of the refolding reaction. **B** Samples taken after 72 h of the refolding reaction. Dotted line indicates the region where the LS1 band was expected. The first lane is the MW marker in kDa.

During the dialysis step, the refolding solution is dialysed against the buffer that does not contain any denaturing agents. The proteins that are correctly folded should remain stable after the buffer exchange and can be purified and concentrated for further experiments. However, if the proteins are not folded correctly or if partially folded intermediates are present in the solution, in absence of the denaturing agents, they will be likely to aggregate and return to the insoluble form.

Buffers used in the dialysis tests explored different salt concentrations and combinations of additives that could stabilise the recombinant protein in the new buffer and push any unstable intermediates towards folding into correct conformation. Unfortunately, in the majority of tests, large amounts of visible aggregates were observed after just a few hours of dialysis. The only two tests where the amount and size of aggregates were significantly reduced, were the tests in which L-arginine was included in the dialysis buffer. The positive effect of arginine on the LS1 construct stability was confirmed by SDS-PAGE, where a clear band can be observed under both reducing and non-reducing conditions, even after filtering off the larger aggregates, proving that the soluble protein is still present in the solution (Fig. 3. 3. 9). In later tests, it was established that the concentration of arginine as low as 50 mM is enough to inhibit protein aggregation.

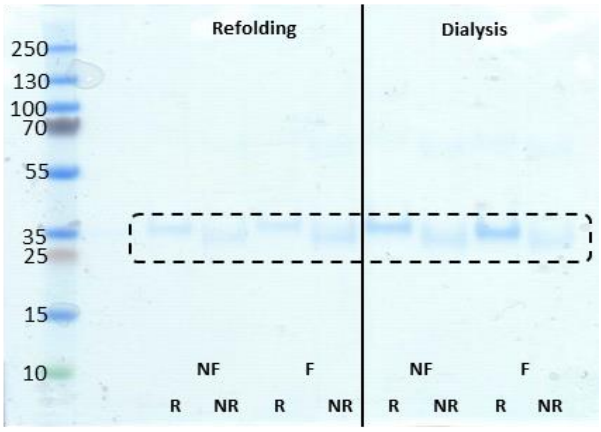


Fig. 3. 3. 9. SDS-PAGE of the samples taken before and after the dialysis in buffer containing 250 mM L-arginine. On the left-hand side are the samples collected after the refolding experiment was completed. On the right-hand side are the samples collected after the dialysis into the new buffer (20 mM Tris, 150 mM NaCl, 10 mM $CaCl_2$, 250 mM arginine, pH 8.0) was completed. Samples for SDS-PAGE were prepared under reducing (R) and non-reducing (NR) conditions, after TCA precipitation. In each case, one of the samples was passed through syringe filter prior to TCA precipitation to remove larger aggregates (F). Dotted line indicates the region where the LS1 band was expected. The first lane is the MW marker in kDa.

L-arginine is an amino acid well-known for its inhibitory role in protein aggregation and it is frequently used to enhance the refolding efficiency[124]. Although the exact mechanism for its action is still unknown, it is suspected that arginine interacts with the protein surface residues, minimising the protein-protein interactions that lead to aggregation. Similarly to guanidinium-HCl, arginine contains a guanidinium group which may account for its denaturant-like activity. It has also been suggested to interact with the hydrophobic protein residues via the three methylene groups in its side chain (Fig. 3. 3. 10)[125,126].

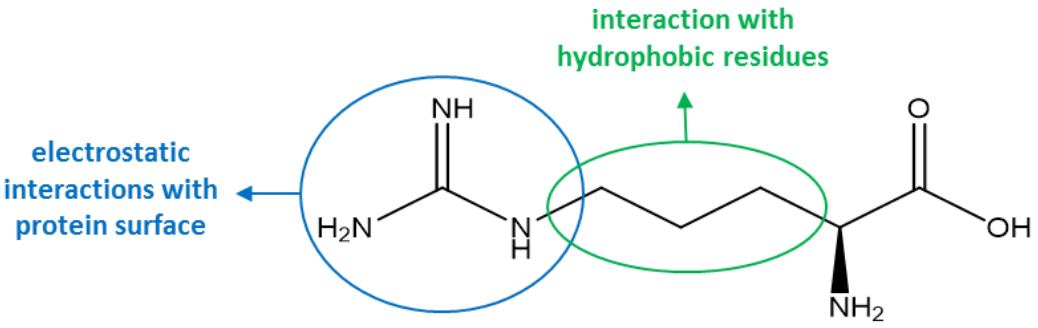


Fig. 3. 3. 10. Chemical structure of L-arginine. Presence of the guanidinium group and the hydrophobic methylene side chain are responsible for the inhibitory activity of L-arginine on protein aggregation.

Despite the fact that protein aggregation was largely reduced in the presence of L-arginine, it was not possible to confirm whether the LS1 construct is truly folded in the correct conformation, or whether even the soluble form of the protein is actually a misfolded intermediate and the presence of arginine simply prevents the aggregation of otherwise misfolded proteins. Filtered sample obtained from dialysis was, in fact, analysed by mass spectrometry and it was not possible to identify individual peaks that could correspond to the soluble LS1 species. It is likely that small aggregates were still present in the sample, despite the presence of arginine. This suspicion was confirmed when, within a day after the dialysis, visible aggregates started to form in the sample stored at 4°C.

Several other tests were performed, attempting to optimise the refolding and dialysis buffers in order to increase the refolding efficiency. Yet again, no positive results were obtained. At this point, an alternative protocol for protein refolding was provided by Dr Karen McClymont from the University of Glasgow. In this method, the protein is refolded whilst bound to the Ni-NTA resin by subsequent washes with buffers containing gradually decreasing urea concentrations. Due to the fact that the protein remains bound to resin as the concentration

of denaturing agent is decreased, aggregation should be prevented and the protein should have more freedom to explore different folding options and find the correct conformation.

Although the idea of on-column refolding sounds very promising and works well for some proteins, the tests on LS1 refolding were, yet again, unsuccessful. Similarly to what was observed in earlier refolding tests, elution of the protein from the Ni-NTA resin was not possible as the resin solidified after treatment with the refolding and elution buffers. A likely explanation for this behaviour is that the protein did not fold correctly when bound to the resin and started to form aggregates as it was being eluted.

Finally, an experiment was performed, attempting to combine different aspects of the two protocols discussed above. In this experiment, the pellets were prepared and solubilised as in the infinite dilution protocol. However, the refolding method from the second protocol was used – the on-column refolding – but this time with decreasing guanidine-HCl concentration. It was thought that the protein may be more stable in the guanidine-HCl buffer and that this buffer could better support the folding. Unfortunately, the results were similar to those observed in case of the on-column refolding with urea-based buffers. The resin formed a solid aggregate and no protein was detected in the eluted fractions.

3. 3. 5. Conclusions

Despite many different methods and varied conditions used in the expression, purification and refolding experiments for different selectin constructs, it was impossible to establish a strategy that would be suitable for expression of these constructs in either yeast or bacterial cells. Large amount of time and resources have been spent on this project, yet none of the results were promising enough to allow for expression of the isotopically labelled proteins necessary for the in-depth NMR studies. For this reason, from this point onwards, direction of the project was changed, exploring the use of Saturation Transfer Difference (STD) NMR and molecular dynamics for the studies of fCS-selectin interactions. This involved continuing expression experiments collaboration with the Oxford Protein Production Facility (OPPF) to investigate the possibility to express the selectin constructs in mammalian cells. This work is described in the following section of this chapter.

3. 4 Selectin construct expression in HEK293 cells

Because of the described problems with expression of selectin constructs in the yeast and bacterial cells, an option of using an alternative expression system was considered. Selectin constructs expressed previously by other groups were produced in mammalian systems and a possibility exists that, due to the complex fold of the lectin and EGF domains, held by multiple disulfide bonds, as well as the post-translational modifications that often help in the regulation of protein folding, expression of selectin constructs in simple, unicellular expression systems may not be achievable.

Isotopically labelled eukaryotic proteins have been previously produced in mammalian cells. For instance, Chinese hamster ovary (CHO) cells were used as a host for expression of ^{13}C,^{15}N-labelled human chorionic gonadotropin and human embryonic kidney (HEK293) cells were used to produce ^{15}N and ^{13}C,^{15}N-labelled rhodopsin[127,128]. However, mammalian expression systems are not a preferred option for expression of isotopically labelled proteins due to the challenges in development and optimisation of efficient experimental protocols for each protein and the high costs of the isotopically enriched media required for mammalian cell culture[129,130].

Despite these limitations, establishing protocols for selectin production in mammalian cells could allow us to obtain the desired selectin constructs which could then be used in some of the NMR experiments, such as STD, or in further crystallographic investigations of selectin-ligand interactions. If an efficient method for selectin construct production in mammalian cells was found, future research could lead to development of protocols for their isotopic labelling to allow for more in-depth NMR studies. For these reasons, collaboration with the Oxford Protein Production Facility (OPPF) was established, as this institution has a longstanding experience in production of proteins from a diverse range of organisms in a variety of expression systems. During this collaboration, a set of selectin constructs for expression in HEK293 cells was designed. A visit to OPPF laboratories undertaken as part of this PhD project involved work on molecular cloning of those constructs into pOPIN vectors and their subsequent transfection into HEK293 cells for small-scale expression tests. These tests allowed to identify a construct suitable for the large-scale expression experiments, kindly undertaken by the OPPF team later in the project.

3. 4. 1. Expression system

Mammalian cells used in the selectin expression tests described below were human embryonic kidney cells (HEK293). This cell line originated from the kidney cells of a human embryo and was the first human cell line transfected with sheared adenovirus DNA. A 4.5-kilobases long fragment of the viral genome inserted into chromosome 19 of the HEK cells, disrupting their morphology and causing them to lose the kidney cell characteristics[131]. For this reason, HEK293 cells are not a good *in vitro* model of kidney cells. They are, however, a popular platform for stable transfection and expression of human proteins as they are easy to grow in culture and transfect[132]. Since their creation in 1973, HEK293 cells and their derivatives have been one of the most frequently used cell lines in molecular biology, used for production of adenoviral vaccines, recombinant proteins, as well as in electrophysiology and neuropharmacology, for analysis of neural synapse formation. Protocols for transfection of HEK293 cells are well-established and the OPPF routinely uses this cell line for expression of recombinant proteins[133]. It was therefore decided that the HEK293 could be a suitable host for expression of one- and two-domain selectin constructs.

3. 4. 2. Construct design

The OPPF has developed a suite of expression vectors (pOPIN) for high-throughput cloning of protein constructs. The pOPIN vectors utilise multiple promoter systems to allow for protein production in all major expression systems without the need to make multiple, host-specific vectors for each construct. The vectors use the pUC origin of replication to obtain high yields of plasmid from *E. coli*, as well as the cytomegalovirus (CMV) enhancer, chicken β-actin promoter and rabbit β-globin poly-A site to increase the protein transcription levels. All of the pOPIN vectors include a cleavable His-tag and offer a variety of fusion proteins to facilitate protein expression and purification[134].

Six selectin constructs were designed, each to be cloned into two types of pOPIN vectors suitable for the use in insect and mammalian cells. The first type of vector used in selectin expression tests was pOPING (Fig. 3. 4. 1. A) – a simple vector that allows for expression of

the desired construct decorated with C-terminal His_6-tag to allow for one-step purification. The second vector used was pOPINEneo-3C-2Strep-His_8 (Fig. 3. 4. 1. B) – a variant of pOPING that contains neomycin phosphotransferase gene, a selectable mammalian marker that confers resistance to the antibiotic G-418, and that allows for expression of a cleavable C-terminal streptavidin-tag followed by a His_8-tag for production of ultrapure proteins. Both vectors contain a signalling sequence that allows for secretion of the transcribed constructs into extracellular media.

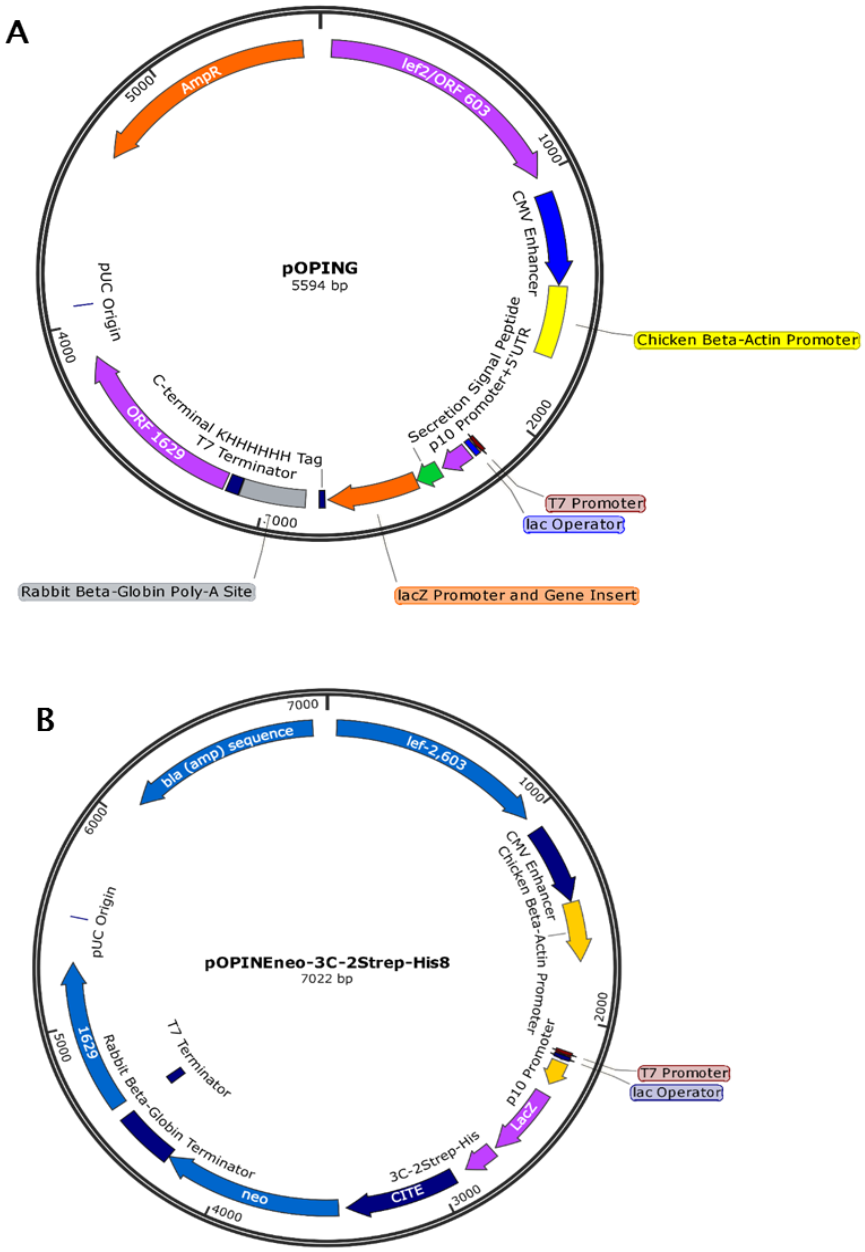


Fig. 3. 4. 1. Maps of vectors used in mammalian expression tests. Vector maps of pOPING (**A**) and pOPINEneo-3C-2Strep-His8 (**B**), highlighting the characteristic features of these vectors such as the promoter and enhancer elements that help to increase levels of protein expression and transcription in mammalian cells. Vector images were produced in SnapGene® Viewer.

Each vector was used to express two types of P- and L-selectin constructs – a single domain construct, consisting of residues Trp1 – Ser121, and a two-domain construct, consisting of residues Trp1 – Val156. As the UniProt entry P14151 for L-selectin contains a ten-residue long propeptide, additional two types of constructs were designed for this protein, with the propeptide sequence preceding residue Trp1. Each construct was assigned an OPPF number for internal reference – the constructs and their numbers are listed in Table 3. 4. 1..

Table 3. 4. 1. List of all selectin constructs designed for mammalian expression tests.

OPPF no.	Protein	Vector	Propeptide	Sequence
20306	P-selectin	pOPING	N/A	Trp1 - Val156
20307	P-selectin	pOPINEneo-3C-2Strep-His8	N/A	Trp1 - Val156
20309	P-selectin	pOPING	N/A	Trp1 - Ser121
20310	P-selectin	pOPINEneo-3C-2Strep-His8	N/A	Trp1 - Ser121
20311	L-selectin	pOPING	Yes	Trp1 - Val156
20312	L-selectin	pOPINEneo-3C-2Strep-His8	Yes	Trp1 - Val156
20313	L-selectin	pOPING	Yes	Trp1 - Ser121
20314	L-selectin	pOPINEneo-3C-2Strep-His8	Yes	Trp1 - Ser121
20315	L-selectin	pOPING	No	Trp1 - Val156
20316	L-selectin	pOPINEneo-3C-2Strep-His8	No	Trp1 - Val156
20317	L-selectin	pOPING	No	Trp1 - Ser121
20318	L-selectin	pOPINEneo-3C-2Strep-His8	No	Trp1 - Ser121

3. 4. 3. Materials and methods

3. 4. 3. 1. Fusion cloning of DNA constructs

The L- and P-selectin gene blocks were based on the sequences deposited in Uniprot database (# 14151 and # 16109, respectively) and synthesised by the Integrated DNA Technologies. Primers were designed to allow for In-Fusion™ cloning of the gene constructs into prepared pOPIN vectors. PCR was performed in 50 µL, using the Phusion Flash Master Mix (F-548L,

Thermoscientific) according to manufacturer's instructions, with 30 pmol of each forward and reverse primers, appended with the In-Fusion™ extensions as appropriate, and 1 µL of genomic DNA (50 ng/µL) as a template per reaction. Amplifications were performed in the Veriti PCR Thermal Cycler (ABI) under the following conditions: 10 s at 98 °C, followed by 29 cycles of 1 s at 98 °C, 5 s at 60 °C and 8 s at 72 °C. The PCR programme ended with a final extension of 2 min at 72 °C. The resulting PCR products were separated by electrophoresis on a 1% TBE agarose gel, visualised with SybrSafe™ (InVitrogen), with 5 µL of Hyperladder™ 1kb (BioLine) as a reference, purified using the Agencourt AMPure XP magnetic-bead based purification method (Beckman) according to the manufacturer's instructions and eluted in 50 µL of EB buffer (10 mM Tris pH 8.0), with yields assessed by electrophoresis.

2 µL of the purified PCR product (10 – 100 ng) and 100 ng of the linearised pOPIN vector were mixed with 1 µL Fusion Enzyme, 2 µL 5x Fusion Buffer (Quick-Fusion Cloning Kit, Biotool) and 4 µL of water. Samples were incubated at 37 °C for 30 min. 40 µL of TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) was added to stop the reaction.

3. 4. 3. 2. Transformation into E. coli and plasmid preparation

2.5 µL of the Fusion Reaction product was added to an aliquot (20 µL) of the *E. coli* OmniMaxII cells. The cells were incubated on ice for 30 min, then heat-shocked for 30 s at 42 °C, returned on ice and diluted with 120 µL of Power Broth (PB). The cells were incubated at 37 °C for 1 h and plated at two different concentrations (50 µL and 5µL) on 1 mL LB Agar plates with 100 µg/mL Ampicillin, 0.02% w/v X-Gal and 1mM IPTG. The plates were allowed to dry and incubated overnight at 37 °C.

The next day, transformed colonies (5 colonies per construct) were used to inoculate 1.2 mL of PB with 100 µg/mL Ampicillin in 96-well, deep-well plates (BD Falcon). The cultures were grown overnight at 37 °C, with shaking at 200 rpm, before harvesting by centrifugation at 5000 g for 10 min at 4 °C. Pelleted cells were then used for plasmid preparation. Plasmids were prepared from the *E. coli* cultures in a 96-well format, using a QIAgen BioRobot 8000 and QIAgen Turboprep kits, according to manufacturer's instructions (QIAgen). The resulting plasmids were screened using 25 µL PCR mix, with 12.5 µL of the Phusion Flash Master Mix (F-548L, Thermoscientific), 10 pmol of the forward primer, and 1 µL of plasmid DNA per reaction. Amplifications were performed in the Veriti PCR Thermal Cycler (ABI) under the

following conditions: 10 s at 98 °C, followed by 29 cycles of 1 s at 98 °C, 5 s at 60 °C and 8 s at 72 °C. The PCR programme ended with a final extension of 2 min at 72 °C. The PCR products were separated by electrophoresis on a 1% TBE agarose gel, visualised with SybrSafe™ (InVitrogen), with 5 µL of Hyperladder™ 1kb (BioLine) as reference.

3. 4. 3. 3. Transfection into HEK293 cells

HEK293 cells were maintained in DMEM (Sigma), supplemented with 10% foetal calf serum (Sigma, Poole, UK), 1x non-essential amino acids and 1mM glutamine (InVitrogen). For expression screening, cells were seeded in 24-well plates, at a density to give 75-80% confluent cells at the time of transfection. The cells were incubated overnight at 37 °C in a 5% CO_2/95% air environment. On the next day, the culture media were discarded and replaced with 1 mL DMEM, supplemented with 2% v/v foetal calf serum, 1x non-essential amino acids and 1 mM glutamine. For each transfection reaction, 2 µL of 1.33 mg/mL GeneJuice® (Novagen) and 60 µL of serum-free DMEM were mixed. Approximately 1 µg of plasmid DNA was added to the mix and incubated at RT for 10 min. The mix was then added to the cells which were then incubated at 37°C for 72-96 h in a 5% CO_2/95% air environment.

Secretion of protein into the media was analysed by harvesting the culture media at 72 h and 96 h post-transfection and running the samples (10 µL) on NuPage Bis-Tris gel (Invitrogen). Following transfer to PVDF membranes, His-tagged proteins were detected by western blotting using an anti-His_6 monoclonal antibody (clone BMG-His-1, Roche Diagnostics) in combination with anti-mouse-horseradish peroxidase secondary antibody (Pierce) and ECL (GE Healthcare).

3. 4. 3. 4. Scale-up and purification

Plasmid DNA for the large-scale transient transfection of HEK293 cells was prepared from a 1 L culture of *E. coli* cells, following the protocols from the PureLink HiPure Megaprep kit (Invitrogen). 4x20 mL of HEK293 cells were transferred to roller bottles for 300 mL scale-up. 75 mL of DMEM, supplemented with 10% v/v foetal calf serum, was added to each roller bottle and incubated at 37 °C for 3 days in the roller rig. After the incubation, the culture media were discarded, 60 mL of fresh DMEM, supplemented with 2% foetal calf serum, was added to each roller bottle and the cells were returned to incubator for 10 min. In a sterile container, 0.6 mg of plasmid DNA were added to 30 mL of serum-free DMEM. 30 mL of serum-free DMEM, containing 1 mg of polyethylenimine (PEI) were added to the mixture and incubated at RT for 10 min. 15 mL of the transfection cocktail were added to each roller bottle and returned to the incubator. The cell cultures were grown for 5-6 days and the secreted proteins were harvested by removing the media from the roller bottles and centrifuging at 6000 g for 20 min. The supernatant was filtered through a 0.45 μm bottle top filter and stored at 4 °C until purification.

Secreted proteins were purified using an automated in-house protocol, implemented on ÄKTA Xpress instrument (GE Healthcare), using previously described programmes[135]. Briefly, batches of media (100 mL) were passed through a 5 mL HisTrap column (GE Healthcare), each followed by a 20 mL wash of 50 mM Tris-HCl, 500 mM NaCl, 30 mM imidazole, pH 7.5, until all media were loaded. The column was washed with a further 20 mL of loading buffer, before elution of bound proteins in 50 mM Tris-HCl, 500 mM NaCl, 500 mM imidazole, pH 7.5. The eluted protein peak was automatically collected in a sample loop and re-injected onto a pre-equilibrated HiLoad 16/60 Superdex 75 column (GE Healthcare). Protein was subjected to size-exclusion chromatography in 20 mM Tris-HCl, 200 mM NaCl, pH 7.5 and fractions collected using a peak detection algorithm. Fractions containing the protein of interest at > 95% pure by SDS-PAGE were pooled, quantified (by A_{280nm}), snap frozen in liquid nitrogen and stored at -80 °C.

3. 4. 4. Results and discussion

3. 4. 4. 1. Construct design and plasmid preparation

As outlined in section 3. 4. 2., four types of constructs were designed for P-selectin and eight types were designed for L-selectin for expression in HEK293 cells. Construct design was similar to the design of constructs used in the previous expression tests. The higher number of L-selectin constructs is due to the presence of the propeptide sequence in the Uniprot gene entry for this protein that could play an important role in the regulation of protein synthesis. Propeptides are commonly found in higher organisms, expressed at the protein's N-terminus to assist in the folding processes and guide the protein to its destination within the cell. Once its role in post-translational processes is completed, the propeptide is cleaved and plays no further role. It is unclear why the propeptide sequence is found only in L-selectin and not in P-and E-selectins but its potential role in the protein folding has to be considered. For this reason, four of the designed L-selectin constructs included the additional ten-residue propeptide sequence at the protein's N-terminus.

Primers were designed to allow for the In-Fusion™ cloning of the gene fragment of interest into the appropriate vector. This method of cloning is fast, efficient and allows for flexible design of gene inserts for cloning into any vectors. For purpose of the In-Fusion™ cloning, 15-bp extensions homologous to the vector sequence at the site of insertion were incorporated into the primers. After PCR amplification with such a set of primers, the gene of interest will contain these extensions and can be incorporated into the homologous site within the linearised vector.

PCR amplification was successful for all constructs, as confirmed by agarose electrophoresis (Fig. 3. 4. 2. A). The resulting DNA fragments were purified using magnetic beads, inserted into linearised pOPING and pOPINEneo-3C-2Strep-His_8 vectors and transformed into *E. coli* cells for generation of high amounts of plasmid DNA, required for transfection of HEK293 cells. For each recombinant strain, five bacterial colonies were tested for the presence of the inserted plasmid. Of the five colonies, two that expressed the plasmid (colonies 3 and 4) were used in the subsequent HEK293 expression tests for each construct (Fig. 3. 4. 2. B).

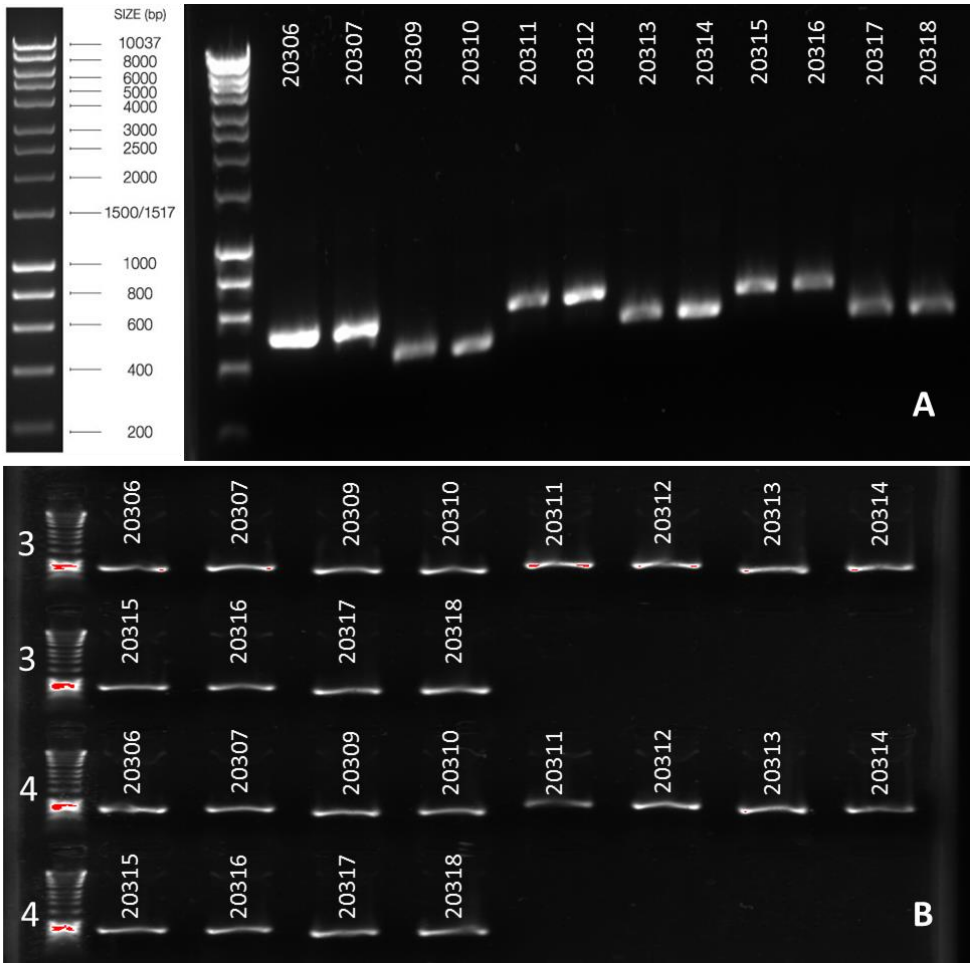


Fig. 3. 4. 2. Agarose electrophoresis of DNA fragments after PCR amplification (A) and of plasmids used for transfection of HEK293 cells (B). A DNA fragments generated in PCR amplification of selected gene sequence for each selectin construct, after the addition of vector-homologous extensions for the In-Fusion™ cloning. **B** Plasmids produced from *E. coli* after the cloning reaction. Shown here are only the results from colonies 3 and 4 that were subsequently used to transfect HEK293 cells. Numbers above the lanes refer to unique OPPF identifiers of the constructs (see Table 3. 4. 1).

3. 4. 4. 2. Small-scale expression tests in HEK293 cells

Transfection of selectin constructs into the HEK293 cells was tested on a small scale, using enhanced Green Fluorescent Protein (eGFP) and an internal standard, OPPF 6218, as positive controls. eGFP expression levels were monitored using fluorescent microscopy. In the cells transfected with the colony 3 plasmid DNA, after 24 h from transfection, close to 50% of the

cells were expressing eGFP (Fig 3. 4 .3. A). Furthermore, proportion of cells expressing eGFP increased to more than 90% (Fig 3. 4. 3. B) after 72 h from transfection. These results confirm that the cells used for transfection were healthy and that there were no issues with the transfection protocol for this set of constructs.

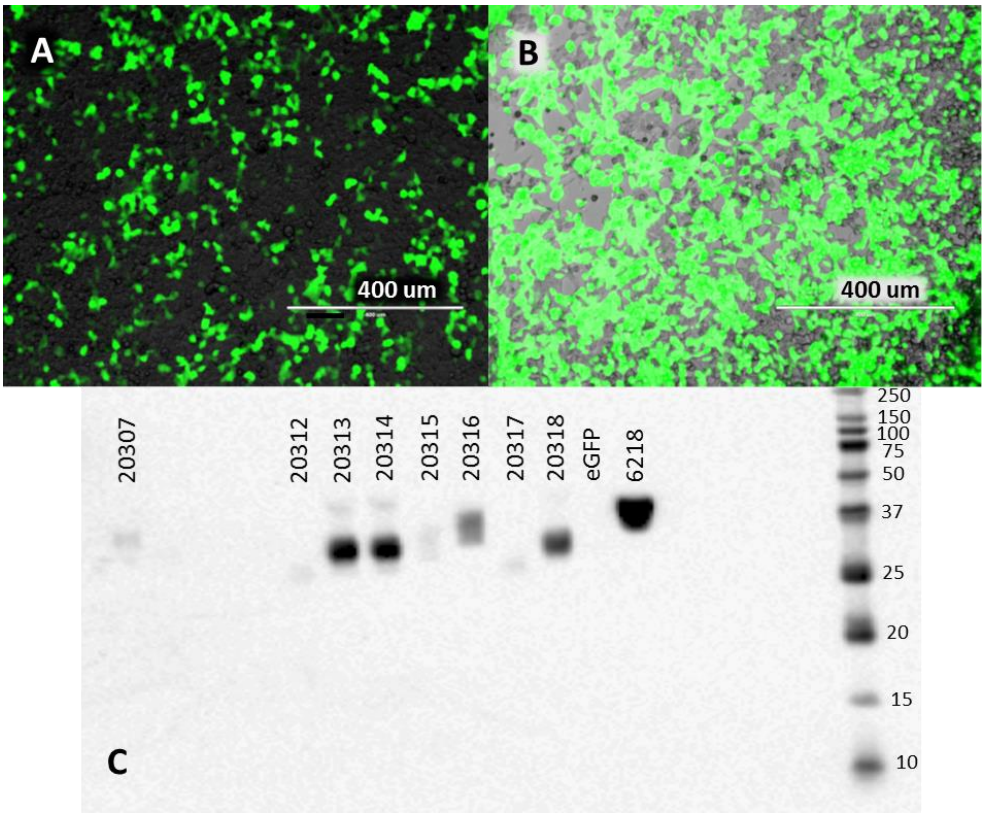


Fig. 3. 4. 3. Results of test expression of selectin constructs in HEK293 cells (colony 3). eGFP expression levels measured by fluorescent microscopy at 24 h (**A**) and 72 h (**B**) after transfection. **C** Western blot of the HEK 293 culture media from the cells transfected with plasmid DNA from colony 3, harvested after 72 h from transfection, with the MW marker on the right hand side (kDa). The lane labels refer to unique OPPF identifiers of the constructs (see Table 3. 4. 1.). OPPF 6218 is included as a positive control.

As the recombinant proteins should be secreted into the culture media, western blot was performed on the culture media harvested at 72 h. Three constructs from colony 3 showed a strong single band on the western blot – all corresponding to the L-selectin 1-domain constructs, 20313, 20314 and 20318 (Fig. 3. 4. 3 C). Plasmid DNA used for transfection of these constructs was sequenced and constructs 20313 and 20318 were confirmed to have the correct genetic sequence. Construct 20313 was found to contain a single missense mutation at residue 130 and was therefore not used in the large scale expression test.

Construct 20313 had a stronger signal on the western blot than 20318, indicating that its expression levels were slightly higher. In addition to that, 20313 is the construct of a simpler design, using only one tag (His_6-tag), which simplifies the purification process. For these reasons, it was decided that construct 20313 will be used for the scale-up expression in HEK293 cells. Interestingly, this is the construct in which the propeptide sequence was included, confirming that this sequence is likely to play a role in the regulation of protein expression.

Two attempts were made to transfect the colony 4 plasmid DNA into HEK 293 cells. However, in each attempt, numbers of the cells were low and the eGFP was expressed in less than 10% of the cells. All of the proteins showed none or very low expression levels when tested by a western blot. Due to time constraints, the focus of further experiments was on the positive results from transfection with the plasmid DNA from colony 3 and on the scale-up expression of the construct 20313.

3. 4. 4. 2. Large-scale expression of the construct 20313

Based on the results from the small-scale expression tests, a large scale (300 mL) expression of the construct 20313 was attempted. The procedure was based on the standard OPPF protocols that are used routinely for protein expression in HEK293 cells. The culture media were collected and purified by the His-Trap and gel filtration chromatography and collected fractions were analysed by SDS-PAGE. As shown in Fig 3. 4. 4, B, fractions A3 – A12 were characterised by a band on SDS-PAGE at approximately 30 kDa. Although this number is higher than the calculated molecular weight of the construct 20313 (16 kDa), it is likely that the protein is glycosylated which would significantly increase its mass. L-selectin has two N-glycosylation sites within the lectin domain and either or both of them could be glycosylated with carbohydrate chains that would account for the additional 14 kDa.

The corresponding region in the gel filtration profile shows a large peak in absorbance value (18 mAU) at retention volume of 55 mL, followed by an even larger peak (24 mAU) at 60 mL. Such a profile would indicate the presence of two species that are close in molecular weight. As the band seen on the SDS-PAGE is also not very well defined, it is possible that there are two species of similar mass present in the fractions from that region. This could be possible if two different L-selectin lectin domain variants were produced by the HEK293 cells. As the

genetic sequence for the construct is known and there can be no splice variants present in the sample, the most likely explanation is the difference in protein glycosylation. There are many options on the types of glycans that can be attached to protein post-translation, varying in structure and size. Slight changes in the structure of these glycans could lead to multiple variants of the protein produced by the cell[136].

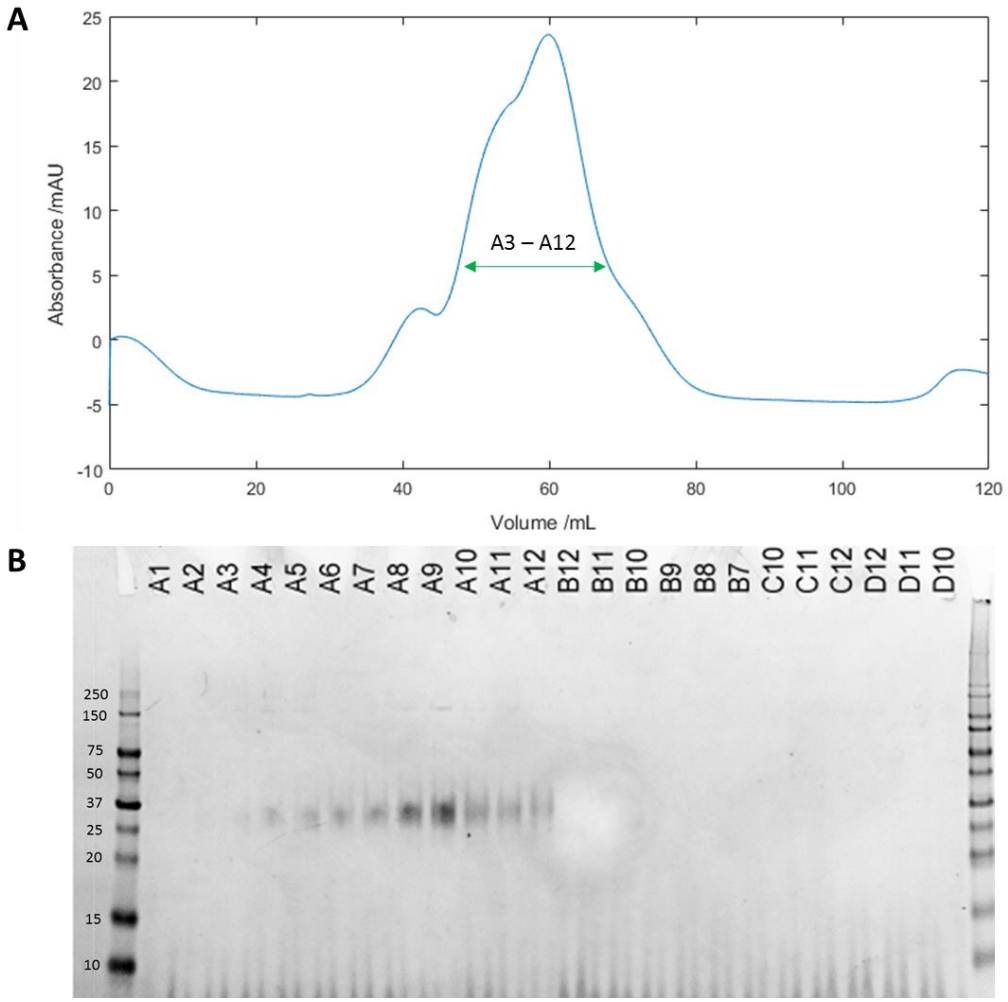


Fig. 3. 4. 4. Results of the gel filtration chromatography of construct 20313. A Gel filtration profile of the culture media from cells expressing construct 20313, following purification on the His-Trap column. Green arrow indicates the region corresponding to fractions A3 – A12 which were combined into the final protein sample. **B** SDS-PAGE of fractions collected in the gel filtration experiment. The first lane from the right is the MW marker (kDa). The lane labels refer to fraction numbers.

Previous studies of L-selectin glycosylation in HEK293 cells have shown that the N-glycan structures found on the lectin domain are mainly diantennary-type glycans, terminating with either Gal or GalNAc residues and containing antennary Fuc residues. One of the glycan determinants shows similarity to Le_x determinants and has a potential to inhibit E-selectin[108]. The pattern of glycosylation can vary between the different cell types and contributes to regulation of cell-cell interactions – L-selectin is believed to interact with the endothelial selectins via its N-glycans[85,137]. Analysis of protein glycans is a very complex field in itself as there are many variants of glycans that can be attached to the protein which leads to a large number of potential combinations. There are methods that can be used to study protein glycosylation, such as the affinity-based assays or MS analysis. The first method requires testing multiple receptors and enzymes in order to determine the type of glycan, whilst the latter one requires the glycans to be cleaved off the protein surface and purified[136]. Such work was too extensive to be undertaken as part of this project and therefore the exact nature of the glycosylation of 20313 construct remains unknown.

Fractions A3 – A12 eluted from the gel filtration column were collected and delivered to the University of Edinburgh. Overall yield of the protein obtained in this experiment was 0.5 mg. The received protein sample (0.7 mg/mL) was used in the selectin-fCS interaction studies by STD NMR, described in the following chapter.

3. 4. 5. Conclusions

Selectin expression tests in HEK293 cells have resulted in production of 0.5 mg of a single domain L-selectin construct. The construct is likely glycosylated, as indicated by the molecular weight of the produced protein which is higher than expected. Expression of the L-selectin single domain construct has never been reported before and protocols described here, used for successful expression of this construct, could be of importance for future research on selectin-ligand interactions. Comparative studies on the binding properties of the single-domain construct when compared to the previously expressed double-domain constructs could reveal important new information about the role of the EGF-like domain in the interactions between selectins and their ligands. It is also possible that, based on this research, protocols for expression of a similar but isotopically labelled protein construct could be developed in future.

Chapter 4: STD NMR studies of the fCS–selectin interactions

4.1. Introduction

There are many NMR techniques that can be used to study protein-ligand interactions. Some of them require a sample of soluble protein labelled with isotopes ^{13}C and ^{15}N. Some others, that are ligand focused, do not require labelled protein. Saturation Transfer Difference (STD) NMR belongs to the second category in that it allows to look at the changes in the ligand ^{1}H signals upon binding to protein. For this reason, it does not require an isotopically labelled sample and very small amounts of protein are sufficient for successful experiments. With a simple 1D homonuclear experiment, STD allows to identify the ligand moieties in direct contact with the protein surface, providing valuable structural information about the protein-ligand interaction in question.

STD NMR was originally designed to allow to screen a library of carbohydrates for binding affinity towards wheat-germ agglutinin[138]. However, during the screening process, it was discovered that this technique allows not only to detect the binding but also to deduce the binding epitope of ligands, an approach which is known as epitope mapping. Regions within the ligand that are in close contact with the protein showed strong STD signal, whilst the regions located further away from the protein surface showed low intensity or no STD signal at all. Thanks to these observations, STD NMR is now a commonly used technique in the discovery and design of new pharmaceutical products[139,140].

In the STD NMR, excess of ligand is added to the protein and two experiments are performed. In the first experiment, a reference ^{1}H spectrum is obtained, where a radio-frequency (RF) pulse is applied in the off-resonance region from both the protein and the ligand signals, usually at around 30 – 40 ppm[139]. In the second experiment, the protein signals are saturated by a selective (on-resonance) RF pulse in a region far from the ligand signals, usually in the region between -4 and 1 ppm. Care needs to be taken to not saturate the ligand signals in the on-resonance experiment. Therefore, if methyl protons are present in the ligand, it may be necessary to saturate in the region below 0 ppm to avoid false positives in the resulting

difference spectrum. No ligand signals are found below 0 ppm, whilst there is still a significant linewidth of protein signals which allows for selective saturation[139]. Alternatively, selective saturation can occur at around +10 ppm, if the ligand has no signals in the aromatic region. In both instances, all of the protein signals are saturated as saturation is able to spread across the protein surface by spin diffusion. If the ligand molecule is positioned close to the protein surface, magnetisation will be transferred to it via ^{1}H-^{1}H cross relaxation. Protons that are located closer to the protein surface will experience more saturation and protons that are located further away will experience less or no saturation. Saturation will cause the signal intensity to decrease. This effect can be observed when the spectra from the two experiments are subtracted[139,140]. The difference spectrum should contain only the signals from the ligand protons that are in close contact with the protein (Fig. 4. 3. 1).

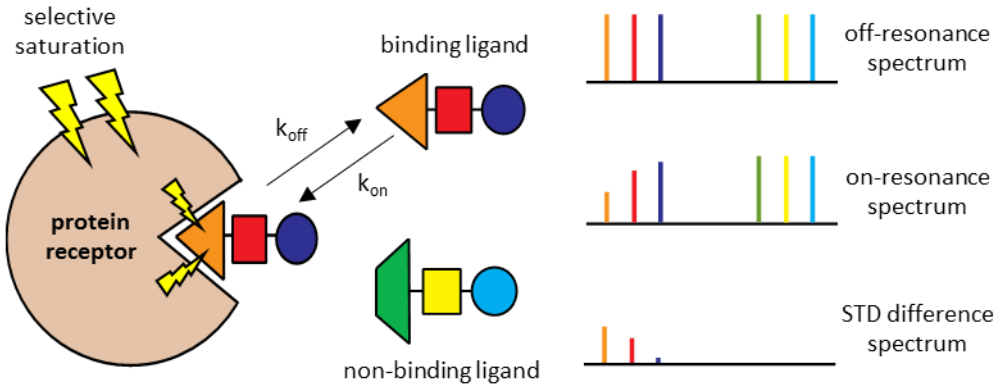


Fig. 4. 3. 1. Schematic representation of the principles of STD NMR. When the protein is saturated with a selective RF pulse, the saturation is transferred onto the bound ligand. The parts of the ligand that are closer to the protein surface will experience stronger saturation transfer and, therefore, will show a stronger signal in the STD difference spectrum. The non-binding ligands will not experience any saturation and will not show in the STD difference spectrum.

The efficiency of any STD experiment depends on the binding kinetics for the given ligand, specifically on the off rate k_{off}. Fast off rate allows the ligand to quickly dissociate from the protein and transfer the magnetisation into solution[140]. This way, if the large excess of ligand is present in the sample, one binding site can saturate many ligand molecules within a short space of time. The ligand is normally used in an approximately 100-fold molar excess over the protein to allow for this effect to take place. This in turn allows for very small (less than 10 μM) amounts of protein to be used in the STD experiments[140]. If k_{off} rates are too low (less than 0.01 s^{-1}), there is not enough exchange between the bound and free ligand and the

saturation transfer becomes inefficient. An additional problem with tight binding is that, as ligand resides in the binding site for a long time, saturation spreads further across the ligand and it becomes impossible to distinguish which protons are those closest to the protein surface[139]. On the other hand, if k_{off} rates are too high (more than 10 000 s^{-1}), the binding becomes too weak and not enough ligand molecules make it to the binding site[140]. The k_{off} values optimal for the STD experiments can be translated to K_D values in the range of 1 mM to 0.1 μM. These are estimated K_D values, based on the assumption that the on rate k_{on} is equal to 10^7 s^{-1}.[140,141]

Success of the STD experiment is dependent on the k_{off} for the ligand, as well as the T1 relaxation for free ligand in solution. Saturation times typically range from 1 – 2 s. Ligands in solution lose saturation mostly by T1 relaxation, which is usually in order of 1s for a small-molecule ligand[140]. Longer T1 relaxation times of the ligand will result in accumulation of saturated ligand molecules in solution and, therefore, in amplification of the STD signal. Large saturation transfer effects will be observed for ligands with large values of k_{off} and long T1 relaxation times. These effects have been previously demonstrated for carbohydrate ligands where a monosaccharide ligand characterised by high k_{off} rate and longer T1 relaxation time showed STD effects close to 3-fold higher than those observed for a larger carbohydrate ligand with a lower k_{off} rate and shorter T_1 relaxation time[142].

Another important condition for a successful STD experiment is that the protein needs to have a molecular weight of at least 10 kDa. In fact, the larger molecular weight, the better STD effect is going to be. This is due to the fact that lager proteins are characterised by slower tumbling in solution and therefore longer correlation time τ_c, which facilitates efficient spin diffusion across the protein surface[140]. Therefore, saturation will be spread more easily for larger protein molecules.

STD NMR has been widely used to study the protein-ligand interactions and one of the first applications was to study the interaction between a type of lectins known as agglutinins and carbohydrate ligands[141,142]. For example, a study performed in 1999 by Mayer and Meyer allowed to define the binding epitope of Lewis^b-hexasaccharide to wheat germ agglutinin, using 1D ^{1}H STD experiments, along with a 2D variant, STD-TOCSY[142]. Similar methods were applied to study the binding of GAGs and sLe^x-type carbohydrates to proteins of the immune system in later years[139,143,144]. In the study undertaken as part of this PhD project, STD experiments were used to investigate the binding between the fCS oligosaccharides from *H. forskali* and four different recombinant selectin constructs. Three different constructs were investigated and both 1D ^{1}H STD and 2D STD-TOCSY experiments were used.

4.2. Material and methods

Recombinant L-selectin constructs were purchased from R&D Systems (ADP2, 200 μg) and Creative BioMart (SELL-7019H, 100 μg). Recombinant P-selectin construct was purchased from Creative BioMart (SELP-474H, 100 μg). The fCS oligosaccharide samples from sea cucumber species *H. forskali* were provided by Glycomar Ltd.

All experiments were performed in Shigemi tubes without the plunger. In each experiment, 100 μg of the protein was dissolved in 450 μL of a 10:1 H_2O/D_2O mixture, producing a sample of ~4 – 7 μM concentration, depending on the construct used. Sodium azide (0.01 %) was added to the protein sample to prevent bacterial growth.

All STD NMR experiments were performed on a Bruker AVANCE III-600 (^{1}H: 600 MHz, ^{13}C: 150 MHz) instrument equipped with a TCI cryo probe. Protein/ligand molar ratios were varied from 1:50 to 1:100. The pulse sequence was BRUKER stddiffesgp.3, with water suppression via a double-pulsed field gradient spin echo (2 ms sinc pulses). The following parameters were used: 12 ppm sweep width, 4 s or 8 s relaxation delay (D1), 4 s presaturation time that was incorporated into the saturation delay and 1.36 s acquisition time. The on- and off-resonance irradiation was set to -4 ppm and 30 ppm, respectively. The irradiation was performed by 78 repetitions of 50 ms Gaussian pulses applied at $\gamma B_1/2\pi$ = 180 Hz. 16 scans were accumulated per one of the 128 averaging cycles, using one-scan alternate accumulation into on- and off-resonance spectra. The overall spectrometer time was 10.6 hours per experiment.

The pulse sequence used for the 2D STD-TOCSY experiment was BRUKER stdmlevesgpph. 2D STD-TOCSY acquires on- and off-resonance saturation in an interleaved manner, where the two transients are subtracted in computer memory. The experiment was recorded with MLEV-17 spin lock field of 40 ms and the relaxation delay of 7.5 s. 256 t_1 increments were collected and the following parameters were used: 12 ppm sweep width in f2 and 8 ppm in f1, 7.5 s relaxation delay (D1), 4s presaturation time and 27 ms acquisition time t_{1max}. The on- and off-resonance irradiation was set to -4 ppm and 30 ppm, respectively. The irradiation was performed by 80 repetitions of 50 ms Gaussian pulses applied at $\gamma B_1/2\pi$ = 180 Hz.

4.3. Results and discussion

A series of STD experiments was undertaken on four different selectin constructs. Three of them were the commercially available constructs consisting of the extracellular domains of L- and P-selectins. The fourth protein used in these experiments was the single-domain L-selectin construct produced at the Oxford Protein Production Facility (see Section 3. 4). The fCS sample used in the STD experiments was a depolymerised fCS (dfCSVI) consisting of a mixture of small oligosaccharide fragments, estimated to have an average MW of 4.8 kDa.

In each experiment, a ^{1}H NMR spectrum was recorded before and after the addition of dfCSVI to the protein sample. The STD spectra were recorded with selective irradiation at -4 ppm. After subtraction of this spectrum from the reference spectrum, peaks with different saturation levels were identified and assigned to the individual protons in the fCS trisaccharide. In this way, it was possible to determine the fCS protons spatially close to the protein surface with their relative degree of saturation. Data obtained for each of the constructs is presented below.

4. 3. 1. fCS binding to the L-selectin construct SELL-7019H

The first selectin construct studied by STD experiments was the recombinant human L-selectin (SELL-7019H) purchased from Creative BioMart. This construct consisted of residues Met1 – Asn332 which constitute the full extracellular part of the protein, including the lectin, EGF-like and the SCR domains. The predicted molecular weight of this protein is 34.5 kDa. However, as the L-selectin produced in mammalian cells is glycosylated, the true molecular weight of the construct is between 55 – 65 kDa.

The reference ^{1}H spectrum and the STD difference spectrum of the SELL-7019H with dfCSVI is presented in Fig. 4. 3. 2. As can be seen in this spectrum, dfCSVI interacts with the extracellular domains of L-selectin. Based on the chemical shifts, it was possible to identify some of the regions within the fCS oligosaccharides that participate in this interaction.

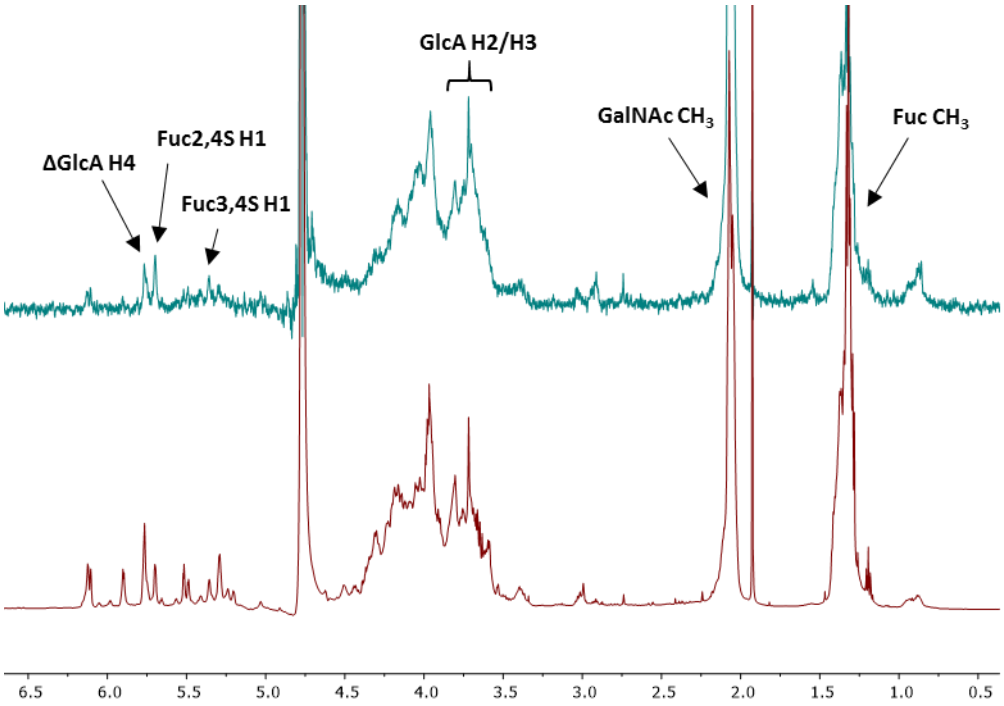


Fig. 4. 3. 2. STD experiment on the mixture of fCS and the SELL-7019H construct. Reference ^{1}H spectrum (bottom) and the STD difference spectrum (top) of the SELL-7019H sample containing dfCSVI.

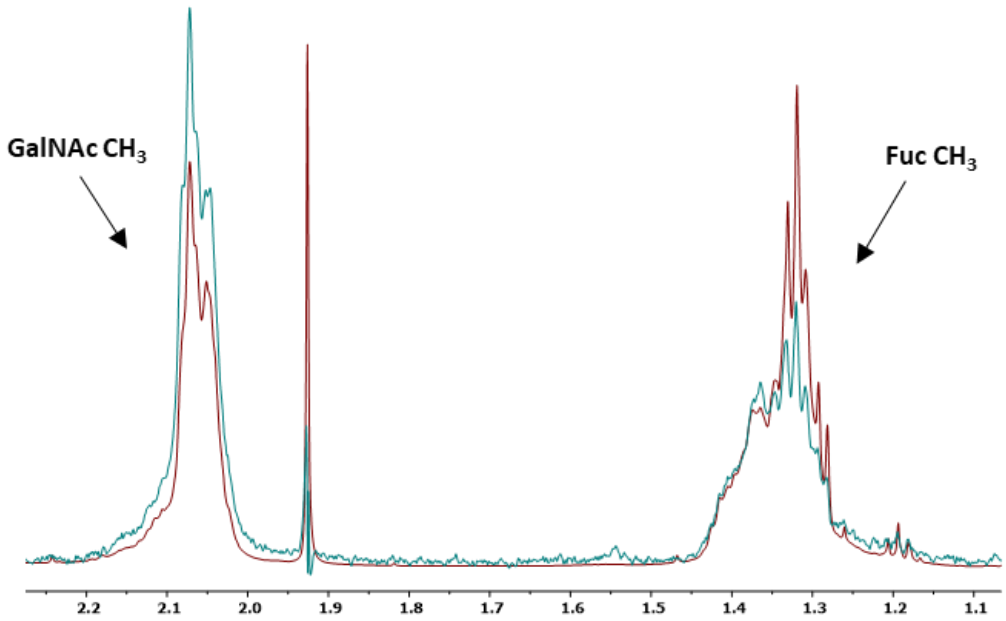


Fig. 4. 3. 3. Expanded section of the NMR spectra from Fig. 4. 3. 2.
Comparison of the STD response for the GalNAc acetyl and Fuc methyl signals. The STD difference spectrum is shown in blue and the ^{1}H reference spectrum is shown in red.

Strong response upon selective saturation of the protein signals was observed for the methyl signals of the galactosamine (GalNAc) and fucose (Fuc) residues. Figure 4. 3. 3 shows an expanded view of these signals – it can be seen that the response in GalNAc acetyl signals is stronger than the response observed in fucose signals. To confirm that this is a true response initiated by the saturation transfer, in another experiment, selective irradiation was performed at 10 ppm – a region far from the chemical shifts of methyl groups. In this variation of the STD experiment, response of the same intensity as in the case of irradiation at -4 ppm was recorded for all of the signals, proving that the methyl signals are not the result of irradiation of carbohydrate signals but of the saturation transfer from the protein. To allow for full relaxation of the methyl signals, experiments with double the relaxation delay (D1 = 8 s) were also performed, as slowly relaxing signals, such as the methyl protons, can give a false response if insufficient relaxation is allowed. This was not the case and identical spectra were obtained, proving that the methyl signals are true STD enhancements.

Assignment of the fCS signals was possible as the oligosaccharide sample used in the STD experiments was previously studied by NMR in the Uhrín group and the STD ^{1}H reference spectra could be compared to the assigned NMR spectra for this sample. All fCS samples have been assigned using a range of 1D and 2D NMR experiments. Although the samples are heterogeneous, it is possible to identify groups of peaks that correspond to specific atoms within the fCS units, for example allowing to distinguish different types of sulfated fucoses. A list of all assigned peaks for this fCS sample is presented in Table 4. 3. 1.

Table 4. 3. 1. List of all assigned peaks for the fCS sample used in the STD experiments.

	H1/C1	H2/C2	H3/C3	H4/C4	H5/C5	H6/C6
Δ glucuronic acid						
Δ1[1]	5.05/103.5	3.96/70.06	4.51/75.6	6.11/111.30		
Δ1[2]	5.02/103.6	nd/ 70.12	nd/75.5	6.12/110.97		
Δ2[1]	4.95/102.7	nd	nd	5.76/106.8		
Δ2[1]	4.92/103.12	3.90/70.3	4.50/76.3	5.76/106.4		
Fucoses (FI 3,4S, FII 2,4S and FIII 4S)						
FI.→Δ1	5.29/98.8	3.98/66.3	4.62/75.3	4.89/79.1	nd	1.31/ nd
FI.→ Δ2	5.28/98.3	3.96/66.5	4.61/75.3	4.91/79.0	nd	
FI.→ U1	5.35/99.2	3.95/66.5	4.54/75.42	5.04/79.38	nd	
FII.→ Δ1	5.51/96.31	4.43/75.07	4.15/66.5	nd/80.75	nd	
FII.→ Δ2	5.49/96.65	4.43/75.02	4.15/66.48	4.70/80.75	nd	
FII.→ U2	5.69/96.6	4.49/75.1	4.16/66.6	4.87/81.13	4.87/nd	1.308/nd
FIII.→ Δ	5.21/98.6	3.81/68.44	4.02/68.50	nd	nd	
FIII.→ Δ	5.19/99.1	3.80/68.44	4.04/68.50	4.63/80.71	nd	
FIII.→ U	5.40/98.50	3.82/68.3	4.04/68.8	nd	nd	
Glucuronic acid						
U1	4.50/103.8	3.63/73.5	3.7/79.2	3.95/nd	nd	
U2	4.50/103.78	3.66/73.87	3.75/77.18	3.96/75.58	nd	
GalNAc						
A	~ 4.6/99.7	~ 4.1/51.4	~ 4.05/77.6	~ 4.85/76.2	3.9-4.2/~72	nd
Aα	5.24/11.34	4.35/49.35	4.20/73.17	4.89	nd	nd
Aβ	4.75/95.05	4.05/nd	4.24/nd	nd	nd	nd
Aβ	4.74/94.73	3.95/52.61	4.21/nd	nd	nd	nd

*Δ designates terminal GlcA residue produced via β-eliminative and photochemical depolymerisation .
**numbers refer to different types of fCS chains, although the full structure of each chain remains unclear.

The ^{1}H 1D spectrum, aligned with the STD ^{1}H reference spectrum for this fCS sample, is shown in Fig. 4. 3. 4. Comparison of these two spectra also allowed to identify impurities present in the protein sample used in this experiment.

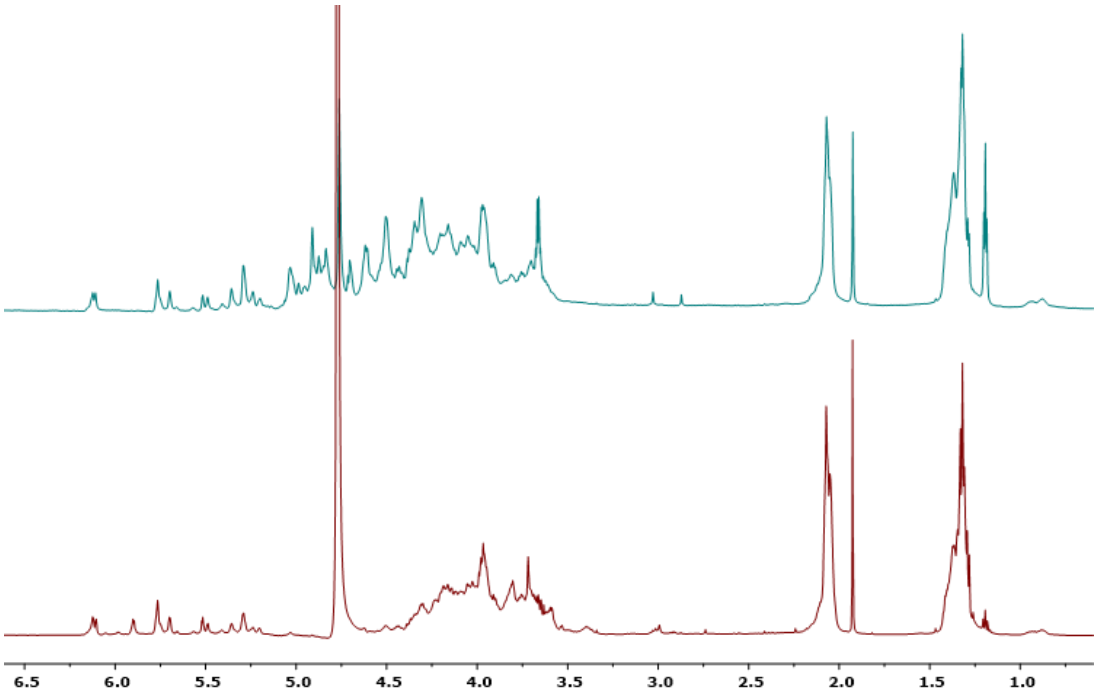


Fig. 4. 3. 4. ^{1}H NMR spectrum of the fCS sample used in this STD experiment.
The ^{1}H spectrum of this fCS sample (top) is aligned with the reference spectrum from the STD experiments, obtained after the addition of SELL-7019H (bottom).

The STD difference spectrum contained multiple signals from fucose. Amongst the fucose signals, the strongest response was recorded for the fucose methyl protons which appear on the spectra as a set of overlapping peaks between 1.28 – 1.43 ppm. Fucose in fCS can exhibit different sulfation patterns and the multiple peaks observed in this region correspond to different types of fucose from oligosaccharides of different lengths and sulfation patterns. STD response is recorded for more than one type of fucose and comparison of peak intensities to those in the original ^{1}H spectrum shows that the level of response varies between the different types of fucose (Fig. 4. 3. 5). Besides the signals of the methyl protons, response for the anomeric protons of both 2,4- and 3,4-sulfated fucose residues was also observed, slightly higher for Fuc2,4S (5.69 ppm) than for Fuc3,4S (5.35 ppm) signal.

Although response was also recorded in other regions of the spectrum, it was not always possible to confirm the identity of the peaks due to high peak overlap. However, it was

possible to establish that the signals in the region between 3.56 – 3.76 ppm most likely correspond to H2 and H3 of the glucuronic acid (GlcA). Strong response was also observed for the H4 signal of the unsaturated glucuronic acid (ΔGlcA) resulting from the depolymerisation of the fCS polysaccharide via β-elimination reaction.

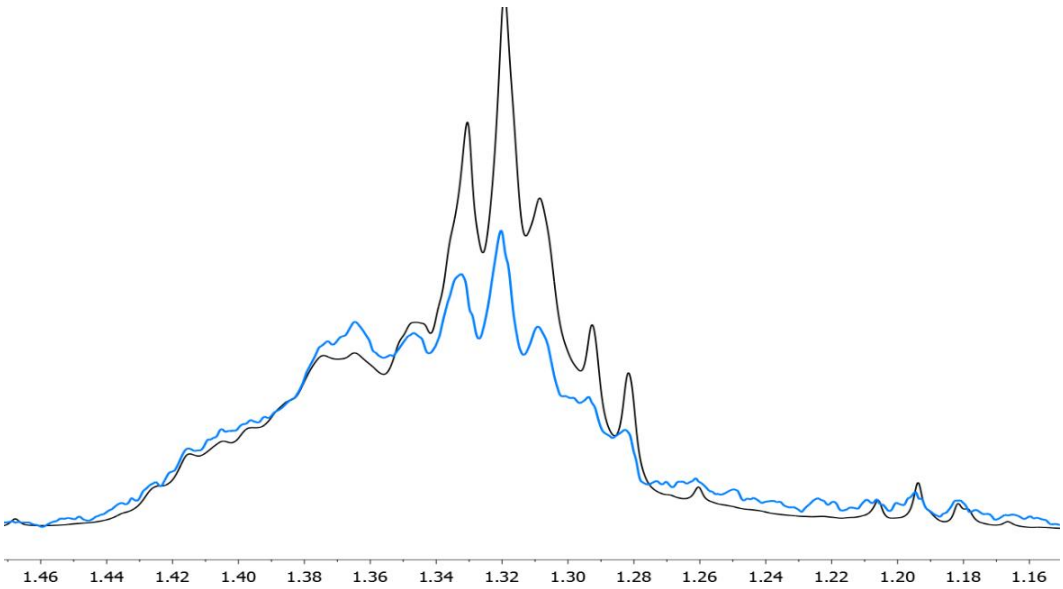


Fig. 4. 3. 5. Magnification of the fucose methyl region of the ^{1}H and STD difference spectra of fCS-SELL-7019H complex. The STD enhancements vary for different types of fucose residues as shown by the differences in intensities of individual peaks in the reference ^{1}H spectrum (black) and the STD difference spectrum (blue).

The absolute values of the STD enhancements are very low (less than 1%). However, the identified regions of the spectra were integrated and the values of integrals were normalised to 100 % for the GalNAc CH_3 signals which showed the strongest STD response. This allowed to calculate the relative values of STD enhancement for all identified peaks and thus to gain information about the relative contribution of individual regions within the fCS to L-selectin binding (Fig. 4. 3. 6).

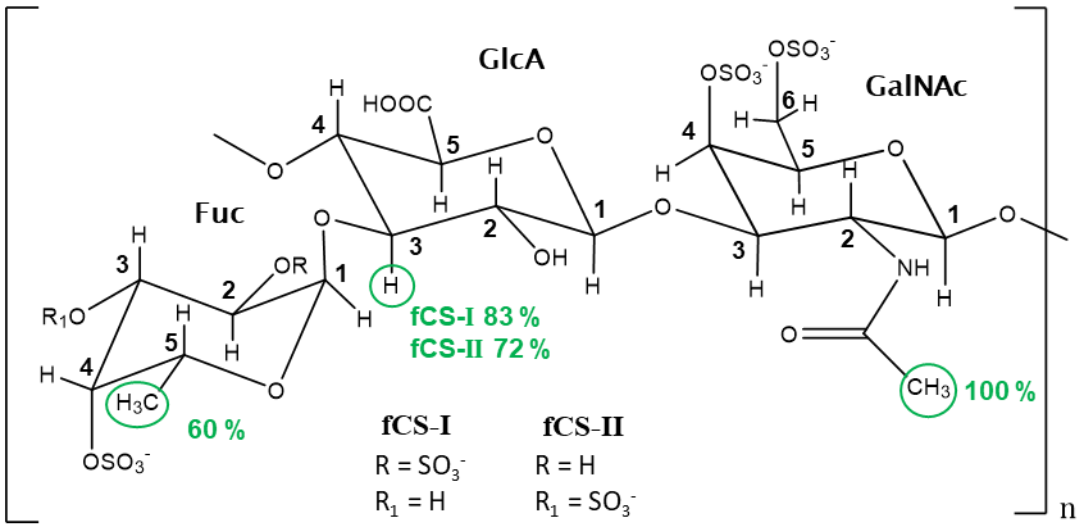


Fig. 4. 3. 6. Structure of the fCS trisaccharide with relative degrees of saturation of individual protons involved in L-selectin binding. Relative percentage saturation (shown in green) for each proton was calculated with respect to the GalNAc CH_3 signal, based on the values of peak integrals.

4. 3. 2. 2D STD–TOCSY experiment on fCS binding to L–selectin

In order to aid the identification of the peaks in the regions of high overlap, a 2D ^{1}H-^{1}H STD-TOCSY experiment was also performed. Due to technical difficulties, it was only possible to perform this analysis on an alternative commercially available L-selectin construct, ADP2. This construct consists of the same portion of the native protein as the SELL-7019H construct. However, the sample contained small-molecule additives which obscured some of the fCS signals and thus the use of SELL-7019H constructs was preferred for the 1D STD experiments.

Similarly as in the case of the 1D STD experiments, in 2D STD-TOCSY experiments two 2D TOCSY spectra are obtained – one with off- and one with on-resonance irradiation of the protein signals. The two spectra are then subtracted from each other which allows to observe correlations between the protons which are in contact with the protein[138]. Use of 2D STD methods can be helpful in cases where mixtures of ligands are used and the STD signals are overlapped in the 1D spectrum[140]. 2D STD TOCSY has been used previously for epitope mapping in protein-carbohydrate systems, for example to study interactions of sialic acid derivatives with rhesus rotavirus[145] or fucosylated hexasaccharide with lectin AAA[140].

The 2D STD-TOCSY experiment allowed to detect the correlation between some of the protons in the overlapped regions of the 1D spectrum (Fig. 4. 3. 7). A set of cross peaks at 1.36-1.40/4.85-4.90 ppm was identified as a correlation between the fucose CH_3 and H5 proton signals. This means that the Fuc H5 protons are also located close to the protein surface. Other identified correlations were between the Fuc2,4S H2 and H3 protons (4.50/4.16 ppm), GlcA H1 and H2 protons (4.50/3.61 ppm) and a terminal GalNAc with a β-reducing end H2 and H3 protons (4.28/3.99 ppm). Therefore, it was possible to obtain further information on the regions of fCS that interact with L-selectin. Cross peaks at 3.94-3.98/4.52-4.57 ppm could not be identified with confidence as they could correspond to correlation between either Fuc3,4S H2 and H3 or ΔGlcA H2 and H3 protons.

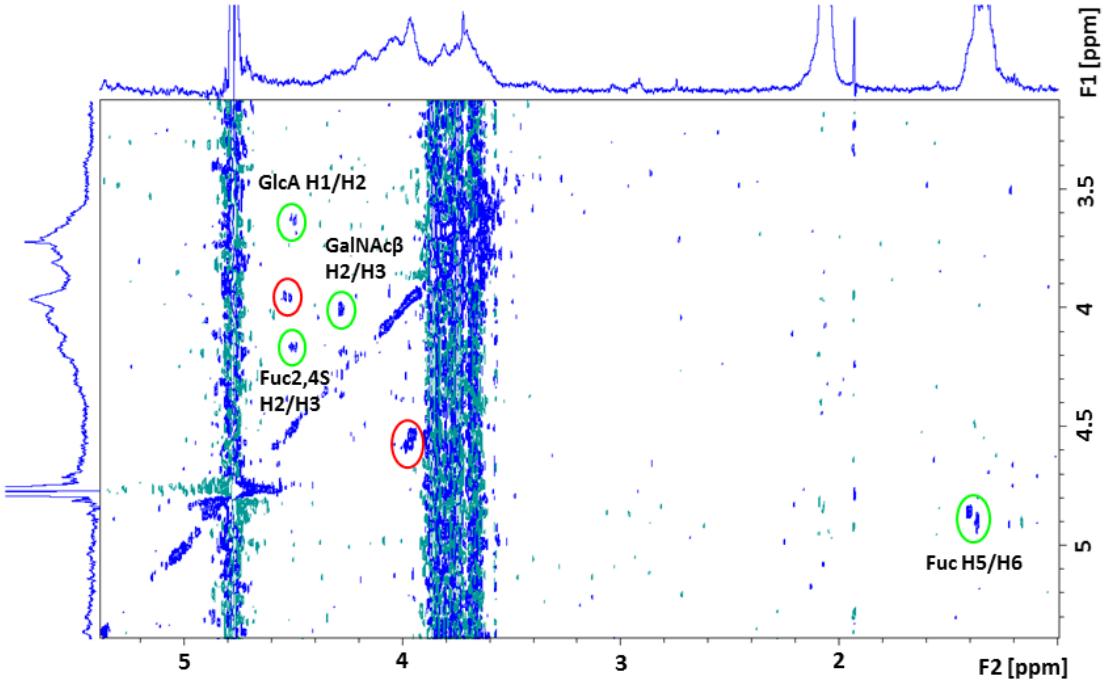


Fig. 4. 3. 7. 2D STD-TOCSY experiment on the mixture of fCS and the ADP2 construct.
2D STD-TOCSY was performed on the L-selectin ADP2 sample containing dfCSVI to confirm the identity of some of the peaks in the regions of high overlap. The identified cross-peaks are highlighted in green. The cross-peaks circled in red could correspond to either Fuc3,4S H2/H3 or ΔGlcA H2/H3 proton correlations.

4. 3. 3. fCS binding to the single-domain L-selectin construct

Further insight into fCS binding to L-selectin was obtained from the 1D STD experiment on the single-domain construct produced at the OPPF. By comparing the results of this experiment with the results obtained for a larger extracellular construct SELL-7019H, it was possible to confirm that the binding occurs via the C-type lectin domain. Slightly lower response was recorded for the fucose signals (66% Fuc H1, 50% Fuc CH_3) than in the case of the larger SELL-7019H construct. GlcA H2 and H3 signals were obscured by a strong signal from the contaminant present in the protein sample (Fig 4. 3. 8).

The above results suggest that the absence of the EGF-like domain does not affect the position of fCS oligosaccharides on the protein surface and only slightly reduces the strength of the fCS-selectin interaction.

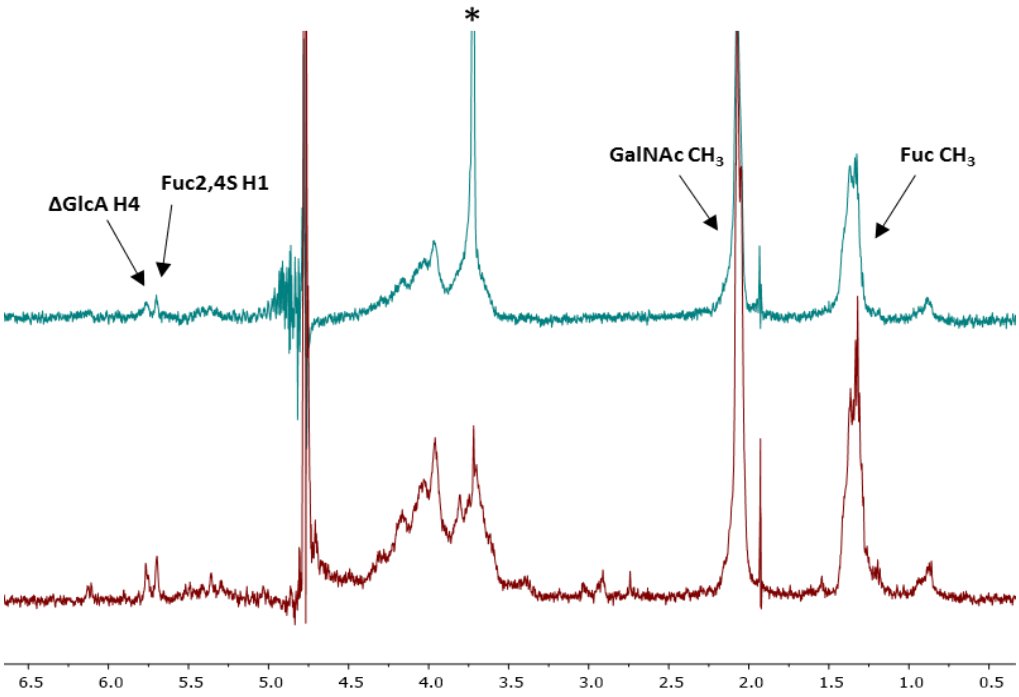


Fig. 4. 3. 8. Comparison of the STD difference spectra for the single domain (top) and the multiple domain (bottom) L-selectin constructs. The two spectra were obtained using the same experimental conditions and the same amount of dfCSVI. High intensity peak at 3.64 ppm (labelled with an asterisk), obscuring possible response from GlcA H2 and H3, is a signal from a contaminant carried over from the protein production. It was not possible to determine the identity of this compound.

4. 3. 4. fCS binding to the P-selectin construct SELP-474H

The final STD experiment tested the binding of dfCSVI to the P-selectin construct SELP-474H, analogous to the L-selectin construct SELL-7019H used in the previous experiments. This construct consists of the residues Met1 – Ala771 and includes all of the P-selectin extracellular domains. As P-selectin contains more SCR domains than L-selectin, the size of this construct is larger, with calculated molecular weight of 81.3 kDa and the apparent weight of 120 kDa due to glycosylation.

The ^{1}H reference and the STD difference spectra for this construct are shown in Fig. 4. 3. 9. As in the case of L-selectin, strong response was recorded for some of the peaks, proving that the fCS oligosaccharides interact with the extracellular domains of P-selectin.

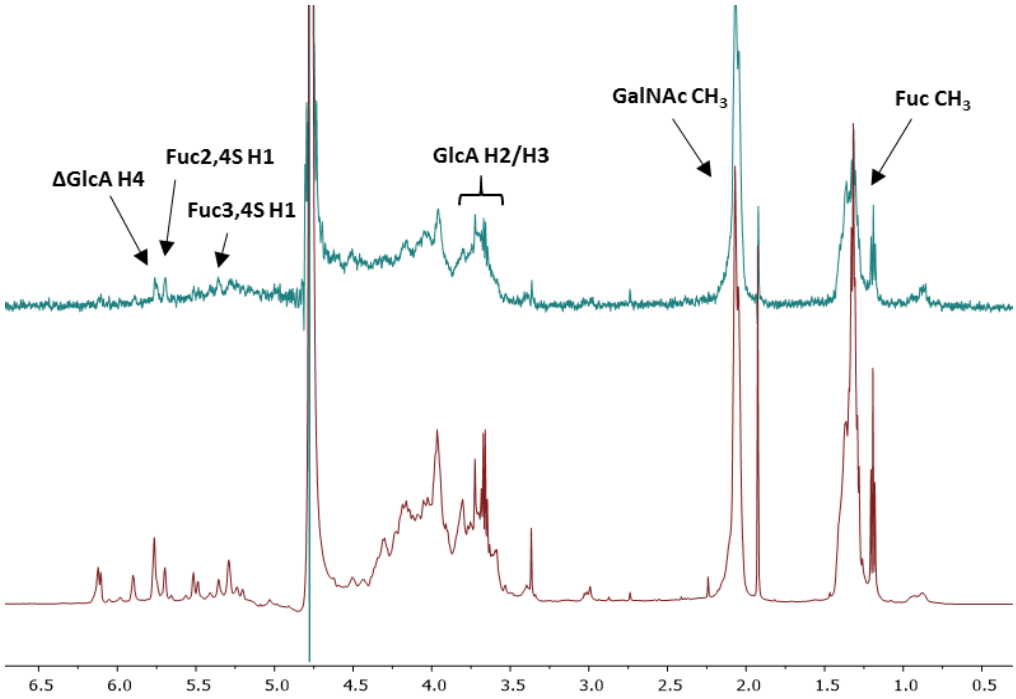


Fig. 4. 3. 9. STD experiment on the mixture of fCS and the SELP-474H construct. Reference ^{1}H spectrum (bottom) and the STD difference spectrum (top) of the SELP-474H sample containing dfCSVI.

The observed response was similar to response recorded for the L-selectin constructs. Again, the strongest response was seen for the methyl signals of GalNAc residues. Fucose methyl proton signals were also identified, as well as the signals corresponding to fucose 2,4S and

3,4S anomeric protons and H4 of ΔGlcA. As in the case of L-selectin, signals between 3.57 and 3.87 ppm likely correspond to GlcA H2 and H3 protons.

As before, relative percentage saturation was determined, by integration of the identified signals and normalisation to GalNAc methyl proton signals to 100% (Fig. 4. 3. 10). The relative saturation values were similar for both proteins, suggesting that the mode of binding of fCS, as well as the binding affinity, are similar for L- and P-selectins.

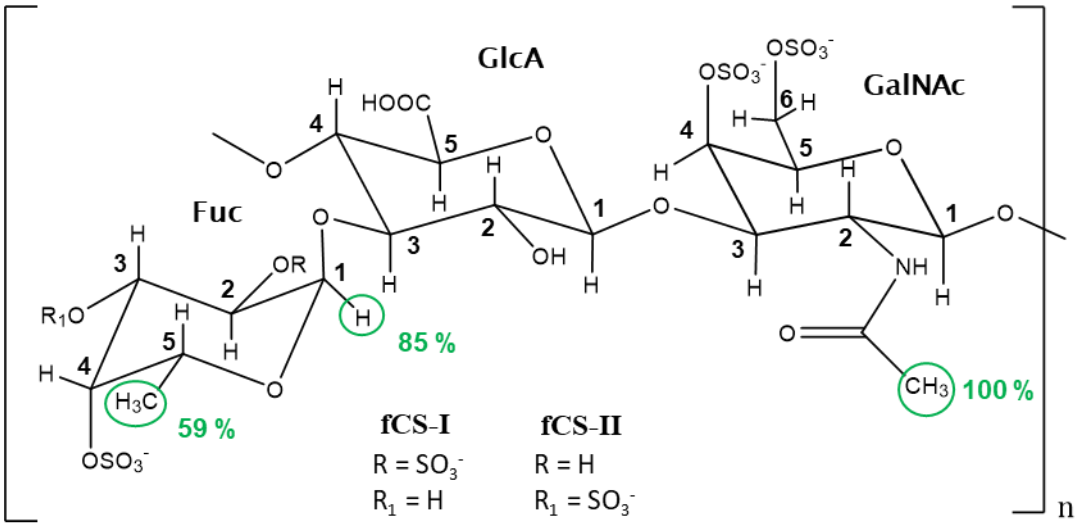


Fig. 4. 3. 10. Structure of the fCS trisaccharide with relative degrees of saturation of individual protons involved in P-selectin binding. Relative percentage saturation (shown in green) for each proton was estimated with respect to GalNAc CH_3 signal.

4.4. Conclusions

Experiments described in this chapter confirmed that fCS oligosaccharides bind to both L- and P-selectins. By performing STD experiments on four different selectin constructs, it was possible to identify the protons in direct contact with selectin surface. It was also possible to compare the binding between the single- and multiple-domain L-selectin constructs and between the analogous L- and P-selectin constructs. No significant differences were observed in the fCS binding by any of these recombinant proteins.

Strong response recorded for GalNAc methyl protons suggests that the N-acetyl group of this residue may be directly involved in selectin binding. Multiple signals from the fucose protons

suggest that this residue also plays an important role in fCS-selectin interaction. In all of the samples, STD enhancement was identified for the Fuc anomeric protons. The 2D STD-TOCSY experiment also confirmed the contribution from the Fuc H2, H3 and H5 protons, indicating that fucose is surrounded by the L-selectin residues from all sides. Other identified regions within the fCS included H1 and H2 protons of the GlcA and H2 and H3 protons of the terminal GalNAc residue. Experiments with the use of isotopically labelled protein would be required to further evaluate the mode of binding of the fCS on the selectin surface.

Chapter 5: *In silico* studies of selectin-carbohydrate interactions

5.1. Introduction

As shown in the previous chapters, studies of the protein-carbohydrate interactions can be very challenging. Carbohydrate-protein complexes are notoriously difficult to crystallise. Moreover, X-ray structures offer limited amount of information as they do not provide many clues on the internal dynamics. Detailed studies by NMR require availability of an isotopically labelled protein which is not always easy to obtain. For this reason, it is becoming increasingly more common to apply computational methods to study protein-carbohydrate binding. Software and algorithms used for the *in silico* studies are becoming more and more sophisticated, allowing to produce fairly reliable predictions about the behaviour of the complex under study.

In absence of an X-ray or NMR structure of the protein-ligand complex, the approach that can be used is to computationally dock the ligand to the existing protein structure by using the available PDB files. Molecular docking uses a set of algorithms to scan through many possible bound complexes and rank them according to an energy scoring function. Predicted contributions from the hydrogen bonding, electrostatic and Van der Waals interactions, repulsive forces, desolvation and torsional entropy are summed together to estimate the free energy of binding (ΔG_{bind})[146]. Complex with the lowest ΔG_{bind} is ranked the most stable and is returned as the most likely candidate, along with a few slightly less stable options. Different docking programmes use different types of scoring functions to rank the results. Their main limitation is the requirement for fast calculation as the algorithm needs to be able to scan through thousands of complexes. Because of this limitation, the binding energy calculations are often inaccurate and the software may struggle to correctly rank the results with similar ligand conformations[146,147]. The task becomes even more challenging in the case of carbohydrate ligands which are characterised by high flexibility and additional anomeric, exoanomeric and gauche effects which are not currently incorporated into the docking algorithms[148,149]. Nevertheless, molecular docking has been successfully used to model GAG-

protein interactions on multiple occasions, for example in investigation of heparin-binding proteins[150–153].

AutoDock and AutoDock Vina are two commonly used molecular docking programmes, frequently applied in the studies of protein-carbohydrate complexes[150–152,154]. Autodock is a suite of non-commercial software that uses a genetic-type algorithm to explore different ligand conformations and rank them with the use of empirical free energy force field[155,156]. Autodock Vina is an improved version of the Autodock, applying a modified global search algorithm and a more efficient scoring function which largely improves the speed and accuracy of the docking calculations[147]. In both programmes, the ligand is treated as flexible, with a possibility to restrain some of the bonds, and the receptor is treated as rigid, with an option to assign some of the residues as flexible. Although it would be preferable to treat both the ligand and the receptor as flexible, such approach is computationally challenging and is not currently implemented in the AutoDock software.

Several factors can decrease the accuracy of the docking. As protein-GAG interactions are electrostatic in nature, with poor geometric complementarity between the oligosaccharides and their binding sites, there are many possibilities for a single oligosaccharide fragment to bind to the protein surface. In some of the previous studies, multiple orientations were identified for GAG ligands as equally plausible based on the estimated energetic contributions[150]. Some proteins have multiple sites able to recognise GAGs and the residues that participate in the interaction are often characterised by long, flexible side chains that can adopt many conformations, difficult to explore using the docking software. Current docking algorithms are optimised for docking of small molecules and lack the tools that support docking of carbohydrate ligands. Finally, as the docking algorithms are stochastic in nature, there is no guarantee that they will succeed in finding the global minimum[154].

For the above reasons, it is important to have a method that allows to verify the docking results and identify the most likely binding conformations. One of the strategies used to validate and optimise the docking results is to perform molecular dynamics (MD) simulations on the obtained complexes. In MD experiments, all atoms are allowed to move which means that the whole system is represented more accurately. The amino acid side chains are treated as flexible and the protein is free to move and alter its conformation to adjust to the changes induced by the ligand binding. By allowing such changes to occur in many steps, MD simulation is a useful tool for checking the stability of the complex and finding its local minimum[146,157]. Moreover, availability of the carbohydrate-specific force fields allows for more accurate predictions of the carbohydrate conformation and dynamic behaviour during

the interaction. MD simulations provide an opportunity to explore the movement of flexible loops within the protein, investigate interdomain dynamics and to observe whether the position of the ligand in the binding site undergoes any changes with time.

An important factor that affects the accuracy of MD predictions is the precision of the force fields determining the energy terms and force constants used to predict the atomic movement, bond lengths, angles and torsions, as well as the non-bonding interactions. Force fields optimised for MD simulations of proteins and carbohydrates are gathered in the Amber Molecular Dynamics Package which also contains numerous programmes and tools for the preparation, set-up and analysis of MD simulations on biomolecules[158,159]. Currently, the best available force field for MD simulations of proteins is ff14SB[160], whilst the force field optimised for carbohydrates is GLYCAM-06, developed by the Woods group[161,162]. The combination of these two force fields have been used previously to model the interactions between carbohydrates and proteins such as the fibroblast growth factor-2, chondroitinase AC, hemagglutinins and galectin-3[154,163,164].

As Autodock and Amber packages were previously used to investigate GAG binding by a variety of receptor proteins, they should be suitable for the studies on the binding of fCS by selectins. Through a collaboration with Prof. Wood's group at the University of Georgia, our group has acquired protocols that allow for preparation of appropriate input files, as well as for molecular docking in Autodock Vina and molecular dynamics simulations on the produced protein-ligand complexes using Amber MD package. With the use of these protocols, it was possible to build different fCS oligosaccharides and to test the nature of their binding to selectins by molecular docking. Results were validated by MD simulations which also provided further insight into the dynamic behaviour of selectin domains and their carbohydrate ligands.

In the following sections, computational studies on selectin-ligand binding will be described. In the first set of experiments, protocols for preparation of the receptor and ligand files, docking experiments and MD simulations were verified using the sLe^x-E-selectin complex for which the X-ray structure was known[66]. This was followed by producing different types of the fCS oligosaccharides and docking them to different selectin structures. Results of these experiments were verified and analysed with the use of MD simulations to provide new information on the nature of the fCS-selectin binding and conformational behaviour of selectin C-type lectin and EGF-like domains in these complexes.

5. 2 Materials and Methods

5. 2. 1. PDB file preparation for MD and docking studies

The initial three-dimensional structures of the sLe^x and fCS oligosaccharides were generated using the Carbohydrate Builder available on the GLYCAM Website. Protein structures used in the docking experiments and MD simulations were extracted from the X-ray structure files deposited in the Protein Data Bank (entries 1ESL and 1G1T for E-selectin, 1G1Q for P-selectin). The produced structures were prepared for MD simulation using tLEaP module of AMBER 2014 software suite. Sulfate charges were obtained from the GLYCAM Website. The force fields used to generate protein and carbohydrate parameters were ff14SB and GLYCAM_06j-1, respectively[160,163].

5. 2. 2. MD simulation set-up

All MD simulations were performed using the CUDA implementation of PMEMD in the AMBER 14 Software Suite[159,165,166]. Structures were solvated in a truncated octahedron periodic box filled with TIP3P water molecules and neutralised by counterions. Prior to MD simulation, energy minimisation was performed in 5000 steps of steepest descent followed by 5000 steps of conjugate gradient using nonbonding cut-off of 10.0 Å with no restraints. The minimisation steps were followed by a heating phase, where the system was brought from 5 to 300 K over 50 ps. A simulation of 10 – 100 ns was performed in the isothermal-isobaric ensemble (NPT) with Langevin temperature coupling with collision frequency parameter $\gamma = 5\ ps^{-1}$ and Berendsen pressure coupling with a time constant of 1.0 ps. The SHAKE algorithm was used to constrain all bonds that contain hydrogen atoms. A 10 Å cut-off was applied to treat non-bonded interactions, the Particle Mesh Ewald (PME) method was introduced for long-range electrostatic interactions treatment and the long-range van der Waals interactions were estimated by a continuum model. Molecular graphics and analysis of the simulation results were performed with the UCSF Chimera package[167].

5. 2. 3. Docking experiments

Protein and carbohydrate files were prepared in AutoDockTools (ADT) 1.5.6 and the docking experiments were performed in AutoDock Vina 1.1.2.[147]. Gasteiger charges were assigned to both protein and ligand files. Hydrogen atoms were added to the protein using ADT and GLYCAM hydrogens were retained in the ligands. Size of grid box was varied, depending on the size of the carbohydrate ligand, to always include the full C-type lectin domain, with the Ca^{2+}-coordination site at the centre and allocating enough space for the ligand to explore different conformational states in solution and positions along the protein surface. The protein was kept rigid in all of the experiments.

5. 2. 4. Free energy calculation

Molecular Mechanics Linearisead Poisson-Boltzman on (MM-PBSA) method was used to calculate the free binding energy for the produced protein-carbohydrate complexes. In this method, the following terms are used to calculate the free energy of protein-ligand binding in solution ($\Delta G_{bind,solv}$): the free energy of protein-ligand binding in vacuum ($\Delta G_{bind,\,vac}$), the free energy of the protein-ligand complex in solution ($\Delta G_{complex,\,solv}$) and the free energies of solvation of the unbound protein ($\Delta G_{protein,solv}$) and ligand ($\Delta G_{ligand,solv}$). From these terms, the free energy of binding is calculated using Equation 1.

$$\Delta G_{bind,solv} = \Delta G_{bind,vac} + \Delta G_{complex,solv} - \left[\Delta G_{protein,solv} + \Delta G_{ligand,solv}\right] \quad (1)$$

Each of the three solvation free energy terms is calculated by solving the linearised Poisson-Boltzman equation and adding an empirical energy term for hydrophobic contributions as shown in Equation 2.

$$\Delta G_{solv} = \Delta G_{electrostatic,\epsilon=80} + \Delta G_{electrostatic,\epsilon=1} + \Delta G_{hydrophobic} \quad (2)$$

The free energy of binding in vacuum can be obtained by calculating the average interaction energy between the receptor and ligand and taking the entropy change upon binding into account (Equation 3) [168,169].

$$\Delta G_{vac} = \Delta E_{molecular\ mechanics} - T\Delta S \quad (3)$$

However, the entropy calculations are computationally costly and in cases where only a comparison of states of similar entropy is required, this term can be omitted[169,170]. For this reason, in the calculations presented here, normal mode analysis for entropy term was not performed.

The average interaction energies for receptor and ligand were obtained from a set of uncorrelated snapshots collected from MD simulation, using the *mm_pbsa.pl* module of Amber 14. For each complex, ΔG_{bind} was calculated for 300 snapshots from the MD trajectory file, taken every 10 ps for the first 3 ns of the simulation. The final ΔG_{bind} value was the average of the calculated ΔG_{bind} values for these snapshots.

5. 3. Results and discussion

5. 3. 1. Sialyl Lewis X and its interaction with E-selectin

In order to establish the suitability of computational protocols for the studies of fCS-selectin interactions, it was deemed necessary to test the procedures on a known carbohydrate ligand with experimental data on its binding to selectins. The obvious choice for selectin studies was the sLe^x tetrasaccharide for which crystal structures, in complex with either E- or P-selectin, are available. As the interaction with E-selectin is known to be stronger, sLe^x binding to this protein was explored in the first set of tests. The main aim of this control study was to build the sLe^x tetrasaccharide using GLYCAM-Web Carbohydrate Builder and to establish whether it is possible to predict its binding site on selectin surface using the molecular docking software Autodock Vina. Provided that the docking experiments were successful, the most suitable docked structure would be minimised and further evaluated by MD simulation using the Amber Molecular Dynamics Package.

5. 3. 1. 1. Preparation of the sLe^x tetrasaccharide

GLYCAM-Web Carbohydrate Builder is an on-line tool developed by the Woods group at the University of Georgia that allows the users to build and minimise a carbohydrate structure of interest, saving the coordinates into a PDB file in a format suitable for the use with Amber MD Package. sLe^x tetrasaccharide was built using the formula DNeup5Aca2-3DGalpb1-4[LFucpa1-3]DGlcpNAcb1-OME. Minimisation produced a conformation that was further minimised with the use of Amber MD Package under GLYCAM06 force field. MD simulation in explicit water was performed for 40 ns. Data obtained from this simulation was analysed to provide information about the stability of the produced structure, its conformation and dynamic behaviour in solution.

The temperature and energy of the system remained stable throughout the simulation (plots are included in Appendix 3). The Root Mean Square Deviation (RMSD) with respect to the lowest energy conformation was calculated. All of the atoms, except for hydrogen, were

included in the calculation. The RMSD plot presented fluctuations of about 1.5 Å over the first 10 ns of the simulation. At close to 10 ns, a sudden change of about 2 Å was observed, after which the system reached a plateau, dropping back to the initial level after about 5 ns (Fig. 5. 3. 1. A). Such a sudden change in RMSD from average value of 1.56 Å to 3.46 Å occurred two more times, each time reaching a stable plateau at the higher RMSD values. This indicates the presence of two distinct conformations where one seems more energetically stable than the other (Fig. 5. 3. 1. A).

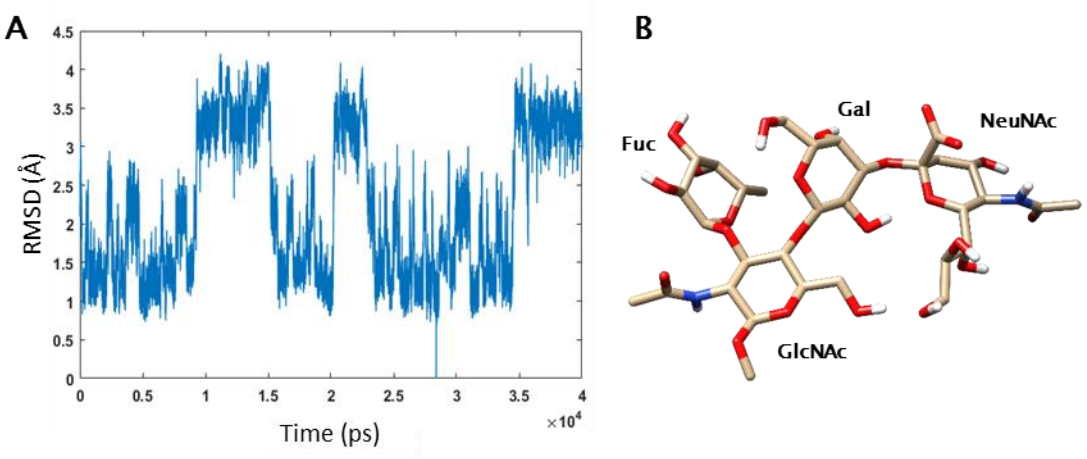


Fig. 5. 3. 1. MD simulation of the produced sLe^x tetrasaccharides. A. Plot of RMSD for sLe^x tetrasaccharide over a 40-ns MD simulation. **B.** The lowest energy conformation of sLe^x tetrasaccharide from the 40-ns simulation.

Following the analysis of the values of dihedral angles over the course of the simulation, changes in RMSD values were correlated to changes in the value of dihedral angle ϕ for glycosidic linkage Neu5Acβ2⟶3Gal from the average value of 59° to -51° (Fig. 5. 3. 2). This result shows that the linkage between sialic acid and galactose is highly flexible and leads to existence of two major conformations of the free tetrasaccharide that are in constant exchange throughout the simulation. Based on the RMSD fluctuations for each of the two species, it appears that the conformations with ϕ of ca. 59° for this linkage are energetically more stable than those with low ϕ values. Analysis of the X-ray structure 1G1T confirms this observation, as the observed dihedral angle ϕ value for this linkage was 57° (Table 5. 3. 1). All of the other glycosidic dihedral angles remained stable, with overall fluctuations of less than 40°, which indicates that the remaining two glycosidic linkages did not undergo any major conformational changes.

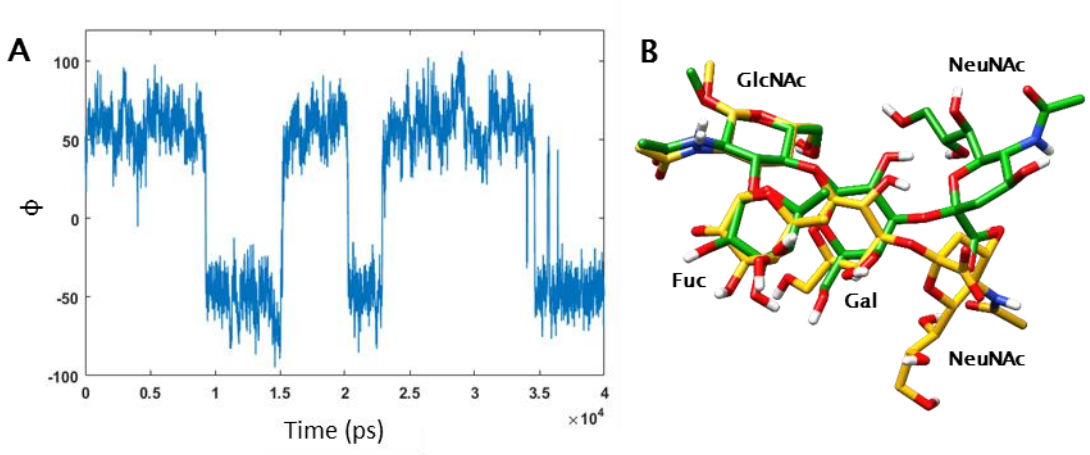


Fig. 5. 3. 2. Changes in dihedral angles for sLe^x tetrasaccharide. A. Changes in dihedral angle ϕ for glycosidic linkage NeuNAcβ2⟶3Gal of sLe^x tetrasaccharide over a 40-ns MD simulation. **B.** Overlay of two conformations of sLe^x tetrasaccharide observed at 1 ns (green) and 10 ns (yellow).

Table 5. 3. 1. Average dihedral angles from the 40-ns MD simulation, compared to the values reported in literature and those extracted from the X-ray structure.

Glycosidic linkage	Torsion angle	Lit. NMR value[171]	Lit. NMR value[172]	X-ray Structure (1G1T)	MD
Galβ1⟶4GlcNAc	**ϕ**	-55	-66	-85	-68.4 ± 0.4
	ψ	120	135	136	130.9 ± 0.3
Fucα1⟶3GlcNAc	**ϕ**	-45	-73	-77	-69.0 ± 0.2
	ψ	-95	-91	-97	-96.0 ± 0.3
NeuNAcβ2⟶3Gal	**ϕ**	N/A**	N/A**	57	22 ± 24
	ψ	N/A**	N/A**	-130	-121 ± 5

*All measurements were made using the following definition of dihedral angles: ϕ = O5 – C1 – OX – C'X, ψ = C1 – OX – C'X – C'(X-1).
**The literature values refer to non-sialylated Le^x determinant.

The average values of dihedral angles for the linkages between glucosamine and galactose (Galβ1⟶4GlcNAc) and glucosamine and fucose (Fucα1⟶3GlcNAc) were also measured and compared to the values previously reported in literature, as well as to the values extracted from the X-ray structure (Table 5. 3. 1.). The average values were within 25° from the experimentally obtained values and it was concluded that the computationally produced oligosaccharide is in a good enough agreement and therefore suitable for the docking studies. Uncertainty of the mean calculation was also determined using the block averaging technique, where the MD trajectory was divided into 10 ns blocks and the mean value of dihedral angles was calculated for each block. Linkages Galβ1⟶4GlcNAc and Fucα1⟶3GlcNAc were characterised by low uncertainty values, consistent with low amount of fluctuations observed for these bonds. Uncertainty was higher for the linkage Neu5Acβ2⟶3Gal, indicating the higher flexibility of this bond.

All conformations occurring over the 40 ns simulation were clustered using the UCSF Chimera Cluster tool. This tool uses pair-wise RMSD fit calculation to cluster all of the included conformations into groups based on their similarity. Total of 4000 conformations were used for this calculation, using every 10^{th} frame in the simulation. This procedure allowed to explore the conformational space of sLe^x tetrasaccharide and to isolate the most commonly occurring conformations to be used in the docking studies. sLe^x oligosaccharide proved to be highly flexible in solution, with over 180 clusters identified. Comparison of the clusters showed that there is high similarity in the values of dihedral angles ϕ and ψ for two of the linkages, Galβ1→4GlcNAc and Fucα1→3GlcNAc, between all of the clusters (Fig. 5. 3. 3. A). Only the linkage NeuNAcβ2→3Gal allows to distinguish between different clusters due to its high flexibility (Fig. 5. 3. 3. B). The most populated cluster (Cluster I) represented 10% of the included conformations, with Clusters II – V representing 7%, 5%, 4% and 3% of the population, respectively. As there is a high overlap between the clusters in the values of dihedral angles, it would be difficult to identify a single conformation unique for each cluster. Five representative conformations were chosen based on the values of dihedral angles closest to the average for a given cluster.

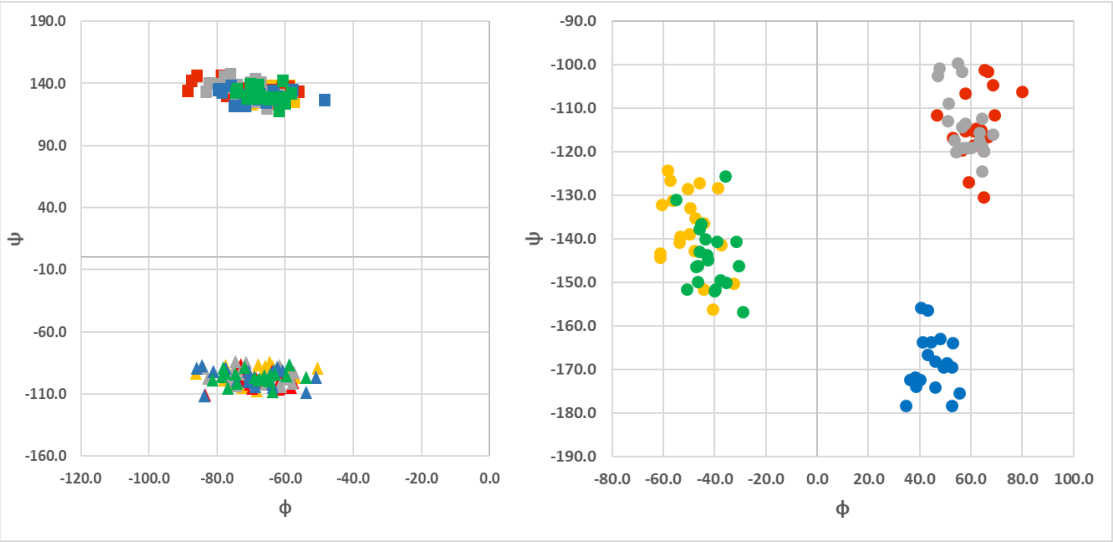


Fig. 5. 3. 3. Distribution of dihedral angles ϕ and ψ within the simulated sLe^x tetrasaccharide. A. Plot of ψ versus ϕ for the glycosidic linkages Galβ1→4GlcNAc (squares) and Fucα1→3GlcNAc (triangles). **B.** Plot of ψ versus ϕ for the glycosidic linkage NeuNAcβ2→3Gal. Five clusters are represented in yellow, red, grey, blue and green for Clusters I to V respectively. 20 representatives were chosen at random from each cluster as data points for this plot. All measurements were made using the following definition of dihedral angles: ϕ = O5 – C1 – OX – C'X, ψ = C1 – OX – C'X – C'(X-1).

5. 3. 1. 2. Docking experiments on sLe^x and E-selectin

The five sLe^x conformations prepared in the previous section were docked to existing X-ray structures of E-selectin. Oligosaccharides were docked initially to E-selectin from the crystal structure 1G1T and then to E-selectin from the structure 1ESL. These two structures have slight differences in the conformation of some of the loops and amino acid side chains and therefore may produce slightly different results upon docking to sLe^x. The protein was kept rigid in all experiments.

All carbohydrate bonds were treated as rotatable in the docking experiments, remaining within the limit of maximal number of 36 rotatable bonds advised for docking in Autodock Vina[147]. Diagram in Fig. 5. 3. 4 shows the approximate size of the search box. It was important to ensure that the search box is large enough to allow the software to explore different possible binding sites, as well as to allow the carbohydrate linkages to freely rotate, exploring different conformations.

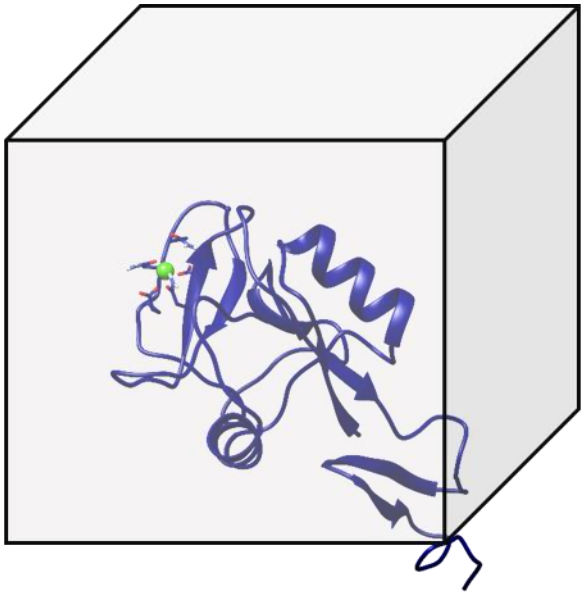


Fig. 5. 3. 4. Approximate size of the search box explored in the docking experiments. The light grey cube represents the search box specified for the docking software to explore possible binding sites and conformations of sLe^x. E-selectin is shown in dark blue.

For each docking experiment, Autodock Vina returned nine results, ranked according to the scoring energy function. A mixture of results was obtained for each input sugar conformation, with the majority of results identifying the Ca^{2+}-coordination site as the sLe^x-binding site. Each of the starting sLe^x conformations generated a slightly different set of results, with many different possible orientations of sLe^x in the binding site (Fig. 5. 3. 5).

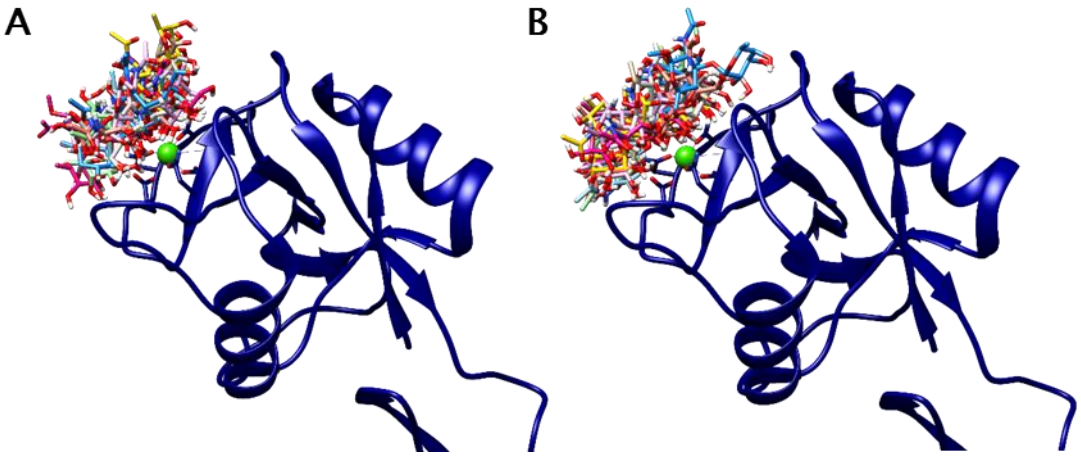


Fig. 5. 3. 5. Two sets of the docking results for sLe^x tetrasaccharide and E-selectin. Autodock Vina returns nine results for each docking experiment. In the above figure, nine results obtained after docking the sLe^x conformations I (**A**) and II (**B**) to E-selectin structure 1G1T are superimposed. Each colour represents a different docking output from the same experiment. The ribbon representation of E-selectin is shown in dark blue.

None of the results were identical to those reported in the X-ray studies. As described in Section 1. 3. 3., X-ray studies have determined that the sLe^x tetrasaccharide coordinates Ca^{2+} ion via Fuc 3- and 4-hydroxyl groups. Only in a few of the docking outputs Ca^{2+} ion was coordinated by fucose and in all of them mode of Ca^{2+} coordination was different to that observed in the X-ray structure. Nevertheless, these structures were evaluated by a set of short (10 – 15 ns) MD simulations using Amber MD Package. Structures that remained stable throughout the simulation were put forward for a longer MD run. Although a few candidates made it through the first set of the simulations, only one was found that did not dissociate or weakened significantly during a longer simulation. This complex was one of the results produced during the docking of sLe^x to the E-selectin structure 1ESL.

The total simulation time for the most stable complex was 70 ns. The complex remained stable throughout the full duration of the simulation, with RMSD maintaining the average value of 1.52 Å with respect to the lowest energy structure. The calculated free binding energy for this complex was equal to -15.7 ± 0.4 kcal mol^{-1}. The negative value of binding energy confirms that the complex is favourable. From to the relationship presented in Equation 4 and the experimentally determined value of K_D for the E-selectin-sLe^x binding (720 nM), it was possible to calculate the theoretical value of ΔG_{bind}.

$$\Delta G_{bind} = RTlnK_D \quad (4)$$

The calculated value for K_D was -4.28 kcal mol^{-1} which is a little higher than the value calculated by the molecular dynamics software. However, the values are close enough to conclude that the energy values calculated by the software can be used to estimate the strength of the interactions between selectins and carbohydrate ligands. It was also noted that ΔG_{bind} decreased at longer simulation times, indicating that the complex became more stable after the first 30 ns of the MD run. The conformation of sLe^x oligosaccharide and the protein-ligand interactions were compared to those reported in the X-ray structure 1G1T by visual assessment and by comparing the interactions formed by individual carbohydrate residues to E-selectin surface (Fig 5. 3. 6).

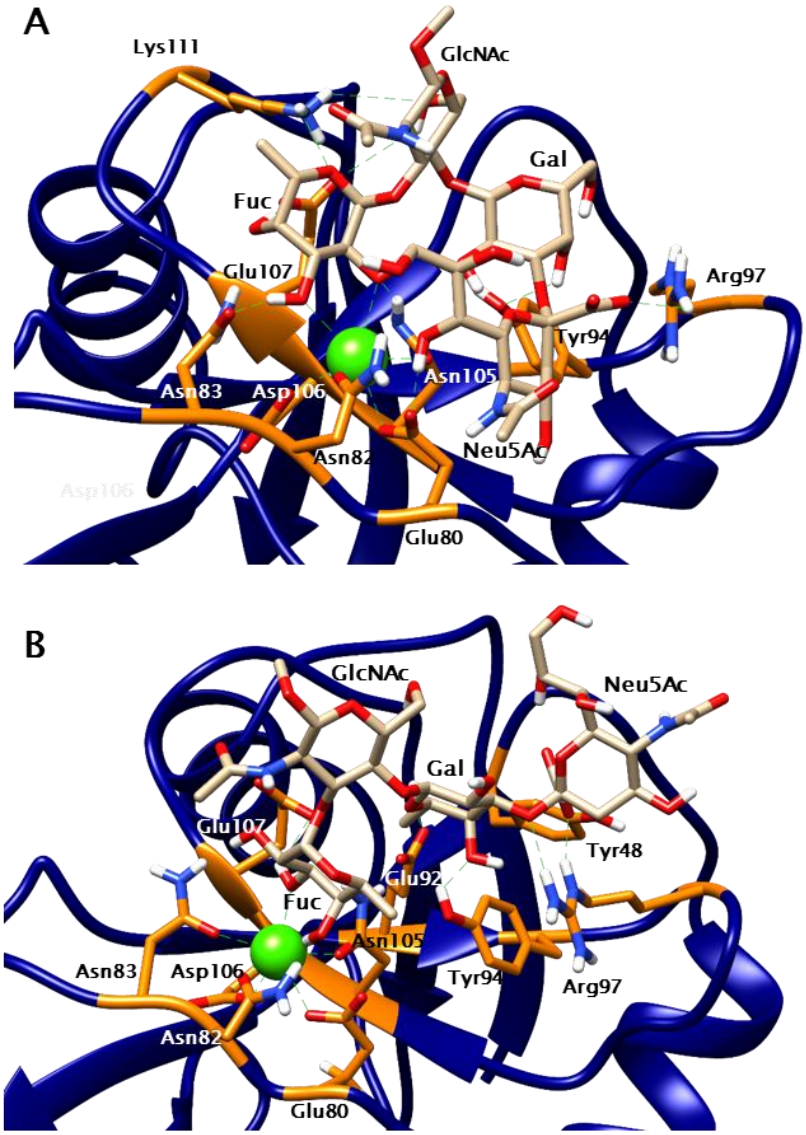


Fig. 5. 3. 6. Comparison of the docking result to the X-ray structure complex. A Ribbon representation of the complex obtained after docking of sLe^x tetrasaccharide to E-selectin structure 1ESL. Shown here are the interactions after the initial minimisation of the complex. **B** The experimentally obtained complex between sLe^x and E-selectin from X-ray structure 1G1T. Proteins are shown in blue, with the interacting residues highlighted in orange. Carbohydrate residues are shown in light grey. Hydrogens are not shown here for clarity.

The sLe^x-E-selectin interactions identified in this complex show many similarities to those reported in the X-ray structure 1G1T. The majority of these interactions remained stable throughout the 70 ns simulation and the carbohydrate remained firmly bound to the protein. Large proportion of the interactions occurs via fucose residue which not only coordinates the Ca^{2+} ion via FucO2 and FucO3 atoms (mean value of 2.79 Å and 2.68 Å, respectively), but also participates in a network of H-bonds that stabilise its position at the Ca^{2+}-coordination site.

FucOH2 group forms a stable H-bond to Asn105 (2.90 Å), FucOH3 – to Asn83 (1.88 Å), FucOH4 – to Glu107 (1.85 Å) and FucO5 – to Lys111 (2.98 Å). A similar network of H-bonding interactions was observed in the X-ray structure, involving residues Asn105, Asn83 and Glu107. In addition to the interactions mediated by fucose, there is also an H-bonding interaction between GalO4 and Tyr94 (2.57 Å) and between the carboxyl group of sialic acid and Arg97 (2.91 Å). These last two interactions were also observed in the X-ray structure. Therefore, the computational methods used in this study allowed to replicate the network of interactions determined experimentally to a satisfactory level. To simplify comparison of the interactions between the two complexes, a schematic representation of these interactions in the docked and X-ray complexes is presented in Fig. 5. 3. 7.

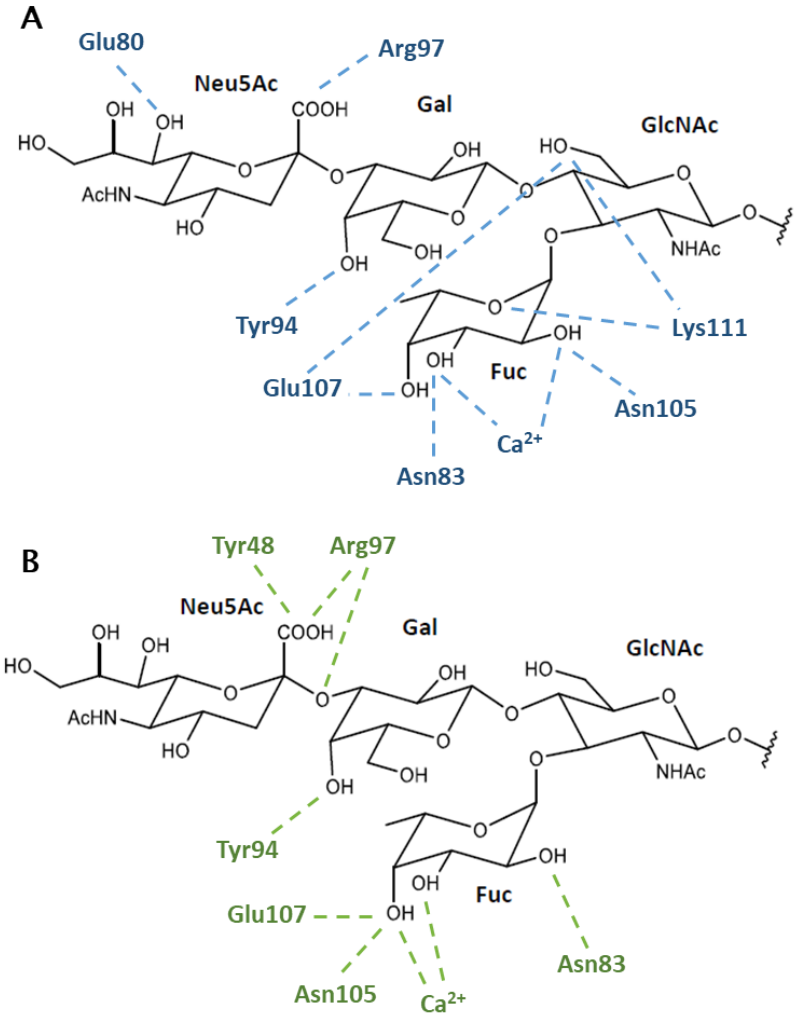


Fig. 5. 3. 7. Schematic representation of the interactions in the docked and the X-ray complexes. A Interactions observed in the docked complex after the initial minimisation. E-selectin residues are shown in blue. **B** Interactions observed in the experimentally obtained complex from X-ray structure 1G1T. E-selectin residues are shown in green. This diagram is provided to simplify the views presented in Fig. 5. 3. 5. and does not reflect the conformation of carbohydrate or orientation of amino acid residues in the binding site.

The main difference between the two structures lies in the orientation of sialic acid residue with respect to the binding site. In the X-ray structure, the carbohydrate conformation is more open, with sialic acid facing away from the Ca^{2+} ion. In the docked structure, sialic acid is folding in towards the Ca^{2+} ion, producing steric hindrance that forces fucose to approach the Ca^{2+}-coordination site from a different angle to what was observed in the X-ray studies. In addition to that, in the docked result, fucose coordinates Ca^{2+} ion via 2- and 3-hydroxyl groups, rather than 3- and 4-hydroxyl groups. Despite these differences, sLe^x remains strongly bound to the protein in such conformation, suggesting that in theory it could be possible for such a mode of binding to occur in nature.

Table 5. 3. 2. compares the dihedral angles between the oligosaccharides in these two complexes. It can be seen that there is a small variation in the values of dihedral angle ϕ for all glycosidic linkages and a large variation in the values of ψ for the glycosidic linkages Galβ1⟶4GlcNAc and Fucα1⟶3GlcNAc. These differences in dihedral angles account for the observed differences in the sLe^x orientation around the Ca^{2+}-coordination site. As the values of ψ in the docked complex are largely different to those reported for the free tetrasaccharide at equilibrium, adopting the proposed bound conformation might involve high energy cost which could explain why such a mode of binding has not been observed experimentally.

Table 5. 3. 2. Comparison of carbohydrate dihedral angles between the X-ray structure 1G1T and the docked sLe^x-E-selectin complex.

Glycosidic linkage	Dihedral angle*	X–ray	Docked complex
Galβ1-4GlcNAc	**Φ**	-85	-69
	Ψ	136	-58
Fucα1-3GlcNAc	**Φ**	-77	-103
	Ψ	-97	93
Neu5Acβ2-3Gal	**Φ**	57	58
	Ψ	-130	-110

*All measurements were made using the following definition of dihedral angles: ϕ = O5 – C1 – OX – C'X, ψ = C1 – OX – C'X – C'(X-1).

Although there are significant differences between the docked complex and the X-ray structure, there are some similarities that indicate that the docking algorithm works reasonably well. Most importantly, Autodock Vina was able to correctly identify the binding

site in the majority of cases as located around the Ca^{2+} ion. Furthermore, a few of the residues involved in the binding of sLe^x were also correctly identified by the docking algorithm. These similarities prove that Autodock Vina can be used to predict the binding site and the amino acid residues involved in the binding of a sLe^x-type oligosaccharide by selectins.

Further docking experiments were performed to establish how the input conformation of the carbohydrate affects the docking result. sLe^x tetrasaccharide from the X-ray structure 1G1T was extracted and docked to both E-selectin files as it was thought that using the true bound conformation may improve the accuracy of the docking algorithm. However, similarly to previous docking experiments, a mixture of results was obtained, with large variation in the orientation of the ligand in the binding site (Fig. 5. 3. 8).

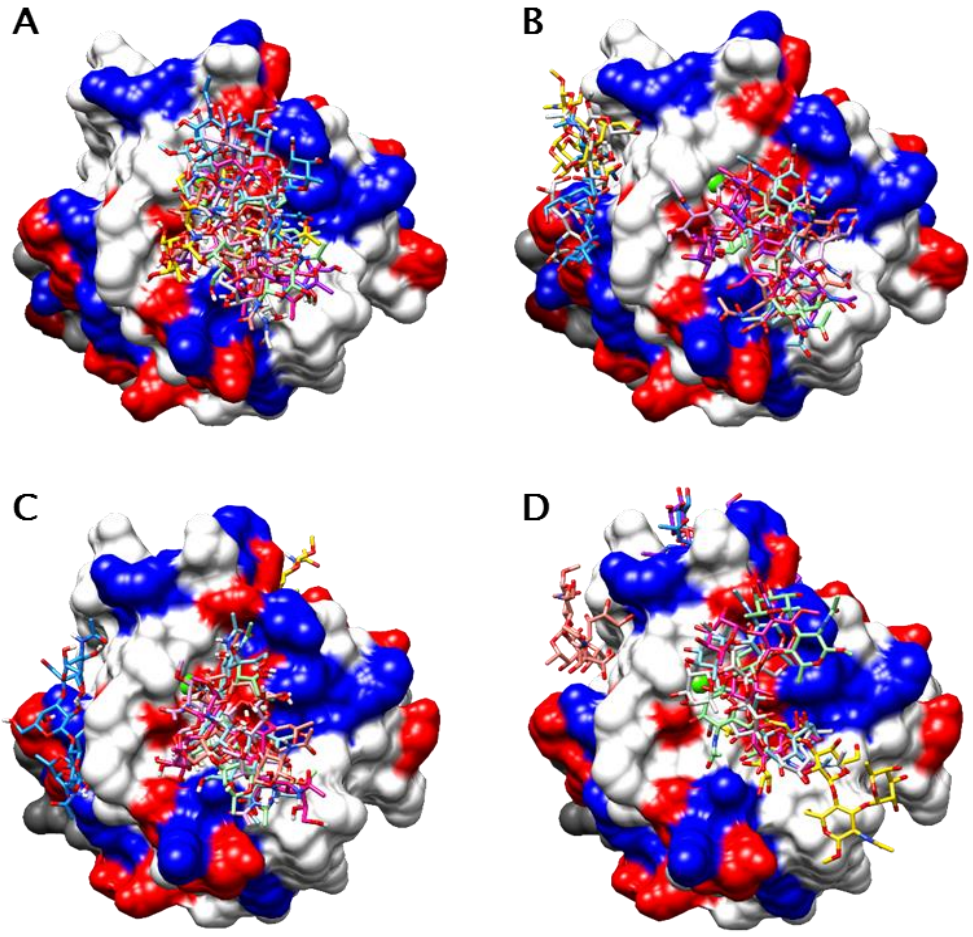


Fig. 5. 3. 8. Docking results for different combinations of the input files. Nine docking results returned by the Autodock Vina are superimposed for four different combinations of the input files. **A** and **C** present docking results for computationally produced sLe^x conformation I and protein structure 1G1T and 1ESL, respectively. **B** and **D** present docking results for the sLe^x conformation extracted from the X-ray structure 1G1T and protein structure 1G1T and 1ESL, respectively. The protein is shown in a surface representation, with positively charged residues highlighted in blue and negatively charged residues – in red.

None of the docked complexes offered a more suitable match for the existing X-ray structure than the complexes produced upon docking of the computationally produced tetrasaccharide. Use of an input conformation identical to the true bound conformation of sLe^x in the docking experiments did not significantly improve the results. After comparison of the docking results for the different sLe^x conformations, it was concluded that, in the case of a fully flexible oligosaccharide ligand such as sLe^x, use of different starting conformations of the carbohydrate does not significantly improve the accuracy of the docking results. Autodock Vina is optimised for the docking of small molecule ligands with a minimal number of rotatable bonds and the large number of rotatable bonds in the carbohydrate ligands leads to large variability in the docked results without a single strongly preferred orientation.

Fig 5. 3. 8 shows how the docking results are distributed around the binding site, highlighting the positively and negatively charged amino acid residues. As these are the residues that participate in H-bonding and electrostatic interactions, it is of no surprise that the carbohydrate preferentially binds to the regions rich in those residues. No clear preference for binding to positively charged residues over the negatively charged residues was observed for sLe^x. It is also noteworthy that, occasionally, results include structures that bind in different regions of the protein (Fig 5. 3. 8. B, C, D). However, in the case of sLe^x, such outcomes were statistically rare, their binding regions did not overlap with each other and they were treated as outliers and not studied in any more detail.

Docking experiments on sLe^x showed that there are many possible outcomes of the docking algorithms that are energetically similar to one another and that it may be difficult to distinguish between these candidate conformations without the help of experimental data. As such data for the fCS-selectin interaction is very scarce, it may not be possible to determine the exact binding site and carbohydrate conformation with confidence. However, MD simulations could be used to test the stability of the proposed binding modes and to use this information to rank them in order of likelihood of such mode of binding occurring *in vivo*.

5. 3. 1. 3. MD study of the X-ray structures of sLe^x–E–selectin complex

For comparison, the sLe^x-E-selectin complex from PDB entry 1G1T was also minimised and simulated for 50 ns using the same protocol as for the docked complex. This experiment allowed to study the stability and dynamic behaviour of the complex obtained in the X-ray experiments and compare it to the docked sLe^x-E-selectin complex.

Protein-ligand interactions reported in the X-ray studies were monitored over the course of the MD simulation. Fucose interactions with the Ca^{2+} ion proved to be very stable, with only slight oscillations of less than 0.5 Å. The distances between the Ca^{2+} ion and FucO3 and O4 both maintained the average value of 2.75 Å. The sLe^x-Ca^{2+} interaction was stabilised by H-bonding interactions between fucose and Glu107 (average of 2.45 Å) and Asn105 (2.82 Å). Arg97 maintained a stable interaction with the carboxyl group of sialic acid of the mean value of 2.79 Å and Tyr94 maintained a weak electrostatic interaction with GalO4 (3.83 Å). These results showed that several intermolecular interactions remained viable throughout the full duration of the simulation. However, the remaining interactions identified in the X-ray studies dissociated within the first 2 ns of the simulation.

RMSD over the 50 ns simulation was calculated with respect to the lowest energy structure. Although the overall RMSD remained stable throughout the majority of the simulation (average value of 1.94 Å), a sudden change of close to 2 Å occurred just before 40 ns. The RMSD was unstable after this point, returning to the initial level towards the end of the simulation. RMSD contributions of sLe^x and the E-selectin were also calculated and allowed to establish that the changes in RMSD by the end of the simulation must correspond to conformational changes within the protein (Fig. 5. 3. 9. A).

This sudden change in the value of RMSD was correlated to a conformational change in the interdomain hinge of E-selectin. For the first 37 ns of the simulation, the C-type lectin and the EGF-like domains maintained the so-called 'bent' conformation, with the interdomain angle of approximately 115 - 120°. At 39 ns, this angle began to increase, as the EGF-like domain rotated away from the lectin domain, to reach a value of 154° at 45 ns (Fig. 5. 3. 9. B). This value of the interdomain angle is consistent with the literature reports of the so-called 'extended' conformation in the X-ray and MD studies on E-selectin[67]. After 45 ns, the EGF-like domain started to rotate back towards the lectin domain, returning to the initial 'bent' conformation. It is noteworthy that the switch between the bent and extended conformations has been linked in literature to the change in orientation of the loop 81-89 and believed to cause changes in affinity for carbohydrate ligands. However, in this experiment, the loop did

not switch to the position reported in the literature at any point of the simulation and no correlation was observed between the interdomain movement and the stability of protein-ligand interactions.

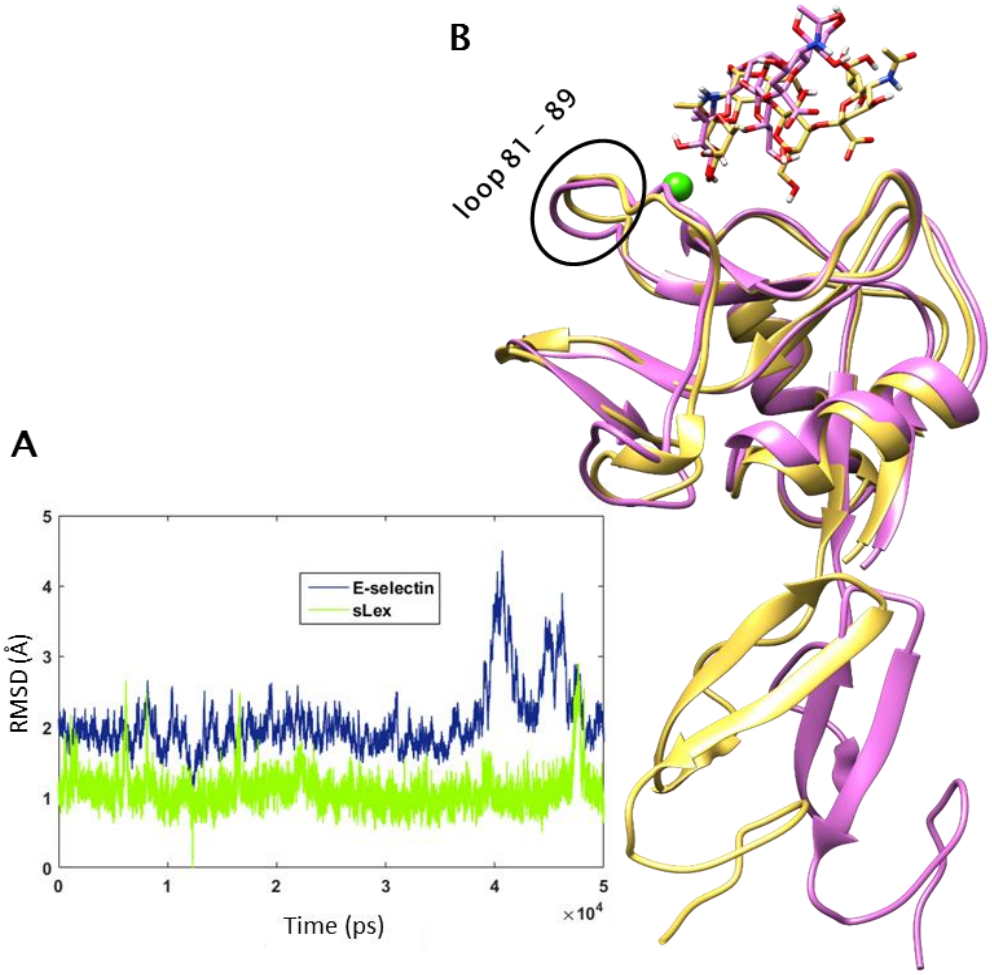


Fig. 5. 3. 9. MD simulation of the sLe^x-E-selectin complex from the X-ray structure 1G1T. **A.** RMSD plot over 50 ns MD simulation for individual E-selectin (blue) and sLe^x (green) contributions. **B.** Ribbon representation of E-selectin at 0 ns (yellow) and 40 ns (purple) into the simulation. Transition between the bent and extended conformations can be observed. No rearrangement is observed in the loop 81 – 89, previously implied to play a role in conformational switch between bent and extended conformations.

The binding free energy was also calculated for the X-ray 1G1T complex, using the same protocol as for the docked complex. However, in this case ΔG_{bind} had a positive value of 4.5 ± 0.3 kcal mol^{-1}. It is surprising that the value is positive, since positive free binding energy values are normally indicative of the complex being unfavourable *in vivo*. However, it has to be emphasised that this is only a predicted value, based on the molecular force fields and calculated energetic contributions. It is important to note that protein-carbohydrate

interactions are inherently weak and, therefore they may be difficult to predict with the use of existing MD force fields, more suitable for studies on stronger-affinity complexes. The MM-PBSA calculations also suffer from inaccuracies in prediction of the entropy contributions, which may cause the ΔG_{bind} values to be overestimated[169,173]. Therefore, the calculated value of ΔG_{bind} is only an estimate and it cannot be treated as an accurate measure of the free binding energy. However, as the same protein-ligand combination is used, the values of ΔG_{bind} can be compared to provide an indication of relative stability of each complex.

1G1T is not the only available X-ray structure of the sLe^x-E-selectin complex, though it is the highest resolution structure available to date. In a more recent study by Preston *et al*., X-ray structure of a larger E-selectin construct, in complex with sLe^x, was obtained (PDB entry 4CSY)[67]. In this study, the ligand was co-crystallised with the protein, rather than soaked into protein crystals, which means that this structure may be a more accurate representation of the actual conformational state of the protein upon interaction with sLe^x. The main difference when compared to the previous structure, is that the E-selectin is present in an extended conformation, with interdomain angle of 136°. Another significant difference is the position of the loop 81 – 89, which is now pointing upwards, closing in on the Ca^{2+}-coordination site. In fact, Glu88 replaces Asn83 as a Ca^{2+}-coordinating residue and forms additional interaction with the fucose residue. The interdomain interactions are weakened in this state, as the EGF-like domain is rotated away from the lectin domain, and small rearrangements are observed in the loops across both domains. It has been speculated that such changes can lead to increased affinity for the sLe^x ligand. To test this hypothesis and compare the dynamic behaviour of the two structures, a 60 ns MD simulation was performed on this complex using the same methods as in the previous experiments.

The energy of the system remained stable and the RMSD measured for the carbohydrate also remained stable throughout the simulation, with the mean value of 1.06 Å with respect to the lowest energy structure (Fig. 5. 3. 10. A). The RMSD for E-selectin underwent large fluctuations over the first 25 ns of the simulation. However, just before 30 ns, the RMSD values stabilised at the mean value of 1.87 Å and remained stable for the rest of the simulation (Fig. 5. 3. 10. A). Further analysis allowed to establish that the fluctuations observed in the first half of the simulation correlated with the movement of the EGF-like domain. Against expectations, the EGF-like domain did not move towards the lectin domain to achieve the bent conformational state, but rotated away from the lectin domain to face the opposite site of the protein, in what could be called a 'hyper-extended' state (Fig. 5. 3. 10. B). This rotation of 50° occurred along a different axis than the opening from the bent to extended state and can be thought of as a twisting movement which brings the C-terminal

tail to the opposite site of the protein than in the bent conformation. In this new conformational state, all residues previously involved in the interactions with the lectin domain, except for Glu135, have moved away from the protein's N-terminus and a new set of interactions has developed between residues Ser2 – Tyr3 of the N-terminal β-sheet and residues Thr136 – Asn138 of the EGF-like domain.

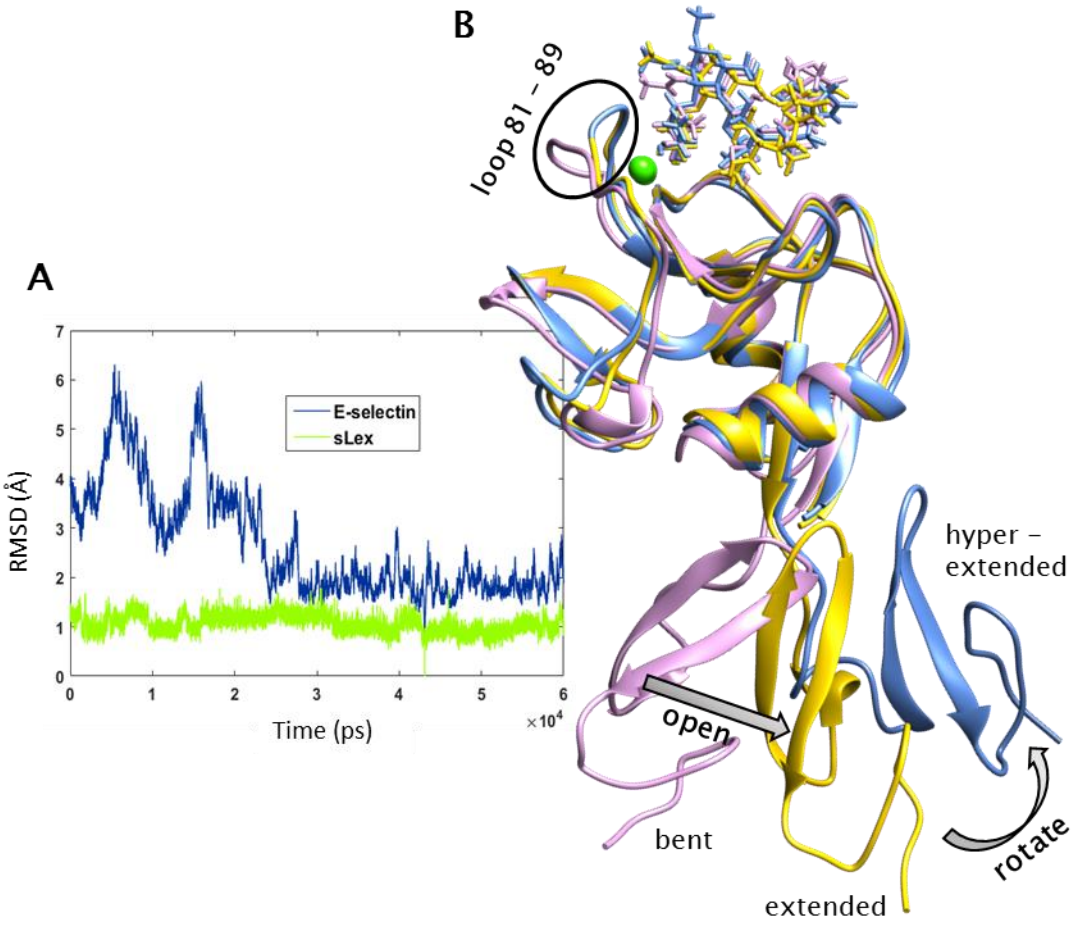


Fig. 5. 3. 10. MD simulation of the sLe^x-E-selectin complex from structure 4CSY. A. RMSD plot for E-selectin (blue) and sLe^x (green) over a 60 ns MD simulation. **B.** Ribbon representation of E-selectin at 0 ns (yellow) and 60 ns (blue) into the simulation. Bent conformation from 1G1T structure is shown for comparison (purple). The interdomain movement occurred just before 20 ns and, after a few ns of equilibration the protein found a stable conformation. EGF-like domain rotated by 50° with respect to its position in the extended conformation to achieve the hyper-extended state. Highlighted loop 83-89 is positioned closer to the binding site in this structure, increasing stability of the sLe^x-E-selectin complex.

Such interdomain movement has not been previously reported for selectins or observed in MD simulations on any other selectin structure used in this study. As the same methods have been applied to study all E-selectin structures, this result suggests that the starting conformation used here has features that promote higher flexibility in the interdomain region. It remains unclear whether this behaviour is affected by the position of the loop 81 – 89. The loop did not switch its position throughout the simulation and maintained the interaction with the Ca^{2+} ion and the Fuc2OH group via residue Glu88 (1.76 Å).

This additional stabilisation from loop 81-89 strengthens the interaction with sLe^x, making it energetically more favourable. The calculated free binding energy of the interaction for the 4CSY complex had a value of -22.8 ± 0.3 kcal mol^{-1}, which is significantly lower than the values calculated for the complex from X-ray structure 1G1T. Although the initial set of interactions was highly similar to those reported in the X-ray structure 1G1T, binding of sLe^x was more stable in the case of 4CSY. The majority of the interactions remained stable for the full length of the simulation, providing the ligand with a larger number of stabilising interactions than in the case of the MD simulation on 1G1T complex. Again, Ca^{2+} ion was coordinated by FucO3 (2.70 Å) and O4 (2.75 Å) and this interaction was further supported by H-bonds to Glu80 (1.77 Å), Glu107 (1.76 Å), Asn82 (2.52 Å) and Asn105 (2.28 Å). The interaction with Tyr94 proved to be much stronger in the case of 4CSY complex, with the mean value of 1.98 Å. Residue Arg97 maintained a stable interaction with GalO3 (2.16 Å) and the carboxyl group of sialic acid (2.71 Å). sLe^x interactions with Tyr48 and Glu92 were also found to be more stable than in the case of 1G1T complex. These findings show that the E-selectin conformation from the X-ray structure 4CSY is able to form more stable interactions with the sLe^x tetrasaccharide.

Oligosaccharide position in the binding site did not undergo any major changes throughout the simulation. All dihedral angles remained very stable, indicating that the conformational switch in the protein did not affect the sLe^x conformation in the binding site. Comparison of average dihedral angles for the two X-ray structures studied here, along with the docked sLe^x-E-selectin complex is shown in Table 5. 3. 3.

Table 5. 3. 3. Comparison of average carbohydrate dihedral angles between the X-ray structures 1G1T, 4CSY and the docked sLe^x-E-selectin complex.

Glycosidic linkage	Dihedral angle	Docked complex	1G1T	4CSY
Galβ1-4GlcNAc	**Φ**	-99	-69	-82
	Ψ	-70	131	135
Fucα1-3GlcNAc	**Φ**	-66	-70	-72
	Ψ	94	-95	-93
Neu5Acβ2-3Gal	**Φ**	61	62	57
	Ψ	-115	-120	-112

*All measurements were made using the following definition of dihedral angles: ϕ = O5 – C1 – OX – C'X, ψ = C1 – OX – C'X – C'(X-1).

5. 3. 1. 3. Conclusion

In summary, the control study has shown that the computationally produced oligosaccharide can be docked to selectin X-ray structures using Autodock Vina and that the docking software correctly identified the binding site as the Ca^{2+}-coordination site in the majority of cases. The software also correctly predicted some of the residues involved in the binding of sLe^x. In the studied complex, Ca^{2+} ion was coordinated by the fucose residue and the binding was supported by interactions with residues Asn83, Tyr94, Arg97, Asn105 and Glu107 which is in agreement with the experimental results. Therefore, it should be possible to use the described *in silico* methods to produce and dock the fCS oligosaccharides to P- and L-selectins to predict the location of the binding site and to identify the protein residues that are most likely involved in the fCS binding. Nevertheless, docking of sLe^x to two different E-selectin structures produced complexes characterised by many possible sugar orientations. Multiple structures had a potential to be stable and, without the support of experimental data, it would be difficult to distinguish between the binding candidates and decide which orientation would be most likely to occur in nature.

In this study, MD simulations were used to validate the docking results and to look at the dynamics of the binding and the conformational changes within the protein and the ligand. The docked complex remained stable throughout the simulation, with no significant changes in either protein or oligosaccharide conformation. Average dihedral angles of the docked

sLe^x were compared to those from the X-ray structures. Large differences in the dihedral angles for glycosidic linkages Galβ1⟶4GlcNAc and Fucα1⟶3GlcNAc in the docked complex indicate that the proposed binding mode is energetically unfavourable when compared to the binding mode observed in the X-ray studies.

MD simulations undertaken in this study were also used to look at the dynamic behaviour of sLe^x-selectin complexes from the existing X-ray structures. It was established that the extended conformation from the structure 4CSY forms more favourable interaction with sLe^x than the bent conformation of structure 1G1T. This is likely due to the difference in position of loop 81-89 which in 4CSY forms a stabilising H-bond to fucose via residue Glu88. Further analysis showed that the interdomain hinge is characterised by a high degree of flexibility in this complex and a new type of EGF-like domain movement was discovered, leading to a 'hyper-extended' conformational state. In this twisting type of movement, the EGF-like domain rotated away by 50° from its position in the extended conformation, bringing the C-terminal tail to the opposite side of the protein and allowing the residues Thr136 – Asn138 to form contacts to the N-terminal region of the C-type lectin domain.

The conformation of sLe^x was in good agreement between the two X-ray structures and did not undergo any major changes throughout the MD simulations. The interdomain movement did not correlate to any changes in the ligand conformation or in any of the sLe^x-E-selectin interactions. Therefore, this study does not confirm the hypothesis that the conformational transition at the interdomain hinge leads to changes in selectin affinity for sLe^x.

5. 3. 2. Docking studies of fCS–selectin interactions

The *in silico* experiments described in this section aimed to investigate the ability of Autodock and MD simulations to predict the binding between fCS oligosaccharides and selectins and to shed light on the possible mechanism of binding between fCS trisaccharides and P-selectin. fCS trisaccharide variants with 2,4-sulfated and 3,4-sulfated fucose were prepared computationally and docked to the available P-selectin structures. The docking results were further evaluated with the use of MD simulations, analogous to those undertaken for the sLe^x tetrasaccharide-E-selectin complexes. The most significant findings of these studies are presented here.

5. 3. 2. 1. Preparation of the fCS oligosaccharides

fCS oligosaccharides were prepared in GLYCAM Carbohydrate Builder, as described for the sLe^x tetrasaccharide. 2,4S trisaccharide was built using the following formula: DGalpNAc[4S,6S]b1-4[LFucp[2,4S]a1-3]DGlcpAb1-OME. Minimisation produced a single file that was further minimised in Amber 14 under GLYCAM06 force field. Data from the MD simulation was analysed as before. The energy of the system remained very stable throughout the 40 ns simulation (sample energy plots are included in Appendix 4). Although RMSD values fluctuated slightly, they maintained the average value of 1.66 Å with respect to the lowest energy conformation (Fig. 5. 3. 11). Average dihedral angles were measured and were found consistent with the values recorded for the corresponding linkages in sLe^x and CS oligosaccharides (Table 5. 3. 4).

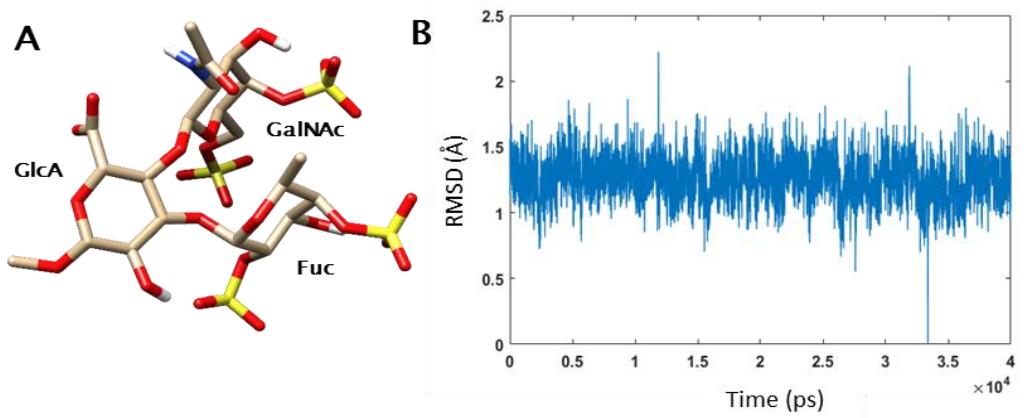


Fig. 5. 3. 11. MD simulation of the fCS 2,4S trisaccharide. A. Lowest energy conformation of the fCS 2,4S trisaccharide over a 40 ns simulation. **B.** Plot of overall RMSD for the fCS 2,4S-trisaccharide with respect to the lowest energy conformation.

Table 5. 3. 4. Average dihedral angles from the 40 ns MD simulation, compared to the values obtained for sLe^x tetrasaccharide and those reported in literature for CS.

Glycosidic linkage	Dihedral angle	2,4S trisaccharide	3,4S trisaccharide	sLe^x MD**	CS Lit.[174]
GalNAcβ1-4GlcA	**ϕ**	-66.6 ± 0.5	-66.2 ± 0.5	-68.4 ± 0.4	-80
	ψ	129.1 ± 0.3	130.4 ± 0.6	130.9 ± 0.3	110
Fucα1-3GlcA	**ϕ**	-69.3 ± 0.6	-71.9 ± 0.7	-69.0 ± 0.2	N/A
	ψ	-94.4 ± 0.2	-95.6 ± 0.3	-96.0 ± 0.3	N/A

*All measurements were made using the following definition of dihedral angles: ϕ = O5 – C1 – OX – C'X, ψ = C1 – OX – C'X – C'(X-1).
**Corresponding 1⟶4 and 1⟶3 linkages in sLe^x tetrasaccharide are compared here.

Oligosaccharide conformations were clustered over the 40 ns period and four most frequently occurring conformations were extracted for the docking studies. UCSF Chimera Cluster tool identified 398 clusters. The top five clusters were populated by a similar number of structures – from 52 structures for the first cluster to 48 for the fifth cluster. The dihedral angles between 20 representative conformations from each cluster are compared in Fig. 5. 3. 12. The representative conformations are highly similar to each other, with only slight differences in the values of dihedral angles. Therefore, it is evident that the UCSF Chimera clustering tool struggles to distinguish between the highly similar carbohydrate conformations as the differences in dihedral angles between the individual conformations are very small. Nevertheless, five representative conformations were chosen based on the values of dihedral angles closest to the average for a given cluster.

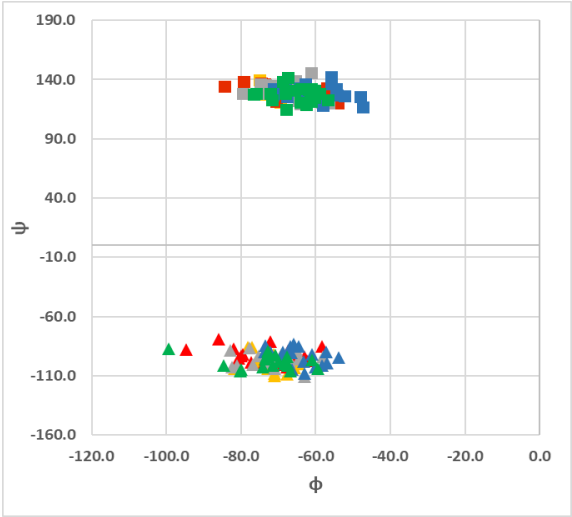


Fig. 5. 3. 12. Distribution of dihedral angles ϕ and ψ within the simulated fCS 2,4S trisaccharide. A. Plot of ψ versus ϕ for the glycosidic linkages GalNAcβ1→4GlcA (squares) and Fucα1→3GlcA (triangles). Five clusters are represented in yellow, red, grey, blue and green for Clusters I to V respectively. 20 representatives were chosen at random from each cluster as data points for this plot. All measurements were made using the following definition of dihedral angles: ϕ = O5 – C1 – OX – C'X, ψ = C1 – OX – C'X – C'(X-1).

5. 3. 2. 2. Docking of the fCS 2,4S trisaccharide to P-selectin

The computationally prepared fCS 2,4S trisaccharide was docked to the P-selectin structure 1G1Q. This structure has been widely studied in literature and was prepared using the same methods as the E-selectin structure 1G1T, described in Section 5. 3. 1. Therefore, direct

comparison could be made between the results of these docking tests and previous results on sLe^x-E-selectin interactions.

Four conformations of trisaccharide were isolated from MD simulation as described above and docked to the P-selectin structure in AutoDock Vina, using the same parameters as for the docking of sLe^x tetrasaccharide to E-selectin. The distribution of the docking results was similar to that observed for sLe^x and E-selectin, with majority of the results (83 %) binding at the Ca^{2+}-coordination site (Fig. 5. 3. 13). However, many different possible orientations of the fCS trisaccharide in the binding site were identified and not all of them bound to protein via Ca^{2+} ion – in some of the results, oligosaccharide formed interactions with amino acid residues on the other side of the binding surface (Fig. 5. 3. 13. B, D). As time and computational resources were limited, it was not possible to analyse all of the returned results in detail by MD and selection criteria had to be established to identify the most promising candidate complexes.

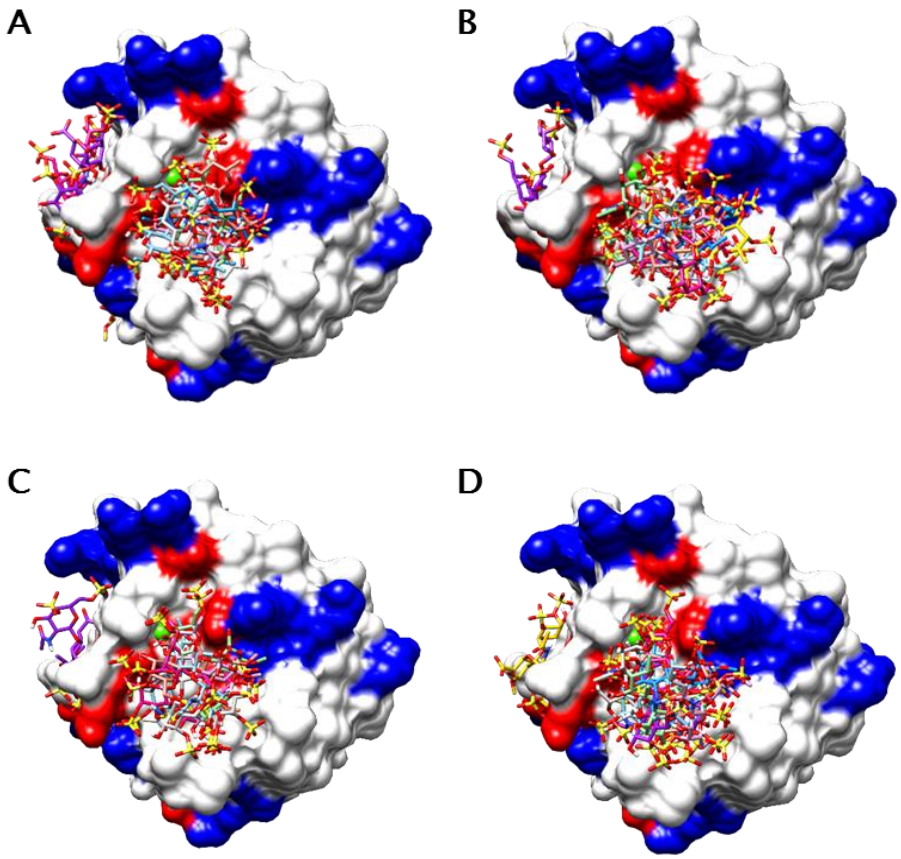


Fig. 5. 3.13. Docking results for four different conformations of fCS 2,4S trisaccharide. Four representative conformations of fCS 2,4S trisaccharide were docked to P-selectin structure 1G1Q. Nine structures returned by Autodock Vina are superimposed for each docking experiment in **A** – **D** (input conformation I – IV, respectively). The protein is shown in a surface representation, with positively charged residues highlighted in blue and negatively charged residues – in red.

The first criterion to be considered was the interaction with Ca^{2+} ion. Due to the fact that fCS binding has been shown to inhibit selectin binding to sLe^x, it is likely that it blocks selectin interaction with native ligands by competitive binding at the Ca^{2+}-coordination site[42]. Moreover, fCS trisaccharide structurally resembles the sLe^x tetrasaccharide, adopting similar 3D conformation which could allow it to mimic sLe^x and to bind to selectins in analogous manner. For this reason, the first step in selection process was to identify the docked complexes where the ligand interacts with the protein by coordinating the calcium ion. That allowed to eliminate close to 50% of the docking results.

The second selection criterion focused on the trisaccharide orientation within the binding site. Previous studies have shown that the unique biological activity of fCS relies on the presence of its fucose branches and the sulfate modifications[30,42,50]. Therefore, it is most likely that fucose will be directly involved in selectin binding and that sulfate groups will contribute to the binding by forming electrostatic interactions with the positively charged protein residues or by hydrogen bonding to polar amino acid side chains or protein backbone. Following on the hypothesis that the fCS trisaccharide unit binds to selectins in a manner similar to sLe^x determinant, it is likely that its interaction with the Ca^{2+}-coordination site would also occur via fucose residue. The STD NMR studies described in the previous chapter have confirmed that fucose is positioned close to the protein surface in fCS-selectin complexes, with the methyl group showing high response in the STD difference spectra. Based on this information, docking results with fucose located close to the protein surface and directly involved in the interaction with the protein were ranked above other results. Moreover, preference was given to the results where the Ca^{2+} ion was coordinated by fucose.

Use of the above selection criteria allowed to reduce the number of candidate binding structures from thirty-six to four. These four results were further evaluated by 15 ns MD simulations to assess the stability and dynamic behaviour of the produced complexes. Two of the structures dissociated during the simulation, leaving only two complexes that could be analysed in more detail. These complexes underwent further 15 ns simulation and their stability was compared using a visual assessment of the interactions supported by the free energy calculation.

5. 3. 2. 3. MD analysis of fCS 2,4S trisaccharide-P-selectin complexes

The first complex (complex I) was characterised by a large number of H-bonding interactions between the protein and the trisaccharide ligand. The initial set of interactions involved residues Tyr48 (2.00 Å), Asn82 (3.00 Å), Ser97 (2.98 Å), Ser99 (1.80 Å), Lys111 (1.77 Å) and Lys113 (1.88 Å) (Fig. 5. 3. 14). Calcium in this complex was coordinated by the Fuc $2SO_3$ group and this interaction remained stable throughout the 30 ns MD simulation, with the average distance of 2.67 Å. H-bonding interaction between Fuc $2SO_3$ group and residue Asn82 also remained stable, with the average distance of 3.24 Å. A weak interaction between Fuc $4SO_3$ and Arg85 developed within the first nanosecond and was maintained until the end of the simulation (3.67 Å). However, all the remaining interactions were broken within the first 15 ns and by the end of the simulation the complex was close to dissociation.

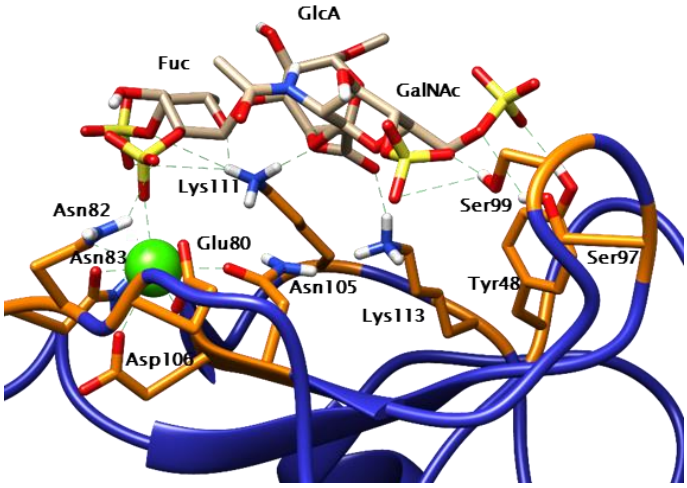


Fig. 5. 3. 14. Structural representation of complex I as predicted by AutoDock. Ribbon representation of P-selectin complexed to the fCS 2,4S trisaccharide. Complex I was one of the highly ranked results from the docking tests. However, the interactions in this complex weakened significantly during the 30 ns MD simulation. Protein is shown in blue, with the interacting residues highlighted in orange. Carbohydrate residues are shown in light grey.

The overall RMSD remained stable for the first 25 ns of the simulation, with a mean value of 1.97 Å with respect to the lowest energy conformation. In the last 5 ns, however, RMSD suddenly increased, maintaining the mean value of 2.97 Å for the rest of the simulation. When RMSD for protein and fCS was measured separately, the change in RMSD was found to be

more dramatic for the trisaccharide than for the protein – increasing from the mean value of 1.37 Å over the first 25 ns to 3.47 Å over the last 5 ns of the simulation (Fig. 5. 3. 15).

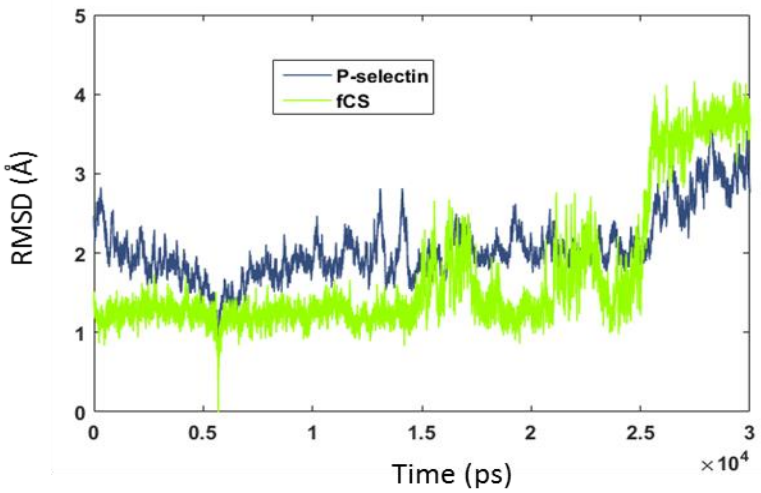


Fig. 5. 3. 15. RMSD plots from the 30 ns MD simulation of complex I. RMSD plots with respect to the lowest energy conformation for the fCS 2,4S trisaccharide (green) and P-selectin (blue) over a 30 ns MD simulation. A large change in the RMSD values can be observed in the final 5 ns of the simulation.

This change in RMSD values was correlated to the dramatic change in trisaccharide orientation in the binding site, leading to dissociation of the GalNAc and GlcA residues from the protein surface. After further analysis, it was linked to the change in dihedral angle ψ for glycosidic linkage Fucα1⟶3GlcA (Fig. 5. 3. 16).

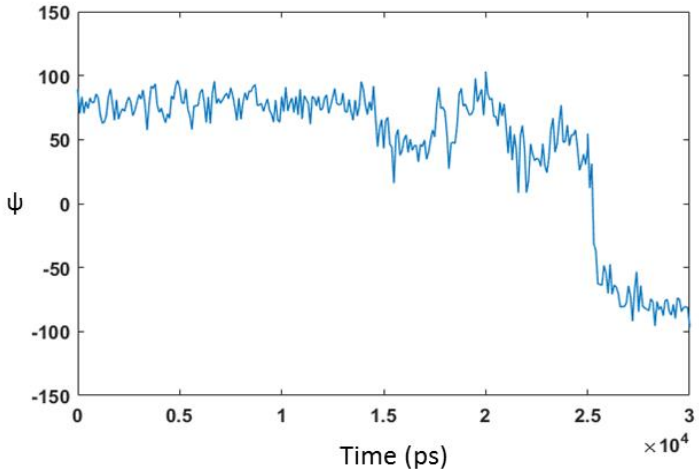


Fig. 5. 3. 16. Changes in dihedral angle ψ for the glycosidic linkage Fucα1-3GlcA. Plot from a 30 ns MD simulation of complex I, showing changes in dihedral angle ψ for the glycosidic linkage Fucα1-3GlcA. These changes correlate with the changes in RMSD observed for the fCS trisaccharide in Fig. 5. 3. 15.

Over the first 25 ns of the simulation, the value of ψ was very high (mean value of 69°) when compared to the value observed for the free fCS 2,4S trisaccharide (-94°). At 25 ns, a sudden change occurred as the value of ψ dropped to -94°, which is in agreement with the average value for the free trisaccharide. Such behaviour indicates that the native conformation of the fCS trisaccharide was significantly altered upon docking to P-selectin and that the return to its equilibrium conformation causes weakening of P-selectin-fCS interactions in this case. It was also noted that the average value of the dihedral angle ψ for the glycosidic linkage GalNAcβ1⟶4GlcA was largely different from the average angle measured for the free trisaccharide (-63° vs 129°) (see Appendix 6). Taken together, these results indicate that complex I is not likely to be a good model for the fCS-P-selectin interactions.

In the second fCS 2,4S trisaccharide-P-selectin complex tested by MD (complex II), Ca^{2+} ion was coordinated by the GalNAc $4SO_3$ group (mean distance of 2.67 Å) (Fig. 5. 3. 17). This interaction was further stabilised by H-bonds between GalNAc $4SO_3$ and Asn105 (2.19 Å), GalNAc O5 and Tyr94 (3.29 Å) and GalNAc CO and Lys111 (2.77 Å). In the complex returned by Autodock, fucose formed multiple H-bonds to Lys111, Lys113 and Ser46. However, only interaction between Fuc $2SO_3$ and Lys113 proved to be stable throughout the full duration of the simulation, with a mean distance of 3.07 Å. Glucuronic acid maintained interactions with Tyr48 (1.87 Å) and Ser97 (3.77 Å) via its carboxyl group.

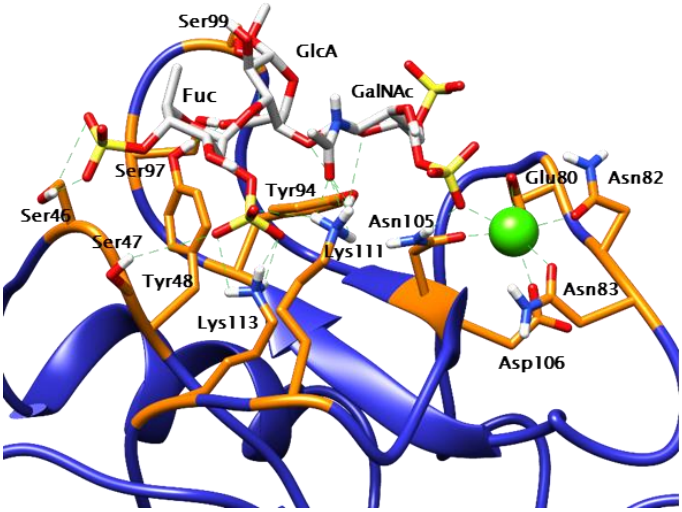


Fig. 5. 3. 17. Complex II between the fCS 2,4S trisaccharide and P-selectin. Ribbon representation of the second P-selectin-fCS trisaccharide complex, also returned by AutoDock Vina in the docking experiments. Protein is shown in blue, with the interacting residues highlighted in orange. Carbohydrate residues are shown in light grey.

Interestingly, a few of the protein residues involved in the fCS binding in this complex were directly involved in the interactions with fCS trisaccharide in complex I, despite the differences in the carbohydrate orientation. Such result suggests that, although AutoDock may struggle with correctly identifying the orientation of the trisaccharide in the binding site, it recognises a set of residues within the binding site that are most likely involved in the fCS trisaccharide binding. Indeed, residues Tyr48, Tyr94, Asn105 and Lys111 are the same residues that support sLe^x binding by E-selectin and Ser97 acts as a replacement for Arg97 which also mediates the E-selectin-sLe^x interaction (see Section 5. 3. 1).

Complex II remained stable throughout the 30 ns MD simulation, maintaining many of the interactions predicted by the docking software. The overall RMSD oscillated around the mean value of 1.86 Å with respect to the lowest energy structure. fCS had a lower average RMSD value (0.98 Å) than the protein (1.88 Å) and smaller changes in RMSD over time (Fig. 5. 3. 18). As the RMSD fluctuations for both the protein and sugar were relatively small, no major conformational changes occurred in either molecule during the 30 ns simulation.

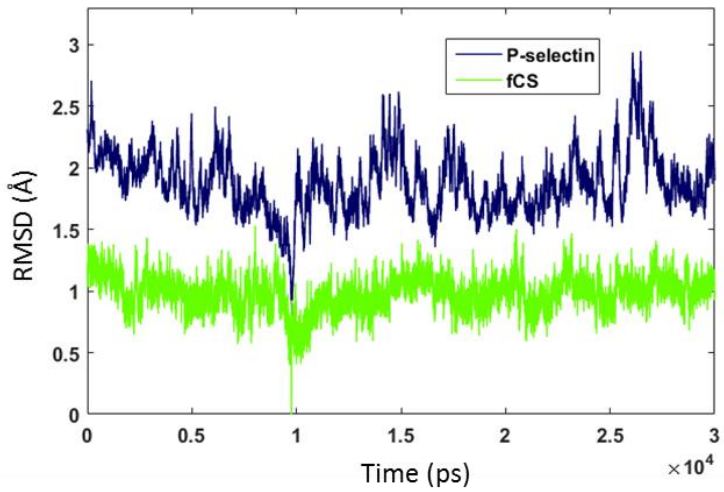


Fig. 5. 3. 18. RMSD plots from MD simulation of Complex II. RMSD plot for the fCS 2,4S trisaccharide (green) and P-selectin (blue) over a 30 ns MD simulation. Unlike in the case of Complex I, RMSD for both the protein and sugar remains stable throughout the full length of the simulation.

This finding was confirmed by the analysis of changes in dihedral angles over time. None of the angles underwent any major changes during the simulation, confirming that the fCS trisaccharide maintained a single stable conformation. However, their average values differed significantly from those recorded for the free trisaccharide during MD equilibration (see Appendix 6). In particular, dihedral angle ψ for the glycosidic linkage Fucα1⟶3GlcA had

a value of 58°, compared to -94° for the unbound trisaccharide and ψ for the linkage GalNAcβ1⟶4GlcA had a value of -72°, compared to 129° for the unbound trisaccharide. This large variation in the values of dihedral angles indicates that the proposed mode of binding would involve substantial reorientation of the equilibrium conformation of the fCS trisaccharide and thus it may not be favourable *in vivo*.

In agreement with the STD NMR experiments, GalNAc residue played an important role in the binding, forming multiple interactions to amino acid residues and with the acetyl group positioned close to the P-selectin surface. Fucose also remained close to the protein surface throughout the simulation. However, fucose methyl group was pointing away from the protein which contradicts the results obtained in the STD NMR experiments.

Free binding energy was also calculated for both complexes. Complex II was found to have ΔG_{bind} of a value of -0.2 ± 0.4 kcal mol^{-1}, lower than the value of 10.4 ± 0.3 kcal mol^{-1} calculated for complex I. Both values are relatively high which may indicate weak binding for the two proposed complexes, with a thermodynamic preference towards complex II. Nevertheless, it is important to note again that the MM-PBSA free energy calculation is not an accurate measure of free binding energy for all systems and can only be taken as a rough estimate and not as a reliable measure of the stability of each complex.

Taken together, the above results suggest that the obtained complexes must be treated with caution. It is possible that the amino acid residues implicated in the binding by Autodock Vina do contribute to fCS 2,4S trisaccharide binding by P-selectin *in vivo*. However, the orientation and conformation of the 2,4S trisaccharide are likely different from those proposed by the docking studies, based on the structural and energetic analysis of the MD data.

For comparison, two of the trisaccharide conformers were docked to P-selectin structure from the PDB entry 1G1S. This structure was produced using the same methods as for the entry 1G1Q, however, in 1G1S P-selectin was crystallised when complexed with PSGL-1 (see Section 1. 3. 3.). The important difference between the two P-selectin structures is that in the 1G1Q entry the protein is present in bent conformation, whilst in the 1G1S entry it exists in the extended conformation. As selectin conformation significantly affected the docking results in the study on sLe^x-E-selectin binding, it was thought that a similar effect could be observed in the docking of fCS oligosaccharides to P-selectin.

PSGL-1 was removed from the PDB structure and the fCS trisaccharides were docked as previously described. This time, trisaccharide bound to P-selectin via the Ca^{2+}-coordination site in a larger proportion of complexes. After applying the selection criteria to the docking

results, five of the eighteen complexes were identified as suitable for further analysis. After a 15-ns MD simulation, two of the complexes dissociated, leaving only three binding options. One complex did not dissociate but, after 15 ns, fCS trisaccharide was no longer bound via Ca^{2+} ion. The remaining two complexes were simulated for another 15 ns, however, the binding weakened significantly throughout the simulation and, in both cases, fucose residue moved away from the protein and did not participate in the binding for most of the simulation.

These results suggest that, although the initial docking results for P-selectin structure 1G1S are more consistent and structurally more promising than upon docking to 1G1Q, the returned complexes are not stable enough to withstand MD simulation longer than 15 ns and do not provide much insight into the fCS-P-selectin binding. Therefore, unlike in the case of sLe^x binding to E-selectin, the extended conformation of the protein did not improve the overall outcome of the docking experiments. It is possible that fCS trisaccharide binds to P-selectin in a different manner to sLe^x and that the selection criteria may need to be changed to obtain a more stable complex. Moreover, alternative starting conformations of the trisaccharide or the protein may need to be explored to improve the docking results. However, due to the time restrictions, it was not possible to investigate such further options in this project.

5. 3. 2. 4. Docking of the fCS 3,4S trisaccharide to P-selectin

After the docking tests on 2,4S trisaccharide, a similar study was performed where 3,4S fCS trisaccharide was docked to P-selectin structure 1G1Q. Two representative conformations from the UCSF Chimera cluster analysis were used in the docking experiments. In 50 % of the results fCS bound to protein at the Ca^{2+}-coordination site, which is a smaller proportion than in the case of 2,4S trisaccharide. In 50 % of the results fCS bound in alternative regions of the protein, either within the C-type lectin domain or at the interdomain hinge (Fig. 5 3. 19). These alternative binding regions were not considered in this study – only the complexes with fCS bound at the Ca^{2+}-coordination site were considered as suitable binding candidates.

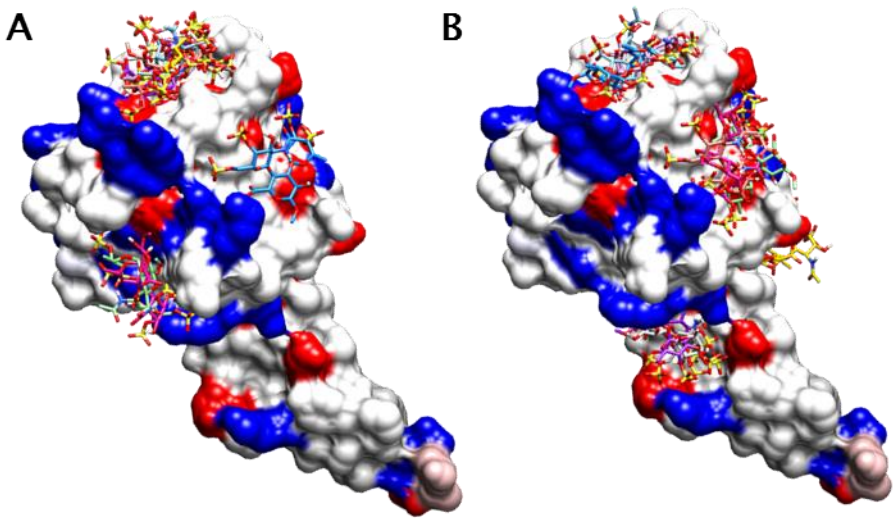


Fig. 5. 3.19. Docking results for two different conformations of fCS 3,4S trisaccharide. Two representative conformations of fCS 3,4S trisaccharide were docked to P-selectin structure 1G1Q. Nine structures returned by Autodock Vina are superimposed for each docking experiment for conformation I (**A**) and II (**B**). The protein is shown in a surface representation, with positively charged residues highlighted in blue and negatively charged residues – in red.

Only in two of the complexes fucose residue was directly involved in the binding and it is these two results that were further tested by MD simulations. Both complexes remained stable throughout the initial set of tests (25 ns) and interacted with the protein via a similar set of amino acid residues, with fucose playing a key role in the binding. As the interactions in these complexes after the 25 ns simulation were promising, it was decided that one of the complexes would be studied by a longer (100 ns) simulation. Due to the computational limitations, only the complex that was calculated to have a slightly lower value of ΔG_{bind} (11.9 ± 0.3 vs. 13.3 ± 0.4 kcal mol^{-1}) was further studied by MD. This complex will be referred to as complex III.

Although the energy remained stable throughout the full duration of the simulation, overall RMSD for complex III showed a large drop from the mean value of 4.91 Å to 2.15 Å just after 65 ns into the simulation. When the RMSD for P-selectin and the fCS was measured separately, it was established that the main contribution to this change comes from the conformational changes within the protein (Fig. 5. 3. 20). RMSD plot for the fCS shows a smaller drop from the mean value of 2.87 Å to 1.07 Å at 30 ns into the simulation.

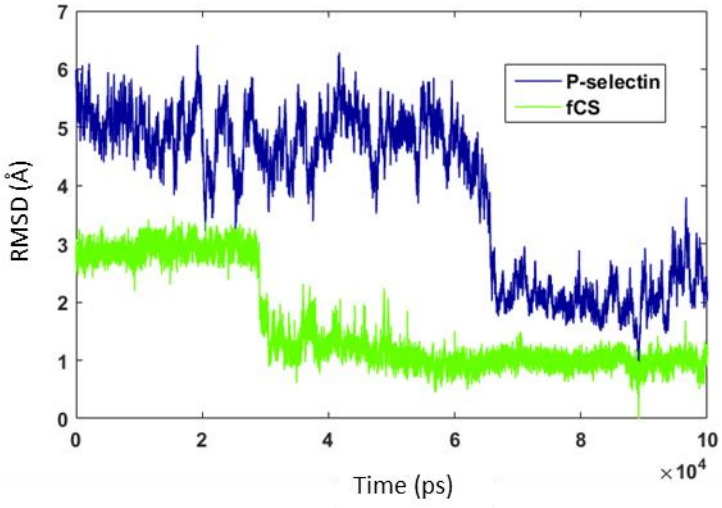


Fig. 5. 3. 20. RMSD plots from MD simulation of Complex III. RMSD plot for the fCS 3,4S trisaccharide (green) and P-selectin (blue) with respect to the lowest energy structure over a 100 ns MD simulation. Conformational changes within the sugar and the protein occur at 30 ns and 65 ns into the simulation, respectively.

The sudden drop in the protein RMSD was correlated to the interdomain movement in the hinge region, similar to that observed for E-selectin in the earlier study on sLe^x-1G1T complex (Section 5. 3. 1. 3.). The interdomain angle changed dramatically, from 109° at 60 ns to 135° at 65 ns, corresponding to conformational transition from the bent to extended state. This observation indicates that it is possible for P-selectin to undergo conformational change between these two states when bound to the 3,4S fCS trisaccharide. It can also be seen that the interdomain movement in P-selectin did not cause any rearrangement in the carbohydrate structure.

The sudden change in the RMSD for the trisaccharide at 30 ns indicates that the fCS also experienced a conformational change during the simulation. Indeed, when the carbohydrate dihedral angles were plotted over time, a large drop in dihedral angle ϕ for the linkage GalNAcβ1⟶4GlcA was observed at 30 ns (Fig. 5. 3. 21). The mean value for this angle over the first 30 ns was 19.4°. After the conformational change has occurred, the mean value dropped to -70° and remained stable for the rest of the simulation.

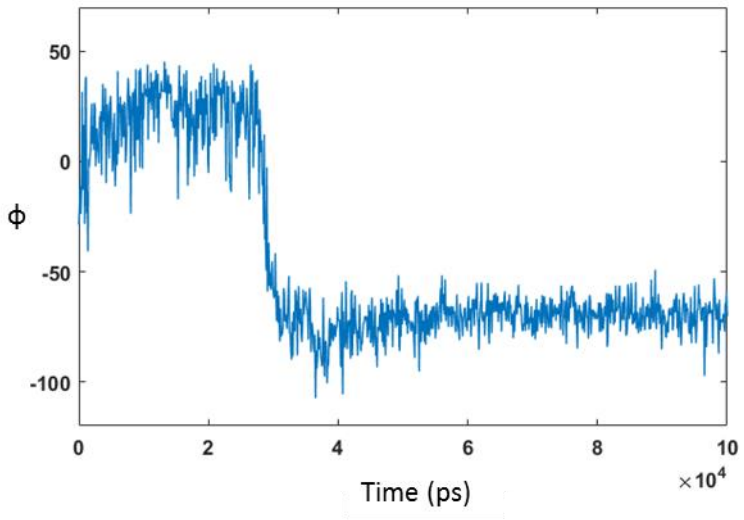


Fig. 5. 3. 21. Changes in dihedral angle ϕ for the glycosidic linkage GalNAcβ1-4GlcA. Plot from a 100 ns MD simulation of complex III, showing changes in dihedral angle ϕ for the glycosidic linkage GalNAcβ1-4GlcA. These changes correlate with the changes in RMSD observed for the fCS trisaccharide in Fig. 5. 3. 20.

Interestingly, after this change, the average dihedral angles measured for the fCS 3,4S trisaccharide in complex with P-selectin were close to those recorded for the unbound 3,4S trisaccharide. This is in contrast to the analysed fCS 2,4S trisaccharide-P-selectin complexes where the dihedral angles for the bound carbohydrate differed significantly from those measured for the free trisaccharide. This observation suggests that complex III may be a better model of the interactions occurring in nature, as it would not require such a large rearrangement of the equilibrium structure of the fCS trisaccharide.

The interactions in complex III involved similar amino acid residues as in the case of fCS 2,4S trisaccharide-P-selectin complexes. Many of these interactions were maintained by the fucose residue which also coordinated the Ca^{2+} ion with its sulfate groups. In the complex returned by Autodock Vina this coordination occurred via Fuc $4SO_3$ group. However, within the first nanosecond of the MD simulation this group was replaced by the Fuc $3SO_3$ group which coordinated the Ca^{2+} ion for the rest of the simulation with the mean distance of 2.68 Å. Other interactions identified by the docking software involved residues Tyr48, Asn82, Glu92, Tyr94, Ser97, Ser99, Asn105, Lys111 and Lys113 (Fig. 5. 3. 22. A). However, the majority of these interactions were broken 30 ns into the simulation, when the trisaccharide underwent the conformational change. From this point onwards, GalNAc and GlcA were positioned too far from the protein surface to maintain any stable interactions with the

protein residues. Only fucose remained firmly attached to the protein, anchoring the sugar to the Ca^{2+}-coordination site. Fuc $3SO_3$ maintained its interaction with Asn82 (mean distance of 3.50 Å), Fuc 2OH formed a stable H-bonding interaction with residue Glu80 (1.78 Å), Fuc O4 – with Tyr94 (3.28 Å) and Asn105 (3.06 Å), and Fuc $4SO_3$ maintained electrostatic interactions with Lys111 (3.84 Å and 3.88 Å) (Fig. 5. 3. 22. B).

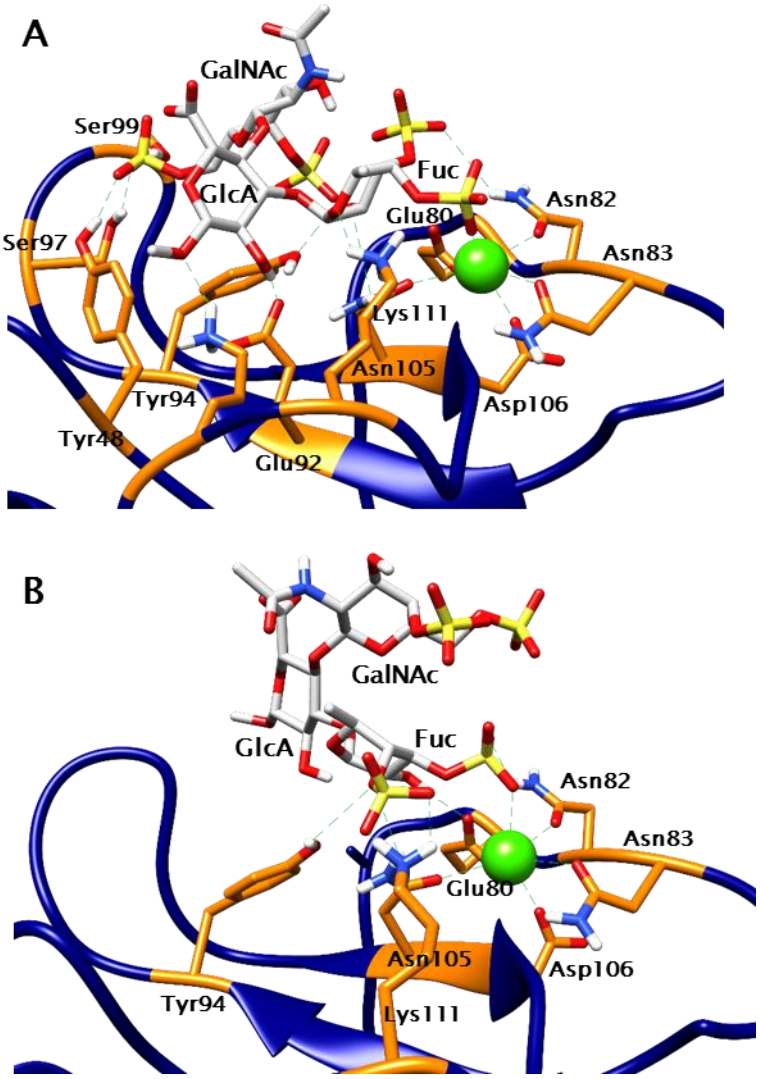


Fig. 5. 3. 22. Initial (A) and final (B) set of interactions for Complex III. Ribbon representation of P-selectin 1G1Q structure bound to fCS 3,4S trisaccharide at the start of the MD simulation (**A**) and after 100 ns of the simulation (**B**). Protein is shown in blue, with the interacting residues highlighted in orange. Carbohydrate residues are shown in light grey.

As can be seen from the comparison of the initial and final interactions, the binding weakened significantly after the conformational change occurred in the trisaccharide. However, fCS

remained bound to the P-selectin for the full 100 ns which indicates that the interactions between fucose and P-selectin are strong in this complex. Additional MD simulation was performed, using different starting coordinates to further investigate the stability of this complex. Again, the complex remained stable for 100 ns and the fucose residue played a crucial role in the interaction. Interestingly, identical conformational change occurred in the carbohydrate at 80 ns of the second MD simulation, again leading to partial dissociation of the fCS from the protein surface. Therefore, although the initial set of interactions seems stable, withstanding 30 ns and 80 ns in two separate MD simulations, these interactions are not strong enough to balance out the thermodynamic drive for the equilibrium conformation of the fCS trisaccharide.

The fact that the fucose residue played an important role in the P-selectin binding and that both Fuc CH_3 group and Fuc H1 remained in close contact with the protein surface is in agreement with the results from STD NMR experiments (Section 4. 3. 4.). However, it is important to note that there were differences between the docked complex and the experimental results – GalNAc residue has dissociated from the protein after the first 30 ns and its CH_3 group did not maintain contacts with the P-selectin surface. GlcA residue also moved away from the protein at 30 ns which means that fucose was the only carbohydrate residue involved in the binding for the majority of the simulation. This is unlikely to be the case in nature and the STD experiments clearly showed that both GalNAc and GlcA residues maintain contacts with P-selectin surface. Therefore, the complex presented here can only partially reflect the interactions occurring *in vivo*. It is possible that the true set of interactions is similar to the one observed in the beginning of the MD simulation. However, the conformation of the fCS trisaccharide would need to be different to the one proposed by Autodock in order to maintain these interactions for a longer time.

5. 3. 2. 5. Docking of the fCS trisaccharide – conclusions

In summary, in the above *in silico* studies on the fCS trisaccharide-P-selectin interactions, Autodock Vina was used to explore the potential binding sites and conformations for two types of fCS trisaccharides – with fucose either 2,4 or 3,4-sulfated. These trisaccharides were docked to P-selectin structure 1G1Q and the resulting complexes were analysed by MD simulations.

The docking studies showed that Autodock Vina predicts that the fCS trisaccharide binds to the Ca^{2+}-coordination site in 83 % of cases for the 2,4S trisaccharide and in 50 % of cases for the 3,4S trisaccharide. However, only in a small proportion of all results, the binding is consistent with the STD NMR experimental results and the current knowledge on selectin-ligand binding. In this study, only the docked complexes that fulfilled the selection criteria underwent further analysis. It is possible that in future studies, the alternative binding options may also need to be considered.

The complete analysis by MD was performed on three complexes: two for 2,4S trisaccharide (complexes I and II) and one for 3,4S trisaccharide (complex III). Complexes I and II underwent a 30 ns MD simulation during which complex II proved to be more stable, maintaining the fCS-P-selectin interactions. However, the carbohydrate dihedral angles in this complex were largely different from those reported for the unbound 2,4S trisaccharide. Complex III proved to be very stable during the initial 25 ns simulation and was further studied by 100 ns MD simulation. During the longer MD simulation, the fCS trisaccharide underwent a conformational change as the value of dihedral angle ϕ for glycosidic linkage GalNAcβ1⟶4GlcA dropped to the equilibrium value for the unbound fCS. Following this conformational change, the GalNAc and GlcA residues dissociated from the protein and the interactions were only maintained by the Fuc residue. Although the interactions with fucose were very stable, the proposed conformation of fCS trisaccharide in this complex is not energetically favourable.

Structural and energetic analysis of these three complexes led to a conclusion that Autodock Vina has a difficulty in predicting a stable carbohydrate conformation during the docking. Therefore, the proposed complexes provide only partial information on the true nature of fCS-selectin binding and the true trisaccharide conformation upon binding is likely to be different to that proposed in the above docking studies. However, the docking results were consistent in their prediction of the amino acid residues that participate in the fCS binding. The residues that were repetitively found to form H-bonding interactions with the fCS trisaccharide were: Tyr48, Glu80, Asn82, Asn83, Tyr94, Ser97, Asn105 and Lys111. Comparison to existing X-ray structures for selectin-sLe^x complexes show that some of these residues participate in selectin interaction with sLe^x. It can therefore be said with confidence that a subset of these residues is involved in the fCS trisaccharide-P-selectin interactions *in vivo*. Further experimental data would be needed to develop better computational models for these interactions and to establish the exact mode of binding between the fCS and P-selectin.

Overall, the *in silico* studies of the interactions between the fCS oligosaccharides and P-selectin provided useful information on the nature of the binding and conformational behaviour of both the protein and the fCS ligands. However, it is clear that Autodock Vina struggles to predict the conformation of fCS trisaccharides in the binding site. It is not surprising, as this docking software is optimised for docking of small ligands, with minimal number of rotatable bonds. If the docking experiments were to be continued in future, it would be necessary to explore the use of a different docking software, more suited for performing calculations on carbohydrate ligands or to alter the docking parameters, for example by reducing the number of rotatable bonds. Finally, the MD studies would provide the most useful information if complementary experimental data was obtained to allow for the use of known experimental restrains in modelling and data analysis.

Chapter 6: Development of GFT NMR experiments for carbohydrate studies

6.1. Introduction

NMR is one of the greatest and most commonly used techniques for structural studies on small, as well as large molecules and is commonly used for elucidation of carbohydrate structure and conformation. Many literature reviews on carbohydrate studies by NMR have been published and it is not within the scope of this thesis to comment on all available techniques. Typically, a combination of 1D and 2D NMR experiments is used to assign the chemical shifts to individual atoms within each carbohydrate ring and to establish the connectivity between the rings[175-178]. For example, a combination of 2D ^{1}H, ^{13}C HSQC, 2D HSQC-TOCSY and HMBC experiments can be used to determine the connectivity between atoms[177,179,180]. A good starting point for such assignment is the anomeric region of the spectra which is well separated from other signals. This region allows to estimate the number of carbohydrate species in the sample and often allows to find correlations with other signals within each carbohydrate ring. Further connections can be found by 'walking' along the carbohydrate chain, following the correlations between individual protons and carbons both within one ring and between two connected rings (HMBC)[177,180]. Other methods, such as NOE experiments, can provide information about the inter-residual linkages and spatial conformation of the oligosaccharides[175,181,182].

Despite the large array of available mono- and multidimensional NMR experiments, determination of carbohydrate structure for oligosaccharides still poses many challenges. Carbohydrate subunits are structurally highly similar to one another. The differences between the individual oligosaccharide chains can be very small, based only on stereochemistry, different pattern of modifications along the carbohydrate rings or connectivity between the individual residues. For this reason, chemical shifts of oligosaccharides are often overlapped and structural elucidation by NMR can become difficult. Moreover, most oligosaccharide samples extracted from natural sources are heterogeneous, which means that, typically, the samples contain mixtures of

oligosaccharides with overlapping signals, adding another level of complexity to the NMR spectra. In many cases, traditional 2D NMR methods are not sufficient to allow to assign the signals and determine the structure of oligosaccharide chains with confidence.

In the ^{1}H dimension, most carbohydrate signals are clustered between ~3.4 – 4.0 ppm. In order to correctly assign the networks of coupled spins using heteronuclear 2D experiments, it is necessary that the signals are well separated in the ^{13}C dimension. This can be problematic in the complex carbohydrate samples, where many species of similar structure are present, such as the depolymerised fCS samples. In such samples, small variations in ^{1}H and ^{13}C chemical shifts cause clusters of cross peaks to appear on the 2D spectra with high overlap in both dimensions. In some cases, such overlap could be resolved by the use of a 3D HSQC-TOCSY experiment[183]. However, acquisition of such an experiment requires long experimental time and the resulting spectra are characterised by reduced resolution. Therefore, 3D experiments offer limited help and are often not practical. A more suitable solution might be to use a modified 2D experiment where the signals are separated from each other in the ^{13}C dimension by a factor originating from the difference in ^{1}H chemical shifts. Such approach is used in the G-matrix Fourier Transform (GFT) NMR spectroscopy – a technique which allows rapid sampling of multidimensional data[184]. In GFT NMR experiments, two or more frequencies are sampled simultaneously by co-incrementing corresponding delays in the radio-frequency pulse sequence – for example, in HSQC-TOCSY experiment mentioned above, ^{1}H and ^{13}C chemical shifts are sampled at the same time. Such experiment is referred to as a (3,2)D HSQC-TOCSY, as it allows to acquire high-resolution 3D spectral data in a 2D experiment, allowing to reduce the dimensionality and experimental time. Data which would normally form a 3rd dimension in a 3D NMR experiment is now encoded within the 2D spectrum, causing splitting of each peak into two peaks. The two resulting peaks are separated from each other by a factor of the corresponding chemical shift from the 3rd dimension (Ω_3) (Fig. 6. 1. 1)[184,185].

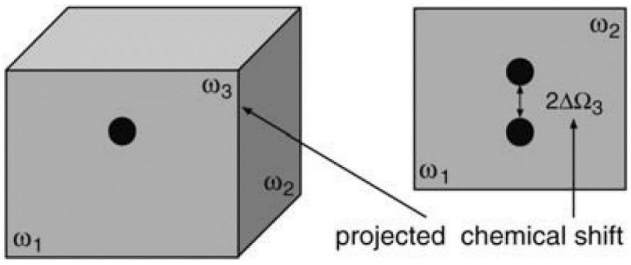


Fig. 6. 1. 1. Schematic representation of the main principle of GFT experiments.
GFT experiments allow to reduce the dimensionality of NMR experiments. Data which would otherwise be acquired as a separate dimension in GFT is encoded within the 2D NMR spectrum, allowing for separation of some of the overlapped signals. Figure taken from Szyperski (2006)[185].

(3,2)D HSQC-TOCSY was previously implemented in the metabolomics studies by Pudakalakatti *et al.*[186]. In this study, two (3,2)D HSQC-TOCSY spectra were produced – in one of them the frequencies in the indirectly detected dimension were offset by $+\Omega_{1H}$, whilst in the second spectrum – by - Ω_{1H} (Fig. 6. 1. 2). Using this approach, it was possible to separate some of the signals, otherwise overlapped in ^{13}C dimension.

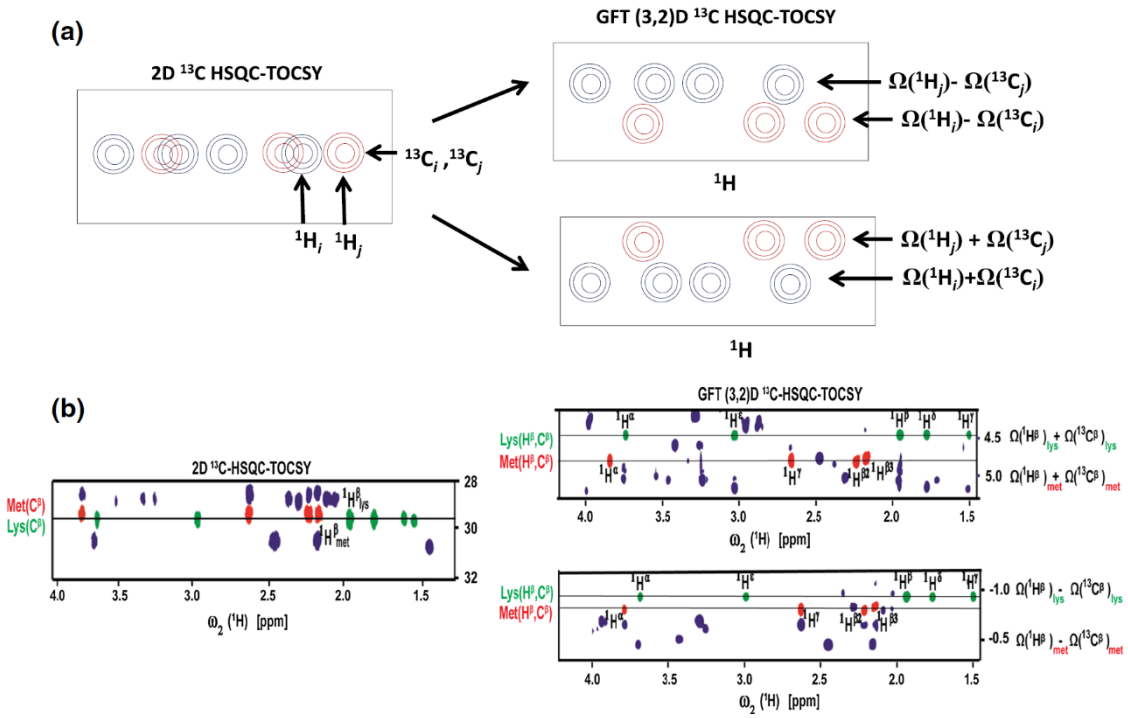


Fig. 6. 1. 2. An illustration of the overlap in 2D HSQC-TOCSY spectrum, resolved by a GFT (3,2)D HSQC-TOCSY experiment. A Schematic representation illustrating the principle of the (3,2)D HSQC-TOCSY experiment. The overlapping signals with ^{13}C frequencies i ad j can be separated by linear combinations of corresponding ^{1}H and ^{13}C. **B** Spectral overlap of two spin systems (Lys and Met) in a conventional HSQC-TOCSY experiment on a sample of 21 metabolites could be resolved by the use of GFT (3,2)D HSQC-TOCSY. Figure taken from Pudakalatti (2014)[186].

In the study undertaken here, a (3,2)D ^{1}H, ^{13}C HSQC-TOCSY experiment was implemented in which a modulation of cross peaks in the indirectly detected dimension (F_1) due to proton-proton couplings was removed. Such modulation is inherent to a regular (3,2)D ^{1}H, ^{13}C HSQC-TOCSY experiment and its removal produces high-resolution singlets in the F_1 dimension, which makes it a suitable tool for analysis of complex samples with clusters of cross peaks. A simple manipulation of the original spectra allows to produce two spectra with signals at either the $\Omega_{13C}+\Omega_{1H}$ or $\Omega_{13C}-\Omega_{1H}$ frequency. This method was tested on a sample of fCS oligosaccharides from *H. forskali* body wall, produced by β-eliminative depolymerisation.

6.2. Material and methods

6. 2. 1. fCS sample preparation

The fCS polysaccharide was extracted from the body wall of a sea cucumber *H. forskali* at Glycomar Ltd. as previously described[55]. The sample was depolymerised using a modified protocol for β-eliminative depolymerisation of carbohydrate in anhydrous solutions[187].

6. 2. 2. NMR spectroscopy

20 mg sample of the produced oligosaccharide mixture was dissolved in 550 μL of D_2O and measured at 300 K. The (3, 2)D ^{1}H, ^{13}C HSQC and (3, 2)D ^{1}H, ^{13}C HSQC-TOCSY spectra of the fCS mixture were acquired on a 4-channel Avance III 800 MHz Bruker spectrometer equipped with a 5 mm TCI CryoProbe™ with automated matching and tuning. The following parameters were used: 2048 and 2048 complex points in t_2 and t_1, respectively, spectral widths of 8 and 130 ppm in F_2 and F_1, yielding t_2 and t_1 acquisition times of 160 and 39.17 ms, respectively. Four scans were acquired for each t_1 increment using a relaxation time of 1.4 s. The overall acquisition time was 7 hours and 47 min (for two interleaved spectra, using 4096 t_1 points). A 40 ms mixing time for the TOCSY transfer used DIPSI-2 pulse sequence[188]. A forward linear prediction to 4096 points was applied in F_1. A zero filling to 4096 was applied in F_2. A cosine square window function was used for apodization prior to Fourier transformation in both dimensions. Identical parameters were used to acquire regular 2D ^{1}H, ^{13}C HSQC and 2D ^{1}H, ^{13}C HSQC-TOCSY spectra (Bruker pulse programs *hsqcedetgpsisp2.3* and *hsqcdietgpsisp.2* in half of the time required for the GFT experiments.

6. 2. 3. Pulse sequence parameters

The following parameters are associated with pulse sequences shown in Fig. 6. 2. 1. Narrow and wide filled rectangles represent 90° and 180° pulses, respectively. Open rectangles with inclined arrows were 500 μs ^{13}C CHIRP pulses, while filled rectangles with and inclined arrows were 2 ms composite ^{13}C CHIRP pulses. The delays used were, $\Delta_1 = 0.5/^1J_{CH} - p14/2$; τ_g =1.2 ms (1 ms gradient and 200 μs recovery), $\Delta_2 = 0.25/^1J_{CH} - \tau_g - p14/2$; $\Delta_3 = 0.25/^1J_{CH} + \tau_g - p14/2 + k*t_1(0)$; $\tau = \tau_g + p_{180}*2*t_1(0)$; $\Delta_4 = 0.25/^1J_{CH} + p14/4$; $\Delta_5 = 0.25/^1J_{CH} - p14/2$; $\Delta_6 = \tau_g - 0.78*p_{90} - DE$, where p14 = 500 μs, p_{90} and p_{180} are 90° and 180° ^{1}H pulses, k is the scaling factor for Ω_{1H} frequencies, $t_1(0)$ is the initial t_1 increment, typically 3 μs and DE is a pre scan delay (10 μs). Unless specified otherwise, the pulses were applied from the x axis; φ_1 = x or y for cosine and sine modulated signals, respectively; φ_2 = 2x, 2(-x); φ_3 = x; φ_4 = x, -x; φ_5 = 2y, 2(-y); Ψ = x, 2(-x), x. Phases φ_3 , φ_4 and Ψ were increased in a TPPI manner by 180° simultaneously with t_1 incrementation; the real and imaginary points were acquired by changing the polarity of the G_1 gradient and increasing by 180° the phase φ_5. Gradient G_3 was 600 μs long, while all other gradients were applied for 1 ms. The following relative strength was used: G_1 = 80%, G_2 = 20.1%, G_3 = 11%, G_4 = -5% and G_5 = 7%. Bruker pulse sequences for the ^{1}H, ^{13}C (3,2)D BIRD^{r,X}-HSQC and ^{1}H, ^{13}C (3,2)D BIRD^{r,X}-HSQC-TOCSY experiments can be found in Appendices 7 and 8, respectively.

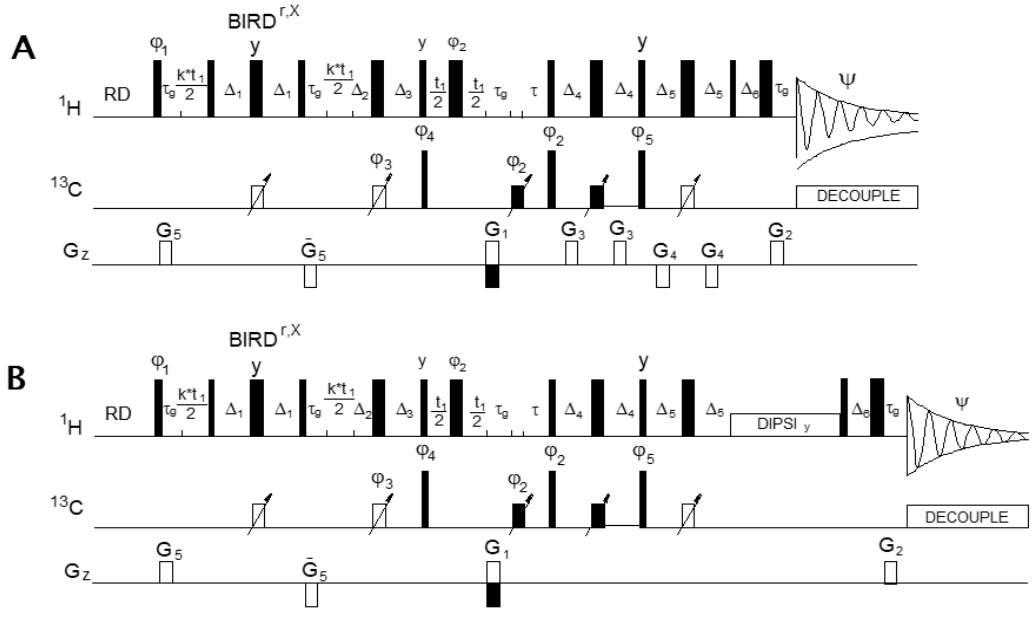


Fig. 6. 2. 1. Pulse sequences of the two GFT NMR experiments.
A ^{1}H, ^{13}C (3,2)D BIRD^{r,X}-HSQC. B ^{1}H, ^{13}C (3,2)D BIRD^{r,X}-HSQC-TOCSY. Parameters are described in section 6. 2. 3. Enlarged version of this figure is included as Appendix 9.

6.3. Results and discussion

6. 3. 1. Pulse sequence

Experiments developed in this study were based on a Bruker implementation of sensitivity-enhanced 2D ^{1}H-^{13}C HSQC[189,190] and 2D ^{1}H-^{13}C HSQC-TOCSY[189-191] experiments and will be referred to as (3,2)D BIRD^{r,X}-HSQC and (3,2)D BIRD^{r,X}-HSQC-TOCSY. In each of the two experiments, ^{1}H chemical shift is labelled during a variable t_1 period. The evolution of the proton-proton and proton-carbon couplings is suppressed by a central BIRD^{r,X} pulse[192,193] (Fig. 6. 2. 2.). This pulse inverts only the magnetisation of protons attached to ^{12}C, thus allowing to label the chemical shift Ω_{1H} of protons attached to ^{13}C.

Depending on the phase of the first 90^{o} ^{1}H pulse, the ^{13}C chemical shifts labelled during the t_1 interval after the transfer of polarisation to carbons, are modulated either by $\cos(2\pi\Omega_{1H}kt_1)$ or $\sin(2\pi\Omega_{1H}kt_1)$ frequencies. Ω_{1H} is the difference between the ^{1}H chemical shift and the ^{1}H carrier frequency and k is the scaling factor between the two simultaneously incremented indirectly detected periods. As a result, cross peaks in (3,2)D BIRD^{r,X}-HSQC-TOCSY spectra appear as in-phase or antiphase doublets centred around ^{13}C chemical shifts of the directly bonded ^{1}H-^{13}C pairs and separated by $2k\Omega_{1H}$ in F_1. This generates $k\Omega_{1H}$ dependent displacement of signals. The cosine and sine modulated spectra are acquired in an interleaved manner and processed to produce two simplified spectra by the addition or subtraction of the original spectra. The simplified spectra thus contain only one part of the F_1 doublets, each as positive signals with increased intensity. They are referred to here as the Ω_{13C} + Ω_{1H} and the Ω_{13C} - Ω_{1H} spectrum. The overall theoretical sensitivity of the (3,2)D BIRD^{r,X}-HSQC-TOCSY experiment is half of that of a regular 2D HSQC-TOCSY experiment. In practise, additional signal-to-noise reduction is observed, which can be attributed to the $^{1}J_{CH}$ mismatch and relaxation effects during the BIRD^{r,X} pulse. Overall, signal-to-noise reduction of 30-45% was measured for the corresponding cross peaks in the (3,2)D BIRD^{r,X}-HSQC spectra of the fCS mixture relative to traditional 2D HSQC spectra.

6. 3. 2. (3,2)D BIRD^{r,X}–HSQC

A partial 2D BIRD^{r,X}-HSQC spectrum for the studied fCS sample is presented in Fig. 6. 3. 1. As can be seen from this figure, 2D HSQC spectra of fCS oligosaccharide mixtures are very complex, with many groups of overlapping peaks. For this reason, 2D HSQC-TOCSY experiments are normally used to allow to find correlations between individual cross peaks. However, many of the signals have the same ^{13}C chemical shift, which hinders the analysis of traditional 2D HSQC-TOCSY spectra in such samples.

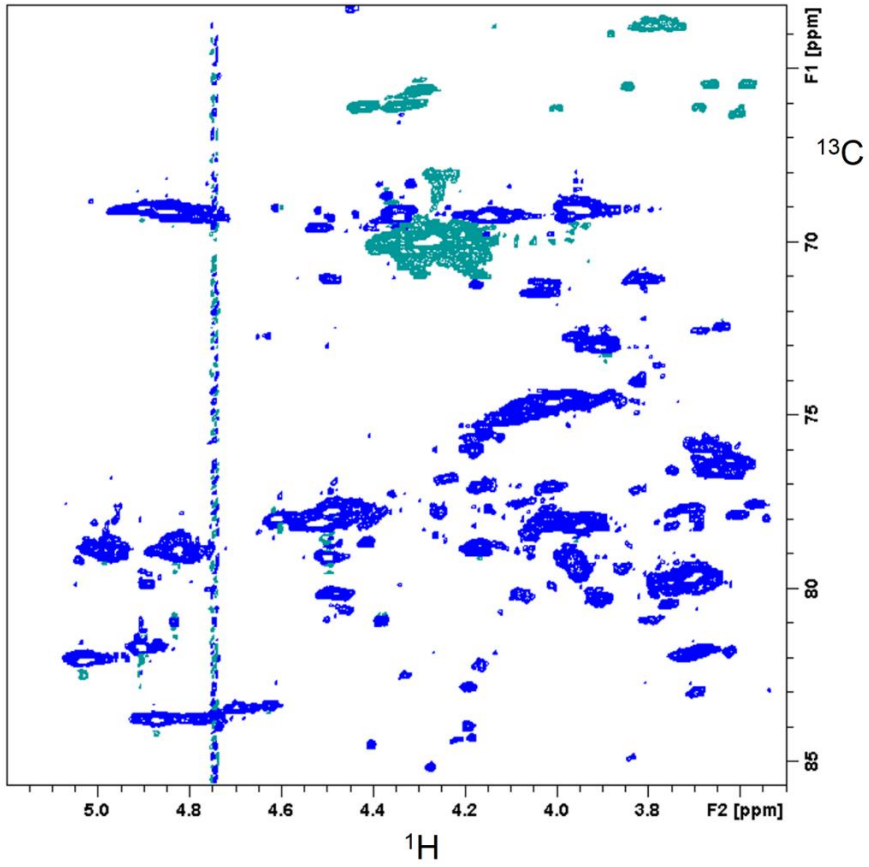


Fig. 6. 3. 1. Partial (non-anomeric region) 2D ^{1}H, ^{13}C BIRD^{r,X}-HSQC spectrum.
Traditional 2D HSQC spectrum of the fCS oligosaccharide mixture under study. As the sample contains multiple oligosaccharide chains of different lengths and composition, the spectrum contains many groups of overlapping peaks and is difficult to analyse.

Although not essential, the (3,2)D BIRD^{r,X}-HSQC experiment helps to identify the position of the direct proton-carbon correlation cross peaks in the (3,2)D BIRD^{r,X}-HSQC-TOCSY spectra and is therefore a useful tool for the analysis of the complex spectra. A complete (3,2)D

BIRD^{r,X}-HSQC spectrum, overlaid with the traditional 2D HSQC spectrum is presented in Fig. 6. 3. 2. A, whilst Fig. 6. 3. 2. B presents an expanded view of the anomeric region.

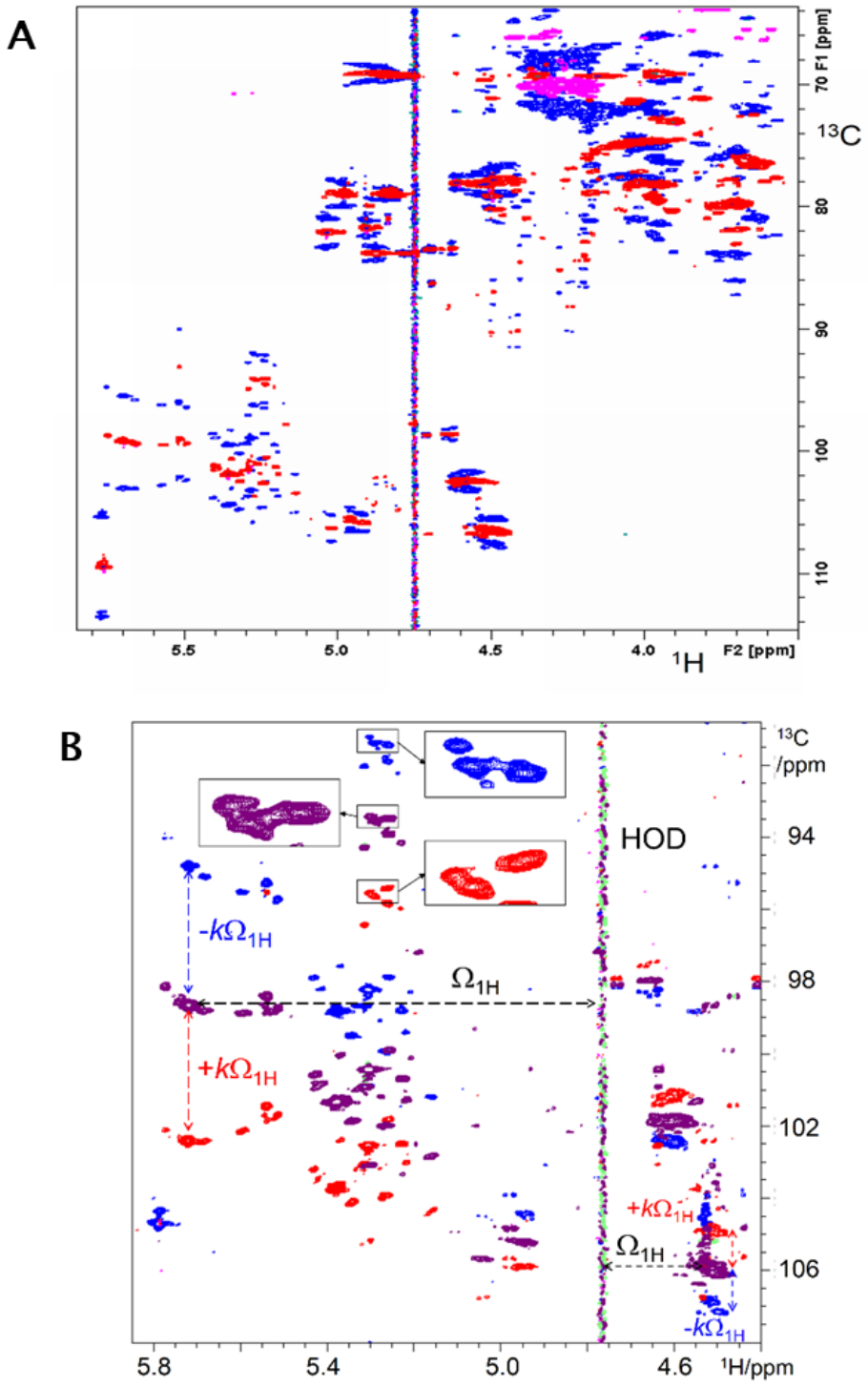


Fig. 6. 3. 2. Overlay of 2D ^{1}H, ^{13}C HSQC spectrum and ^{1}H, ^{13}C (3,2)D BIRD^{r,X}-HSQC spectrum. A Complete spectrum of the fCS sample – the 2D HSQC spectrum is shown in red and the (3,2)D BIRD^{r,X}-HSQC spectrum is shown in blue. **B** Expansion of the anomeric region of the spectrum. The 2D HSQC spectrum is shown in purple and the two parts of the (3,2)D BIRD^{r,X}-HSQC spectrum, $\Omega_{13C} + \Omega_{1H}$ and $\Omega_{13C} - \Omega_{1H}$, are shown in red and blue, respectively.

Cross peaks in the F_1 dimension of the (3,2)D BIRD^{r,X}-HSQC spectrum are displaced by $\pm\Omega_{1H}$ relative to the cross peaks in a regular HSQC spectrum (k = 1 was used). In this instance, the ^{1}H carrier frequency was set on the HOD signal, resulting in the cross peaks to the left (positive Ω_{1H}) and right (negative Ω_{1H}) of the water signal to swap the direction of the displacement. Expansions in the rectangles allow to see how the cross peaks which were overlapped in the traditional HSQC spectrum became resolved in the GFT version of this experiment (Fig, 6. 3. 2. B). Depending on the local topology of the cross peaks, increased resolution is achieved in either the $\Omega_{13C} + \Omega_{1H}$ or the $\Omega_{13C} - \Omega_{1H}$ part of the spectrum or in both.

6. 3. 2. (3,2)D BIRD^{r,X}–HSQC–TOCSY

In the (3,2)D BIRD^{r,X}-HSQC-TOCSY spectrum cross peaks are displaced from their positions in a traditional 2D HSQC-TOCSY experiment in the same manner as in the (3,2)D BIRD^{r,X}-HSQC spectrum: by either $+\Omega_{1H}$ or $-\Omega_{1H}$. A complete (3,2)D BIRD^{r,X}-HSQC-TOCSY spectrum, overlaid with the regular 2D BIRD^{r,X}-HSQC-TOCSY spectrum is shown in Fig. 6. 3. 3.

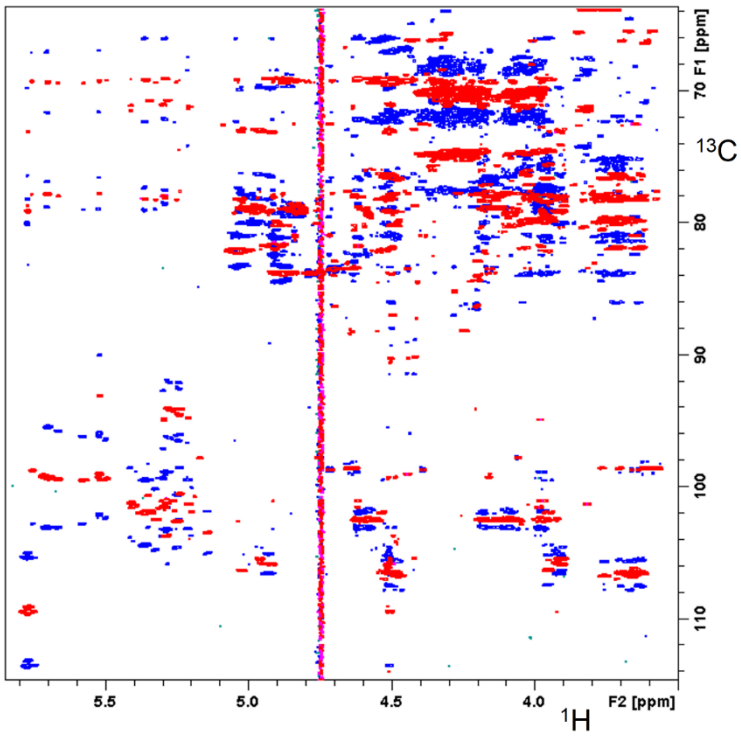


Fig. 6. 3. 3. Overlay of 2D ^{1}H, ^{13}C HSQC-TOCSY spectrum and ^{1}H, ^{13}C (3,2)D BIRD^{r,X}-HSQC-TOCSY spectrum. The 2D HSQC-TOCSY spectrum is shown in red and the (3,2)D BIRD^{r,X}-HSQC-TOCSY spectrum is shown in blue.

In order to demonstrate how the GFT experiments can help to resolve some of the problematic regions of the carbohydrate spectra, analysis of a fragment of the (3,2)D BIRD^{r,X}-HSQC-TOCSY spectrum of the fCS oligosaccharide mixture is presented in Fig. 6. 3. 4. Even in this small fragment of the spectrum, a large number of peaks can be seen which may initially make the analysis difficult. In this figure, three different spectra are overlaid: traditional 2D HSQC and 2D HSQC-TOCSY and a (3,2)D BIRD^{r,X}-HSQC-TOCSY. Although there is a large group of peaks on the right hand side of the spectrum, corresponding to the direct and TOCSY cross peaks for H5/H6 of GalNAc, this group of peaks will be ignored for the purpose of this discussion. Peaks of special interest here are the four groups of cross peaks shown in pink that resonate at ~66.5 ppm in F_1, located between the two pink arrows. These peaks are direct correlation cross peaks from a 2D HSQC spectrum. In a regular 2D HSQC-TOCSY spectrum, TOCSY correlation peaks from those signals will all appear at ~66.5 ppm, which makes it impossible to distinguish which of the signals are correlated with each other. However, because of their different position along the F_2 dimension, these signals will be displaced to different positions along the F_1 axis in the (3,2)D HSQC and (3,2)D HSQC-TOCSY experiments (indicated by black arrows). This allows to separate the TOCSY correlations to different lines and, therefore, to assign the spectrum more easily.

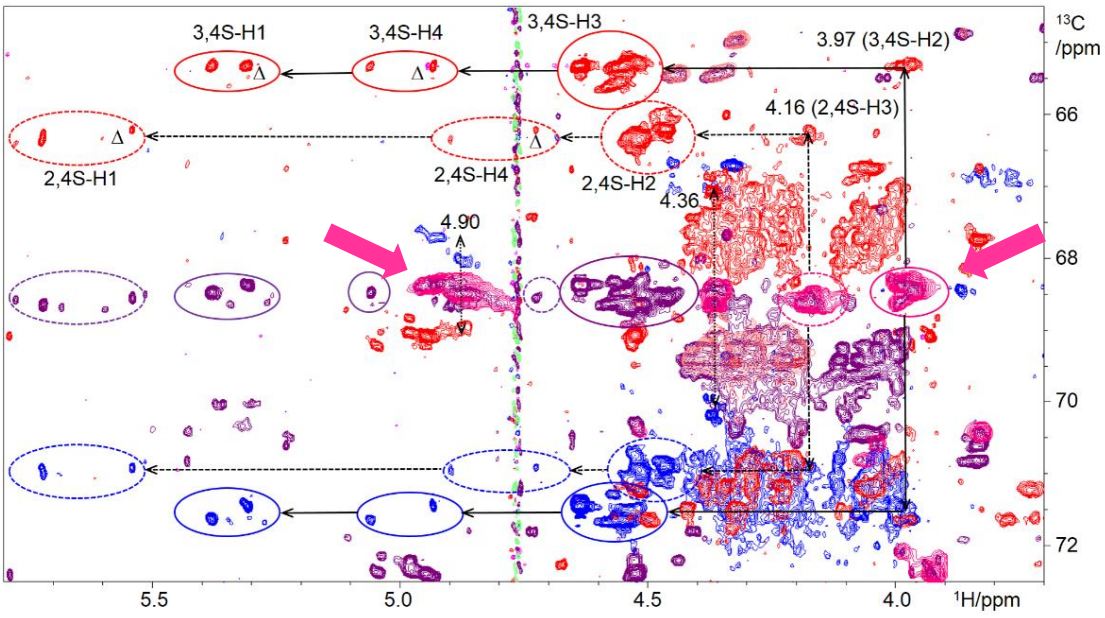


Fig. 6. 3. 4. Overlay of the regular 2D ^{1}H, ^{13}C HSQC spectrum (pink), 2D ^{1}H, ^{13}C HSQC-TOCSY spectrum (purple) and ^{1}H, ^{13}C (3,2)D BIRD^{r,X}-HSQC-TOCSY spectra (red and blue). The thick pink arrows indicate four groups of cross peaks that overlap in the ^{13}C dimension, while the vertical and horizontal black arrows show the displacement and the positions of the direct and the TOCSY (circled) cross peaks, respectively.

In this example, the TOCSY cross peaks are stronger than the direct correlation peaks, as the magnetisation was efficiently transferred from the ^{13}C- to ^{12}C-attached protons during a 40 ms TOCSY mixing time. The HSQC cross peaks at the δ_{1H} of 4.90 and 4.36 ppm do not show any TOCSY correlations in the displayed region of the spectra and will be discussed later. On the other hand, signals resonating at 4.16 and 3.97 ppm show several HSQC-TOCSY correlations (purple circles), such as the cross peaks overlapping in the proton dimension of a regular HSQC-TOCSY spectrum between 4.44 and 4.66 ppm. This overlap is removed in the (3,2)D BIRD^{r,X}-HSQC-TOCSY spectra (cross peaks circled by full and dashed red and blue lines). Analysis of the spectra showed that the original HSQC cross peaks belong to protons H3 and H2 of fucose 2,4-sulfated (4.16 ppm) and 3,4-sulfated (3.97 ppm), respectively. These protons show TOCSY transfers to protons H2 from 2,4S fucose (4.56 – 4.44 ppm, dashed circle) and H3 from 3,4S fucose (4.66 – 4.49 ppm, full line circle), respectively. Their separation in the (3,2)D BIRD^{r,X}-HSQC-TOCSY spectra is a consequence of the unique ^{1}H chemical shifts of H3 (4.16 ppm) and H2 (3.97 ppm) protons.

Transfer of magnetisation from the H3 and H2 protons is seen to continue to protons H4 and H1 of their individual monosaccharide rings, an observation that allowed their assignment. These (H4, H1)/C2 and (H4, H1)/C3 cross peaks are also circled in Fig. 6. 3. 4 and appear in pairs between 4.7 and 5.7 ppm. The observed doubling of signals with distinct ^{1}H chemical shifts within each circle is a consequence of the depolymerisation. Cleavage of the glycosidic bond generated fucose linked to a terminal modified GlcA monosaccharide with a C4-C5 double bond (ΔGlcA), while the original inner GlcA units are still present to some extent. A comparison with the chemical shifts of the native fCS allowed to establish that the H1 and H4 chemical shifts of the terminal fucose linked to ΔGlcA are always up to 0.2 ppm smaller. Their cross peaks are labelled with a Δ symbol in Fig. 6. 3. 4.

The other two groups of cross peaks at ~66.5 ppm in F_1 and at 4.87 and 4.36 ppm in F_2, were assigned to H5/C5 of fucose, based on their TOCSY transfer to H6 protons that produced H6/C5 cross peaks in the empty region of the HSQC-TOCSY spectra (1.30 – 1.45/~66.5 ppm). The fact that both of these groups of signals belong to H5 of fucose was, however, only established through the analysis of (3,2)D BIRD^{r,X}-HSQC-TOCSY spectra. Figure 6. 3. 5 illustrates how the ambiguity of a regular HSQC-TOCSY experiment, where it is not clear whether the transfer originated on one or both of these protons, is removed in the (3,2)D BIRD^{r,X}-HSQC-TOCSY spectra.

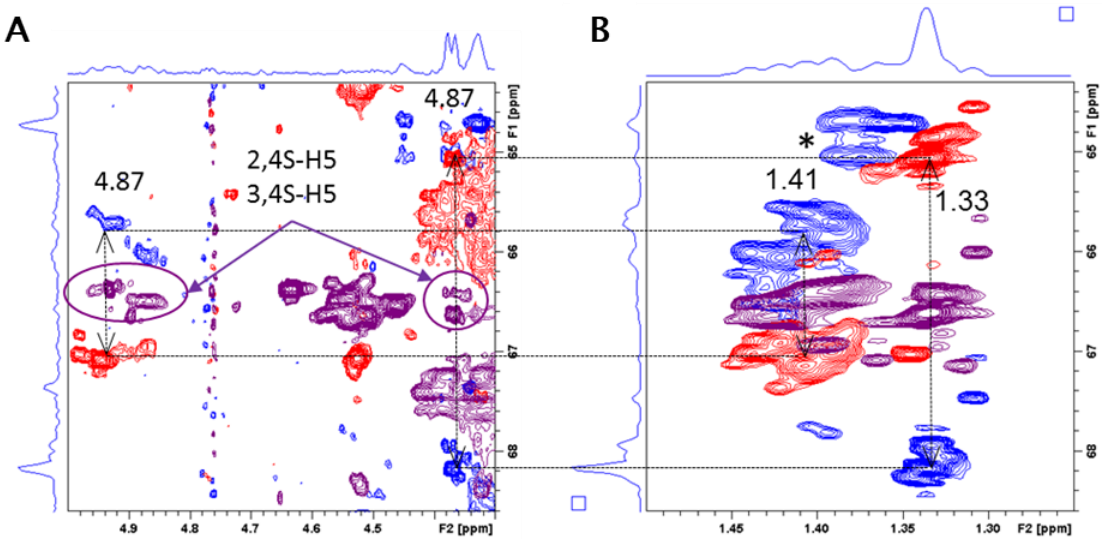


Fig. 6. 3. 5. Overlay of the regular 2D ^{1}H, ^{13}C HSQC-TOCSY (purple) spectrum and ^{1}H, ^{13}C (3,2)D BIRD^{r,X}-HSQC-TOCSY spectra (blue and red), focusing on H5/C5 (A) and H6/C5 (B) cross peaks. The black arrows indicate displacement along the vertical axis and horizontal lines connect the H5 and H6 cross peaks. Blue cross peaks labelled by an asterisk indicate additional CH_3 signals.

In the native fCS polysaccharide fucose ring is stacked above the neighbouring GalNAc residue, enabling a formation of an unusual hydrogen bond from its H5 proton to the ring oxygen of GalNAc, as described previously[55,194]. During depolymerisation, β-elimination breaks the GalNAc(1→4)GlcA glycosidic bond, making fucose a terminal unit in the newly formed oligosaccharides. This results in significant lowering of the H5 chemical shift (by ca. 0.51 ppm) and is indicative of successful depolymerisation. As seen above, the H1 and H4 of fucose are also affected in a similar manner, although to a lesser degree.

The (3,2)D BIRD^{r,X}-HSQC-TOCSY experiment also enabled assignment of H6 protons of both forms of fucose. As the coupling constant between protons H4 and H5 of fucose is close to zero, no transfer is possible from H5 to H4 protons and the correlations to H6 protons are strong. As seen in Fig. 6. 3. 5, protons H6 appear at around 1.41 and 1.33 ppm, for native and terminal fucose respectively, following the chemical shift trends seen for the other protons of fucose in different environments. Correlations in the opposite direction, from H6 to H5, are also resolved in a (3,2)D BIRD^{r,X}-HSQC-TOCSY spectrum presented in Fig. 6. 3. 6. In the regular HSQC-TOCSY spectrum, these peaks were overlapped and it was not possible to deduce whether these signals belong to the inner or terminal fucose.

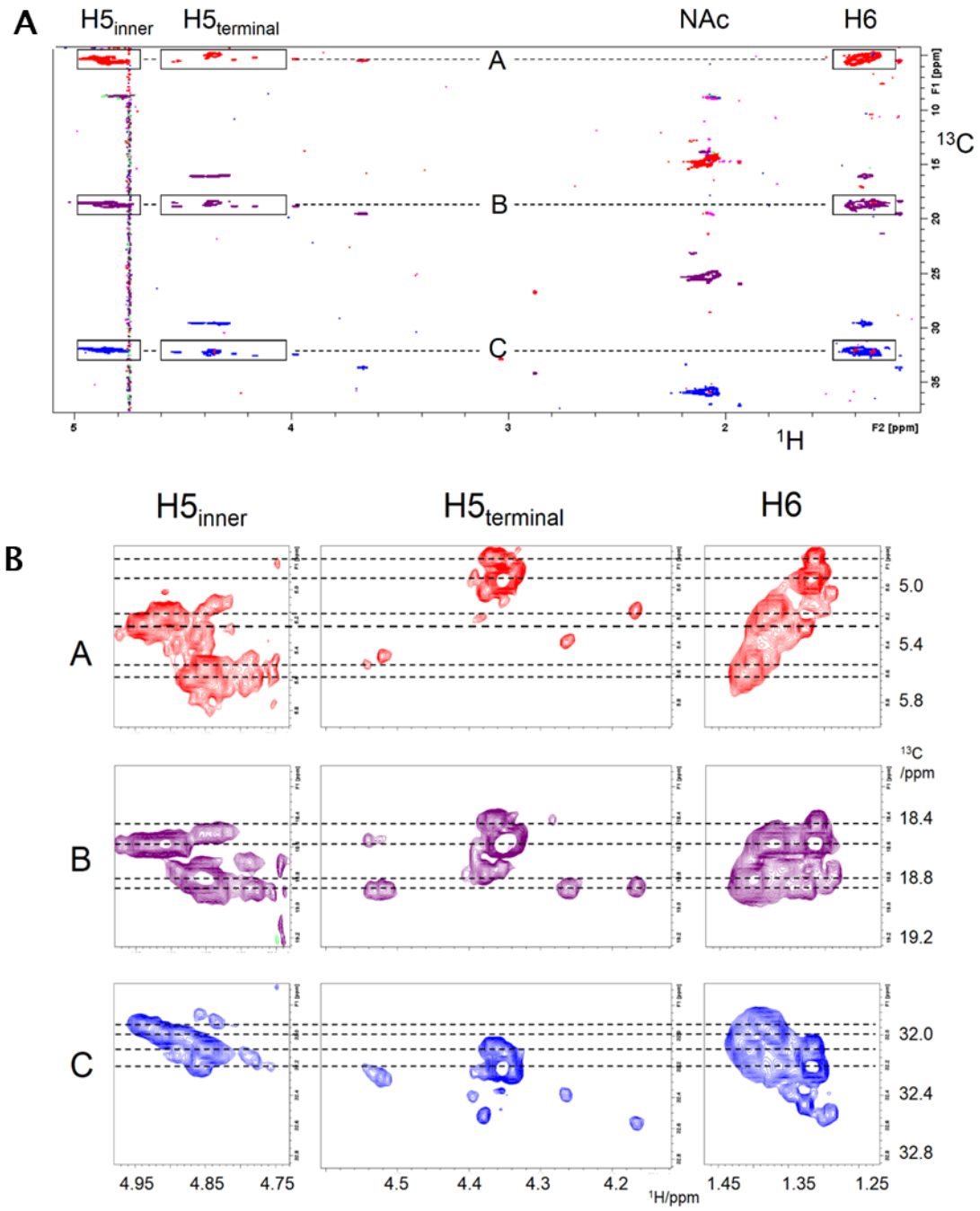


Fig. 6. 3. 6. The methyl region of 2D ^{1}H, ^{13}C HSQC-TOCSY spectrum (purple) and ^{1}H, ^{13}C (3,2)D BIRD^{r,X}-HSQC-TOCSY spectra (blue and red). A Overview of the methyl region for the depolymerised fCS mixture, as well as the correlated H5 protons. **B** Expanded view of the groups of peaks highlighted in **A**. Displacement of the H5 peaks in (3,2)D BIRD^{r,X}-HSQC-TOCSY spectra allows to assign the connections to specific H6 signals.

Upon the examination of the (3,2)D BIRD^{r,X}-HSQC-TOCSY it became clear that it is useful to inspect both the $\Omega_{13C} + \Omega_{1H}$ and $\Omega_{13C} - \Omega_{1H}$ spectra. Peak separation in each spectrum follows a different pattern, depending on the pattern of correlations between ^{1}H and ^{13}C signals in the original 2D HSQC-TOCSY spectrum. If ^{13}C chemical shifts are identical, separation of signals is identical in both spectra (Fig. 6. 3. 7. A). However, in cases when the proton and carbon chemical shifts in the overlapped areas both increase, better cross peak separation is achieved in the $\Omega_{13C} - \Omega_{1H}$ spectrum (Fig. 6. 3. 7. B). On the other hand, in cases when ^{13}C chemical shift decreases and ^{1}H chemical shift increases (inverse correlation), better cross peak separation is achieved in the $\Omega_{13C} + \Omega_{1H}$ spectrum (Fig. 6. 3. 7. C). Inspecting both spectra is also useful when the displaced peaks become overlapped with other peaks in one spectrum, as a clearer view can be provided by the second spectrum.

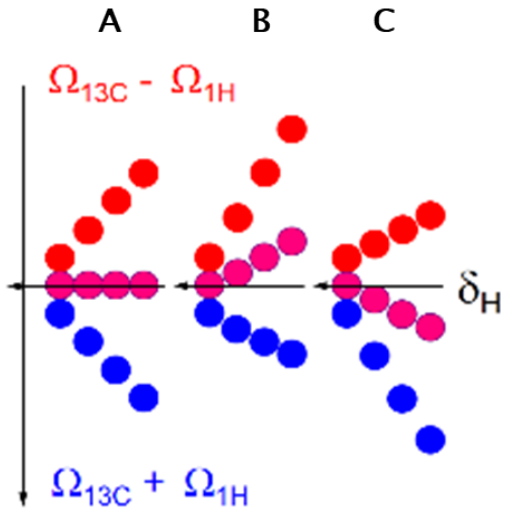


Fig. 6. 3. 7. A schematic representation of different patterns of peak displacement in the two ^{1}H, ^{13}C (3,2)D BIRD^{r,X}-HSQC-TOCSY spectra. A The pattern of peak separation in two spectra (red and blue for $\Omega_{13C} - \Omega_{1H}$ and $\Omega_{13C} + \Omega_{1H}$, respectively) is identical when ^{13}C chemical shifts in the original HSQC spectrum (pink) are the same for all peaks. **B** Better separation is obtained in the $\Omega_{13C} - \Omega_{1H}$ spectrum (red) when ^{1}H and ^{13}C chemical shifts increase at the same time. **C** Better separation is obtained in the $\Omega_{13C} + \Omega_{1H}$ spectrum (blue) when ^{1}H and ^{13}C chemical shifts are inversely correlated.

These points are further illustrated in Fig. 6. 3. 8, where an overlay of four partial spectra is presented. Focusing on the overlapping 2,4S-H2 and 3,4S H3 2D HSQC cross peaks at ~4.52 ppm (pink circle) and the 2,4S-H3 HSQC-TOCSY cross peaks at ~4.17 ppm (red and blue circles), it can be seen that the separation of cross peaks is better in the $\Omega_{13C} - \Omega_{1H}$ (red)

spectrum than in the $\Omega_{13C} + \Omega_{1H}$ spectrum (blue). This is a consequence of the topology of the 2,4S-H2 HSQC cross peaks indicated by the dotted arrows.

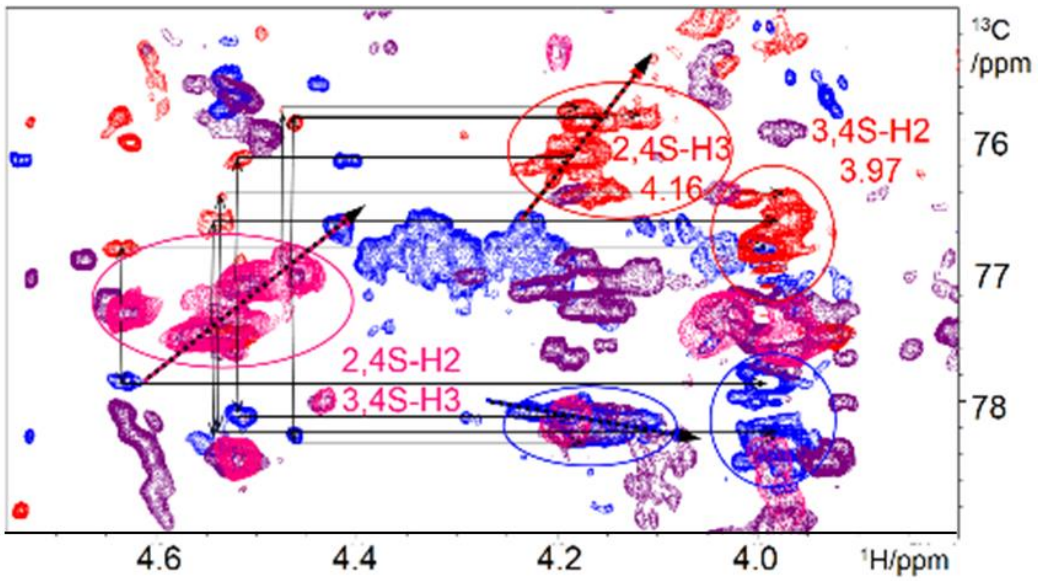


Fig. 6. 3. 8. Overlay of the regular 2D ^{1}H, ^{13}C HSQC (pink), 2D ^{1}H, ^{13}C HSQC-TOCSY (purple) spectrum and ^{1}H, ^{13}C (3,2)D BIRD^{r,X}-HSQC-TOCSY spectra (blue and red), focusing on H5/C5 (A) and H6/C5 (B) cross peaks. The black arrows indicate displacement along the vertical axis and horizontal lines connect the TOCSY cross peaks. The dotted arrows indicate a different degree of cross peak separations in the 2D HSQC and consequently also in the (3,2)D HSQC-TOCSY in $\Omega_{13C} - \Omega_{1H}$ (red) and $\Omega_{13C} + \Omega_{1H}$ (blue) spectra.

6.3. Conclusions

The presented methodology introduces a new application of a powerful 2D ^{1}H, ^{13}C HSQC-TOCSY experiment to carbohydrate samples with signal overlap in the ^{13}C dimension. The proposed (3, 2)D BIRD^{r,X}-HSQC-TOCSY experiment preserves high digital resolution in the indirectly detected (F_1) dimension while accepting some sensitivity loss. The method was illustrated on the assignment of fucose resonances in a mixture of oligosaccharides containing FucX(1→3)GlcA and FucX(1→3)ΔGlcA disaccharide fragments. The achieved resolution also made possible a more detailed assignment of individual resonances belonging to 2,4 or 3,4 sulfated fucose in different fCS oligosaccharides and accounted for the effects of reducing rings, which are a consequence of depolymerisation.

Although illustrated here on the TOCSY transfer, the same principles can also be applied to a NOESY experiment. In conclusion, the (3, 2)D BIRD^{r,X}-HSQC-TOCSY/NOESY are a powerful addition to the limited arsenal of NMR techniques specifically designed to enable structure elucidation of carbohydrate mixtures.

Chapter 7: Summary and future directions

The main purpose of this thesis was to study the interactions between the fCS oligosaccharides from the sea cucumber species *H. forskali* and L- and P-selectins. This research aimed to provide new information on the nature of the fCS-selectin binding and to reveal structural data that could be used in the development of new drugs targeting selectins in treatment of conditions such as inflammation, cancer metastasis and HIV. The fCS-selectin interactions were studied using two different methods – the Saturation Transfer Difference (STD) NMR and the *in silico* modelling of the fCS-selectin complexes.

STD experiments performed with the use of the commercially available L- and P-selectin constructs confirmed that the fCS oligosaccharides bind to the extracellular domains of these two proteins. Comparison of the STD response upon fCS binding to those two constructs indicates that the mode of binding is likely similar for both proteins. It was also possible to identify the regions within the fCS that are located in close proximity to the selectin surface upon binding. The largest response was observed for the methyl groups of galactosamine and fucose residues and from the fucose anomeric proton. Additional 2D STD-TOCSY experiment allowed to confirm the identity of other signals, such as protons H2, H3 and H5 of fucose and H1 and H2 of glucuronic acid. It was also possible to establish that fCS oligosaccharides of different sulfation pattern bind to selectins with different affinities.

An STD experiment on the fCS binding to the L-selectin construct consisting of only the C-type lectin domain was also performed. This experiment allowed to establish that the mode of binding is similar for a single-domain construct as for the larger constructs of L-selectin. No significant changes in the STD spectrum were observed in absence of the EGF-like and the SCR domains. This is an interesting observation as it is currently believed that the EGF-like domain plays a role in selectin-ligand binding and it has been shown that this domain influences the ability of cells to attach to the vascular endothelium. Unfortunately, the STD experiments do not provide enough structural information to elucidate the exact mechanism of the binding and make a detailed comparison for both constructs. It appears however, that the presence of the EGF-like domain is not required for the fCS-L-selectin binding to occur. Further STD experiments could be performed in future, using the same selectin constructs and testing their binding to native L-selectin ligand such as sLe^x. Use of other techniques, such as the isothermal titration calorimetry (ITC) or the surface plasmon resonance (SPR),

could also be explored in future to further investigate and compare the binding between the two L-selectin constructs and fCS oligosaccharides.

Further insight into the fCS-selectin interactions was obtained from a series of computational studies on the binding between the fCS trisaccharides to P-selectin. Docking of the fCS oligosaccharides to P-selectin PDB structure 1G1Q allowed to establish that the binding site for fCS 2,4S and 3,4S trisaccharide is most likely the Ca^{2+}-coordination site and that the P-selectin residues predicted to participate in the binding are Tyr48, Glu80, Asn82, Asn83, Tyr94, Ser97, Asn105 and Lys111. These results indicate that both fCS trisaccharides inhibit P-selectin function by binding to the Ca^{2+}-coordination site and preventing its interaction with the natural selectin ligand sialyl Lewis X (sLe^x). The computational studies also allowed to explore the dynamic behaviour of P-selectin upon binding to these two different trisaccharides. It was found that the P-selectin is able to undergo a conformational switch from the bent to extended state upon binding to fCS 3,4S trisaccharide at the Ca^{2+}-coordination site and that this movement does not cause any changes in the conformation of the ligand.

In addition to the new information about the fCS-P-selectin binding, new insight into the binding between the sLe^x and the E-selectin was also obtained. The MD studies confirmed that sLe^x binds with higher energetic preference to the extended conformation of E-selectin (PDB structure 4CSY) than to the bent conformation (PDB structure 1G1T), and that the interactions were more stable upon binding to the 4CSY structure. A possible reason for this behaviour could be the position of the loop 83 – 90, which in the 4CSY structure points towards the carbohydrate and interacts with the fucose residue. The position of this loop has been previously correlated to the affinity of selectins for their ligands in both E- and P-selectins. However, none of the computational studies undertaken in this work have detected a switch in position of this loop in either the bent or extended conformation of these two proteins.

It was also found that in the 4CSY structure, the E-selectin exhibits a higher degree of flexibility in the interdomain region and adopts a new conformational 'hyper-extended' state. This conformation is achieved by a twisting movement of the EGF-like domain which turns by 50° degrees from its position in the extended conformation. After switching to this position, residues Thr136 – Asn138 of the EGF-like domain were able to develop interactions with the N-terminal region of the protein. As it was not the main focus of this study, the reason for this movement was not studied in more detail here. However, as the 4CSY structure is the only selectin structure for which the movement was observed, it would be an

interesting point of future research to determine the structural features within this protein that allow for such movement. Similarly as in the case of the fCS-P-selectin binding, the changes in protein conformation did not lead to any observable changes in the conformation of the carbohydrate ligand.

Although the computational studies described here allowed to obtain new information about the fCS-selectin binding, they suffered from some imperfections. The docking software struggled to predict stable conformation of the oligosaccharides upon binding to selectins. Autodock Vina is a software more suitable for docking of smaller ligands to proteins with a clearly defined binding pocket. Future studies could test the use of different types of the docking software such as Dock, ICM, FlexX or Glide. Freezing the oligosaccharide in a single conformation during the docking and making it flexible only for the MD component of the simulation could also help in finding more stable docked complexes. For larger oligosaccharides, it could be possible to dock small segments of the carbohydrate in increments - such approach has previously been used to dock large ligands in Autodock[195].

In order to obtain more detailed structural information on the fCS-selectin binding, it would be desirable to have an isotopically labelled sample of the L- or P-selectin extracellular domains that could be used in the protein-observed NMR experiments. In such experiments, the ligand is titrated into the protein sample and changes in the chemical shifts of the protein signals are monitored, which allows to identify the amino acid residues that interact with the ligand. Therefore, another aim of this work was to develop protocols for expression of ^{13}C,^{15}N-labelled L- and P-selectin constructs consisting of either the C-type lectin domain only or of both the C-type lectin and the EGF-like domains. Extensive work was undertaken in which many protocols for expression of these constructs in either the *P. pastoris* or *E. coli* were tested. Unfortunately, none of these tests were successful and, therefore, it was not possible to establish a suitable protocol for expression of the isotopically labelled selectin constructs. However, the experimental work described here can provide guidance for future attempts of expression of selectin constructs. It is possible that these proteins could be expressed if a different construct design was implemented. For example, expression levels of the soluble protein in *E. coli* could be increased by using a different fusion partner or by co-expression of molecular chaperones that assist in protein folding processes. In some cases, use of a weaker promoter can help to increase the yield of soluble protein in *E. coli* cytoplasm. Alternatively, intracellular expression in *P. pastoris* could be tested, which would not require the addition of the N-terminal signalling sequence, believed to interfere with the correct folding of selectin constructs described here. Different strains of *P. pastoris* could also be

used. For example, strain SMD1168 is deficient in proteases and, therefore, could help to prevent protein degradation during expression.

Despite the lack of success with expression of selectin constructs in the bacterial and yeast cells, work undertaken in this project allowed to express a single-domain L-selectin construct in HEK293 cells, in collaboration with the Oxford Protein Production Facility. This construct was successfully used in the STD NMR experiments and could be used in the future studies on the L-selectin-ligand binding. For example, mutant versions of this construct could be produced to test the role of individual amino acid residues in the binding of selectin C-type lectin domain to the fCS oligosaccharides, verifying the findings of the computational studies. As proteins expressed in mammalian cells have previously been labelled with ^{13}C and ^{15}N isotopes, it could be possible to develop protocols for production of isotopically labelled L-selectin construct in HEK293 cells in the future.

Overall, this project provided a new insight into the interaction between the L- and P-selectins and the fCS oligosaccharides. The information obtained in this work could help in development of future experiments on expression of selectin constructs, the studies of selectin-ligand interactions and potentially in the design of future drugs.

Further work undertaken as part of this project allowed to develop a new NMR method for the study of carbohydrate mixtures, incorporating the concepts of G-matrix Fourier Transform (GFT) NMR into 2D HSQC and 2D HSQC-TOCSY experiments. The method allows to separate peaks in the regions of high spectral overlap in the ^{13}C dimension, based on the difference in the ^{1}H chemical shifts. This method was applied to the study of an fCS sample, containing a mixture of oligosaccharides produced via β-eliminative depolymerisation. It allowed to assign individual resonances of 2,4- and 3,4-sulfated fucose in different fCS oligosaccharides and account for the effects of reducing rings. The proposed (3, 2)D $BIRD^{r,X}$-HSQC-TOCSY experiment preserves high digital resolution and could be used in analysis of mixtures containing carbohydrates or small molecules. Other variants of this method, for example incorporating NOESY experiment, could be developed in future to further assist with analysis of complex carbohydrate mixtures.

References

1. Yamada S., Sugahara K., Özbek S.: **Evolution of glycosaminoglycans. Comparative biochemical study.** *Commun. Integr. Biol.* **4:2**, 150-158 (2011).
2. Imberty, A., Lortat-Jacob, H., Pérez, S.: **Structural view of glycosaminoglycan–protein interactions.** *Carbohydr. Res.* **342,** 430–439 (2007).
3. Gandhi, N. S., Mancera, R. L. **The structure of glycosaminoglycans and their interactions with proteins.** *Chem. Biol. Drug Des.* **72,** 455–482 (2008).
4. Esko, J. D., Kimata, K., Lindahl, U.: **Proteoglycans and Sulfated Glycosaminoglycans.** in *Essentials of Glycobiology* (ed. Varki, A. et al.)(Cold Spring Harbor Laboratory Press, 2009).
5. Xu, X., Dai, Y. **Heparin: an intervenor in cell communication**. *J. Cell. Mol. Med.* **14,** 175–180 (2010).
6. Sasisekharan, R., Raman, R., Prabhakar, V.: **Glycomics Approach to Structure-Function Relationships of Glycosaminoglycans.** *Annu. Rev. Biomed. Eng.* **8,** 181–231 (2006).
7. Volpi, N.: **Dermatan sulfate: Recent structural and activity data.** *Carbohydr. Polym.* **82,** 233–239 (2010).
8. Prydz, K.: **Determinants of Glycosaminoglycan (GAG) Structure.** *Biomol*ecules **5,** 2022–1120 (2015)
9. Gray, E., Hogwood, J., Mulloy, B.: **The Anticoagulant and Antithrombotic Mechanisms of Heparin.** In: Lever, R., Mulloy, B., Page, C (eds) *Heparin - A Century of Progress. Handbook of Experimental Pharmacology*, **207**, (2012) Springer, Berlin, Heidelberg.
10. Khorana, A. A., Sahni, A., Altland, O. D., Francis, C. W.: **Heparin inhibition of endothelial cell proliferation and organization is dependent on molecular weight.** *Arterioscler. Thromb. Vasc. Biol.* **23,** 2110–2115 (2003).
11. Wei, M., Tai, G., Gao Y., Li, N., Huang, B., Zhou, Y., Hao, S., Zeng, X.: **Modified heparin inhibits P-selectin-mediated cell adhesion of human colon carcinoma cells to immobilized platelets under dynamic flow conditions.** *J. Biol. Chem.* **279,** 29202–29210 (2004).
12. Afratis, N., Gialeli, C., Nikitovic, D., Tsegenidis, T., Karousou, E., Theocharis, A. D., Pavão, M. S., Tzanakakis, G. N., Karamanos, N. K.: **Glycosaminoglycans: key players in cancer cell biology and treatment.** *FEBS J.* **279,** 1177–1197 (2012).
13. Theocharis, A. D., Tsolakis, I., Tzanakakis, G. N. & Karamanos, N. K.: **Chondroitin Sulfate as a Key Molecule in the Development of Atherosclerosis and Cancer Progression.** *Adv Pharmacol.* **53,** 281–295 (2006).
14. Galtrey, C. M., Fawcett, J. W.: **The role of chondroitin sulfate proteoglycans in regeneration and plasticity in the central nervous system.** *Brain Res. Rev.* **54,** 1–18 (2007).
15. Zou, X. H., Foong, W. C., Cao, T., Bay, B. H., Ouyang, H. W., Yip G. W.: **Chondroitin sulfate in palatal wound healing.** *J. Dent. Res.* **83,** 880–885 (2004).
16. Trowbridge, J. M., Gallo, R. L.: **Dermatan sulfate: new functions from an old glycosaminoglycan.** *Glycobiology* **12,** 117R–125R (2002).
17. Cardin, A. D., Weintraub, H. J.: **Molecular modeling of protein-glycosaminoglycan interactions.** *Arterioscler. Thromb. Vasc. Biol.* **9,** 21–32 (1989).
18. Hileman, R. E., Fromm, J. R., Weiler, J. M., Linhardt, R. J.: **Glycosaminoglycan-protein interactions: definition of consensus sites in glycosaminoglycan binding proteins.** *BioEssays* **20,** 156–167 (1998).

19. Colliec-Jouault, S., Bavington, C., Delbarre-Ladrat, C.: **Heparin-like Entities from Marine Organisms. in *Heparin***. *Handb Exp Pharmacol.* **207,** 423–449 (2012).
20. Kozlowski, E. O., Gomes, A. M., Silva, C. S., Pereira, M. S., de Vilela Silva, A. C. E. S., Pavão, M. S. G.: **Structure and Biological Activities of Glycosaminoglycan Analogs from Marine Invertebrates: New Therapeutic Agents?** in *Glycans in Diseases and Therapeutics* (ed. Pavão, M. S. G.) 159–184 (Springer Berlin Heidelberg, 2011).
21. Pavão, M. S. G. **Glycosaminoglycans analogs from marine invertebrates: structure, biological effects, and potential as new therapeutics.** *Front. Cell. Infect. Microbiol.* **4,** (2014).
22. Pavão, M. S. G., Aiello, K. R., Werneck, C. C., Silva, L. C., Valente, A. P., Mulloy, B., Colwell, N. S., Tollefsen, D. M., Murão P. A.: **Highly Sulfated Dermatan Sulfates from Ascidians. Structure versus anticoagulant activity of these glycosaminoglycans.** *J. Biol. Chem.* **273,** 27848–27857 (1998).
23. Mansour, M. B., Dhahri, M., Hassine, M., Ajzenberg, N., Venisse, L., Ollivier, V., Chaubet, F., Jandrot-Perrus, M., Maaroufi., R. M.: **Highly sulfated dermatan sulfate from the skin of the ray Raja montagui: anticoagulant activity and mechanism of action.** *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **156,** 206–215 (2010).
24. Kozlowski, E. O., Pavao, M. S. G., Borsig, L.: **Ascidian dermatan sulfates attenuate metastasis, inflammation and thrombosis by inhibition of P-selectin.** *J. Thromb. Haemost.* **9,** 1807–1815 (2011).
25. Vieira, R. P., Mulloy, B., Mourão, P. A.: **Structure of a fucose-branched chondroitin sulfate from sea cucumber. Evidence for the presence of 3-O-sulfo-beta-D-glucuronosyl residues.** *J. Biol. Chem.* **266,** 13530–13536 (1991).
26. Yoshida, K., Minami, Y., Nemoto, H., Numata, K., Yamanaka, E.: **Structure of DHG, a depolymerized glycosaminoglycan from sea cucumber, Stichopus japonicus.** *Tetrahedron Lett.* **33,** 4959–4962 (1992).
27. Bordbar, S., Anwar, F., Saari, N.: **High-Value Components and Bioactives from Sea Cucumbers for Functional Foods—A Review.** *Mar. Drugs* **9,** 1761–1805 (2011).
28. San Miguel-Ruiz, J. E., García-Arrarás, J. E.: **Common cellular events occur during wound healing and organ regeneration in the sea cucumber Holothuria glaberrima.** *BMC Dev. Biol.* **7,** 115 (2007).
29. Herencia, F., Ubeda, A., Ferrándiz, M. L., Terencio, M. C., Alcaraz, M. J., Garcia-Carrascosa, M., Capaccioni, R., Payá, M.: **Anti-inflammatory activity in mice of extracts from Mediterranean marine invertebrates.** *Life Sci.* **62,** PL115-120 (1998).
30. Mourão, P. A. S., Pereira, M. S., Pavão M. S., Mulloy, B., Tollefsen, D. M., Mowinckel, M. C., Abildgaard, U.: **Structure and Anticoagulant Activity of a Fucosylated Chondroitin Sulfate from Echinoderm. Sulfated fucose branches on the polysaccharide account for its high anticoagulant action.** *J. Biol. Chem.* **271,** 23973–23984 (1996).
31. Mourão, P. A., Giumarães, B., Mulloy, B., Thomas, S., Gray, E.: **Antithrombotic activity of a fucosylated chondroitin sulphate from echinoderm: sulphated fucose branches on the polysaccharide account for its antithrombotic action.** *Br. J. Haematol.* **101,** 647–652 (1998).
32. Mourão, P. A., Pereira, M. S.: **Searching for alternatives to heparin: sulfated fucans from marine invertebrates.** *Trends Cardiovasc. Med.* **9,** 225–232 (1999).
33. Mourão, P. A. Boisson-Vidal, C., Tapon-Bretaudière, J., Drouet, B., Bros, A., Fischer, A.: **Inactivation of thrombin by a fucosylated chondroitin sulfate from echinoderm.** *Thromb. Res.* **102,** 167–176 (2001).
34. Fonseca, R. J. C., Mourão, P. A. S.: **Fucosylated chondroitin sulfate as a new oral antithrombotic agent.** *Thromb. Haemost.* **96,** 822–829 (2006).
35. Buyue, Y., Sheehan, J. P. **Fucosylated chondroitin sulfate inhibits plasma thrombin generation via targeting of the factor IXa heparin-binding exosite.** *Blood* **114,** 3092–3100 (2009).

36. Fonseca, R. J. C., Santos, G. R. C., Mourão, P. A. S.: **Effects of polysaccharides enriched in 2,4-disulfated fucose units on coagulation, thrombosis and bleeding. Practical and conceptual implications.** *Thromb. Haemost.* **102,** 829–836 (2009).
37. Fonseca, R. J. C., Oliveira, S. N., Pomin, V. H., Mecawi, A. S., Araujo, I. G., Mourão, P. A.: **Effects of oversulfated and fucosylated chondroitin sulfates on coagulation. Challenges for the study of anticoagulant polysaccharides.** *Thromb. Haemost.* **103,** 994–1004 (2010).
38. Chen, S., Li, G., Wu, N., Guo, X., Liao, N., Ye, X., Liu, D., Xue, C., Chai, W.: **Sulfation pattern of the fucose branch is important for the anticoagulant and antithrombotic activities of fucosylated chondroitin sulfates.** *Biochim. Biophys. Acta* **1830,** 3054–3066 (2013).
39. Liu, X. Hao, J., Shan, X., Zhang, X., Zhao, X., Li, Q., Wang, X., Cai, C., Li, G., Yu, G.: **Antithrombotic activities of fucosylated chondroitin sulfates and their depolymerized fragments from two sea cucumbers.** *Carbohydr. Polym.* **152,** 343–350 (2016).
40. Tapon-Bretaudière, J., Chabut, D., Zierer, M., Matou, S., Helley, D., Bros, A., Mourão, P. A., Fischer, A. M.: **A fucosylated chondroitin sulfate from echinoderm modulates in vitro fibroblast growth factor 2-dependent angiogenesis.** *Mol. Cancer Res.* **1,** 96–102 (2002).
41. Tian, F. Zhang, X., Tong, Y., Yi, Y., Zhang, S., Li, L., Sun, P., Lin, L., Ding, J., **PE, a new sulfated saponin from sea cucumber, exhibits anti-angiogenic and anti-tumor activities in vitro and in vivo.** *Cancer Biol. Ther.* **4,** 874–882 (2005).
42. Borsig, L., Wang, L., Cavalcante, M. C. M., Cardilo-Reis, L., Ferreira, P. L., Mourão, P. A. S., Esko J. D., Pavão.: **Selectin Blocking Activity of a Fucosylated Chondroitin Sulfate Glycosaminoglycan from Sea Cucumber. Effect on tumour metastasis and neutrophil recruitment.** *J. Biol. Chem.* **282,** 14984–14991 (2007).
43. Zhao, Y., Daohai, Z., Wang, S., Tao, L., Wang, A., Chen, W., Zhu, Z., Zheng, S., Gao, X., Lu, Y.: **Holothurian Glycosaminoglycan Inhibits Metastasis and Thrombosis via Targeting of Nuclear Factor-κB/Tissue Factor/Factor Xa Pathway in Melanoma B16F10 Cells.** *PLoS ONE,* **8**(2), e56557 (2013).
44. Yue, Z., Wang, A., Zhu, Z., Tao, L., Li, Y., Zhou, L., Chen, W., Lu, Y.: **Holothurian glycosaminoglycan inhibits metastasis via inhibition of P-selectin in B16F10 melanoma cells.** *Mol. Cell. Biochem.* **410,** 143–154 (2015).
45. Beutler, J. A., McKee, T. C., Fuller, R. W., Tischler, M., Cardellina II, J. H., Snader, K. M., McCloud, T. G., Boyd, M. R.: **Frequent Occurrence of HIV-Inhibitory Sulphated Polysaccharides in Marine Invertebrates.** *Antivir. Chem. Chemother.* **4,** 167–172 (1993).
46. Lian, W., Wu, M., Huang, N., Gao, N., Xiao, C., Li, Z., Zhang, Z., Zheng, Y., Peng, W., Zhao, J.: **Anti-HIV-1 activity and structure-activity-relationship study of a fucosylated glycosaminoglycan from an echinoderm by targeting the conserved CD4 induced epitope.** *Biochim. Biophys. Acta* **1830,** 4681–4691 (2013).
47. McClure, M. O., Moore, J. P., Blanc, D. F., Scotting, P., Cook, G. M., Keynes, R. J., Weber, J. N., Davies, D., Weiss, R. A.: **Investigations into the mechanism by which sulfated polysaccharides inhibit HIV infection in vitro.** *AIDS Res. Hum. Retroviruses* **8,** 19–26 (1992).
48. Huang, N., Wu, M. Y., Zheng, C. B., Zhu, L., Zhao, J. H., Zheng, Y. T.: **The depolymerized fucosylated chondroitin sulfate from sea cucumber potently inhibits HIV replication via interfering with virus entry.** *Carbohydr. Res.* **380,** 64–69 (2013).
49. Pangestuti, R., Arifin, Z.: **Medicinal and health benefit effects of functional sea cucumbers.** *Journal of Traditional and Complementary Medicince*. 1-11 (2017).
50. Pomin, V. H.: **Holothurian Fucosylated Chondroitin Sulfate.** *Mar. Drugs* **12,** 232–254 (2014).
51. Vieira, R. P., Mourão, P. A.: **Occurrence of a unique fucose-branched chondroitin sulfate in the body wall of a sea cucumber.** *J. Biol. Chem.* **263,** 18176–18183 (1988).

52. Chen, S., Changhu, X., Li'ang, Y., Qingjuan, T., Guangli, Y., Wengang, C.: **Comparison of structures and anticoagulant activities of fucosylated chondroitin sulfates from different sea cucumbers.** *Carbohydr. Polym.* **83,** 688–696 (2011).
53. Wu, M., Huang, R., Wen, D., Gao, N., He, J., Li, Z., Zhao, J.: **Structure and effect of sulfated fucose branches on anticoagulant activity of the fucosylated chondroitin sulfate from sea cucumber Thelenata ananas.** *Carbohydr. Polym.* **87,** 862–868 (2012).
54. Luo, L. Wu, M., Xu, W., Lian, W., Xiang, J., Lu, F., Gao, N., Xiao, C., Wang, S., Zhao, J.: **Comparison of Physicochemical Characteristics and Anticoagulant Activities of Polysaccharides from Three Sea Cucumbers.** *Mar. Drugs* **11,** 399–417 (2013).
55. Panagos, C., Thomson, D. S., Moss, C., Hughes, A. D., Kelly, M. S., Liu, Y., Chai, W., Venkatasamy, R., Spina, D., Page, C. P., Hogwood, J., Woods, R. J., Mulloy, B., Bavington, C. D., Uhrín, D.: **Fucosylated chondroitin sulfates from the body wall of the sea cucumber Holothuria forskali. Conformation, selectin binding and biological activity.** *J. Biol. Chem.* **41**, 28284-98 (2014).
56. Ruggiero, J., Vieira, R. P., Mourão, P. A.: **Increased calcium affinity of a fucosylated chondroitin sulfate from sea cucumber.** *Carbohydr. Res.* **256,** 275–287 (1994).
57. Franette, P. S., Wagner, D. D.: **Insights into selectin function from knockout mice.** *Thromb. Haemost.* **78,** 60–64 (1997).
58. Steeber, D. A., Subramanian, H., Grailer, J. J., Conway, R. M., Storey, T. J.: **L-selectin-mediated leukocyte adhesion and migration.** in *Adhesion Molecules: Function and Inhibition* (ed. Ley, K.) 27–70 (Birkhäuser Basel, 2007).
59. Chen, A., Engel, P., Tedder, T. F.: **Structural requirements regulate endoproteolytic release of the L- selectin (CD62L) adhesion receptor from the cell surface of leukocytes.** *J. Exp. Med.* **182,** 519–530 (1995).
60. Bevilacqua, M. P., Stengelin, S., Gimbrone, M. A., Seed, B.: **Endothelial leukocyte adhesion molecule 1: an inducible receptor for neutrophils related to complement regulatory proteins and lectins.** *Science* **243,** 1160–1165 (1989).
61. Kansas, G. S.: **Selectins and their ligands: current concepts and controversies.** *Blood* **88,** 3259–3287 (1996).
62. Chase, S. D., Magnani, J. L., Simon, S. I.: **E-selectin ligands as mechanosensitive receptors on neutrophils in health and disease.** *Ann. Biomed. Eng.* **40,** 849–859 (2012).
63. Katayama, Y., Hidalgo, A., Furie, B. C., Vestweber, D., Furie, B., Frenette, P. S.: **PSGL-1 participates in E-selectin-mediated progenitor homing to bone marrow: evidence for cooperation between E-selectin ligands and alpha4 integrin.** *Blood* **102,** 2060–2067 (2003).
64. Tedder, T. F., Steeber, D. A., Chen, A., Engel, P.: **The selectins: vascular adhesion molecules.** *FASEB J.* **9,** 866–873 (1995).
65. Graves, B. J., Crowther, R. L., Chandran, C., Rumberger, J. M., Li, S., Huang, K. S., Presky, D. H., Familletti, P. C., Wolitzky, B. A., Burns, D. K.: **Insight into E-selectin/ligand interaction from the crystal structure and mutagenesis of the lec/EGF domains.** *Nature* **367,** 532–538 (1994).
66. Somers, W. S., Tang, J., Shaw, G. D., Camphausen, R. T.: **Insights into the Molecular Basis of Leukocyte Tethering and Rolling Revealed by Structures of P- and E-Selectin Bound to SLeX and PSGL-1.** *Cell* **103,** 467–479 (2000).
67. Preston, R. C., Jakob, R. P., Binder, F. P. C., Sager, C. P., Ernst, B., Maier, T.: **E-selectin ligand complexes adopt an extended high-affinity conformation.** *J. Mol. Cell Biol.* **8,** 62–72 (2016).
68. Freedman, S. J., Sanford, D. G., Bachovchin, W. W., Furie, B. C., Baleja, J. D., Furie, B.: **Structure and function of the epidermal growth factor domain of P-selectin.** *Biochemistry* **35,** 13733–13744 (1996).
69. Mehta-D'souza, P., Klopocki, A. G., Oganesyan, V., Terzyan, S., Mather, T., Li, Z., Panicker, S. R., Zhu, C., McEver, R. P.: **Glycan Bound to the Selectin Low Affinity State Engages**

Glu-88 to Stabilize the High Affinity State under Force. *J. Biol. Chem.* **292,** 2510–2518 (2017).
70. Weis, W. I., Taylor, M. E., Drickamer, K.: **The C-type lectin superfamily in the immune system.** *Immunol. Rev.* **163,** 19–34 (1998).
71. Zelensky, A. N., Gready, J. E.: **The C-type lectin-like domain superfamily.** *FEBS J.* **272,** 6179–6217 (2005).
72. Cummings, R. D., McEver, R. P.: **C-type Lectins.** in *Essentials of Glycobiology* (eds. Varki, A. et al.) (Cold Spring Harbor Laboratory Press, 2009).
73. Lou, J., Yago, T., Klopocki A. G., Mehta, P., Chen, W, Zarnitsyna, V. I., Bovin, N. V., Zhu, C., McEver, R. P.: **Flow-enhanced adhesion regulated by a selectin interdomain hinge.** *J. Cell Biol.* **174,** 1107–1117 (2006).
74. Springer, T. A.: **Structural basis for selectin mechanochemistry.** *Proc. Natl. Acad. Sci.* **106,** 91–96 (2009).
75. Wouters, M. A. Rigoutsos, I, Chu, C. K., Feng, L. L., Sparrow, D. B., Dunwoodie, S. L.: **Evolution of distinct EGF domains with specific functions.** *Protein Sci. Publ. Protein Soc.* **14,** 1091–1103 (2005).
76. Rodgers, S. D., Camphausen, R. T., Hammer, D. A.: **Tyrosine Sulfation Enhances but Is Not Required for PSGL-1 Rolling Adhesion on P-Selectin.** *Biophys. J.* **81,** 2001–2009 (2001).
77. Poppe, L., Brown, G. S., Philo, J. S., Nikrad, P. V., Shah, B. H.: **Conformation of sLex Tetrasaccharide, Free in Solution and Bound to E-, P-, and L-Selectin**. *J. Am. Chem. Soc.* **119,** 1727–1736 (1997).
78. Wilkins, P. P., Moore, K. L., McEver, R. P., Cummings, R. D.: **Tyrosine Sulfation of P-selectin Glycoprotein Ligand-1 Is Required for High Affinity Binding to P-selectin.** *J. Biol. Chem.* **270,** 22677–22680 (1995).
79. Pouyani, T., Seed, B.: **PSGL-1 recognition of P-selectin is controlled by a tyrosine sulfation consensus at the PSGL-1 amino terminus.** *Cell* **83,** 333–343 (1995).
80. Sako, D., Comess, K. M., Barone, K. M., Camphausen, R. T., Cumming, D. A., Shaw, G. D.: **A sulfated peptide segment at the amino terminus of PSGL-1 is critical for P-selectin binding.** *Cell* **83,** 323–331 (1995).
81. Goetz, D. J., Greif, D. M., Ding, H., Camphausen, R. T., Howes, S., Comess, K. M., Snapp, K. R., Kansas, G. S., Luscinskas, F. W.: **Isolated P-selectin Glycoprotein Ligand-1 Dynamic Adhesion to P- and E-selectin.** *J. Cell Biol.* **137,** 509–519 (1997).
82. McEver, R. P., Zhu, C.: **Rolling cell adhesion.** *Annu. Rev. Cell Dev. Biol.* **26,** 363–396 (2010).
83. Hemmerich, S., Leffler, H., Rosen, S. D.: **Structure of the O-Glycans in GlyCAM-1, an Endothelial-derived Ligand for L-selectin.** *J. Biol. Chem.* **270,** 12035–12047 (1995).
84. Hidalgo, A., Peired, A. J., Wild, M. K., Vestweber, D., Frenette, P. S.: **Complete identification of E-selectin ligands on neutrophils reveals distinct functions of PSGL-1, ESL-1, and CD44.** *Immunity* **26,** 477–489 (2007).
85. Zöllner, O., Lenter M. C., Blanks, J. E., Borges, E., Steegmaier, M., Zerwes, H.-G., Vestweber, D.: **L-Selectin from Human, but Not from Mouse Neutrophils Binds Directly to E-Selectin.** *J. Cell Biol.* **136,** 707–716 (1997).
86. Sreeramkumar, V., Leiva, M., Stadtmann, A., Pitaval, C., Ortega-Rodríguez, I., Wild, M. K., Lee, B., Zarbock, A., Hidalgo, A.: **Coordinated and unique functions of the E-selectin ligand ESL-1 during inflammatory and hematopoietic recruitment in mice.** *Blood* **122,** 3993–4001 (2013).
87. Alon, R., Hammer, D. A., Springer, T. A.: **Lifetime of the P-selectin-carbohydrate bond and its response to tensile force in hydrodynamic flow.** *Nature* **374,** 539–542 (1995).
88. Finger, E. B., Puri, K. D., Alon, R., Lawrence, M. B., von Andrian, U. H., Springer, T. A.: **Adhesion through L-selectin requires a threshold hydrodynamic shear.** *Nature* **379,** 266–269 (1996).

89. Lawrence, M. B., Kansas, G. S., Kunkel, E. J., Ley, K.: **Threshold Levels of Fluid Shear Promote Leukocyte Adhesion through Selectins (CD62L,P,E).** *J. Cell Biol.* **136,** 717–727 (1997).
90. Thomas, W. E., Trintchina, E., Forero, M., Vogel, V., Sokurenko, E. V.: **Bacterial adhesion to target cells enhanced by shear force.** *Cell* **109,** 913–923 (2002).
91. Doggett, T. A., Girdhar, G., Lawshé, A., Schmidtke, D. W., Laurenzi, I. J., Diamond, S. L., Diacovo, T. G.: **Selectin-like kinetics and biomechanics promote rapid platelet adhesion in flow: the GPIb(alpha)-vWF tether bond.** *Biophys. J.* **83,** 194–205 (2002).
92. Marshall, B. T., Long, M., Piper, J. W., Yago, T., McEver, R. P., Zhu, C.: **Direct observation of catch bonds involving cell-adhesion molecules.** *Nature* **423,** 190–193 (2003).
93. Dwir, O., Kansas, G. S., Alon, R.: **An activated L-selectin mutant with conserved equilibrium binding properties but enhanced ligand recognition under shear flow.** *J. Biol. Chem.* **275,** 18682–18691 (2000).
94. Phan, U. T., Waldron, T. T., Springer, T. A.: **Remodeling of the lectin–EGF-like domain interface in P- and L-selectin increases adhesiveness and shear resistance under hydrodynamic force.** *Nat. Immunol.* **7,** 883–889 (2006).
95. Lou, J., Zhu, C.: **A Structure-Based Sliding-Rebinding Mechanism for Catch Bonds.** *Biophys. J.* **92,** 1471–1485 (2007).
96. Waldron, T. T., Springer, T. A.: **Transmission of allostery through the lectin domain in selectin-mediated cell adhesion.** *Proc. Natl. Acad. Sci.* **106,** 85–90 (2009).
97. Dwir, O., Solomon, A., Mangan, S., Kansas, G. S., Schwartz, U. S., Alon, R.: **Avidity enhancement of L-selectin bonds by flow.** *J. Cell Biol.* **163,** 649–659 (2003).
98. Natoni, A., Macauley, M. S., O'Dwyer, M. E.: **Targeting Selectins and Their Ligands in Cancer.** *Front. Oncol.* **6,** (2016).
99. Mannori, G., Crottet, P., Cecconi, O., Hanasaki, K., Aruffo, A., Nelson, R. M., Varki, A., Bevilacqua, M. P.: **Differential colon cancer cell adhesion to E-, P-, and L-selectin: role of mucin-type glycoproteins.** *Cancer Res.* **55,** 4425–4431 (1995).
100. Kim, Y. J., Borsig, L., Han, H.-L., Varki, N. M., Varki, A.: **Distinct Selectin Ligands on Colon Carcinoma Mucins Can Mediate Pathological Interactions among Platelets,** Leukocytes, and Endothelium. *Am. J. Pathol.* **155,** 461–472 (1999).
101. Borsig, L., Wong, R., Hynes, R. O., Varki, N. M., Varki, A.: **Synergistic effects of L- and P-selectin in facilitating tumor metastasis can involve non-mucin ligands and implicate leukocytes as enhancers of metastasis.** *Proc. Natl. Acad. Sci. U. S. A.* **99,** 2193–2198 (2002).
102. Köhler, S., Ullrich, S., Richter, U., Schumacher, U.: **E-/P-selectins and colon carcinoma metastasis: first in vivo evidence for their crucial role in a clinically relevant model of spontaneous metastasis formation in the lung.** *Br. J. Cancer* **102,** 602–609 (2010).
103. Läubli, H., Borsig, L.: **Selectins as mediators of lung metastasis.** *Cancer Microenviron. Off. J. Int. Cancer Microenviron. Soc.* **3,** 97–105 (2010).
104. Bendas, G., Borsig, L.: **Cancer Cell Adhesion and Metastasis: Selectins, Integrins, and the Inhibitory Potential of Heparins.** *Int. J. Cell Biol.* **2012,** e676731 (2012).
105. Borsig, L., Vlodavsky, I., Ishai-Michaeli, R., Torri, G., Vismara, E.: **Sulfated Hexasaccharides Attenuate Metastasis by Inhibition of P-selectin and Heparanase.** *Neoplasia N. Y. N* **13,** 445–452 (2011).
106. Perez, S., MouhousRiou, N., Nifantev, N. E., Tsvetkov, Y. E., Bachet, B., and Imberty, A.: **Crystal and molecular structure of a histo-blood group antigen involved in cell adhesion: The Lewis x trisaccharide**. *Glycobiology* **6**, 537-542 (1996).
107. Poppe, L., Brown, G. S., Philo, J. S., Nikrad, P. V., and Shah, B. H. **Conformation of sLe(x) tetrasaccharide, free in solution and bound to E-, P-, and L-selectin**. *J. Am. Chem. Soc.* **119**, 1727-1736 (1997)

108. Wedepohl, S. Kaup, M., Riese, S. B., Berger, M., Dernedde, J., Tauber, R., Blanchard, V.: **N-Glycan Analysis of Recombinant L-Selectin Reveals Sulfated GalNAc and GalNAc–GalNAc Motifs.** *J. Proteome Res.* **9,** 3403–3411 (2010).
109. de Marco, A.: **Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli.** *Microb. Cell Factories* **8,** 26 (2009).
110. Cereghino, J. L.. Cregg, J. M.: **Heterologous protein expression in the methylotrophic yeast Pichia pastoris.** *FEMS Microbiol. Rev.* **24,** 45–66 (2000).
111. Daly, R., Hearn, M. T. W.: **Expression of heterologous proteins in Pichia pastoris: a useful experimental tool in protein engineering and production.** *J. Mol. Recognit.* **18,** 119–138 (2005).
112. Yao, Y., Wang, J., Viroonchatapan, N., Samson, A., Chill, J., Rothe, E., Anglister, J., Zuo-Zhong, W.: **Yeast Expression and NMR Analysis of the Extracellular Domain of Muscle Nicotinic Acetylcholine Receptor α Subunit.** *J. Biol. Chem.* **277,** 12613–12621 (2002).
113. Huang, Y., Zhang, Y., Wu, Y., Wang, J., Liu, X., Dai, L., Wang, L., Yu, M., Mo, W.: **Expression, Purification, and Mass Spectrometric Analysis of 15N, 13C-Labeled RGD-Hirudin, Expressed in Pichia pastoris, for NMR Studies.** *PLOS ONE* **7,** e42207 (2012).
114. Cregg, J. M., Cereghino, J. L., Shi, J., Higgins, D. R.: **Recombinant protein expression in Pichia pastoris.** *Mol. Biotechnol.* **16,** 23–52 (2000).
115. Ang, R. P., Teoh, L. S., Chan, M. K., Miswan, N., Khoo, B. Y.: **Comparing the expression of human DNA topoisomerase I in KM71H and X33 strains of Pichia pastoris.** *Electron. J. Biotechnol.* **21,** 9–17 (2016).
116. Krainer, F. W, Dietzsch, C., Hajek, T., Herwig, ., Spadiut, O., Glieder, A.: **Recombinant protein expression in *Pichia pastoris* strains with an engineered methanol utilization pathway.** *Microbial Cell Factories*. **11**:22 (2012)
117. Cregg, J. M.: **Expression in the methylotropic yeast Pichia pastoris.** in *Gene Expression Systems* 157–191 (Academic Press, 1999).
118. Rosano, G. L., Ceccarelli, E. A.: R**ecombinant protein expression in Escherichia coli: advances and challenges.** *Front. Microbiol.* **5,** (2014).
119. Costa, S., Almeida, A., Castro, A., Domingues, L.: **Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system.** *Front. Microbiol.* **5,** (2014).
120. Sørensen, H. P., Mortensen, K. K.: **Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli.** *Microb. Cell Factories* **4,** 1 (2005).
121. Yamaguchi, S., Yamamoto, E., Mannen, T., Nagamune, T., Nagamune, T.: **Protein refolding using chemical refolding additives.** *Biotechnol. J.* **8,** 17–31 (2013).
122. Middelberg, A. P. J.: **Preparative protein refolding.** *Trends Biotechnol.* **20,** 437–443 (2002).
123. Okumura, M., Saiki, M., Yamaguchi, H., Hidaka, Y.: **Acceleration of disulfide-coupled protein folding using glutathione derivatives.** *FEBS J.* **278,** 1137–1144 (2011).
124. Thomson, C. A., Olson, M., Jackson, L. M., Schrader, J. W.: **A Simplified Method for the Efficient Refolding and Purification of Recombinant Human GM-CSF.** *PLOS ONE* **7,** e49891 (2012).
125. Ghosh, R., Sharma, S., Chattopadhyay, K.: **Effect of Arginine on Protein Aggregation Studied by Fluorescence Correlation Spectroscopy and Other Biophysical Methods.** *Biochemistry (Mosc.)* **48,** 1135–1143 (2009).
126. Shukla, D., Trout, B. L.: **Interaction of Arginine with Proteins and the Mechanism by Which It Inhibits Aggregation.** *J. Phys. Chem. B* **114,** 13426–13438 (2010).
127. Lustbader, J. W., Birken, S., Pollak, S., Pound, A., Chait, B. T., Mirza, U. A., Ramnarain, S., Canfield, R. E., Brown, J. M.: **Expression of human chorionic gonadotropin uniformly labeled with NMR isotopes in Chinese hamster ovary cells: an advance toward rapid determination of glycoprotein structures.** *J. Biomol. NMR* **7,** 295–304 (1996).
128. Werner, K., Richter, C., Klein-Seetharaman, J., Schwalbe, H.: **Isotope labeling of mammalian GPCRs in HEK293 cells and characterization of the C-terminus of**

bovine rhodopsin by high resolution liquid NMR spectroscopy. *J. Biomol. NMR* **40,** 49–53 (2008).
129. Dutta, A., Saxena, K., Schwalbe, H., Klein-Seetharaman, J.: **Isotope Labeling in Mammalian Cells.** *Methods Mol. Biol.* **831,** 55–69 (2012).
130. Takahashi, H., Shimada, I.: **Production of isotopically labeled heterologous proteins in non-E. coli prokaryotic and eukaryotic cells.** *J. Biomol. NMR* **46,** 3 (2010).
131. Graham, F. L., Smiley, J., Russell, W. C., Nairn, R.: **Characteristics of a human cell line transformed by DNA from human adenovirus type 5.** *J. Gen. Virol.* **36,** 59–74 (1977).
132. Thomas, P., Smart, T. G.: **HEK293 cell line: A vehicle for the expression of recombinant proteins.** *J. Pharmacol. Toxicol. Methods* **51,** 187–200 (2005).
133. Stepanenko, A. A., Dmitrenko, V. V.: **HEK293 in cell biology and cancer research: phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution.** *Gene* **569,** 182–190 (2015).
134. Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D. I., Owens, R. J.: **A versatile ligation-independent cloning method suitable for high-throughput expression screening applications.** *Nucleic Acids Res.* **35,** e45 (2007).
135. Nettleship, J. E., Rahman-Huq, N., Owens, R. J.: **The production of glycoproteins by transient expression in Mammalian cells.** *Methods Mol. Biol.* **498,** 245–263 (2009).
136. Roth, Z., Yehezkel, G., Khalaila, I.: **Identification and Quantification of Protein Glycosylation.** *Int. J. Carbohydr. Chem.* **2012,** e640923 (2012).
137. Picker, L. J., Warnock, R. A., Burns, A. R., Doerschuk, Berg, E. L., Butcher, E. C.: **The neutrophil selectin LECAM-1 presents carbohydrate ligands to the vascular selectins ELAM-1 and GMP-140.** *Cell* **66,** 921–933 (1991).
138. Mayer, M., Meyer, B.: **Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy.** *Angew. Chem. Int. Ed.* **38,** 1784–1788 (1999).
139. Meyer, B., Peters, T.: **NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors.** *Angew. Chem. Int. Ed Engl.* **42,** 864–890 (2003).
140. Haselhorst, T., Lamerz, A.-C., Itzstein, M. von.: **Saturation transfer difference NMR spectroscopy as a technique to investigate protein-carbohydrate interactions in solution.** *Methods Mol. Biol. Clifton NJ* **534,** 375–386 (2009).
141. Mayer, M., Meyer, B.: **Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy.** *Angew. Chem. Int. Ed.* **38** (12), 1784–1787 (1999).
142. Mayer, M., Meyer, B.: **Group Epitope Mapping by Saturation Transfer Difference NMR To Identify Segments of a Ligand in Direct Contact with a Protein Receptor.** *J. Am. Chem. Soc.* **123**, 6108–6117 (2001).
143. Künze, G., Gehrcke, J.-P., Pisabarro, M. T., Huster, D.: **NMR characterization of the binding properties and conformation of glycosaminoglycans interacting with interleukin-10.** *Glycobiology* **24,** 1036–1049 (2014).
144. Herfurth, L., Ernst, B., Wagner, B., Ricklin, D., Strasser, D. S., Magnani, J. L., Benie, A., J., Peters, T.: **Comparative Epitope Mapping with Saturation Transfer Difference NMR of Sialyl Lewisa Compounds and Derivatives Bound to a Monoclonal Antibody.** *J. Med. Chem.* **48,** 6879–6886 (2005).
145. Haselhorst, T., Blanchard, H., Frank M., Kraschnefski M. J., Kiefel, M. J., Szyczew, A., J., Dyason, J. C., Fleming, F., Holloway, G., Coulson, B., S., von Itzstein, M.: **STD NMR spectroscopy and molecular modelling investigation of the binding of N-acetylneuraminic acid derivatives to rhesus rotavirus VP8* core.** *Glycobiology* **17** (1), 68-81 (2006).
146. Alonso, H., Bliznyuk, A. A., Gready, J. E.: **Combining docking and molecular dynamic simulations in drug design.** *Med. Res. Rev.* **26,** 531–568 (2006).
147. Trott, O., Olson, A. J.: **AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading.** *J. Comput. Chem.* **31,** 455–461 (2010).
148. Sapay, N., Nurisso, A., Imberty, A.: **Simulation of carbohydrates, from molecular docking to dynamics in water.** *Methods Mol. Biol. Clifton NJ* **924,** 469–483 (2013).

149. Nivedha, A. K., Makeneni, S., Foley, B. L., Tessier, M. B., Woods, R. J.: **Importance of ligand conformational energies in carbohydrate docking: Sorting the wheat from the chaff.** *J. Comput. Chem.* **35,** 526–539 (2014).
150. Bitomsky, W., Wade, R. C.: **Docking of Glycosaminoglycans to Heparin-Binding Proteins: Validation for aFGF, bFGF, and Antithrombin and Application to IL-8.** *J. Am. Chem. Soc.* **121,** 3004–3013 (1999).
151. Samsonov, S. A., Teyra, J., Pisabarro, M. T.: **Docking glycosaminoglycans to proteins: analysis of solvent inclusion.** *J. Comput. Aided Mol. Des.* **25,** 477–489 (2011).
152. Samsonov, S. A., Gehrcke, J.-P., Pisabarro, M. T.: **Flexibility and Explicit Solvent in Molecular-Dynamics-Based Docking of Protein–Glycosaminoglycan Systems.** *J. Chem. Inf. Model.* **54,** 582–592 (2014).
153. Takaoka, T., Mori, K., Okimoto, N., Neya, S., Hoshino, T.: **Prediction of the Structure of Complexes Comprised of Proteins and Glycosaminoglycans Using Docking Simulation and Cluster Analysis.** *J. Chem. Theory Comput.* **3,** 2347–2356 (2007).
154. Samsonov, S. A., Pisabarro, M. T.: **Computational analysis of interactions in structurally available protein–glycosaminoglycan complexes.** *Glycobiology* **26,** 850–861 (2016).
155. Cosconati, S., Forli, S., Perryman, A. L., Harris, R., Goodsell, D. S., Olson, A. J.: **Virtual Screening with AutoDock: Theory and Practice.** *Expert Opin. Drug Discov.* **5,** 597–607 (2010).
156. Forli, S., Huey, R., Pigue, M. E., Sanner, M. F., Goodsell, D., S., Olson, A. J.: **Computational protein-ligand docking and virtual drug screening with the AutoDock suite.** *Nat. Protoc.* **11,** 905–919 (2016).
157. Meng, X.-Y., Zhang, H.-X., Mezei, M., Cui, M.: **Molecular Docking: A powerful approach for structure-based drug discovery.** *Curr. Comput. Aided Drug Des.* **7,** 146–157 (2011).
158. Salomon-Ferrer, R., Case, D. A., Walker, R. C.: **An overview of the Amber biomolecular simulation package.** *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **3,** 198–210 (2013).
159. Case, D. A., Babin, V., Berryman, J. T., Betz, R. M., Cai, Q., Cerutti, D. S., Cheatham, T. E., Darden, T. A., Duke, R. E., Gohlke, H., Goetz, A. W., Gusarov, S., Homeyer, N., Janowski, P., Kaus, J., Kolossváry, I., Kovalenko, A., Lee, T. S., LeGrand, S., Luchko, T., Luo, R., Madej, B., Merz, K. M., Paesani, F., Roe, D. R., Roitberg, A., Sagui, C., Salomon-Ferrer, R., Seabra, G., Simmerling, C. L., Smith, W., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X., Kollman, P. A.: **AMBER 14**, University of California, San Francisco (2014) Available at: http://ambermd.org/.
160. Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., Simmerling, C.: **ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB.** *J. Chem. Theory Comput.* **11,** 3696–3713 (2015).
161. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., Woods, R. J.: **GLYCAM06: A generalizable biomolecular force field. Carbohydrates.** *J. Comput. Chem.* **29,** 622–655 (2008).
162. Fadda, E., Woods, R. J.: **Molecular simulations of carbohydrates and protein-carbohydrate interactions: motivation, issues and prospects.** *Drug Discov. Today* **15,** 596–609 (2010).
163. Diehl, C., Genheden, S., Modig, K., Ryde, U., Akke, M.: **Conformational entropy changes upon lactose binding to the carbohydrate recognition domain of galectin-3.** *J. Biomol. NMR* **45,** 157–169 (2009).
164. Newhouse, E. I., Xu, D., Markwick, P. R. L., Amaro, R. E., Pao, H. C., Wu, K. J., Alam, M., McCammon, J. A., Li., W. W.: **Mechanism of Glycan Receptor Recognition and Specificity Switch for Avian, Swine, and Human Adapted Influenza Virus Hemagglutinins: A Molecular Dynamics Perspective.** *J. Am. Chem. Soc.* **131,** 17430–17442 (2009).
165. Götz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., Walker, R. C.: **Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born.** *J. Chem. Theory Comput.* **8,** 1542–1555 (2012).

166. Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S., Walker, R. C.: **Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald.** *J. Chem. Theory Comput.* **9,** 3878–3888 (2013).
167. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., Ferrin, T. E.: **UCSF Chimera--a visualization system for exploratory research and analysis.** *J. Comput. Chem.* **25,** 1605–1612 (2004).
168. Zhan, D., Yu, L., Jin, H., Guan, S., Han, W.: **Molecular modeling and MM-PBSA free energy analysis of endo-1,4-β-xylanase from Ruminococcus albus 8.** *Int. J. Mol. Sci.* **15,** 17284–17303 (2014).
169. Genheden, S., Ryde, U.: **The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities.** *Expert Opin. Drug Discov.* **10,** 449–461 (2015).
170. Yuriev, E., Farrugia, W., Scott, A. M., Ramsland, P. A.: **Three-dimensional structures of carbohydrate determinants of Lewis system antigens: Implications for effective antibody targeting of cancer.** *Immunol. Cell Biol.* **83,** 709–717 (2005).
171. Miller, K. E., Mukhopadhyay, C., Cagas, P., Bush, C. A.: **Solution structure of the Lewis x oligosaccharide determined by NMR spectroscopy and molecular dynamics simulations.** *Biochemistry (Mosc.)* **31,** 6703–6709 (1992).
172. Azurmendi, H. F., Martin-Pastor, M., Bush, C. A.: **Conformational studies of Lewis X and Lewis A trisaccharides using NMR residual dipolar couplings.** *Biopolymers* **63,** 89–98 (2002).
173. Hou, T., Wang, J., Li, Y., Wang, W.: **Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations.** *J. Chem. Inf. Model.* **51,** 69–82 (2011).
174. Cherney, M. M., Lecaille, F., Kienitz, M., Nallaseth, F. S., Li, Z., James, M. N., Brömme, D.: **Structure-activity analysis of cathepsin K/chondroitin 4-sulfate interactions.** *J. Biol. Chem.* **286,** 8988–8998 (2011).
175. Bubb, W. A.: **NMR spectroscopy in study of carbohydrates: Characterizing the structural complexity.** *Concepts Magn. Reson Part A.* **19A (1),** 1-19 (2003).
176. Duus, J. O., Gotfredsen., C. H., Bock, C.: **Carbohydrate structural determination by NMR spectroscopy: Modern methods and limitations.** *Chem. Rev.* **100,** 4589-4614 (2000).
177. Panagos, C., Thomson, D., Bavington, C., Uhrín, D.: **Structural characterisation of oligosaccharides obtained by Fenton-type radical depolymerisation of dermatan sulfate.** *Carbohydr. Poly.* **87,** 2086–2092 (2012).
178. Sato, H., Fukae, K., Kajihara, Y.: **2D Selective-TOCSY-DQFCOSY and HSQC-TOCSY NMR experiments for assignment of a homogeneous asparagine-linked triantennary complex-type undecasaccharide.** *Carbohydr. Res.* **343,** 1333–1345 (2008).
179. de Beer, T., van Zuylen C. W. E. M., Hård, C., Boelens, R., Kaptein, R., Kamerling, J. P., Vliegenthart, J. F. G.: **Rapid and simple approach for the NMR resonance assignment of the carbohydrate chains of an intact glycoprotein.** *FEBS Letters.* **348,** 1–6 (1994).
180. Roslund, M. U., Tähtinen, P., Niemitz, M., Sjöholm, R.: **Complete assignments of the ^{1}H and ^{13}C chemical shifts and $J_{H,H}$ coupling constants in NMR spectra of D-glucopyranose and all D-glucopyranosyl-D-glucopyranosides.** *Carbohydr. Res.* **343,** 101–112 (2008).
181. von Halbeek, H.: **NMR developments in structural studies of carbohydrates and their complexes.** *Curr. Opin. Struct. Biol.* **4,** 697-709 (1994).
182. Rodríguez-Carvajal, M. A., du Penhoat, C. H., Mazeau, K., Doco, T., Pérez, S.: **The three-dimensional structure of the mega-oligosaccharide rhamnogalacturonan II monomer: a combined molecular modeling and NMR investigation.** *Carbohydr. Res.* **338,** 651-671 (2003).
183. Krishnamurthy, V. V.: **Sensitivity-enhanced 3D HSQC-TOCSY experiments.** *J. Magn. Reson Series B.* **106,** 170-177 (1995).
184. Seho, K., Szyperski, T..: **GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information.** *J. Am. Chem. Soc.* **125,** 1385-1393 (2003).

185. Szyperski, T.: **Principles and applications of projected multidimensional NMR spectroscopy - G-matrix Fourier Transform NMR.** In: Arrondo, J. L. R., Alonso, A. (eds) *Advanced Techniques in Biophysics.* Springer Series in Biophysics, **10,** Springer, Berlin, Heidelberg (2006).
186. Pudakalakatti, S. M., Dubey, A., Jaipuria, G., Shubhashree, U., Adiga, S. K., Moskau, D., Atreya, H. S.: **A fast NMR method for resonance assignments: application to metabolomics.** *J. Biomol. NMR.* **58,** 165-173 (2014).
187. Brodaczewska, N., Košťálová, Z., Uhrín, D.: **(3,2)D ^{1}H, ^{13} BIRD^{r,x}-HSQC-TOCSY for NMR structure elucidation of mixtures: application to complex carbohydrates.** *J. Biomol. NMR.* https://doi.org/10.1007/s10858-018-0163-8 (2018).
188. Rucker, S. P., Shaka, A. J.: **Broad-band homonuclear cross polarization in 2D NMR using DIPSI-2.** *Mol. Phys.* **68,** 509-517 (1989).
189. Kay, L. E., Keifer, P., Saarinen, T.: **Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity.** *J. Am. Chem. Soc.* **114,** 10663-10665 (1992).
190. Schleucher, J., Schwendinger, M., Sattler, M., Schmidt, P., Schedletzky, O., Glaser, S. J., Sorensen, O. W., Griesinger, C.: **A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed-field gradients..** *J. Biomol. NMR.* **4,** 301-306 (1994).
191. Palmer, A. G., Cavanagh, J., Wright, P. E., Rance, M.: **Sensitivity improvement in proton-detected 2-dimensional heteronuclear correlation NMR spectroscopy.** *J. Magn. Reson.* **93,** 151-170 (1991).
192. Garbow, J. R., Weitekamp, D. P., Pines, A..: **Bilinear rotation decoupling of homonuclear scalar interactions.** *Chem. Phys. Lett.* **93,** 504-509 (1982).
193. Uhrín, D., Brisson, J. R., Maclean, L. L., Richards, J. C., Perry, M. B.: **Application of 1D and 2D NMR techniques to the structure elucidation of the O-polysaccharide from *Proteus mirabilis* O-57.** *J. Biomol. NMR.* **4,** 615-630 (1994).
194. Aeschbacher, T., Zierke, M., Smiesko, M., Collot, M., Mallet, J. M., Ernst, B., Allain, F. H. T., Schubert, M.: **A secondary structural element in a wide range of fucosylated glycoepitopes.** *Chem-A Eur. J.* **23,** 11598-11610 (2017).
195. Dhanik, A., McMurray, J. S., Kavraki L. E.: **DINC: A new AutoDock-based protocol for docking large ligands.** *BMC Struct. Biol.* **13**, Suppl 1:S11 (2013).
196. Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Séverine, D., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., Stockinger, H.: **ExPASy: SIB bioinformatics resource portal.** *Nucleic Acids Res.* **40,** W597-603 (2012).

Appendix 1

Amino acid sequences of L and P-selectin constructs

LS1:

WTYHYSEKPMNWQRARRFCRDNYTDLVAIQNKAEIEYLEKTLPFSRSYYWIGIRKIGGIWTWVGT
NKSLTEEAENWGDGEPNNKKNKEDCVEIYIKRNKDAGKWNDDACHKLKAALCYTAS

LS2:

WTYHYSEKPMNWQRARRFCRDNYTDLVAIQNKAEIEYLEKTLPFSRSYYWIGIRKIGGIWTWVGT
NKSLTEEAENWGDGEPNNKKNKEDCVEIYIKRNKDAGKWNDDACHKLKAALCYTASCQPWSCS
GHGECVEIINNYTCNCDVGYYGPQCQFVIQ

PS1:

WTYHYSTKAYSWNISRKYCQNRYTDLVAIQNKNEIDYLNKVLPYYSSYYWIGIRKNNKTWTWVG
TKKALTNEAENWADNEPNNKRNNEDCVEIYIKSPSAPGKWNDEHCLKKKHALCYTAS

PS2:

WTYHYSTKAYSWNISRKYCQNRYTDLVAIQNKNEIDYLNKVLPYYSSYYWIGIRKNNKTWTWVG
TKKALTNEAENWADNEPNNKRNNEDCVEIYIKSPSAPGKWNDEHCLKKKHALCYTASCQDMSCS
KQGECLETIGNYTCSCYPGFYGPECEYVRE

Appendix 2

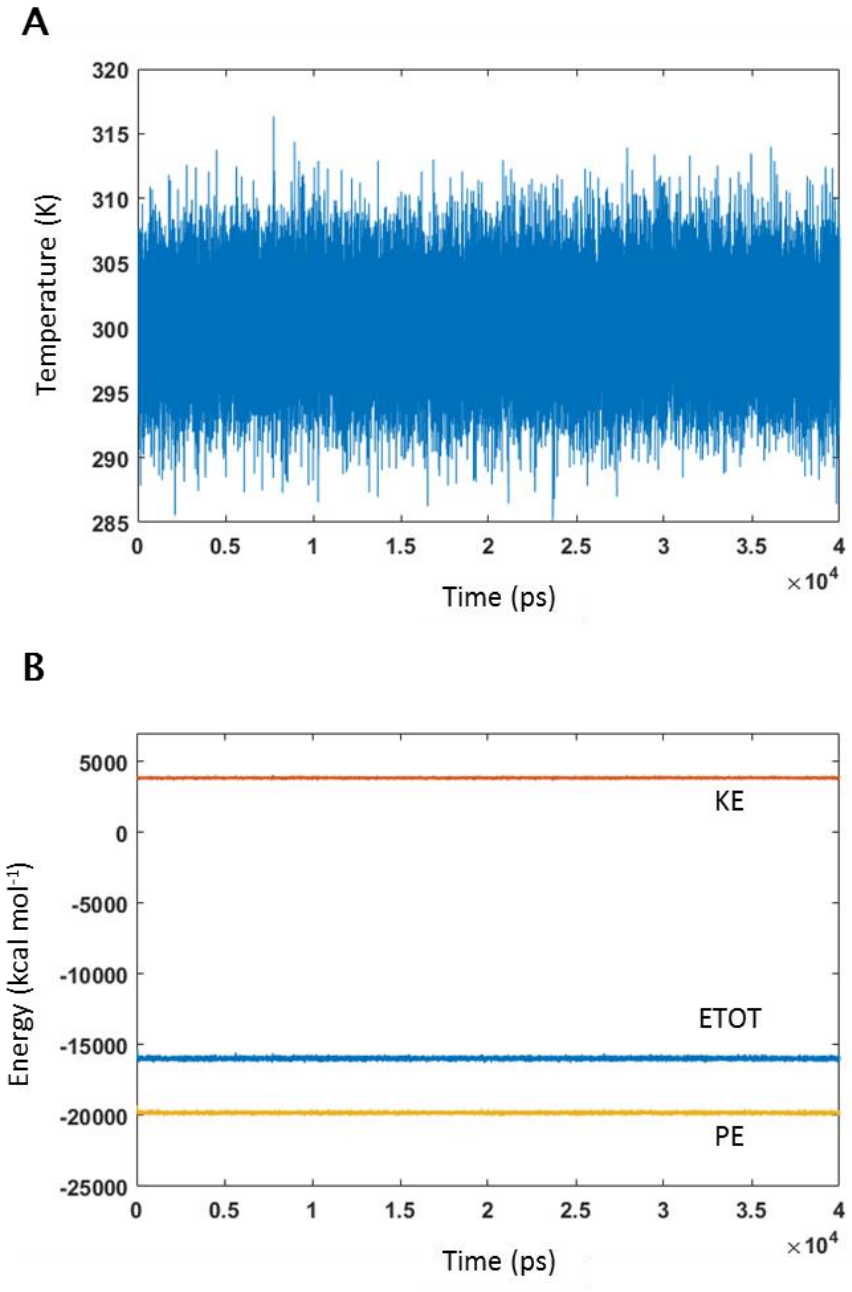
Calculated molecular weight (MW) and isoelectric point (pI) values for selectin constructs expressed in *P. pastoris* and *E. coli*

Host	Construct	MW (kDa)	pI
Pichia pastoris	**LS1**	14.5	8.6
	LS2	18.6	7.0
	PS1	14.5	9.1
	PS2	18.7	8.0
Escherichia coli	**LS1**	28.1	6.0
	LS2	32	5.9

All values presented here were calculated using the ExPASy ProtParam On-line Tool from SIB (Swiss Institute of Bioinformatics)[196].

Appendix 3

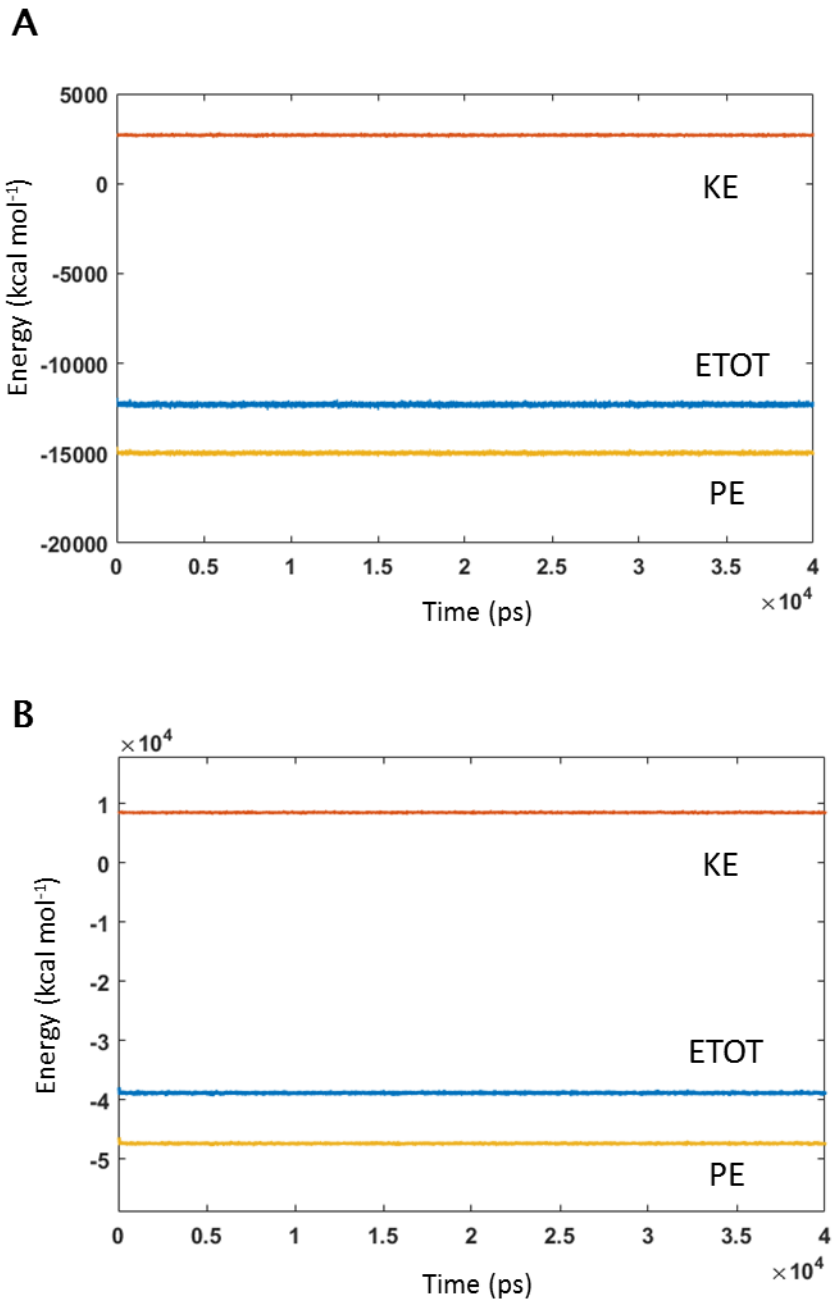
Temperature and energy plots for MD simulation on computationally produced sLe^x tetrasaccharide



Temperature (**A**) and energy (**B**) plots for sLe^x tetrasaccharide over a 40 ns MD simulation. Total energy (ETOT) is shown in blue, kinetic energy (KE) – in red, and potential energy (PE) – in yellow. It should be noted that these values correspond to the whole solvated system and are not accurate measures of the actual energy of the system.

Appendix 4

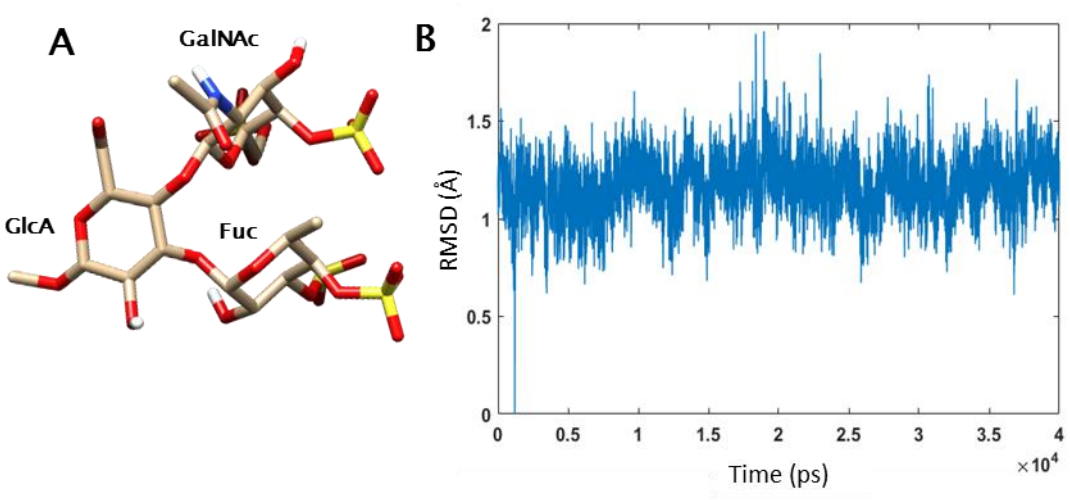
Sample energy plots for MD simulations on fCS oligosaccharides



Energy plots for fCS 2,4S trisaccharide (**A**) and nonasaccharide (**B**) over a 40 ns MD simulation. Total energy (ETOT) is shown in blue, kinetic energy (KE) – in red, and potential energy (PE) – in yellow. It should be noted that these values correspond to the whole solvated system and are not accurate measures of the actual energy of the system.

Appendix 5

Structure and MD analysis of fCS 3,4S trisaccharide



A. Lowest energy structure of the fCS 3,4S trisaccharide over a 40 ns MD simulation. **B.** Plot of overall RMSD for the fCS 3,4S trisaccharide with respect to the lowest energy structure.

Appendix 6

Average dihedral angles for fCS trisaccharide complexes

Glycosidic linkage	Dihedral angle	sLe^x	2,4S free	3,4S free	2,4S complex I	2,4S complex II	3,4S complex III
GalNAcβ1-4GlcA	**ϕ**	-70	-67	-66	-88	-106	-45
	ψ	130	129	130	-63	-72	129
Fucα1-3GlcA	**ϕ**	-79	-69	-72	-73	-88	-68
	ψ	-104	-94	-96	44	72	-96

Appendix 7

Pulse sequence of 2D ^{1}H, ^{13}C (3,2)D $BIRD^{r,X}$–HSQC experiment

```
;(3, 2)D BIRD(r,X)-HSQC
;based on Bruker hsqcetgpsisp2.2
;avance-version (12/01/11)
;Interleaved HSQC with additional cosine or sine modulation of 13C chemical shift
;by omega(1H)
;2D H-1/X correlation via double inept transfer
;   using sensitivity improvement
;phase sensitive using Echo/Antiecho-TPPI gradient selection
;with decoupling during acquisition
;using trim pulses in inept transfer
;using shaped pulses for all 180degree pulses on f2 - channel
;with gradients in back-inept
;
; N. Brodaczewska, Z. Košťálová, D. Uhrín

;A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn.
;   Reson. 93, 151-170 (1991)
;L.E. Kay, P. Keifer & T. Saarinen, J. Am. Chem. Soc. 114,
;   10663-5 (1992)
;J. Schleucher, M. Schwendinger, M. Sattler, P. Schmidt, O. Schedletzky,
;   S.J. Glaser, O.W. Sorensen & C. Griesinger, J. Biomol. NMR 4,
;   301-306 (1994)
;
;$CLASS=HighRes
;$DIM=2D
;$TYPE=
;$SUBTYPE=
;$COMMENT=
;$RECOMMEND=y

#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

"p2=p1*2"
"d4=1s/(cnst2*4)"
"d11=30m"
"d12=20u"

"d0=3u"
"d25=3u"

"in0=inf1/2"
"in25=cnst1*in0"

"l0=0"
```

```
"DELTA=p16+d16+p2+d0*2-4u"
"DELTA1=p16+d16-p1*0.78+de"
;"DELTA1=p16+d16-p1*0.78+de+4u"
"DELTA2=d4-larger(p2,p14)/2-4u"
"TAU1=d4-p16-d16-larger(p2,p14)/2-4u"
"TAU2=d4+p16+d16-larger(p2,p14)/2-4u+2*d25"
"DELTA3=d24-cnst17*p24/2-p19-d16-4u"
"DELTA4=d4-larger(p2,p14)/2-p16-d16-4u"
"DELTA5=d4*2-larger(p2,p14)/2-4u"


"acqt0=0"
baseopt_echo

1 ze
 d11 pl12:f2
2 d11 do:f2

# ifdef PRESAT
 4u BLKGRAD
 d12 pl9:f1
 d1 cw:f1 ph29
 d12 do:f1
 d12 pl1:f1
 50u UNBLKGRAD
# else
 4u BLKGRAD
 d1
 50u UNBLKGRAD
# endif


if "l0 %2 == 0"
   {
  (p1 ph2)         ; sin
   }
 else
   {
  (p1 ph1)         ; cos
   }

 d25
 p16:gp5
 d16

 (p1 ph1)     ;BIRD(r,X)
 DELTA5
 4u
 (center (p2 ph2) (p14:sp3 ph1):f2 )
 4u
 DELTA5
 (p1 ph1)
```

```
 p16:gp5*-1
 d16
 d25

 TAU1 pl0:f2
 4u
 (center (p2 ph1) (p14:sp3 ph6):f2 )
 4u
 TAU2 pl2:f2
 p28 ph1
 4u
 (p1 ph2) (p3 ph3):f2    ; transfer to 13C
 d0
 (p2 ph7)
 d0
 p16:gp1*EA
 d16 pl0:f2
 (p24:sp7 ph8:r):f2
 4u
 DELTA pl2:f2
 (center (p1 ph1) (p3 ph4):f2 )
 4u
 p19:gp3
 d16
 DELTA3 pl0:f2
 (center (p2 ph1) (p24:sp7 ph9:r):f2 )
 4u
 DELTA3 pl2:f2
 p19:gp3
 d16
 (center (p1 ph2) (p3 ph5):f2 )
 4u
 p16:gp4
 d16
 DELTA4 pl0:f2
 (center (p2 ph1) (p14:sp3 ph1):f2 )
 4u
 DELTA4
 p16:gp4
 d16
 (p1 ph1)
 DELTA1
 (p2 ph1)
 4u
 p16:gp2
 d16 pl12:f2
; 4u BLKGRAD
 go=2 ph31 cpd2:f2
 d11 do:f2 mc #0 to 2
   F1I(iu0, 2)
   F1EA(calgrad(EA) & calph(ph5, +180), caldel(d0, +in0) & caldel(d25, +in25) &
calph(ph3, +180) & calph(ph6, +180) & calph(ph31, +180))
 4u BLKGRAD
exit
```

```
ph1=0
ph2=1
ph3=0 2
ph4=0 0 2 2
ph5=1 1 3 3
ph6=0
ph7=0 0 2 2
ph8=0 0 2 2
ph9=0
ph29=0
ph31=0 2 2 0

;pl0 : 0W
;pl1 : f1 channel - power level for pulse (default)
;pl2 : f2 channel - power level for pulse (default)
;pl3 : f3 channel - power level for pulse (default)
;pl12: f2 channel - power level for CPD/BB decoupling
;sp3: f2 channel - shaped pulse (180degree inversion)
;spnam3: Crp60,0.5,20.1
;sp7: f2 channel - shaped pulse (180degree refocussing)
;spnam7: Crp60comp.4
;p1 : f1 channel -  90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p3 : f2 channel -  90 degree high power pulse
;p14: f2 channel - 180 degree shaped pulse for inversion
;    = 500usec for Crp60,0.5,20.1
;p16: homospoil/gradient pulse                    [1 msec]
;p19: gradient pulse 2                          [500 usec]
;p22: f3 channel - 180 degree high power pulse
;p24: f2 channel - 180 degree shaped pulse for refocussing
;    = 2msec for Crp60comp.4
;p28: f1 channel - trim pulse
;d0 : incremented delay (2D)                      [3 usec]
;d25 : incremented omega(1H) evolution delay  [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d4 : 1/(4J)XH
;d11: delay for disk I/O                         [30 msec]
;d16: delay for homospoil/gradient recovery
;d24: 1/(8J)XH for all multiplicities
;    1/(4J)XH for XH
;cnst2: = J(XH)
;cnst17: = -0.5 for Crp60comp.4
;inf1: 1/SW(X) = 2 * DW(X)
;in0: 1/(2 * SW(X)) = DW(X)
;in25 = cnst1*in0
;nd0: 2
;ns: 1 * n
;ds: >= 16
;td1: twice the usual number of experiments
;FnMODE: echo-antiecho
;cpd2: decoupling according to sequence defined by cpdprg2
;pcpd2: f2 channel - 90 degree pulse for decoupling sequence
```

```
;use gradient ratio:  gp 1 : gp 2 : gp 3 : gp 4
;                80 : 20.1 :   11 :   -5    for C-13
;                80 :  8.1 :   11 :   -5    for N-15

;for z-only gradients:
;gpz1: 80%
;gpz2: 20.1% for C-13, 8.1% for N-15
;gpz3: 11%
;gpz4: -5%
;gpz5: 7%

;use gradient files:
;gpnam1: SMSQ10.100
;gpnam2: SMSQ10.100
;gpnam3: SMSQ10.100
;gpnam4: SMSQ10.100
;gpnam5: SMSQ10.100


;cnst17: Factor to compensate for coupling evolution during a pulse
;       (usually +1). A positive factor indicates that coupling
;       evolution continues during the pulse, whereas a negative
;       factor is necessary if the coupling is (partially) refocussed.
```

Appendix 8

Pulse sequence of 2D ^{1}H, ^{13}C (3,2)D $BIRD^{r,X}$–HSQC – TOCSY experiment

```
;(3, 2)D BIRD(r,X)-HSQC-TOCSY
;Based on Bruker hsqcdietgpsisp.2
;avance-version (12/01/11)
;Interleaved HSQC-TOCSY with additional cosine or sine modulation of 13C chemical shift
;by omega(1H)
;2D H-1/X correlation via double inept transfer
;   using sensitivity improvement and DIPSI2
;   for homonuclear Hartman-Hahn mixing
;phase sensitive using Echo/Antiecho-TPPI gradient selection
;with decoupling during acquisition - using f2
;using trim pulses in inept transfer
;using shaped pulses for all 180degree pulses on f2 – channel

; N. Brodaczewska, Z. Košťálová, D. Uhrín

;
;$CLASS=HighRes
;$DIM=2D
;$TYPE=
;$SUBTYPE=
;$COMMENT=
;$RECOMMEND=y

#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

"p2=p1*2"
"d4=1s/(cnst2*4)"
"d11=30m"
"d12=20u"

"d0=3u"
"d25=3u"

"in0=inf1/2"
"in25=cnst1*in0"

"l0=0"

"FACTOR1=(d9/(p6*115.112))/2"
"l1=FACTOR1*2"
"d12=20u"

"DELTA=p16+d16+p2+d0*2-4u"
"DELTA1=p16+d16-p1*0.78+de"
```

```
;"DELTA1=p16+d16-p1*0.78+de+4u"
"DELTA2=d4-larger(p2,p14)/2-4u"
"TAU1=d4-p16-d16-larger(p2,p14)/2-4u"
"TAU2=d4+p16+d16-larger(p2,p14)/2-4u+2*d25"
"DELTA3=d24-cnst17*p24/2-4u"
"DELTA4=d4-larger(p2,p14)/2-4u-2*p3-6u-p16-d16"
"DELTA5=d4*2-larger(p2,p14)/2-4u"
"DELTA6=d4-larger(p2,p14)/2-4u-p1*0.78+de"


"acqt0=0"
baseopt_echo

1 ze
 d11 pl12:f2
2 d11 do:f2

#ifdef PRESAT
 4u BLKGRAD
 d12 pl9:f1
 d1 cw:f1 ph29
 d12 do:f1
 d12 pl1:f1
 50u UNBLKGRAD
#else
 4u BLKGRAD
 d1
 50u UNBLKGRAD
#endif


if "l0 %2 == 0"
   {
  (p1 ph2)    ; sin
   }
 else
   {
  (p1 ph1)    ; cos
   }

 d25
 p16:gp5
 d16 pl0:f2

 (p1 ph1)     ;BIRD(r,X)
 DELTA5
 4u
 (center (p2 ph2) (p14:sp3 ph1):f2 )
 4u
 DELTA5
 (p1 ph1)

 p16:gp5*-1
 d16
```

```
d25

TAU1
4u
(center (p2 ph1) (p14:sp3 ph6):f2 )
4u
TAU2 pl2:f2
p28 ph1
4u
(p1 ph2)

(p3 ph3):f2  ; transfer to 13C
d0
(p2 ph7)
d0
p16:gp1*EA
d16 pl0:f2
(p24:sp7 ph4):f2
4u
DELTA pl2:f2
(center (p1 ph1) (p3 ph4):f2 )
4u
DELTA3 pl0:f2
(center (p2 ph1) (p24:sp7 ph1):f2 )
4u
DELTA3 pl2:f2
(center (p1 ph2) (p3 ph5):f2 )
4u
DELTA2 pl0:f2
(center (p2 ph1) (p14:sp3 ph1):f2 )
4u
DELTA2 pl10:f1

                                                    ;begin DIPSI2
4 p6*3.556 ph22
  p6*4.556 ph24
  p6*3.222 ph22
  p6*3.167 ph24
  p6*0.333 ph22
  p6*2.722 ph24
  p6*4.167 ph22
  p6*2.944 ph24
  p6*4.111 ph22

  p6*3.556 ph24
  p6*4.556 ph22
  p6*3.222 ph24
  p6*3.167 ph22
  p6*0.333 ph24
  p6*2.722 ph22
  p6*4.167 ph24
  p6*2.944 ph22
  p6*4.111 ph24
```

```
  p6*3.556 ph24
  p6*4.556 ph22
  p6*3.222 ph24
  p6*3.167 ph22
  p6*0.333 ph24
  p6*2.722 ph22
  p6*4.167 ph24
  p6*2.944 ph22
  p6*4.111 ph24

  p6*3.556 ph22
  p6*4.556 ph24
  p6*3.222 ph22
  p6*3.167 ph24
  p6*0.333 ph22
  p6*2.722 ph24
  p6*4.167 ph22
  p6*2.944 ph24
  p6*4.111 ph22
  lo to 4 times l1
                                                    ;end DIPSI2

  4u pl1:f1
  (p1 ph1)

  DELTA1
  (p2 ph1)
  4u
  p16:gp2
  d16 pl12:f2
; 4u BLKGRAD

  go=2 ph31 cpd2:f2
  d11 do:f2 mc #0 to 2
    F1I(iu0, 2)
    F1EA(calgrad(EA) & calph(ph5, +180), caldel(d0, +in0) & caldel(d25, +in25) &
calph(ph3, +180) & calph(ph6, +180) & calph(ph31, +180))
exit


ph1=0
ph2=1
ph3=0 2
ph4=0 0 2 2
ph5=1 1 3 3
ph6=0
ph7=0 0 2 2
ph22=3
ph24=1
ph29=0
ph31=0 2 2 0


;pl0 : 0W
```

```
;pl1 : f1 channel - power level for pulse (default)
;pl2 : f2 channel - power level for pulse (default)
;pl10: f1 channel - power level for TOCSY-spinlock
;pl12: f2 channel - power level for CPD/BB decoupling
;sp3: f2 channel - shaped pulse (180degree inversion)
;spnam3: Crp60,0.5,20.1
;sp7: f2 channel - shaped pulse (180degree refocussing)
;spnam7: Crp60comp.4
;p1 : f1 channel -  90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p3 : f2 channel -  90 degree high power pulse
;p6 : f1 channel -  90 degree low power pulse
;p14: f2 channel - 180 degree shaped pulse for inversion
;    = 500usec for Crp60,0.5,20.1
;p16: homospoil/gradient pulse                    [1 msec]
;p24: f2 channel - 180 degree shaped pulse for refocussing
;    = 2msec for Crp60comp.4
;p28: f1 channel - trim pulse                     [1 msec]
;d0 : incremented delay (2D)                      [3 usec]
;d25 : incremented omega(1H) evolution delay  [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d4 : 1/(4J(XH))
;d9 : TOCSY mixing time
;d11: delay for disk I/O                          [30 msec]
;d16: delay for homospoil/gradient recovery
;d24: 1/(8J)XH for all multiplicities
;    1/(4J)XH for XH
;cnst1: scaling factor for omega(1H) relative to omega(13C)
;cnst2: = J(XH)
;cnst17: = -0.5 for Crp60comp.4
;l1: loop for DIPSI cycle: ((p6*115.112) * l1) = mixing time
;inf1: 1/SW(X) = 2 * DW(X)
;in0: 1/(2 * SW(X)) = DW(X)
;in25 = cnst1*in0
;nd0: 2
;ns: 4 * n
;ds: >= 16
;td1: twice the usual number of experiments
;FnMODE: echo-antiecho
;cpd2: decoupling according to sequence defined by cpdprg2
;pcpd2: f2 channel - 90 degree pulse for decoupling sequence

;use gradient ratio:        gp 1 : gp 2
;                            80 : 20.1      for C-13
;                            80 :  8.1      for N-15

;for z-only gradients:
;gpz1: 80%
;gpz2: 20.1% for C-13, 8.1% for N-15
;gpz5: 7%

;use gradient files:
;gpnam1: SMSQ10.100
;gpnam2: SMSQ10.100
```

```
;gpnam5: SMSQ10.100

;cnst17: Factor to compensate for coupling evolution during a pulse
;       (usually +1). A positive factor indicates that coupling
;       evolution continues during the pulse, whereas a negative
;       factor is necessary if the coupling is (partially) refocussed.
```

Appendix 9

Pulse sequence of 2D ^{1}H, ^{13}C (3,2)D BIRD^{r,X}–HSQC (A) and 2D ^{1}H, ^{13}C (3,2)D BIRD^{r,X}–HSQC–TOCSY (B) experiments